International Journal for Quality in Health Care 2003; **Volume 15, Number 6**: pp. 473–485 10.1093/intqhc/mzg078

'Emerge': benchmarking of clinical performance and patients' experiences with emergency care in Switzerland

DAVID L. B. SCHWAPPACH[1], ANNETTE BLAUDSZUN[2], DIETER CONEN[3], HEINZ EBNER[4], KLAUS EICHLER[2] AND MARC-ANTON HOCHREUTENER[2]

[1]Department of Health Policy and Management, Faculty of Medicine, University Witten/Herdecke, Witten, Germany, [2]Verein Outcome, Zurich, [3]Department of Internal Medicine, Aarau, Switzerland, [4]Koeck, Ebner and Partner, Vienna, Austria

Abstract

Objective. To assess the effects of uniform indicator measurement and group benchmarking followed by hospital-specific activities on clinical performance measures and patients' experiences with emergency care in Switzerland.

Design. Data were collected in a pre–post design in two measurement cycles, before and after implementation of improvement activities. Trained hospital staff recorded patient characteristics and clinical performance data. Patients completed a questionnaire after discharge/transfer from the emergency unit.

Setting. Emergency departments of 12 community hospitals in Switzerland, participating in the 'Emerge' project.

Subjects. Eligible patients were entered into the study (18 544 in total: 9174 and 9370 in the first and second cycles, respectively), and 2916 and 3370 patients returned the questionnaire in the first and second measurement cycles, respectively (response rates 32% and 36%, respectively).

Main outcome measures. Clinical performance measures (concordance of prospective and retrospective assessment of urgency of care needs, and time intervals between sequences of events) and patients' reports about care provision in emergency departments (EDs), measured by a 22-item, self-administered questionnaire.

Results. Concordance of prospective and retrospective assignments to one of three urgency categories improved significantly by 1%, and both under- and over-prioritization, were reduced. The median duration between ED admission and documentation of post-ED disposition fell from 137 minutes in 2001 to 130 minutes in 2002 ($P < 0.001$). Significant improvements in the reports provided by patients were achieved in 10 items, and were mainly demonstrated in structures of care provision and perceived humanity.

Conclusion. Undertaken in a real-world setting, small but significant improvements in performance measures and patients' perceptions of emergency care could be achieved. Hospitals accomplished these improvements mainly by averting strong outliers, and were most successful in preventing series of negative events. Uniform outcomes measurement, group benchmarking, and data-driven hospital-specific strategies for change are suggested as valuable tools for continuous improvement. Several hospitals have already implemented the developed measures in their internal quality systems and subsequent measurements are projected.

Keywords: benchmarking, emergency care, patient satisfaction, quality of health care, safety

In 2000, the Swiss Federal Social Insurance Office and the 'Verein Outcome' (Outcome Association) initiated the nationwide project 'Emerge' to assess and improve the quality of care provided by Swiss emergency departments (EDs). The Outcome Association is a non-profit organization that was founded by the health authority of the Canton Zurich as part of a broad initiative for the improvement of hospital quality of care [1]. The primary aim of the project was to establish a nationwide pilot benchmarking project and to evaluate the methodology and feasibility for the initiation of quality improvement activities in hospitals. The focus was set on emergency care departments as these provide highly interdisciplinary care and are among the most sensitive, challenging, and risk-entailing areas of hospital care. In addition, EDs are

Address reprint requests to David L. B. Schwappach, Department of Health Policy and Management, Faculty of Medicine, University Witten/Herdecke, Alfred-Herrhausen Strasse 50, 58448 Witten, Germany. E-mail: davids@uni-wh.de

of central importance for the corporate image and economic success of individual hospitals.

'Emerge' comprised seven phases: (1) selection of interested hospitals, participating on a voluntary basis; (2) joint development of a set of clinical performance indicators agreed upon by all parties; (3) establishment of a measurement system, development of measurement tools (manuals, training), and design of data collection instruments; (4) data collection in a first measurement cycle; (5) benchmarking of results and definition of shared, quantitative targets; (5) initialization of hospital-specific improvement activities; (6) data collection in a second measurement cycle; and (7) benchmarking of results. The normative framework of the project included the focusing of results, the orientation on the patient's perspective, an interdisciplinary, cooperative approach in the development of indicators, and feasibility of measurement in routine care. All parties agreed that interpretation of measurement results and benchmarking should be guided by a culture of organizational learning rather than individual blame. The developed indicator set includes two main components: objective measures that evaluate clinical performance in terms of speed and accuracy of patient assessment, and patients' experiences with care provided by EDs. In this contribution, we report the results of two measurement cycles, pooled across the 12 participating community hospitals. We analyze whether uniform indicator measurement and benchmarking followed by hospital-specific activities and changes in processes or structures of care resulted in improved clinical performance measures and patients' reports.

Patients and methods

The Swiss emergency health care system

Patients in Switzerland have direct, unlimited access to primary care physicians in an ambulatory care setting and free access to outpatient specialist treatment unless they are insured with managed care organizations with gatekeeper systems (currently $<10\%$ of the population). There are no barriers to hospital-based EDs and non-health related factors often affect decisions whether to seek care in an ED rather than in primary care. Especially in urban areas, and during nights or weekends, EDs are often used as substitutes for primary care physicians by patients. EDs in Switzerland are very diverse and have no uniform structure. The level of training of physicians working in EDs varies largely and depends mainly on the size of hospitals and internal policies. In larger hospitals, ED physicians are often specialists, certified in emergency or intensive care. In smaller hospitals, ED care is usually provided by residents without formal, postgraduate emergency care qualifications, closely supervised by higher-qualified physicians (attendings or seniors). Residents have usually attended qualified in-house training.

