On the quality of diagnostic hospital discharge data for medical practice assessment: an experiment in a pediatric department

Prins, H.

Hilco Prins

On the Quality of Diagnostic Hospital Discharge Data for Medical Practice Assessment

An Experiment in a Pediatric Department

On the Quality of Diagnostic Hospital Discharge Data for Medical Practice Assessment

Hilco Prins

On the Quality of Diagnostic Hospital Discharge Data for Medical Practice Assessment

An Experiment in a Pediatric Department

On the Quality of Diagnostic Hospital Discharge Data for Medical Practice Assessment; an Experiment in a Pediatric Department
PhD thesis, University of Amsterdam, The Netherlands
ISBN: 978-94-6108-344-9

Cover design: Tijmen, Hielke, Jelmer en Emma Prins, Martin van Wijngaarden
Print: Gildeprint Drukkerijen

The printing of this thesis is supported by a grant from Stichting Bazis. Also the time given by Windesheim University of Applied Sciences and its Departments of Nursing and ICT-Innovations in Healthcare, and the support given by the Department of Medical Informatics at the Academic Medical Center of the University of Amsterdam in order to work on this thesis is gratefully acknowledged.

On the Quality of Diagnostic Hospital Discharge Data for Medical Practice Assessment

An Experiment in a Pediatric Department

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. D.C. van den Boom
ten overstaan van een door het college voor promoties ingestelde
commissie, in het openbaar te verdedigen in de Aula der Universiteit
op vrijdag 2 november 2012, te 11:00 uur

door

Hilbert Prins

geboren te Hoogeveen

Promotiecommissie:

Promotores:	Prof. dr. ir. A. Hasman
	Prof. dr. H.A. Büller
Overige leden:	*Prof. dr. A. Abu-Hanna*
	Dr. W.T.F. Goossen
	Dr. J.B. Reitsma
	Prof. dr. R.J.P.M. Scholten
	Prof. dr. F.A. Wijburg

Faculteit der Geneeskunde

Voor

mijn Ouders
Maaike
Tijmen, Hielke, Jelmer en Emma

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

This chapter is the introduction to this thesis. Paragraph 1.1 describes the context and background of the study. Paragraph 1.2 provides the problem, research questions and outline of the thesis.

1.1 DOMAIN

1.1.1 Era of Assessment and Accountability

In the Western World, after an era of expansion in the 1950s and 1960s, and an era of cost containment in the 1970s and 1980s, medical care entered a new era in the 1990s: the era of assessment and accountability [1]. Cost containment instruments alone, like price-policy or budgeting, did not lead to the intended cost control. As a consequence of the ageing population and the continuous development of new medical technologies, the volume and costs of medical services are still increasing. In many countries the growth of costs in health care exceeds the growth of the economy so that an increasing part of their gross national product is spent on health care [2].

The discovery that some medical services are not appropriately used or have no positive effect on the health status of patients [3], and the discovery of the existence of variability in many medical services without differences in outcome [4-6], led to activities to distinguish medical services which are effective, efficient and safe from the other ones. Randomized controlled trials [7], medical technology assessment [8] and systematic reviews and meta-analyses of clinical studies [9, 10] are examples of these activities. Further, the development of practice guidelines, preferably based on systematically collected evidence and patients' preferences [11, 12], can bring the knowledge about effectiveness and efficiency of medical practices to the physician. A practice guideline is a *"systematically developed statement to assist practitioner and patient decisions about appropriate health care for specific circumstances"* [13]. These statements *"provide an intellectual vehicle through which the profession can distill the lessons of research and clinical experiences and pool the knowledge and preferences of many people into conclusions about appropriate practice."* [14] However, providing guidelines alone is not enough. Physicians have also to comply with this knowledge, leading to evidence based practice [15] and better patient safety [16]. Nowadays, there is an urge to follow practice guidelines when indicated. When the specific circumstances are met, one has to comply with the guideline; only well-founded deviations from the guideline

are permitted. Unfortunately, where applicable, this knowledge is not always put into practice. For example, implementation of practice guidelines appears only moderately successful [17-19]. Grimshaw et al. [20, 21] conclude that guidelines do improve medical practice, but only when they are introduced under rigorous evaluations.

Doubts about the effectiveness and efficiency of daily medical practice and the large attention of the mass media for medical errors have decreased the trust of society in the medical profession. In reaction to the one-sided attention for costs, a counter movement asked more attention for the quality of care [22, 23]. Therefore it is more and more expected that physicians account for their activities [24]. Questions as: which services have been provided, what was the quality of the services and what has been done to assess and assure the quality of the services, have to be answered by health care institutions and physicians. These questions are particularly asked by three parties concerned: 1) the governments and 2) third-party payers who both consider themselves as patrons of patients, premium payers and taxpayers, and 3) the patients themselves who increasingly show signs of having a professed opinion and who are more and more unionized. These questions fit within the framework of quality assessment and quality assurance. Systematic, retrospective assessment of daily medical practice offers possibilities to answer these questions on quality of care and allows physicians to account for their medical practice.

1.1.2 Quality Assessment and Quality Assurance

The Institute of Medicine [13] defines health care quality as *"the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge"*. Gemke [25] defines quality assessment as *"the critical appraisal of the measured results of a health care program, in comparison with formulated objectives"*. These objectives can be formulated by means of standards of quality which are the results of clinical scientific research or consensus of experts [26, 27]. Standards of quality are authoritative statements concerning 1) minimum levels of acceptable performance or results, 2) excellent levels of performance or results, or 3) the range of acceptable levels of performance or results [13].

Quality assurance includes more: it demands that action is taken to protect, maintain or improve performance, dependent on the critical appraisal of the

measured results. In this respect Lomas [28] defined quality assurance as *"the measurement of health care activity, and the outcomes of that activity, in order to identify whether the objectives of that activity are being achieved and, when this is not the case, to respond with effective action to reduce the deviations from the objectives"*.

Starting from this definition three principal activities of quality assurance can be distinguished: measurement of health care activities and/or outcomes of those activities, comparison of these measured activities or outcomes with standards, and responses to proposed changes. These three activities are performed in a cyclic manner. An important means to measure health care activities and outcomes is the use of performance indicators [29]. A performance indicator is a systematically developed quantitative measure that can be used to assess and improve health care activities and outcomes for which standards are set [30].

Practice guidelines, performance indicators and standards of quality are strongly related. From a practice guideline performance indicators can be derived. In order to assess medical practice, for aspects of care which are measured by performance indicators, standards of quality have to be agreed upon.

Dependent on the results of the assessment, actions for change have to be undertaken oriented towards patients, the physicians and / or the health care system within which the care is given. The assessment also can lead to changes in the guidelines, indicators or standards, especially when these are based on expert opinion or consensus in case there is no scientific evidence available. Performance indicators or standards have to be adapted if one's own patient population deviates from average.

1.1.3 Structure, Process and Outcome Indicators

Donabedian [26] classifies variables related to quality of care, into three categories: structure, process and outcome. For variables in each of these three categories performance indicators can be developed. Structure indicators measure attributes of the setting in which care takes place. An example is the number of nurses divided by the number of beds of an orthopedic department. Process indicators measure activities performed during the course of patient care. An example is the percentage of patients starting mobilization the day after a hip replacement. Outcome indicators measure the effect of care on the health status of the patient. The percentage of patients that can walk independently without pain one month

after a hip replacement is an example of an outcome indicator. A good structure increases the likelihood of a good process, and a good process increases the likelihood of a good outcome.

1.1.4 Professionalization

Besides accountability, Klazinga [31] mentions another important reason for medical practice assessment: professionalization. As a professional, a physician is interested in the quality of his or her work and in ways to improve it further. Assessing the work and learn from it is a means to improve the quality [26, 32].

The most important aims of assessment are threefold:

1. Prevention. Knowing that one's own practice is subject to assessment will be an extra stimulus to act carefully and to act according to the newest insights in what can be considered as good clinical practice. If performance indicators are derived from practice guidelines, the assessment will be an incentive to act according to the guidelines [31, 33];
2. Education [34]. If the practice of an individual physician or group of physicians incorrectly deviates from guidelines, it is important to recognize this in order to change practice for the better in future. Physicians can learn from their own mistakes, but clearly it would be beneficial when physicians could also learn from the mistakes of others;
3. Incentive. It is possible that structure variables of the clinical environment make it difficult for physicians to act according to a practice guideline. If this is observed it should be an incentive to change structure [35].

By being preventive, educational and incentive, assessment will lead to quality improvement of medical practice.

1.1.5 Assessing medical practice in hospitals

The assessment of medical practice in hospitals can be done at three levels.

At the first level, cases of individual patients are analyzed. The emphasis lies on an elaborate analysis of activities performed by individual physicians for one or only a few patients in case there is a suspicion that something seriously went wrong. When the patients are known, computerized patient records can be used for the

selection of cases. In order to analyze the case(s), the additional use of the paper medical record will usually be necessary.

At the second level, groups of patients are analyzed that have one or more clinically relevant and important attributes in common. The attributes are related to a disease and/or a therapy, e.g. “admitted with suspected meningitis”, “having acute lymphatic leukemia” or “underwent a gastroscopy”. Because of the expected similarities of diagnostic and therapeutic activities for these patients, it is possible to follow the care process of the group as a whole from admission till discharge, supplemented with outpatients’ follow-up. The emphasis lies on medical activities instantiated by individual physicians or a group of physicians. By means of process criteria and related standards the congruence between performance and practice guidelines can be measured. Because this type of assessment doesn’t need many cases it can be applied locally, and since it gives insight in what physicians do (or let do) compared to what is desired, it is an excellent educational tool [(36)]. Often also some outcome criteria are used, but only to detect big deviations from the quality standards. Most of the time there is only one or a limited number of specialties involved. In order to make this kind of analysis feasible, it is highly desirable to be able to select and analyze the cases on the basis of electronically recorded patient data, such as diagnoses, procedures and test results.

Finally care is analyzed at the level of (a department of) the hospital. The goal is not to follow a group of patients from admission till discharge, but to analyze a specific activity or event within a specific time interval, especially how often it is done or has happened compared to other time periods or other hospitals. The issue at stake can be quality but also only costs. Examples are the number of X-rays or total hip replacements performed, the number of nosocomial infections, percentage of re-admissions and mortality. Emphasis lies on outcome measurements. It is very well possible that the selected cases are very diverse with regard to their disease related attributes and that many specialties are involved. In that case case-mix adjustment is necessary for a valid comparison over time or between hospitals. Many cases are needed and adequate statistical techniques are required to make valid inferences. Differences in outcome over time or between hospitals give rise to further research which can lead, dependent on the results, to new practice policies. However, since in this type of assessment the performed activities are considered as a black box, it is often difficult to discover the reasons for differences in outcome [(36)]. Probably in this case assessment at the second level will be helpful.

Selecting and analyzing cases at this third level is only feasible when patient data are electronically available.

Each of the three types of assessment has its own advantages and disadvantages. The disadvantages of one may be compensated by the advantages of another type. Therefore, the three types of assessment should not be considered as competitors but as complementary to each other. It is important to launch the three types of assessment in such a way that with limited effort a maximum effect on quality of care can be attained.

1.1.6 Assessment in the Netherlands

Also Dutch hospitals and medical specialists highly value assessment of their medical practice [31]. The attention for quality has considerably increased during the last decade by the establishment of the Quality of Health Care Institutions Act [37], which obliges health care organizations to develop quality systems. The introduction of market elements in the Dutch health care system has also contributed to the increased attention to quality. In addition, the Scientific Board for Government Policy [38] explicitly recommended as a goal for the next decades the improvement of the effectiveness and efficiency of medical (and other health care) practices performed on individual patients. It is considered important to warrant quality and accessibility of health care for everyone. Therefore, testing the effectiveness and efficiency of medical practice and the translation of these results into practice guidelines is considered very important. With regard to the medical specialists it is expected that they themselves will develop practice guidelines, which will constitute the basis for quality assessment.

1.1.7 Data quality

At each level it is necessary, in order to assess medical practice, to have high-quality data about the patients, their disease-related attributes, activities performed or initiated by physicians and patient events [26, 39, 40].

When the data, needed for this purpose, are recorded in a computerized system correctly, completely, with enough detail, timely, standardized and according to an adequate data model, the analysis can be performed completely and efficiently. However, not all data of the care process are recorded electronically. Unfortunately, in many hospital information systems, data about history taking and physical examination are still lacking [41], also in the Netherlands. Furthermore,

when recorded electronically, some of the data are not recorded completely, correctly, with enough detail or timely. The data quality of test results can be regarded as good as these data are used for daily patient care and usually are reported electronically. Major procedures that are usually performed in operating rooms are reasonably well-coded; minor procedures that are routinely performed on wards or in radiology departments are generally under-coded [42]. The quality of diagnostic hospital discharge data stands in bad repute among physicians and health care researchers. The discharge data concerning the diagnoses play no role in daily patient care and this is possibly the main reason why there are doubts about the reliability of the registration [43-57]. Furthermore, patient data are stored in several subsystems. Unreliable and fragmented data could hamper the use of patient data for medical practice assessment [40, 58-60].

However, the hospital discharge data registry is the only registry of diagnostic data so far that cover all hospitalizations. This complete coverage is a major advantage for the use of the data across several specialties, patient groups and hospitals.

1.2. SCOPE

1.2.1 Object of study

In this thesis the object of study is the use of routinely collected and electronically recorded patient data for the assessment by medical specialists themselves of their medical practice for specific, clinically defined patient groups. The study is limited to pure medical considerations concerning medical practice. Other important factors for quality of care such as attitude, communication skills and patient satisfaction, are not taken into account.

By taking the lead in the assessment of their care, the specialists can keep quality assurance activities in their own hands, especially when they make this transparent. Besides, to improve one's own quality is an important characteristic of professionals. For these reasons, assessment of their medical practice for clinically defined patient groups (the second level mentioned in § 1.1.5) can be very attractive for medical specialists. This assessment focuses on aspects of care that correspond with the way physicians reason and act and that can be influenced by the physician's choices. The specialists are closely involved with all the patients in these groups and assessment at this level provides a more systematic evaluation of clinical care than case reviews. Another important advantage of assessment at this

level is that practice guidelines are also developed for the same -clinically defined- patient groups.

1.2.2 Problem description

It is not clear beforehand what physicians want to know about their own medical practice in order to evaluate it. Therefore, it is also not clear which patient data are needed for medical practice assessment. The electronic availability and usability of data can only be determined when the information needs of physicians are known.

In medical practice assessment of specific, clinically defined patient groups, diagnostic data play an important role because process and outcome indicators are often disease specific. For case selection as well as process and outcome measures several forms of diagnoses are important. Patient groups are often defined by diagnoses which implies that patient cases should be selected based on their diagnostic data. Data about complications, which form a special type of diagnoses, can be used to get insight in important outcome indicators. For the interpretation of outcome indicators, insight in comorbidities, also a special form of diagnoses, can be necessary. Since the pediatricians at the Academic Medical Center in Amsterdam, the Netherlands had serious doubts about the reliability of diagnostic data, we were especially interested in the quality of diagnostic data and sought ways to increase the reliability of the data.

1.2.3 Aim of the study

In order explore the possibilities of medical practice assessment using electronically available patient data, a research project was started by the Department of Medical Informatics and the Department of Pediatrics at the Academic Medical Center, Amsterdam.

The aim of the study was fivefold:

1. To get insight in the information needs of the physicians for the assessment of their medical practice of specific patient groups;
2. To test whether patient data needed for medical practice assessment are electronically available and usable;
3. Since, as mentioned above, it was expected that the data quality of the discharge registry was not optimal, it was hypothesized that increased influence of the physician would be beneficial and lead to a better

diagnosis registry. Therefore the third aim of the study was: To find a way to incorporate a diagnosis registry into the clinical care process;

4. To test whether incorporating the diagnosis registry into the clinical care process improves diagnostic data quality;
5. To see whether the results of our study correspond to those published in the literature. Therefore the fifth aim of the study was: To get insight, based on a systematic review of the literature, in diagnostic data quality and in the factors that influence this data quality.

1.2.4 Main research questions

We performed five studies. The research questions for each study are presented below.

Study 1

In this case study we investigated which performance indicators are needed for the assessment of the medical practice of children with suspected or proven meningitis who were not premature neonates or patients with cancer. In this study we analyzed the availability of those data needed to determine the value of the performance indicators and the usability (defined as availability of complete and accurate data in a standardized form) of electronically recorded patient data to automatically determine the values of the performance indicators. We were interested in the following:

1. Which performance indicators, case-mix and exploratory information are needed by physicians for medical practice assessment?
2. Are the required data electronically available and usable for medical practice assessment?

Study 2

In this study we describe the redesign of the process of diagnostic coding used by a pediatric department. The goal was to improve the completeness and accuracy of the diagnostic data. We addressed the following questions:

1. How can the diagnostic discharge registration be incorporated into the care process?

2. What is the effect of the physicians' involvement on the quality of diagnosis coding?

Study 3

In study 3 we studied the quality of the redesigned diagnostic coding process in more detail. The research question was:

1. Would physician coding and the integration of the diagnosis registration with the communication process, improve completeness, correctness, specificity and timeliness of diagnostic data?

Study 4

In this study we investigated the influence of physician involvement in diagnosis encoding in the long run. Research questions were:

1. Are diagnoses encoded more specifically?
2. Does the number of coded diagnoses increase?
3. Are any effects sustainable over time?

Study 5

In this systematic review we investigated the quality of diagnostic inpatient hospital discharge data as reported in scientific journals in order to examine whether the results of our study correspond to those published in the literature. We investigated:

1. Which gold standards and designs were used to assess data quality?
2. What completeness and correctness values were reported?
3. Which factors influence the data quality of studies?
4. What are determinants of data quality reported in studies?
5. What is the evidence about the consequences of data quality for medical practice assessment?
6. Are diagnostic data appropriate for quality of care purposes?

1.3 OUTLINE OF THIS THESIS

In chapter 2 we analyze availability and quality of patient data in the hospital information system of the AMC, for the assessment of medical practice concerning children with suspected meningitis.

In chapter 3 we describe a project with the goal on the one hand to improve the accuracy of the diagnosis registration and on the other hand to accelerate discharge letter writing. This chapter describes the redesign process of the form-based encoding by the medical record coder, by involving pediatricians, and by developing a new discharge-letter linked encoding procedure. Furthermore the coding performance of pediatricians in the new situation is evaluated.

In chapter 4 we tested our hypothesis that integration of the diagnosis registration into the communication process with GPs combined with physician encoding improves completeness, correctness, specificity and timeliness of diagnostic data.

Chapter 5 describes a time series study covering twelve consecutive years. In the first four years, the usual form-based encoding by the medical record coder was in use and in the last eight years, the discharge letter-linked encoding by pediatricians.

Chapter 6 is a systematic review investigating the quality of diagnostic inpatient hospital discharge data as reported in scientific literature. The question to be answered was whether the quality of the diagnostic data increased as a function of time and which factors influenced the quality.

In chapter 7 we discuss the findings of this thesis.

REFERENCES

1. Relman AS. Assessment and accountability: the third revolution in medical care [editorial]. New England Journal of Medicine. 1988;319(18):1220-2.
2. Adang EM, Ament A, Dirksen CD. Medical technology assessment and the role of economic evaluation in health care. Journal of Evaluation in Clinical Practice. 1996;2(4):287-94.
3. Leape LL. Unnecessary surgery. Health Services Research. 1989;24(3):351-407.
4. Blumenthal D. The variation phenomenon in 1994 [editorial; comment]. New England Journal of Medicine. 1994;331(15):1017-8.
5. Detsky AS. Regional variation in medical care [editorial; comment]. New England Journal of Medicine. 1995;333(9):589-90.

6. Vayda E. A comparison of surgical rates in Canada and in England and Wales. New England Journal of Medicine. 1973;289(23):1224-9.

7. Greenfield S, Kravitz R, Duan N, Kaplan SH. Heterogeneity of treatment effects: implications for guidelines, payment, and quality assessment. Am J Med. 2007 Apr;120(4 Suppl 1):S3-9.

8. Fuchs VR, Garber AM. The new technology assessment [published erratum appears in N Engl J Med 1991 Jan 10;324(10):136] [see comments]. New England Journal of Medicine. 1990;323(10):673-7.

9. Chalmers I, Haynes B. Reporting, updating, and correcting systematic reviews of the effects of health care [see comments]. Bmj. 1994;309(6958):862-5.

10. Eysenck HJ. Meta-analysis and its problems. Bmj. 1994;309(6957):789-92.

11. Eddy DM. Clinical decision making: from theory to practice. Rationing by patient choice. Jama. 1991;265(1):105-8.

12. Eddy DM. Clinical decision making: from theory to practice. Guidelines for policy statements: the explicit approach. Jama. 1990;263(16):2239-40, 43.

13. Institute of Medicine (Field M, Lohr, KN. eds). Guidelines for clinical practice: from development to use. Washington DC: National Academic Press; 1992.

14. Eddy DM. Clinical decision making: from theory to practice. Practice policies --what are they? [see comments]. Jama. 1990;263(6):877-8, 80.

15. Chapman NH, Lazar SP, Fry M, Lassere MN, Chong BH. Clinicians adopting evidence based guidelines: a case study with thromboprophylaxis. BMC Health Serv Res. 2011 [Epub];11:240.

16. Arah OA, Klazinga NS. How safe is the safety paradigm? Qual Saf Health Care. 2004 Jun;13(3):226-32.

17. Delamothe T. Wanted: guidelines that doctors will follow [editorial]. Bmj. 1993;307(6898):218.

18. Latoszek-Berendsen A, Tange H, van den Herik HJ, Hasman A. From clinical practice guidelines to computer-interpretable guidelines. A literature overview. Methods Inf Med. 2010;49(6):550-70.

19. Gagliardi AR, Brouwers MC, Palda VA, Lemieux-Charles L, Grimshaw JM. How can we improve guideline use? A conceptual framework of implementability. Implement Sci. 2011;6:26.

20. Grimshaw JM, Russell IT. Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations [see comments]. Lancet. 1993;342(8883):1317-22.

21. Grimshaw J, Eccles M, Russell I. Developing clinically valid practice guidelines. Journal of Evaluation in Clinical Practice. 1995;1(1):37-48.

22. Angell M, Kassirer JP. Quality and the medical marketplace--following elephants [editorial; comment] [see comments]. New England Journal of Medicine. 1996;335(12):883-5.

23. Chassin MR. Quality of health care. Part 3: improving the quality of care. New England Journal of Medicine. 1996;335(14):1060-3.

24. Blumenthal D, Epstein AM. Quality of health care. Part 6: The role of physicians in the future of quality management [see comments]. New England Journal of Medicine. 1996;335(17):1328-31.

25. Gemke RJBJ. Outcome assessment of pediatric intensive care: principles and applications. Utrecht: University Utrecht; 1994.

26. Donabedian A. The quality of care. How can it be assessed? Jama. 1988;260(12):1743-8.

27. Donabedian A. The role of outcomes in quality assessment and assurance [see comments]. Qrb Quality Review Bulletin. 1992;18(11):356-60.

28. Lomas J. Quality assurance and effectiveness in health care: an overview [editorial]. Quality Assurance in Health Care. 1990;2(1):5-12.

29. Bernstein SJ, Hilborne LH. Clinical indicators: the road to quality care? Joint Commission Journal on Quality Improvement. 1993;19(11):501-9.

30. Prins H, Kruisinga FH, Buller HA, Zwetsloot-Schonk JH. Availability and usability of data for medical practice assessment. Int J Qual Health Care. 2002 Apr;14(2):127-37.

31. Klazinga NS. Quality management of medical specialist care in The Netherlands. Rotterdam: Erasmus University; 1996.

32. Kritchevsky SB, Simmons BP. Continuous quality improvement. Concepts and applications for physician care [see comments]. Jama. 1991;266(13):1817-23.

33. Casparie AF. Guidelines for medical care; the relationship between medical decision making, technology assessment and quality assurance [editorial]. Netherlands Journal of Medicine. 1988;33(1-2):1-4.

34. McIntyre N, Popper K. The critical attitude in medicine: the need for a new ethics. British Medical Journal Clinical Research Ed. 1983;287(6409):1919-23.

35. Goud R, van Engen-Verheul M, de Keizer NF, Bal R, Hasman A, Hellemans IM, et al. The effect of computerized decision support on barriers to guideline implementation: a qualitative study in outpatient cardiac rehabilitation. Int J Med Inform. 2010 Jun;79(6):430-7.

36. McIntyre N. Evaluation in clinical practice: problems, precedents and principles. Journal of Evaluation in Clinical Practice. 1995;1(1):5-13.

37. Ministry of Public Health WaS. The Quality of Health Care Institutions Act. The Hague: SDU Uitgeverij; 1996.

38. Scientific Board for Government Policy. Public Health 52; Reports to the government. The Hague: SDU Uitgeverij; 1997.

39. Wyatt JC. Clinical data systems, Part 1: Data and medical records [see comments]. Lancet. 1994;344(8936):1543-7.

40. Wyatt J. Acquisition and use of clinical data for audit and research. Journal of Evaluation in Clinical Practice. 1995;1(1):15-27.

41. Jha AK, DesRoches CM, Campbell EG, Donelan K, Rao SR, Ferris TG, et al. Use of electronic health records in U.S. hospitals. N Engl J Med. 2009 Apr 16;360(16):1628-38.

42. Quan H, Parsons GA, Ghali WA. Validity of procedure codes in International Classification of Diseases, 9th revision, clinical modification administrative data. Med Care. 2004 Aug;42(8):801-9.

43. Hohnloser JH, Kadlec P, Puerner F. Coding clinical information: analysis of clinicians using computerized coding. Methods of Information in Medicine. 1996;35(2):104-7.

44. Colin C, Ecochard R, Delahaye F, Landrivon G, Messy P, Morgon E, et al. Data quality in a DRG-based information system. International Journal for Quality in Health Care. 1994;6(3):275-80.

45. Lu JH, Lin FM, Shen WY, Chen SJ, Hwang BT, Wu SI, et al. Data quality of a computerized medical birth registry. Medical Informatics. 1994;19(4):323-30.

46. Mosbech J, Jorgensen J, Madsen M, Rostgaard K, Thornberg K, Poulsen TD. [The national patient registry. Evaluation of data quality]. Ugeskrift for Laeger. 1995;157(26):3741-5.

47. Weissman C. Can hospital discharge diagnoses be used for intensive care unit administrative and quality management functions? [see comments]. Critical Care Medicine. 1997;25(8):1320-3.

48. Hasan M, Meara RJ, Bhowmick BK. The quality of diagnostic coding in cerebrovascular disease. International Journal for Quality in Health Care. 1995;7(4):407-10.

49. Chewning SJ, Nussman DS, Griffo ML, Kiebzak GM. Health care information processing: how accurate are the data? Journal of the Southern Orthopaedic Association. 1997;6(1):8-16.

50. Davenport RJ, Dennis MS, Warlow CP. The accuracy of Scottish Morbidity Record (SMR1) data for identifying hospitalised stroke patients. Health Bulletin. 1996;54(5):402-5.

51. Leibson CL, Naessens JM, Brown RD, Whisnant JP. Accuracy of hospital discharge abstracts for identifying stroke. Stroke. 1994;25(12):2348-55.

52. Chancellor AM, Swingler RJ, Fraser H, Clarke JA, Warlow CP. Utility of Scottish morbidity and mortality data for epidemiological studies of motor neuron disease. Journal of Epidemiology & Community Health. 1993;47(2):116-20.

53. MacIntyre CR, Ackland MJ, Chandraraj EJ, Pilla JE. Accuracy of ICD-9-CM codes in hospital morbidity data, Victoria: implications for public health research. Australian & New Zealand Journal of Public Health. 1997;21(5):477-82.

54. Hobbs FD, Parle JV, Kenkre JE. Accuracy of routinely collected clinical data on acute medical admissions to one hospital. British Journal of General Practice. 1997;47(420):439-40.

55. Beattie TF. Accuracy of ICD-9 coding with regard to childhood accidents. Health Bulletin. 1995;53(6):395-7.

56. Benesch C, Witter DM, Jr., Wilder AL, Duncan PW, Samsa GP, Matchar DB. Inaccuracy of the International Classification of Diseases (ICD-9-CM) in identifying the diagnosis of ischemic cerebrovascular disease. Neurology. 1997;49(3):660-4.

57. Geraci JM, Ashton CM, Kuykendall DH, Johnson ML, Wu L. International Classification of Diseases, 9th Revision, Clinical Modification codes in discharge abstracts are poor measures of complication occurrence in medical inpatients. Medical Care. 1997;35(6):589-602.

58. Iezzoni LI. Assessing quality using administrative data. Annals of Internal Medicine. 1997;127(8 (Pt 2)):666-74.

59. McKee M. Routine data: a resource for clinical audit? Quality in Health Care. 1993(2):104-11.

60. Safran C, Chute CG. Exploration and exploitation of clinical databases. International Journal of Bio-Medical Computing. 1995;39(1):151-6.

CHAPTER 2

AVAILABILITY AND USABILITY OF DATA FOR MEDICAL PRACTICE ASSESSMENT

Hilco Prins, Frea Kruisinga, Hans Büller, Bertie Zwetsloot-Schonk

International Journal for Quality in Health Care. 2002 Apr;14(2):127-37

ABSTRACT

Objective: We analyzed availability and usability of the electronic patient data required for assessment of medical practice for a specific patient group.

Design: Case study in which physicians defined performance indicators and additional exploratory information. Data availability in the hospital information system was determined. Data usability was evaluated based on reason for recording, administrative procedures and comparison with paper data.

Setting: A 155 bed pediatric department in a public academic medical center.

Study participants: Pediatricians and children with suspected meningitis.

Main outcome measures: Availability and usability of electronic patient data. Usability criteria were standardization, completeness and correctness.

Results: A total of 14 performance indicators were defined. Of 39 data items required for indicator quantification, 29 were available, and 19 were usable without manual handling. Completeness and correctness of registration of reason for admission and discharge diagnoses were insufficient, leading to problematic patient selection and complication detection. Time-points of patient events were incorrect or not available. Data regarding outpatient diagnosis, signs and symptoms, indications for test ordering and medication administration were missing. Test result reports were not adequately standardized. Based on electronic patient data, five out of 14 performance indicators could be quantified reliably, but only after patient selection problems were overcome. For exploratory information, 16 out of 25 required data items were available and 13 were usable.

Conclusions: Availability and usability of electronic patient data are insufficient for physician-led and detailed assessment of medical practice for specific patient groups. Extended registration of reason for admission will improve patient selection and assessment of diagnostic process.

Keywords: Data Collection, Data Quality, Hospital Information System, Meningitis, Outcome Measurement, Pediatrics, Process Measurement

2.1 INTRODUCTION

Assessment of medical practice for clinically defined patient groups may be used for improvements in quality of care [1]. Medical practice is the diagnostic, therapeutic and follow-up decisions and services of physicians. Performance indicators may assist medical practice assessment [2-4]. Performance indicators are systematically developed quantitative measurements that can be used to assess appropriateness of specific health care decisions, services and outcomes [5]. Using performance indicators, aspects of care can be quantified and the resulting values can be compared to standards [5]. A standard is a chosen level of performance that has to be met or surpassed.

For quantification of performance indicators, reliable data about patient characteristics, care process and outcomes are a prerequisite [6, 7]. Correct interpretation of performance indicator values requires insight into case-mix [8, 9]. If indicators point to below standard care, additional information should be retrieved for further exploration. For practical reasons, the required data should be in electronic and standardized form [10]. Hospital information systems (HIS) may be appropriate as data source [11-13].

In the present case study we analyzed availability and quality of patient data in HIS, for assessment of medical practice for a specific patient group. We were interested in the following:

1. Which performance indicators, case-mix and exploratory information should be selected for medical practice assessment?
2. Are required data electronically available and usable for medical practice assessment?

2.2 MATERIALS & METHODS

2.2.1 Study Design, Setting and Materials

In 1996, a case study was performed at the Department of Pediatrics of the Academic Medical Center (AMC) in Amsterdam. The AMC is a university hospital with an integrated HIS [14]. This means that from a central patient module, electronically available patient data can be examined. Workstations are available in every important clinical workplace. The clinical use of this HIS is limited to examination of test results and patient history. This history consists of earlier

diagnoses and discharge letters. For other documentation the paper medical record is the main source. Furthermore, the system is used for administrative and billing reasons, e.g. diagnosis and procedure registration. An outpatient diagnosis registry, medication prescription system and order management system are under construction. The AMC is quite unique in the sense that, besides discharge diagnoses, reason for admission is also coded and recorded. Reason for admission is defined as diagnosis, symptom, sign or injury that, *at the time of admission*, was considered as reason for admission.

The pediatric department is a tertiary center with 155 beds. The case study concerned the assessment of medical practice for children with suspected or proven meningitis and who were not premature neonates or patients with cancer. Nine pediatricians and two medical informaticians were involved in the medical practice assessment process. We retrieved patient data from the HIS and paper medical records retrospectively.

2.1.2 Methods

The pediatricians formulated performance indicators for local use and assessed their medical practice during four meetings. Before each meeting the pediatricians were asked to provide pre-specified input. To support pediatricians, the medical informaticians searched and summarized literature, extracted and analyzed patient data, prepared the meetings and structured the results. The meetings led to consensus regarding:

1. A flow chart of the care process;
2. A set of performance indicators;
3. The accompanying standards;
4. Data availability and usability, plus quality of medical practice.

Our method is further elaborated below.

Performance Indicators, Case-mix and Exploratory Information

To obtain an agreed overview of relevant medical decisions and activities and to lay an unambiguous foundation for the rest of the project, the care process was modeled. Nine pediatricians filled out questionnaires about diagnostic and therapeutic activities in response to an exemplary clinical case of suspected

meningitis, and were subsequently interviewed based on local guidelines. For every pediatrician a flow chart [15] was constructed. The 19 aspects on which opinions differed were agreed upon by majority votes after thorough discussion during the first consensus meeting. In the resulting flow chart, 21 different patient states, 39 decisions and 46 activities were made explicit.

We then provided the pediatricians with summarized literature about performance indicators [16]. Based on the flow chart, each pediatrician formulated performance indicators on special forms [17]. We asked them not to take data availability into consideration. The pediatricians formulated 63 performance indicators, of which 29 were unique: 20 process and nine outcome indicators. During the second consensus meeting, indicators were discussed and tested against the RUMBA criteria: relevance, understandable, measurable, formulated in behavioral terms, and acceptable [18, 19]. This resulted in 14 performance indicators. The pediatricians also agreed on three case-mix parameters influencing interpretation of provided care.

Subsequently, each pediatrician defined standards based on literature with quantitative clinical findings, and personal experience and knowledge about local circumstances and patient population. During the third consensus meeting, definitive standards were set for their own clinical setting. Pediatricians also defined exploratory information for each performance indicator, in case provided care deviates from the standard.

Based on the defined performance indicators, case-mix, and exploratory information, necessary data items were listed.

Availability and Usability of Data

Subsequently, patient selection, quantification of performance indicators, gathering of case-mix and exploratory information, and presentation of results to the pediatricians took place. During these activities it was determined whether data items were available in the HIS, and whether they were usable for medical practice assessment.

Patient selection is a first and important step in indicator quantification. Criteria for patient selection were; age ≤ 18 years, treatment by pediatricians, having (suspected) meningitis as reason for admission or meningitis as one of the discharge diagnoses, but not being a premature neonate or patient with cancer. Therefore, the following data items were required: birth date, admission date,

specialty of admitting physician, reason for admission, discharge diagnoses and ward. The HIS functioned as sampling frame. We selected all patients who had an ICD-9-CM [20] meningitis code as the reason for admission or discharge diagnosis and who also fulfilled the other criteria. The results will show that this selection strategy was not sufficient and that another, more laborious and less effective, strategy was necessary to continue the project.

After patient selection, data for indicator quantification, case-mix, and exploratory information were collected from the HIS and, when not available, from paper medical records. During data collection, usability of electronically available data was estimated. Usability was estimated based on how data were collected at the source, administrative procedures for recording, original reason for which data were recorded, and comparison with paper data whenever possible. Insight into procedures for collecting and recording data was acquired by interviewing pediatricians, secretaries and a medical record coder. The paper medical record served as the gold standard only for reason for admission and diagnoses. Other data are recorded in either the HIS or in the paper medical record. Test results found in paper records are printouts of the HIS and could thus not serve as a gold standard. We determined standardization, completeness, and correctness of the data. Standardization refers to the use of a controlled terminology and structured recording. This makes automatic handling possible. Completeness is the proportion of true data that is recorded. Correctness is the proportion of recorded data that is true. With our method only rough estimations of completeness and correctness were possible. On the basis of these estimations we determined whether performance indicators could be quantified reliably.

In the fourth consensus meeting, we provided information about availability and usability of data and about the provided care. During this meeting our interpretation about data quality was discussed and agreed upon by the pediatricians. Subsequently, with the data limitations in mind, the pediatricians assessed their own medical practice based on the quantification of performance indicators and in view of the defined standards, case-mix, and exploratory information.

2.3 RESULTS

2.3.1 Performance Indicators, Case-mix and Exploratory Information

Fourteen performance indicators with standards were defined. These cover important aspects of care from admission to outpatient follow-up. Ten relate to process and four to outcomes. Of the ten process indicators, five refer to diagnostic, three to therapeutic, and two to follow-up activities. The 14 indicators with standards are listed below.

Diagnostic process indicators (CSF, cerebrospinal fluid):

1. $$\frac{\text{Number of children with suspected meningitis having a lumbar puncture} < 3 \text{ hours after admission}}{\text{Number of children with suspected meningitis having a lumbar puncture}} \geq 0.75$$

2. $$\frac{\text{Number of children with suspected meningitis having a lumbar puncture}}{\text{Number of children with suspected meningitis}} \geq 0.95$$

3. $$\frac{\text{Number of children with suspected meningitis having CSF cytology}}{\text{Number of children with suspected meningitis having a lumbar puncture}} = 1.00$$

4. $$\frac{\text{Number of children with suspected meningitis having CSF/serum glucose ratio measured}}{\text{Number of children with suspected meningitis having a lumbar puncture}} \geq 0.95$$

5. $$\frac{\text{Number of children with suspected meningitis having CSF culture}}{\text{Number of children with suspected meningitis having a lumbar puncture}} \geq 0.95$$

Therapeutic process indicators:

6. $$\frac{\text{Number of children with (suspected) meningitis receiving antibiotics} < 3 \text{ hours after arrival}}{\text{Number of children with (suspected) meningitis receiving antibiotics}} \geq 0.75$$

7. $$\frac{\text{Number of children with (suspected) meningitis started with antibiotics according to protocol}}{\text{Number of children with (suspected) meningitis started with antibiotics}} = 1.00$$

8. $$\frac{\text{Number of children with meningitis having antibiotics adjusted to antibiogram}}{\text{Number of children with meningitis having antibiogram}} \geq 0.80$$

Follow-up indicators:

9. $$\frac{\text{Number of children with meningitis having outpatient visit} < 8 \text{ weeks after discharge}}{\text{Number of children with meningitis}} \geq 0.90$$

10. $$\frac{\text{Number of children with meningitis having hearing test between 4 and 12 weeks after discharge}}{\text{Number of children with meningitis}} \geq 0.90$$

Outcome indicators:

11. $$\frac{\text{Number of children with meningitis having length of stay} < 21\text{ days}}{\text{Number of children with meningitis}} \geq 0.75$$

12. $$\frac{\text{Number of children with meningitis having residual neurologic impairments}}{\text{Number of children with meningitis}} \leq 0.20$$

13. $$\frac{\text{Number of children with meningitis having residual hearing impairments}}{\text{Number of children with meningitis}} \leq 0.25$$

14. $$\frac{\text{Number of children with meningitis dying during admission}}{\text{Number of children with meningitis}} \leq 0.07$$

The pediatricians selected *severity of illness* at presentation, *pathogenic organism* and *age* as necessary case-mix information. Table 1 shows the desired exploratory information.

2.3.2 Availability and Usability of Data

Patient Selection

The selection of patients with suspected meningitis based on registration of reason for admission and discharge diagnoses failed for several reasons. According to the rules, the reason for admission should contain an ICD-9-CM meningitis code in cases of admission with suspected meningitis, even if the eventual diagnosis appears to be another disease (which is the case in approximately two third of the patients). However, as the medical record coder informed us, reason for admission in this situation is often, for the sake of convenience, equated with the most important discharge diagnosis. The fact that a child has been admitted with suspected meningitis is lost. Sometimes the non-disease-specific ICD-9-CM code V718 'Observation for other specified suspected conditions' or an ICD-9-CM code for a symptom that contributes to the suspicion is recorded. In another study we have already shown that the registration of principal and secondary diagnoses was not complete and not correct [21].

Because of the registration shortcomings, complete patient selection could not be obtained. Therefore, another additional strategy had to be applied. We also selected children who had the V718 code or an ICD-9-CM code for symptoms, relevant to the suspected disease in the reason for admission. Furthermore, we selected all

children who underwent a lumbar puncture but who were not staying on the neonatology and oncology wards. For the remaining children we verified the presence of (suspected) meningitis based on information in paper medical records. The selection procedure with resulting patient numbers is presented in Figure 1. Based on registration of ICD-9-CM meningitis codes alone, 39 instead of 102 patients would have been selected.

Table 1: Defined exploratory information related to relevant performance indicators.

Defined exploratory information for process indicators	**PI[1]**
Time interval between arrival and first contact with pediatrician	1,6
Percentage of children with contraindication for lumbar puncture (coagulation disturbance, intra-cranial mass effect or cardio-respiratory instability); Percentage of children with lumbar puncture elsewhere provided	2
Percentage children for whom CSF[2] cytology is ordered but not (successfully) performed	3
Percentage children with only serum glucose; Percentage children with only CSF glucose; Percentage children for whom CSF and serum glucose is ordered but result not available; Percentage children with CSF and serum glucose but ratio not calculated	4
Percentage of children for whom CSF culture is ordered but result not available	5
Time interval between arrival and result CSF cytology, and between prescription and administration of antibiotics	6
Reasons to deviate from antibiotic protocol	7
Reasons not to adjust to antibiogram	8
Percentage of no-shows; Percentage of children with follow-up in another hospital	9,10
Type of hearing test related to age	10
Defined exploratory information for outcome indicators	**PI**
Mean length of stay per pathogenic organism; Percentage and type of complications; Length of stay in preceding hospital; Mortality	11
Percentage of children with preceding hospital care elsewhere	11,14
Severity of neurological impairments; Type of neurological tests; Percentage neurological impairments per pathogenic organism; Percentage neurological impairments per age category; Percentage neurological impairments early developed from onset; Percentage neurological impairments per severity of illness category	12
Severity of hearing impairments; Type of hearing tests; Percentage hearing impairments per pathogenic organism; Percentage hearing impairments per age category; Percentage hearing impairments early developed from onset; Percentage hearing impairments per severity of illness category	13
Mortality per pathogenic organism, per age category and per severity of illness category	14

[1] PI = Performance Indicator; [2] CSF = Cerebro Spinal Fluid

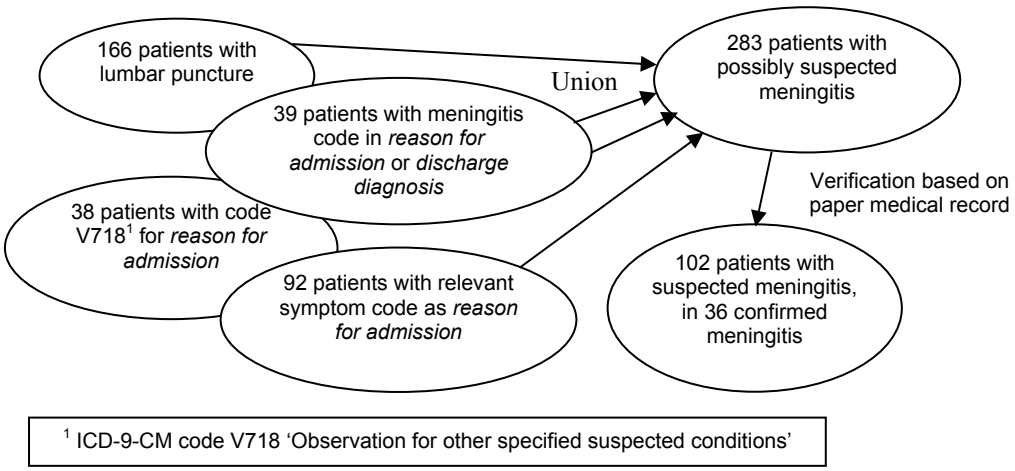

Figure 1: Selection procedure with quantities of children with suspected meningitis.

