LUNDS UNIVERSITET
Medicinska fakulteten

Lund University, faculty of medicine

Course: LÄKM01

Spring term 2012

Evaluation of checklists as a diagnostic tool in the Emergency Department: a pilot study

Pär Areskoug and Peter Hammarlund

Supervisors: Eric Dryver, MD and Ulf Ekelund, MD, PhD, Associate professor

2012-05-29

Abstract

Introduction: The Emergency Department (ED) is a complex environment. Patients present with many different complaints with alternating severity. This setting makes time efficiency and diagnostic accuracy crucial. The use of checklists in medicine has been successful in several medical fields, such as in reducing surgical complications and in decreasing catheter-related bloodstream infections in the intensive care unit, but has yet to be evaluated in the ED environment. This pilot study aimed to evaluate whether the use of checklists as a diagnostic tool in the ED could improve efficiency and be of benefit in the diagnostic process.

Methods: Checklists were created for nine common patient complaints for use in the medical history uptake, the physical examination and for ordering relevant investigations. An evaluation period of six weeks after the checklist implementation was compared to a control period prior to the checklist implementation, and the subpopulation of patients on whom checklists were used was compared to the same control period. Time measures were used as indicators of efficiency. The proportion of patients leaving without being seen by a physician (LWBS) and admission rate were also used. Additionally, a substudy was conducted where five fairly inexperienced physicians were studied one week not using and one week using the checklists. Time measures, rate of admission and probability assessments for differential diagnoses and admission were used as outcome measures.

Results: Checklists were used in 219 cases (7.9 % of eligible cases in the evaluation period). There was significantly longer patient length of stay (with time waiting for admission excluded) in the checklist subpopulation compared to the control period. Median times were 04:25 and 05:01, respectively ($p<0.001$). The checklist users consisted of a larger proportion of junior physicians than normal (64.4 % vs 41.4 % respectively). The substudy consisted of 61 cases, 36 in which checklists were used. The physicians were not less efficient and slightly better at probability assessments of differential diagnoses ($p= 0.018$) when checklists were used.

Conclusion: Checklist usage appeared to be associated with impaired time efficiency, but the result was confounded by a high proportion of inexperienced physicians using them. The results in a more controlled setting showed no difference in time measures, and a benefit regarding probability assessment of differential diagnoses when physicians used checklists. This proves an interesting point for further investigation.

Introduction

Patients seek care in the Emergency Department (ED) because of a wide variety of problems caused by up to 10 000 diagnoses.[1] The main task of physicians working in the ED is to identify the patients suffering from diagnoses where urgent treatment decreases morbidity and mortality. For example, patients suffering from severe sepsis benefit from early antibiotic treatment[2] and patients suffering from stroke benefit from acute thrombolysis[3]. It follows that the diagnostic work-up of patients in the ED ought to be accurate as well as time-efficient. Diagnostic errors and delayed patient management (e.g. caused by crowding and poorly functioning routines) are both threats to patient safety.

The frequency and impact of diagnostic mistakes is unclear but are thought to be substantial.[4-6] The rate of diagnostic errors in the ED ranges from 0.6 to 12% according to one review[7], and another study estimates that 5% of autopsies reveal that the patient's death might have been avoided by appropriate diagnosis and treatment.[8] Diagnostic errors can be divided into 1) no fault errors, 2) system related errors and 3) cognitive errors.[5] Examples of these include 1) unusual presentation of a disease, 2) organizational flaws and 3) premature closure (i.e. failure to continue considering reasonable alternatives after an initial diagnosis is reached). Cognitive errors seldom appear as the lone cause of misdiagnosis, but are present in a high percentage of cases.[5,6]

Delayed patient management in the ED results from factors that increase patient management time and factors that increase the rate of patient presentation to the ED. Crowding (i.e. high patient occupancy) is a major issue facing modern EDs[9-14] and the result when the rate of presentation to the ED exceeds the rate of discharge. A commonly used definition of crowding is prolonged length of stay (LOS).[9] Crowding results from the large number and high complexity of patients seeking care in the ED[14] combined with high rates of in-hospital occupancy[9,15]. It is associated with increased patient morbidity and mortality[10,11,13,15], decreased adherence to guidelines[16] and increased risk of preventable medical errors.[12] It is also associated with an increase in the proportion of patients leaving without being seen (LWBS) and a decrease in patient satisfaction.[17]

Strategies to improve diagnostic accuracy and timely patient management are sorely needed in the ED. To increase efficiency and quality, EDs are using triage systems like the Medical Emergency Triage and Treatment System (METTS)[18,19], patient streaming into different diagnostic/treatment algorithms[20] and integrated care pathways (ICP). ICPs are structured multidisciplinary documents that direct care (i.e. guidelines). Several studies have reported that ICPs improve certain aspects of patient management, although reductions in morbidity and mortality have yet to be demonstrated.[21-25]

Another strategy that has been gaining interest is the use of checklists. Checklists have been successfully used for a long time in the aviation industry.[26] In medicine, checklists have been shown to reduce central line infections in the intensive care unit (ICU) [27], the World Health Organization (WHO) surgical safety checklist have been proven to significantly decrease surgical complications worldwide [28,29] and beneficial effects have also been shown in other medical areas.[30-32] However, the potential benefits of checklists as a diagnostic tool and a tool to expedite care in the ED are poorly studied. One study discusses the potential use of checklists to reduce diagnostic errors but does not provide any experimental data.[33]

The aim of this study was to evaluate whether the implementation of complaint-specific checklists in the ED can benefit the diagnostic process and improve the efficiency of patient management. Since this was a pilot study, an additional study aim was also to assess whether the study design can effectively evaluate checklist implementation in the ED.