Indicators and measurement

Clinical performance in terms of speed and accuracy of patient assessment was assessed by seven indicators, of which six measure the time intervals between the sequences of events from admission to established diagnosis (Box 1). One indicator captures the accuracy of the first medical decision to prioritize patients according to urgency of care needs. ED physicians triaged patients in one of three urgency classes ('emergent', 'urgent', or 'non-urgent') in the first medical evaluation based on vital signs (and cardiogram if appropriate), visual assessment of clinical presentation, and suspected diagnosis. Patients were re-assessed retrospectively by more-qualified physicians after completed diagnostic work-up based upon all clinical information available at the end of emergency care, i.e. with documentation of post-ED disposition. The indicator reflects the concordance of initial and subsequent re-assessment, and the frequency of up- and down-grading (Box 1). No protocols or algorithms were used for the classification of urgency of care needs. However, hospital staff had been extensively trained in measurement in a series of workshops preceding both measurement cycles. For the assessment of urgency of care needs (triage), detailed lists of typical conditions and clinical presentations falling in the respective urgency of care need category were prepared and discussed by professionals of the participating hospitals in the workshops. For the assessment of patients' experiences, a 22-item questionnaire was developed using qualitative methods in samples of former patients from the participating hospitals and hospital staff [2]. This questionnaire uses 'report'-like questions rather than ratings of satisfaction. As others have noted, responses to such 'report'-like questions are readily interpretable, and actions to be taken to improve quality can often be derived directly [3].

Patients and setting

Patients were recruited among all patients admitted to the emergency units of the study hospitals from 18th April to 15th May 2001 (first cycle) or from 3rd April to 1st May 2002 (second cycle), either by self-selection or referral by any health care facility. All in- and outpatients were eligible in case they were considered to have an 'emergency care problem', either from the professional's or the patient's perspective. EDs of 12 community hospitals in Switzerland participated, comprising small rural and large urban regional hospitals, and two university hospitals. The mean number of beds in these institutions is 334 (range 106–1060) and the mean number of emergency admissions in 2000 was 11 725 per hospital (range 3000–24 755).

Data collection

Demographic, clinical, and performance data were collected via specific data sheets by hospital staff: attending physicians, responsible for the diagnostic and therapeutic decisions, assigned acuity scores at presentation and at re-assessment. Physicians recorded information on time points relating to acuity assessment, diagnostic work-up, and treatment. Time points relating to administrative procedures, patient information and nursing care were recorded by nurses. Each hospital nominated a person to be in charge of ensuring and supervising

Box 1 Definitions of performance measures

Indicator	Description
1. Over-, under-, and accurate prioritization of urgency of care needs Category I: highest priority (acute life-threatening situation, e.g. myocardial infarction, polytrauma) Category II: high priority (acute severe condition, e.g. acute appendicitis, hip fracture) Category III: low priority (moderate to mild condition,e.g. superficial cut, rash)	a. Rate of agreement between first and second assessments: proportion of patients classified in the same category at both assessments[1] b. Rate of over-prioritization: proportion of patients downgraded from category I or category II in the initial assessment to categories II and III in the second assessment c. Rate of under-prioritization: proportion of patients upgraded from category II or category III in the initial assessment to categories I and II in the second assessment
2. Time interval from ED admission to initial medical assessment (assignation of course of diagnostic and therapeutic interventions)	
3. Time interval from ED admission to second medical assessment and diagnosis (end of emergency care and decision about discharge/ transfer destination)	
4. Time interval from initial to second medical assessment	
5. Time interval from ED admission to begin of nursing activities	
6. Time interval from ED admission to history taking	
7. Time interval from ED admission to provision of information to the patient about the diagnosis and subsequent treatment course	

[1]The re-assessment (i.e. second assessment) seeks a second retrospective appraisal of the initial situation ('considering all the information available now, was the situation life-threatening?') rather than an evaluation of the initial assessment ('did the situation seem to be life threatening?').

correct data collection. The 'Verein Outcome' provided comprehensive additional support to ensure data quality. This included recurrent training in data collection for hospital staff before measurements, a manual describing the indicators and the data collection procedure, answering frequently asked questions, hotline support during measurement phases, and data controlling. Data sheets were completed at discharge and transferred to the Verein Outcome office. Patients' evaluations of care provided by EDs were obtained via self-administered questionnaires dispensed by the attending carers at the end of ED stay (at discharge or during the 3 days following hospital admission). The procedure followed a standardized course that was described and explained in the measurement manuals and exercised in workshops. Patients (and parents of paediatric patients) received the questionnaire with a covering letter and a reply-paid envelope to complete either before discharge or at home, and were asked to return the completed questionnaire in the closed envelope to hospital staff or to post it to the specified neutral postbox address. Hospital staff were encouraged to remind patients of the questionnaire 12 h after hand-out, and were advised on appropriate communication and presentation of the questionnaire. In case patients felt unable to complete the questionnaires alone, they were encouraged to ask their accompanying persons, or, if necessary, hospital staff to help them. Questionnaires were available in German, Italian, and French. Foreign patients were offered support by interpreters. Questionnaires were confidential and quasi-anonymous (coded by code number), and responses were related to clinical data sheets by code number.

Processing and analyzing data

Random samples of transmitted data sheets were subject to systematic data controlling.

The initial 5% of datasets expected to be transferred by each hospital and subsequent weekly 5% random samples of each hospital's transferred datasets were checked for eligibility, completeness, and a set of pre-defined plausibility tests, consecutively. These included checks for contradicting data (e.g. discharge of males to the department of obstetrics), checks for double information (e.g. more than one post-ED disposition noted), and plausibility of time measurements (e.g. whether the hour noted for admission was before the hour

noted for the triage decision). In addition, the number of transferred cases was compared to the expected number of cases on an individual hospital level, based on historical data. Hospitals were informed weekly on the quality of data they provided. A commercial data processing company scanned data sheets and questionnaires, and returned data on file. Data sheets and questionnaires were then merged by code number.

Statistical analyses

Responders to the questionnaire were compared with non-responders in terms of several variables using unpaired *t*-tests and χ^2 tests. These procedures were also applied to test for differences in performance data between the years. Differences in questionnaire responses between cycles were analyzed using Mann–Whitney tests. Differences in subsamples were tested using the χ^2 or Fisher's exact test in case of at least one frequency < 5. We also calculated a 'problem score' for survey data: for each question, responses were dichotomized as 0 (no problem) or 1 (a serious problem, represented by the least favorable response code) [4]. Patients that reported one or more problems, i.e. those who scored '1' on any question, were considered to have experienced an ED admission 'with a patient-evaluated problem'. We then compared the fraction of patients that reported a patient-evaluated problem between the years. Multiple logistic regression (forward selection) was used to adjust simultaneously for confounding factors. Confidence intervals (CIs) are reported on the 95% level. The level of statistical significance was set at $P < 0.05$. Data were analyzed with the statistical software Stata 8 [5].