Table 2 shows that of the 39 patients with an ICD-9-CM meningitis code, 31 did indeed have meningitis. Note that there is an overlap, e.g. some patients with lumbar puncture have relevant ICD-9-CM codes.

Table 2: Number of patients selected from HIS based on selection criteria and their true status according to the paper medical record.

Selection Criterion HIS[1] (n=283)		**Status according to paper medical record**	
		Suspected Meningitis (n=102)	Meningitis (n=36)
ICD code	meningitis[2] (n=39)	38	31
	V718[3] (n=38)	7	2
	symptoms[4] (n=92)	21	2
Lumbar Puncture (n=166)		81	22

[1] HIS = Hospital Information System
[2] All ICD-9-CM meningitis codes (found in reason for admission or discharge diagnoses)
[3] ICD-9-CM code defined as: 'Observation for other specified suspected conditions' (found in reason for admission)
[4] Selection of ICD-9-CM codes for symptoms relevant for suspected disease (found in reason for admission)

Suppose the selection was based on all possibly relevant ICD-9-CM codes. From Table 3 it can be derived that recall (or sensitivity) is 0.60 and precision (or

positive predicted value) 0.37. Lumbar puncture as a selection criterion leads to 41 extra patients with true (suspected) meningitis.

Table 3: Number of patients selected based on ICD-9-CM codes by patients' disease status according to the paper medical records.

		(Suspected) Meningitis[1]		
		+	-	
ICD code[2]	+	61	104	165
	-	41	6973	7014
		102	7077	7179[3]

[1] According to paper medical record
[2] ICD-9-CM meningitis codes (in reason for admission or discharge diagnoses), code V718 (in reason for admission) and codes for symptoms relevant for suspected meningitis (in reason for admission)
[3] Total number of admissions in sample frame

Using these data
- Recall (or sensitivity) is calculated as 61/102 = 0.60
- Precision (or positive predictive value) is calculated as 61/165 = 0.37

If we evaluate the diagnosed meningitis patients only, then selection based on ICD-9-CM meningitis codes alone, whether as reason for admission or discharge diagnosis, results in a 0.86 recall and 0.79 precision (Table 4).

Table 4: Number of patients selected based on ICD-9-CM meningitis codes by patients' meningitis status according to the paper medical records.

		Meningitis[1]		
		+	-	
ICD code[2]	+	31	8	39
	-	5	7135	7140
		36	7143	7179[3]

[1] According to paper medical record
[2] ICD-9-CM meningitis codes (in reason for admission or discharge diagnoses)
[3] Total number of admissions in sample frame

Using these data
- Recall (or sensitivity) is calculated as 31/36 = 0.86
- Precision (or positive predictive value) is calculated as 31/39 = 0.79

Table 5 is based on findings obtained during the selection procedure. As registration of reason for admission was inadequate, additional data were needed. However, signs, symptoms, and test indications were not recorded electronically. Although the activity 'performance of lumbar puncture' itself is not recorded, lumbar puncture was considered to be performed if we found evidence of CSF testing, the results of which are virtually always reported through the HIS.

Table 5: Availability and usability of data needed to select patients.

Data item(s)	**Available**				**Usable**
		Standardized	*Complete*	*Correct*	
Patient: birth date	y	y	y	y	y
Admission: date	y	y	y	y	y
Specialty of admitting physician: type	y	y	y	y	y
Reason for admission: type	y	y	n	n	n
Inpatient diagnosis: type	y	y	n	n	n
Inpatient ward: type	y	y	y	y	y
Sign/Symptom: type	n	-	-	-	-
Additional test: indication (lumbar puncture)[1]	n	-	-	-	-
Additional test: type (lumbar puncture)	y	y	y	y	y

[1] Order management system under construction

Performance Indicator Quantification

Table 6 shows data needed to quantify performance indicators, without data exclusively needed for patient selection (birth date, specialty of physician, reason for admission, and ward). It is noteworthy that time-points of clinical events are either not recorded or are recorded incorrectly. Often when a time-point is recorded, it is an administrative time, which does not reflect the precise time-point of the event. A diagnosis date is equated with the discharge date, which probably does not coincide with the actual moment of diagnosis. As a result, we have no information about whether a diagnosis was a complication that originated during course of admission, or whether the diagnosis was already at hand at the moment of admission. Time and date of medication administration are not available. Of the neurological and hearing impairments found in the paper medical records (n = 6 and n=2, respectively) none was present in the diagnosis registration. Many of the

result reports do not use standard terminology or are not structured, preventing automatic handling.

Table 6: Availability and usability of data needed to quantify the performance indicators.

Data item(s)	**For PI**[1]	**Available**				**Usable**
			Standar-dized	*Com-plete*	*Cor-rect*	
Encounter: arrival date / time[2]	6	n	-	-	-	-
Outpatient visit: date	9	y	y	y	y	y
Admission: date	1,11	y	y	y	y	y
Admission: time	1	y	y	y	n	n
Discharge: date	9-11	y	y	y	y	y
Inpatient death: date	14	y	y	y	y	y
Inpatient diagnosis: type	8-14	y	y	n	n	n
Inpatient diagnosis: date	12,13	y	y	y	n	n
Outpatient diagnosis: type / date	12,13	n	-	-	-	-
Medication: type	6-8	n	-	-	-	-
Medication: administration date	6-8	n	-	-	-	-
Medication: administration time	6	n	-	-	-	-
Sign/Symptom: type / date	12,13	n	-	-	-	-
Additional test: type						
-lumbar puncture	1-5	y	y	y	y	y
-hearing test[3]	10,13	y	y	y	y	y
-neurological tests	12	y	y	y	y	y
Additional test: performance date						
-lumbar puncture	1-5,7	y	y	y	y	y
-hearing test	10,13	y	y	y	y	y
-neurological tests	12	y	y	y	y	y
Additional test: performance time						
-lumbar puncture	1	y	y	y	n	n
Additional test: result						
-CSF[4] cytology	3,(8-14)[5]	y	y	y	y	y
-CSF glucose; -Serum glucose[6]	4	y	y	y	y	y
-CSF culture						
-virology	(8-14)[5]	y	n	y	y	n
-bacteriology	5,(8-14)[5]	y	n	y	y	n
-antibiogram	8	y	n	y	y	n
-neurological tests	12	y	n	y	y	n
-hearing test	13	y	n	y	y	n
Additional test: result date						
-antibiogram	8	y	y	y	y	y
-neurological tests	12	y	y	y	y	y
-hearing tests	13	y	y	y	y	y

[1] PI = Performance Indicator
[2] Encounter arrival time is not necessarily equal to admission time. Arrival time is the time a patient enters the hospital whether or not he/she will be admitted. Admission time is the time a patient enters the ward where he/she will be admitted. When a patient first visit the emergency room or outpatient clinic followed by an admission, the two time-points can differ substantially.
[3] BAER (Brainstem Auditory Evoked Response) or Audiogram.
[4] CSF = Cerebro-Spinal Fluid
[5] To verify diagnosis; electronically only available for patients who underwent lumbar puncture in own hospital.
[6] CSF/serum glucose ratio not electronically available as these measures are done in two different laboratories with different information systems. Both measures are available electronically separately.

For patient selection and quantification of indicators, 39 different data items have been considered for use; 29 were available and 19 usable.

Case-mix and Exploratory Information

Table 7 shows the data items regarding exploratory information needed when care deviates from standards. Data items for case-mix information are included. Data on severity of illness at the moment of admission are not available. Results regarding pathogenic organism can be obtained from laboratory results. Information on medication prescription, reason of deviation from protocol, no show and care provided elsewhere (important in case of transfer) were unavailable. A total of 45 data items were needed, of which 29 were available and 20 usable.

Table 7: Availability and usability of data to explain deviation from the standard.

Data item(s)	**For PI**[1]	**Available**				**Usable**
			Standar-dized	*Com-plete*	*Cor-rect*	
Patient: birth date	12-14	y	y	y	y	y
Encounter: arrival date / time[2]	1,6	n	-	-	-	-
Outpatient visit: no-show	9	n	-	-	-	-
Admission: date	11-14	y	y	y	y	y
Discharge: date	11	y	y	y	y	y
Inpatient: place of origin	2,11	y	y	y	y	y
Inpatient: disposition	9,10	y	y	y	y	y
Admission: severity of illness[3]	12-14	n	-	-	-	-
Inpatient death: date	11	y	y	y	y	y
Specialty: first contact time	1,6	n	-	-	-	-
Inpatient diagnosis: type	2,8,11-14	y	y	n	n	n
Inpatient diagnosis: date	12-14	y	y	y	n	n
Outpatient diagnosis: type / date[4]	12,13	n	-	-	-	-

Medication: type[5]	6	n	-	-	-	-
Medication: prescription time[5]	6	n	-	-	-	-
Medication: reason to deviate from protocol	7	n	-	-	-	-
Medication: reason not to adjust to antibiogram	8	n	-	-	-	-
Sign/Symptom: type	2,11	n	-	-	-	-
Sign/Symptom: severity	12,13	n	-	-	-	-
Sign/Symptom: date	2,12,13	n	-	-	-	-
Additional test: type						
-coagulation test; -CT[6] scan; -ECG[7]	2	y	y	y	y	y
Additional test: order date						
-CSF[8] cytology	3	y	y	y	y	y
-CSF glucose; Serum glucose	5	y	y	y	y	y
-CSF culture	5	y	y	y	y	y
Additional test: performance date						
-hearing test[9]	13	y	y	y	y	y
-no show	10	n	-	-	-	-
-neurological tests	12	y	y	y	y	y
Additional test: result						
-CSF culture						
-virology; -bacteriology	11-14	y	n	y	y	n
-neurological tests	12	y	n	y	y	n
-hearing test	13	y	n	y	y	n
-coagulation test; -CT scan; -ECG	3	y	n	y	y	n
Additional test: result date						
-CSF cytology	6	y	y	y	y	y
-coagulation test; -CT scan; -ECG	2	y	y	y	y	y
Additional test: result time						
-CSF cytology	6	y	y	y	y	y
Care elsewhere before admission[10]	2,11	n	-	-	-	-

[1] PI = Performance Indicator

[2] Encounter arrival time is not necessarily equal to admission time. Arrival time is the time a patient enters the hospital whether or not he/she will be admitted. Admission time is the time a patient enters the ward where he/she will be admitted. When a patient first visit the emergency room or outpatient clinic followed by an admission, the two time-points can differ substantially.

[3] Severity of illness at the moment of admission.

[4] Outpatient diagnosis registry under construction

[5] Medication prescription system under construction

[6] CT = Computer Tomography

[7] ECG = Electro Cardiogram.

[8] CSF = Cerebro Spinal Fluid

[9] Lumbar puncture in another hospital; admission and discharge date of preceding hospital.

[10] To verify diagnosis, but electronically only available for patients who underwent lumbar puncture in own hospital.

Tables 5, 6 and 7 show availability and usability of data for patient selection, performance indicator quantification, and exploratory information. There is some redundancy, e.g. admission date is needed for patient selection, quantification of performance indicators and exploratory information. For case-mix and exploratory information, 25 new data items were added, of which 16 were available and 13 usable. Combining all data items leads to 64 different data items, of which 45 were available and 32 usable.

Based on availability and usability of data, the possibility of quantifying performance indicators reliably is presented in Table 8. Even if it were possible to select patients reliably, five of the fourteen performance indicators could not be quantified reliably.

Table 8: Possibility to quantify performance indicators reliably.

PI number	Quantifiable[1]	Explanation
1	y	
2	y	
3	y	
4	y	If CSF[2]-glucose and blood-glucose available, then glucose ratio is supposed
5	y	
6	n	Hospital arrival and administration time of antibiotics are not recorded
7	n	Administration of antibiotics is not recorded
8	n	Administration of antibiotics is not recorded
9	y	Only for children not referred to other hospitals after treatment
10	y	Only for children not referred to other hospitals after treatment
11	y	
12	n	Conclusions EEG[3] are reported in free text; signs, symptoms and outpatient diagnosis registration is lacking; only for children not referred to other hospitals after treatment
13	n	Conclusions hearing tests are reported in free text, outpatient diagnosis registration is lacking; only for children not referred to other hospitals after treatment
14	y	

[1] Provided that suspected meningitis patients have been selected successfully.
[2] CSF = Cerebro Spinal Fluid
[3] EEG = Electro Encephalogram

2.4 DISCUSSION

We studied availability and usability of electronic data for medical practice assessment of children with suspected meningitis. Pediatricians defined 14 performance indicators, case-mix, and exploratory information. Of the 39 data

items needed for patient selection and indicator quantification 29 were electronically available and 19 usable without manual handling. Reason for admission and diagnoses were incomplete and incorrectly recorded. This seriously hampered patient selection and detection of complications. Time-points of clinical events and interventions were either not available or incorrect. Outpatient diagnosis, signs and symptoms, indications for tests and data about medication administration were missing. Many test result reports were not adequately standardized. Therefore, even if it were possible to select patients reliably, five of the 14 performance indicators could not be quantified. For case-mix and exploratory information, 25 additional data items were needed, of which 16 were available and 13 usable. Data about severity of illness, medication prescription, reasons for deviation from protocol, no show and care provided elsewhere were particularly likely to be missing.

This medical practice assessment was meant for internal use only, contrary to some areas where performance of hospitals, managed care organizations or individual physicians are reported publicly [22-24]. This local, internal use allows medical practice assessment at a specific and detailed level. On a larger scale, this detailed assessment is probably not possible. However, according to our pediatricians, only a detailed assessment does justice to the complex processes. Our study empirically supports Palmer's conclusion [8] that "many different process-based measures are needed to comprehensively assess quality, and many process-based measures require detailed clinical data currently found only in medical records".

Our study has some methodological limitations. Most importantly, we performed a case study in one hospital, with a specific HIS, and based on one patient group. Therefore, our evaluation of data quality is specific to the chosen hospital. Another hospital may have a different pattern of data availability and usability. The choice for another patient group would have led to other performance indicators. Despite these limitations, we believe that our study demonstrates the practical difficulties in implementing ongoing performance measurements using available patient data. These practical difficulties are fairly universal. Many studies evaluated quality of a limited data set. Results of these studies are often consistent with our estimates. For example, we assumed quality of demographic patient data to be complete and correct. This is in agreement with findings of other studies [25-28]. We assumed registration of admission and discharge date to be good. Horbar and Leahy [27] and Teikari and Raivio [28] reported an error rate of about 5 – 10%. We reported

problems with discharge diagnoses, as do other studies [26, 29-36]. We evaluated the quality of procedural codes as being positive. Cooper et al [37] and Schwartz et al [38] concluded that hospital-based procedural codes are a reasonably accurate source of data for process and outcomes analyses of gastro-intestinal hemorrhage and perinatal care, respectively. No studies evaluating quality of the whole data set needed for medical practice assessment have been found.

Another limitation of this study is the determination of completeness and correctness of data. We estimated these quality aspects (as suggested by [39]) based on procedures for collecting and recording data, and on original reason for recording. As with much data available electronically, a gold standard could not be constructed, and there were no other means to evaluate data quality in this retrospective study. Many data are available either electronically or on paper. The problem of constructing a true gold standard for electronic clinical data has already been mentioned by Brennan and Stead [40]. We could construct a gold standard only for reason for admission and for discharge diagnoses. Therefore, we attached great value to validation of our estimates by the pediatricians. For comparison between electronic and paper representation of reason for admission and diagnoses the term "concordance" is more appropriate than "gold standard" [41]. However, we believe that in our hospital the paper representation gives a better depiction of the real status of the patient than the electronic representation, which has no function in daily patient care.

In the results section, problems with the registration of suspected meningitis are described. But there are other, more fundamental, problems too. Firstly, the ICD-9-CM provides no possibility to describe *'suspected meningitis'*. Also the registry itself does not provide the possibility of indicating the status of selected ICD-9-CM codes. This means that no distinction can be made between patients admitted with suspected meningitis and patients admitted with proven meningitis. This last situation occurs frequently in a tertiary care hospital. Secondly, only one reason for admission can be recorded in our HIS. In case where suspected meningitis was part of a differential diagnosis but not the immediate working diagnosis, it will not be recorded as such. Not many institutions record reason for admission. We found no other study about data quality of reason for admission. Trepka et al [42] concluded that only 38.3% of the persons with an ICD-9-CM tuberculosis code as one of the discharge diagnoses did actually have tuberculosis. This was due to the fact that in

the registration no distinction could be made between suspected tuberculosis and diagnosed tuberculosis.

Despite problems with data availability and data quality, the project still had clear benefits. The new idea to flowchart each physician's process uncovered more practice variability than the pediatricians were aware of. Importantly, the indicator reporting showed, as far as we could measure, that the provided care resembled desired care, except regarding hearing follow-up. Only 31% of children with meningitis underwent a hearing test 4 - 12 weeks after discharge.

Our method of developing performance indicators and assessing medical practice was comparable to the ten-step monitoring and evaluation process of the Joint Commission on Accreditation of Healthcare Organizations [(2)], but we organized the steps in four consensus meetings. This was done to minimize the time burden on pediatricians. In our experience, in four meetings, pediatricians are able to develop performance indicators and assess their medical practice, provided that they are supported by experts on medical practice assessment and patient documentation. To take data quality explicitly into account was important to improve the credibility of this quality of care project.

We purposely started a study on a patient group defined by a suspected disease. In this way it was possible to analyze data availability and usability for the assessment of care from first contact to follow-up. The idea of assessing diagnostic processes by studying a patient group defined by a "suspected" disease has not been discussed previously in literature. Assessing patient groups defined by established diagnoses limits the possibility of assessing the diagnostic process. It leads to a situation in which patients admitted with a suspected disease, but who are eventually found to have another disease, are not taken into account when assessing the diagnostic process. Especially for serious diseases that have to be ruled out in case of suspicion, assessment of the diagnostic process can only be done meaningfully if all patients with the suspected disease are included. Defining patient groups by diagnostic procedure alone is not a good alternative. Firstly, not many diagnostic procedures are disease specific. This means that using it as a selection criterion will also lead to selection of not-intended patients. Registration of test indication can address this problem. Secondly, there can be a contra-indication for the procedure, leading to not selecting intended patients. We did use lumbar puncture as a selection criterion, but in addition to the diagnosis-based criteria. This could have introduced selection bias. Patients with a contra-indication

or who (incorrectly) did not undergo a lumbar puncture could have been overlooked if the diagnosis registration failed for these patients. This means that the indicator depicting the performance of a lumbar puncture probably scores higher than its true value. From a methodological point of view it is not correct to use a dependent variable as selection criterion.

It is difficult to define a direct relation between a diagnostic process and expected patient outcomes. In general, patients will benefit from a quick and adequate diagnostic process. Specific outcomes are highly dependent on the eventual diagnosis and chosen therapy. Therefore, the quality of the diagnostic process in our case study is defined in terms of timely and appropriate actions. Further, to keep the assessment manageable, we defined outcome indicators only for those patients with confirmed meningitis.

Patients with suspected meningitis were not an easy group to study. Firstly, it was not always easy to establish whether the diagnosis was suspected meningitis. For this, we used explicit criteria during the medical record-based verification. Secondly, some of the patients were admitted with already proven meningitis. Furthermore, there was great variability in severity of illness and pathogenic organisms. In fact there were several subgroups. A solution could be to define more specific performance indicators, e.g. performance indicator 11: length of stay for meningococcal meningitis. This does not influence the total set of required data, but for performance indicators themselves, more data items are required. This can result in indicators that are even more difficult to quantify. Many of the defined exploratory information are performance indicators themselves, such as time interval between prescription and administration of antibiotics. However, most of them are only applicable in small subgroups.

We conclude from our data that pediatricians define quality in terms of appropriateness and timeliness of interventions, and patient outcomes. In general, this means that patient conditions, interventions, and exact timing of both need to be registered. From the perspective of pediatricians, detailed information is needed to assess their own medical practice. Part of this information cannot be delivered from today‘s clinical information systems. The main problems are incomplete and incorrect registration of discharge diagnoses, time-points that are not or incorrectly recorded, some patient conditions that are not recorded and some test results that are not standardized. For assessment of the diagnostic process, a reliable registration of reason for admission is necessary. Not many hospitals do register

reason for admission and if they do, the completeness and correctness of the information is questionable. Besides clinical information systems, use of paper medical records is necessary for additional data and verification. Although information supply is a problem, participation of physicians in a quality of care project leads to better awareness of important aspects of care and may uncover possibilities for improvement.

ACKNOWLEDGEMENT

We thank Martien Limburg for his substantial comments on the manuscript.

REFERENCES

1. McIntyre N. Evaluation in clinical practice: problems, precedents and principles. J Eval Clin Pract 1995;1(1):5-13.
2. Joint Commission on Accreditation of Healthcare Organizations. An introduction to Quality Improvement in Health Care; The transition from QA to CQI. Oakbrook Terrace, Illinois: JCAHO, 1991.
3. Kazandjian VA, Wood P, Lawthers J. Balancing science and practice in indicator development: the Maryland Hospital Association Quality Indicator (QI) project. Int J Qual Health Care 1995;7(1):39-46.
4. Nadzam DM, Turpin R, Hanold LS, White RE. Data-driven performance improvement in health care: the Joint Commission's Indicator Measurement System (IMSystem). Jt Comm J Qual Improv 1993;19(11):492-500.
5. Institute of Medicine (Field M, Lohr, KN. eds). Guidelines for clinical practice: from development to use. Washington DC: National Academic Press, 1992.
6. Iezzoni LI, Foley SM, Daley J, Hughes J, Fisher ES, Heeren T. Comorbidities, complications, and coding bias. Does the number of diagnosis codes matter in predicting in-hospital mortality? [see comments]. JAMA 1992;267(16):2197-203.
7. Wyatt J. Acquisition and use of clinical data for audit and research. J Eval Clin Pract 1995;1(1):15-27.
8. Palmer RH. Process-based measures of quality: the need for detailed clinical data in large health care databases. Ann Intern Med 1997;127(8 Pt 2):733-8.
9. Richardson D, Tarnow-Mordi WO, Lee SK. Risk adjustment for quality improvement. Pediatrics 1999;103(1 Suppl E):255-65.
10. Iezzoni LI. Assessing quality using administrative data. Ann Intern Med 1997;127(8 (Pt 2)):666-74.
11. Safran C, Chute CG. Exploration and exploitation of clinical databases. Int J of Biomed Comput 1995;39(1):151-6.

12. McKee M. Routine data: a resource for clinical audit? Quality in Health Care 1993(2):104-11.

13. McKee M, James P. Using routine data to evaluate quality of care in British hospitals. Med Care 1997;35(10 Suppl):OS102-11.

14. Bakker AR. An integrated hospital information system in The Netherlands. MD Comput 1990;7(2):91-7.

15. Committee on Standardization of Clinical Algorithms. Proposal for clinical algorithm standards. Society for Medical Decision Making Committee on Standardization of Clinical Algorithms. Med Decis Making 1992;12(2):149-54.

16. Joint Commission on Accreditation of Healthcare Organizations. Characteristics of clinical indicators. Qual Rev Bull 1989;15(11):330-9.

17. Lehmann R. Joint Commission Forum. Forum on clinical indicator development: a discussion of the use and development of indicators. QRB Qual Rev Bull 1989;15(7):223-7.

18. Casparie AF. Handbook quality of care (in Dutch). Utrecht: De Tijdstroom, 1993.

19. Klazinga NS. Quality management of medical specialist care in The Netherlands. Rotterdam: Erasmus University, 1996.

20. Commission on Professional and Hospital Activities. The International Classification of Diseases, ninth revision, Clinical Modification. Ann Arbor: Commission on Professional and Hospital Activities, 1978.

21. Prins H, Buller HA, Zwetsloot-Schonk JH. Effect of discharge letter-linked diagnosis registration on data quality. Int J Qual Health Care 2000;12(1):47-57.

22. Green J, Wintfeld N. Report cards on cardiac surgeons. Assessing New York State's approach [see comments]. N Engl J Med 1995;332(18):1229-32.

23. Epstein AM. Rolling down the runway: the challenges ahead for quality report cards. JAMA 1998;279(21):1691-6.

24. Spoeri RK, Ullman R. Measuring and reporting managed care performance: lessons learned and new initiatives. Ann Intern Med 1997;127(8 Pt 2):726-32.

25. Brewster D, Crichton J, Muir C. How accurate are Scottish cancer registration data? Br J Cancer 1994;70(5):954-9.

26. Colin C, Ecochard R, Delahaye F, et al. Data quality in a DRG-based information system. Int J Qual Health Care 1994;6(3):275-80.

27. Horbar JD, Leahy KA. An assessment of data quality in the Vermont-Oxford Trials Network database. Control Clin Trials 1995;16(1):51-61.

28. Teikari JM, Raivio I. Validity of hospital discharge registry in acute glaucoma. Acta Ophthalmol (Copenh) 1987;65(5):589-90.

29. Benesch C, Witter DM, Jr., Wilder AL, Duncan PW, Samsa GP, Matchar DB. Inaccuracy of the International Classification of Diseases (ICD-9-CM) in identifying

the diagnosis of ischemic cerebrovascular disease [published erratum appears in Neurology 1998 Jan;50(1):306]. Neurology 1997;49(3):660-4.

30. Dixon J, Sanderson C, Elliott P, Walls P, Jones J, Petticrew M. Assessment of the reproducibility of clinical coding in routinely collected hospital activity data: a study in two hospitals. J Public Health Med 1998;20(1):63-9.

31. Fisher ES, Whaley FS, Krushat WM, et al. The accuracy of Medicare's hospital claims data: progress has been made, but problems remain. Am J Public Health 1992;82(2):243-8.

32. Gibson N, Bridgman SA. A novel method for the assessment of the accuracy of diagnostic codes in general surgery. Ann R Coll Surg Engl 1998;80(4):293-6.

33. Green J, Wintfeld N. How accurate are hospital discharge data for evaluating effectiveness of care? Med Care 1993;31(8):719-31.

34. Pladevall M, Goff DC, Nichaman MZ, et al. An assessment of the validity of ICD Code 410 to identify hospital admissions for myocardial infarction: The Corpus Christi Heart Project [see comments]. Int J Epidemiol 1996;25(5):948-52.

35. Raiford DS, Perez Gutthann S, Garcia Rodriguez LA. Positive predictive value of ICD-9 codes in the identification of cases of complicated peptic ulcer disease in the Saskatchewan hospital automated database. Epidemiology 1996;7(1):101-4.

36. Romano PS, Mark DH. Bias in the coding of hospital discharge data and its implications for quality assessment. Med Care 1994;32(1):81-90.

37. Cooper GS, Chak A, Lloyd LE, Yurchick PJ, Harper DL, Rosenthal GE. The accuracy of diagnosis and procedural codes for patients with upper GI hemorrhage. Gastrointestinal Endoscopy 2000;51(4 Pt 1):423-6.

38. Schwartz RM, Gagnon DE, Muri JH, Zhao QR, Kellogg R. Administrative data for quality improvement. Pediatrics 1999;103(1 Suppl E):291-301.

39. Lodder H, Bakker AR, Zwetsloot JHM. Hospital Information Systems; Technical Choices. In: Bemmel van JH, Musen MA, eds. Handbook of Medical Informatics. Houten, The Netherlands: Bohn Stafleu Van Loghum, 1997:

40. Brennan PF, Stead WW. Assessing data quality: from concordance, through correctness and completeness, to valid manipulatable representations [editorial; comment]. J Am Med Inform Assoc 2000;7(1):106-7.

41. Stein HD, Nadkarni P, Erdos J, Miller PL. Exploring the degree of concordance of coded and textual data in answering clinical queries from a clinical data repository. J Am Med Inform Assoc 2000;7(1):42-54.

42. Trepka MJ, Beyer TO, Proctor ME, Davis JP. An evaluation of the completeness of tuberculosis case reporting using hospital billing and laboratory data; Wisconsin, 1995. Ann Epidemiol 1999;9(7):419-23.

CHAPTER 3

REDESIGN OF DIAGNOSTIC CODING IN PEDIATRICS: FROM FORM-BASED TO DISCHARGE LETTER-LINKED

Hilco Prins, Hans Büller, Bertie Zwetsloot-Schonk

Perspectives in Health Information Management. 2004;1(1):1-10

ABSTRACT

Diagnostic coding after hospital discharge is mainly based on abstracting of paper medical records by medical record coders. Studies show that the quality of these data is often moderate, possibly because discharge registries play no role in daily patient care. Timely writing of discharge letters is needed to support continuity of care, at least in the Netherlands. This article describes the redesign and evaluation of diagnosis registration and discharge letter writing at a Dutch pediatric department.

Formerly, pediatricians at this department completed discharge forms. However, many forms were completed with insufficient information or not at all. Pediatricians now provide diagnoses with codes in a special heading of the discharge letter. The medical record coder checks and corrects this diagnosis heading. A list of diagnoses for pediatrics, based on ICD-9-CM, was developed and alphabetically ordered into one booklet used by pediatricians when dictating discharge letters. A reminder system for in-time writing of letters was implemented.

Since 1995, this discharge letter-linked registration has proven to be applicable in daily care. How accurately pediatricians filled in the diagnosis heading was analyzed during two periods. In 1995, 25% of the diagnoses were initially (before adjustments made by the medical record coder) not coded or incorrectly coded; nine percent of these shortcomings could be attributed to the pediatricians. In 1997, 67% of the diagnoses were initially not coded or incorrectly coded; 37% of these shortcomings were attributable to pediatricians. Initially, only half of the letters was written within six weeks after discharge. The correction function of the medical record coder is indispensable.

Keywords: Diagnosis; Registries; Patient Discharge; Forms and Records Control; Medical Record Administrators; Hospital, Pediatric

3.1 INTRODUCTION

Process, Use and Quality of Diagnosis Registration

Diagnosis registration in hospitals usually takes place after discharge of the patient by abstracting of the paper medical record and encoding of the resulting diagnostic information by a medical record coder into ICD-9-CM [1] or ICD-10 [2]. The codes are electronically recorded, primarily for statistical, policy, or reimbursement reasons. Studies that measured quality of diagnostic data in a broad domain of medicine show that both completeness and correctness are less than 100% [3-10]. These inaccuracies hamper use of these data, for example, for next patient encounters [11], communication about patients [8] and assessment of medical practice [12-16]. Most likely, when diagnosis registration is integrated with daily medical practice, data quality will improve [17, 18]. Use of a computerized patient record (CPR) may provide substantial impetus [19], although this is not guaranteed [20, 21]. CPRs are implemented in high care environments, like neonatology and intensive care, that are characterized by technical orientation, automatic monitoring and supportive data recording by specialized nurses. In many other departments, it will probably be years before physicians will record diagnostic information electronically.

Discharge letters play an important role in continuity of care in some countries. The specialist sends a discharge letter to the general practitioner (GP) to inform him or her about the course during admission, the treatment, and the condition of the patient. Availability of discharge letters may decrease readmission rates [22]. However, several problems are reported, especially late receipt of the letters [23-25] and missing information [24, 26], including admission and discharge diagnoses [27].

This paper is based on a research project conducted in the department of pediatrics at the Academic Medical Center (AMC) in the Netherlands. The purpose of the project was to improve the accuracy of the diagnosis registration and accelerate discharge letter writing. This paper describes the redesign process, especially the involvement of the pediatricians in it, the new registration procedure, and the evaluation of the coding performance of pediatricians in the new situation.

3.2 METHODS

The Problem Case: Former Diagnosis Registration at a Pediatric Department

The pediatric department of the AMC, a tertiary teaching hospital in Amsterdam, has 155 beds and handles 4,500 admissions a year. In the AMC, diagnoses of admitted patients are collected and translated into ICD-9-CM and subsequently sent to the Dutch hospital discharge registry. This registry collects, for statistical purposes, data of hospital admissions on demographic and medical characteristics of patients.

Figure 1 presents an overview of the former registration procedure at the pediatric department. After each responsibility period, wherein one specialty is responsible for the total medical care for a patient, physicians collected diagnostic data in free text on paper forms. Medical record coders in a central department encoded the physicians' descriptions and recorded the codes in the hospital information system (HIS). A previously detected problem was that the pediatricians recorded many diagnoses inaccurately, or not at all, on the paper forms. Therefore, the medical record coder customarily used discharge letters to optimize diagnosis registration; however, feedback was never given to physicians. The pediatricians already gave much attention to the structure and content of the discharge letters. After a brief notification at discharge, the letter had to be sent to the GP within six weeks, a goal that was often not achieved.

Several types of diagnoses could be collected: reason for admission, primary diagnosis, and, when applicable, secondary diagnoses, complications, and cause of injury. In case of more than one responsibility period during hospital admission, medical record coders decided which primary diagnosis was the principal diagnosis of the admission. Diagnoses were coded according to the eight-digit AMC Diagnosis Codes (ADC), a local extension of ICD-9-CM. The first digit indicates whether a code is a disease (D), visit (V), external cause of injury (E), or morphology (M) code. The last digit is a control digit. The other six digits can be used for the diagnosis itself. Thus, depending on the length of the ICD-9-CM code, one, two, or three digits are available for extension.

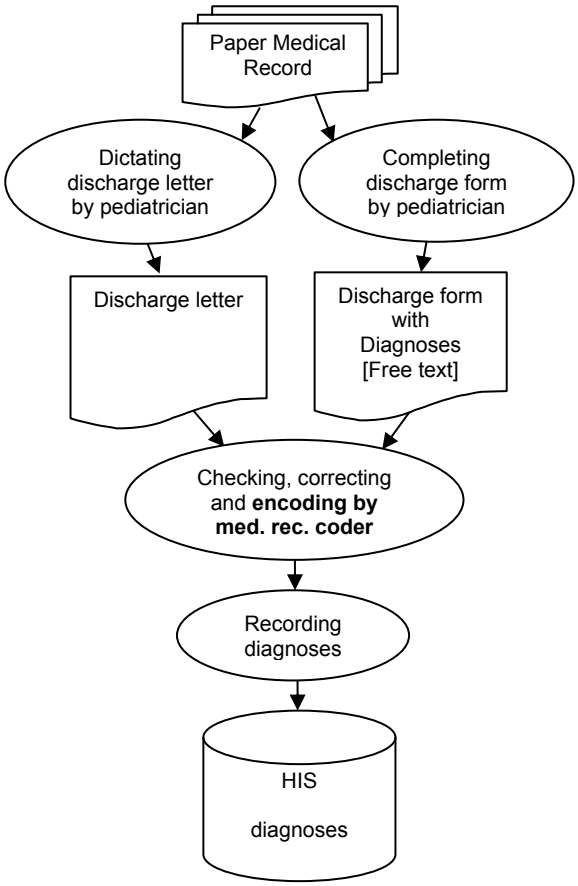

Figure 1: Former Procedure of Diagnosis Registration.

The medical staff of the pediatric department felt a need for improved feedback about their medical care for the purpose of quality assessment. It was felt that the reliability of routinely collected diagnostic data was insufficient; furthermore, the coding system lacked specificity (level of detail). The medical staff felt they themselves should code diagnoses on the level of detail used in daily practice. In 1993, redesign of the diagnosis registration process was begun to make the diagnosis registration complete, correct, specific, and timely enough to support assessment of medical practice.

Therefore, the following prerequisites were formulated. The registration:

1. has a role in the care process;

2. is integrated into the work of pediatricians;
3. does not increase the administrative workload of pediatricians;
4. forms a unique collection of diagnoses for both internal and external use;
5. provides the ability to convert diagnoses to ICD-9-CM;
6. takes place within six weeks after discharge;
7. forms a basis for further automation and fits within developments towards a CPR.

Diagnosis Registration Redesign

To meet the prerequisites, a project group worked out six changes in the registration process. The group consisted of a pediatrician, medical informatician, medical record coder, epidemiologist, administrative staff members, and a junior researcher. During the development, nine pediatric subspecialty groups were involved. The process from idea to implementation took two years. The six changes were:

1. **Completing the form with free text was replaced by completing a special heading on the discharge letter with standardized diagnoses and their ICD-9-CM-based local codes** (Figure 2). Because of the role in continuity of care, medical staff believed that a discharge letter-linked diagnosis registration would enhance completeness and correctness of the registration.
2. **A reminder system for in-time writing of letters was developed.** Letters were to be sent to the GP, and diagnoses recorded in the HIS, within six weeks. If necessary, the pediatrician received a first reminder four weeks after discharge.
3. **The discharge letters new style became available on the local PC network.** The discharge letter with diagnosis heading might act as a first information source in case of readmission.
4. **Instead of the medical record coder, pediatricians themselves encoded the diagnoses.** As pediatricians know their patients, they are in a better position to choose the appropriate codes.
5. **In order to support pediatricians in finding the relevant diagnosis codes, and to meet the necessary level of detail, a portable booklet was developed**

with a list of pediatric diagnoses in alphabetic order (Figure 3). The new list had to be a subset of the ADC. In cooperation with the pediatricians, a selection of the relevant diagnoses was taken from the ADC; within that selection, further specifications were made. Appendix A describes the procedure.

6. **The role of the medical record coder was decentralized from the medical administration department to the pediatrics department**, leading to better communication between coder and pediatricians. The coder spent less time encoding and more time advising and checking.

MEDICAL REGISTRATION HEADING		
Item	**Description**	**Code**
referred by	general practitioner	
diagnosis history	Gastro esophageal reflux Asthma, bronchial Allergy	05301049 04939025 09953009
reason for admission	Stridor, inspiratory	07861015
principal diagnosis	Epiglottitis	04643002
secondary diagnoses	none	
complications	none	
cause of injury	irrelevant	

Figure 2: Example of a Completed Medical Registration Heading at the Bottom of the Discharge Letter.

If, for example, a junior physician writes a discharge letter under the new diagnosis registration process, the junior physician dictates the letter and selects appropriate diagnoses with codes from the booklet. Next, a secretary types the letter with diagnoses and codes in the medical registration heading. Subsequently, a senior pediatrician checks the letter. The medical record coder checks the correctness, completeness and specificity of the diagnoses. If letter and heading have been agreed upon, the countersigned letter is sent to the GP and a copy is inserted into the medical record. The electronic version is placed on the local PC network. The codes are stored in the HIS by the medical record coder (Figure 4).

05781002	Melaena
07816003	Meninigismus
	Meningitis
03200009	due to Haemophilus
00360005	due to Meningococcus
03201005	due to Pneumococcus
03203004	due to Staphylococcus
03202007	due to Streptococcus
00130001	due to Tuberculosis
03209003	bacterial, unspecified
00479004	viral, unspecified
03229005	unspecified
07598389	Menkes' Disease
02703013	Menkes' Syndrome (Maple Syrup Disease)

Figure 3: Part of booklet with diagnoses for pediatrics in the AMC.

Implementation

The new procedure was introduced, with full commitment of the head of the department, at two staff meetings. In addition, written instructions about how to complete the diagnosis heading were incorporated in the diagnosis booklet. After implementation, the medical record coder gave feedback to the pediatricians about their coding. The new system was implemented in February 1995.

Evaluation

We evaluated how accurately and when pediatricians filled in the diagnosis heading of the discharge letters. For this, completed diagnosis headings were analyzed, and interviews were held.

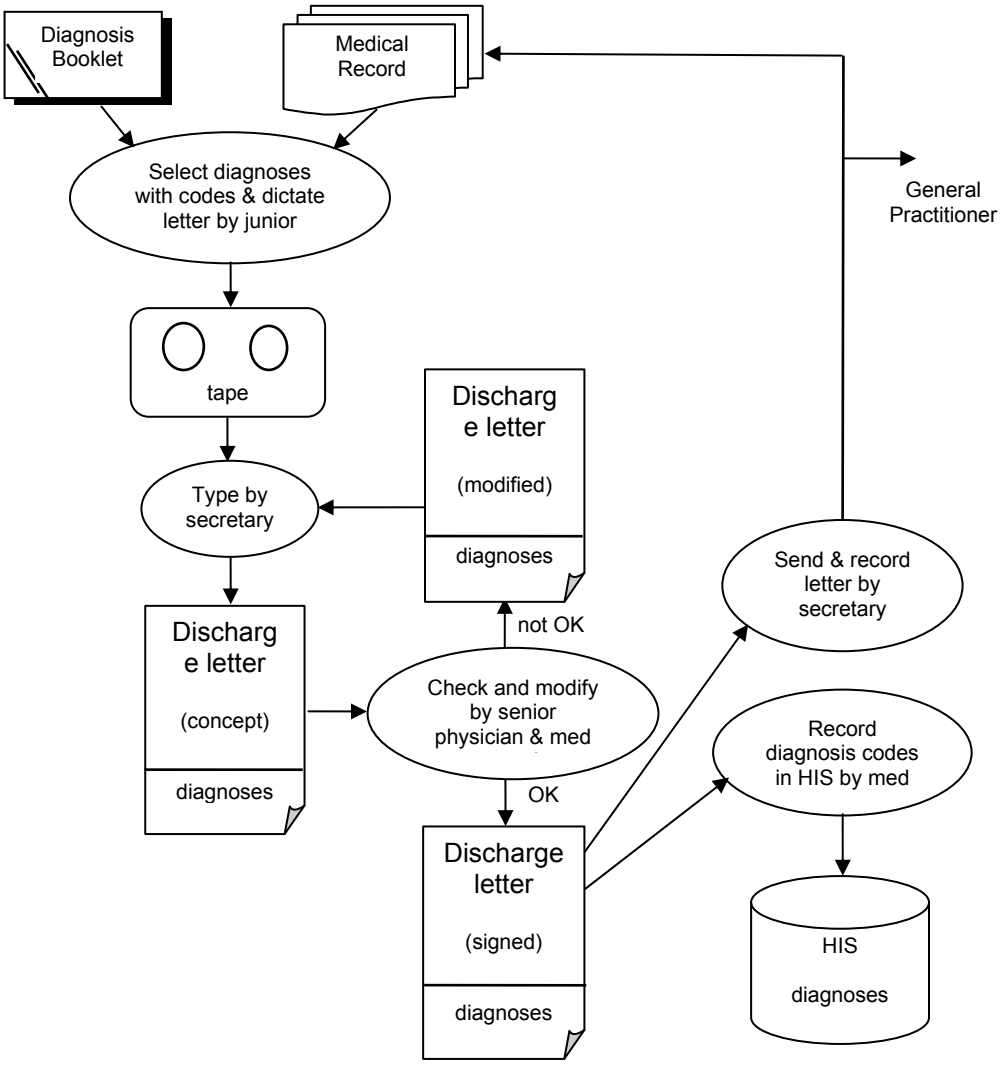

Figure 4: New Procedure of Diagnosis Registration

To analyze the diagnosis headings, the medical record coder completed an evaluation form (Figure 5) after every first version of the 276 discharge letters of patients discharged from February 1995 to May 1995. This analysis was done at an early stage of the implementation to find out where improvements had to be made. This analysis was repeated by taking a small sample of 30 discharge letters of patients discharged in June 1997. Unfortunately, because of staff problems, it was not possible to use the same time period (February to May) in 1997. The time

interval between February and May 1995 was split into two periods in order to analyze whether a positive learning curve existed. Furthermore, the results of the period from February to May 1995 were compared to the results of June 1997 in order to evaluate the long-term effect. The z-test with continuity correction was used to compare the periods.