Methods

This pilot study was conducted in the Emergency Department (ED) of Skane's University Hospital, Lund, Sweden. Lund's ED receives approximately 65 000 patient visits per year. Excluding orthopaedic complaints, the nine of the most common presenting complaints are abdominal pain, chest pain, dyspnea, neurological deficits, head injury, headache, dizziness, transient loss of consciousness (TLoC) and allergic reactions. Checklists for these complaints were developed by Dr E.D. (specialist in Emergency Medicine) on the basis of articles that investigated the diagnostic value of clinical information and with the input of Emergency Medicine colleagues when evidence was lacking. The checklists specify the information from the history, physical examination and tests that should routinely be gathered from all patients presenting to the ED with the specific complaint, as well as the additional information that should be gathered when specific diagnostic hypotheses are suspected (Appendix 1a and 1b). For example, the checklist for abdominal pain includes asking the patient about pain radiation; the checklist specifies that the blood pressure should be checked in both arms if aortic dissection is suspected.

Complaint-specific reference documents for each checklist were available on the ED's computer desktops, and provided evidence-based guidance on how to interpret the gathered information. For example, the checklist for dyspnea recommends a routine assessment of the jugular venous pressure. The associated reference document states that an elevated jugular venous pressure has a likelihood ratio (LR) of around 5 for congestive heart failure (thus increasing the odds of the diagnosis by a factor of 5) whereas a normal pressure has a LR of 0.66 (thus decreasing the odds for congestive heart failure by a third). The content of the checklists and reference documents were open for additional changes during the whole period, since the goal was to have the most current information possible.

The checklists were implemented in the ED on the 12th of March 2012. They were to be attached, by a triage nurse, to the charts of all patients presenting to the ED with one of the complaints listed above, and were also made available in the room were the physicians dictate their notes and on every computer desktop within a distinct folder. Their use was encouraged, but not mandatory. Since the checklists were assumed to be user-friendly and self-instructing, they were

not formally introduced. Short oral information was however given during the first weeks. The first two weeks were considered to be an implementation period and no data was collected during this period. The data collection for the evaluation period started on the 26th of March and ended on the sixth of May. The period between the second of January and the eleventh of March was chosen as a control period prior to checklist implementation.

Inclusion criteria for the study were: 1) a presenting complaint for which a checklist had been developed and 2) a priority of 2-4 according to the Medical Emergency Triage and Treatment System (METTS)[18]. Data was collected from the ED patient registration system "Patientliggaren" using the statistic program "QlikView". "Patientliggaren" contains basic patient information (such as sex, age, presenting complaint), patient times in the ED (for example total length of stay, time from arrival to first physician contact etc.) and whether the patients were admitted, discharged or left without being seen by a physician (LWBS). Information as to whether a checklist was used during patient management was entered separately in a data spreadsheet.

Time-efficiency was evaluated by extracting data relating to the following three time variables (formatted as hours:minutes): 1) modified length of stay (mLOS), defined in this study as time from patient arrival to discharge or the decision for admittance , 2) time to physician contact (TTPC), defined as the time from arrival until first physician contact and 3) physician's management time (PMT), defined as the time between the first physician contact to discharge/decision for admittance. The relationship between these measures are: mLOS = TTPC + PMT. Additionally, the proportion of the following variables were analyzed: 1) patients with mLOS <4 h, 2) patients with TTPC <1 h, 3) patients who left without being seen (LWBS) by a physician and 4) admittance to a ward.

Furthermore, every checklist contained a three-question evaluation to be filled in by the user. The questions were: 1) Did the checklist help you in the management of your patient? (graded 1-6, 1 being "not at all" and 6 being "very much"), 2) Do you want them to remain after the evaluation period? ("no", "no opinion" or "yes") and 3) What is your medical title?

Since the frequency of checklist usage during the study period was lower than expected, a smaller and more controlled substudy with the same inclusion criteria was conducted. Five

physicians (four junior physicians and one first year resident) working two consecutive weeks in the ED were identified. They were instructed to work during the first week without the use of checklists and during the second week using the checklists (non-optional). During these two weeks, time spent with the patient during the initial physician contact (a variable not present in "Patientliggaren") was measured by direct observation as a complement to the existing variables extracted from "Patientliggaren". In order to assess the quality of patient management, physicians were asked to grade the probability of 3-9 relevant diagnoses/syndromes (depending on the presenting complaint) and the probability for admittance using a simple grading system composed by the authors and dr E.D. (Appendix 2). These gradings (i.e. probability assessments) were conducted before and after consulting a senior physician.

For example, if a patient presented with transient loss of consciousness (TLoC) the physicians were asked to assess the probabilities of the diagnostic hypotheses cardiogenic syncope, seizure, orthostatic hypotension and reflex syncope on a scale of 1-5 (1- excluded, 2- low probability, 3- compatible with, 4- probable, 5- very probable). The sum of the differences between the assigned probabilities for each differential diagnosis before and after consulting a senior colleague was calculated. Thus if the physician assigned probabilities of 1, 3, 1 and 2 for the four diagnostic hypotheses before consulting a senior colleague, and probabilities of 2, 4, 1 and 2 after the consultation, the sum of the differences in probability would be (2-1) + (4-3) + (1-1) + (2-2) = 2 (example shown in Appendix 2). This sum of differences was used as an outcome measure. The same method was used for admittance probability, which was also assessed on a scale from 1 to 5 (1- no, 2- probably not, 3- maybe, 4- probably, 5- yes), and was compared to if the patient actually was admitted or not (5 or 1 respectively). Small absolute differences between gradings were taken to represent better initial probability assessment (0 being the optimum), using the assessment from a more experienced colleague as the "gold standard."

For statistical analysis, the program IBM SPSS Statistics 20 was used. A p-value <0.05 was considered statistically significant. For time variables, Mann Whitney U- test were used to calculate p-values since they did not follow a normal distribution curve. For variables consisting of proportions, Fisher's exact test were used.

Results

Study populations

The number of patients included during the control and evaluation periods were 4373 and 2770, respectively. During the evaluation period, checklists were properly used on merely 219 (7.9 %) patients. Figure 1 shows the patient distribution and the ways of comparison.

Table 1 shows baseline characteristics for the control and evaluation populations. The populations were comparable, with the only statistically significant difference being a higher median age in patients with dizziness in the evaluation population (p= 0.047). When comparing the baseline characteristics for the control population vs the cases in which checklists were used, there were no statistically significant differences (see table 2).