Results

Performance data were obtained for 9174 and 9370 patients in the first and second measurement cycles, respectively. Characteristics of patients are provided in Table 1. Distributions of patients' age and sex do not significantly differ between measurements. However, patients in the second measurement were less likely to be self-selected for emergency care, were more often transferred to a non-intensive care unit, and were less likely to be transferred to immediate surgery. Also, the distribution of conditions and affected organ-systems differs slightly between measurements. An analysis of the diagnostic interventions and tests ordered in 2001 and 2002 shows that no general increase in procedures can be observed. Patients in 2002 were slightly more likely to have magnetic resonance imaging compared with 2001 (0.7% and 0.4%, respectively; $P = 0.006$), and less likely to have echocardiography (1.9% and 2.4%, respectively; $P = 0.047$). There were no differences in the utilization of X-ray, laboratory, sonography, duplexsonography, endoscopy, computer tomography, interventional radiology, or consultation of specialists. A significant shift in the qualification of staff establishing the first preliminary diagnosis could be observed with fewer interns and residents, and more fellows, attendings, and seniors performing the first assessment in 2002 compared with 2001. This effect was even more pronounced for qualification of professionals in charge of the second assessment, indicating a change in hospitals' practice styles. Based on the results of the first measurement cycle, subsequent benchmarking, and process analyses, a number of improvement activities were initiated in individual hospitals covering a wide range of targets, from investment in ED structures to professional education and organization of care (Box 2).

Clinical performance measures

Small, but significant improvements were achieved in all performance measures (Table 2). Concordance of first and second assignment to one of three urgency categories improved by 1%. Both under- and over-prioritization were reduced. After adjusting for confounders (age, sex, admission mode, underlying condition, qualifications of physicians assessing and re-assessing acuity, post-ED disposition) via multiple logistic regression, the odds ratio for any misclassification in 2002 compared with 2001 was 0.82 ($P < 0.001$; CI 0.73–0.91). The adjusted odds ratio for under-prioritization, i.e. upgrading of urgency of care needs in the second assessment, in 2002 compared with 2001 is 0.62 ($P < 0.001$; CI 0.49–0.79). The major source of improvement is an increase in appropriate classifications of 'urgent' patients (90.6% in 2001 compared with 93.1% in 2002; $P < 0.001$). Correct identification of 'emergent' patients with the first assessment increased from 83.4% in 2001 to 85.1% in 2002 ($P = 0.558$). Retrospective classification of acuity (second assessment) was strongly associated with post-ED disposition ($P < 0.001$). The hospitalization rate ranged from 94.4% in emergent, to 79.2% in urgent and 23.9% in non-urgent patients.

All time intervals between the sequences of events from admission to established diagnosis decreased significantly, whereas subsequent intervals from diagnosis to documentation of post-ED disposition could not be shortened. In sum, the median duration between ED admission and documentation of post-ED disposition fell from 137 minutes in 2001 to 130 minutes in 2002 ($P = 0.002$). Again, patients that were evaluated as 'urgent' in the preliminary assessment profited most, with a decrease in median total time from ED admission to documentation of post-ED disposition from 190 minutes in 2001 to 178 minutes in 2002 ($P < 0.001$). An analysis of means, medians, and distribution of time intervals shows that hospitals achieved these improvements mainly by cutting down intervals at the upper extremes (outliers).

Results of questionnaires

Questionnaires were obtained from 2916 patients in the first measurement, and from 3370 patients in the second (response rates 32% and 36%, respectively; $P < 0.001$). Comparison of responders and non-responders reveals that responders were slightly older (49 versus 47 years; $P < 0.001$), less likely to be self-selected for ED admission (48.5% versus 52.5%; $P < 0.001$) or transferred by ambulance or rescue services (11.0% versus 13.2%; $P < 0.001$), and were more likely to be classified as 'urgent' (28.6% versus 26.1%; $P < 0.001$).

Table I Patient characteristics by measurement cycle

Characteristics	Number of patients	
	2001 ($n=9174$)[1] [n (%)]	2002 ($n=9370$)[1] [n (%)]
Age, years ($P=0.406$)		
≤5	232 (2.6)	204 (2.2)
6–15	428 (4.7)	416 (4.5)
16–25	1261 (13.9)	1287 (13.9)
26–39	1999 (22.1)	1996 (21.5)
40–65	2695 (29.7)	2773 (29.9)
66–80	1502 (16.6)	1555 (16.8)
≥81	946 (10.4)	1037 (11.2)
Sex ($P=0.068$)		
Female	4119 (45.5)	4340 (47.0)
Male	4939 (54.5)	4930 (53.0)
Source of referral[2] ($P=0.04$)		
Self-referral	4587 (51.6)	4619 (50.6)
GP, as emergency	2181 (24.6)	2292 (25.1)
GP, as non-emergency	102 (1.2)	150 (1.6)
Other external physician	643 (7.2)	615 (6.7)
Ambulance and rescue services	1088 (12.3)	1156 (12.7)
Other	283 (3.2)	293 (3.2)
Condition category (as of second assessment)[3] ($P<0.001$)		
Blood and lymphatic	125 (1.4)	119 (1.3)
Infectious and parasitic	444 (4.9)	588 (6.4)
Endocrine and metabolic	84 (0.9)	82 (0.9)
Mental	172 (1.9)	198 (2.2)
Nervous	631 (7.0)	648 (7.1)
Eye	42 (0.5)	67 (0.7)
Ear	137 (1.5)	88 (1.0)
Circulatory	907 (10.1)	913 (9.9)
Respiratory	434 (4.8)	503 (5.5)
Digestive	1111 (12.4)	1171 (12.7)
Genitourinary	368 (4.1)	366 (3.9)
Obstetric	18 (0.2)	23 (0.3)
Injuries without fracture	2399 (26.7)	2314 (25.2)
Injuries with fracture	907 (10.1)	946 (10.3)
Poisoning	126 (1.4)	145 (1.6)
Other	1083 (12.1)	1020 (11.1)
Discharge/transfer destination[2] ($P=0.029$)		
Intensive care	318 (3.5)	286 (3.1)
Immediate surgery	481 (5.4)	486 (5.3)
Non-intensive care unit	2650 (29.5)	2862 (31.4)
Ambulatory care	5323 (59.3)	5315 (58.2)
Other hospital	186 (2.1)	157 (1.7)
Death during EC	13 (0.1)	20 (0.2)

GP, general practitioner; EC, emergency care.
[1]Data may not sum to total n because of missing information for some subjects.
[2]χ^2 $P<0.05$.
[3]χ^2 $P<0.01$.