- Discharge date
- Check date of letter
- Total number of diagnoses per letter
- Medical record coder agrees with Medical Registration Heading: y / n
- Number of mistakes found and of what kind:
  - relevant diagnosis missing
  - incorrect diagnosis
  - diagnosis in wrong category
  - diagnosis code not found by physician, but listed in the diagnosis booklet
- Number of diagnoses rightly not found because not listed in the booklet
- Letter sent to the secretary / physician

Figure 5: Items on the Evaluation Form filled out by Medical Record Coder.

In June 1995, four months after introduction, five junior physicians, a senior pediatrician, and the medical record coder were interviewed by a research assistant (see Figure 6 for questions asked). The semi-open interviews were directed at obtaining insight into physicians' opinions about their new role, the feasibility of the procedure, the effect that physicians and medical record coder expected on data quality and their opinions on which factors of the procedure contribute to this effect.

3.3 RESULTS

Table 1 gives the distribution of the number of diagnoses in the text of the discharge letters, according to the medical record coder. The number of diagnoses per letter ranged from two to 11. Two is the minimum number of diagnoses, as "reason for admission" and "principal diagnosis" are obligatory.

- What is your opinion about the new diagnosis registration procedure?
- What are, from your point of view, advantages of the new diagnosis registration compared to the former procedure?
- What are, from your point of view, disadvantages of the new diagnosis registration compared to the former procedure?
- Do you experience bottlenecks that hamper feasibility of the new procedure?
- What is the effect of the reminder system on writing your discharge letters?
- Do you think that the final diagnostic data quality (completeness, correctness and specificity) really has increased?
  - If yes, which factors do you think contribute to better data quality (physician encoding, diagnosis booklet, check by senior physician / medical record coder, integration with discharge letter)?
  - If no, why not?

Figure 6: List of questions asked at the Semi open Interviews of Pediatricians and Medical Record Coder.

Table 1: Distribution of Number of Diagnoses per Discharge Letter, according to Medical Record Coder.

Number of diagnoses per discharge letter	Number of discharge letters				
	Feb '95	March'95	Apr '95	May '95	June '97
2	38	37	35	36	12
3	14	19	17	14	4
4	14	8	5	12	5
5	5	4	3	1	5
6	1	3	1	0	0
7	1	0	0	0	1
8	2	0	0	0	0
9	3	0	0	2	1
10	0	0	0	0	2
11	1	0	0	0	0
Mean number of diagnoses	3.37	2.83	2.65	2.85	3.90
Standard deviation	2.21	1.25	1.23	1.53	2.37

Table 2: Results of the Evaluation at the Level of Diagnoses by the Medical Record Coder about how the Pediatricians filled out the Medical Registration Heading of the Discharge Letter.

	Feb. - May '95			**June '97**
Number of diagnoses in text of discharge letters	*Feb - March '95* # (prop[3].)	*Apr - May '95* # (prop.)	*Whole period* # (prop.)	# (prop.)
all	467	347	814	115
- incorrectly or not coded in heading	114 (.24)	90 (.26)	204 (.25)	77 (.67)[1]
• diagnosis incorrect	12 (.03)	8 (.02)	20 (.02)	0 (.00)
• diagnosis in wrong category (e.g. principal instead of secondary)	19 (.04)	1 (.00)[1]	20 (.02)	0 (.00)
• diagnosis text or code missing	83 (.18)	81 (.23)[2]	164 (.20)	77 (.67)[1]
✔ diagnosis text and code missing	6 (.01)	3 (.01)	9 (.01)	21 (.18)[1]
✔ diagnosis given in text; code not found, although in booklet	15 (.03)	12 (.03)	27 (.03)	22 (.19)[1]
✔ diagnosis given in text; code rightly not found, because not in booklet	62 (.13)	66 (.19)[2]	128 (.16)	34 (.30)[1]
▸ code relevant for booklet, according to pediatrician	15 (.03)	12 (.03)	27 (.03)	1 (.01)
▸ code not relevant for booklet, according to pediatrician	47 (.10)	54 (.16)[2]	101 (.12)	33 (.29)[1]

[1] statistically significant at a 1 percent level
[2] statistically significant at a 5 percent level
[3] proportion

Table 2 compares the findings of the first and second halves of the February to May 1995 period; also the whole 1995 period is compared to the June 1997 period. In the second half of the February to May 1995 period, the number of diagnoses in the incorrect category decreased, and the number of diagnoses for which text or code was missing increased; the reason for the latter observation was that diagnoses and codes were rightly not found in the booklet but, according to the pediatricians, also were not very relevant to be entered into the booklet. In February to May 1995, 204 of all 814 diagnoses (25 percent) found in the text of the discharge letter were, according to the medical record coder, initially coded inaccurately; 40 (5 percent) of them were incorrectly coded (20) or wrongly categorized (20) – both situations that can be attributed to the physicians - and 164 (20 percent) were missing. However, 128 codes (16 percent) were missing because they were not listed in the diagnosis booklet; the other 36 missing codes (4 percent) can be attributed to the physicians - nine times diagnosis text and code were

missing, and 27 times diagnosis text was given but there was no available code in the booklet. In 27 cases (3 percent), the pediatricians found it important to list the missing diagnosis in the next version of the booklet, especially diagnoses in neonatology. In 1997, 77 of 115 diagnoses (67 percent) were initially not coded, 34 codes (30 percent) because they were not listed in the diagnosis booklet. Shortcomings attributable to pediatricians increased from nine percent in February to May 1995 to 37 percent in June 1997.

Table 3: Results of the Evaluation at the Level of the Discharge Letters by the Medical Record Coder about how the Pediatricians filled out the Medical Registration Heading of the Discharge Letter.

	Feb. - May '95			**June '97**
Number of discharge letters	*Feb - March '95* # (prop[2].)	*Apr - May '95* # (prop.)	***Whole period*** # (prop.)	# (prop.)
all	150	126	276	30
-headings with incorrect/missing diagnoses/codes	75 (.50)	68 (.54)	143 (.52)	27 (.90)[1]
•letters without diagn. information in heading	0 (.00)	0 (.00)	0 (.00)	5 (.17)[1]
•letters with incorrect code or category	21 (.14)	10 (.08)	31 (.11)	0 (.00)[1]
•letters with missing code	65 (.43)	59 (.47)	124 (.45)	27(.90)[1]
✔ missing, though code in booklet	18 (.12)	12 (.09)	30 (.11)	13 (.43)[1]
✔ missing, code not in booklet	47 (.31)	47 (.37)	94 (.34)	14 (.47)
▸ code should be in booklet	12 (.08)	12 (.10)	24 (.09)	1 (.03)
▸ intentionally not in booklet	36 (.24)	37 (.29)	73 (.26)	13 (.43)
letters sent back to senior pediatrician with feedback	53 (.35)	22 (.17)[1]	75 (.27)	0 (.00)[1]

[1] statistically significant at a 1 percent level

[2] proportion

Table 3 gives results at the level of discharge letters. From February to May 1995, 143 of 276 first versions of the letters (52 percent) showed one or more shortcomings. There were hardly differences between first and second halves of this period, except for decrease in letters sent back to the senior physician for feedback. In 1997, 27 of 30 letters (90 percent) contained missing codes; in 5 letters (17 percent), the registration headings was not filled in at all.

Table 4 shows that in the period from February to May 1995, each month fewer than half of the discharge letters were sent to the GP within six weeks after

discharge. This time interval is equivalent for recording diagnoses in HIS. In June 1997, 28 of 30 letters (93 percent) were sent to the GP within six weeks.

Table 4: Distribution of Time Interval between Discharge and sending Letter to General Practitioner.

Time interval (weeks)	Number of discharge letters				
	Feb '95	*Mar '95*	*Apr '95*	*May '95*	*June '97*
1	3	7	2	0	1
2	11	9	3	1	0
3	21	8	1	4	4
4	14	0	3	5	8
5	7	3	3	7	8
6	2	5	1	8	7
7	4	4	5	8	1
8	4	6	6	8	1
9	2	3	5	4	
10	2	9	4	2	
11	1	4	5	3	
12	2	4	7	5	
13	1		10	1	
14	2	1	5	2	
15				4	
16		2		2	
17		6			
18	1				
19				1	
≥ 20	1x21w 1x27w		1x22w		
Average time interval (days)	35.63	48.83	62.30	54.86	31.17
Standard deviation	32.77	34.56	30.53	27.08	8.57

From the interviews in June 1995, four months after implementation, we learned that pediatricians had complaints about the time needed to find diagnoses in the booklet and about not finding diagnoses. There was resistance to and lack of enthusiasm for the changes. Writing discharge letters is generally seen as a burden, which was made even more complicated by diagnostic coding. The reminder system resulted, in the experience of the pediatricians, in writing a letter where formerly no letter was written at all. The knowledge that a senior pediatrician would be checking their letters compelled junior physicians to fill in the diagnosis heading more seriously. However, junior physicians criticized the lack of feedback.

The medical record coder sent fewer and fewer letters back to the senior pediatrician. The pediatrician felt that this practice of feedback was a burden and took too much time. Although the medical record coder thought that diagnostic data in the letter was not nearly as complete and correct as possible, she thought that the new procedure was an improvement. Formerly, many forms were not filled in at all or filled in superficially. All believed that completeness, correctness, and specificity of the diagnosis registration had increased (although not enough) as a result of the more active way the pediatricians dealt with the registration, the use of the booklet, and the checks. The pediatricians did not believe that embedding the diagnosis registration in the discharge letter as such contributed to the quality of the registration.

3.4 DISCUSSION

It is important to have a high-quality diagnosis registration. As this was not the case at a pediatric department in Amsterdam, a project was initiated to improve the registration. Today, instead of the medical record coder, pediatricians themselves encode diagnoses. After discharge, diagnoses are described in standardized form with codes in a heading of the discharge letter. In order to support the pediatricians, a booklet with alphabetically ordered diagnoses was developed. This booklet contains a selection and further specification of the ICD-9-CM. A reminder system was implemented to stimulate discharge letter writing within six weeks. The role of the medical record coder shifted from encoding to checking and advising.

Completion of the heading by the pediatrician was evaluated. In the first four months, 25 percent of the diagnoses were not coded or inaccurately coded; 9 percent of these shortcomings could be attributed to the pediatricians, 16 percent to incompleteness of the diagnosis booklet. As a consequence, more than half of the first versions of the letters contained shortcomings. In the course of 1995, there was no positive learning curve despite feedback. In 1997, 67 percent of the diagnoses were not coded or incorrectly coded; 37 percent of these shortcomings could be attributed to the pediatricians, 30 percent to incompleteness of the diagnosis booklet. Consequently, almost all first versions of the letters contained shortcomings. This represented a considerable downswing. In 1995, fewer than half of the letters was written, and subsequently accompanying diagnoses recorded, within the predefined six weeks. In 1997, more than 90 percent of the letters were written within six weeks.

From interviews early after implementation, it was learned that, although the diagnosis heading was not nearly as complete, correct and timely as possible, the new procedure was seen as an improvement. Formerly, many forms were not filled in at all or filled in superficially. The medical record coder had to complete or correct almost all forms. Unfortunately, we could not verify this, as all forms were destroyed.

The diagnosis booklet was not complete. Some diagnoses, for example, rare diseases, were intentionally not listed in the booklet. However, in an academic medical center, rare diseases occur more frequently than in nonacademic hospitals. The neonatology department did not participate in the project; consequently, relevant diagnoses were not included. However, many children from the neonatology department are later carried over to the pediatric department.

Three reasons were advanced for the deterioration seen in 1997. Firstly, through the lack of feedback and diminishing attention, the pediatricians became more lax with regard to the registration. Second, new junior physicians had entered the department since the introduction and were not familiar with the new registration. There was no adequate training program. Thirdly, although the primary goal of the new registration was to enable assessment of medical practice, structural assessment of medical practice was not implemented. In 1996, a medical practice assessment project was performed [(14)]. It showed difficulties in availability and usability of electronic data and had no sequel. To guarantee a certain level of data quality, the medical record coder was emphatically instructed to check and correct the diagnostic data critically based on the text of the discharge letter.

Conforming to sociotechnical guidelines [(28)], we gave the diagnosis registration a role in daily care, integrated it with physicians' workflow, involved users in the development and gave careful attention to commitment and introduction. Still, the role of the pediatricians in the process was not very successful. The linkage to the discharge letter was probably not a good choice. Physicians experience writing letters as a burden and additional coding as a complication of this task. This does not support in-time writing and accurate coding. Apparently, the role of the diagnosis heading in communication is not considered very important and thus gave physicians insufficient incentives, especially in combination with a lack of consequences of low data quality. This study shows the difference between theoretical design and practical work. In practice, processes are optimized to reach short-term goals and interest.

We expected a better quality of the discharge letter-linked diagnosis registration compared to the form-based registration. This study showed that encoding and filling in the diagnoses in a special heading of the discharge letter by the pediatricians was not optimal. It is important to realize that after checking, the medical record coder corrected imperfections she found. Whether the combined effort of pediatrician and medical record coder leads to better data quality can only be tested with a before and after study in which electronically recorded diagnostic data are compared to a gold standard. This study has been performed [8] and showed that completeness of form-based diagnosis registration was 0.51 and of discharge letter-linked diagnosis registration 0.54. Correctness was 0.65 and 0.67 respectively. Completeness is the proportion of all relevant diagnoses that is recorded. Correctness is the proportion of recorded diagnoses that is true. It was concluded that the discharge letter-linked diagnosis registration did not provide a better basis for assessment of medical practice.

When Dutch hospitals change over to the ICD-10, the AMC children's list has to be adapted. However, in doing so, the essence of the process of new diagnosis registration will remain the same. Besides ICD, many local or disease-specific diagnosis classifications are available. Specialists prefer to use their own classification. If conversion tables with translation to ICD codes become available, these classifications can be built into our concept. For communication with the GP, conversion of ICD-codes to International Classification of Primary Care (ICPC) [29] is attractive. When discharge letters are sent electronically, the GP can record diagnostic data automatically in his or her information system. GPs prefer this kind of discharge summary [30].

The solution chosen here can be implemented within the contemporary information infrastructure of hospitals. It is a step towards the CPR, as the physician selects standardized diagnosis descriptions and the electronically recorded diagnoses get a role in the care process. A specific diagnosis list or classification is an essential prerequisite for the CPR. However, given the somewhat disappointing results, it is questionable whether our experiment is worth following.

Other initiatives found in literature

Other initiatives have been undertaken to improve the quality of diagnostic data. Cox et al. [31] designed and tested a checklist to improve encoding of acute myocardial infarction by a medical record coder. Unfortunately, this solution to a

particular encoding problem does not provide a solution to a generic encoding problem at a pediatric department where a broad range of diagnoses are encountered. Hohnloser et al. [17] tested the use of a computerized browsing and encoding tool by clinicians. The tool was part of a CPR. They concluded that the tool could increase data quality and the volume of documented data. However, the test was performed in an ICU/CCU environment where CPR yields profit more obviously than most other medical departments. Delamarre et al. [32] described an automated coding system of free text patient discharge summaries from the field of coronary diseases into the ICD-9-CM classification. This system seems to offer possibilities, but it cannot be implemented in short term at a pediatric department, as the system covers only a limited domain. Besides, it is necessary to structure the discharge summary in such a way that the system can classify a diagnosis as "reason of admission", "principal diagnosis", or "secondary diagnosis". Moreover, it is questionable whether in the future the discharge letter will be a basis for diagnosis registration or whether the diagnosis registration will form input for the discharge letter. The latter will probably be the case [33, 34]. Van Walraven and Demers [10] showed that using a high-quality clinical database instead of reviewing medical records improves the correctness and completeness of diagnostic coding. Lorenzoni et al [35] showed that training of medical record coders improves quality of data abstracted from the medical record. Arts et al. [36] showed that training physicians in data definition and extraction is an effective way to improve quality of intensive care data, including diagnoses.

3.5 CONCLUSION

Within the contemporary information infrastructure of hospitals, discharge letter-linked diagnosis registration appears feasible in routine practice, but the process is not sufficient to improve the final data quality. If the physicians views the diagnosis registration as having only an additive role in communication with other health care providers after discharge, the correction function of the medical record coder is indispensable.

Quality of diagnosis registration is difficult to manage. Continuous attention from both the physician and medical record coder seems necessary. When diagnostic data are used for medical practice assessment and therefore affect physicians, a positive effect on the quality of the data can be expected. As McKee [13] states: "The feedback loop must be closed".

REFERENCES

1. National Center for Health Statistics. International Classification of Diseases, 9th Revision, Clinical Modification, 6th edition. Hayattsville, Maryland: NCHS; 2003.
2. World Health Organization. International Statistical Classification of Diseases and Related Health Problems; Tenth Revision. Geneva: World Health Organization; 1992.
3. Cleary R, Beard R, Coles J, Devlin B, Hopkins A, Schumacher D, et al. Comparative hospital databases: value for management and quality. Qual Health Care 1994;3(1):3-10.
4. Elkin PL, Ruggieri AP, Brown SH, Buntrock J, Bauer BA, Wahner-Roedler D, et al. A randomized controlled trial of the accuracy of clinical record retrieval using SNOMED-RT as compared with ICD9-CM. Proc AMIA Symp 2001:159-63.
5. Fisher ES, Whaley FS, Krushat WM, Malenka DJ, Fleming C, Baron JA, et al. The accuracy of Medicare's hospital claims data: progress has been made, but problems remain. American Journal of Public Health 1992;82(2):243-8.
6. Gittelsohn A, Senning J. Studies on the reliability of vital and health records: I. Comparison of cause of death and hospital record diagnoses. Am J Public Health 1979;69(7):680-9.
7. Martinez Jimenez R, Garcia Benavides F. [Hospital morbidity statistics: the accuracy of the diagnosis reported in the discharge registry book]. Med Clin (Barc) 1991;96(20):765-8.
8. Prins H, Buller HA, Zwetsloot-Schonk JH. Effect of discharge letter-linked diagnosis registration on data quality. Int J Qual Health Care 2000;12(1):47-57.
9. Quan H, Parsons GA, Ghali WA. Validity of information on comorbidity derived rom ICD-9-CCM administrative data. Med Care 2002;40(8):675-85.
10. van Walraven C, Demers SV. Coding diagnoses and procedures using a high-quality clinical database instead of a medical record review. J Eval Clin Pract 2001;7(3):289-97.
11. van Walraven C, Seth R, Laupacis A. Dissemination of discharge summaries. Not reaching follow-up physicians. Can Fam Physician 2002;48:737-42.
12. Iezzoni LI. Assessing quality using administrative data. Annals of Internal Medicine 1997;127(8 (Pt 2)):666-74.
13. McKee M. Routine data: a resource for clinical audit? Quality in Health Care 1993(2):104-11.
14. Prins H, Kruisinga FH, Buller HA, Zwetsloot-Schonk JH. Availability and usability of data for medical practice assessment. Int J Qual Health Care 2002;14(2):127-37.
15. Safran C, Chute CG. Exploration and exploitation of clinical databases. International Journal of Bio-Medical Computing 1995;39(1):151-6.
16. Wyatt J. Acquisition and use of clinical data for audit and research. Journal of Evaluation in Clinical Practice 1995;1(1):15-27.

17. Hohnloser JH, Puerner F, Soltanian H. Improving clinician's coded data entry through the use of an electronic patient record system: 3.5 years experience with a semiautomatic browsing and encoding tool in clinical routine. Computers & Biomedical Research 1996;29(1):41-7.

18. Grobe SJ. Informatics: the infrastructure for quality assessment and quality improvement [editorial]. Journal of the American Medical Informatics Association 1995;2(4):267-8.

19. Porcelli PJ, Jr. An electronic documentation system increases diagnostic code capture for very low birth weight infants. Proc AMIA Symp 2001:543-7.

20. Hogan WR, Wagner MM. Accuracy of data in computer-based patient records. Journal of the American Medical Informatics Association 1997;4(5):342-55.

21. Arts DG, De Keizer NF, Scheffer GJ. Defining and improving data quality in medical registries: a literature review, case study, and generic framework. J Am Med Inform Assoc 2002;9(6):600-11.

22. van Walraven C, Seth R, Austin PC, Laupacis A. Effect of discharge summary availability during post-discharge visits on hospital readmission. J Gen Intern Med 2002;17(3):186-92.

23. Bolton P, Mira M, Kennedy P, Lahra MM. The quality of communication between hospitals and general practitioners: an assessment. J Qual Clin Pract 1998;18(4):241-7.

24. Christensen KA, Mainz J, Kristensen E. [Communication between primary and secondary sector and its significance for patient care. Evaluation based on admission records and discharge letters of patients admitted acutely to a department of internal medicine]. Ugeskr Laeger 1997;159(48):7141-5.

25. Jorgensen FS, Kjaergaard J, Blegvad K. [Evaluation of the discharge summary by general practitioners. A questionnaire study]. Ugeskr Laeger 1990;152(42):3056-9.

26. Raval AN, Marchiori GE, Arnold JM. Improving the continuity of care following discharge of patients hospitalized with heart failure: is the discharge summary adequate? Can J Cardiol 2003;19(4):365-70.

27. van Walraven C, Weinberg AL. Quality assessment of a discharge summary system [see comments]. Cmaj 1995;152(9):1437-42.

28. Berg M. Patient care information systems and health care work: a sociotechnical approach. Int J Med Inf 1999;55(2):87-101.

29. Hofmans-Okkes IM, Lamberts H. The International Classification of Primary Care (ICPC): new applications in research and computer-based patient records in family practice. Family Practice 1996;13(3):294-302.

30. Brazy JE, Langkamp DL, Brazy ND, De Luna RF. Do primary care physicians prefer dictated or computer-generated discharge summaries? Am J Dis Child 1993;147(9):986-8.

31. Cox JL, Melady MP, Chen E, Naylor CD. Towards improved coding of acute myocardial infarction in hospital discharge abstracts: a pilot project. Canadian Journal of Cardiology 1997;13(4):351-8.

32. Delamarre D, Burgun A, Seka LP, Le Beux P. Automated coding of patient discharge summaries using conceptual graphs. Methods of Information in Medicine 1995;34(4):345-51.

33. Adams DC, Bristol JB, Poskitt KR. Surgical discharge summaries: improving the record. Ann R Coll Surg Engl 1993;75(2):96-9.

34. Smith RP, Holzman GB. The application of a computer data base system to the generation of hospital discharge summaries. Obstet Gynecol 1989;73(5 Pt 1):803-7.

35. Lorenzoni L, Da Cas R, Aparo UL. The quality of abstracting medical information from the medical record: the impact of training programmes. International Journal for Quality in Health Care 1999;11(3):209-13.

36. Arts DG, Bosman RJ, de Jonge E, Joore JC, de Keizer NF. Training in data definitions improves quality of intensive care data. Crit Care 2003;7(2):179-84.

APPENDIX A

Strategy to Develop a Portable Booklet with a List of Pediatric Diagnoses in Alphabetic Order

1. Diagnoses used at the pediatric department of the Leiden University Medical Centre from 1978 to 1992 were taken as a starting point. This set encompassed about 1,700 different ICD-9-CM-based codes. Each diagnosis code was provided with a subspecialty code.
2. A draft AMC children's list was composed by selecting the matching ADC diagnoses. Subspecialty codes of the Leiden children's list were also incorporated.
3. Using the subspecialty codes, the draft AMC list was divided into subspecialty lists. Every subspecialty criticized its own list with the following comments: (a) diagnosis has to be made more specific; (b) diagnosis has to be formulated differently; (c) diagnosis can be omitted; and (d) missing diagnosis has to be added.
4. The subspecialties pediatric cardiology, oncology, and genetics had their own, non ICD-9-CM-based, classifications. Diagnoses in these classifications were translated to the ADC. Where diagnoses were more specific than those of the ADC, ADC codes were extended. The translated lists were finalized with comments from the subspecialties concerned.
5. Lists from steps three and four were combined, resulting in a list of 2,480 diagnoses. In order to make it easy for pediatricians to find diagnoses, syntax rules were applied and led to clustering of diagnoses, for example, diagnoses that relate to meningitis (Figure 3). For oncology, the topography (T) codes and associated morphology (M) codes were listed separately. Overlapping and uncommon diagnoses were omitted. The booklet contained 1,560 diagnoses, of which 468 were newly specified.

CHAPTER 4

EFFECT OF DISCHARGE LETTER-LINKED DIAGNOSIS REGISTRATION ON DATA QUALITY

Hilco Prins, Hans Büller, Bertie Zwetsloot-Schonk

International Journal for Quality in Health Care. 2000 Feb;12(1):47-57

ABSTRACT

Objective: Diagnostic data are essential for the assessment of medical practice: they are needed for retrieval of clinical cases and describing co-morbidity and complications. In most Western countries, diagnosis registration in hospital information systems is based mainly on completing forms after patient discharge. As this registration plays no role in patient care, data quality is usually unsatisfactory. To improve data quality, we redesigned the process of diagnosis registration at a pediatric department, and now pediatricians provide diagnoses with codes in a separate registration heading of the discharge letter. We compared the quality of this discharge letter-linked diagnosis registration to the quality of the previous form-based registration.

Design: Retrospective study with blinded before-after measurement. Re-abstracted diagnosis descriptions of the text of discharge letters were taken as gold standard.

Setting: A pediatric department in an academic medical center

Study participants: From each registration period, 60 admissions were selected randomly. Mean age of the patients was 4.5 (SD $\pm$ 5.5) and 5.2 (SD $\pm$ 5.2) years for the old and new situation respectively. Mean length of stay was 8.8 (SD $\pm$ 11.0) and 7.2 (SD $\pm$ 12.4) days.

Intervention: Discharge letter-linked diagnosis registration

Main outcome measures: Completeness and correctness, both at three-digit level of ICD-9-CM

Results: Completeness of form-based diagnosis registration was 51% (95% CI, 44-58%) and of discharge letter-linked diagnosis registration 54% (95% CI, 47-60%). Correctness was 65% (95% CI, 58-72%) and 67% (95% CI, 60-74%) respectively.

Conclusions: The discharge letter-linked diagnosis registration does not provide a better basis for assessment of medical practice than the form-based diagnosis registration.

Keywords: abstracting and indexing, data collection, diagnosis, patient discharge, pediatrics, quality assurance

4.1 INTRODUCTION

An important method of improving quality of care within hospitals is assessment of medical care for patient groups [1]. When using a hospital information system (HIS) for assessment [2-4], the recorded diagnostic data must be of high quality [5, 6] in order to retrieve cases [7-11] and describe co-morbidity and complications [5, 12-17]. Diagnostic data form the framework within which the care provided can be assessed.

Important aspects of diagnostic data quality are completeness and correctness. Completeness is the proportion of true diagnoses that are recorded; correctness is the proportion of recorded diagnoses that are true. When retrieving cases, incompleteness leads to false negatives and incorrectness to false positives. Specificity and timeliness also determine usefulness of diagnostic data for assessment. Specificity in this study is the proportion of correctly recorded diagnoses that contain all available diagnostic information and timeliness is the time interval between recognizing and recording diagnoses.

Stimulated by the institution of the Dutch Quality of Health Care Institutions Act [18], pediatricians of the Academic Medical Center (AMC), Amsterdam, decided to assess their medical practice using routinely collected patient data [19]. The AMC is a university hospital and the pediatric department consists of 155 beds. A general feeling of unease arose among the pediatricians about current diagnostic data quality. As in most Western hospitals, collecting diagnostic data in the AMC is based on forms, completed by physicians, after patient discharge. Subsequently, medical record coders encode the diagnosis descriptions according to ICD-9-CM and record codes in HIS. The pediatricians stated that this procedure negatively influences data quality. First, the diagnostic data do not play a role in daily patient care. This leads to a situation in which completing discharge-forms is not given high priority. Second, medical record coders are unaware of additional diagnostic facts that might influence selection of appropriate codes.

Studies in several countries show that completeness of routinely collected diagnostic data in hospitals varies from 0.50 to 0.90 [10, 11, 20-38] and correctness from 0.30 to 0.95 [9-11, 20-26, 28-33, 35-52]. These results depend on study design, registration use and registration process.

Because of differences in study design, comparability is limited. Some studies use a disease specific registry as gold standard [20, 22, 28-30, 32], other studies use medical

records, re-abstracted or not. Some studies are limited to one disease [20, 22, 28, 29, 31, 33, 36, 37, 39, 42-45, 52] whereas others take a broader domain. Differences are also found in setting, limitation of diagnostic categories, and operational definition of completeness and correctness. It seems that a disease specific registration as gold standard [20, 22, 28-30, 32] and a limitation to severe diseases and principal diagnosis [24] lead to better diagnostic data quality. When diagnostic data are used for financial compensation, data quality is higher than without this use, but problems still remain [53, 54].

Notable is that *principal diagnosis*, *secondary diagnosis*, *complications* and *nature of injury* are recorded, but no study mentions *reason for admission*. However, for assessment of medical practice particularly *reason for admission* is important, because medical activities should also be judged from this perspective.

Some studies were limited to the measurement of correctness [11, 39, 42, 44]. In these studies, cases were retrieved based on specific ICD-9-CM codes and then compared to the gold standard. In this sort of study completeness cannot be measured. However, correctness without completeness is of limited use in assessing data quality. Specificity (level of detail of the code) is almost never measured or is implicitly part of correctness [41]. Timeliness is never measured.

Many studies conclude that form-based diagnostic discharge data should not be used, or should be used with great caution, for quality of care measurement. Several suggestions have been formulated on how to improve data quality, such as systematic audit of data quality [20, 24, 41, 42], information feedback to physicians [21] and education of physicians [20, 22, 42]. However, few intervention studies comparing different registration procedures have been performed. Yeoh [38] implemented physician encoding at a pediatric department. Correctness increased from 0.54 to 0.85. Hohnloser [27] implemented computer based registration in daily care process at an intensive / critical care unit; here, completeness increased from 0.48 to 0.82. However, the method was not fully appropriate as no gold standard was used: numbers of codes were compared to numbers of diagnoses in the discharge summary.

To improve data quality at the pediatric department of the AMC, we redesigned diagnosis registration. After implementation in routine practice we tested our hypothesis that integration of the diagnosis registration with communication about patients combined with physician encoding, improves completeness, correctness, specificity and timeliness of diagnostic data.

4.2 METHODS

An intervention study was performed with blinded before and after measurement.

4.2.1 Intervention

The old situation is described in the introduction and reported graphically in Figure 1a. The new situation is reported in Figure 1b. In this situation pediatricians themselves encode diagnoses. After discharge, standardized descriptions of diagnoses with codes are reported in a medical registration heading at the bottom of the discharge letter. This letter plays an important role in communication between pediatrician and general practitioner. Besides *principal diagnosis*, *secondary diagnosis*, *complications* and *nature of injury*, also *reason for admission* is recorded. The letter with this heading is checked by one medical record coder and one supervisor. In order to support pediatricians, a pediatric diagnosis booklet has been developed in close consultation with the pediatricians. This booklet has an alphabetical list of selected and further specified ICD-9-CM codes with descriptions. Six digits are available for a code. This means that compared to ICD-9-CM, one to three digits are available for local extensions.

In two meetings, pediatricians and residents were instructed. During the first months after introduction, regular deliberations between medical record coder and pediatricians took place. Furthermore, electronic versions of the letters were made available on the Intranet. Consequently the new registration facilitates patient information retrieval in case of readmission. A mechanism was implemented to remind pediatricians to write discharge letters within 6 weeks, which is policy in the AMC. Participating specialties were general pediatrics and seven pediatric subspecialties (Table 1). The new registration started February 1, 1995. Up to the time of writing approximately 7000 discharge letters in the new style have been produced.

4.2.2 Case selection

We estimated both completeness and correctness in the old situation to be 0.65. In order to use data for medical practice assessment we judged that completeness and correctness should be at least 0.90. With a power of 0.80 52 cases per group would be needed to demonstrate this meaningful difference [55]. We randomly selected 60 admissions with discharge dates between September and December 1994 (sample ‘94) and 60 admissions with discharge dates between September and December

1995 (sample '95) with electronic versions of the discharge letter. By choosing 'September to December' for both registrations, seasonal influences were avoided. The sampling frame consisted of admissions with only one responsibility period for which one of the participating subspecialties bore responsibility. A responsibility period is a period during which one medical specialty carries the principal responsibility for medical care. Registration of diagnoses takes place after every period of responsibility.

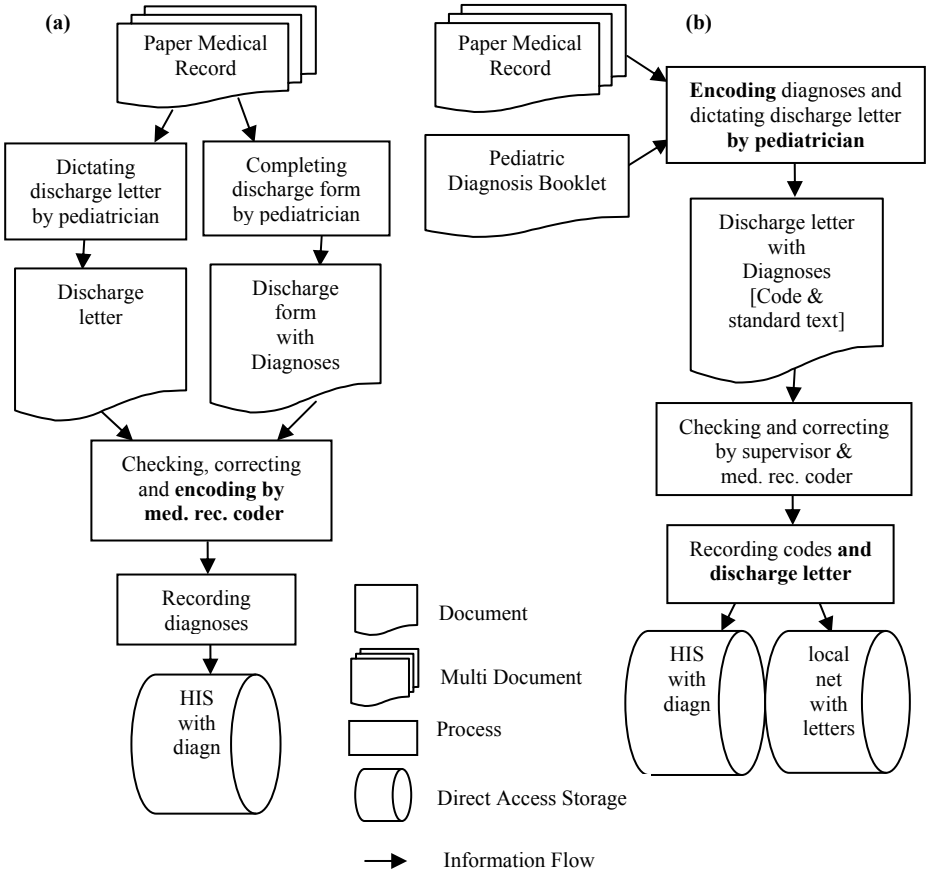

Figure 1 (a): Form-based diagnosis registration. **(b)**: Discharge letter-linked diagnosis registration.

4.2.3 Data Collection

For each case selected the following data were collected: patient's age and sex, length of stay, number of words in discharge letter, whether the letter was written by pediatrician or resident, responsible subspecialty, admission day and diagnosis codes in HIS with category indication.

4.2.4 Gold standard

The text of the discharge letter was basis for the gold standard. In the electronic versions of the letters of 1995 we antedated all dates by 1 year and removed the medical registration heading for blinding. One pediatrician marked parts of the text referring to relevant diagnoses. The pediatrician also stated to what category (*reason for admission*, *principal diagnosis*, *secondary diagnosis*, or *complication*) each diagnosis belonged. Rules for diagnosis marking were formulated (Box 1).

Box 1: Rules for diagnosis marking by pediatrician.

1. Read letter.
2. Read letter again and mark relevant diagnoses.
3. Indicate per diagnosis whether it is *reason_for_admission, principal_diagnosis*, *secondary_diagnosis* or *complication*.
4. The following definitions and procedures apply:

Reason_for_admission: Diagnosis, symptom, sign or injury that, at admission day, was considered as *reason_for_admission*.
Procedure: Code one diagnosis as *reason_for_admission* with '*R'*. If more than one *reason_for_admission*, indicate also most important. *Reason_for_admission* may also be *principal_diagnosis*.
Principal_diagnosis: Diagnosis that, at discharge, is most important reason for treatment.
Procedure: Code one diagnosis as *principal_diagnosis* with '*P'*. It is allowed that a diagnosis is *reason_for_admission ('R')* and *principal_diagnosis ('P')*.
Secondary_diagnosis: Every relevant diagnosis other than *reason_for_admission*, *principal_diagnosis* or *complication*
Procedure: Code up to five diagnoses as *secondary_diagnosis* with *'S'*. If more than five secondary diagnoses, indicate five most important.
Complication: Diagnosis that developed as result of hospital stay or received treatment.
Procedure: Code up to three diagnoses as *complication* with *'C'*. If more than three complications, indicate three most important. Handle definition strictly. In case of doubt choose *secondary_diagnosis*.

One blinded expert medical record coder checked whether every diagnosis had been recorded in HIS and whether diagnosis code in HIS could be recognized in the corresponding letter. Rules were formulated for the expert coder (Box 2): first he coded the marked diagnosis descriptions twice - once according to the list of locally extended codes available in 1994 and once according to the list available in 1995. HIS codes were already presented on the evaluation forms. Then, for every admission the coder matched, at the 3-digit level of ICD-9-CM, his own codes with HIS codes. Alternative codes could be correct. If there was a match, the coder checked whether the match was also true at 6-digit level of ICD-9-CM and whether the diagnostic category was correctly indicated in the HIS.

Box 2: Rules for checking diagnoses by expert medical record coder.

1. Read letter.
2. Code every marked diagnosis description as specifically as possible: once according to the list of locally extended codes available in 1994, and once according to the list available in1995.
3. Check correctness of HIS codes by comparing them with your own codes at 3-digit level of ICD-9-CM. If correct, tick the appropriate HIS code on evaluation form. Alternative codes may also be correct!
4. If answer in 3 is true, check specificity of every code in HIS at 4- to 6-digit level: for 1994 and for 1995.
5. Check whether faults have been made at 4- to 6-digit level.
6. Check whether HIS code was classified in the correct category.
7. If HIS codes on evaluation form not ticked but letter consists diagnostic information compatible with code, mark relevant text and copy text to evaluation form: repeat rules 2 to 6.

4.2.5 Measures

We made operational and formalized quality aspects described in the *Introduction* as follows:

Completeness is the proportion of marked diagnoses in discharge letters that, at 3-digit level of ICD-9-CM, are coded in the corresponding admission records of HIS.

$$\text{Completeness } (cat) = \frac{\sum_{i=1}^{n} \left| l(D_{i,cat}) \cap l(C_i) \right|}{\sum_{i=1}^{n} \left| D_{i,cat} \right|}$$

where: ALL_CATEGORIES = {reason_for_admission, principal_diagnosis, secondary_diagnosis}

$cat \subseteq$ ALL_CATEGORIES

D_i,cat : set of diagnosis codes obtained by re-coding the marked diagnosis descriptions that belong to *cat* in discharge letter of admission *i*, e.g. $D_i,\{reason\_for\_admission\}$ is the set of diagnosis codes obtained by re-coding the marked reason for admission descriptions of admission *i* and $D_i,ALL\_CATEGORIES$ is the set of diagnosis codes obtained by re-coding all marked diagnosis descriptions of admission *i*

C_i : set of diagnosis codes in the HIS record of admission *i*
n : number of selected admissions

l : function that returns a set of 3-digit level ICD-9-CM codes, e.g. *l*({786.010, 490.000}) = {786, 490}

$\left| S \right|$: number of elements in set S, e.g. $\left|$ {786.010, 490.000} $\right| = 2$

Correctness is the proportion of diagnosis codes in the HIS that, at 3-digit level of ICD-9-CM, have matching diagnosis descriptions in corresponding discharge letters.

$$\text{Correctness } (cat) = \frac{\sum_{i=1}^{n} \left| l(D_i) \cap l(C_i, cat) \right|}{\sum_{i=1}^{n} \left| C_i, cat \right|}$$

where: D_i : set of diagnosis codes obtained by re-coding marked diagnosis descriptions in discharge letter of admission *i*

C_i,cat : set of diagnosis codes that belong to *cat* in the HIS record of admission *i*, e.g. $C_i,\{reason\_for\_admission\}$ is the set of reason for admission codes in the HIS record of admission *i* and $C_i,ALL\_CATEGORIES$ is the set of all diagnosis codes in the HIS record of admission *i*

Specificity is the proportion of correct HIS codes that contain, as far as possible with the local 6-digit codes, all additional diagnostic information in the text of corresponding discharge letters.

$$\text{Specificity } (cat) = \frac{\sum_{i=1}^{n} \left| m(D_i) \cap m(C_i, cat) \right|}{\sum_{i=1}^{n} \left| l(D_i) \cap l(C_i, cat) \right|}$$

where: m : function that returns a set of 6-digit level local extended ICD-9-CM codes available at time of recording the codes, e.g. $m(\{786.010, 490.000\}) = \{786.010, 490.000\}$

Timeliness is the proportion of admissions with time-intervals less than six weeks between dates of discharge and recording codes in HIS. This criterion is derived from the policy to write discharge letters within six weeks.

$$\text{Timeliness} = \frac{\left| \{ adm \in ADMISSIONS \mid t(adm) \leq 6 \} \right|}{n}$$

where: *ADMISSION* : set of all admissions

t(adm) : time interval (in weeks) between discharge and recording codes in HIS for admission *adm*

As complications are not recorded as such, but as *secondary diagnoses*, no distinction has been made between these categories in the analysis. It soon became clear that physicians do not know the exact definition of *complication* and rarely use the term because of negative associations.