Control period vs evaluation period

Table 3 shows outcome measures (time variables, proportion admitted and proportion of LWBS) for the control and evaluation populations. There were no statistically significant differences when comparing all complaints together. Among patients with headache, there was a statistically significant shorter TTPC in the evaluation group (median time 2:32 vs 2:01, p= 0.02). Among patients with head injury, the proportion of patients with mLOS <4 h (48.0% vs 58.7%, p= 0.024) was higher in the evaluation group and so was the rate of admission (16.7% vs 25.4 %, p=0.026). Among patients with TLoC, there was a lower rate of admission in the evaluation group (56.1% vs 39.5%, p=0.023). There were no significant differences when comparing the other patient complaints.

Control period vs cases in which checklists were used

When comparing the control population with the evaluation subpopulation in which checklists were used, there were significant differences for all three time variables in favor of the control group (see Table 4). Median mLOS was 04:25 vs 05:01 (p<0.001). Median TTPC was 01:52 vs

02:20 (p=0.005). Finally, median PMT was 01:49 vs 02:14 (p=0.001). There were also significant differences in favor of the control group for the proportion of mLOS<4h (43.9% vs 33.8%, p=0.003) and the proportion of TTPC<1h (28.4% vs 16.1%, p<0.001).

When comparing the individual presenting complaints, only abdominal pain, chest pain and dyspnea were evaluated because of the low user frequencies of the other checklists. Of these, only dyspnea had statistically significant results, but the number of dyspnea checklists that were properly used was only 32. Median mLOS and PMT were both longer in the checklist group (04:15 vs 05:41, p<0.001 and 02:03 vs 03:40, p=0.001, respectively), and the proportion of mLOS<4h and proportion of TTPC<1h were both lower in the checklist group (47.2% vs 15.6%, p<0.001).

The substudy

The substudy compared the management of 25 patients without the use of checklists to the management of 36 patients with the use of checklists. There were no statistically significant differences between the groups when it comes to baseline data (Tables 5a and 5b).

There were no significant differences between populations in regards to the outcome variables used in the main study (i.e. time variables, proportion admitted and proportion of LWBS). There were no significant differences in time spent with the patient or probability assessment of admission. However, there was a significantly smaller median difference in probability assessment of diagnostic hypotheses before and after discussion with a senior physician (2.5 vs 1 in median difference, p=0.018), indicating an improvement in probability assessment of diagnoses when using checklists (Table 6).

User feedback

The total number of evaluations received was 135 (125 fully completed). As shown in diagram 1 a high percentage of the checklist evaluators were junior physicians (64.4%). The mean and median responses to the question “Did the checklist help you in the management of your

patient?" (1-6) were 3.35 and 3, respectively. To the question "Do you want them to remain after the evaluation period?" 58.3% answered "Yes", 7.1 % answered "No" and 34.6 % answered "No opinion". 82 evaluations were from junior physicians and 43 from licensed physicians. Junior physicians wanted the checklists to remain in 61.0 % of cases and did not want them to remain in 8.5% of cases. Licensed physicians wanted them to remain in 51.2% of cases and did not want them to remain in 4.7% of cases (Diagrams 2a, 2b and 2c).

Discussion

The main findings of this study were that checklists were used in only 7.9% of eligible patients during the evaluation period, that the use of checklists was associated with increased mLOS, TTPC, PMT, admission rate and a more junior physician title, and that the majority of physicians who used the checklists wanted them to remain after the study period. The main findings of the substudy were that the use of checklists did not affect patient management time but did improve the assessment of diagnostic probabilities.

Checklist usage

The most striking result of this study was the very low user rate of the checklists. Checklists were used in only 7.9% (n=219) of eligible patients, making the results of this study hard to interpret. This rate of use is much lower than expected and somewhat disappointing. Only the checklists that could properly be linked to a specific patient (i.e. marked with patient ID) were included and checklists without any signs of usage (i.e. checkmarks or handwritten notes) were excluded since it could not be proven that they were used. Some physicians used the checklists as a support without writing or putting checkmarks on them, suggesting that the actual user rate was greater than 7.9% (estimated to around 15-20%, based on weekly counts of handed-in checklists). Nevertheless, such a user rate remains poor.

There are several factors that affect the implementation of new guidelines and decision-making tools, such as this study's checklists.[33-35] In this study the information prior to the start of implementation was sparse, partly due to lack of time and partly due to the logistical hurdles when attempting to convey information to the large number of physicians who work in the ED. The assumption was that user-friendly and fairly self-instructing checklists would not require an in-depth introduction. However, the lack of information probably made it easier to forget using the checklists. Checklists were to be attached to the patient charts by the triage nurse. Since this was a new routine without any obvious benefit to the triage staff, it was easily forgotten. Likewise, some physicians felt a lack of incentive for checklist usage and argued that the checklists would be more beneficial if the gathered information could be transferred directly into the patient records, thereby reducing administrative time. If the checklists had been better

integrated into the workflow, they might have been more accessible to the clinician. Some physicians also expressed that the high work rate in the ED made it hard to learn a new concept of information gathering.

Control period vs evaluation period

When the control population was compared with the evaluation population as a whole (i.e. all presenting complaints combined), there were no statistically significant differences in time variables, rate of admission or proportion of LWBS. This result is unremarkable since the proper checklist usage rate was only 7.9%, and a dramatic checklist effect would be required to result in a detectable difference.

However, when specific complaints were examined, differences in management times were detected among patients presenting because of headache, head injury and TLoC (Table 3). It seems unlikely that these findings resulted from the use of checklists. On the one hand, the low user frequency of the checklists for these specific complaints (n= 13, 7 and 8, respectively) makes it unlikely that the results are attributable to checklist usage. On the other hand, some of the results argue against a checklist effect. For instance, TTPC among patients with headache was significantly lower in the evaluation population while mLOS was not affected. It is unlikely that the checklist would affect a time variable that starts and stops before the checklist is used, unless the other time variables in the ED are affected. Finally, due to the large number of analyses performed, some differences will be statistically significant by chance alone.