Responders were also less likely to receive ambulatory treatment (55.6% versus 60.4%; $P<0.001$) and more likely to be transferred to immediate surgery (6.2% versus 4.9%; $P<0.001$) or non-intensive care wards (33.9% versus 28.7%; $P<0.001$). Relative differences between responders and non-responders did not change between the years. There were no differences in response mode (self-completed versus other options) between questionnaires returned in the first and second cycle.

Box 2 Examples of initiated improvement activities

Target: professional education
Systematic advanced training of residents in emergency care
Systematic introductory training in ED care for newly entering residents
Installation of regular interdisciplinary training on ED-specific topics, e.g. management of anxiety and pain
Enhanced clinical supervision of residents
Target: process development
Implementation of guidelines for the assessment and treatment of certain clinical conditions/clinical situations
Process re-engineering to increase speed of triage and reduction of waiting times
Introduction of pain protocols and dolometers
Installation of a risk management and error detection system
Target: communication
Training in appropriate communication to improve manner, diction, and timing when communicating with ED patients
Development of internal conventions for the avoidance of X-ray evaluation and teaching while patients present
Development of internal conventions regarding explicit nomination of professionals in charge of the information on patients with unclear diagnoses
Development of internal conventions on when, how, and by whom patients are informed about whether and when they are allowed to eat or drink
Development of a communication concept regarding waiting times
Target: human resources and organizational development
Change in practice style: all surgical/trauma patients must be evaluated by the surgical department's chairperson
Development of guidelines for time intervals between alerting medical staff and arrival at the ED
Recruitment of additional staff
Target: buildings and investments
Investment in new beds, cubicles, and curtains
Establishment and equipment of dedicated triage rooms
Acquisition of snack bars and drink dispensers

Detailed responses to the questionnaire are provided in Tables 3–6. Significant improvements in the reports provided by patients could be achieved for 10 items. Improvements were mainly demonstrated in reports relating to structures of care provision and perceived humanity. The rate of patients with a problem score of 1 (i.e. reporting any ED problem) decreased from 41.3% in 2001 to 36.2% after the benchmarking intervention (CI of the difference 3–7%; $P<0.001$), a relative decrease of 14%. After adjusting for age, sex, admission mode, final urgency classification, diagnosis, post-ED disposition, and survey response mode, the odds ratio of reporting any problem in 2002 versus 2001 was 0.81 (CI 0.72–0.90; $P<0.001$). An analysis of the distribution of problem scores on the individual patient level reveals that not only the fraction of responders that experienced any ED problem, but also the number of problems reported per patient, decreased. While 12% experienced more than three serious ED-related problems in 2001, this fraction decreased to 9% in 2002 ($P=0.012$).

There are major differences in the experiences with ED care according to initial urgency classification on the content level. Calculated over both measurement cycles, the top three items emergent patients were most likely to report negative experiences with were, in descending order, contradictions between different staff members (item 7), ED rooms (item 1), and information about diagnosis (item 8). Urgent and non-urgent patients were most likely to report negative experiences with preservation of privacy (item 5), opportunities for eating and drinking (item 6), and waiting times (item 19). Improvements in patients' experiences and perceptions of care were mainly achieved in patients initially evaluated as 'non-urgent'. The fraction reporting one or more ED problems decreased from 42.3% (2001) to 36.9% (2002) in 'non-urgent' patients ($P<0.001$), from 40.0% to 35.7% in 'urgent' patients ($P=0.043$), and from 28.9% to 28.6% in 'emergent' patients ($P=0.962$).

Although on the aggregate level decreases in average actual waiting times coincide with improvements in subjectively perceived waiting times (item 19) and information about the causes underlying delays (item 21), actual waiting times are only moderately correlated with patients' perceptions of waiting times on the individual patient level (Spearman's rho $\rho=0.26$; $P<0.001$).

Discussion

In this study, undertaken in a real-world setting, small but significant improvements in performance measures and patients' perceptions of emergency care could be achieved by uniform outcomes measurement, group benchmarking, and hospital-specific strategies for change. In contrast to other studies that have evaluated frequency of diagnostic errors or concordance of ED and discharge diagnoses, we assessed identification of patients' urgency of care needs as an indicator of ED performance before the establishment of diagnosis [6–8]. In

Table 2 Changes in performance measures between two measurements

Performance measure	2001 ($n=9174$)[1] [n (%)]	2002 ($n=9370$)[1] [n (%)]
Accuracy of prioritization[2]		
1a. Accurate prioritization[3]	8066 (90.4)	8344 (91.5)
1b. Under-prioritization	198 (2.2)	136 (1.5)
1c. Over-prioritization	655 (7.3)	641 (7.0)
Average time interval between stages of care (minutes)		
2. Admission to first assessment		
Mean	34.1	28.5
95% confidence interval	33.1–35.2	27.9–29.1
Median[4]	20	20
Patients with missing data (%)	4	4
3. Admission to second assessment		
Mean	104.5	93.4
95% confidence interval	102.2–106.8	91.5–95.4
Median[4]	75	70
Patients with missing data (%)	5	4
4. First to second assessment		
Mean	74.2	68.7
95% confidence interval	71.9–76.5	66.7–70.7
Median[5]	45	45
Patients with missing data (%)	4	3
5. Admission to nursing activities		
Mean	11.3	7.2
95% confidence interval	10.4–12.2	6.8–7.7
Median[4]	5	2
Patients with missing data (%)	5	5
6. Admission to history taking		
Mean	25.0	20.5
95% confidence interval	24.0–26.0	20.0–21.0
Median[4]	15	15
Patients with missing data (%)	5	5
7. Admission to diagnosis/treatment information provided to patient		
Mean	114.5	103.5
95% confidence interval	112.0–117.1	101.0–106.0
Median[4]	85	77
Patients with missing data (%)	5	5