4.2.6 Statistical analysis

To compare patient and admission characteristics, we used the z-test and t-test with continuity correction for dichotomous variables, respectively continuous variables. The z-test was also used to compare results of both registrations. The X^2 test was used to compare both groups with regards to day of admission and treating specialty. The Mann-Whitney test was used to compare age and length of stay.

4.3 RESULTS

Case Selection

From participating specialties, 512 patients were discharged from September to December 1994. Of these admissions, 470 (92%) had one responsibility period with a mean of 2.57 (SD $\pm$ 0.98) diagnoses in the HIS. To select 60 admissions with electronic discharge letters, we had to retrieve randomly 173 admissions. From September to December 1995, the participating specialties discharged 582 patients. Of these admissions, 535 (92%) had one responsibility period with a mean of 2.87 (SD $\pm$ 1.29) diagnoses in the HIS. To select 60 admissions with electronic discharge letters, we retrieved randomly 92 admissions.

Table 1: Patient and admission characteristics of selected cases.

Variable	**Sample '94** (N = 60)	**Sample '95** (N = 60)
Mean age (years ± SD)	4.5 ± 5.5	5.2 ± 5.2
Sex: male/female	30/30	37/23
Mean length of stay (days ± SD)	8.8 ± 11.0	7.2 ± 12.4
Mean number of words/letter (± SD)	528.4 ± 289.0	472.8 ± 199.9
Number of letters written by (%):		
Pediatrician	2 (3)	4 (7)
Resident	58 (97)	56 (93)
Number in subspecialty (%):		
General pediatrics	38 (63)	37 (62)
Pediatric gastro-enterology	7 (12)	2 (3)
Pediatric nephrology	6 (10)	8 (13)
Pediatric hematology and immunology	4 (6)	1 (2)
Pediatric pulmonology	4 (6)	3 (5)
Pediatric metabolic disorders	1 (2)	4 (7)
Pediatric cardiology	0 (0)	4 (7)
Pediatric endocrinology	0 (0)	1 (2)
Number of admissions per day (%):		
Saturday	6 (10)	4 (7)
Sunday	1 (2)	4 (7)
Monday	9 (15)	7 (12)
Tuesday	12 (20)	15 (25)
Wednesday	11 (18)	10 (17)
Thursday	14 (23)	11 (18)
Friday	7 (12)	9 (15)

Patient and admission characteristics

Table 1 shows patient and admission characteristics of the samples. There were no statistically significant differences.

Number of diagnoses in discharge letter and HIS

Table 2 shows the number of marked diagnoses in discharge letters and the number of diagnoses in HIS. Initially, the re-abstractor marked 117 and 104 reasons for admission for sample ’94 and sample ’95 respectively. However, per admission only the most important reason for admission was taken into account. For both samples, the number of diagnoses in discharge letters is higher than the number in the HIS. The difference is explained by the difference in number of secondary diagnoses.

Table 2: Number of diagnoses in discharge letters and HIS.

	Number of diagnoses (mean)			
	Sample ’94 (n = 60)		**Sample ’95** (n = 60)	
Diagnostic category	***Letters***	***HIS***	***Letters***	***HIS***
All categories	209 (3.48)	165 (2.75)	221 (3.68)	177 (2.95)
Reason for admission	60 (1.00)	60 (1.00)	60 (1.00)	60 (1.00)
Principal diagnosis	60 (1.00)	60 (1.00)	60 (1.00)	60 (1.00)
Secondary diagnosis	89 (1.48)	45 (0.75)	101 (1.68)	57 (0.95)

Completeness

Completeness of the two methods of diagnosis registration is presented in Table 3. In the definition used, category indication in the letter and category indication in the HIS are not necessarily the same. If we make this a requirement, the proportions in Table 3 are 0.03 to 0.10 lower. There are no statistically significant differences in completeness between the two methods of registration. Notable is that completeness of *secondary diagnoses* is lower than completeness of *reason for admission* and *principal diagnoses*.

In only 27% (95% CI, 16-40) of the admissions, all marked discharge letter diagnoses were recorded in the HIS. This is true for the old and the new method of registration.

Table 3: Completeness of diagnosis registrations.

	Completeness diagnosis registration (95% CI)	
Diagnostic category in discharge letter	**Sample '94** (n = 60) ***(Form-based)***	**Sample '95** (n = 60) ***(Discharge letter-linked)***
All categories	107/209=0.51 (0.44-0.58)	119/221=0.54 (0.47-0.60)
Reason for admission	35/60=0.58 (0.45-0.71)	34/60=0.57 (0.43-0.69)
Principal diagnosis	37/60=0.62 (0.48-0.74)	40/60=0.67 (0.53-0.78)
Secondary diagnosis	35/89=0.39 (0.29-0.50)	45/101=0.45 (0.35-0.54)

Correctness

Correctness of diagnosis registrations is given in Table 4. Also in this definition category indication of a diagnosis code in the HIS and category indication of corresponding diagnosis in the letter are not necessarily the same. If this demand is made, proportions in Table 4 are 0.03 to 0.09 lower. There are no statistically significant differences between the registrations.

Table 4: Correctness of diagnosis registrations.

	Correctness diagnosis registration (95% CI)	
Diagnostic category in HIS	**Sample '94** (n = 60) ***(Form-based)***	**Sample '95** (n = 60) ***(Discharge letter-linked)***
All categories	107/165=0.65 (0.58-0.72)	119/177=0.67 (0.60-0.74)
Reason for admission	35/60=0.58 (0.45-0.71)	36/60=0.60 (0.47-0.72)
Principal diagnosis	41/60=0.68 (0.55-0.80)	43/60=0.72 (0.59-0.83)
Secondary diagnosis	31/45=0.69 (0.53-0.82)	40/57=0.70 (0.57-0.82)

In the form-based as well as in the discharge letter-linked diagnosis registration 58 incorrect codes were found (Table 5). From the perspective of correctness one may denote *nature of incorrectness* '3' to '5' (Table 5) as correct. This leads to overall correctness of 0.78 (95% CI, 0.72-0.85) for form-based and 0.77 (95% CI, 0.71-0.83) for discharge letter-linked registration. Consequently, this leads to overall completeness of 0.62 (95% CI: .55-.68) for form-based and also 0.62 (95% CI: .55-.68) for discharge letter-linked registration. It appeared that most inaccuracies stem from hasty and imprecise completion of the form (in the old situation) or medical registration heading (in the new situation) by pediatricians.

Table 5: Analysis of incorrect codes.

Nature of incorrectness	Number of Incorrect Codes	
	Form-based	*Discharge letter-linked*
1. Code from correct chapter of ICD-9-CM, but not consistent with patient's disease, e.g. Acute Bronchitis, viral (466.0) instead of Viral Pneumonia, unspecified (480.9)	17	8
2. Code related to patient's disease, but not justifying patient's problem, e.g. Respiratory malfunction arising from mental factors (306.1) instead of Hyperventilation (786.01)	8	13
3. Incorrect double recording of code as *reason for admission* and *principal diagnosis*	16	11
4. Incorrect double recording of code as *secondary diagnosis* and *complication*	2	1
5. Unnecessary code, e.g. Functional digestive disorders, not elsewhere classified (564), while in history slight constipation irrelevant for admission	4	5
6. No relationship between code and patient's problems	11	12
7. Diagnosis information received too late (special code for this)	0	8
Total	58	58

With regards to the form-based diagnosis registration, in 32% (95% CI, 20-45) of the admissions, all diagnosis codes in the HIS match with diagnosis descriptions in discharge letters. For discharge letter-linked registration this is 40% (95% CI, 28-54).

Specificity

In Table 6 specificity of both registrations is presented. The form-based registration has higher rates, albeit not statistically significant. Specificity was assessed against possibilities of the list of locally extended codes at the moment of use. For example, a patient had Laryngo-tracheo-bronchitis: in 1994, code 490.000 (Bronchitis, inclusive Tracheobronchitis NOS) was specific at 6-digit level. In 1995, code 490.002 (Laryngo-tracheo-bronchitis) was added to the list. This means that in 1995, code 490.000 was not specific for the patient.

Timeliness

Time interval between discharge and recording diagnosis codes in the HIS are presented in Figure 2. The proportion of admissions for which it is true that diagnoses are recorded within six weeks after discharge (our definition of timeliness) are 0.47 (95% CI, 0.34-0.60) and 0.53 (95% CI, 0.40-0.66) for form-

based and letter-linked registration respectively. However, after 24 weeks all letters are written in the old situation and only 87% in the new situation. This unexpected result can be explained as follows. Before February 1995 letters were either produced soon after discharge or not at all. From February 1995 letters that were not produced soon after discharge, were, due to the reminders, still written after a relatively long interval.

Table 6: Specificity of diagnosis registrations.

	Specificity diagnosis registration (95%CI)	
Diagnostic category	**Sample '94** (n = 60) ***(Form-based)***	**Sample '95** (n = 60) ***(Discharge letter-linked)***
All categories	101/107=0.94 (0.88-0.98)	100/118=0.85 (0.78-0.91)
Reason for admission	32/35=0.91 (0.77-0.98)	30/36=0.83 (0.67-0.94)
Principal diagnosis	40/41=0.96 (0.87-1.00)	36/43=0.84 (0.69-0.93)
Secondary diagnosis	29/31=0.94 (0.79-0.99)	34/39=0.87 (0.73-0.96)

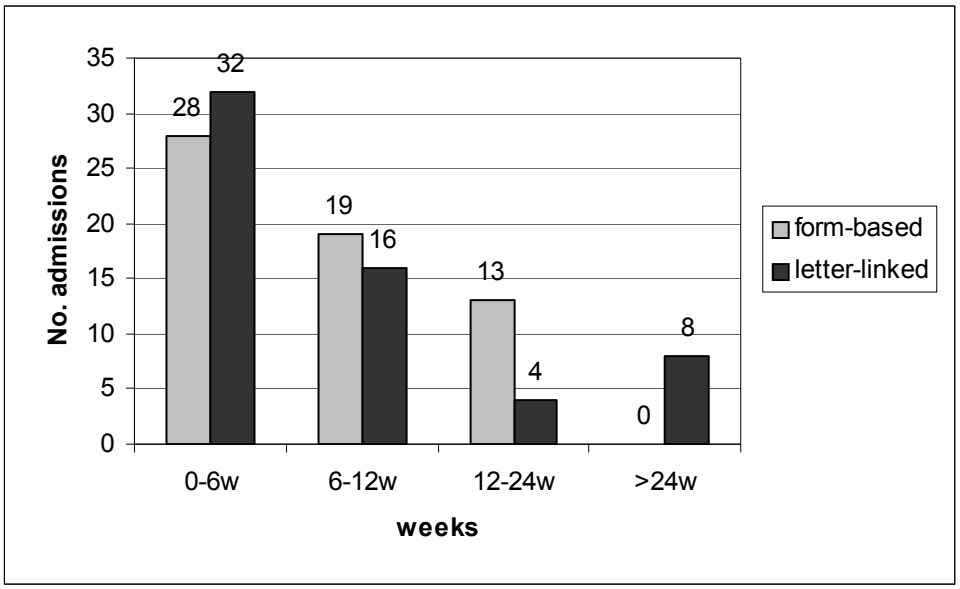

Figure 2: Time interval between discharge and recording diagnostic codes.

4.4 DISCUSSION

Diagnostic data of high quality are essential for the assessment of medical practice. To improve data quality in a hospital information system, we redesigned diagnosis registration at a pediatric department. We compared quality of the discharge letter-linked diagnosis registration to quality of the previous form-based registration. We performed a retrospective study with blinded before and after measurement. Diagnostic information in the text of the discharge letter was taken as gold standard after separate and independent abstraction. At the 3-digit level of ICD-9-CM, completeness of the discharge letter-linked diagnosis registration was 54% (95% CI, 47-60) and correctness 67% (95% CI, 60-74). This was similar to the previous form-based registration.

4.4.1 Study design

For each of the two observation periods 60 admissions were selected, each of the two samples having about two hundred diagnoses. It is clear that no meaningful improvement occurred.

The selection procedure may have introduced selection bias. In order to identify 60 appropriate cases, far more admissions had to be retrieved from the form-based registration than from the discharge letter-linked registration. In the period of the form-based registration more often a letter was not written after discharge than in the period of the discharge letter-linked registration. However, comparison of both samples on important patient and admission characteristics revealed no relevant differences.

We measured quality of both registrations in cases there was an electronic discharge letter available. Quality of both registrations in cases where there were no electronic discharge letters remains unknown, but we think that it is more likely to be worse than better. As the proportion of available discharge letters increased, the overall diagnostic data quality in the new situation is likely to be improved.

We constructed a gold standard by re-abstracting diagnoses descriptions from the text of discharge letters. Because this re-abstracting process is more thorough than routine abstracting process, re-abstracted diagnoses are assumed to be correct [56]. A prerequisite is strict control on re-abstracting, as registration quality is wholly dependent upon validity of the re-abstracted data. We did this by formulating rules for the pediatrician who re-abstracted diagnoses.

Whether text of discharge letters is a good basis for a gold standard can be discussed. In The Netherlands, the discharge letter is an important tool in communication between medical specialist and general practitioner, since the general practitioner has a pivotal role in patient care. At the department where this study was performed, discharge letter quality has received much attention for many years. The letter gives a complete outline of the admission and has a fixed structure. Apart from the medical registration heading there are no differences between letters of both periods. An advantage of using the discharge letter as the gold standard, over using the paper medical record, is that the electronic version of the discharge letter provides opportunities for blinding for the period and thus for the registration method.

We found no other studies in literature in which two methods of diagnosis registration were compared and re-abstractor and reviewer were blinded for registration method.

Inter-coder variability is a well-known phenomenon [49, 50]. We assumed that the expert reviewer was capable of assessing whether an alternative code was acceptable or not. The rules we made for the review process supported this. It is unlikely that the Hawthorne effect played a role in the study. Pediatricians were not told that the registrations would be evaluated. Moreover, in the included periods no extra attention was given to registration process.

The criterion of matching at 3-digit level of ICD-9-CM is rather crude. An incorrectly recorded code and a true code may both belong to the same group of diseases - e.g. recorded code is 462xxx (Acute Pharyngitis), but true code is 466xxx (Acute Bronchitis): both belong to “Acute respiratory Infections”. But it can also mean that recorded and true code belong to completely different groups of diseases, e.g. recorded code is 462xxx, but true code is 458xxx (Hypotension). In Table 5 we provide information on the degree of incorrectness.

4.4.2 Consequences for assessment of medical practice

Patient retrieval based on diagnostic information in the HIS will result in false positive and false negative cases. In order to remove irrelevant cases, verification based on medical records or discharge letters is necessary. Afterwards, re-abstracting diagnoses from medical records or discharge letters is important to get a valid idea of patients’ diseases. These verification and re-abstracting activities are time consuming.

In patient retrieval there is typically a tradeoff between *precision* (the proportion of retrieved patients that are relevant) and *recall* (the proportion of relevant patients retrieved) [57]. Precision is a function of registration correctness and choice of codes by which patients are retrieved. Recall is a function of registration completeness and choice of codes by which patients are retrieved. When only one disease-specific code is used to retrieve patients, precision will be relatively high and recall will be relatively low. The use of more, disease-related, codes will lead to decreased precision and increased recall.

It is likely that severe cases of a disease are more often coded than mild cases. This means that the retrieved patients may not represent the whole group.

4.4.3 Improving diagnostic data

In the HIS, per admission one *reason for admission* can and must be recorded. During the re-abstracting process it appeared that often more than one *reason for admission* is relevant, e.g. a complex of signs and symptoms or a list of differential diagnoses. We therefore advocate the possibility to record more than one reason for admission.

Our hypothesis that linking diagnosis registration to the discharge letter would improve registration quality could not be demonstrated. The promise of these data to have a communication and patient information retrieval function, and possibly to have a function in quality assessment, management and research, is no sufficient incentive. Maybe it is the case that diagnostic data recorded after discharge of patients are not suitable for the assessment of medical practice for patient groups.

Probably the only way to improve the quality of diagnostic data is to incorporate the registration in the daily care process. Finland is unique in that diagnosis codes are recorded by physicians in paper medical records during the care process [29]. In other countries arrival of computerized patient records will provide the opportunity to implement rigorous quality improvement of diagnosis registration. However, early studies indicate that computerized patient records alone is a valuable but insufficient condition for high quality data [58]. If we wish to evaluate daily care with routinely collected patient data we have to accept that these studies do not meet high standards required for scientific clinical research.

ACKNOWLEDGEMENT

We thank Peter van Halem for assessing the diagnoses, and Ameen Abu-Hanna and Martien Limburg for their comments on the manuscript.

REFERENCES

1. McIntyre N. Evaluation in clinical practice: problems, precedents and principles. J Eval Clin Pract 1995;1(1):5-13.
2. Safran C, Chute CG. Exploration and exploitation of clinical databases. Int J of Biomed Comput 1995;39(1):151-6.
3. McKee M. Routine data: a resource for clinical audit? Quality in Health Care 1993(2):104-11.
4. McKee M, James P. Using routine data to evaluate quality of care in British hospitals. Med Care 1997;35(10 Suppl):OS102-11.
5. Iezzoni LI, Foley SM, Daley J, Hughes J, Fisher ES, Heeren T. Comorbidities, complications, and coding bias. Does the number of diagnosis codes matter in predicting in-hospital mortality? [see comments]. JAMA 1992;267(16):2197-203.
6. Wyatt J. Acquisition and use of clinical data for audit and research. J Eval Clin Pract 1995;1(1):15-27.
7. Henry OA, Gregory KD, Hobel CJ, Platt LD. Using ICD-9 codes to identify indications for primary and repeat cesarean sections: agreement with clinical records. Am J Public Health 1995;85(8 Pt 1):1143-6.
8. Cherkin DC, Deyo RA, Volinn E, Loeser JD. Use of the International Classification of Diseases (ICD-9-CM) to identify hospitalizations for mechanical low back problems in administrative databases. Spine 1992;17(7):817-25.
9. Campos-Outcalt DE. Accuracy of ICD-9-CM codes in identifying reportable communicable diseases. Qual Assur Util Rev 1990;5(3):86-9.
10. Pladevall M, Goff DC, Nichaman MZ, et al. An assessment of the validity of ICD Code 410 to identify hospital admissions for myocardial infarction: The Corpus Christi Heart Project [see comments]. Int J Epidemiol 1996;25(5):948-52.
11. Raiford DS, Perez Gutthann S, Garcia Rodriguez LA. Positive predictive value of ICD-9 codes in the identification of cases of complicated peptic ulcer disease in the Saskatchewan hospital automated database. Epidemiology 1996;7(1):101-4.
12. Deyo RA, Cherkin DC, Loeser JD, Bigos SJ, Ciol MA. Morbidity and mortality in association with operations on the lumbar spine. The influence of age, diagnosis, and procedure. J Bone & Joint Surgery - Am Vol 1992;74(4):536-43.
13. Siegler EL, Stineman MG, Maislin G. Development of complications during rehabilitation. Arch Intern Med 1994;154(19):2185-90.

14. Faciszewski T, Johnson L, Noren C, Smith MD. Administrative databases' complication coding in anterior spinal fusion procedures. What does it mean? Spine 1995;20(16):1783-8.

15. Iezzoni LI, Daley J, Heeren T, et al. Identifying complications of care using administrative data. Med Care 1994;32(7):700-15.

16. Romano PS, Roos LL, Jollis JG. Adapting a clinical comorbidity index for use with ICD-9-CM administrative data: differing perspectives. J Clin Epidemiol 1993;46(10):1075-9; discussion 1081-90.

17. Hoore DW, Bouckaert A, Tilquin C. Practical considerations on the use of the Charlson comorbidity index with administrative data bases. J Clin Epidemiol 1996;49(12):1429-33.

18. Ministry of Public Health. The Quality of Health Care Institutions Act. The Hague: SDU Uitgeverij, 1996.

19. Büller HA, Zwetsloot-Schonk JHM. [Project proposal Department Emma Children Hospital / Children AMC: "Information for evaluation of patientcare"]. Amsterdam: Academic Medical Center, 1992:

20. Chancellor AM, Swingler RJ, Fraser H, Clarke JA, Warlow CP. Utility of Scottish morbidity and mortality data for epidemiological studies of motor neuron disease. J Epidemiol Community Health 1993;47(2):116-20.

21. Colin C, Ecochard R, Delahaye F, et al. Data quality in a DRG-based information system. Int J Qual Health Care 1994;6(3):275-80.

22. Davenport RJ, Dennis MS, Warlow CP. The accuracy of Scottish Morbidity Record (SMR1) data for identifying hospitalised stroke patients. Health Bull (Edinb) 1996;54(5):402-5.

23. Dixon J, Sanderson C, Elliott P, Walls P, Jones J, Petticrew M. Assessment of the reproducibility of clinical coding in routinely collected hospital activity data: a study in two hospitals. J Public Health Med 1998;20(1):63-9.

24. Fisher ES, Whaley FS, Krushat WM, et al. The accuracy of Medicare's hospital claims data: progress has been made, but problems remain. Am J Public Health 1992;82(2):243-8.

25. Geraci JM, Ashton CM, Kuykendall DH, Johnson ML, Wu L. International Classification of Diseases, 9th Revision, Clinical Modification codes in discharge abstracts are poor measures of complication occurrence in medical inpatients. Med Care 1997;35(6):589-602.

26. Green J, Wintfeld N. How accurate are hospital discharge data for evaluating effectiveness of care? Med Care 1993;31(8):719-31.

27. Hohnloser JH, Puerner F, Soltanian H. Improving coded data entry by an electronic patient record system. Methods Inf Med 1996;35(2):108-11.

28. Leibson CL, Naessens JM, Brown RD, Whisnant JP. Accuracy of hospital discharge abstracts for identifying stroke. Stroke 1994;25(12):2348-55.

29. Mahonen M, Salomaa V, Brommels M, et al. The validity of hospital discharge register data on coronary heart disease in Finland. Eur J Epidemiol 1997;13(4):403-15.

30. McClish DK, Penberthy L, Whittemore M, et al. Ability of Medicare claims data and cancer registries to identify cancer cases and treatment. Am J Epidemiol 1997;145(3):227-33.

31. Nielsen GL, Sorensen HT, Pedersen AB, Sabroe S. Analyses of data quality in registries concerning diabetes mellitus--a comparison between a population based hospital discharge and an insulin prescription registry. J Med Syst 1996;20(1):1-10.

32. Kahn LH, Blustein J, Arons RR, Yee R, Shea S. The validity of hospital administrative data in monitoring variations in breast cancer surgery. Am J Public Health 1996;86(2):243-5.

33. Pietila K, Tenkanen L, Manttari M, Manninen V. How to define coronary heart disease in register-based follow-up studies: experience from the Helsinki Heart Study. Ann Med 1997;29(3):253-9.

34. Romano PS, Mark DH. Bias in the coding of hospital discharge data and its implications for quality assessment. Med Care 1994;32(1):81-90.

35. Roos LL, Sharp SM, Cohen MM. Comparing clinical information with claims data: some similarities and differences [see comments]. J Clin Epidemiol 1991;44(9):881-8.

36. Stegmayr B, Asplund K. Measuring stroke in the population: quality of routine statistics in comparison with a population-based stroke registry. Neuroepidemiology 1992;11(4-6):204-13.

37. Vestberg K, Thulstrup AM, Sorensen HT, Ottesen P, Sabroe S, Vilstrup H. Data quality of administratively collected hospital discharge data for liver cirrhosis epidemiology. J Med Syst 1997;21(1):11-20.

38. Yeoh C, Davies H. Clinical coding: completeness and accuracy when doctors take it on [see comments]. BMJ 1993;306(6883):972.

39. Benesch C, Witter DM, Jr., Wilder AL, Duncan PW, Samsa GP, Matchar DB. Inaccuracy of the International Classification of Diseases (ICD-9-CM) in identifying the diagnosis of ischemic cerebrovascular disease. Neurology 1997;49(3):660-4.

40. George AM, Maddocks GB. Accuracy of diagnostic content of hospital activity analysis in infectious diseases. BMJ 1979;1(6174):1332-4.

41. Gibson N, Bridgman SA. A novel method for the assessment of the accuracy of diagnostic codes in general surgery. Ann R Coll Surg Engl 1998;80(4):293-6.

42. Hasan M, Meara RJ, Bhowmick BK. The quality of diagnostic coding in cerebrovascular disease. Int J Qual Health Care 1995;7(4):407-10.

43. Iezzoni LI, Burnside S, Sickles L, Moskowitz MA, Sawitz E, Levine PA. Coding of acute myocardial infarction. Clinical and policy implications. Ann Intern Med 1988;109(9):745-51.

44. Johnston ID, Bleasdale C, Hind CR, Woodcock AA. Accuracy of diagnostic coding of hospital admissions for cryptogenic fibrosing alveolitis. Thorax 1991;46(8):589-91.

45. Kennedy GT, Stern MP, Crawford MH. Miscoding of hospital discharges as acute myocardial infarction: implications for surveillance programs aimed at elucidating trends in coronary artery disease. Am J Cardiol 1984;53(8):1000-2.

46. Lindblad U, Rastam L, Ranstam J, Peterson M. Validity of register data on acute myocardial infarction and acute stroke: the Skaraborg Hypertension Project. Scand J Soc Med 1993;21(1):3-9.

47. MacIntyre CR, Ackland MJ, Chandraraj EJ, Pilla JE. Accuracy of ICD-9-CM codes in hospital morbidity data, Victoria: implications for public health research. Aust N Z J Public Health 1997;21(5):477-82.

48. MacIntyre CR, Ackland MJ, Chandraraj EJ. Accuracy of injury coding in Victorian hospital morbidity data. Aust N Z J Public Health 1997;21(7):779-83.

49. Mosbech J, Jorgensen J, Madsen M, Rostgaard K, Thornberg K, Poulsen TD. [The national patient registry. Evaluation of data quality]. Ugeskr Laeger 1995;157(26):3741-5.

50. Schwartz RJ, Nightingale BS, Boisoneau D, Jacobs LM. Accuracy of e-codes assigned to emergency department records. Acad Emerg Med 1995;2(7):615-20.

51. Smith MW. Hospital discharge diagnoses: how accurate are they and their international classification of diseases (ICD) codes? N Z Med J 1989;102(876):507-8.

52. Teikari JM, Raivio I. Validity of hospital discharge registry in acute glaucoma. Acta Ophthalmol (Copenh) 1987;65(5):589-90.

53. Assaf AR, Lapane KL, McKenney JL, Carleton RA. Possible influence of the prospective payment system on the assignment of discharge diagnoses for coronary heart disease [see comments]. N Engl J Med 1993;329(13):931-5.

54. Hsia DC, Ahern CA, Ritchie BP, Moscoe LM, Krushat WM. Medicare reimbursement accuracy under the prospective payment system, 1985 to 1988 [see comments]. JAMA 1992;268(7):896-9.

55. Dawson-Saunders B, Trapp RG. Basic and Clinical Biostatistics. Englewood Cliffs, New Yersey: Prentice-Hall, 1990.

56. Goldberg J, Gelfand HM, Levy PS. Registry evaluation methods: a review and case study. Epidemiol Rev 1980;2:210-20.

57. Rijsbergen van K. Information Retrieval. (second ed.) London: Butterworths, 1979.

58. Hogan WR, Wagner MM. Accuracy of data in computer-based patient records. J Am Med Inform Assoc 1997;4(5):342-55.

CHAPTER 5

LONG TERM IMPACT OF PHYSICIAN ENCODING SUPPORTED BY A SPECIALTY SPECIFIC LIST OF DISEASES ON DETAIL AND NUMBER OF RECORDED DIAGNOSES

Hilco Prins, Arie Hasman

Methods of Information in Medicine. 2011;50(2):115-23

SUMMARY

Objectives: To improve the recording of diagnostic discharge data, pediatricians encoded diagnostic information as part of discharge letter writing supported by a pediatric list of ICD-9-CM-based codes. We evaluated the effect of this new policy on level of detail and number of recorded diagnoses.

Methods: We compared proportions of specific principal diagnoses and numbers of secondary diagnoses of the four years before with the eight years after introduction.

Results: Immediately after introduction, half of the diagnoses for which both generic and specific codes existed, was coded specific. In later years this proportion remained stable at 0.35 ($p < 0.05$). Diagnoses that fall under the pediatrician's own subspecialty had more often a specific code than diagnoses that do not. The mean number of secondary diagnoses per admission increased from 0.7 before introduction to 1.4 in the third year after introduction ($p < 0.05$) but gradually fell back to 0.7. This increase and decrease was mainly due to diagnoses that did not fall under the pediatrician's own subspecialty. The extra codes in individual discharge summaries had added informational value.

Conclusions: Discharge letter-linked encoding by pediatricians supported by a pediatric list of diseases, leads initially to increased detail and number of diagnoses with added informational value. When attention diminishes, especially the level of detail and number of secondary diagnoses that do not fall under one's own subspecialty decrease. The level of detail of principal diagnoses remains stable because of the advantage for pediatricians of having specific diagnostic codes falling under their own subspecialty.

Keywords: Pediatrics; Physicians; Medical Record Administrators; International Classification of Diseases; Registries

5.1 INTRODUCTION

Diagnostic coding and medical practice assessment

In most Western countries, electronic diagnosis registration in hospitals is usually carried out by a medical record coder after discharge of the patient [1]. He or she encodes diagnostic information using ICD-9-CM [2] or ICD-10 [3]. The codes are recorded for statistical, policy or reimbursement reasons but also for epidemiological and clinical research and medical practice assessment [1]. For these purposes relevant electronic patient data should be available and reliable [4-8]. However, incorrect and incomplete diagnostic discharge data are still a clearly recognized problem [9].

With respect to medical practice assessment, also the level of detail of the recorded diagnoses has an impact on the usability of these data. For example, the type and duration of antimicrobial therapy in case of bacterial meningitis depend on the type of bacterial pathogen [10]. Also prognosis in terms of mortality or neurological and hearing sequelae and length of stay differ per bacterial pathogen [11]. For medical practice assessment, process measures as choice of therapy and outcome measures as complications, mortality and length of stay are important indicators [12]. Diagnostic coding should therefore take into account details necessary for the proper application of these indicators. This is especially true for the coding of the principal diagnosis of an admission.

The number and kind of recorded secondary diagnoses per admission are also important for medical practice assessment. A distinction should be made between comorbidities and complications [13]. Comorbidities may explain additional therapy or may influence outcomes and should therefore be taken into account when interpreting indicators for quality of care [14]. Complications may form outcome indicators themselves [15].

An international consortium addressed new and ongoing challenges associated with using administrative data in health services research [1]. They identified ‘interventional studies to enhance coding quality’ as one of the potential areas of research in this field.

In this paper we present a long time pre-post study to determine the long term effects of a system intervention concerning the level of detail of principal diagnoses and the number and kind of secondary diagnoses.

Problem case

Because they are responsible for quality of care, pediatricians at the Academic Medical Centre (AMC), The Netherlands, wanted to assess their medical care. Part of the necessary information could be determined from the discharge registry. However, the pediatricians doubted whether the diagnostic discharge data adequately reflected the patients' conditions. This doubt was initially based on their experience with the registration process and was later empirically confirmed in a study where the pediatricians defined performance indicators. We used these indicators to evaluate the availability and usability of the required electronic patient data [(12)].

The coding procedure was form-based. After discharge of a patient, a junior physician had to complete a paper discharge form with information in free text about, among others, reason for admission, principal diagnosis, secondary diagnoses and complications. This diagnostic information was encoded and recorded in the hospital information system (HIS) by a medical record coder. For encoding the AMC Diagnosis Codes (ADC), a local extension of the ICD-9-CM, was used. In the ADC, six digits are available for a diagnostic code; depending on the length of the original ICD-9-CM code, one, two, or three digits are available for local extensions.

However, the discharge form was often rudimentarily completed by the pediatricians due to low priority. Only in a few cases the medical record coder had the disposal of the discharge letter to verify, and if necessary revise and complete, the diagnostic information on the discharge form. At that time the discharge letter was not always written, was often written long after discharge and was not available in the HIS. Further, since the medical coder does not know the patient, he or she may be unaware of additional diagnostic facts that might have influenced the selection of appropriate codes. The pediatricians also felt that the ADC lacked clinical detail for their domain. They thought that diagnoses should be coded at the level of detail needed in daily practice to form the framework within which the care provided can be assessed in a valid way.

Intervention

To improve data quality, the pediatricians decided that diagnoses should be coded by themselves and that diagnosis registration should be integrated in their daily

medical practice. They also wanted to extend the ADC with more clinical details. Measures taken were [16]:

A. Completing the discharge form with free text was replaced by completing a section on the discharge letter with predefined text and corresponding codes. A discharge letter always has to be sent to the general practitioner who plays a pivotal role in the Dutch health care system. Therefore, the discharge letter contains –next to the above mentioned section- a detailed and structured overview of the admission;

B. Instead of the medical record coder, pediatricians themselves encode the diagnoses. Junior pediatricians encode under the supervision of senior pediatricians;

C. In order to support the pediatricians, a portable booklet was developed with a list of alphabetically ordered pediatric diseases with their codes. The list was developed with the help of the pediatricians. A set of 1,700 ICD-9-CM codes used at the pediatric department of another university hospital from 1978 to 1992 was taken as starting point. Based on their knowledge of the characteristics of their own patient population, pediatricians created a list consisting of diseases selected from this set and added other diseases that were relevant to them. The added diseases could be existing diseases from the ADC, or new, more specific ones. The specific codes in the list that were not present in the ADC yet, were added to the ADC. The booklet therefore represented a subset of the ADC. It contained 1560 diseases, including 468 new, more specific diseases;

D. The medical record coder was decentralized to the Pediatric Department to bring in coding expertise by supporting, evaluating, correcting and completing the pediatricians' coding and giving feedback;

E. The new procedure was introduced with full commitment of the head of the department during two staff meetings. In addition, written instructions about how to complete the diagnosis heading were incorporated in the diagnosis booklet. These instructions consisted of definitions of the several types of diagnoses and were in line with ICD-9-CM coding rules. The instructions stressed the importance of coding as detailed as possible and the use of the booklet as the preference list for coding.

After implementation, most of the times a junior physician dictates a discharge letter and selects appropriate diseases with codes from the booklet. If the junior physician cannot find an appropriate disease in the booklet, he or she enters the diagnosis in free text. Next, a secretary types the letter with diseases and codes

(when present) in the intended section. Subsequently, a senior pediatrician checks the letter, including the diagnosis section. Then, the medical record coder checks the correctness, completeness and level of detail of the selected diagnoses. If applicable, she encodes the free text descriptions; first she consults the booklet, but if it does not contain an appropriate code, she will use the ADC. If letter and section have been agreed upon, the countersigned letter is sent to the general practitioner. The electronic version is placed on the hospital information system (HIS). The codes are entered in the HIS by the medical record coder.

It is estimated that the new procedure on average required two minutes extra time for the pediatrician per discharge. Due to the shift in work the medical record coder will not need extra time.

We have investigated the effect of this new procedure on data quality in terms of completeness and correctness of the diagnostic codes stored in the HIS [17]. We did a medical chart review in the year before and after the implementation. Completeness was measured as the proportion of all relevant diagnoses in the medical chart that was actually coded. Correctness was measured as the proportion of coded diagnoses that was documented in the medical chart. For the principal diagnoses, completeness of the old and new registration was 0.62 and 0.67 respectively and correctness at the 3-(and 6-)digit level of ICD-9-CM was 0.68 (0.65) and 0.72 (0.60) respectively. However, the results of the new discharge letter-linked registration were not statistically different from the old form-based registration, but in line with internationally reported data quality [18, 19]. In the old situation the correctness of the 3-digit level coding was only 0.03 larger than that of the 6-digit level. In the new situation this difference was 0.12. This larger difference was due to the fact that in the new situation many more specific codes were used. Since the correctness was determined using the blinded re-abstraction of the medical charts by a colleague senior pediatrician as the gold standard, this determination of correctness resembles the determination of the inter-rater reliability. Since the correctness decreased at the 6-digit level we assume that also the inter-rater reliability at the 6-digit level was somewhat lower than at the 3-digit level.

The completeness of secondary diagnoses was in the year after implementation 0.45 and statistically not better than the year before. It may be that the intervention only leads to an increase in completeness after a longer period of time. Therefore we were interested in the long term effects.

Objectives

We tested the hypotheses that, due to the new procedure:

1. The specific diagnoses will substitute their 'parents', and thus more generic, diagnoses, since:
    - the extension of the classification with more specific codes was the express wish of pediatricians;
    - specific codes were listed in the booklet which became the preference list for encoding;
2. The number of recorded diagnoses will increase, since:
    - more diagnostic information is available because of the pediatricians' contribution and the better availability of the discharge letter for verification and correction by the medical record coder;
    - completeness was rather low;
3. The effect will be permanent, since:
    - the registration is integrated with the communication and care process;
    - a correcting and feedback function of the medical record coder is part of the procedure, so a continuous learning process will develop;
    - the diagnostic data will be used for regular assessment of pediatric care; therefore high quality data is a prerequisite.

5.2 METHODS

5.2.1 Setting, Study Design and Materials

The study was performed at the pediatric department of the Academic Medical Center (AMC) in Amsterdam. This department has 155 beds. The AMC is a tertiary teaching and university hospital.

A time series study was performed covering twelve consecutive years. In the first four years (1991-1994), the common form-based encoding by the medical record coder was in use and in the last eight years (1995-2002), the discharge letter-linked encoding by pediatricians. During these years there were no legislation changes in

The Netherlands that could bias the data, e.g. increased coding due to monetary incentives. After 2002, the announced and meanwhile introduced prospective payment system with its own registration procedures led to uncertainty and a temporarily lower priority for registration of discharge data. Only in recent years discharge data are becoming more important again.

We limited the study to admissions for which the participating subspecialties were solely responsible. Participating subspecialties were: general pediatrics, pediatric cardiology, endocrinology, gastroenterology, hematology, metabolism, nephrology, oncology and pulmonology. Pediatric oncology was excluded since this subspecialty wrote a discharge letter only after an episode of care and not after each admission. Moreover, in 1995 pediatric oncology was faced with two other measures that influenced their diagnosis registration: 1) in case of chemotherapy a combination of neoplasm and procedure code had to be used, instead of code V58.1 alone, and 2) besides the neoplasm code also a morphology code had to be added.

We further limited the study mainly to long stay admissions. A long stay admission is an admission that normally needs at least one overnight stay. A daycare admission is an admission for an intervention that needs nursing care for some hours, but no overnight stay. Long stay and daycare admission belong both to inpatient care. For daycare admissions the form-based registration remained in use. However, since the pediatric booklet was also used for daycare admissions, these admissions were used to evaluate the effect on the recording of principal diagnoses of using the booklet alone. Long stay and daycare both require one and only one principal diagnosis and are thus comparable in this sense. The number of secondary diagnoses recorded for long stay patients differs substantially from the number of secondary diagnoses for daycare patients; secondary diagnoses are hardly recorded for daycare and we therefore did not determine the effect of the booklet in this case.

In order to determine whether the baseline characteristics of the admissions were comparable over the years, for each long stay admission we collected information about the length of stay, responsible subspecialty and the ICD-9-CM chapter corresponding to the recorded principal diagnosis.

We retrieved the retrospective data from the hospital information system.

5.2.2 Main outcome measures

To test the hypotheses we analyzed the data as follows:

For hypothesis 1, we limited our analysis to principal diagnoses. Since the number of principal diagnoses per admission is always exactly one, it is possible to determine the degree of' replacement' of generic codes by specific codes. We computed for each year the "proportion of specific codes" as:

{N recorded specific codes$_p$} /

{(N recorded specific codes$_p$) + (N recorded generic codes$_p$ that have specific codes)}

where *N* stands for "number of" and *codes$_p$* for "codes of principal diagnoses". "Specific codes" are defined here as "locally extended codes". "Generic codes" are defined as "ICD-9-CM codes". In the determination of this proportion, generic codes for which no specific codes exist were excluded.

For hypothesis 2, we limited our analysis to secondary diagnoses, since the number of secondary diagnoses per admission can vary (from 0 to 20). We computed for each year the "mean number of codes per admission" as:

{N recorded codes$_s$} / {N admissions}

where *N* stands for "number of" and *codes$_s$* for "codes of secondary diagnoses".

For hypothesis 3, we compared the above defined proportions and means of subsequent years after introduction.

5.2.3 Data analysis

We computed the mean length of stay per quarter and did a time series analysis with seasonal decomposition and subsequently a linear regression analysis. For detecting differences in proportions of specific principal diagnoses in subsequent years, we used the Pearson Chi-Square test and subsequently the Marascuilo procedure to compare the proportions pairwise with an overall level of significance of 0.05. For detecting differences in the mean number of secondary diagnoses per admission in subsequent years we used the One-Way ANOVA test with Tukey's honestly significant difference test as post hoc test for pairwise multiple comparisons with an overall level of significance of 0.05. We used SPSS® 14.0 to analyze the data. The Marascuilo procedure was calculated by hand.

We additionally investigated whether the subspecialties tend to encode principal diagnoses that fall under their domain in a more specific way than diagnoses that do not. In Table 1, the coding scheme for this analysis is shown. Further we investigated whether serious secondary diagnoses with a high impact on outcome of care –predictive comorbidities- were more often coded after the introduction of the new procedure. In total 27 diagnosis groups were identified as highly predictive of death within 1 year of hospital discharge for children between 1 and 14 years of age [20]. For these diagnosis groups we determined the mean number of secondary diagnostic codes -not qualified as complications- per year. We subdivided both secondary diagnoses and predictive comorbidities into subspecialty and non-subspecialty diagnoses according to the coding scheme in Table 1.

Table 1: Coding scheme used to determine the subspecialty related diagnoses.

Subspecialty	**Subspecialty related diagnoses originates from ICD-9-CM Chapter***
pediatric cardiology	7: Diseases of Circulatory System; 14: Congenital Anomalies
pediatric endocrinology	3: Endocrine, Nutritional, and Metabolic Diseases and Immunity Disorders
pediatric gastroenterology	9: Diseases of Digestive System
pediatric hematology	4: Diseases of Blood and Blood Forming Organs
general pediatrics	1: Infectious and Parasitic Diseases; 8: Diseases of Respiratory System; 16: Signs, Symptoms and Ill-Defined Conditions; 17: Injury and Poisoning
pediatric metabolism	3: Endocrine, Nutritional, and Metabolic Diseases and Immunity Disorders
pediatric nephrology	10: Diseases of Genitourinary System
pediatric pulmonology	8: Diseases of Respiratory System

* Diagnoses in other ICD-9-CM Chapters are considered to be not-subspecialty related

5.3 RESULTS

Table 2 shows baseline characteristics of the admissions. The number of admissions was of the same magnitude over the years. Analysis of the mean length of stay showed a stable seasonal influence over the years, and a decreasing trend ($p < 0.01$). Analysis of the distribution of responsible subspecialties and chapters of ICD-9-CM where the principal diagnoses originated from showed a notable decrease in the proportion of admissions falling under the responsibility of general pediatrics and also a decrease in the proportion of admissions concerning a respiratory disease.