Control period vs cases in which checklist were used

When comparing the control population with the population in which checklists were used, there were significant differences in favor of the control population for the outcome measures mLOS, TTPC, PMT, proportion of mLOS<4h and proportion of TTPC<1h. No significant difference was seen for admission rates. Without further analysis, such results suggest that checklist usage negatively impacts efficiency in the ED. However, junior physicians constituted a large proportion of checklist users (64.4% according to the user feedback, Diagram 1), compared to

normal (approximately 41.4 %, when calculating the title proportion for three random days from each period). Junior physicians are generally slower, partly because of inexperience and partly because they, at this particular ED, are obliged to discuss every case with a senior physician before making any major decisions. Waiting for this consultation can sometimes be lengthy. Such confounders caution against a hasty conclusion that checklists impair time-efficiency.

When subdividing the cases according to presenting complaint, only dyspnea had in itself statistically significant results; mLOS and PMT were both significantly longer in the checklist population while the proportion of mLOS<4h and the proportion of TTPC<1h were both lower in the checklist population (Table 4). Additionally, there was a trend towards lower admission rate in the checklist group (54.8 % vs 37.5 %, p=0.069). Although not statistically significant, this could mean that instead of investigating the patients while they were hospitalized, the complete diagnostic work-up was done in the ED. This could mean that even if efficiency was impaired, it resulted in fewer unnecessary admissions. However, further studies are required.

Aside from dyspnea, chest pain and abdominal pain, the complaint-specific rates of checklist usage were too low for a meaningful analysis. No differences in time-efficiency were found for either abdominal or chest pain. This could indicate that the dyspnea checklist needs improvement, but could also indicate that it was the only checklist that actually had an impact.

For the completeness of the evaluation, the population in which checklists were used and the rest of the evaluation population were compared. The results were virtually identical to those found when the control population was compared with the population in which checklists were used. This finding is not surprising since the control population and the evaluation population in total were similar regarding both baseline and outcome measures (and that no checklist was used in 92.1% of the evaluation population).

The substudy

The substudy was conducted to palliate for a low checklist usage rate. The substudy's design provided an opportunity to evaluate whether the checklists actually made patient management

less efficient, adjusted for the uneven distribution of junior physicians when comparing cases in which checklists were used versus the control and the evaluation populations. During the substudy, only fairly inexperienced physicians were included, which made the results less confounded by physician experience but also less generalizable. The results of this substudy showed that time-efficiency, measured as mLOS, proportion of mLOS<4h, TTPC, proportion of TTPC<1h and PMT, was not impaired by checklist usage. In fact, all outcome variables were similar to or better in the checklists group (Table 6) although the differences were not statistically significant. This challenges the results from the main study and is of import for potential future studies.

Probability assessment (measured as the sum of the differences between assigned probabilities for diagnostic hypotheses prior to and after consultation with a senior physician) showed a statistically significant difference favoring checklist usage. This result could indicate that checklists improved physicians' differential diagnosis abilities, since a smaller absolute difference would mean that the initial assessment better corresponds to the assessment of a more experienced physician. However, this difference was only 1.5 steps on the scale. For example, this could be the difference from "probable" to "low probability"/"compatible with" for a single diagnose/syndrome (and depending on the presenting complaint there were 3-9 diagnoses/syndromes to evaluate). The clinical relevance of this result can therefore be questioned, but is nevertheless an improvement and therefore it shows promise for further studies with larger groups. However, due to the large number of analyses performed and the low number of used checklists, this result could be a chance occurrence. There were no differences regarding admission assessment (i.e. checklists did not alter the physicians ability to judge whether admission to a ward was necessary).

There were some limitations. Since the substudy was conducted after the implementation of the checklists, four of five physicians had already had the opportunity to use them before the initiation of the substudy. Although rather unlikely, this could have biased the result at the disadvantage of the checklists since the physicians might already have learned their content, therefore contaminating the "non-checklist"-group. Additionally, the participating physicians might have become more familiar with the probability scoring system the more times they used it, which could have lead to a better probability assessment in the checklist week due to "learning

the system". However, one of the aims with the checklists was to teach physicians to think in terms of probability and likelihood ratios. Additionally, one can question whether the use of a senior colleague's probability assessment as a "gold standard" is adequate. Preferably, other measures such as rate of diagnostic errors should have been used. This would however require larger study groups and access to patient medical records. Another limitation is the low number of cases (n=25 and 36 in total) which prompted the analysis of all presenting complaints in the same group, despite the fact that there were a somewhat unequal distribution of presenting complaints between the groups. For example, there were no cases of neurological deficits in the group without checklists, and six in the group with checklists (table 5b). The numbers of cases were however to low to use in statistical calculations. Also, since the physicians were observed there could potentially be a Hawthorne effect (i.e. an improvement caused by the simple fact of being observed). Since they were observed during both weeks, it is unlikely that it would affect the result significantly.

Ideally, the substudy should have extended over a longer period, excluding at least one week at the start of the first period (allowing the physicians to adapt to the probability scoring system) and at least one week at the start of the second period (allowing physicians to properly be introduced to the use of checklists and their reference documents).

User feedback

The average grade on the question "Did the checklist help you in the management of your patient?" was 3.35 out of 6, suggesting a moderate perceived benefit. When evaluating this result, one has to bear in mind that many cases are pretty straight-forward (especially for experienced physicians). The way the question was formulated, the physician could feel that the use of a checklist did not help in a specific case, although appreciating having it as a support. This is reflected in the fact that 58.3% of the users wanted the checklists to remain after the evaluation period, and only 7.1% did not.

When comparing the feedback from junior physicians and licensed physicians, junior physicians were more positive to the future existence of checklists after the evaluation period than the licensed physicians (61.0 vs 51.2 %). This is not surprising since junior physicians probably

benefit slightly more from a diagnostic tool, although more experienced physicians also can be interested in diagnostic aids.

One limitation is that physicians who did not like the checklists only used them a few times or not at all, thus leaving very little or no user feedback. These results are therefore biased. Instead of having the user feedback on the checklists, surveys could have been performed using separate forms. The current evaluation method was used, since a higher user rate was initially expected (which probably would have given a less biased result).