[1]Data may not sum to total n because of missing information for some subjects.
[2]χ^2 $P<0.001$.
[3]Definitions of indicators are given in Box 1.
[4]Mann–Whitney rank sum test $P< 0.001$.
[5]Mann–Whitney rank sum test $P = 0.014$.

terms of patient safety in fast-track areas, the latter may be more important. Participating hospitals used data for misclassified patients and ultimate clinical outcome in grand rounds and internal supervision as a tool for continuous learning and risk management. However, one has to consider that not all disconcordances between initial and re-assessment are due to 'errant' triage decisions. It is instead one of the key characteristics of emergency care to prioritize patients at high risk for life-threatening conditions that have not yet been confirmed. The fact that frequencies of both up- and down-grading decreased is promising. It indicates that hospitals did in sum not increase safety at the cost of efficiency by systematically shifting initial assessments towards higher urgency classes. Results of measurement of time intervals and questionnaires imply that improvements in speed and patients' reports were mainly achieved by averting strong, negative outliers and series of negative experiences, and by targeting concentrated deficits.

As others, we observed an association between urgency of care needs and patients' experiences, with non-urgent patients

Table 3 Patient survey responses related to their perceptions of safety and staff technical skills [*n* (2001) = 2916; *n* (2002) = 3370][1]

No.	Question	Response category								*P*-value
		Yes, absolutely		Somewhat		Little		No, not at all		
		2001 [*n* (%)]	2002 [*n* (%)]	2001 [*n* (%)]	2002 [*n* (%)]	2001 [*n* (%)]	2002 [*n* (%)]	2001 [*n* (%)]	2002 [*n* (%)]	
9	Did you trust the accuracy of the diagnosis made?	2173 (77.7)	2603 (79.9)	501 (17.9)	555 (17.0)	84 (3.0)	69 (2.1)	39 (1.4)	32 (1.0)	0.025[2]
10	Did you think that all necessary tests and examinations had been carried out?	2340 (82.3)	2759 (83.8)	387 (13.6)	437 (13.3)	71 (2.5)	68 (2.1)	47 (1.7)	27 (0.8)	0.068
15	Did you have the impression that the staff had good professional skills?	2436 (85.7)	2820 (86.0)	357 (12.6)	417 (12.7)	35 (1.2)	29 (0.9)	16 (0.6)	14 (0.4)	0.667

[1]Data may not sum to total *n* because of missing information for some subjects.
[2]Mann–Whitney test $P < 0.05$.

more frequently providing negative reports of care provision [9,10]. While these patients also benefited most in terms of patient satisfaction, improvements in performance measures were primarily achieved in patients initially evaluated as urgent. We found little effects on care provided to emergent patients and the problems reported by these patients. One has to take into account, however, that the problem score does, by definition, only capture changes in responses that affect the least favorable response code, and systematic improvements within the other answer categories are not detected. In addition, equal importance is attached to all questions. It is a normative decision which response codes should be evaluated as being 'problematic'. Although only responses in the least favorable category were used to calculate problem scores in this study, all three categories indicating 'non-optimal' experiences are used for benchmarking. This approach is rooted in the understanding that every non-optimal experience offers the potential for improvement. Conversely, to increase discriminatory power and to identify subpopulations at increased risk for serious ED experiences, we report the problem score as a binary measure contrasting the least favorable response code.

Consistent with the literature, this study also provides further evidence that actual waiting times and patients' perceptions of waiting times are distinct concepts. As others have noted, perceptions of waiting times and satisfaction with waiting times may be strongly influenced by expected waiting times, expressive quality, and provision of information about the causes underlying delays, rather than being determined by actual waiting times [11–14]. Participating hospitals achieved improvements in patients' subjective evaluation of waiting times with systematic changes in ED information and communication patterns.

The study also has some limitations: the major disadvantage is that the reliability and validity of the clinical performance measures, in particular of triage decisions, were not assessed. A number of studies have shown triage decisions based on three-tier triage scales to suffer from poor interrater reliability [15–19]. As we did not study the final outcomes of ED admissions, one may question the clinical significance of triage decisions as well as the accuracy of the retrospective second assessment. Although we cannot provide direct evidence, there are a number of reasons that suggest the reliability and validity of the indicators to be higher than the figures reported elsewhere. The initial triage decision at ED admission determines the prioritization of patients, and the urgency and speed at which further procedures are to be undertaken. The identification of these needs and the avoidance of delays are of vital importance in a number of conditions frequently seen in EDs (e.g. initiation of thrombolysis in acute myocardial infarction, or prevention of perforation in acute appendicitis). The strong association we found between triage status and post-ED disposition can be considered as an indirect indication of its construct validity. The retrospective assessment was performed by more qualified and experienced physicians, based on all relevant clinical information and after the post-ED disposition has been documented. It was therefore well accepted in the project that the second assessment would

Table 4 Patient survey responses related to aspects of the ED care structure [*n* (2001) = 2916; *n* (2002) = 3370][1]