Table 2: Baseline characteristics of the long stay admissions during the study years.

Year	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
Number of Admissions	1250	1288	1217	1243	1328	1387	1431	1451	1369	1355	1321	1202
Mean LOS* (Std. Dev)	7.79 (11.97	(11.11)	7.29 (11.48)	6.70 (11.46)	6.62 (10.38)	6.99 (13.28)	6.61 (11.01)	6.11 (8.67)	6.27 (10.83)	6.10 (10.74)	5.80 (9.25)	6.00 (10.20)
Prop Adm. under Gen Ped's**	0.72	0.71	0.69	0.69	0.57	0.58	0.57	0.59	0.54	0.54	0.51	0.44
Prop Adm. with Resp Princ Dx***	0.20	0.22	0.24	0.20	0.19	0.21	0.17	0.18	0.17	0.17	0.15	0.14

* LOS : Length of stay in day's

** Proportion admissions under responsibility of General Pediatrics

*** Proportion admissions with principal diagnoses part of ICD-9-CM Chapter 8 "Diseases of Respiratory System"

Figure 1 shows the use of specific codes for principal diagnoses. In case of long stay, immediately after the introduction of the new policy, half of the diagnoses for which both generic and specific codes existed, was coded specific (proportion 0.486). Compared to 1994 (prop. 0.030) there was a significant increase in all later years, but also a significant decrease in 1997 (prop. 0.380) compared to 1996 (prop. 0.477) (overall $p < .05$). This decrease is largely caused by a substantial decrease of specific coding of non-subspecialty diagnoses. For daycare, a moderate increase is seen. Compared to 1994 (prop. 0.004) there was a significant increase in the proportion of specific codes in all later years, and also a significant increase in 1999 (prop. 0.154) compared to 1998 (prop. 0.073) (overall $p < .05$).

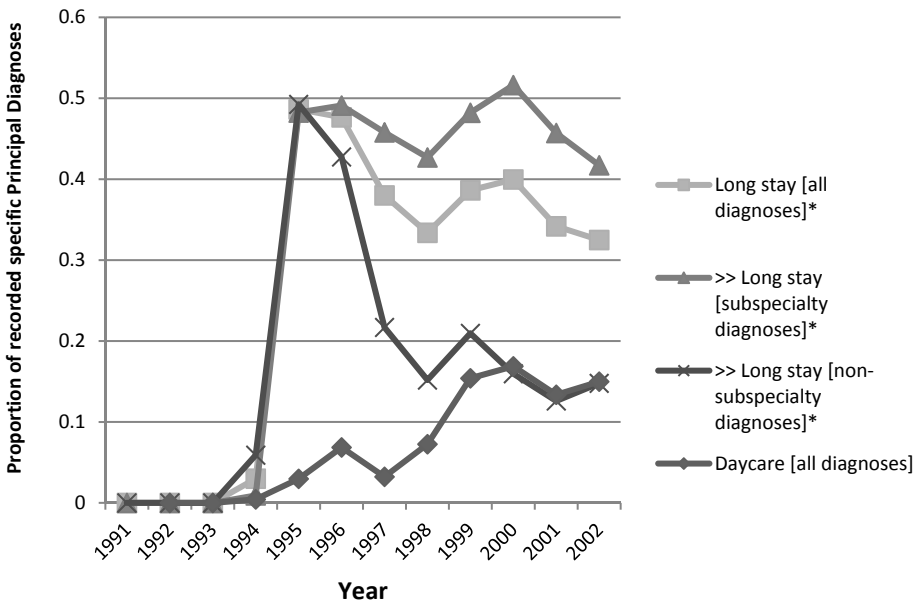

Figure 1: Proportion of recorded specific principal diagnoses (as defined in the section on "main outcome measures") for long stay and daycare. * Long stay diagnoses are also subdivided into subspecialty and non-subspecialty diagnoses according to schema in Table 1. * The proportion of specific principal diagnoses in the recorded long stay diagnoses is the weighted average of the proportions for subspecialty and non-subspecialty long stay diagnoses.

Figure 2 shows that generic codes having specific codes in the pediatric booklet were less used after implementation of the new procedure, and that the use of specific codes listed in the booklet increased. Also longer existing specific codes (in the ADC) not listed in the booklet were used after the implementation, although they were not used before introduction of the booklet. Figure 3 shows that generic subspecialty codes decreased and specific subspecialty diagnoses increased after the intervention, which was much less the case for generic and specific non-subspecialty diagnoses after the intervention.

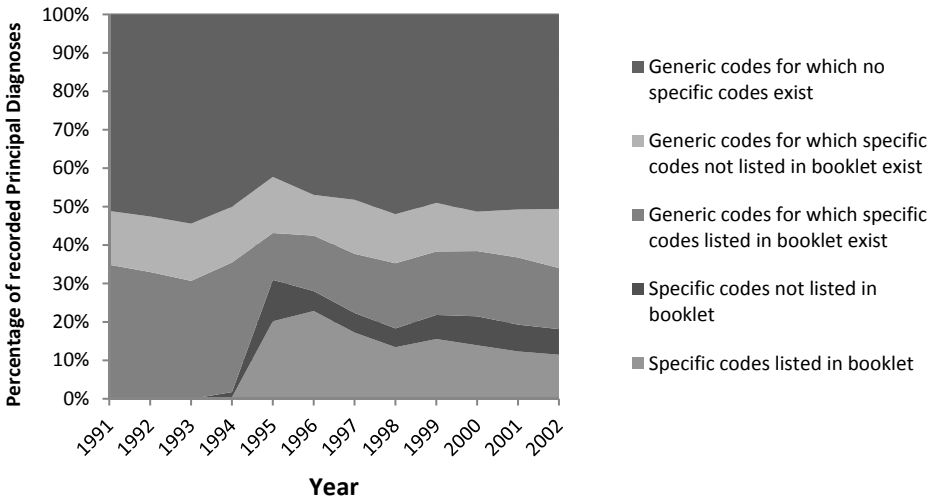

Figure 2: Percentage of recorded specific and generic principal diagnoses per year for long stay, subdivided into specific codes listed or not listed in the booklet and generic codes with or without specific codes.

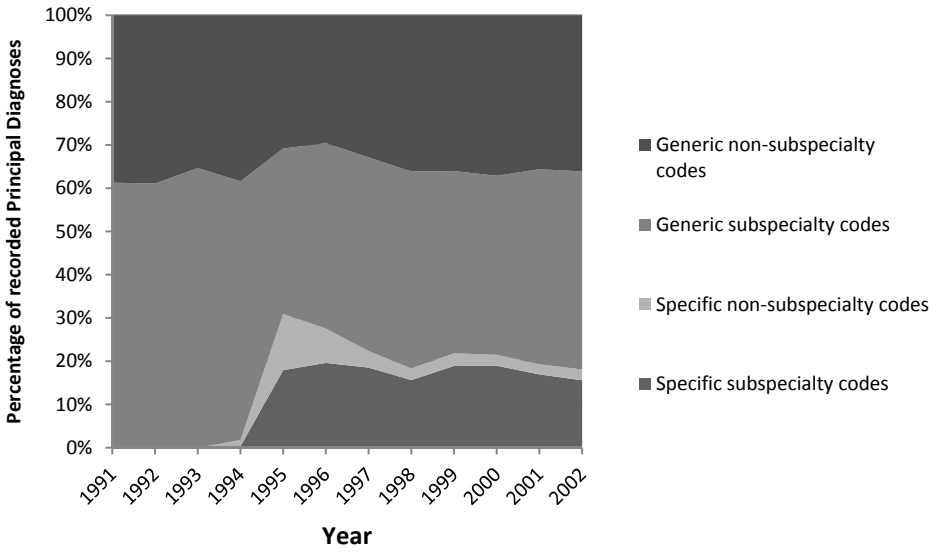

Figure 3: Percentage of recorded principal diagnoses per year for long stay, subdivided into generic and specific subspecialty codes, and generic and specific non-subspecialty codes.

Tukey's analysis of homogeneous subsets shows that the mean number of secondary diagnoses in the years 1996 till 2001 (>0.96) was significantly higher than the mean number of secondary diagnoses during the period 1991 till 1995 and 2002 (about 0.7), and the peak year 1997 (1.41) differed significantly from all other years (overall $p < .05$). The increase and decrease of the mean number of recorded secondary diagnoses is largely caused by the increase and decrease of non-subspecialty diagnoses and to a lesser extent by the increase and decrease of subspecialty diagnoses (Figure 4).

We observed that the mean number of secondary diagnoses almost doubled. Analysis of the diagnoses within the same discharge summaries shows that this increase in the number of secondary diagnoses concerned new information and was not due to redundancies, or the use of subclasses or siblings. The situation was not substantially different from the situation before the introduction of the new registration. We saw an increase in the use of codes from Chapter 16: "Signs, Symptoms and Ill-defined Conditions". Sometimes it was a serious condition belonging to a predictive diagnostic group (see below). Other times it was a sign, symptom or condition belonging to a diagnosis that was also coded.

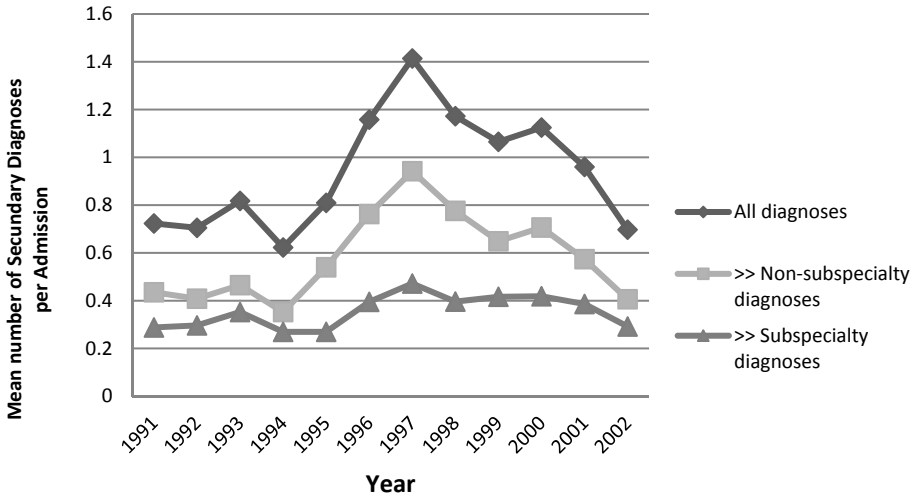

Figure 4: Mean number of recorded secondary diagnoses per admission per year for long stay, also subdivided into subspecialty and non-subspecialty codes.

The mean number of recorded predictive comorbidities [20] increased even more than the mean number of all secondary diagnoses, but also finally ended at the level before implementation (Figure 5). These comorbidities were severe diseases or conditions, such as brain cancer, shock, heart failure and pneumonitis. The increase was not caused by "double" coding of predictive comorbidities within individual discharge summaries but by an increase of the percentage of admissions with a coded predictive comorbidity: from about 10% before 1995 to about 20% in 1997/1998. The increase and decrease of the mean number of predictive comorbidities for the non-subspecialty diagnoses was somewhat more pronounced than the increase and decrease of the mean number of predictive comorbidities for the subspecialty diagnoses.

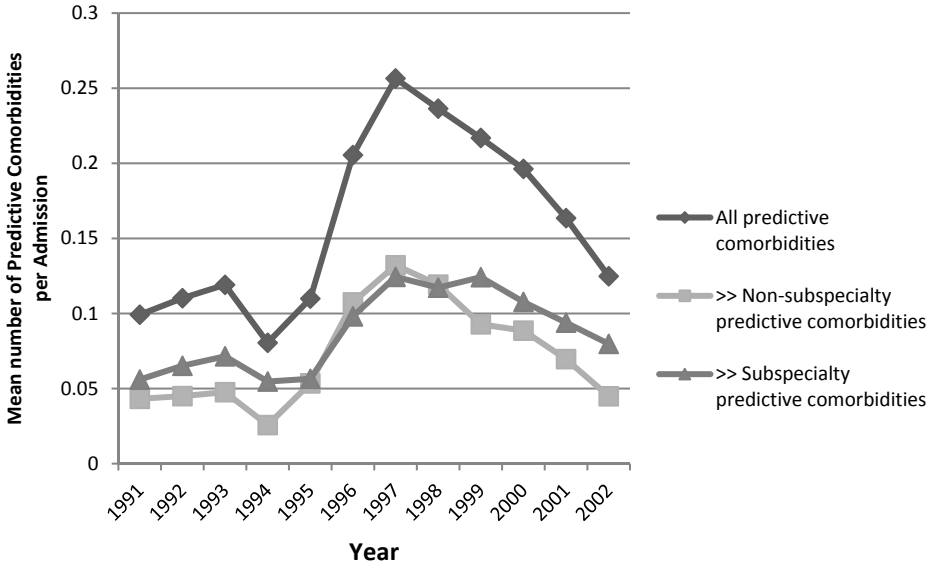

Figure 5: Mean number of predictive comorbidities per admission per year for long stay, subdivided into subspecialty and non-subspecialty codes.

5.4 DISCUSSION

We evaluated the effect and durability of a new diagnostic discharge data registration with increased contributions of pediatricians, on the detail and number of recorded discharge diagnoses. Diagnostic encoding was integrated with discharge letter writing, supported by a pediatric specific booklet of ICD-9-CM-

based diseases. This booklet contained more specific diseases based on the express wish of pediatricians.

Immediately after introduction of the new policy, half of the principal diagnoses for which both generic and specific codes existed, was coded specific. In later years this ratio remained stable at 0.35. The mean number of secondary diagnoses also increased substantially but more gradually, and returned after several years to the level before introduction. This was largely caused by the increase and decrease of the recorded non-subspecialty codes. Also an increase and decrease in a small but important subset of secondary diagnoses (the predictive comorbidities) was observed. The increase in number of secondary diagnoses really provided new information and was not due to the entry of redundant codes.

Methodological issues

We did not carry out an experiment, but evaluated retrospectively routine practice. This means that we had no control over factors other than the new procedure that could influence the diagnostic registration. Some influences could be ruled out by excluding possibly affected cases, like admissions under the responsibility of pediatric oncologists for which simultaneously additional registration rules were introduced. There were also factors that could not be ruled out. In 1992, day care was introduced and in 1995 outpatient treatments, affecting all participating subspecialties. Though difficult to quantify, this could have increased the mean complexity of the remaining long stay admissions with a concomitant increase in the mean number of secondary diagnoses. However, in 1992 there is no increase in mean number of secondary diagnoses, and its decline in 2001 and 2002 does not fit with the still increasing number of day care admissions. Further, the decreased proportion over the years of admissions under the responsibility of general pediatrics and admissions with a principal diagnosis belonging to ICD-9-CM Chapter 8 “Diseases of Respiratory System”, could have had an impact on the level of detail and number of recorded secondary diagnoses (see Table 2). However, the pattern and magnitude over the years of both level of detail and number of recorded diagnoses per admission for these two subsets of admissions were similar to that of all admissions. Further, the decreasing length of stay fits in the general tendency of more efficient care and earlier discharge, case mix being equal.

The present study is quite unique in the sense that we did a time series study over twelve consecutive years with an intervention after four years. So we were able to

study the durability of the effects on the long term. Further, because of the local extensions we were able to analyze the effect of the intervention on the level of detail of the recorded diagnoses. Moreover, by coupling ICD-9-CM chapters with pediatric subspecialties we were able to study the relation between subspecialty related diagnoses and both the detail and number of codes.

Separate effect of discharge letter and booklet

It appears difficult to separate the effect of discharge letter-linked registration and booklet on the recording of pediatric diseases. For daycare admissions, where the discharge letter-linked registration is not in use, the use of specific codes gradually increases although it never reached the level of long stay. This gradual increase may be the result of contamination. The pediatricians complete discharge forms with specific diagnosis descriptions they remember from the discharge letter-linked procedure and the medical record coder selects specific codes she knows from the booklet. It seems that the discharge letter-linked registration catalyzes the selection of specific codes. It forces the physicians to encode, and thus to use the booklet. Moreover, the medical record coder is better informed through the availability of the detailed, well structured discharge letter with registration section. Because of this, even if the physician cannot find a relevant code, she will be able to add a code. This could also explain the use of specific codes not listed in the booklet, although they already existed but were not used before.

The introduction of the booklet has an effect. Additional analyses, not presented here, showed a clear shift in the recording in favor of codes listed in the booklet, even if we correct for the new specific codes that did not exist before introduction of the booklet.

Explanation

Diagnoses that fall under a pediatrician's subspecialty are most of the times the result of the diagnostic process instigated by the responsible pediatrician, so of their own diagnoses they know the details. Pediatricians wanted to code diagnoses on the level of detail needed for their work in daily practice. Each subspecialty could define new specific codes for their own field to be listed in the pediatric booklet. Thus, the new specific codes fill in a need, an important socio-technical factor for success [(21)]. The discharge letter-linked registration and booklet facilitate specific code finding and documentation. In the booklet the specific codes and their

generic code are clustered. The encoding of exactly one principal diagnosis is obligatory. Although a generic code may require less thinking than differentiating between various more specific codes, once the pediatricians are familiar with the codes, it takes hardly extra time to encode a specific code instead of a generic code. As the number of relevant specific codes for their own subspecialty is limited, pediatricians know them. Furthermore, the knowledge of the medical record coder about the availability of specific codes is lasting, important in cases where she has to complement the encoding of the physician. Principal diagnoses most of the time fall under the subspecialty of the responsible pediatricians. Their motivation and the above mentioned conditions lead to a stable and durable percentage of specific principal diagnoses.

Secondary diagnoses most of the time do not fall under one's own subspecialty. Registration of these diagnoses is less important for the responsible pediatrician, especially if diagnoses are not used for quality of care purposes. Besides, the number of codes outside their subspecialty is immense. Furthermore, it costs more effort to encode additional diagnoses and documenting secondary diagnoses is optional. Therefore, it took a year of motivation and feedback before the number of recorded secondary diagnoses increased. The gradual increase indicates a learning curve. In situations where pediatricians are not used to encode secondary diagnoses carefully, or where the attention for the registration policy diminishes, especially the number of secondary diagnoses outside the pediatrician's subspecialty will seriously be affected.

In later years attention, training and reinforcement diminished due to the fact that:
- the pediatrician who did the pioneering work left the organization so that the clinical leadership in this respect disappeared;
- junior physicians and staff members were replaced by new ones that were not given solid instruction;
- feedback on physicians' encoding was lacking: it became too laborious and time-consuming, and led to unwanted delay in sending discharge letters;
- the data were not used for medical practice assessment, as originally intended.

We reported earlier about the accuracy with which the pediatricians filled in the diagnostic section of the discharge letter [16]. Two years after introduction, most of the time pediatricians reported the diagnostic information in free text on the diagnostic section, but often failed to encode this information. After the millennium, according to the medical record coder, even the textual information

about secondary diagnoses was lacking increasingly in the intended section and only available in the text of the letter. Lack of positive incentives and the limited time available for encoding each case, prevented her to correct this fully.

Additional remarks

Not every extra diagnosis probably has an added informational value. However, in consideration of the earlier observed incompleteness of secondary diagnoses in the year before and after implementation, we believe that the extra secondary diagnoses overall have increased the value of the registry for medical practice assessment. However, more coding could also have led to proportionally more false positives and therefore could have influenced correctness negatively.

Pediatricians hardly qualified secondary diagnoses as complications, a well-known phenomenon [15]. This could mean that some predictive comorbidities were in fact complications. There are algorithms to automatically qualify a secondary diagnosis as a complication, based on the probability of the relationship with the principal diagnosis [22]. Should such algorithms be applied to our data the mean number of complications would probably be higher and possibly fluctuate with the total number of recorded secondary diagnoses.

In the literature the role of physicians and medical record coders in diagnostic coding has been discussed. Surján [23] has constructed a framework of the coding process. The first steps in the framework require knowledge about medicine and the patient, ask for careful patient documentation, and is therefore typically the domain of the physician [24]. The last steps require knowledge about the structure and rules of the ICD and is therefore the domain of the medical record coder. According to Surján, physicians mainly work with diagnoses and not with diseases [23]. Whereas a diagnosis is a specific medical problem of a patient, a disease is "a name of an abstract entity, based on some common features observed in different patients grouped together for some reason". The induction from diagnoses to disease coding requires medical knowledge and knowledge on categorical structures but is often neglected. Ideally physician and medical record coder should work together here. Santos et al [25] reported that coding quality could be improved by coders engaging in a variety of role behaviors and increased coder interactions with medical staff. Other means to improve the coding process are education [26] and feedback [27].

5.5 CONCLUSIONS

There is a need for detailed encoding of diagnoses. Extension of the classification of diseases according to the wishes of pediatricians and easy access to this classification truly lead to a more detailed encoding of the principal diagnosis, especially of diagnoses that belong to the pediatricians' own subspecialties. In combination with the discharge letter-linked coding procedure, this leads to a long-lasting use of detailed codes. A substantial effect on the number of secondary diagnoses per admission could only be shown after a longer period of feedback and attention, but this effect gradually diminishes on the long term when feedback and attention are lacking. The increase of secondary diagnoses did lead to added informational value.

Especially the number of secondary diagnoses that do not fall under one's own subspecialty will fall back in the long term. Availability of a detailed and well-structured discharge letter leads to a better informed medical record coder. However, in the long term the medical record coder is not able to compensate for the diminishing contribution of physicians to the encoding process. The number of recorded secondary diagnoses falls back again to the initial level, and with that the added value for medical practice assessment disappears. The level of detail of principal diagnoses remains stable because of the advantage for pediatricians of having specific diagnostic codes falling under their own subspecialty.

A combined effort of pediatricians and medical record coder leads to more coded diagnostic information than the effort of a medical record coder alone.

REFERENCES

1. De Coster C, Quan H, Finlayson A, et al. Identifying priorities in methodological research using ICD-9-CM and ICD-10 administrative data: report from an international consortium. BMC Health Serv Res 2006;6:77.
2. National Center for Health Statistics. International Classification of Diseases, 9th Revision, Clinical Modification, 6th edition. Hayattsville, Maryland: NCHS; 2003.
3. World Health Organization. International Statistical Classification of Diseases and Related Health Problems; Tenth Revision. Geneva: World Health Organization; 1992.
4. McKee M. Routine data: a resource for clinical audit? Quality in Health Care 1993:104-11.
5. Safran C, Chute CG. Exploration and exploitation of clinical databases. International Journal of Bio-Medical Computing 1995;39:151-6.

6. Wyatt J. Acquisition and use of clinical data for audit and research. Journal of Evaluation in Clinical Practice 1995;1:15-27.

7. Iezzoni LI. Assessing quality using administrative data. Annals of Internal Medicine 1997;127:666-74.

8. Coory M, Thompson B, Baade P, Fritschi L. Utility of routine data sources for feedback on the quality of cancer care: an assessment based on clinical practice guidelines. BMC Health Serv Res 2009;9:84.

9. Rosamond WD, Chambless LE, Sorlie PD, et al. Trends in the sensitivity, positive predictive value, false-positive rate, and comparability ratio of hospital discharge diagnosis codes for acute myocardial infarction in four US communities, 1987-2000. Am J Epidemiol 2004;160:1137-46.

10. Tunkel AR, Hartman BJ, Kaplan SL, et al. Practice guidelines for the management of bacterial meningitis. Clin Infect Dis 2004;39:1267-84.

11. Lazzarini L, Toti M, Fabris P, et al. Clinical features of bacterial meningitis in Italy: a multicenter prospective observational study. J Chemother 2008;20:478-87.

12. Prins H, Kruisinga FH, Buller HA, Zwetsloot-Schonk JH. Availability and usability of data for medical practice assessment. Int J Qual Health Care 2002;14:127-37.

13. Naessens JM, Huschka TR. Distinguishing hospital complications of care from pre-existing conditions. International Journal for Quality in Health Care 2004;16 Suppl 1:i27-35.

14. Pine M, Jordan HS, Elixhauser A, et al. Modifying ICD-9-CM coding of secondary diagnoses to improve risk-adjustment of inpatient mortality rates. Medical Decision Making 2009;29:69-81.

15. Quan H, Parsons GA, Ghali WA. Assessing accuracy of diagnosis-type indicators for flagging complications in administrative data. Journal of Clinical Epidemiology 2004;57:366-72.

16. Prins H, Buller HA, Zwetsloot-Schonk JH. Redesign of Diagnostic Coding in Pediatrics: From Form-based to Dicharge Letter Linked. Perspectives in Health Information Management 2004;1:1-10.

17. Prins H, Buller HA, Zwetsloot-Schonk JH. Effect of discharge letter-linked diagnosis registration on data quality. Int J Qual Health Care 2000;12:47-57.

18. Cleary R, Beard R, Coles J, et al. Comparative hospital databases: value for management and quality. Qual Health Care 1994;3:3-10.

19. Pietila K, Tenkanen L, Manttari M, Manninen V. How to define coronary heart disease in register-based follow-up studies: experience from the Helsinki Heart Study. Annals of Medicine 1997;29:253-9.

20. Tai D, Dick P, To T, Wright JG. Development of pediatric comorbidity prediction model. Arch Pediatr Adolesc Med 2006;160:293-9.

21. Berg M. Patient care information systems and health care work: a sociotechnical approach. Int J Med Inf 1999;55:87-101.

22. Kronman MP, Hall M, Slonim AD, Shah SS. Charges and lengths of stay attributable to adverse patient-care events using pediatric-specific quality indicators: a multicenter study of freestanding children's hospitals. Pediatrics 2008;121:e1653-9.

23. Surjan G. Questions on validity of International Classification of Diseases-coded diagnoses. Int J Med Inform 1999;54:77-95.

24. O'Malley KJ, Cook KF, Price MD, Wildes KR, Hurdle JF, Ashton CM. Measuring diagnoses: ICD code accuracy. Health Services Research 2005;40:1620-39.

25. Santos S, Murphy G, Baxter K, Robinson KM. Organisational factors affecting the quality of hospital clinical coding. HIM J 2008;37:25-37.

26. de Lusignan S, Hague N, Brown A, Majeed A. An educational intervention to improve data recording in the management of ischaemic heart disease in primary care. J Public Health (Oxf) 2004;26:34-7.

27. De Lusignan S, Stephens PN, Adal N, Majeed A. Does feedback improve the quality of computerized medical records in primary care? J Am Med Inform Assoc 2002;9:395-401.

CHAPTER 6

APPROPRIATENESS OF ICD-CODED DIAGNOSTIC INPATIENT HOSPITAL DISCHARGE DATA FOR MEDICAL PRACTICE ASSESSMENT: A SYSTEMATIC REVIEW

Hilco Prins, Arie Hasman

Accepted for publication in revised form in Methods of Information in Medicine, Sept 2012

ABSTRACT

Objectives: We performed a systematic review to investigate the quality of diagnostic hospital discharge data (DHDD) in order to gain insight in the usefulness of these data for medical practice assessment. We investigated the methods used to evaluate data quality, factors that determine data quality and its consequences for medical practice assessment.

Methods: We performed a sensitive search and selected studies in which both completeness (or sensitivity: SENS) and correctness (or positive predictive value: PPV) were measured. We used the random-effects model to calculate the pooled SENS and PPV and to explore the effect of a number of covariates.

Results: The 101 included studies were heterogeneous on coding practice, diagnoses and evaluation methods. We distinguished six typical study designs. We found an average SENS of 0.67 (95%CI: 0.62-0.73) and PPV of 0.76 (95%CI: 0.73-0.79); for some single disease studies SENS was significantly higher and for comorbidity and complication studies SENS was significantly lower. PPV was significantly higher for Scandinavian countries than for other countries. Recoding compared to re-abstracting of the medical record as a gold standard gave a significantly lower PPV. Diagnostic data were considered appropriate by the authors of the studies for quality of care purposes when both SENS and PPV were at least 0.85. Only 13% of the studies fulfilled this criterion.

Conclusions: Variability in quality of care between settings can easily be overshadowed by the unknown variability in data quality. However, the use of DHDD by physicians to evaluate their own medical practice may be useful. But only if physicians are willing to critically interpret the meaning of the information for their medical practice assessment.

Keywords: ICD, diagnostic data, discharge data, sensitivity, positive predictive value

6.1 INTRODUCTION

If physicians want to assess retrospectively their medical care for certain patient groups in order to improve their medical practice, information about the diagnoses of the patients is of utmost importance. In most Western countries, diagnoses for each admission in a hospital are coded and electronically recorded after discharge of the patient for statistical, policy or reimbursement reasons [(1)]. Nowadays, this diagnostic information is coded using ICD-9-CM [(2)] or ICD-10 [(3)]. The use of these diagnostic hospital discharge data (DHDD) for medical practice assessment is appealing because of its availability, ease of access, standardization of documentation, complete coverage of all admissions and low additional cost [(4)]. However, high data quality is not self-evident. Since the 1970s, many studies have been conducted to evaluate the quality of these data for various purposes. One of the first major studies was performed in 1974 by the Institute of Medicine [(5)]. This study examined the agreement between the principal diagnoses of the hospital discharge data and the corresponding diagnostic information within the medical records. It was concluded that: "*Diagnosis-specific discrepancies are of sufficient magnitude to preclude use of such* (diagnostic hospital discharge, authors) *data for detailed research and evaluation"*. It also stated that: *"These findings may be particular timely, since increasingly important decisions about the content of medical care and levels of reimbursement may be based on such information*". But is this situation still the case, or did the quality of data improve in later years? The objective of our review is to get insight in the appropriateness of diagnostic hospital discharge data for medical practice assessment.

Performance indicators are a promising tool to help physicians in a hospital assess the quality of their specialistic care for patient groups with certain clinical conditions [(6)]. These performance indicators have to be translated into data retrieval queries to identify in the hospital information system (HIS) the patients with the clinical conditions of interest and to detect within this group of patients events such as comorbidities or complications. Often the clinical conditions of interest are diagnoses, with comorbidities and complications being special forms of diagnoses. In order to get reliable performance indicators, the hospital diagnostic discharge data should be of high quality [(7-11)], especially in terms of completeness and correctness. To qualify the data in terms of completeness and correctness, a gold standard (GS) is needed to compare the data with, resulting in 'criterion validity' [(12)]. Using a GS, a diagnostic code can be true positive (TP), false positive (FP),

false negative (FN) or true negative (TN), and this outcome can typically be placed in a 2x2 table. Completeness is then equivalent to sensitivity (SENS: TP/(TP+FN)) which is here [13]:

> *The proportion of patients' hospitalizations with a certain diagnosis according to the GS for which the corresponding diagnostic code is present in the hospital discharge record.*

Correctness is equivalent to the positive predictive value (PPV: (TP)/(TP+FP)) which is [13]:

> *The proportion of patients' hospitalizations with a certain diagnostic code in the hospital discharge record for which the corresponding diagnosis is present according to the GS.*

Both completeness and correctness are required to describe the quality of clinical data, since there is typically a tradeoff between them [14]. Generous allocation of a diagnostic code can lead to high completeness at the cost of correctness. Careful allocation of a diagnostic code can lead to high correctness at the cost of completeness. The magnitude of these effects depends on the prevalence of the diagnosis. When identifying cases for medical practice assessment, incomplete data leads to undetected cases and incorrect data to unwanted cases. With regards to complications, incomplete and incorrect data lead to underestimation and bias of adverse outcomes respectively [15, 16]. With regards to comorbidities, incomplete and incorrect data lead to underestimation and bias of risks of adverse outcomes respectively. In the literature, no criteria can be found about minimum requirements for completeness or correctness of diagnostic data so that they can be used for medical practice assessment.

This systematic review investigates the quality of diagnostic inpatient hospital discharge data as reported in scientific journals. We selected only studies in which -at least- both completeness and correctness were measured. Our research question was: Are ICD-Coded diagnostic inpatient hospital discharge data appropriate for medical practice assessment? We investigated:

1. the characteristics of the settings where the evaluation took place;
2. the gold standards and designs used to assess completeness and correctness;
3. the values for completeness and correctness and relation with kind of diagnosis, setting, and evaluation method;
4. the determinants of data quality that were reported by the researchers;

5. evidence about the consequences of the data quality for medical practice assessment;
6. the appropriateness of the data for the objectives of the studies, especially for quality of care.

6.2 METHODS

6.2.1 Selection procedure

We selected the papers in three steps. First, we used PubMed to perform a sensitive search. Our query was composed of four sets of terms, combined with "AND": (1) terms about quality, (2) terms about coded diagnostic hospital discharge data, (3) title words about quality or coded diagnostic hospital discharge data and (4) relevant MeSH headings. We used a previously defined set of eligible studies to optimize the query, which was executed on March 21st, 2006 (see Appendix A for the complete query).

Second, a manual selection of studies resulting from step one was carried out by the two researchers by applying in- and exclusion criteria (see below) to the title and abstract in order to obtain a set of possibly eligible studies. Possibly eligible were those studies for which the abstracts made not clear that they could be excluded. A set of 2081 (out of a total of 4040) abstracts was independently evaluated by both researchers. In case of disagreement, both researchers had to come to consensus and the interpretation of the criteria was discussed to increase interrater reliability. The rest of the abstracts was split up; one half was evaluated by one and the other half by the other researcher.

Thirdly, the two researchers independently screened the resulting set of possibly eligible studies based on full-text papers, again using the list of in- and exclusion criteria. In case of disagreement, both researchers had to come to consensus.

6.2.2 In- and exclusion criteria

Included were studies that fulfilled all of the following criteria; the study:

- Evaluated ICD coded diagnostic hospital discharge / administrative / claims data (when DRG's were evaluated, the study could be included if also the quality of the underlying ICD codes was evaluated)
- Reported the evaluation of inpatient hospital discharge codes separately in case also other discharge codes (e.g. outpatient or physician claims data) were evaluated.
- Used a GS as a reference for data quality
- Measured both completeness (SENS) and correctness (PPV)
- Reported absolute numbers for TP, FP and FN or gave sufficient information to compute these numbers*.

_____

* This makes it possible to carry out a meta-analysis

Excluded were studies that fulfilled one or more of the following criteria:

- Reviews;
- Only about, or mingled with out-patient data*;
- Only about, or mingled with procedural data*;
- Only about, or mingled with physician claims data*;
- Only about, or mingled with E-codes (external cause of injury) *, **;
- With an 'experimental' design instead of evaluation of routinely collected data;
- With PPV and SENS not sharing the same set of TP's.

_____

* If a study also includes, and explicitly reports about inpatient diagnostic codes, we did not exclude the study.

** ICD-codes from Chapter 'Injury and Poisoning' (ICD-9-CM) do not belong to this exclusion-criterion.

6.2.3 Data extraction

Both researchers independently extracted the necessary data from the selected full-text papers and the results were discussed in order to come to consensus. We used a data extraction form (see Appendix B). Data were extracted about:

- Setting of study (country, type of hospital, number of hospitals, who coded, etc)
- Routine use of data;
- Objectives of data evaluation;
- Kind of diagnostic data;
- Data evaluation methods;
- Numbers of TP, FP, FN and –if measured- TN
- Data quality in terms of SENS, PPV and –if possible- SPEC;
- Authors' conclusions about determinants of data quality and usability of data.

For each study we extracted or reconstructed one 2x2 table. If the authors themselves reported overall measures of TP, FP, FN (and –if possible- TN) and the corresponding SENS and PPV (and –if possible- SPEC[1]), even when several disease subcategories were distinguished, we extracted the overall numbers. If no overall measures were reported, we pooled TP's, FP's, FN's (and –if possible- TN's) of disease subcategories into one 2x2 table, computed the SENS, PPV (and –if possible- SPEC) and labeled the pooled data with an appropriate disease category. If numbers were given over several periods, we extracted only the numbers of the latest period. In case of intervention studies, we only used the post intervention data.

6.2.4 Data analysis

We did a meta-analysis to estimate the overall SENS and PPV averaged over the studies. Since we found that the studies were very heterogeneous (because of different hospitals and coding settings, populations, diseases and evaluation methods), we used the random-effects model of Dersimonian & Laird [(17)] instead of a fixed-effects model to calculate the pooled SENS and PPV. We used the I^2-statistic to measure the heterogeneity of the studies. The I^2-statistic represents the percentage of variability in SENS and PPV estimates due to heterogeneity rather than chance. The random-effects model explicitly takes between-study variance into account and estimates this variance. To determine the weighted average of SENS (or PPV) the estimate of SENS (or PPV) of each individual study is multiplied by a weight being the inverse of the sum of the variance of SENS (or PPV) of the individual study and the estimate of the between-study variance (Tau-squared). Since this between study variance was also included when calculating the

[1] Specificity: The proportion of patients' hospitalizations without a certain diagnosis according to the GS for which the corresponding diagnostic code is not present in the hospital discharge record.

95% confidence interval (95% CI), the 95% CI of the random model is wider than the 95% CI of a fixed-effects model that does not assume between-study variance.

The random-effects model was also used to determine the effect of a number of covariates on both SENS and PPV (subgroup analysis). We limited our subgroup analysis to predefined covariates related to kind of diagnostic data, period of original coding, setting and methodological issues. We hypothesized that disease category, coding year, country and type of GS would influence the observed data quality. Within the studies that used the medical record as GS we also analyzed the influence of blinding and recoding for GS construction.

We calculated mean SENS and PPV with their 95% CI with the random-effects model using Excel based on Borenstein et al [18]. We compared mean SENS and PPV of the study's subgroups with the two sample t-test.

Many authors of the included papers discuss determinants of diagnostic data quality. Our inventory of determinants was limited to those that appeared to be significant after a statistical analysis like Pearson Chi-square tests or multivariate regression analyses.

6.3 RESULTS

6.3.1 The search

Our PubMed search yielded 4040 results, out of which 283 were labeled, based on title and abstract, as "possibly eligible". Agreement between the two researchers over the first 2081 abstracts was 0.917 and Kappa was 0.651. After a manual review of 276 full-text papers (seven papers could not be obtained), 101 studies were included based on our eligibility criteria. Agreement between the two researchers about the full-text based paper selection was 0.826 and Kappa was 0.645. The reasons for exclusion are given in Table 1.

Figure 1 shows the number of studies included per publication year. The number of evaluation studies about diagnostic data quality increased since the nineties. The increased wish to critically determine the usability of DHDD for secondary purposes coincides with the increased call for transparency of care and accountability (in terms of quality and costs). The studies were published in epidemiological, quality of care or general health care journals in 62 cases, and published in specialty orientated journals in 39 cases.

Table 1: Numbers of inclusions and exclusions with reason.

Source	Title + Abstract	Full-text
Number of studies	4040	283
Reason for exclusion		
- Full-text paper could not be obtained	-	7
- Not about ICD-coded hospital discharge data	2828	9
- ICD codes are used in study, but not evaluated	745	36
- Review study	6	1
- Inpatient hospital diagnostic data not (separately) evaluated, but about (or mingled with) outpatient, physician, procedure or 'cause of injury' data	71	37
- No (correct) criterion validity of routinely collected data	77	30
- TP, FP, FN not (all) available or not possible to calculate	30	62
Number of (possibly) eligible studies	(283)	101

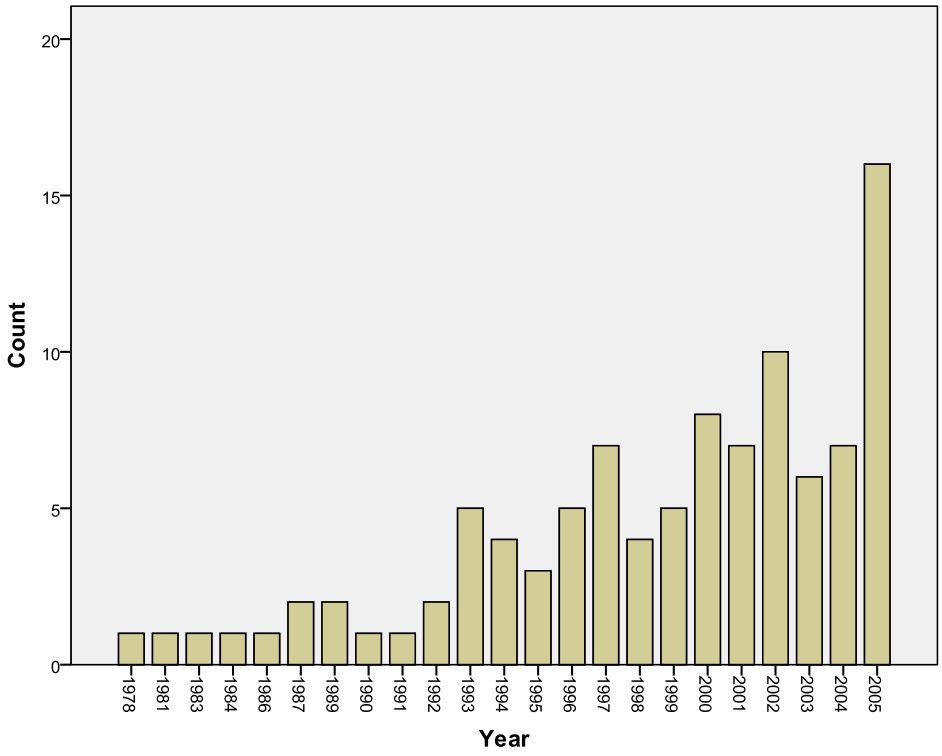

Figure 1: Number of included studies as a function of the year of publication (2006 was excluded since this year was not yet finished when the search was performed).

6.3.2 Characteristics of the (coding) settings where the evaluation took place

Figure 2 shows that almost half of the studies came from the USA; the rest of the studies came from Anglo-Saxon, Scandinavian and some other European countries.