Outcome measures

Measuring quality in the ED is difficult, which is why it is mainly measured in time variables. As presented in an article by Heyworth, the National Health Service in the United Kingdom has defined several quality indicators.[36] Of these, "Total time in the ED" (in this study excluding the time waiting for an in-hospital bed), "LWBS" and "time to seeing a decision-making clinician" are used in this study.

Four hours is a cut-off commonly used for evaluation of total time in the ED, known as the "four hour target", although it has been debated.[9,36] However, it was used as an outcome measure in this study, as the proportion of mLOS<4h.

LWBS was used in this study only for the control period vs the whole evaluation period, since all patients that checklists had been used on by definition had met a physician, thereby making this variable useless when comparison included the subpopulation in which checklists were used. LWBS can not directly be affected by checklists, but is still useful when comparing the two whole groups since an effect on the time variables indirectly can alter the rate of LWBS.

Also, time to physician contact (TTPC) was used. A cut-off of one hour was used (proportion of TTPC<1h), since an indicator in the United Kingdom is that this median time should not be above 60 minutes.[36] In this study median TTPC was well over 1 h, however priority 1 patients (who are seen immediately at arrival) were not included which would have lowered the median. Whether this is a relevant outcome measure or not can be debated. As with LWBS this outcome measure cannot be affected directly by checklists, but indirectly by an effect on the other time

variables. However, since time to treatment initiation often is crucial[2,3] this is a highly relevant variable for patient safety. The variable "time to initiation of treatment" would have been of more interest than TTPC, but unfortunately it was not available in "Patientliggaren".

Additionally, the variable PMT (i.e. time from meeting a physician to being discharged/put on queue for admission) was used since it reflects the time frame in which checklists most likely would exert their effect. This is perhaps the most interesting variable for evaluating the effect of the checklists. This is however not an optimal measure of efficiency, since PMT is affected by time spent waiting for consultations and investigations (e.g. an x-ray). Additionally, physicians can manage several patients at a time which could mean that even if the median PMT for one specific physician is long, he or she can still manage many patients during a shift and thereby be efficient. An outcome measure in future studies could for instance be the average number of patients a physician manages during a shift.

Finally, the rate of admission to a ward was used. The results of this is however hard to evaluate, since both an increase and a decline in admission rate could be due to a checklist effect. Checklists could give confidence to discharge more patients safely, but also lead to admitting more patients if finding additional diagnoses that otherwise would have been missed.

The outcome measures used are mostly based on time variables. Ideally, other quality measures such as complication rate and rate of diagnostic errors should have been used. In a sufficiently powered future study, mortality would also be of high interest. Although checklists have the potential of increasing time-efficiency, their major effect in the long run probably lies in reducing diagnostic errors. However in this study setting it was not feasible to use more quality-orientated measures since it would have required time-consuming reviews of patient medical records.

Checklists in other studies

Pronovost et al showed in their study on use of a checklist in central venous catheterising, conducted in 103 ICUs, that the use of a checklist significantly decreased the level of bloodstream infections from a median of 2.7 per 1000 catheter days to a median of 0 after 3 months, which was sustained during the whole study period of 18 months.[27] These remarkable results aroused the interest of checklists in other medical areas.[26] One example is the study by Haynes et al on the WHO surgical safety checklist, which reduced death rate from surgery from 1.5 to 0.8 % and reduced surgical complications from 11 to 7 %.[28] One strength of this study was its generalizability, since it was conducted at eight sites worldwide, in both urban and rural hospitals. The same authors, with Weiser as main author, showed that this beneficial effect also applied urgent cases[29] and a study by de Vries et al showed beneficial results from a similar surgical safety checklist.[30] Renzi et al showed that a paper stamp checklist for asthma improved primary care physician knowledge of guidelines and reduced number of hospital visits. However there were no effects on patient outcomes.[31] Another study by Wolff et al showed that checklists and reminders could improve the adherence to important guidelines for patients with stroke and acute myocardial infarction.[32]

Although these studies show favourable effects of checklists for medical procedures or specific diagnoses, none of them focuses on checklists as a diagnostic tool. Ely and colleagues raises the subject of the use of checklists in this area.[33] They present three different types of diagnostic checklists, one of which is a differential diagnosis checklist. It lists virtually all differential diagnoses to consider in patients with a specific presenting complaint. For example, the checklist for sinus tachycardia consists of a list of 29 diagnoses to consider. Although developing checklists for 46 complaints and revising them during two years of clinical use, they have not been formally evaluated. This type of checklist is probably useful in view of reducing diagnostic errors (e.g. reducing premature closure), but seems time consuming and prone to cause unnecessary investigations.

Checklists in this study

The checklists used in this study have a different structure than the ones described in the study by Ely et al.[33] Rather than listing differential diagnoses, they consist of important questions in the medical history uptake and important parts of the physical examination, as well as other relevant investigations to consider. The content is sought to be in line with Bayes Theorem and the use of likelihood ratios (LR) [37,38], a concept that is used (both consciously and unconsciously) in the ED, where probability assessments are central in the diagnostic process. The aim in the development of the checklists for this study was to mainly include questions and investigations that had a clinical relevance in affecting the probability of differential diagnoses (e.g. a sufficiently high or low LR). For example, rectal examination was not a part of the routine physical examination for patients with abdominal pain, since performing this examination routinely on all abdominal pain patients has doubtful evidence[39,40]. One large meta-analysis on showed no clinical relevance of the exam when suspecting appendicitis [41]. However, when considering certain diagnoses (e.g. bowel obstruction), the checklist prompted its use.

The checklists were open for continuous revision, and underwent some layout changes and minor content changes during the evaluation period. Although this is not ideal when evaluating their effect, one of the major points of these checklists was their continuous improvement and update based on suggestions from the users.

These checklists were created for the complex environment of the ED, and are therefore different than the better studied checklists for specific diagnoses and medical procedures. A problem can be that patients can present with multiple presenting complaints at the same time, making it difficult for the physician to know which checklist to use. Additionally, some physicians found the checklists to be restrictive in the medical history uptake on elderly, demented or confused patients since these patients may have trouble answering direct questions. On the other hand, one might find the structure they bring reassuring in difficult cases.