No.	Question	Response category										*P*-value
		Yes, absolutely		Somewhat		Little		No, not at all		Not applicable[2]		
		2001 [*n* (%)]	2002 [*n* (%)]	2001 [*n* (%)]	2002 [*n* (%)]	2001 [*n* (%)]	2002 [*n* (%)]	2001 [*n* (%)]	2002 [*n* (%)]	2001 [*n* (%)]	2002 [*n* (%)]	
1	Did you feel cramped in the rooms of the emergency department?	219 (7.6)	194 (5.7)	232 (8.1)	259 (7.8)	680 (23.6)	746 (22.5)	1750 (60.7)	2114 (63.8)	N/O	N/O	0.006[4]
2	Was it sufficiently warm in the emergency department?	2796 (75.7)	2526 (75.7)	523 (18.1)	578 (17.3)	135 (4.7)	166 (5.0)	37 (1.3)	66 (2.0)	N/O	N/O	0.651
3	Did you experience any draught in the emergency department?	69 (2.4)	72 (2.2)	66 (2.3)	85 (2.6)	367 (12.7)	416 (12.5)	2386 (82.6)	2748 (82.7)	N/O	N/O	0.897
4	Were the beds comfortable?	1166 (41.2)	1434 (44.0)	900 (31.8)	1022 (31.3)	161 (5.7)	166 (5.1)	80 (2.8)	83 (2.5)	525 (18.5)	557 (17.1)	0.016[3]
5	Were you sufficiently shielded during examinations so that other patients could not overhear?	1149 (40.5)	1413 (43.0)	580 (20.5)	725 (22.1)	426 (15.0)	473 (14.4)	402 (14.2)	369 (11.2)	277 (9.8)	305 (9.3)	0.004[4]
6	Were there sufficient opportunities for you to eat and drink?	527 (20.2)	692 (22.6)	97 (3.7)	128 (4.2)	72 (2.8)	83 (2.7)	143 (5.5)	129 (4.2)	1765 (67.8)	2032 (66.3)	0.019[3]
19	Did you have to wait for a long time within the emergency department?	381 (13.4)	375 (11.3)	577 (20.2)	641 (19.3)	685 (24.0)	784 (23.6)	1208 (42.4)	1516 (45.7)	N/O	N/O	0.002[4]

N/O, response option not offered.
[1]Data may not sum to total *n* because of missing information for some subjects.
[2]Response wording varied with question, e.g. one response to question 5 read 'There were no other patients'.
[3]Mann–Whitney test $P < 0.05$.
[4]Mann–Whitney test $P < 0.01$.

Table 5 Patient survey responses related to communication [*n* (2001) = 2916; *n* (2002) = 3370][1]

No.	Question	Response category										*P*-value
		Yes, absolutely		Somewhat		Little		No, not at all		Not applicable[2]		
		2001 [*n* (%)]	2002 [*n* (%)]	2001 [*n* (%)]	2002 [*n* (%)]	2001 [*n* (%)]	2002 [*n* (%)]	2001 [*n* (%)]	2002 [*n* (%)]	2001 [*n* (%)]	2002 [*n* (%)]	
7	In your opinion, were there any contradictions between different staff members involved in the examinations?	209 (7.5)	207 (6.4)	N/O	N/O	N/O	N/O	2099 (75.1)	2455 (75.8)	488 (17.5)	575 (17.8)	0.104
8	Was it difficult for you to obtain information about the diagnosis made?	125 (4.4)	151 (4.6)	269 (9.5)	263 (8.0)	401 (14.1)	493 (15.0)	2048 (72.0)	2378 (72.4)	N/O	N/O	0.623
11	In case no final diagnosis could be made, was this sufficiently explained to you?	1028 (39.2)	1257 (41.2)	335 (12.8)	387 (12.7)	76 (2.9)	78 (2.6)	48 (1.8)	31 (1.0)	1139 (43.4)	1300 (42.6)	0.059
12	Were the staff kind and friendly?	2678 (93.1)	3130 (94.1)	170 (5.9)	168 (5.1)	18 (0.6)	17 (0.5)	10 (0.4)	10 (0.3)	N/O	N/O	0.100
20	Was the course of examinations and procedures sufficiently explained to you?	2063 (72.8)	2470 (75.3)	572 (20.2)	611 (18.6)	119 (4.2)	133 (4.1)	79 (2.8)	65 (2.0)	N/O	N/O	0.021[3]
21	Were the causes underlying potential delays sufficiently explained to you?	898 (32.4)	1153 (36.2)	430 (15.5)	458 (14.4)	198 (7.2)	196 (6.2)	173 (6.3)	164 (5.2)	1069 (38.6)	1216 (38.2)	<0.001[4]

[1]Data may not sum to total *n* because of missing information for some subjects.
[2]Response wording varied with question, e.g. one response to question 21 read 'I did not experience delays'.
[3]Mann–Whitney test $P < 0.05$.
[4]Mann–Whitney test $P < 0.001$.
N/O, response option not offered.

Table 6 Patient survey responses related to care and support [*n* (2001) = 2916; *n* (2002) = 3370][1]

No.	Question	Response category										*P*-value
		Yes, absolutely		Somewhat		Little		No, not at all		Not applicable[2]		
		2001 [*n* (%)]	2002 [*n* (%)]	2001 [*n* (%)]	2002 [*n* (%)]	2001 [*n* (%)]	2002 [*n* (%)]	2001 [*n* (%)]	2002 [*n* (%)]	2001 [*n* (%)]	2002 [*n* (%)]	
13	Were you treated as human, and not as a number?	2542 (88.3)	2993 (90.1)	232 (8.1)	247 (7.4)	56 (2.0)	35 (1.1)	48 (1.7)	47 (1.4)	N/O	N/O	0.020[3]
14	Did you perceive the information provided as truthful?	2530 (88.9)	2994 (90.9)	277 (9.7)	275 (8.4)	26 (0.9)	16 (0.5)	13 (0.5)	9 (0.3)	N/O	N/O	0.008[4]
16	When you were anxious, were you sufficiently supported by staff?	879 (31.9)	1095 (34.3)	206 (7.5)	248 (7.8)	43 (1.6)	54 (1.7)	28 (1.0)	26 (0.8)	1596 (58.0)	1768 (55.4)	0.552
17	Did the staff take enough measures to relieve your pain?	1663 (59.4)	1964 (60.5)	321 (11.5)	355 (10.9)	69 (2.5)	86 (2.7)	65 (2.3)	57 (1.8)	682 (24.4)	784 (24.2)	0.276
18	Did you have the feeling that staff were reachable whenever you wanted?	2102 (74.3)	2483 (75.8)	544 (19.2)	606 (18.5)	138 (4.9)	140 (4.3)	46 (1.6)	49 (1.5)	N/O	N/O	0.163
22	Had your accompanying persons enough opportunities to support you during waiting times?	1615 (57.8)	1940 (60.3)	174 (6.2)	203 (6.3)	66 (2.4)	56 (1.7)	48 (1.7)	43 (1.3)	889 (31.8)	973 (30.3)	0.100

[1]Data may not sum to total *n* because of missing information for some subjects.
[2]Response wording varied with question, e.g. one response to question 16 read 'I was not anxious'.
[3]Mann–Whitney test $P < 0.05$.
[4]Mann–Whitney test $P < 0.01$.
N/O, response option not offered.

accurately reflect the patient's 'true' state in the vast majority of cases.