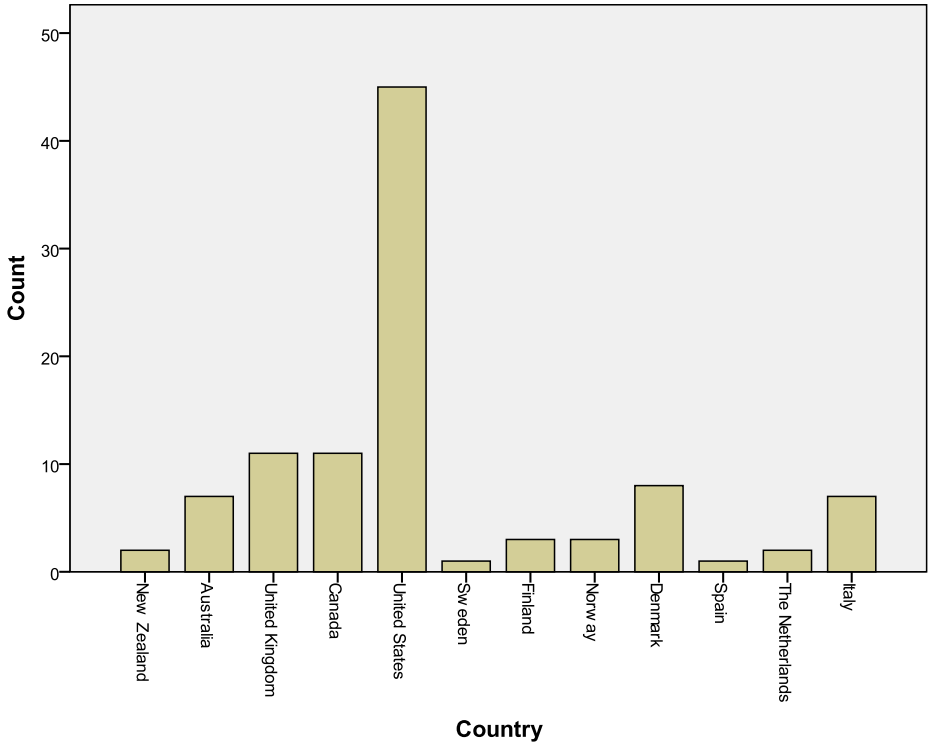

Figure 2: Number of included studies per country. Countries are grouped (Anglo Saxon, Scandinavian, and other European countries) and sorted by count within group.

In 37 studies, the evaluation was performed in a single hospital setting. In 64 studies the evaluation setting consisted of more than one hospital: 18 studies with 2-10 hospitals, 22 studies with 11-100 hospitals, 7 studies with > 100 hospitals and 17 with an unknown number of hospitals. In 39 studies the evaluation took place only in general hospitals, in 16 studies only in university hospitals, in 32 studies in a combination of general and university hospitals and in 32 studies the type of hospital was not specified.

Table 2 shows, by country, the reasons for coding and who is coding. Reasons for coding were not always, or not always completely, reported. Notably, reimbursement as reason for coding is almost limited to studies done in the United States and other Anglo Saxon countries, and the doctor as coder is relatively often seen in Scandinavian studies.

In 8 studies, the use of ICD-8 was evaluated. In respectively 34, 49 and 10 studies ICD-9, ICD-9-CM and ICD-10 were evaluated.

Table 2: Reasons for coding and who is coding by country.

Country	**Reasons for coding***						**Who is coding**		
	Poli-cy	*Manage-ment*	*Reimburse-ment*	*Qua-lity*	*Re-search*	*N.S***	*Doc-tor*	*Traine d coder*	*N.S** *
United States (n=45)	24	5	29	15	7	5	1	32	12
Other Anglo Saxon countries (n=31)	20	2	12	9	7	6	2	21	8
Scandinavian countries (n=15)	12	0	1	4	4	3	6	1	8
Other European countries (n=10)	6	0	3	1	1	1	3	2	5
All countries (n=101)	62	7	45	29	19	15	12	56	33

* Reasons for coding categories not mutually exclusive

** N.S.: Not Specified

6.3.3 Gold standards and designs used to assess completeness and correctness

First we distinguished studies that evaluated diagnostic data quality of 1) all kinds of diseases together, 2) a single disease or 3) all or some diseases in a subgroup of patients who a) had a specific disease or b) underwent a specific procedure. In case of a subgroup of patients having a specific disease the interest is on the quality of the data concerning comorbidities or complications. Diseases in a subgroup of patients who underwent a specific procedure can be comorbidities or complications as well as principal diagnoses.

Secondly we distinguished studies according to the GS that was used: 1) the medical record, 2) a disease specific registry or 3) prospectively collected diagnostic data.

When the medical record was used as GS, the re-abstracted or recoded medical record was compared with the diagnostic hospital discharge codes. Electronically available test results were considered by us to be part of the medical record. Re-abstracting means that relevant information from the medical record was retrieved, but not recoded. Re-abstracting took place blinded or not blinded for the original codes although blinding was not always (clearly) specified. After re-abstracting the re-abstractor decides whether the re-abstracted data confirm the original code. In case the re-abstractor is not blinded for the original code the results of re-abstraction may be biased. In recoding, the medical record is re-abstracted and again coded, and the newly obtained codes are compared with the hospital discharge codes. Recoding is always done blind for the original codes. The medical record was used in 61 studies (60%) as main GS. Re-abstracting took place in 35 of these studies of which 14 blinded. Recoding took place in the other 26 studies; in 12 studies a medical record coder was the re-coder, in 9 studies a doctor (in 5 studies not specified). Agreement between two or more re-abstracters or re-coders was reported in 21 studies. In 23 studies, clinical criteria (instead of, or in addition to ICD criteria) were applied to re-abstract or re-code. Clinical criteria are criteria acknowledged by a professional organization of specialists, to diagnose a disease.

A disease specific registry as GS was used in 31 studies (31%). In order to match the hospital discharge data with the registry data, a patient ID in combination with a time frame was used for record linkage in 20 studies; in 4 studies more or less anonymous patient data without ID with a time frame was used for record linkage, e.g. birth date, sex and Zip code. In 7 studies the matching procedure was not specified. A time frame was used to limit matching to those disease registry events having a date stamp close to the period of hospitalization (e.g. within 28 days before or after the admission date [19]).

Prospectively collected data as GS was used in 11 studies (11%). In these studies already available diagnostic data prospectively collected for clinical research or other purposes during hospitalization of the patients, were ‘gratefully’ used as GS.

There is a relationship between the type of GS and the range of diseases studied, shown in Table 3. In studies where all kinds of diseases were combined, only the medical record was used as GS. Most of the studies evaluating diseases in clinical

subgroups also use the medical record as GS. Studies that evaluated single diseases used the medical record, a disease specific registry or prospectively collected data.

When matching the DHDD with the GS data, studies differ in the number of digits of the ICD codes that have to match exactly for agreement; studies compared the GS data at the 5-, 4-, or 3-digit level of codes, or with a group of codes in respectively 5, 13, 31 and 52 studies.

Table 3: Number of studies as a function of type of GS and range of diseases.

		GS - Main			
		Medical Record	*Disease Specific Registry*	*Prospectively collected data*	**Total**
Range of diseases	**All kinds of diseases**	18	0	0	18
	Single disease	14	25	8	47
	Diseases in clinical subgroup	29	4	3	36
Total		61	29	11	101

Looking at Table 3, we infer the following typical designs to evaluate diagnostic data quality in terms of completeness and correctness:

Design 1-**all kinds of diseases, GS is medical record:** 18 studies [5, 13, 16, 20-34]

A (random/stratified/every n^{th} case/consecutive) sample of hospitalizations is drawn from a routine discharge database. Then the coded diagnoses are compared with those in the corresponding recoded medical record. In case of re-abstraction it is determined whether the information confirms the code. All data are represented by one 2x2 table, resulting in one SENS and PPV per study. Some large studies also report SENS and PPV for a number of single diagnoses with relatively high prevalence.

Design 2a-**single disease, GS is medical record:** 14 studies [35-48]

First a GS is determined/constructed for a specific diagnosis (e.g. all true myocardial infarctions) based on a comprehensive medical record review with abstracting or recoding. Then all cases with the hospital discharge code(s) of interest are compared with the GS.

Design 2b-**single disease, GS is disease specific registry:** 25 studies [19, 49-72]

The GS is based on a disease specific registry. All cases from the hospital discharge database with the relevant code(s) are compared with the GS, sometimes combined with a medical record review for extra or additional (in case of non-matching) verification.

Design 2c-**single disease, GS is prospectively collected diagnostic data:** 8 studies [73-80]

The GS is constructed prospectively and independently of coding practice, e.g. cases included in a clinical study of a specific disease. Then all cases with the relevant code(s) from the hospital discharge database are compared with the GS.

Design 3a- **all or some kinds of diseases in patients having a specific disease:** 12 studies [81-92]

(A sample of) all admissions in which the patient had a specific disease is obtained and for these selected cases a verification takes place of the coded comorbidities or complications using a GS, most of the times the medical record.

Design 3b- **all or some kinds of diseases in patients who underwent a specific procedure:** 24 studies: [93-116]

(A sample of) all admissions in which the patient underwent a specific procedure is obtained and for these selected cases a verification takes place of principal diagnosis, comorbidity or complication data using a GS, most of the times the medical record.

The collection of studies contained three intervention studies: doctors that code [34], a discharge letter-linked diagnosis registration [13] and a database for clinical use as origin of discharge data [33].

In general one can conclude that the studies were very heterogeneous and that there is no standard procedure of how to determine diagnostic data quality. However, each study can be classified in one of the six typical designs that can be distinguished based on the range of diseases evaluated and the type of GS used.

6.3.4 Values for completeness and correctness and relation with kind of diagnoses, settings and evaluation methods

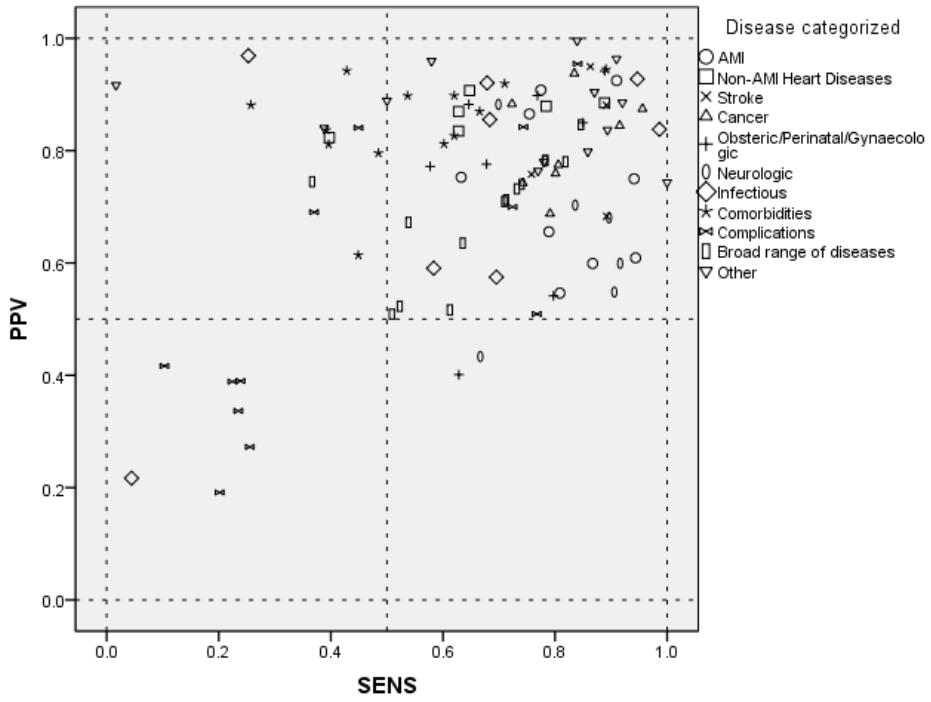

Figure 3: Scatter plot of SENS and PPV for Disease Categories

Figure 3 shows the SENS and PPV of each study for the disease categories that we distinguished. Figure 4 shows the mean SENS and PPV with 95% CI (random-effects model) of all studies together and of subgroups determined by the covariates 'disease category', 'range of diseases', 'years of diagnostic coding', 'country', 'GS' and GS construction. The subgroup analyses for country and GS were limited to the studies that evaluated a single disease (n=47) since only within this set of diseases the four groups of countries and three types of GS's were reasonably distributed over the studies. Studies that evaluated all kinds of diseases (n=18) or diseases in a clinical subgroup (n=36) were almost always done in the USA or other Anglo Saxon countries and not in Scandinavian or other European countries. For the effect of blinding and recoding for GS construction we of course

limited the set to studies that used the medical record as GS and we further excluded all 'Single disease' studies since these studies were overrepresented in the non-blinded re-abstracting studies. The I^2-statistic for all studies together as well as for subgroups of studies was almost always > 98% which means that the studies (even when limited to subgroups based on common characteristics) were highly heterogeneous.

Complications and comorbidities had a statistically significantly lower mean SENS than AMI, Stroke, Cancer and Neurological diseases ($p<0.01$) had. Complications also had a statistically significantly lower mean PPV than Stroke, Cancer, Comorbidities and Other single diseases ($p<0.01$) had. Studies that evaluated diagnoses in clinical subgroups (21 of these 36 studies evaluated comorbidities or complications) had a statistically significantly lower mean SENS than studies that evaluated a single disease or all kinds of diseases combined. Complication and comorbidity studies were strongly correlated with clinical subgroup studies and secondary diagnoses. Studies that only evaluated secondary diagnoses (22 of these 23 studies evaluated comorbidities or complications) had a statistically significantly lower mean SENS (0.50; 95%CI 0.41 – 0.59) than studies that evaluated only principal diagnoses (0.73; 95%CI 0.70 – 0.77, n=14) or a combination of principal and secondary diagnoses (0.72; 95%CI 065 – 0.79, n=64) (results not presented in Figure 4).

We did not find a statistically significant improvement in later years compared to earlier years of coding ('years' as a dichotomous factor), while the characteristics of the studies (disease categories, countries, GS used etc.) in earlier and later years were quite comparable. Related to the year of coding, also the version of ICD had no influence on SENS or PPV (results not shown here). With regards to countries, no statistically significant difference was found for mean SENS, but mean PPV was statistically significantly higher for Scandinavian studies ($p<0.01$).

There were no statistically significant differences between types of GS. Of studies that used the medical record as GS, studies with recoding had statistically significantly lower mean PPV than studies with blinded re-abstracting ($p<0.01$). Application of clinical criteria instead of ICD-criteria for re-abstracting or recoding had no effect on SENS or PPV (results not shown here).

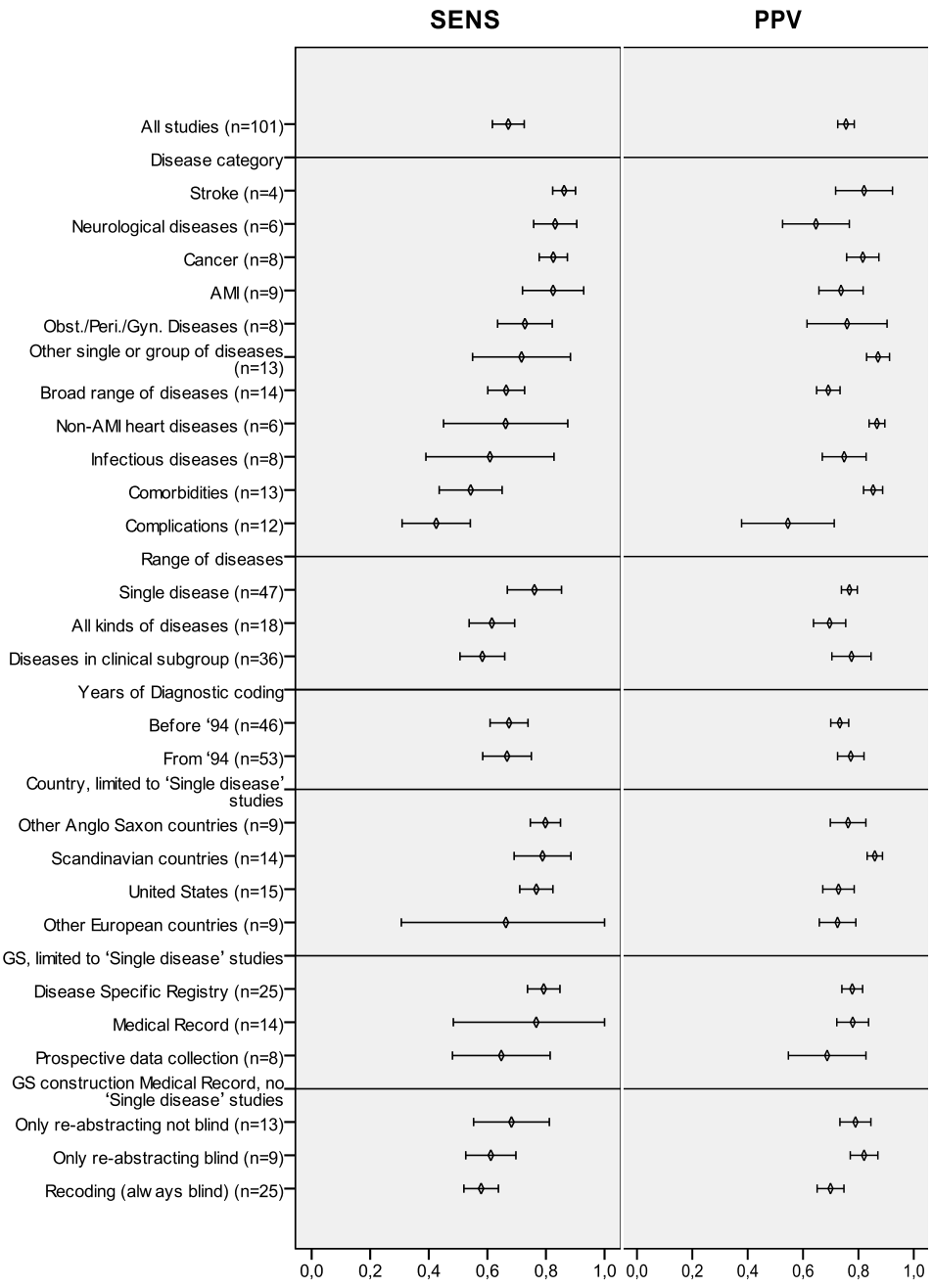

Figure 4: Forrest plot of SENS and PPV with mean and 95% CI (random-effects model) for all studies together and for subgroups based on covariates.

6.3.5 Determinants of data quality reported by the researchers

Several authors reported on determinants of data quality obtained via a statistical analysis like Pearson Chi-square tests or multivariate regression analyses. We categorized these determinants in diagnostic type dependent, disease dependent, disease manifestation dependent, patient dependent and hospital dependent determinants. Since these determinants of the quality of diagnostic data were obtained in separate studies they may not be generalizable.

With regards to the diagnosis dependent determinants it was found that data quality is higher for:

- principal than for secondary diagnoses [50, 107].

With regards to the disease dependent determinants it was found that data quality is higher for:

- more severe than for less severe diseases [104, 112];
- symptomatic than for asymptomatic diseases [103];
- common than for rare disorders [24].

With regards to the disease manifestation dependent determinants it was found that data quality is higher for:

- patients in a later stage of a disease [65];
- patients having longer present [48], more frequently occurring [48], more severe [81], or more [78] manifestations of the disease;
- patients having no history of the disease [51] (explanation: true cases of AMI with history were often falsely coded as CHF);
- patients having a disease confirmed with an important test than for patients where such a test had not been performed [78].

With regards to patient dependent determinants it was found that data quality is higher for:

- infants than for older children [55];
- neonatal than for maternal patients [115];
- patients discharged alive than for patients who died during hospitalization [81] (possible explanation: when patients died the need to code completely was less felt);
- patients who also have comorbidities [55, 65];
- patients not having other, more severe diseases [43];

- patients undergoing surgery for the disease [65, 117];
- patients having a clear risk-factor for the disease (e.g. smoking for AMI) [51].

Some findings regarding patient dependent determinants are conflicting. Data quality is found to be higher for younger patients [40, 70, 78] on one hand and for patients over 63 [37] or 64 years old [40, 104] on the other hand. The latter may be due to the influence of the Medicare reimbursement system for patients over 64 years old where DHDD are necessary for payment and the on average higher disease severity in patients over 64 years old. Some studies found that data quality is higher for male patients [51, 63, 69], other studies for female patients [37]. Some studies found that data quality is higher for white patients [69], other studies for non-Hispanic white patients [37]. This can possibly be explained by the fact that acute myocardial infarction (AMI) is better diagnosed in white male patients than in female patients [118] and non-(Hispanic) white patients [119], resulting in higher data quality for white males in case of AMI. As a consequence, the ICD-code for heart failure would be used more in female and non-(Hispanic) white patients, resulting in higher SENS (and thus data quality) for this code. Both shorter length of stay (LOS) and longer LOS were associated with better data quality. Shorter LOS is on average correlated with less comorbidity which is often underreported. This can explain why a shorter LOS results in higher completeness. On the other hand, a longer LOS can also lead to higher completeness since in general more test results are available before discharge of the patient [22].

With regards to hospital dependent determinants it was found that data quality is higher for:

- public than for private hospitals [29, 41, 90];
- teaching/university than for non-teaching/university hospitals [94, 107];
- regional than for local hospitals [63];
- urban than for rural hospitals [90];
- big than for small hospitals [46];
- hospitals having a higher volume of procedures [108].

Several multicenter studies reported differences in data quality between hospitals or regions for the same diseases. Some researchers reported differences between specialties [40], others between hospitals or regions [31, 63, 65, 68, 69, 78, 89]. Often it is not clear where these differences come from, but authors speculate about reasons that have to do with the local coding practices and settings: differences in diagnostic

practices, differences in case-mix, clarity and structure of documentation of diagnoses in the medical record, training of medical record coders, access to patient data, time per case to code and coding instructions.

Some researchers explicitly reported factors that did not influence data quality: year of coding [101], insurance status [60], code position [39], age [101], gender [40, 101] and ethnicity [60].

Of the three studies that evaluated interventions to improve data quality, Yeoh et al [34] found that the participation of doctors in coding leads to higher accuracy (SENS and PPV). Prins et al [13], however, could not prove their hypothesis that the participation of doctors in coding by means of a discharge letter-linked diagnosis registration would improve data quality. Van Walraven et al [33] found that coding by doctors with the help of a clinical database (instead of the standard chart review by medical record coders) significantly improved completeness and correctness of secondary diagnoses.

6.3.6 Impact of data quality on medical practice assessment

Losina et al [108] reported a substantial overestimation of the true effect of avascular necrosis (AVN) on the risk of perioperative dislocation when using AVN diagnoses derived from hospital discharge data compared to this risk when using diagnoses derived from medical record data. They also found that hospital discharge diagnoses led to an 80% overestimation of the association between low functional status three years after total hip replacement and rheumatoid arthritis. This finding was the result of selection bias due to the more sensitive coding of severe cases.

Rinaldi et al [80] observed a large difference in case fatalities in FP and FN stroke cases (respectively, one month after onset of the disease, 32.7%–36.8% vs. 6.9%–21.1%). According to them, this may signify that patients could be misclassified on the basis of clinical severity, by coding different high mortality disorders as stroke cases, and not coding minor strokes as such. As a consequence the case fatality rate of ischemic stroke could be overestimated.

Sapsford et al [46] on the other hand, found that hospital coding misses a substantial proportion (22.5%) of AMI cases, but without any apparent systematic bias, and thus provides a suitably representative and robust basis for audit.

Romano et al [111] found that half of the difference in risk-adjusted complication rates between low and high outlier hospitals was attributable to reporting variation.

6.3.7 Appropriateness of the data for the objectives of the studies

Based on the conclusions of the authors in each individual study, we qualified the authors' opinion about the appropriateness of the DHDD for their specified (secondary) purposes as 'yes', 'no' or 'doubts'**.** The purposes could be categorized into 'Financial', 'Quality of Care', 'Research' and 'Surveillance'. Then, we computed the mean SENS and PPV with 95% CI (random-effects model) for the different qualifications of appropriateness (not for Financial purpose due to low number of studies: n=5). It appeared that the criteria for the appropriateness for quality of care purposes were stricter than for other purposes, see Figure 5. If we take the lower limit of the 95% CI as a minimum, SENS and PPV should both be at least 85% for quality of care purposes. In only 13 [19, 34, 40, 48, 49, 53, 59, 66, 70, 89, 91, 96, 108] of all 101 included studies the data quality met this criterion.

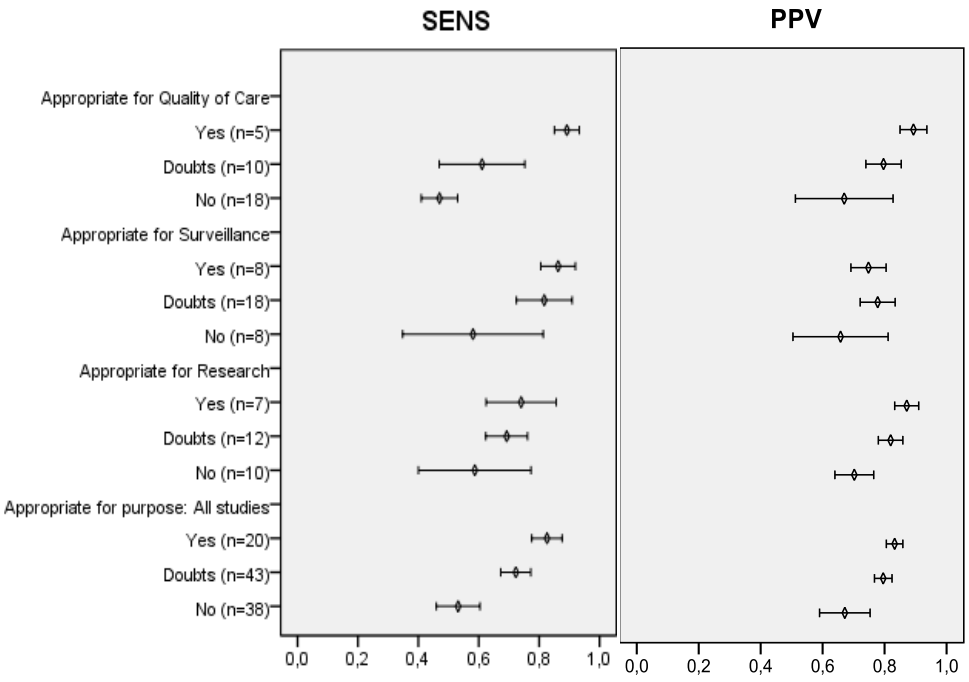

Figure 5: Forest plot of Mean SENS and PPV with 95% CI (random-effects model) by appropriateness for Quality of Care, Surveillance and Research purposes.

6.4 DISCUSSION

6.4.1 Heterogeneity of studies

We carried out a systematic review to get insight in the appropriateness of diagnostic hospital discharge data for medical practice assessment. The retrieved studies were very heterogeneous due to variations in settings and coding practices, the type of diagnoses and diseases studied and the data quality evaluation methods used. In Scandinavian countries for example, the diagnoses are often encoded by physicians, whereas in other countries this is often done by a medical record coder. Some studies evaluated the data quality of a broad range of diseases, whereas other studies evaluated data quality of a single disease or diseases within a clinically defined subgroup of patients. The medical record, disease specific registries and prospectively collected data were used as GS. We could distinguish six typical designs based on range of diseases studied and GS used.

6.4.2 Findings

There was a great variability in observed SENS and PPV between the individual studies. Our meta-analysis using the random-effects model resulted in an average SENS of 0.67 (95% CI: 0.62-0.73) and PPV of 0.76 (95% CI: 0.73-0.79). For some single disease studies SENS was significantly higher and for comorbidity and complication studies SENS was significantly lower. When the medical record was used to construct a GS, recoding of diagnoses gave a significantly lower PPV than blind re-abstracting of diagnoses. PPV was significantly higher for Scandinavian countries compared to non-Scandinavian countries. Determinants of data quality, obtained via statistical analysis, could be categorized in diagnostic type dependent, disease dependent, disease manifestation dependent, patient dependent and hospital dependent determinants. Severity of a disease and severity of the manifestations of a disease lead to a higher chance of the disease to be coded. Studies that evaluated the effect of data quality on quality of care estimates showed that the use of diagnostic hospital discharge data can easily lead to an overestimation of sentinel outcomes. The cases selected for medical practice assessment based on these data represented the more severe cases. In studies with quality of care purposes, the authors stated that the diagnostic data were appropriate when both SENS and PPV were at least 85%. Only 13% of all included studies fulfilled this criterion.

6.4.3 Coding performance and data quality

SENS and SPEC are an expression of coding performance and are independent of prevalence. They indicate how well coders are able to detect and code diagnoses (SENS) and how well coders are able not to code non-existing diagnoses (SPEC). As such, SENS and SPEC also are an expression of data quality. However, in most studies SPEC is not reported or roughly estimated, because for a correct measurement a lot of (extra) work has to be done. In this case it should be evaluated for all admissions whether (certain) diseases certainly not present are actually not coded (the TN's). However, theoretically it can be argued that SPEC in most of the studies is very high, about 0.98 or 0.99. Studies in which SPEC was measured, showed this. Romano et al [113] for example wrote: *"We did not report specificity ... because this parameter was never below 97%, and nearly always exceeded 99%"*. PPV is determined by the values of SENS, SPEC and prevalence. Since prevalence is a factor that cannot be influenced by the coding process, PPV is not an expression of coding performance. Nevertheless, PPV is a very valuable measure for data quality and determines the informational value or usability of the codes in a specific setting, namely the chance that the code is a true representation of a diagnosis.

6.4.4 Do differences in SENS and PPV indicate different coding performance?

Studies evaluating some specific diseases showed higher mean values of SENS than other studies. This was also true for 'single disease' and 'principal diagnosis' studies (these studies were highly correlated with specific diseases). We were curious whether the higher mean SENS indeed was due to a better coding performance or instead indicated a more generous (but not better) coding (a threshold effect). In case of better coding we – mathematically - expect also a higher PPV. The magnitude of this increase depends on the value of the prevalence [120]. In case of more generous coding we expect a higher SENS but lower SPEC (threshold effect), and thus a less than – mathematically - expected increase or even a decrease of PPV. The lower the prevalence, the bigger the chance that PPV decreases. Since the often unknown prevalence of many diseases will be quite low, a more generous coding will usually lead to an increase of SENS and a decrease of PPV. The higher SENS for some subgroups in our analyses were accompanied with an unchanged or increased (but not statistically significant) PPV. Since differences in prevalence between subgroups may exist [120], we only cautiously conclude that coding performance is better in these subgroups. Differences

between subgroups of studies are observational in nature and are prone to bias and confounding [121]. Other factors could be responsible for the differences.

We found that mean PPV is significantly higher for Scandinavian countries compared to the USA and other countries while mean SENS did not significantly differ. We also found that studies using a recoded medical record as GS had a significantly lower mean PPV, but no significantly different mean SENS than studies using a blindly re-abstracted medical record. If prevalence would explain the differences in PPV, then coding performance would not differ between the two groups of studies. Coding performance can be represented by the diagnostic odds ratio (DOR). This DOR can be expressed in terms of SENS and SPEC with the formula: (Sensitivity/(1-Sensitivity))/((1-Specificity)/Specificity) and is independent of prevalence [122]. The property of DOR that we will use is that DOR will be almost constant for the different values of SENS and SPEC [122, 123] that result from the threshold effect. Different DOR values result from different coding performances. We will now show that the assumption of equal coding performance (no differences in DOR) will lead to improbable differences in the prevalence of the considered diseases. If, for example, we assume a DOR value of 100 (which is a typical value) for both the USA and Scandinavian studies, then SPEC can be calculated from the values of DOR and SENS, giving a SPEC of 0.968 for the USA and 0.964 for Scandinavia. The PPV can then be used to calculate the prevalence of the disease, resulting in a prevalence of 0.10 (USA) and 0.218 (Scandinavia) respectively (see column 4 and 5 of Table 4). For the assumed DOR value this would mean that prevalence of the disease in Scandinavian countries has to be more than two times the prevalence of the disease in the USA, which is unlikely. Similar analyses with values of DOR ranging from 5 to 10000 showed a factor 1.5 to 2.5 difference in the resulting prevalences.

If we assume that the prevalence of diseases in Scandinavia and the USA are the same then we get values for DOR that are rather different, indicating a better coding performance in the Scandinavian countries. If for example we assume a prevalence (PREV) of 0.10 for both the USA and Scandinavian studies, then SPEC is 0.968 and 0.986 (calculated from the PPV values), resulting in a DOR of 100 (USA) and 257 (Scandinavia) respectively (see column 6 and 7 of Table 4). This would mean that DOR for Scandinavian countries is more than two and a half times the DOR of the USA, which is unlikely when the coding performance would be the same. Similar analyses with values of PREV ranging from 0.001 to 0.5

showed a factor 2.5 - 3 difference in resulting DOR's. Thus the different values of SENS and PPV found in the studies in the USA and Scandinavia are probably not only due to a threshold effect and therefore indicate a better coding performance in Scandinavian countries, possibly due to physicians' involvement in the coding process.

Applying the same sensitivity analysis on recoding compared to blinded re-abstracting also produces an unlikely two times lower prevalence and DOR for recoding (see Table 4). Recoding therefore leads to a lower observed coding performance than re-abstracting. Dixon et al [22] and Prins et al [13] showed that a substantial part of the disagreements between DHDD and the GS could be explained by the fact that DHDD represents conditions that were closely related to, but not covered by the codes representing the GS. Compared to a recoded GS, matching with a re-abstracted GS leaves more room for subjective assessment of the DHDD and may be influenced by knowledge of the GS, known as test review bias [124]. This makes it plausible to assume that, given the observed PPV's, re-abstracting compared to recoding leads to an overestimation of data correctness.

Table 4. Effects of recoding and country on the sensitivity and positive predictive value. In addition, sensitivity analyses for effects on specificity and prevalence/DOR using a baseline value for DOR and prevalence respectively.

Factor	**Mean SENS**	**Mean PPV**	**Specificity and prevalence if DOR stays the same (100)**		**Specificity and DOR if prevalence stays the same (0.10)**	
			specificity	*prevalence*	*specificity*	*DOR*
USA	0.7670	0.7281	0.9681	0.1002	0.9682	100.16
Scandinavia	0.7883	0.8593	0.9641	0.2176	0.9857	256.67
Recoding	0.5786	0.6992	0.9865	0.0516	0.9724	48,30
Re-abstracting, blinded	0.6118	0.8204	0.9845	0.1038	0.9851	104.34

DOR = diagnostic odds ratio

6.4.5 Unexpected Findings

For rare diseases we expect very low PPV's. This is the logical consequence of our expectation of a similar high SPEC but a much lower PREV than for common diseases. However, several authors report PPV's for rare diseases comparable to, or only somewhat lower than for common diseases (e.g. Bogliun et al [50]: PPV of 0.55

for Guillan-Barrè Syndrome; Beghi et al [36]: PPV of 0.60 for ALS (Amyotropic Lateral Sclerosis); Chancellor et al [52]: PPV of 0.70 for Motor Neuron Disease). Possibly, medical record coders may not consider a code for a rare disease unless there is a very clear indication in the medical record, while they may consider a code for a common disease even when there is a less clear indication. This will result in a higher SPEC for rare diseases compared to common diseases which can explain the higher than expected PPV for rare diseases.

6.4.6 Gold Standard

The observed data quality is partly the result of the GS quality which is debatable. Romano et al [112] showed that not only coding, but also recoding for GS construction is susceptible to interrater variability. They compared DHDD and independently recoded ICD-9-CM data with complications abstracted from the medical record by clinicians using detailed criteria. The recoded data captured 56% of all severe complications, whereas DHDD data captured 44%. According to Dixon et al [22], expert recoders may have had access to information added to the notes after the local coding had been done. This indicates that retrospective GS construction and original coding may not be based on the same information. One can also question whether the GS really includes all true cases. Sometimes the capture-recapture method is used, e.g. Nielsen 1996 [66], leading to a somewhat higher estimate of the number of true diagnoses and thus to a lower SENS. Some studies possibly measured data accuracy at the patient level including several hospitalizations, although this is not clearly specified. If so, it can lead to overestimation of SENS and PPV from the point of view of data quality at admission level.

6.4.7 Improvement over the years

We did not find an improvement in data quality over the years. However, Pajunen et al [68] saw an increase in data quality in Finland in the years 1998-2002 compared to 1988-1992. Leibson et al [125] showed an increase of SENS over time (1970 – 1989) in the USA while PPV remained the same, due to the introduction of the prospective payment system in 1982. In our meta-analysis, data from only few studies date from 1981 or earlier, so we were not able to analyze the effect of the prospective payment system. We also did not find an improvement of data quality for successive versions of the ICD. Quan et al [126] compared ICD-9-CM and ICD-

10 coding of the same set of hospitalizations, but could also not demonstrate that ICD-10 did perform better than ICD-9-CM.

6.4.8 International Classification of Diseases

Some authors criticize the limitations of the ICD. Romano et al [111] mentioned the vague definitions of ICD codes, Bogliun et al [50] stated that Guillain-Barre syndrome is included in code 357.0 and does not have a separate diagnostic code; Tetsche et al [70] warned that borderline tumors could not be excluded from the code for ovarian cancer, and McNaughton et al [75] stated in general that in ICD there is no one single code for a disorder, the codes are not mutually exclusive and definitions are more based on pathological than on clinical information. Cimino [4] argues that some of the problems of ICD-coded data relate to the design aspects of the terminology, such as lack of detail limited by the restrictive nature of the numbering system, the strict hierarchical structure and changes of meaning of terms when the terminology is updated.

6.4.9 Limitations of the study

We limited our review to a search in PubMed and to studies written in the English language. Queries not limited to the English language, could have led to the inclusion of some additional studies from countries that also reported studies in the English language, like Denmark, Spain and Italy. However, we were not able to understand these languages and to use such studies for the review. We developed the search strategy by an iterative process learning about relevant subject headings and text words to identify additional search terms from retrieved studies that seemed potentially relevant. We also selected some key articles that met the inclusion criteria for the review to note common text words and their variants (such as synonyms, abbreviations and spelling variants) as well as subject headings the database indexers assigned to the articles. This is in line with the advice of the Cochrane Collaboration [127]. However, there are not many subject headings with a meaning related to our topic, so we used many text words and tried many queries in order to find a balance between the number of articles needed to read and a sensitive search.

Publication bias may play an important role in this field. Settings where value is placed on data quality have a higher chance of critically evaluating their diagnostic

data and publish the results. Therefore, our overall picture based on publications may even be too optimistic.

The strength of our review is that we could analyze the effects of factors that could not be analyzed in individual studies, e.g. the type of GS used. So, we could also determine typical designs to evaluate diagnostic data quality.

6.4.10 Can hospital discharge data be used for medical practice assessment?

The observed data quality of diagnostic hospital discharge data is a function of the diagnostic process, diagnoses documentation, coding practice, characteristics of the disease, manifestation of the disease in patients, prevalence of the diagnosis, and the way data quality is measured. Studies were highly heterogeneous with respect to these factors and showed highly variable data quality. The effect of several forms of bias remains unclear.

We conclude that quality of diagnostic hospital discharge data leaves much to be desired. On average, completeness is about 67% and correctness about 76%. For some single diseases it is somewhat better, for comorbidities and especially complications it is far worse. In only 13% of the studies completeness and correctness are both at least 85%. Despite all the efforts to improve data quality, it did in general not lead to better data quality in the course of the years. The use of diagnostic hospital discharge data can easily lead to a biased idea of the quality of care. On the one hand cases selected for quality of care assessment possibly represent the more severe cases with a chance of overestimating sentinel outcomes, especially when comorbidities are also underreported. On the other hand, complication data are usually incomplete thus leading to an underestimation of sentinel outcomes. These shortcomings can differ between settings and thus complicates comparisons between hospitals or geographical areas. True variability in quality of care between settings can easily be overshadowed by the unknown variability in data quality which makes it very difficult to interpret the observed variability in the quality of care based on diagnostic hospital discharge data.

Despite the moderate data quality, we think that the use of these data by physicians to assess their own medical practice may be useful. Of all stakeholders, physicians have the best insight into the quality of their diagnostic data and its implications for the interpretation of performance indicators. Physicians are able to compare the information with their own experiences and can reason about what the information

means for the quality of their care. However, this will only work when they are willing to critically reflect on their own medical practice.

ACKNOWLEDGEMENT

We thank prof. dr. A.H. Zwinderman for his support of the statistical analyses.

REFERENCES

1. De Coster C, Quan H, Finlayson A, Gao M, Halfon P, Humphries KH, et al. Identifying priorities in methodological research using ICD-9-CM and ICD-10 administrative data: report from an international consortium. BMC Health Serv Res. 2006;6:77.
2. National Center for Health Statistics. International Classification of Diseases, 9th Revision, Clinical Modification, 6th edition. Hayattsville, Maryland: NCHS; 2003.
3. World Health Organization. International Statistical Classification of Diseases and Related Health Problems; Tenth Revision. Geneva: World Health Organization; 1992.
4. Cimino JJ. High-quality, standard, controlled healthcare terminologies come of age. Methods Inf Med. 2011;50(2):101-4.
5. Demlo LK, Campbell PM, Brown SS. Reliability of information abstracted from patients' medical records. Med Care. 1978 Dec;16(12):995-1005.
6. Prins H, Kruisinga FH, Buller HA, Zwetsloot-Schonk JH. Availability and usability of data for medical practice assessment. Int J Qual Health Care. 2002 Apr;14(2):127-37.
7. McKee M. Routine data: a resource for clinical audit? Quality in Health Care. 1993(2):104-11.
8. Safran C, Chute CG. Exploration and exploitation of clinical databases. International Journal of Bio-Medical Computing. 1995;39(1):151-6.
9. Wyatt J. Acquisition and use of clinical data for audit and research. Journal of Evaluation in Clinical Practice. 1995;1(1):15-27.
10. Iezzoni LI. Assessing quality using administrative data. Annals of Internal Medicine. 1997;127(8 (Pt 2)):666-74.
11. Coory M, Thompson B, Baade P, Fritschi L. Utility of routine data sources for feedback on the quality of cancer care: an assessment based on clinical practice guidelines. BMC Health Serv Res. 2009;9:84.
12. Romano PS. Can administrative data be used to compare the quality of health care? Med Care Rev. 1993 Winter;50(4):451-77.
13. Prins H, Buller HA, Zwetsloot-Schonk JH. Effect of discharge letter-linked diagnosis registration on data quality. Int J Qual Health Care. 2000 Feb;12(1):47-57.

14. Hogan WR, Wagner MM. Accuracy of data in computer-based patient records. J Am Med Inform Assoc. 1997 Sep-Oct;4(5):342-55.

15. Pine M, Jordan HS, Elixhauser A, Fry DE, Hoaglin DC, Jones B, et al. Modifying ICD-9-CM coding of secondary diagnoses to improve risk-adjustment of inpatient mortality rates. Medical Decision Making. 2009 Jan-Feb;29(1):69-81.

16. Quan H, Parsons GA, Ghali WA. Assessing accuracy of diagnosis-type indicators for flagging complications in administrative data. J Clin Epidemiol. 2004 Apr;57(4):366-72.

17. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials. 1986 Sep;7(3):177-88.

18. Borenstein M, Hedges L, Rothstein H. Meta-analysis: fixed effect vs. random effects. 2007 [updated 2007; cited 2012 May 20th]; Available from: http://www.sciencedownload.net/demodownload/Fixed%20effect%20vs.%20random%20effects.pdf.

19. Madsen M, Davidsen M, Rasmussen S, Abildstrom SZ, Osler M. The validity of the diagnosis of acute myocardial infarction in routine statistics: a comparison of mortality and hospital discharge data with the Danish MONICA registry. J Clin Epidemiol. 2003 Feb;56(2):124-30.

20. Demlo LK, Campbell PM. Improving hospital discharge data: lessons from the National Hospital Discharge Survey. Med Care. 1981 Oct;19(10):1030-40.