Comments on checklists in general

The idea that a simple checklist on a paper could cause the major improvements shown by, for instance, Pronovost et al[27] and Haynes et al[28], is appealing, but clearly not the only factor that has contributed to their success in previous studies. As discussed in a commentary by Bosk et al (with Pronovost as co-author)[42]: "*The mistake of the 'simple checklist' story is in the assumption that a technical solution (checklists) can solve an adaptive (sociocultural) problem*". This means that only a checklist is not sufficient. It requires a general mentality of safety thinking and teamwork that is not easily achieved. Therefore only the implementation of a checklist might not be enough. The concept has to be well-anchored and accepted among all personnel categories before a major impact can be expected. This may for instance require a better implementation strategy than the one used in this study.

The concept of checklists has been subject to a lot of skepticism.[26] Physicians may feel that a checklist would not be beneficial since they already consider themselves to have a sufficient level of knowledge and experience, thus not needing additional clinical tools. In some cases, this can be a dangerous attitude, since overconfidence contributes to diagnostic errors.[7]

One might also be worried that checklists might turn physicians into "robots", mindlessly following instructions and neglecting the individual characteristics of patients. Similar concerns were raised when clinical algorithms came into practice,[33] although they are today widely used. Checklists are tools, not a replacement of the physician. The checklists in this particular study are for aid in the information gathering, the interpretation of the information is up to the individual physician. When using a checklist, the physician must (as always) adapt to the specific situation.

As discussed in the study by Ely et al, pilots have learned not to rely on their memories. For them, checklists in all aspects of their work are mandatory. If future studies can prove beneficial effects of checklists as a diagnostic tool, then the question should be raised whether they should be mandatory in medicine as well. As put by Ely et al, "*pilots no longer feel insulted when reminded by their copilots to release the elevator locks*".

Conclusion

In this study, the implementation of checklists as a diagnostic tool in the ED did not increase efficiency. On the contrary, their use seemed to increase the time patients spent in the ED. However, since a major confounder altered the outcome (i.e. a higher proportion of inexperienced physicians as checklist users) these results have little validity. The more controlled substudy (adjusted for level of inexperience) showed no negative influence on efficiency when using checklists, and showed promising results regarding probability assessment of differential diagnoses, although with unclear clinical relevance. The generalizability of the substudy can be questioned, since it mostly consisted of junior physicians.

Nonetheless, the results of this pilot study warrant further investigation in a more carefully prepared setting. One should take into consideration the limitations and confounders of this pilot study when developing future study designs. Emphasis should also be put on the pre-implementation information, thereby gaining approval of the concept of diagnostic aids. Since the results from the substudy showed promising results, a similar study design in a larger scale might be of interest in the future. Quality measures such as complication rate and rate of diagnostic errors should also be evaluated. If beneficial effects are found, the generalizability should be assessed through implementation in other EDs.

Acknowledgements

The authors would like to extend thanks to Eric Dryver, Ulf Ekelund, Arash Mokhtari Karchegani, Christina Möllerstedt, Ing-Marie Nilsson, Erik Uddman, Martin Galeano, Joakim Olofsson and the five physicians who participated in the substudy. Thank you for your help.

References

1. Glatter RD, Martin RE, Lex J. How Emergency Physicians Think - Highlights of the Fourth Mediterranean Emergency Medicine Congress (MEMC IV)

http://www.hsi.gatech.edu/erfuture/images/f/fc/HowEPsThink.pdf. 2008. Reviewed: May 15th 2012.

2. Gaieski DF, Mikkelsen ME, Band RA, et al. Impact of time to antibiotics on survival in patients with severe sepsis or septic shock in whom early goal-directed therapy was initiated in the emergency department. *Crit Care Med.* Apr 2010;38(4):1045-1053.
3. Hacke W, Kaste M, Bluhmki E, et al. Thrombolysis with alteplase 3 to 4.5 hours after acute ischemic stroke. *N Engl J Med.* Sep 2008;359(13):1317-1329.
4. Newman-Toker DE, Pronovost PJ. Diagnostic errors--the next frontier for patient safety. *JAMA.* Mar 2009;301(10):1060-1062.
5. Graber ML, Franklin N, Gordon R. Diagnostic error in internal medicine. *Arch Intern Med.* Jul 2005;165(13):1493-1499.
6. Kachalia A, Gandhi TK, Puopolo AL, et al. Missed and delayed diagnoses in the emergency department: a study of closed malpractice claims from 4 liability insurers. *Ann Emerg Med.* Feb 2007;49(2):196-205.
7. Berner ES, Graber ML. Overconfidence as a cause of diagnostic error in medicine. *Am J Med.* May 2008;121(5 Suppl):S2-23.
8. Shojania KG, Burton EC, McDonald KM, Goldman L. Changes in rates of autopsy-detected diagnostic errors over time: a systematic review. *JAMA.* Jun 2003;289(21):2849-2856.
9. Higginson I. Emergency department crowding. *Emerg Med J.* Jan 2012.
10. Guttmann A, Schull MJ, Vermeulen MJ, Stukel TA. Association between waiting times and short term mortality and hospital admission after departure from emergency department: population based cohort study from Ontario, Canada. *BMJ.* 2011;342:d2983.
11. Richardson DB. Increase in patient mortality at 10 days associated with emergency department overcrowding. *Med J Aust.* Mar 2006;184(5):213-216.
12. Epstein SK, Huckins DS, Liu SW, et al. Emergency department crowding and risk of preventable medical errors. *Intern Emerg Med.* Apr 2012;7(2):173-180.
13. Pines JM, Pollack CV, Diercks DB, Chang AM, Shofer FS, Hollander JE. The association between emergency department crowding and adverse cardiovascular outcomes in patients with chest pain. *Acad Emerg Med.* Jul 2009;16(7):617-625.
14. Morris ZS, Boyle A, Beniuk K, Robinson S. Emergency department crowding: towards an agenda for evidence-based intervention. *Emerg Med J.* Jun 2011.
15. Sprivulis PC, Da Silva JA, Jacobs IG, Frazer AR, Jelinek GA. The association between hospital overcrowding and mortality among patients admitted via Western Australian emergency departments. *Med J Aust.* Mar 2006;184(5):208-212.
16. Diercks DB, Roe MT, Chen AY, et al. Prolonged emergency department stays of non-ST-segment-elevation myocardial infarction patients are associated with worse adherence to the American College of Cardiology/American Heart Association guidelines for management and increased adverse events. *Ann Emerg Med.* Nov 2007;50(5):489-496.
17. Vieth TL, Rhodes KV. The effect of crowding on access and quality in an academic ED. *Am J Emerg Med.* Nov 2006;24(7):787-794.
18. Widgren BR, Jourak M. Medical Emergency Triage and Treatment System (METTS): a new protocol in primary triage and secondary priority decision in emergency medicine. *J Emerg Med.* Jun 2011;40(6):623-628.
19. Farrokhnia N, Göransson KE. Swedish emergency department triage and interventions for improved patient flows: a national update. *Scand J Trauma Resusc Emerg Med.* 2011;19:72.
20. Kinsman L, Champion R, Lee G, et al. Assessing the impact of streaming in a regional emergency department. *Emerg Med Australas.* Jun 2008;20(3):221-227.