Reliability studies have usually used paper cases (vignettes) to measure triage performance and have commonly excluded 'typical ED patients', patients with life-threatening diseases or comorbidities, all of which are included in our study. As opposed to other studies, characteristics of the initial triage process as applied in our study, namely visual assessment by physicians, clinical judgment, and suspected diagnosis based on knowledge of vital signs, have been shown to increase accuracy in terms of post-ED disposition prediction, especially in patients from vulnerable populations (the very young and the elderly) [20–23]. Aimed at developing a common understanding of the acuity scale, hospital staff participated in intensive measurement workshops and indicator discussions, were trained in the application of the scale, and consented on detailed lists of typical conditions. Because this stimulated awareness and sensibility, it can itself be regarded as an intervention that increases reliability. Finally, the relative frequency of assigned triage scores per hospital remained stable between the years (data not shown). With poor reliability, one would expect large variability in the fraction of patients allocated to each of the urgency classes by hospitals in the two measurement cycles.

Another limitation of performance measurement relates to the qualification of professionals that conducted the re-assessment of urgency of care needs. The shift towards higher qualifications in the second measurement cycle may have introduced a detection bias, with more misclassifications detected in the second cycle. This in turn would have artificially decreased the magnitude of improvements achieved in accurate initial assessments.

A more general concern is that surveys applied to ED patients directly after the care experience suffer from bias towards less injured patients, as the severely ill are less likely to respond. We also noted some response bias, with patients admitted via external physicians, evaluated as 'urgent', and those discharged to in-hospital wards being more likely to respond. Although a number of studies have reported comparatively low or even poorer rates, the response rates are dissatisfying and make the interpretation of results difficult [10,24,25]. Finally, we could not relate improvement activities to performance measures and patients' experiences on the individual hospital level. Based on project conventions, hospital-specific data are not disclosed to the public. Concerns about league tabling, i.e. the public release of rankings of institutions based on single performance measures with the effect of 'naming and shaming', would inhibit motivation and set incentives to develop evading strategies, thereby resulting in poor data quality and endangering any culture of continuous learning and improvement [26,27]. The disadvantage is, however, that failures and successes of specific improvement strategies cannot be presented and their impact on results is averaged.

In summary, some improvements in ED performance and patients' experiences could be achieved by participating hospitals in a design of data-driven quality improvement activities with recurrent plan–do–check–act cycles. A number of studies have reported on activities to improve quality of ED care and patient satisfaction, and several interventions have been suggested to improve quality of care, mainly in the field of professional education, staffing, and management of patient flow [4,28–32]. The key differences in our approach are that hospital staff are actively involved not only in benchmarking of results but also in measurement, and that clinical performance measures as well as patient-reported experiences serve as quality indicators. As noted by the authors of the *Harvard Emergency Department Quality Study*, which used a 2-year intervention period in a comparable design, improvements achieved per cycle may only be moderate with rapid cycling [4]. However, we chose to provide hospitals with timely feedback on the results of the actions taken to allow for adjustment of effective strategies and to stimulate institutional learning. Current achievements should therefore not be regarded as final. Several participating hospitals have already implemented the developed measures in their continuous quality management systems and subsequent measurements are projected. Future cycles and analyses for trend will elucidate whether this approach also proves to be successful in the long-term.

Acknowledgements

The authors express their gratitude and respect to the institutions that participated in the 'Emerge' project and provided the data for this study. The invaluable commitment of countless individuals who collected the data, managed measurements, and contributed to discussion of processes and results in addition to their regular activities is greatly appreciated. They are represented here by the contributing institutions and the professionals in charge of outcome measurement: Spital Aarberg (Dr T. Ritschard, M. Friedrich, Dr C. Klaiber); Universitätsspital Kantonsspital Basel (Dr B. Martina, R. Herold, J. Wuhrmann); Spital Bülach (Dr M. Osusky, Dr C. Lauber); Kantonsspital Chur (Dr T. Wieland, R. Lorez, Dr A. Frutiger); Klinik Hirslanden (R. Bartczak, Dr F. Horber, M. Rauber); Universitätsspital Inselspital (Dr F. Nohl, M. Jost, Prof. H. Gerber, Dr M. Hess); Spital Interlaken (R. Enezian, Dr M. Studer, Dr H. Schaad); Spital Lachen (G. Kurpjuhn, Dr I. Guldenschuh, Dr A. Pellegrino, Dr L. Stäger); Ospedale Regionale la Carità Locarno (Dr L. Gabutti, Dr S. Schlunke, L. Merlini); Stadtspital Waid (S. Vetter, M. Herzog, E. Seiler, Prof. H. Bühler); Kantonsspital Winterthur (Dr K. Käch, U. Müller, Dr R. Imoberdorf); Spital Zollikerberg (A. Keller, Dr R. Spalinger, Dr P. Guyer). We also thank patients for taking the time to complete the survey and sharing their experiences of ED care. Finally, we wish to acknowledge the valuable comments of two anonymous referees.

Contributors

M.A.H. and H.E. developed the principal design of the 'Emerge' project. M.A.H., D.C., K.E., A.B., and H.E. participated in the development of performance measures. A.B. and K.E. developed the measurement system, supervised measurements and data collection, and provided support activities to hospitals.

H.E. and D.C. served as methodological consultants during the study period. D.L.B.S. conceived and designed the methodology and conception, analyzed and interpreted the data, and wrote up the paper. A.B., K.E., and M.A.H. contributed to the conception of the study. Each author critically discussed and reviewed the final paper. D.L.B.S. is guarantor.