21. Cleary R, Beard R, Coles J, Devlin B, Hopkins A, Schumacher D, et al. Comparative hospital databases: value for management and quality. Qual Health Care. 1994 Mar;3(1):3-10.

22. Dixon J, Sanderson C, Elliott P, Walls P, Jones J, Petticrew M. Assessment of the reproducibility of clinical coding in routinely collected hospital activity data: a study in two hospitals. J Public Health Med. 1998 Mar;20(1):63-9.

23. Doremus HD, Michenzi EM. Data quality. An illustration of its potential impact upon a diagnosis-related group's case mix index and reimbursement. Med Care. 1983 Oct;21(10):1001-11.

24. Elkin PL, Ruggieri AP, Brown SH, Buntrock J, Bauer BA, Wahner-Roedler D, et al. A randomized controlled trial of the accuracy of clinical record retrieval using SNOMED-RT as compared with ICD9-CM. Proc AMIA Symp. 2001:159-63.

25. Fisher ES, Whaley FS, Krushat WM, Malenka DJ, Fleming C, Baron JA, et al. The accuracy of Medicare's hospital claims data: progress has been made, but problems remain. Am J Public Health. 1992 Feb;82(2):243-8.

26. Frohnert BK, Lussky RC, Alms MA, Mendelsohn NJ, Symonik DM, Falken MC. Validity of hospital discharge data for identifying infants with cardiac defects. J Perinatol. 2005 Nov;25(11):737-42.

27. Kashner TM. Agreement between administrative files and written medical records: a case of the Department of Veterans Affairs. Med Care. 1998 Sep;36(9):1324-36.

28. Mehanni M, Loughman E, Allwright SP, Prichard J. The hospital in-patient enquiry scheme: a study of data accuracy and capture. Ir Med J. 1995 Jan-Feb;88(1):24-6.

29. Preen DB, Holman CD, Lawrence DM, Baynham NJ, Semmens JB. Hospital chart review provided more accurate comorbidity information than data from a general practitioner survey or an administrative database. J Clin Epidemiol. 2004 Dec;57(12):1295-304.

30. Quan H, Parsons GA, Ghali WA. Validity of information on comorbidity derived rom ICD-9-CCM administrative data. Med Care. 2002 Aug;40(8):675-85.

31. Ricketts D, Hartley J, Harries W, Hitchin D. Who should code orthopaedic inpatients? A comparison of junior hospital doctors and coding clerks. Ann R Coll Surg Engl. 1993 Nov;75(6 Suppl):203-6.

32. Smith MW. Hospital discharge diagnoses: how accurate are they and their international classification of diseases (ICD) codes? N Z Med J. 1989 Sep 27;102(876):507-8.

33. van Walraven C, Demers SV. Coding diagnoses and procedures using a high-quality clinical database instead of a medical record review. J Eval Clin Pract. 2001 Aug;7(3):289-97.

34. Yeoh C, Davies H. Clinical coding: completeness and accuracy when doctors take it on. BMJ. 1993 Apr 10;306(6883):972.

35. Aronsky D, Haug PJ, Lagor C, Dean NC. Accuracy of administrative data for identifying patients with pneumonia. Am J Med Qual. 2005 Nov-Dec;20(6):319-28.

36. Beghi E, Logroscino G, Micheli A, Millul A, Perini M, Riva R, et al. Validity of hospital discharge diagnoses for the assessment of the prevalence and incidence of amyotrophic lateral sclerosis. Amyotroph Lateral Scler Other Motor Neuron Disord. 2001 Jun;2(2):99-104.

37. Goff DC, Jr., Pandey DK, Chan FA, Ortiz C, Nichaman MZ. Congestive heart failure in the United States: is there more than meets the I(CD code)? The Corpus Christi Heart Project. Arch Intern Med. 2000 Jan 24;160(2):197-202.

38. Johnson RL, Gabella BA, Gerhart KA, McCray J, Menconi JC, Whiteneck GG. Evaluating sources of traumatic spinal cord injury surveillance data in Colorado. Am J Epidemiol. 1997 Aug 1;146(3):266-72.

39. Khand AU, Shaw M, Gemmel I, Cleland JG. Do discharge codes underestimate hospitalisation due to heart failure? Validation study of hospital discharge coding for heart failure. Eur J Heart Fail. 2005 Aug;7(5):792-7.

40. Leone MA, Capponi A, Varrasi C, Tarletti R, Monaco F. Accuracy of the ICD-9 codes for identifying TIA and stroke in an Italian automated database. Neurol Sci. 2004 Dec;25(5):281-8.

41. Lofthus CM, Cappelen I, Osnes EK, Falch JA, Kristiansen IS, Medhus AW, et al. Local and national electronic databases in Norway demonstrate a varying degree of validity. J Clin Epidemiol. 2005 Mar;58(3):280-5.

42. Mascioli SR, Jacobs DR, Jr., Kottke TE. Diagnostic criteria for hospitalized acute myocardial infarction: the Minnesota experience. Int J Epidemiol. 1989 Mar;18(1):76-83.

43. Movig KL, Leufkens HG, Lenderink AW, Egberts AC. Validity of hospital discharge International Classification of Diseases (ICD) codes for identifying patients with hyponatremia. J Clin Epidemiol. 2003 Jun;56(6):530-5.

44. Pladevall M, Goff DC, Nichaman MZ, Chan F, Ramsey D, Ortiz C, et al. An assessment of the validity of ICD Code 410 to identify hospital admissions for myocardial infarction: The Corpus Christi Heart Project. Int J Epidemiol. 1996 Oct;25(5):948-52.

45. Rodella S, Picoco C, Stanzial F, Turazza M, Donati LF. Cancer registration: a feasibility study in northern Italy. The Collaborative Group of Pathologists for Cancer Registration in Verona. Eur J Cancer. 1994;30A(8):1128-33.

46. Sapsford RJ, Lawrance RA, Dorsch MF, Das R, Jackson BM, Morrell C, et al. Identifying acute myocardial infarction: effects on treatment and mortality, and implications for National Service Framework audit. Qjm. 2003 Mar;96(3):203-9.

47. Segal JB, Powe NR. Accuracy of identification of patients with immune thrombocytopenic purpura through administrative records: a data validation study. Am J Hematol. 2004 Jan;75(1):12-7.

48. Vestergaard M, Obel C, Henriksen TB, Christensen J, Madsen KM, Ostergaard JR, et al. The Danish National Hospital Register is a valuable study base for epidemiologic research in febrile seizures. J Clin Epidemiol. 2006 Jan;59(1):61-6.

49. Austin PC, Daly PA, Tu JV. A multicenter study of the coding accuracy of hospital discharge administrative data for patients admitted to cardiac care units in Ontario. Am Heart J. 2002 Aug;144(2):290-6.

50. Bogliun G, Beghi E. Validity of hospital discharge diagnoses for public health surveillance of the Guillain-Barre syndrome. Neurol Sci. 2002 Sep;23(3):113-7.

51. Boyle CA, Dobson AJ. The accuracy of hospital records and death certificates for acute myocardial infarction. Aust N Z J Med. 1995 Aug;25(4):316-23.

52. Chancellor AM, Swingler RJ, Fraser H, Clarke JA, Warlow CP. Utility of Scottish morbidity and mortality data for epidemiological studies of motor neuron disease. J Epidemiol Community Health. 1993 Apr;47(2):116-20.

53. Davenport RJ, Dennis MS, Warlow CP. The accuracy of Scottish Morbidity Record (SMR1) data for identifying hospitalised stroke patients. Health Bull (Edinb). 1996 Sep;54(5):402-5.

54. Ellekjaer H, Holmen J, Kruger O, Terent A. Identification of incident stroke in Norway: hospital discharge data compared with a population-based stroke register. Stroke. 1999 Jan;30(1):56-60.

55. Graff Stensballe L, Kristensen K, Nielsen J, Aaby P. Diagnosis coding in The Danish National Patient Registry for respiratory syncytial virus infections. Scand J Infect Dis. 2005;37(10):747-52.

56. Hexter AC, Harris JA, Roeper P, Croen LA, Krueger P, Gant D. Evaluation of the hospital discharge diagnoses index and the birth certificate as sources of information on birth defects. Public Health Rep. 1990 May-Jun;105(3):296-307.

57. Hunt JP, Baker CC, Fakhry SM, Rutledge RR, Ransohoff D, Meyer AA. Accuracy of administrative data in trauma. Surgery. 1999 Aug;126(2):191-7.

58. Hunt JP, Cherr GS, Hunter C, Wright MJ, Wang YZ, Steeb G, et al. Accuracy of administrative data in trauma: splenic injuries as an example. J Trauma. 2000 Oct;49(4):679-86; discussion 86-8.

59. Joakimsen RM, Fonnebo V, Sogaard AJ, Tollan A, Stormer J, Magnus JH. The Tromso study: registration of fractures, how good are self-reports, a computerized radiographic register and a discharge register? Osteoporos Int. 2001 Dec;12(12):1001-5.

60. Kahn LH, Blustein J, Arons RR, Yee R, Shea S. The validity of hospital administrative data in monitoring variations in breast cancer surgery. Am J Public Health. 1996 Feb;86(2):243-5.

61. Kennedy GT, Stern MP, Crawford MH. Miscoding of hospital discharges as acute myocardial infarction: implications for surveillance programs aimed at elucidating trends in coronary artery disease. Am J Cardiol. 1984 Apr 1;53(8):1000-2.

62. Madsen KM, Schonheyder HC, Kristensen B, Nielsen GL, Sorensen HT. Can hospital discharge diagnosis be used for surveillance of bacteremia? A data quality study of a Danish hospital discharge registry. Infect Control Hosp Epidemiol. 1998 Mar;19(3):175-80.

63. Mahonen M, Salomaa V, Brommels M, Molarius A, Miettinen H, Pyorala K, et al. The validity of hospital discharge register data on coronary heart disease in Finland. Eur J Epidemiol. 1997 Jun;13(4):403-15.

64. Martin CA, Hobbs MS, Armstrong BK. Identification of non-fatal myocardial infarction through hospital discharge data in Western Australia. J Chronic Dis. 1987;40(12):1111-20.

65. McClish DK, Penberthy L, Whittemore M, Newschaffer C, Woolard D, Desch CE, et al. Ability of Medicare claims data and cancer registries to identify cancer cases and treatment. Am J Epidemiol. 1997 Feb 1;145(3):227-33.

66. Nielsen GL, Sorensen HT, Pedersen AB, Sabroe S. Analyses of data quality in registries concerning diabetes mellitus--a comparison between a population based hospital discharge and an insulin prescription registry. J Med Syst. 1996 Feb;20(1):1-10.

67. Norgaard M, Skriver MV, Gregersen H, Pedersen G, Schonheyder HC, Sorensen HT. The data quality of haematological malignancy ICD-10 diagnoses in a population-based hospital discharge registry. Eur J Cancer Prev. 2005 Jun;14(3):201-6.

68. Pajunen P, Koukkunen H, Ketonen M, Jerkkola T, Immonen-Raiha P, Karja-Koskenkari P, et al. The validity of the Finnish Hospital Discharge Register and

Causes of Death Register data on coronary heart disease. Eur J Cardiovasc Prev Rehabil. 2005 Apr;12(2):132-7.

69. Rosamond WD, Chambless LE, Sorlie PD, Bell EM, Weitzman S, Smith JC, et al. Trends in the sensitivity, positive predictive value, false-positive rate, and comparability ratio of hospital discharge diagnosis codes for acute myocardial infarction in four US communities, 1987-2000. Am J Epidemiol. 2004 Dec 15;160(12):1137-46.

70. Tetsche MS, Norgaard M, Skriver MV, Andersen ES, Lash TL, Sorensen HT. Accuracy of ovarian cancer ICD-10 diagnosis in a Danish population-based hospital discharge registry. Eur J Gynaecol Oncol. 2005;26(3):266-70.

71. Toniolo P, Pisani P, Vigano C, Gatta G, Repetto F. Estimating incidence of cancer from a hospital discharge reporting system. Rev Epidemiol Sante Publique. 1986;34(1):23-30.

72. Warren JL, Feuer E, Potosky AL, Riley GF, Lynch CF. Use of Medicare hospital and physician data to assess breast cancer incidence. Med Care. 1999 May;37(5):445-56.

73. Guevara RE, Butler JC, Marston BJ, Plouffe JF, File TM, Jr., Breiman RF. Accuracy of ICD-9-CM codes in detecting community-acquired pneumococcal pneumonia for incidence and vaccine efficacy studies. Am J Epidemiol. 1999 Feb 1;149(3):282-9.

74. Marrie TJ, Durant H, Sealy E. Pneumonia--the quality of medical records data. Med Care. 1987 Jan;25(1):20-4.

75. McNaughton H, Wadsworth K. Assessing the accuracy of hospital admission and discharge diagnosis of traumatic brain injury in a New Zealand hospital. N Z Med J. 2000 May 26;113(1110):184-6.

76. Moro ML, Morsillo F. Can hospital discharge diagnoses be used for surveillance of surgical-site infections? J Hosp Infect. 2004 Mar;56(3):239-41.

77. Pietila K, Tenkanen L, Manttari M, Manninen V. How to define coronary heart disease in register-based follow-up studies: experience from the Helsinki Heart Study. Ann Med. 1997 Jun;29(3):253-9.

78. Porta M, Costafreda S, Malats N, Guarner L, Soler M, Gubern JM, et al. Validity of the hospital discharge diagnosis in epidemiologic studies of biliopancreatic pathology. PANKRAS II Study Group. Eur J Epidemiol. 2000 Jun;16(6):533-41.

79. Ragnarson Tennvall G, Apelqvist J, Eneroth M. The inpatient care of patients with diabetes mellitus and foot ulcers. A validation study of the correspondence between medical records and the Swedish Inpatient Registry with the consequences for cost estimations. J Intern Med. 2000 Nov;248(5):397-405.

80. Rinaldi R, Vignatelli L, Galeotti M, Azzimondi G, de Carolis P. Accuracy of ICD-9 codes in identifying ischemic stroke in the General Hospital of Lugo di Romagna (Italy). Neurol Sci. 2003 Jun;24(2):65-9.

81. Birman-Deych E, Waterman AD, Yan Y, Nilasena DS, Radford MJ, Gage BF. Accuracy of ICD-9-CM codes for identifying cardiovascular and stroke risk factors. Med Care. 2005 May;43(5):480-5.

82. Geraci JM, Ashton CM, Kuykendall DH, Johnson ML, Wu L. International Classification of Diseases, 9th Revision, Clinical Modification codes in discharge abstracts are poor measures of complication occurrence in medical inpatients. Med Care. 1997 Jun;35(6):589-602.

83. Hsu VP, Staat MA, Roberts N, Thieman C, Bernstein DI, Bresee J, et al. Use of active surveillance to validate international classification of diseases code estimates of rotavirus hospitalizations in children. Pediatrics. 2005 Jan;115(1):78-82.

84. Kohli HS, Knill-Jones RP. How accurate are SMR1 (Scottish Morbidity Record 1) data? Health Bull (Edinb). 1992 Jan;50(1):14-23; discussion 9-31.

85. Lee DS, Donovan L, Austin PC, Gong Y, Liu PP, Rouleau JL, et al. Comparison of coding of heart failure and comorbidities in administrative and clinical data for use in outcomes research. Med Care. 2005 Feb;43(2):182-8.

86. Lydon-Rochelle MT, Holt VL, Cardenas V, Nelson JC, Easterling TR, Gardella C, et al. The reporting of pre-existing maternal medical conditions and complications of pregnancy on birth certificates and in hospital discharge data. Am J Obstet Gynecol. 2005 Jul;193(1):125-34.

87. Mears SC, Bawa M, Pietryak P, Jones LC, Rajadhyaksha AD, Hungerford DS, et al. Coding of diagnoses, comorbidities, and complications of total hip arthroplasty. Clin Orthop Relat Res. 2002 Sep(402):164-70.

88. Mont MA, Mears SC, Jones LC, Rajadhyaksha AD, Krackow AM, Bawa M, et al. Is coding of diagnoses, comorbidities, and complications in total knee arthroplasty accurate? J Arthroplasty. 2002 Sep;17(6):767-72.

89. Parrish KM, Holt VL, Connell FA, Williams B, LoGerfo JP. Variations in the accuracy of obstetric procedures and diagnoses on birth records in Washington State, 1989. Am J Epidemiol. 1993 Jul 15;138(2):119-27.

90. Powell H, Lim LL, Heller RF. Accuracy of administrative data to assess comorbidity in patients with heart disease. an Australian perspective. J Clin Epidemiol. 2001 Jul;54(7):687-93.

91. Romano PS, Mark DH. Bias in the coding of hospital discharge data and its implications for quality assessment. Med Care. 1994 Jan;32(1):81-90.

92. Thomas SK, Brooks SE, Mullins CD, Baquet CR, Merchant S. Use of ICD-9 coding as a proxy for stage of disease in lung cancer. Pharmacoepidemiol Drug Saf. 2002 Dec;11(8):709-13.

93. Azimuddin K, Rosen L, Reed JF, 3rd. Computerized assessment of complications after colorectal surgery: is it valid? Dis Colon Rectum. 2001 Apr;44(4):500-5.

94. Hawker GA, Coyte PC, Wright JG, Paul JE, Bombardier C. Accuracy of administrative data for assessing outcomes after knee replacement surgery. J Clin Epidemiol. 1997 Mar;50(3):265-73.

95. Ballaro A, Oliver S, Emberton M. Do we do what they say we do? coding errors in urology. BJU Int. 2000 Mar;85(4):389-91.

96. Edouard L, Rawson NS. Reliability of the recording of hysterectomy in the Saskatchewan health care system. Br J Obstet Gynaecol. 1996 Sep;103(9):891-7.

97. Best WR, Khuri SF, Phelan M, Hur K, Henderson WG, Demakis JG, et al. Identifying patient preoperative risk factors and postoperative adverse events in administrative databases: results from the Department of Veterans Affairs National Surgical Quality Improvement Program. J Am Coll Surg. 2002 Mar;194(3):257-66.

98. Faciszewski T, Broste SK, Fardon D. Quality of data regarding diagnoses of spinal disorders in administrative databases. A multicenter study. J Bone Joint Surg Am. 1997 Oct;79(10):1481-8.

99. Cadwallader HL, Toohey M, Linton S, Dyson A, Riley TV. A comparison of two methods for identifying surgical site infections following orthopaedic surgery. J Hosp Infect. 2001 Aug;48(4):261-6.

100. Faciszewski T, Johnson L, Noren C, Smith MD. Administrative databases' complication coding in anterior spinal fusion procedures. What does it mean? Spine. 1995 Aug 15;20(16):1783-8.

101. Flum DR, Koepsell TD. Evaluating diagnostic accuracy in appendicitis using administrative data. J Surg Res. 2005 Feb;123(2):257-61.

102. Fox KM, Reuland M, Hawkes WG, Hebel JR, Hudson J, Zimmerman SI, et al. Accuracy of medical records in hip fracture. J Am Geriatr Soc. 1998 Jun;46(6):745-50.

103. Humphries KH, Rankin JM, Carere RG, Buller CE, Kiely FM, Spinelli JJ. Co-morbidity data in outcomes research: are clinical data derived from administrative databases a reliable alternative to chart review? J Clin Epidemiol. 2000 Apr;53(4):343-9.

104. Jollis JG, Ancukiewicz M, DeLong ER, Pryor DB, Muhlbaier LH, Mark DB. Discordance of databases designed for claims payment versus clinical information systems. Implications for outcomes research. Ann Intern Med. 1993 Oct 15;119(8):844-50.

105. Kjaergaard J, Clemmensen IH, Thomsen BL, Storm HH. Validity of diagnoses of and operations for nonmalignant gynecological conditions in the Danish National Hospital Registry. J Clin Epidemiol. 2002 Feb;55(2):137-42.

106. Korst LM, Gregory KD, Gornbein JA. Elective primary caesarean delivery: accuracy of administrative data. Paediatr Perinat Epidemiol. 2004 Mar;18(2):112-9.

107. Li J, Morlet N, Semmens J, Gavin A, Ng J. Coding accuracy for endophthalmitis diagnosis and cataract procedures in Western Australia. The Endophthalmitis Population Study of Western Australia (EPSWA): second report. Ophthalmic Epidemiol. 2003 Apr;10(2):133-45.

108. Losina E, Barrett J, Baron JA, Katz JN. Accuracy of Medicare claims data for rheumatologic diagnoses in total hip replacement recipients. J Clin Epidemiol. 2003 Jun;56(6):515-9.

109. Lydon-Rochelle MT, Holt VL, Nelson JC, Cardenas V, Gardella C, Easterling TR, et al. Accuracy of reporting maternal in-hospital diagnoses and intrapartum procedures in Washington State linked birth records. Paediatr Perinat Epidemiol. 2005 Nov;19(6):460-71.

110. Malenka DJ, McLerran D, Roos N, Fisher ES, Wennberg JE. Using administrative data to describe casemix: a comparison with the medical record. J Clin Epidemiol. 1994 Sep;47(9):1027-32.

111. Romano PS, Chan BK, Schembri ME, Rainwater JA. Can administrative data be used to compare postoperative complication rates across hospitals? Med Care. 2002 Oct;40(10):856-67.

112. Romano PS, Schembri ME, Rainwater JA. Can administrative data be used to ascertain clinically significant postoperative complications? Am J Med Qual. 2002 Jul-Aug;17(4):145-54.

113. Romano PS, Yasmeen S, Schembri ME, Keyzer JM, Gilbert WM. Coding of perineal lacerations and other complications of obstetric care in hospital discharge data. Obstet Gynecol. 2005 Oct;106(4):717-25.

114. Roos LL, Sharp SM, Cohen MM. Comparing clinical information with claims data: some similarities and differences. J Clin Epidemiol. 1991;44(9):881-8.

115. Taylor LK, Travis S, Pym M, Olive E, Henderson-Smart DJ. How useful are hospital morbidity data for monitoring conditions occurring in the perinatal period? Aust N Z J Obstet Gynaecol. 2005 Feb;45(1):36-41.

116. Weintraub WS, Deaton C, Shaw L, Mahoney E, Morris DC, Saunders C, et al. Can cardiovascular clinical characteristics be identified and outcome models be developed from an in-patient claims database? Am J Cardiol. 1999 Jul 15;84(2):166-9.

117. Mikolajczyk K, Szabatin M, Rudnicki P, Grodzki M, Burger C. A JAVA environment for medical image data analysis: initial application for brain PET quantitation. Med Inform (Lond). 1998 Jul-Sep;23(3):207-14.

118. Albarran JW, Clarke BA, Crawford J. 'It was not chest pain really, I can't explain it!' An exploratory study on the nature of symptoms experienced by women during their myocardial infarction. J Clin Nurs. 2007 Jul;16(7):1292-301.

119. Correa-de-Araujo R, Stevens B, Moy E, Nilasena D, Chesley F, McDermott K. Gender differences across racial and ethnic groups in the quality of care for acute myocardial infarction and heart failure associated with comorbidities. Womens Health Issues. 2006 Mar-Apr;16(2):44-55.

120. Macaskill P, Gatsonis C, Deeks JJ, Harbord RM, Takwoingi Y. Analysing and Presenting Results. In: Deeks JJ, Bossuyt PM, Gatsonis C, editors. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy: The Cochrane Collaboration; 2010.

121. DerSimonian R, Kacker R. Random-effects model for meta-analysis of clinical trials: an update. Contemp Clin Trials. 2007 Feb;28(2):105-14.

122. Glas AS, Lijmer JG, Prins MH, Bonsel GJ, Bossuyt PM. The diagnostic odds ratio: a single indicator of test performance. J Clin Epidemiol. 2003 Nov;56(11):1129-35.

123. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol. 2005 Oct;58(10):982-90.

124. Reitsma JB, Rutjes AWS, Whiting P, Vlassov VV, Leeflang MMG, Deeks JJ. Assessing methodological quality. In: Deeks JJ, Bossuyt PM, Gatsonis C, editors. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy. 1.0.0 ed: The Cochrane Collaboration; 2009.

125. Leibson CL, Naessens JM, Brown RD, Whisnant JP. Accuracy of hospital discharge abstracts for identifying stroke. Stroke. 1994 Dec;25(12):2348-55.

126. Quan H, Li B, Saunders LD, Parsons GA, Nilsson CI, Alibhai A, et al. Assessing validity of ICD-9-CM and ICD-10 administrative data in recording clinical conditions in a unique dually coded database. Health Services Research. 2008 Aug;43(4):1424-41.

127. Vet de HCW, Eisinga A, Riphagen II, Aertgeerts B, Pewsner D. Searching for Studies. In: Deeks JJ, Bossuyt PM, Gatsonis C, editors. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy. 0.4 ed: The Cochrane Collaboration; 2008.

APPENDIX A: QUERY

Query in PubMed, executed on 21 march 2006:

(Forms and Records Control[MeSH] OR Patient Discharge [MeSH] OR Data Collection[MeSH] OR Registries[MeSH] OR Medical Records[MeSH] OR Abstracting and Indexing[MeSH] OR Classification[MeSH] OR Vocabulary, Controlled[MeSH] OR Sensitivity and Specificity[MeSH] OR Automatic Data Processing[MeSH] OR Information Storage and Retrieval[MeSH] OR Hospital Information Systems[MeSH])

AND

(administrative data OR claims data OR discharge register OR discharge registry OR discharge summary OR diagnostic code OR diagnostic codes OR diagnostic coding OR discharge code OR discharge codes OR discharge coding OR discharge record OR discharge records OR discharge data OR clinical coding OR morbidity coding OR (international classif* disease*[TIAB]) OR ICD*)

AND

(quality OR agree* OR valid* OR invalid* OR accura* OR inaccura* OR complet* OR incomplet* OR correct* OR incorrect* OR reliab* OR unreliab* OR value OR compare OR comparing OR sensitive OR sensitivity OR specific OR specificity OR precision)

AND

(quality[TITLE] OR agree[TITLE] OR disagree[TITLE] OR agreement[TITLE] OR disagreement[TITLE] OR valid*[TITLE] OR invalid*[TITLE] OR reliab*[TITLE] OR unreliab*[TITLE] OR accura*[TITLE] OR inaccura*[TITLE] OR complet*[TITLE] OR incomplet*[TITLE] OR relevan*[TITLE] OR irrelevan*[TITLE] OR consisten*[TITLE] OR inconsisten*[TITLE] OR precis*[TITLE] OR imprecis*[TITLE] OR correct*[TITLE] OR incorrect*[TITLE] OR exact*[TITLE] OR regist*[TITLE] OR computerized[TITLE] OR "patient chart"[TITLE] OR code[TITLE] OR codes[TITLE] OR coding[TITLE] OR coder[TITLE] OR miscoding[TITLE] OR database*[TITLE] OR morbidity[TITLE] OR discharge[TITLE] OR diagnostic[TITLE] OR ICD[TITLE] OR data[TITLE] OR identify[TITLE] OR identifying[TITLE] OR classif*[TITLE] OR claims[TITLE] OR administrative[TITLE] OR registr*[TITLE] OR "predictive value"[TITLE] OR summary[TITLE] OR summaries[TITLE] OR sensitive[TITLE] OR sensitivity[TITLE] OR specific[TITLE] OR specificity[TITLE])

Limits: only items with abstracts, English

APPENDIX B : DATA EXTRACTION FORM SYSTEMATIC REVIEW DIAGNOSTIC DATA ACCURACY

Article ID:

First Author:

Year:

Type of Data

Type of Data: ☐ Routine discharge data ☐ Hospital claims data

Setting, Local Coding Practice and Data Use

Country:

Hospital*: ☐ University ☐ General ☐ Specialized ☐ Not specified

Number of Hospitals: ☐ Single center ☐ Multi center

Period of evaluation: - - till - - (DD-MM-YYYY)

ICD version evaluated*: ☐ ICD-7 ☐ ICD-8 ☐ ICD-9 ☐ ICD-9-CM ☐ ICD-10

Coder*: ☐ trained med record coder ☐ Doctor ☐ Nurse ☐ Clerk ☐ Not specified

Sources for Coding*: ☐ Paper medical record ☐ Electronic medical record
☐ Clinical information systems ☐ Discharge form
☐ Discharge letter ☐ Not specified

Routine Data Use*: ☐ Reimbursement ☐ Quality of care
☐ Research ☐ Management
☐ Patient care ☐ Other, namely:
☐ Not specified

Evaluation Method

Authors' perspective*: ☐ Quality of care ☐ Research
☐ Surveillance ☐ Statistics
☐ Other: ☐ Not specified

Design*: ☐ Retrospective / ☐ Prospective†
☐ Intervention study

Case Selection: ☐ All
☐ Sample
☐ Random
☐ Non random

Gold Standard*: ☐ Medical record / ☐ Discharge letter/summary†
☐ Re-coded / ☐ Re-abstracted (if yes: ☐ By standardized form)
☐ Blinded for routine coding
☐ Clinical criteria
☐ By two or more independent experts
☐ Test result
☐ Disease specific database
☐ Research dataset
☐ Other, namely:

Main criteria for Correctness: ☐ 3-Digits ☐ 4-Digits ☐ 5-Digits
☐ Group of codes ☐ Other:
☐ Not specified

Determinants of data quality: ☐ Analyzed; if yes, by (statistical technique)*:
☐ Pearson's Chi-square test
☐ Multivariate regression analysis
☐ Other, namely:

Data Accuracy

Diagnostic category: ☐ Primary / principle ☐ Secondary / comorbidities
☐ Complications ☐ Reason for admission
☐ Combination ☐ Not specified

Disease category: ☐ "All" diseases / No specific diseases
☐ Myocardial Infarction (MI)
☐ Stroke
☐ Comorbidity / complication associated with specific disease:

☐ Disease associated with specific procedure:
☐ Other, namely:

Overall Results

Pooled 2 x 2 Table with absolute numbers for TP, FP, FN and TN:

Pooled Sens:

Pooled Spec:

Pooled PPV:

Pooled NPV:

Main Conclusions

About data quality: ☐ Sufficient for purpose / Authors' perspective

Note:

About determinants of data quality

Determinants are:

a)

b)

c)

d)

e)

Other conclusion(s):

Note:

CHAPTER 7

DISCUSSION

7.1 INTRODUCTION

Analysis of delivered medical care can be helpful in the pursuit of high quality of care. The electronic capturing of patient data offers the possibility of using the computer to acquire relatively easy insight into the quality of the delivered care for a group of patients with similar clinical characteristics, for example by reporting on the basis of performance indicators. Diagnostic data are critical for analyzing and assessing medical care: many treatment options are diagnosis specific and diagnostic data involve co morbidities and complications. Thus, diagnostic information is important for the interpretation of the delivered care and its outcomes.

This thesis contributes to the discussion of the question whether electronically captured patient data are sufficiently reliable to provide insight into the delivered care. In this thesis the attention was especially focused on electronic diagnostic data as part of the hospital discharge data that are collected in many countries for almost all hospitalizations.

In this chapter we discuss the findings of the studies presented in the chapters two till six and answer the main research questions of this thesis stated in chapter one. Further we place the findings and conclusions in the context of present developments within the field of medical informatics.

7.2 FINDINGS FROM THE STUDIES

7.2.1 Availability and usability of patient data

In chapter 2 we analyzed availability and usability (defined as availability of complete and accurate data in a standardized form) of patient data in the hospital information system of the Academic Medical Centre (AMC), Amsterdam, for the assessment of medical practice concerning children with suspected meningitis for use by the pediatricians themselves. We were interested in the following questions:

1. Which performance indicators, case-mix and exploratory information are needed by physicians for medical practice assessment?
2. Are the required data electronically available and usable for medical practice assessment?

In chapter 2 it is explained which performance indicators, case-mix and exploratory information were selected, based on consensus among pediatricians. Pediatricians of the AMC define quality in terms of appropriateness and timeliness of interventions, and in terms of patient outcomes [1]. In general, this means that patient conditions and interventions and the times that these conditions were observed and the interventions were carried out need to be recorded. The pediatricians needed detailed information to assess their own medical practice. They defined 14 performance indicators plus case-mix, and exploratory information. Of the 39 data items needed for patient selection and indicator quantification 29 were electronically available and 19 usable without manual handling. Reason for admission and diagnoses were incompletely and incorrectly recorded. This seriously hampered patient selection and detection of complications. Time-points of clinical events and interventions were either not available or incorrect. Outpatient diagnosis, signs and symptoms, indications for tests and data about medication administration were missing. Many test result reports were not adequately standardized. In total the value of 9 performance indicators could be determined, but only if it were possible to select patients reliably. For case-mix and exploratory information, 25 additional data items were needed, of which 16 were available and 13 usable. Data about severity of illness, medication prescription, reasons for deviation from the protocol, no show and care provided elsewhere were particularly likely to be missing.

Assessing patient groups defined by established diagnoses limits the possibility of assessing the diagnostic process. It leads to a situation in which patients admitted with a suspected disease, but who are eventually found to have another disease, are not taken into account when assessing the diagnostic process. Especially for serious diseases that have to be ruled out in case of suspicion, assessment of the diagnostic process can only be done meaningfully if all patients with the suspected disease are included. Therefore, a reliable registration of reason for admission is necessary. Not many hospitals did register the reason for admission and if they did, the completeness and correctness of the information was questionable.

Although information supply was a problem, the participation of physicians in this quality of care project led to better awareness of important aspects of care and did uncover possibilities for improvement.

The overall conclusion is that at the time of study not enough usable data items were electronically available to determine the performance indicators needed for

assessing the quality of medical practice. Since the assessment was for internal use the performance indicators were more detailed than when used for public reporting of the performance of hospitals, or individual physicians. Some of the data not available or usable at the time of the study are available or even usable today, e.g. outpatient diagnoses and medication prescription. However, even today the overall conclusion will not be very different.

7.2.2 Redesign of the diagnosis registration

In chapter 3 we described a project at the AMC with the objective to improve the accuracy of the diagnosis registration on the one hand and to accelerate discharge letter writing on the other. The chapter describes the redesign process, especially the involvement of the pediatricians in it, the new registration procedure, and the evaluation of the coding performance of pediatricians in the new situation. In the chapter we addressed the questions:

1. How can the diagnostic discharge registration be incorporated into the care process?
2. What is the effect of the physicians' involvement on the quality of diagnosis coding?

We linked diagnostic coding to discharge letter writing [(2)]. After implementation in routine practice, pediatricians provided diagnoses with codes in a special heading of the discharge letter. The medical record coder checked and corrected this diagnosis heading. A list of diagnoses for pediatrics, based on ICD-9-CM, was developed and alphabetically ordered into a booklet and was used by pediatricians when dictating discharge letters. Within the information infrastructure of hospitals, the discharge letter-linked diagnosis registration appeared feasible in routine practice, but the contribution of the pediatricians to coding was limited due to the low priority given to it. If the physician views the diagnosis registration as having only an additive role in the communication with other health care providers after discharge, the correction function of the medical record coder appears to be indispensable.

The quality of the diagnosis registration appeared difficult to manage. Continuous attention from both the physician and medical record coder seemed necessary.

7.2.3 Evaluation of diagnostic data quality

In chapter 4 we tested our hypothesis that integration of the diagnosis registration into the communication process with GPs combined with physician encoding would improve diagnostic data quality. The research question was:

1. Would physician coding and the integration of the diagnosis registration with the communication process, improve completeness, correctness, specificity and timeliness of diagnostic data?

Our hypothesis that increased influence of the physician would be beneficial and would improve the quality of diagnostic data could not be corroborated [(3)]. Completeness of the form-based diagnosis registration was 51% (95% CI, 44-58%) and of the discharge letter-linked diagnosis registration 54% (95% CI, 47-60%). Correctness was 65% (95% CI, 58-72%) and 67% (95% CI, 60-74%) respectively. That the readily-accessible diagnostic data have a communication and reminder function, and possibly a function in quality assessment, management and research, was apparently not a sufficient incentive for the physicians to improve data quality. Maybe it is the case that diagnostic data recorded after discharge of patients are not suitable for the assessment of the medical practice of patient groups. Probably the only way to improve the quality of diagnostic data is to incorporate the registration in the daily care process. We suggested that the appearance of computerized patient records would provide the opportunity to improve the quality of the diagnosis registration. However, early studies indicated that computerized patient records are a valuable but insufficient addition for obtaining high quality data. We concluded that if we wish to evaluate daily care with routinely collected patient data we probably have to accept that these studies do not meet the high standards required for scientific clinical research.

In the hospital discharge registry of the AMC, per admission one *reason for admission* had to be recorded. During the re-abstracting process it appeared that often more than one *reason for admission* is relevant, e.g. a complex of signs and symptoms or a list of differential diagnoses. We therefore advocated the possibility to record more than one reason for admission.

7.2.4 Long-term effects of the diagnostic coding redesign

In chapter 4 we concluded that we could not corroborate our hypothesis that increased influence of the physician would be beneficial and would improve the

quality of diagnostic data. In chapter 5 we looked in a more detailed way to the data and investigated long-term effects of the introduction of the discharge letter-linked coding procedure. Chapter 5 describes a time series study covering twelve consecutive years. In the first four years, the usual form-based encoding by the medical record coder was in use and in the last eight years, it was the discharge letter-linked encoding by pediatricians. In this study we investigated the influence of physician involvement in diagnosis encoding in the long run. Research questions were:

1. Are diagnoses encoded more specifically?
2. Does the number of coded diagnoses increase?
3. Are any effects sustainable over time?

There was a need for detailed encoding of diagnoses. Extension of the classification of diseases according to the wishes of pediatricians and easy access to this classification truly led to a more detailed encoding of the principal diagnosis, especially of diagnoses that belong to the pediatricians' own subspecialties [4]. In combination with the discharge letter-linked coding procedure, this led to a long-lasting use of detailed codes. A substantial effect on the number of secondary diagnoses per admission could only be shown after a longer period of feedback and attention, but this effect gradually diminished on the long term when feedback and attention were lacking. The increase of secondary diagnoses did lead to added informational value.

Especially the number of secondary diagnoses that do not fall under one's own subspecialty fell back in the long term. Availability of a detailed and well-structured discharge letter led to a better informed medical record coder. However, in the long run the medical record coder was not able to compensate for the diminishing contribution of physicians to the encoding process. The number of recorded secondary diagnoses fell back again to the initial level and with that the added value for medical practice assessment disappears. The level of detail of principal diagnoses remained stable because of the advantage pediatricians have of specific diagnostic codes falling under their own subspecialty.

We concluded that the combined effort of pediatricians and medical record coder led to more coded diagnostic information than the effort of the medical record coder alone.

7.2.5 Systematic Review about diagnostic data quality

Chapter 6 describes a systematic review investigating the quality of diagnostic inpatient hospital discharge data as reported in scientific journals in order to examine whether the results of our study correspond to those published in the literature. We investigated:

1. Which gold standards and designs were used to assess data quality?
2. What completeness and correctness values were reported?
3. Which factors influence the data quality of studies?
4. What are determinants of data quality reported in studies?
5. What is the evidence about the consequences of data quality for medical practice assessment?
6. Are diagnostic data appropriate for quality of care purposes?

The observed data quality of diagnostic hospital discharge data is a function of the diagnostic process, diagnosis documentation, coding practice, characteristics of the disease, manifestation of the disease in patients, prevalence of the diagnoses, and the way data quality is measured [5]. Studies were highly heterogeneous with respect to these factors and showed highly variable data quality. The effect of several forms of bias remained unclear.

We concluded that the quality of diagnostic hospital discharge data leaves much to be desired. On average, completeness was about 67% and correctness about 76%. For some single diseases it was somewhat better, for comorbidities and especially complications it was far worse. In only 13% of the studies completeness and correctness were both at least 85%. Despite all the efforts to improve data quality, it did in general not lead to better data quality in the course of the years. The use of diagnostic hospital discharge data can easily lead to a biased idea of the quality of care. On the one hand cases selected for quality of care assessment possibly represent the more severe cases with a chance of overestimating sentinel outcomes, especially when comorbidities are also underreported. On the other hand, complication data are usually incomplete thus leading to an underestimation of sentinel outcomes. These shortcomings can differ between settings and thus complicates comparisons between hospitals or geographical areas. True variability in quality of care between settings can easily be overshadowed by the unknown

variability in data quality which makes it very difficult to interpret the observed variability in the quality of care based on diagnostic hospital discharge data.

We can conclude that the quality of diagnostic encoding in Amsterdam is similar to that reported in the literature.

7.3 MEANING OF OUR FINDINGS IN THE CONTEXT OF TODAY

The quality of diagnostic hospital discharge data is still an important issue given the fact that many scientific papers on this subject are published to date. They show that the diagnostic data quality is still problematic. Recently for example, Lindenauer et al [6] suggest that the decline in mortality rate of patients hospitalized with pneumonia as determined using data from the 2003-2009 releases of the US Nationwide Inpatient Sample represents a shift in diagnostic coding of pneumonia from principal to secondary diagnoses instead of a true decline in mortality rate. Januel et al [7] found that in two teaching and one non-teaching hospital in Switzerland, the sensitivity of the 17 Charlson co-morbidities was only 36.5% in 1999, 42.5% in 2001 and 42.8% in 2003 which means low completeness. Stein et al [8] found in two urban academic medical centers in the US that the sensitivity of ICD-9-CM codes for COPD exacerbations was very low and varied – dependent on the chosen set of codes – from 12% to 25%, and the positive predictive value varied from 81% to 97%. They concluded that relying on ICD-9-CM codes alone to identify patients hospitalized for COPD may be problematic. Burns et al [9] did a systematic review limited to Great Britain and principal diagnoses (and principal procedures) in which they found a median accuracy for principal diagnoses of 80.3%. They also found that since the 2002 introduction of Payment by Results, principal diagnoses accuracy has improved from 74% to 96%. They did not measure data quality in terms of completeness and correctness, but in terms of percentage agreement between routine coding and gold standard.

Given all the efforts made worldwide in the past to increase diagnostic data quality, one might well wonder whether trying to improve the quality with the current discharge registry system is not a dead end.

Perhaps one should look at other opportunities to improve data quality. Besides for quality of care, quality of electronically captured diagnostic data is also important for other purposes. In direct care diagnostic information has to be recorded for information sharing and transfer to benefit coordination, cooperation and continuity

of care. This is also true for other patient data. Moreover, correct, complete and timely recording of the conditions of the patient is a prerequisite for automatic decision support. For the administration of a healthcare organization it can be important to know the size of patient categories. In case of financial reimbursement systems like DRG, the reimbursement is (partly) based on diagnostic data.