21. Olsson LE, Karlsson J, Ekman I. The integrated care pathway reduced the number of hospital days by half: a prospective comparative study of patients with acute hip fracture. *J Orthop Surg Res.* 2006;1:3.

22. Read SJ, Levy J. Effects of care pathways on stroke care practices at regional hospitals. *Intern Med J.* Oct 2006;36(10):638-642.

23. Cunningham S, Logan C, Lockerbie L, Dunn MJ, McMurray A, Prescott RJ. Effect of an integrated care pathway on acute asthma/wheeze in children attending hospital: cluster randomized trial. *J Pediatr.* Mar 2008;152(3):315-320.

24. Waller SL, Delaney S, Strachan MW. Does an integrated care pathway enhance the management of diabetic ketoacidosis? *Diabet Med.* Apr 2007;24(4):359-363.

25. Pettie JM, Dow MA, Sandilands EA, Thanacoody HK, Bateman DN. An integrated care pathway improves the management of paracetamol poisoning. *Emerg Med J.* May 2011.

26. Gawande A. *The Checklist Manifesto - How To Get Things Right.* 2010.

27. Pronovost P, Needham D, Berenholtz S, et al. An intervention to decrease catheter-related bloodstream infections in the ICU. *N Engl J Med.* Dec 2006;355(26):2725-2732.

28. Haynes AB, Weiser TG, Berry WR, et al. A surgical safety checklist to reduce morbidity and mortality in a global population. *N Engl J Med.* Jan 2009;360(5):491-499.

29. Weiser TG, Haynes AB, Dziekan G, et al. Effect of a 19-item surgical safety checklist during urgent operations in a global patient population. *Ann Surg.* May 2010;251(5):976-980.

30. de Vries EN, Prins HA, Crolla RM, et al. Effect of a comprehensive surgical safety system on patient outcomes. *N Engl J Med.* Nov 2010;363(20):1928-1937.

31. Renzi PM, Ghezzo H, Goulet S, Dorval E, Thivierge RL. Paper stamp checklist tool enhances asthma guidelines knowledge and implementation by primary care physicians. *Can Respir J.* 2006 May-Jun 2006;13(4):193-197.

32. Wolff AM, Taylor SA, McCabe JF. Using checklists and reminders in clinical pathways to improve hospital inpatient care. *Med J Aust.* Oct 2004;181(8):428-431.

33. Ely JW, Graber ML, Croskerry P. Checklists to reduce diagnostic errors. *Acad Med.* Mar 2011;86(3):307-313.

34. Harrison MB, Légaré F, Graham ID, Fervers B. Adapting clinical practice guidelines to local context and assessing barriers to their use. *CMAJ.* Feb 2010;182(2):E78-84.

35. Légaré F, Ratté S, Gravel K, Graham ID. Barriers and facilitators to implementing shared decision-making in clinical practice: update of a systematic review of health professionals' perceptions. *Patient Educ Couns.* Dec 2008;73(3):526-535.

36. Heyworth J. Emergency medicine-quality indicators: the United Kingdom perspective. *Acad Emerg Med.* Dec 2011;18(12):1239-1241.

37. Gallagher EJ. Clinical utility of likelihood ratios. *Ann Emerg Med.* Mar 1998;31(3):391-397.

38. McGee S. Simplifying likelihood ratios. *J Gen Intern Med.* Aug 2002;17(8):646-649.

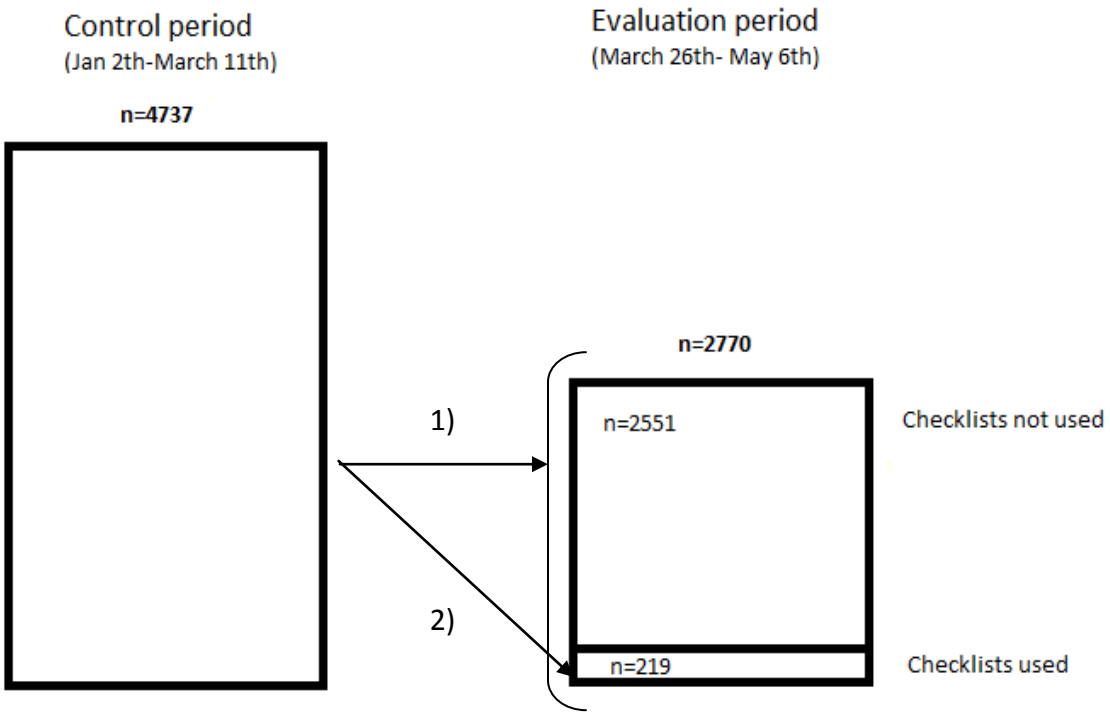
39. O'Brien M. *Chapter 74 Acute abdominal pain. In: Tintinalli JE, Stapczynski JS, Cline DM, Ma OJ, Cydulka RK, Meckler GD. editors. Emergency medicine: A comprehensive study guide. 7th ed. New York: McGraw-Hill.* 2011.