References

1. Schwappach DLB, Hochreutener M-A, Conen D, Eichler K, Blaudszun A, Koeck CM. Outcome measurement in the Canton of Zurich. From theory to practice—progress report on six years of development, implementation and experience with routine inpatient outcome measurement. Muri: SGGP (Swiss Society for Health Policy), 2003.
2. Rehm J, Blaudszun A, Frick U *et al*. The development of a questionnaire to assess patient satisfaction with emergency units of Swiss hospitals. *Soz Praventivmed* 2003, in press.
3. Sixma HJ, Kerssens JJ, Campen CV, Peters L. Quality of care from the patients' perspective: from theoretical concept to a new measuring instrument. *Health Expect* 1998; **1**: 82–95.
4. Burstin HR, Conn A, Setnik G *et al*. Benchmarking and quality improvement: the Harvard Emergency Department Quality Study. *Am J Med* 1999; **107**: 437–449.
5. StataCorp. *Stata Statistical Software: Release 8.0*. College Station, TX: Stata Corporation, 2003.
6. Guly HR. Diagnostic errors in an accident and emergency department. *Emerg Med J* 2001; **18**: 263–269.
7. Wogan JM. ED follow-up: a comparison of admission and discharge diagnoses. *Am J Emerg Med* 2001; **19**: 249–251.
8. Chellis M, Olson J, Augustine J, Hamilton G. Evaluation of missed diagnoses for patients admitted from the emergency department. *Acad Emerg Med* 2001; **8**: 125–130.
9. Hansagi H, Carlsson B, Brismar B. The urgency of care need and patient satisfaction at a hospital emergency department. *Health Care Manage Rev* 1992; **17**: 71–75.
10. Trout A, Magnusson AR, Hedges JR. Patient satisfaction investigations and the emergency department: what does the literature say? *Acad Emerg Med* 2000; **7**: 695–709.
11. Thompson DA, Yarnold PR. Relating patient satisfaction to waiting time perceptions and expectations: the disconfirmation paradigm. *Acad Emerg Med* 1995; **2**: 1057–1062.
12. Thompson DA, Yarnold PR, Williams DR, Adams SL. Effects of actual waiting time, perceived waiting time, information delivery, and expressive quality on patient satisfaction in the emergency department. *Ann Emerg Med* 1996; **28**: 657–665.
13. Yarnold PR, Michelson EA, Thompson DA, Adams SL. Predicting patient satisfaction: a study of two emergency departments. *J Behav Med* 1998; **21**: 545–563.
14. Naumann S, Miles JA. Managing waiting patients' perceptions: the role of process control. *J Manag Med* 2001; **15**: 376–386.
15. Gill JM, Reese CL, Diamond JJ. Disagreement among health care professionals about the urgent care needs of emergency department patients. *Ann Emerg Med* 1996; **28**: 474–479.
16. Nakagawa J, Ouk S, Schwartz B, Schriger DL. Interobserver agreement in emergency department triage. *Ann Emerg Med* 2003; **41**: 191–195.
17. Reilly BM, Evans AT, Schaider JJ, Wang Y. Triage of patients with chest pain in the emergency department: a comparative study of physicians' decisions. *Am J Med* 2002; **112**: 95–103.
18. Wuerz R, Fernandes CM, Alarcon J. Inconsistency of emergency department triage. Emergency Department Operations Research Working Group. *Ann Emerg Med* 1998; **32**: 431–435.
19. Travers DA, Waller AE, Bowling JM, Flowers D, Tintinalli J. Five-level triage system more effective than three-level in tertiary emergency department. *J Emerg Nurs* 2002; **28**: 395–400.
20. Brillman JC, Doezema D, Tandberg D, Sklar DP, Skipper BJ. Does a physician visual assessment change triage? *Am J Emerg Med* 1997; **15**: 29–33.
21. Cooper RJ, Schriger DL, Flaherty HL, Lin EJ, Hubbell KA. Effect of vital signs on triage decisions. *Ann Emerg Med* 2002; **39**: 223–232.
22. Schoenenberger RA, Conzelmann M, Dubach UC, Schwander J. Quality of emergency room triage of medical inpatients to an acute care clinic or chronic health care facilities. *J Gen Intern Med* 1992; **7**: 321–327.
23. Waldrop RD, Harper DE, Mandry C. Prospective assessment of triage in an urban emergency department. *South Med J* 1997; **90**: 1208–1212.
24. Boudreaux ED, Ary RD, Mandry CV, McCabe B. Determinants of patient satisfaction in a large, municipal ED: the role of demographic variables, visit characteristics, and patient perceptions. *Am J Emerg Med* 2000; **18**: 394–400.
25. Sun BC, Adams JG, Burstin HR. Validating a model of patient satisfaction with emergency care. *Ann Emerg Med* 2001; **38**: 527–532.
26. Thomson RG, McElroy H, Kazandjian VA. Maryland Hospital Quality indicator project in the United Kingdom: an approach for promoting continuous quality improvement. *Qual Health Care* 1997; **6**: 49–55.
27. Shojania KG, Showstack J, Wachter RM. Assessing hospital quality: a review for clinicians. *Eff Clin Pract* 2001; **4**: 82–90.
28. Cobelas C, Cooper C, Ell M, Hawthorne G, Kennedy M, Leach D. Quality management and the Emergency Services Enhancement Program. *J Qual Clin Pract* 2001; **21**: 80–85.
29. Fernandes CM, Christenson JM. Use of continuous quality improvement to facilitate patient flow through the triage and fast-track areas of an emergency department. *J Emerg Med* 1995; **13**: 847–855.
30. Hall MF, Press I. Keys to patient satisfaction in the emergency department: results of a multiple facility study. *Hosp Health Serv Admin* 1996; **41**: 515–532.
31. Jelinek GA, Mountain D, O'Brien D *et al*. Re-engineering an Australian emergency department: can we measure success? *J Qual Clin Pract* 1999; **19**: 133–138.
32. Wallis LA, Guly HR. Improving care in accident and emergency departments. *Br Med J* 2001; **323**: 39–42.

Accepted for publication 31 July 2003