It is a challenge to get physicians actually documenting all relevant patient data. The more favorable the balance between benefit and effort of documentation, the greater the chance of success. Those who are documenting are also those who should benefit. It is plausible that this is best achieved when physicians have to document patient data only once for the benefit of the direct patient care, for obtaining automated decision support and to automatically share information with other health care professionals. The added value for the care process of such things as information exchange and decision support could contribute to a complete, correct, timely and specific documentation. It would be very nice if management, reimbursement, research and quality of care data could automatically be derived from this one-time documentation. Then the likelihood that the completeness and correctness of electronic patient documentation is influenced in a negative way is reduced. After all, it is for the direct care itself of primary importance to sketch an accurate picture of the patient's condition and the provided care which means that there is less room for data optimization for the secondary purposes (such as for the benefit of reimbursement or performance indicator optimization, also called 'upcoding' or 'gaming') because it is soon at the expense of good clinical care. However, this 'enter data once, reuse multiple times paradigm' asks for a high level of standardization during data gathering in the clinical context [10, 11]. Clinical archetypes such as detailed clinical models (DCM) can contribute because they specify care in such a way that the data content is consistent across the entire care chain with its many sites and applications [11, 12]. Each Detailed Clinical Model includes all data attributes and potential terminology bindings that are useful to describe a single, discrete clinical concept for use in a broad range of clinical scenarios [12]. Examples of DCMs include diagnosis, adverse reaction, medication order and symptom.

In the case of the 'enter data once, reuse multiple times paradigm' it is a challenge to select automatically the necessary data for quality of care purposes from the systems that are primarily intended for patient care. The systems for direct patient care often have such a complicated data structure and offer so much room for

variation in patient documentation, that it might be desirable or even necessary to derive one or more new datasets for the secondary purposes. This probably needs human expertise in order to direct, monitor or rectify the automated data selection process.

We learned that a new version of the ICD did not bring a substantial improvement in data quality. Often the ICD is criticized because of its limited possibilities to document patient data as specific as needed for the care process and because of its strict hierarchical structure without representing a rich medical knowledge [13]. The Systematized Nomenclature of Medicine, Clinical Terms (SNOMED CT®) is considered the most comprehensive, multilingual clinical terminology in the world. When implemented in software, SNOMED CT represents clinically relevant information consistently, reliably and comprehensively as an integral part of the electronic health record [13]. However, looking at the reasons for disagreement between gold standards and routine coding, another terminology system alone would probably not solve the problems completely.

Although electronic health records, terminologies like SNOMED-CT and new developments like DCMs can stimulate, they still do not guarantee that comorbidities or complications are documented completely or correctly (since documentation of these items cannot be obligatory since not each patient has them) or that principal diagnoses are documented correctly. The quality of the medical record, be it electronic or paper, is a significant issue in the management of the quality of care. Perhaps 'medical record keeping' should be a performance indicator in itself.

In the current situation where medical practice assessment is dependent on discharge data for the diagnostic information, the medical record coder is an indispensible factor in maintaining data quality. Physician engagement in diagnostic coding also contributes to data quality. Both are necessary, but not sufficient as this thesis shows. Should the necessary diagnostic data be derived from electronic patient documentation, we foresee another role for the medical record coder: the healthcare information specialist. He can play a role in the preparation of new datasets for the secondary purposes. This asks for human expertise in order to direct, monitor or rectify the automated data selection process. This new expert can also evaluate data quality and highlight gaps and inaccuracies in the primary health care professionals' documentation and thus makes a contribution to the quality of the patient documentation. He is trained in patient

care, patient documentation systems, terminology (systems), and in reimbursement, management and quality of care. Creating a secondary dataset still introduces a risk of data-optimization ('gaming'). Ethics would therefore be part of the training of the health information specialist. Of course, the clinicians stay responsible for adequate patient documentation.

Further monitoring of the quality of the current diagnostic discharge data is needed in order to be informed about the data quality. However, clinically significant improvements can only be expected from developments as mentioned above. What data quality can be attained eventually, can only be uncovered empirically. In such a new situation the evaluation of data quality should be based on a review of the electronic medical record: are relevant diagnoses explicitly and correctly recorded as such and are the right data selected for medical practice assessment? An alternative method could be to compare the primary documentation with another prospective, parallel but independent data collection, e.g. a data collection for a scientific study. Another method could be the use of fake patients who visit wards and for which the associated data recording is checked afterwards. No matter what method is used for evaluating data quality in the future, we advocate periodic measurement over a longer period of both completeness (or sensitivity) and correctness (or positive predictive value) of the diagnostic data necessary for medical practice assessment. Blinding and recoding by experts should be the standard for gold standard construction.

7.4 CONCLUSIONS

What this study has clearly shown is that electronic patient documentation is the work of men. In particular the registration of diagnoses which are interpretations of clinicians about what is going on with the patient depends on the completeness, accuracy and clarity with which the care provider captures this. Despite the promise of advanced terminology, documentation models, patient records and care process support systems, this thesis shows that data quality and usability of the data for several purposes will never be self-evident. The cooperation of the clinicians is necessary. Data quality will need never lasting attention, to begin with the training of junior healthcare professionals, but also by the monitoring of data quality in health care settings and the support by healthcare information specialists.

On the basis of our own research and a systematic review of studies published in scientific journals, we conclude that the quality of these diagnostic data in terms of

completeness and correctness leaves much to be desired. The systematic review showed also that this data quality can differ significantly per setting. This makes the use of these electronically captured data in order to determine the quality of patient care debatable. Measured differences in quality of care between settings can be explained by true differences in delivered care as well as by differences in data quality. The problem is that the quality of the data in each setting is almost always unknown [14]. In the interpretation of the measured differences it is difficult to take the data quality into account, making the comparison of settings an impossible task.

Despite the moderate data quality, we think that the use of these data by physicians to assess their own medical practice is one of the least worse applications. Of all stakeholders physicians have the best insight into the quality of their diagnostic data and its implications for the interpretation of performance indicator reports. Physicians are able to compare the information with their own experiences and can reason about what the information means for the quality of their care. However, this will only work when they are also willing to critically reflect on their own medical practice.

REFERENCES

1. Prins H, Kruisinga FH, Buller HA, Zwetsloot-Schonk JH. Availability and usability of data for medical practice assessment. Int J Qual Health Care. 2002 Apr;14(2):127-37.
2. Prins H, Buller HA, Zwetsloot-Schonk JH. Redesign of Diagnostic Coding in Pediatrics: From Form-based to Dicharge Letter Linked. Perspectives in Health Information Management. 2004;1(1):1-10.
3. Prins H, Buller HA, Zwetsloot-Schonk JH. Effect of discharge letter-linked diagnosis registration on data quality. Int J Qual Health Care. 2000 Feb;12(1):47-57.
4. Prins H, Hasman A. Long-term impact of physician encoding on detail and number of recorded diagnoses. Support by a speciality specific list of diseases. Methods Inf Med. 2011;50(2):115-23.
5. Prins H, Hasman A. Appropriateness of ICD-Coded Diagnostic Inpatient Hospital Discharge Data for Medical Practice Assessment: a Systematic Review. submitted. 2012.
6. Lindenauer PK, Lagu T, Shieh MS, Pekow PS, Rothberg MB. Association of diagnostic coding with trends in hospitalizations and mortality of patients with pneumonia, 2003-2009. JAMA. 2012 Apr 4;307(13):1405-13.
7. Januel JM, Luthi JC, Quan H, Borst F, Taffe P, Ghali WA, et al. Improved accuracy of co-morbidity coding over time after the introduction of ICD-10 administrative data. BMC Health Serv Res. 2011;11:194.

8. Stein BD, Bautista A, Schumock GT, Lee TA, Charbeneau JT, Lauderdale DS, et al. The validity of International Classification of Diseases, Ninth Revision, Clinical Modification diagnosis codes for identifying patients hospitalized for COPD exacerbations. Chest. 2012 Jan;141(1):87-93.

9. Burns EM, Rigby E, Mamidanna R, Bottle A, Aylin P, Ziprin P, et al. Systematic review of discharge coding accuracy. J Public Health (Oxf). 2012 Mar;34(1):138-48.

10. Goossen W. Representing knowledge, data and concepts for EHRS using DCM. Stud Health Technol Inform. 2011;169:774-8.

11. Goossen W, Goossen-Baremans A, van der Zel M. Detailed clinical models: a review. Healthc Inform Res. 2010 Dec;16(4):201-14.

12. Coyle JF, Mori AR, Huff SM. Standards for detailed clinical models as the basis for medical data exchange and decision support. Int J Med Inform. 2003 Mar;69(2-3):157-74.

13. Cimino JJ. High-quality, standard, controlled healthcare terminologies come of age. Methods Inf Med. 2011;50(2):101-4.

14. van Walraven C, Bennett C, Forster AJ. Administrative database research infrequently used validated diagnostic or procedural codes. J Clin Epidemiol. 2011 Oct;64(10):1054-9.

A physician, as a professional, is interested in the quality of his work and in ways to improve it further. Systematic, retrospective assessment of daily medical practice offers physicians possibilities to improve their quality of care and to account for their medical practice. In this thesis the object of study is the usability of routinely collected and electronically recorded patient data for the assessment by medical specialists themselves of their medical practice for specific, clinically defined patient groups. This study was carried out by the Department of Medical Informatics and the Department of Pediatrics at the Academic Medical Center, Amsterdam.

Chapter 1 introduces the way in which physicians can assess their medical practice and presents the aims of this thesis. To assess medical practice, high-quality data about the patients, their disease-related attributes, activities performed or initiated by physicians and patient events in the course of time are necessary. When the data are recorded in a computerized system desired analyses can be carried out efficiently. However, not all data of the care process are recorded electronically. Also some electronically available data may not be recorded completely, correctly, with enough detail, timely or standardized. The quality of diagnostic hospital discharge data stands in bad repute among physicians and health care researchers. These data play no role in daily patient care and this is possibly the main reason why there are doubts about the reliability of the registration. However, the hospital discharge data registry is the only registry of diagnostic data so far that covers all hospitalizations. This complete coverage is a major advantage for the use of the data for the assessment of medical practice for specific patient groups.

Diagnostic data play an important role in medical practice assessment because process and outcome indicators are often disease specific. Data about complications, a special type of diagnosis, can be used to get insight in important

outcome indicators. For the interpretation of process and outcome indicators, insight in comorbidities, also a special form of diagnosis, can be necessary.

Since the pediatricians had serious doubts about the reliability of the diagnosis data, we were especially interested in the quality of diagnosis data and sought ways to increase the reliability of the data. The aim of the study was fivefold:

1. To get insight in the information needs of the physicians for the assessment of their medical practice of specific patient groups;
2. To test whether patient data needed for medical practice assessment are electronically available and usable.
3. To find a way to incorporate a diagnosis registry into the clinical care process in order to get a better diagnosis registry;
4. To test whether incorporating the diagnosis registry into the clinical care process improves diagnostic data quality;
5. To get insight, based on a systematic review of the literature, in diagnostic data quality elsewhere and in the factors that influence this data quality. In this way we could see whether the results of our study correspond to those published in the literature.

In **chapter 2** we investigated which performance indicators pediatricians want to use for the assessment of their medical practice of children with suspected or proven meningitis. A performance indicator is a systematically developed quantitative measure that can be used to assess and improve health care activities and outcomes for which norms are defined. We analyzed the electronically availability of those patient data needed to automatically determine the values of the performance indicators and evaluated the usability -defined as availability of complete and accurate data in a standardized form- of these data.

A total of 14 performance indicators were defined. Part of the necessary data needed for determining the values of the indicators could not be delivered by the clinical information systems. Of 39 data items required for indicator quantification, 29 were available of which 19 were usable without manual handling. The main problems are among others incomplete and incorrect registration of diagnoses, time-points that are not or incorrectly recorded, medication administration and indications for test ordering that are not electronically recorded, and some test results that are not standardized. For assessment of the diagnostic process, a

reliable registration of reason for admission is necessary. Not many hospitals do register reason for admission. Availability and usability of electronic patient data are insufficient for physician-led, detailed assessment of their medical practice for specific patient groups. The use of paper medical records is necessary for additional data and verification.

In **chapter 3** we describe a project with the goal to improve the accuracy of the diagnosis registration. Formerly, pediatricians completed discharge forms. However, many forms were completed with insufficient information or not at all. In the new situation pediatricians provide diagnoses with codes in a special heading of the discharge letter. The medical record coder checks and corrects this diagnosis heading. A list of diagnoses for pediatrics, based on ICD-9-CM, was developed and alphabetically ordered into a booklet used by pediatricians when dictating discharge letters. Immediately after implementation, 25% of the diagnoses were initially (before adjustments were made by the medical record coder) not coded or incorrectly coded by the pediatricians; nine percent of these shortcomings could be attributed to the pediatricians. Two years later, 67% of the diagnoses were initially not coded or incorrectly coded; 37% of these shortcomings were attributable to pediatricians.

Within the current information infrastructure of hospitals, discharge letter-linked diagnosis registration appears feasible in routine practice. However, if physicians are of the opinion that this form of diagnosis registration plays only a minor role in the communication with other health care providers after discharge, the correction function of the medical record coder remains indispensable.

In **chapter 4** we tested our hypothesis that diagnostic coding by physicians in combination with the integration of the diagnosis registration and discharge letter writing, improves completeness and correctness of diagnostic data. We compared the quality of this discharge letter-linked diagnosis registration with the quality of the previous form-based registration. A retrospective study was performed with blinded before-after measurement. Re-abstracted diagnosis descriptions of the text of discharge letters were taken as gold standard. For both registration methods 60 admissions were selected randomly. Completeness and correctness, both at the three-digit level of ICD-9-CM, were determined. Completeness of form-based diagnosis registration was 51% (95% CI, 44-58%) and of discharge letter-linked diagnosis registration 54% (95% CI, 47-60%). Correctness was 65% (95% CI, 58-72%) and 67% (95% CI, 60-74%) respectively. Our hypothesis that linking

diagnosis registration to the discharge letter would improve diagnostic data quality could not be demonstrated.

Chapter 5 presents a time series study covering twelve consecutive years. In the first four years, the usual form-based encoding by the medical record coder was in use and in the last eight years, the discharge letter-linked encoding by pediatricians. In this study we evaluated the effect in the long run of the new policy on the level of detail and the number of recorded diagnoses. Immediately after introduction, half of the diagnoses for which both generic and specific codes existed, was coded specific. In later years this proportion remained stable at 0.35 ($p < 0.05$). Diagnoses that fall under the pediatrician's own subspecialty had more often a specific code than diagnoses that did not. The mean number of secondary diagnoses per admission increased from 0.7 before introduction to 1.4 in the third year after introduction ($p < 0.05$) but gradually fell back to 0.7. This increase and decrease was mainly due to diagnoses that did not fall under the pediatrician's own subspecialty. The extra codes in individual discharge summaries had added informational value. The medical record coder is not able to compensate for the diminishing contribution of physicians to the encoding process. With the decrease of the number of recorded secondary diagnoses, the added value of the earlier increase for medical practice assessment disappears. The level of detail of principal diagnoses remains stable because of the advantage for pediatricians of having specific diagnostic codes falling under their own subspecialty. A combined effort of pediatricians and medical record coder leads to more coded diagnostic information than the effort of a medical record coder alone.

Chapter 6 describes a systematic review investigating the quality of diagnostic inpatient hospital discharge data as reported in the scientific literature in order to gain insight in the usefulness of these data for medical practice assessment. We investigated the methods used to evaluate data quality, factors that determine data quality, data quality itself and its consequences for medical practice assessment. We selected studies in which both completeness (sensitivity) and correctness (positive predictive value) were measured. The random-effects model was used to calculate mean completeness and correctness and to explore the effect of a number of covariates. We included 101 studies. We could distinguish six typical study designs. We found a mean completeness of 0.67 (95%CI: 0.62-0.73) and correctness of 0.76 (95%CI: 0.73-0.79). Completeness was significantly lower for comorbidities and complications than for some single diseases. Completeness of

complications and comorbidities barely reaches 50%. Correctness was significantly higher for Scandinavian countries than for other countries. Recoding compared to re-abstracting of the medical record as a gold standard gave a significantly lower correctness. Diagnostic data were considered appropriate by the authors of the studies for quality of care purposes when both completeness and correctness were at least 0.85. In only 13% of the studies this criterion was fulfilled.

The observed data quality of diagnostic hospital discharge data is a function of the diagnostic process, diagnoses documentation, coding practice, characteristics of the disease, prevalence of the disease in the population, manifestation of the disease in patients and the way data quality is measured. Studies were highly heterogeneous with respect to these factors and showed highly variable data quality.

The quality of diagnostic hospital discharge data leaves much to be desired. Despite all the efforts to improve data quality, it did in general not lead to better data quality in the course of the years. The use of diagnostic hospital discharge data can easily lead to a biased idea of the quality of care. True variability in quality of care between settings can easily be overshadowed by the unknown variability in data quality which makes it very difficult to interpret the observed variability in the quality of care.

The principle findings of this thesis are summarized and discussed in **chapter 7**. The quality of diagnostic hospital discharge data is still an important issue given the fact that many scientific papers on this subject are published to date. Given all the efforts made worldwide to increase diagnostic data quality, one might well wonder whether trying to improve the quality of the current discharge registry system is not a dead end.

It is a challenge to get physicians actually documenting all relevant patient data in electronic form. The more favorable the balance between benefit and effort of documentation, the greater the chance of success. Those who are documenting should benefit. This is probably best achieved when physicians have to document patient data only once for the benefit of the direct patient care, for obtaining automated decision support and to automatically share information with other health care professionals. It would be very nice if management, reimbursement, research and quality of care data could automatically be derived from this one-time documentation. Clinical archetypes can contribute because they specify care in such a way that the data content is consistent across the entire care chain with its many sites and applications. However, even then systems for direct patient care

often have such a complicated data structure and offer so much room for variation in patient documentation, that it might be desirable or even necessary to derive one or more new datasets for secondary purposes. Despite the moderate data quality, we think that the use of these data by physicians to assess their own medical practice may be useful. Of all stakeholders, physicians have the best insight into the quality of their diagnostic data and its implications for the interpretation of indicator reports. Physicians are able to compare the information with their own experiences and can reason about what the information means for the quality of their care. However, this will only work when they are also willing to critically reflect on their own medical practice.

Een arts is, als professional, geïnteresseerd in de kwaliteit van zijn werk en in manieren om deze verder te verbeteren. Systematische, retrospectieve beoordeling van de eigen dagelijkse medische praktijk biedt artsen de mogelijkheid om de kwaliteit van zorg te verbeteren en verantwoording af te leggen over het eigen medisch handelen. Het onderwerp van studie in dit proefschrift betreft de bruikbaarheid van routinematig verzamelde en elektronisch vastgelegde patiëntgegevens voor de beoordeling door medisch specialisten zelf van het eigen medisch handelen bij specifieke, klinisch gedefinieerde patiëntengroepen. Deze studie werd uitgevoerd door de afdeling Medische Informatiekunde en de afdeling Kindergeneeskunde van het Academisch Medisch Centrum te Amsterdam.

Hoofdstuk 1 introduceert de manier waarop artsen hun medisch handelen kunnen beoordelen en presenteert de doelstellingen van dit proefschrift. Om medisch handelen te kunnen beoordelen is het nodig om over hoogwaardige gegevens te beschikken van patiënten, hun ziekte gerelateerde kenmerken, de door de artsen uitgevoerde of geïnitieerde verrichtingen en de medische toestandsveranderingen die zich in de loop van de tijd voordoen bij patiënten. Wanneer de gegevens worden geregistreerd in een computersysteem kunnen gewenste analyses efficiënt worden uitgevoerd. Echter, niet alle gegevens van het zorgproces worden elektronisch vastgelegd. Daarnaast wordt een aantal elektronisch beschikbare gegevens mogelijk niet volledig, juist, met voldoende detail, tijdig of gestandaardiseerd vastgelegd. In het bijzonder heeft de kwaliteit van diagnosegegevens als onderdeel van de ziekenhuis ontslagregistratie geen goede reputatie onder artsen en gezondheidszorgonderzoekers. Deze gegevens spelen geen rol in de dagelijkse patiëntenzorg en dit is mogelijk de belangrijkste reden waarom er twijfels zijn over de betrouwbaarheid van de registratie. Echter, de ziekenhuis ontslagregistratie is tot dusver de enige registratie van diagnosegegevens die alle ziekenhuisopnamen dekt. Deze volledige dekking is een

belangrijk voordeel voor het gebruik van deze gegevens voor het beoordelen van het medisch handelen bij specifieke patiëntengroepen.

Diagnosegegevens spelen een belangrijke rol in de beoordeling van het medisch handelen omdat gewenste activiteiten en verwachte resultaten vaak ziektespecifiek zijn. Gegevens over complicaties, een speciaal type diagnose, kunnen worden gebruikt om inzicht te krijgen in belangrijke uitkomstindicatoren. Voor de interpretatie van proces- en uitkomstindicatoren kan inzicht in nevendiagnosen, ook een bijzonder type diagnose, noodzakelijk zijn.

Omdat de kinderartsen ernstige twijfels hadden over de betrouwbaarheid van de diagnosegegevens, waren we vooral geïnteresseerd in de kwaliteit van de diagnosegegevens en naar manieren om de betrouwbaarheid van deze gegevens te verhogen. Het doel van de studie was vijfvoudig:

1. Inzicht verkrijgen in de informatiebehoefte van de artsen voor de beoordeling van hun medisch handelen bij specifieke patiëntengroepen;
2. Testen of de patiëntgegevens die nodig zijn voor het beoordelen van het medisch handelen elektronisch beschikbaar en bruikbaar zijn;
3. Een manier vinden om de diagnoseregistratie te integreren in het klinisch zorgproces om daarmee een betere diagnoseregistratie te verkrijgen;
4. Testen of het integreren van de diagnoseregistratie in het klinisch zorgproces de kwaliteit van de diagnosegegevens verbetert;
5. Inzicht verkrijgen, op basis van een systematische beoordeling van de literatuur, in de kwaliteit van diagnosegegevens elders en in factoren die deze kwaliteit beïnvloeden. Op deze wijze konden we nagaan of de resultaten van onze studie overeenkwamen met die in de literatuur.

In **hoofdstuk 2** hebben we onderzocht welke prestatie-indicatoren kinderartsen willen gebruiken voor de beoordeling van hun medische handelen bij kinderen met een vermoeden op, of met bewezen hersenvliesontsteking. Een prestatie-indicator is een systematisch ontwikkelde kwantitatieve maat die gebruikt kan worden voor het beoordelen en verbeteren van zorgactiviteiten en –uitkomsten waarvoor een norm is vastgesteld. We analyseerden vervolgens de elektronische beschikbaarheid van de patiëntgegevens die nodig zijn om de waarden van de prestatie-indicatoren vast te kunnen stellen en onderzochten de bruikbaarheid - gedefinieerd als de

beschikbaarheid van volledige en juiste gegevens in een gestandaardiseerde vorm - ervan.

Er werden 14 prestatie-indicatoren gedefinieerd. Een deel van de noodzakelijke data voor het bepalen van de waarden van de indicatoren kon niet worden geleverd door de klinische informatiesystemen. Van de 39 data-items die nodig waren voor indicator kwantificering, waren er 29 beschikbaar, waarvan 19 geschikt waren om te gebruiken zonder handmatige bewerking. De belangrijkste problemen waren onder andere onvolledige en onjuiste registratie van diagnosen, tijdstippen die verkeerd of niet waren geregistreerd, het niet vastleggen van het daadwerkelijk toedienen van medicatie en van de indicaties voor het aanvragen van testen, en een aantal testresultaten dat niet gestandaardiseerd was vastgelegd. Voor de beoordeling van het diagnostisch proces is een betrouwbare registratie van de opnamereden nodig. Niet veel ziekenhuizen registreren de reden van opname. De beschikbaarheid en bruikbaarheid van elektronische patiëntgegevens zijn onvoldoende voor een gedetailleerde beoordeling door artsen van hun medische handelen bij specifieke patiëntengroepen. Het gebruik van de papieren medische dossiers is nodig voor aanvullende gegevens en verificatie.

In **hoofdstuk 3** beschrijven we een project dat als doel heeft de nauwkeurigheid van de diagnoseregistratie te verbeteren. Vroeger vulden kinderartsen hiervoor ontslagformulieren in. Echter, veel formulieren werden onvolledig of helemaal niet ingevuld. In de nieuwe situatie vermelden de kinderartsen de diagnosen met hun codes in een speciale rubriek op de ontslagbrief. De medisch codeur controleert deze rubriek en corrigeert eventueel. Een lijst van diagnosen voor de kindergeneeskunde, gebaseerd op de ICD-9-CM, werd ontwikkeld en alfabetisch gerangschikt in een boekje dat wordt gebruikt door de kinderartsen als zij hun ontslagbrieven dicteren.

Onmiddellijk na de implementatie was 25% van de diagnosen in eerste instantie (voordat correcties werden aangebracht door de medisch codeur) niet gecodeerd of verkeerd gecodeerd door de kinderartsen; negen procent van deze tekortkomingen konden worden toegeschreven aan de kinderartsen. Twee jaar later was 67% van de diagnosen in eerste instantie niet gecodeerd of niet correct gecodeerd; 37% van deze tekortkomingen kon worden toegeschreven aan de kinderartsen.

Binnen de huidige informatie-infrastructuur van ziekenhuizen lijkt de ontslagbriefgekoppelde diagnoseregistratie haalbaar in de dagelijkse praktijk. Echter, als artsen van mening zijn dat deze vorm van diagnoseregistratie slechts

een kleine rol speelt in de communicatie met andere zorgverleners na ontslag, blijft de correctie functie van de medisch codeur onmisbaar.

In **hoofdstuk 4** hebben we de hypothese getest dat diagnosecodering door de specialisten, in combinatie met de integratie van de diagnoseregistratie en het schrijven van de ontslagbrief voor huisartsen, de volledigheid en juistheid van de diagnosegegevens verbetert. We vergeleken de kwaliteit van deze ontslagbriefgekoppelde diagnoseregistratie met de kwaliteit van de vorige formuliergebaseerde registratie. Een retrospectieve studie werd uitgevoerd met blinde voor- en nameting. Opnieuw geïnterpreteerde en geselecteerde diagnosebeschrijvingen uit de tekst van ontslagbrieven werden als gouden standaard gebruikt. Voor beide registratiemethoden werden 60 opnames willekeurig geselecteerd. Volledigheid en juistheid, beide op het driecijferig niveau van de ICD-9-CM, werden bepaald. Volledigheid van de formuliergebaseerde diagnoseregistratie was 51% (95% BI, 44 tot 58%) en van de ontslagbriefgekoppelde diagnoseregistratie 54% (95% BI, 47 tot 60%). Juistheid was respectievelijk 65% (95% BI, 58 tot 72%) en 67% (95% BI, 60-74%). Onze hypothese dat het koppelen van de diagnoseregistratie aan de ontslagbrief de kwaliteit van de diagnosegegevens zou verbeteren kon niet worden aangetoond.

Hoofdstuk 5 presenteert een tijdreeks studie voor twaalf opeenvolgende jaren. In de eerste vier jaar was de gebruikelijke formuliergebaseerde codering door de medisch codeur in gebruik en in de laatste acht jaar de ontslagbriefgekoppelde codering door kinderartsen. In deze studie evalueerden we het effect op lange termijn van het nieuwe beleid op het aantal vastgelegde diagnosen en het detailniveau ervan.

Onmiddellijk na de introductie werd de helft van de diagnosen waarvoor zowel generieke en specifieke codes bestonden, specifiek gecodeerd. In latere jaren bleef de proportie stabiel op 0,35 ($p < 0,05$). Diagnosen die onder de eigen subspecialisatie van de kinderarts vielen, hadden vaker een specifieke code dan diagnosen die daar niet onder vielen. Het gemiddelde aantal nevendiagnosen per opname nam toe van 0,7 vóór de invoering tot 1,4 in het derde jaar na introductie ($p < 0,05$), maar viel geleidelijk terug tot 0,7. Deze stijging en daling was voornamelijk het gevolg van diagnosen die niet onder de eigen subspecialisatie van de kinderarts vielen. De extra codes per opname leverden nieuwe diagnostische informatie op.

Medische codeurs kunnen de afnemende bijdrage van artsen aan het coderingsproces niet compenseren. Met de daling van het aantal vastgelegde nevendiagnosen, verdwijnt ook de toegevoegde waarde van de eerdere toename voor het beoordelen van het medisch handelen. De mate van detail van de hoofddiagnosen blijft stabiel vanwege het voordeel dat kinderartsen hebben van specifieke diagnostische codes die onder hun eigen subspecialisatie vallen. Een gecombineerde inspanning van kinderartsen en medisch codeur leidt tot meer gecodeerde diagnostische informatie dan de inspanning van een medisch codeur alleen.

Hoofdstuk 6 beschrijft een systematische beoordeling van de wetenschappelijke literatuur met betrekking tot de kwaliteit van de diagnosegegevens in ziekenhuis ontslagregistraties om inzicht te krijgen in de bruikbaarheid van deze gegevens voor het beoordelen van het medisch handelen. We onderzochten de methoden die worden gebruikt om de kwaliteit van gegevens te evalueren, factoren die de kwaliteit van de gegevens beïnvloeden, de gegevenskwaliteit zelf en de gevolgen ervan voor het beoordelen van het medisch handelen. We selecteerden studies waarin zowel volledigheid (sensitiviteit) als juistheid (positieve voorspellende waarde) waren gemeten. Het random-effects model werd gebruikt om de gemiddelde volledigheid en juistheid te berekenen en het effect van een aantal factoren te verkennen. We includeerden 101 studies. We konden zes typische studieopzetten onderscheiden. We vonden een gemiddelde volledigheid van 0,67 (95% BI: 0,62 - 0,73) en juistheid van 0,76 (95% BI: 0,73 - 0,79). Volledigheid was significant lager voor comorbiditeiten en complicaties dan voor enkele individuele ziekten. Volledigheid van complicaties en comorbiditeiten bereikte nauwelijks de 50%. Juistheid was significant hoger voor de Scandinavische landen dan voor andere landen. Hercoderen in vergelijking met herinterpreteren van het medisch dossier als gouden standaard gaf een significant lagere juistheid.

Diagnosegegevens werden door de auteurs van de studies geschikt geacht voor het beoordelen van het medisch handelen wanneer zowel volledigheid als juistheid tenminste 0,85 waren. In slechts 13% van de studies werd aan dit criterium voldaan.

De waargenomen kwaliteit van de diagnose gegevens is een functie van het diagnostisch proces, de diagnose documentatie, de coderingspraktijk, kenmerken van de ziekte, de prevalentie van de ziekte onder de populatie, de manifestatie van de ziekte bij de patiënten en de wijze waarop gegevenskwaliteit gemeten wordt.

Studies waren zeer heterogeen met betrekking tot deze factoren en toonden een zeer variabele gegevenskwaliteit.

De kwaliteit van de diagnosegegevens in de ziekenhuis ontslagregistraties laat te wensen over. Ondanks alle inspanningen om gegevens te verbeteren leidde dit in de loop der jaren niet tot een betere kwaliteit van gegevens. Het gebruik van de diagnosegegevens uit de ziekenhuis ontslagregistratie kan gemakkelijk leiden tot een vertekend beeld van de kwaliteit van zorg. Echte verschillen in kwaliteit van zorg tussen instellingen kunnen gemakkelijk worden overschaduwd door de onbekende verschillen in de kwaliteit van gegevens waardoor het lastig is om waargenomen verschillen in kwaliteit van zorg te interpreteren.

De belangrijkste bevindingen van dit proefschrift worden samengevat en besproken in **hoofdstuk 7**. De kwaliteit van diagnosegegevens in de ziekenhuis ontslagregistraties is nog steeds een belangrijk onderwerp gezien de vele wetenschappelijke publicaties die tot op de dag van vandaag over dit onderwerp verschijnen. Gezien alle inspanningen die wereldwijd zijn gedaan om de kwaliteit van diagnosegegevens te verhogen, kan men zich afvragen of pogingen om de kwaliteit van de gegevens binnen het huidige registratiesysteem te verbeteren niet een doodlopende weg vormen.

Het is een uitdaging om artsen daadwerkelijk alle relevante patiëntgegevens in elektronische vorm te laten documenteren. Hoe gunstiger de verhouding tussen voordeel en inspanning van de documentatie, hoe groter de kans op succes. Degenen die documenteren moeten ook profiteren. Dit wordt waarschijnlijk het beste bereikt wanneer artsen patiëntgegevens slechts één keer hoeven vast te leggen ten behoeve van de directe patiëntenzorg, voor het verkrijgen van geautomatiseerde beslissingsondersteuning en het automatisch delen van informatie met andere hulpverleners. Het zou heel mooi zijn als informatie ten behoeve van management, financiering, onderzoek en kwaliteit van zorg automatisch kunnen worden afgeleid uit de eenmalige documentatie. Klinische archetypen kunnen hieraan bijdragen omdat zij zorg zo specificeren dat de gegevensinhoud consistent is over de gehele zorgketen met de vele locaties en toepassingen. Echter, ook dan hebben systemen voor directe patiëntenzorg vaak zo'n ingewikkelde datastructuur en bieden zo veel ruimte voor variatie in patiëntdocumentatie, dat het wenselijk of zelfs noodzakelijk is één of meer nieuwe datasets te creëren voor secundaire doeleinden. Ondanks de matige kwaliteit van diagnosegegevens, denken we dat het gebruik van deze gegevens door artsen om

hun eigen medische handelen te beoordelen van nut kan zijn. Van alle betrokkenen, hebben artsen het beste inzicht in de kwaliteit van hun diagnosegegevens en de gevolgen daarvan voor de interpretatie van indicator rapportages. Artsen zijn in staat om de informatie te vergelijken met hun eigen ervaringen en kunnen beredeneren wat de informatie betekent voor de kwaliteit van hun zorg. Dit zal echter alleen werken als ze bereid zijn om ook kritisch te reflecteren op hun eigen medisch handelen.

Blij ben ik dat het gelukt is mijn proefschrift te mogen verdedigen en dat ook nog eens binnen vier jaar; in drie jaar en 5845 dagen om precies te zijn. Dat is niet voor iedere promovendus weggelegd. Daarom wil ik iedereen die heeft bijgedragen aan deze bijzondere prestatie[1] hartelijk danken.

Het promotietraject was voor mij de tocht der tochten en velen hebben een aandeel gehad in het bereiken van de finish. Zo moest de tocht eerst uitgeschreven en georganiseerd worden. Ik mocht starten zonder specifieke voorbereiding. In het begin van de tocht schaatsten begeleiders mee die mij in het donker de weg wezen, maar die op een bepaald moment afsloegen om een andere route te nemen. Het was inmiddels licht geworden en op eigen houtje reed ik verder. Ik week vaak af van de route vanwege het mooie ijs elders, maar altijd waren er lieden die me weer in de juiste richting duwden. Meerdere schaatsers reden een stukje mee en deden kopwerk. Ervaren schaatsers gaven advies over aanpak en schaatstechniek. Onderweg werd de hoofdtrainer tweemaal gewisseld. Onder elke trainer boekte ik progressie. Desondanks werd ik voorbijgestreefd door schaatsers die veel later waren gestart. ‘Jij haalt de finish ook wel’, riepen ze me na. Ook achter de schermen waren mensen mij goedgezind. Zo vernam ik dat een schaatsster die het felbegeerde kruisje al in bezit had, haar trainer had getipt dat ik ergens halverwege voort ploeterde en een nieuwe trainer nodig had. Er waren mensen die ervoor zorgden dat de route beter begaanbaar werd door het ijs te vegen en obstakels te verwijderen. Ik kreeg adviezen over tijdschema’s, hoewel ik die vaak in de wind sloeg. Windwakken heb ik niet altijd kunnen ontwijken, maar telkens werd ik van droge kleren voorzien. Als ik een tijdje langs de kant zat om bij te komen, kwam er

[1] Als kind droomde ik ervan ooit eens een bijzondere prestatie neer te zetten. Op mijn tiende ging ik bij wielrennen, maar finishte steevast als laatste. Hardlopen ging mij beter af, maar een groot talent was ik niet. Schaatsen op natuurijs was mijn grote passie. Op 26 februari 1986 reed ik de Elfstedentocht onder een schuilnaam. Die tocht leerde mij dat het halen van een doel vooral een kwestie is van niet opgeven en niemand tegenkomen die je tegenhoudt vanwege het overschrijden van een tijdslimiet.

altijd wel weer iemand die me op de been hielp en een duwtje in de rug gaf. Ook waren er omstanders die voor de nodige afleiding zorgden. Overal waar ik langs kwam waren mensen belangstellend en enthousiast. Op het eind werd mijn stempelkaart gecontroleerd en geldig bevonden: gelukkig, het was niet voor niets geweest.

Nooit heb ik het idee gehad dat ik de finish niet zou halen, ik vertrouwde erop dat het goed zou komen. Ik putte moed uit de trajecten waar ik de slag weer goed te pakken kreeg. Ik besef echter terdege dat ik het kruisje in ontvangst mag nemen omdat nooit iemand zei: “Stop er maar mee”.

Mijn grote dank gaat uit naar een ieder die zich in de vergelijking herkent. In het bijzonder wil ik mijn promotoren, paranimfen en familie noemen.

Prof. dr. ir. A. Hasman, beste Arie, jij was de trainer die me ergens halverwege het traject in 2006 onder je hoede nam. Als één van de laatsten raakte je betrokken bij mijn promotietraject, maar als één van de eersten bracht je mij beginselen van de medische informatiekunde bij. Dat was begin jaren ’90 toen ik nog in Maastricht studeerde en jij daar hoogleraar was. Je was voor mij de ideale begeleider en promotor: kritisch en geduldig waardoor teksten konden uitgroeien tot volwaardige artikelen. Vele discussies hebben we gevoerd en vele verkenningen zijn gedaan waardoor de inhoud enorm aan kwaliteit heeft gewonnen. Zelf was je altijd heel snel met het reageren op een volgende versie van een manuscript. Bovenal bleef je begeleiden, ook al ging jij met emeritaat en ik met horten en stoten vooruit. Ik ben je daar heel dankbaar voor.

Prof. dr. H.A. Büller, beste Hans, jij was één van die begeleiders van het eerste uur die een andere route nam. Je was destijds heel enthousiast en verzorgde de band met de kindergeneeskunde. Ik stel het daarom zeer op prijs dat je in het zicht van de finish weer betrokken bent, nu als promotor.

Beste Jan, fijn dat je mijn paranimf wilt zijn na eerder al ceremoniemeester te zijn geweest op de trouwdag van Maaike en mij. Sinds de middelbare school zijn we hele goede vrienden. Inmiddels zijn de verjaardagsfeestjes van jou, Hanneke en Jens geliefde uitjes voor ons gezin geworden. We gingen vroeger samen op fietsvakantie en die slopende ritten bergop waren volgens jou goed voor het doorzettingsvermogen. Ik heb daar veel aan gehad. Beste Jos, ik ben blij jouw vurige wens om ooit eens paranimf te zijn in vervulling te kunnen laten gaan. Natuurlijk ben ik ook blij dat je mij als collega terzijde wilt staan. Nu maar hopen

dat ook je wens om een vraag te mogen beantwoorden in vervulling gaat… Ik vrees er nu toch aan te moeten geloven om samen met jou een marathon te lopen.

Lieve pa en ma, heel dankbaar ben ik voor de ruimte die jullie mij van jongs af aan gaven om me te ontplooien in richtingen die ik zelf verkoos. Jullie stonden onvoorwaardelijk klaar om mij en mijn 'broertje' en 'zusje' te helpen en boden altijd een veilige thuishaven. Dat geldt nu ook voor onze gezinnen. Deze 'wetenschap' is belangrijk geweest voor het promotietraject.

Lieve schoonouders, jullie hebben mij gestimuleerd weer te gaan studeren en hadden veel belangstelling voor het onderzoek. Bovenal hebben jullie een fantastische dochter op de wereld gezet die mij in vuur en vlam zette en waarmee ik een gezin stichtte. Veel dank voor jullie gastvrijheid en het ondersteunen van ons drukke gezin zodat ik meer tijd had voor het onderzoek.

Beste Harry & Anke, Emmy & Kasper en Joost & Lilian, veel dank voor jullie belangstelling als broer, zus, zwagers en schoonzussen. De vele, gezellige familiebijeenkomsten gaven telkens welkome afleiding. Gelukkig hadden Laura, Marnix, Jorwen, Bodil, Paiva, Anne, Lotte en Stan vooral belangstelling voor mijn capaciteiten als tikker of voetballer en vroegen zij niet naar de vorderingen van mijn onderzoek.

Lieve Tijmen, Hielke, Jelmer en Emma, wat ben ik trots op jullie. Jullie zijn geboren in de loop van mijn promotieonderzoek en het mooie is dat jullie de plechtige verdediging bewust mee kunnen maken. Bovendien konden jullie helpen bij de omslag van dit proefschrift. Dat maakt het voor mij extra waardevol. Ik hoop dat jullie mij het oneigenlijk gebruik van ouderschapsverlof niet kwalijk nemen.

Lieve, lieve Maaike, ik besef hoe ongelooflijk veel geluk ik heb met jou aan mijn zijde. Als er één persoon op de wereld is die het recht had te zeggen: 'Nu is het genoeg', dan ben jij het. Je hebt er nooit gebruik van gemaakt. Je steunde me van begin tot eind en door dik en dun terwijl jij ook je scholing en werk had als verpleegkundige en daarnaast vier kinderen het leven schonk die op al jouw aandacht en zorg konden rekenen. Jouw tocht vroeg om net zo'n lange adem als die van mij. 'Hora est' luidt ook voor jou.

Hilco Prins

Heino, september 2012

Hilbert Prins werd op 31 januari 1965 geboren in Hoogeveen. Zijn ouders gaven hem de roepnaam Hilco.

Hilco behaalde in 1983 zijn VWO diploma aan het Menso Alting College te Hoogeveen. Na anderhalf jaar fysiotherapie en tijdelijk werk volgde hij vanaf 1986 de inservice opleiding tot A-verpleegkundige in het Sophia Ziekenhuis te Zwolle. Na diplomering studeerde hij van 1989 tot 1993 Gezondheidswetenschappen aan Rijksuniversiteit Limburg waar hij keuzeonderwijs volgde op het terrein van de informatievoorziening in de gezondheidszorg.

In november 1993 startte hij zijn promotieonderzoek bij de afdeling Klinische Informatiekunde in het Academisch Medisch Centrum (AMC) aan de Universiteit van Amsterdam. Hij participeerde vanaf het begin in het onderwijs bij de studie Medische Informatiekunde en werd in 1995 onderwijscoördinator. In eerste instantie had hij een aanstelling als AIO, later als universitair docent. Na tweemaal eerder te zijn genomineerd, won hij in 2002 de MFAS-onderwijsprijs.

In 2003 startte hij als docent bij de opleiding Verpleegkunde aan Hogeschool Windesheim te Zwolle waar hij in 2012 als hogeschoolhoofddocent werd benoemd. Zijn aandachtsgebieden zijn wetenschappelijk en praktijkgericht onderzoek als basis voor het verpleegkundig handelen en ICT toepassingen in de zorg. Naast onderwijsuitvoering deed hij veel aan onderwijsontwikkeling waaronder projectleiding curriculumvernieuwing. Sinds 2009 is hij tevens als onderzoeker verbonden aan het lectoraat ICT-innovaties in de zorg.

In 1988 leerde hij Maaike van Gaal kennen en in 2000 trouwden zij. Zij kregen vier kinderen: Tijmen, Hielke, Jelmer en Emma.

Hilco hoopt, ‘ijs en wederdienende’, op 2 november 2012 te promoveren.