40. Quaas J, Lanigan M, Newman D, McOsker J, Babayev R, Mason C. Utility of the digital rectal examination in the evaluation of undifferentiated abdominal pain. *Am J Emerg Med.* Nov 2009;27(9):1125-1129.

41. Andersson RE. Meta-analysis of the clinical and laboratory diagnosis of appendicitis. *Br J Surg.* Jan 2004;91(1):28-37.

42. Bosk CL, Dixon-Woods M, Goeschel CA, Pronovost PJ. Reality check for checklists. *Lancet.* Aug 2009;374(9688):444-445.

Figure 1. Patient distribution and ways of comparison



1) Comparison between the control period and the whole evaluation period

2) Comparison between the control period and cases in which checklists were properly used

Table 1. Baseline characteristics for the control period and the evaluation period

Complaint	Visits		Visits per day		Median age		Female sex		Prio 2		Prio 3		Prio 4	
	Control	Evaluation	Control	Evaluation	Control	Evaluation	Control	Evaluation	Control	Evaluation	Control	Evaluation	Control	Evaluation
	Numbers of				*Years*						*Percent*			
Abd. Pain	1421	949	20.3	22.6	46 (30-66)	45 (30-63)	60.3	59.6	20.3	21.3	73.0	72.6	6.7	6.1
Allergy	130	73	1.9	1.7	38 (25-54)	38 (23-54)	66.9	71.2	51.5	43.8	41.5	53.4	6.9	2.7
Chest pain	895	546	12.8	13.0	59 (43-72)	58 (45-71)	44.8	46.9	54.4	53.1	42.6	45.2	3.0	1.6
Dizziness	261	143	3.7	3.4	**60 (43-76)**	**66 (50-77)**	56.7	57.3	28.4	21.7	65.1	67.1	6.5	11.2
Dyspnea	672	416	9.6	9.9	70 (48-81)	70 (53-83)	55.1	54.6	55.5	53.1	40.9	44.7	3.6	2.2
Headache	242	145	3.5	3.5	39 (31-57)	44 (31-54)	59.9	66.9	29.3	36.6	56.2	48.3	14.5	15.2
Head injury	281	189	4.0	4.5	49 (24-75)	45 (24-75)	55.2	46.0	27.4	36.0	61.9	56.6	10.7	7.4
Neur. Deficit	348	228	5.0	5.4	68 (50-80)	69 (53-81)	50.6	46.9	51.4	47.8	45.7	48.7	2.9	3.5
TLoC	123	81	1.8	1.9	65 (43-79)	67 (43-83)	54.5	59.3	44.7	43.2	48.0	49.4	7.3	7.4
TOTAL	4373	2770	62.5	66.0	56 (36-73)	55 (36-73)	55.0	55.1	38.2	37.6	55.9	57.2	5.9	5.2
p-value					0.865		0.981		0.582		0.281		0.246	

P-values are shown for the comparison of the total values of the control period and the evaluation period. Bold types indicates values that were statistically significant ($p<0.05$). Median age is presented with interquartile range within parentheses. TLoC= transient loss of consciousness.

Table 2. Baseline characteristics for the control period and the cases in which checklists were used

Complaint	Visits		Median age		Female sex		Prio 2		Prio 3		Prio 4	
	Control	Checklists	Control	Checklists	Control	Checklists	Control	Checklists	Control	Checklists	Control	Checklists
	Numbers of		*Years*						*Percent*			
Abd. pain	1421	85	46 (30-66)	41 (25-60)	60.3	57.6	20.3	22.4	73.0	72.9	6.7	4.7
Allergy	130	2	-	-	-	-	-	-	-	-	-	-
Chest pain	895	44	59 (43-72)	60 (47-72)	44.8	52.3	54.4	45.5	42.6	54.5	3.0	0.0
Dizziness	261	10	-	-	-	-	-	-	-	-	-	-
Dyspnea	672	32	70 (48-81)	63 (48-80)	55.1	59.4	55.5	40.6	40.9	56.2	3.6	3.1
Headache	242	13	-	-	-	-	-	-	-	-	-	-
Head injury	281	7	-	-	-	-	-	-	-	-	-	-
Neur. Deficit	348	18	-	-	-	-	-	-	-	-	-	-
TLoC	123	8	-	-	-	-	-	-	-	-	-	-
TOTAL	4373	219	56 (36-73)	54 (36-73)	55.0	57.5	38.2	32.0	55.9	61.6	5.9	6.4
p-value			0.773		0.487		0.064		0.108		0.768	

P-values are shown for the comparison of the total values of the control period and the cases in which checklists were used. Median age is presented with interquartile range within parentheses. TLoC= transient loss of consciousness. Data is not shown for presenting complaints with n<20, since statistical analyses were not feasible.

Table 3. Outcomes for the control period and the evaluation period

Complaint	Visits		Median mLOS		Proportion of mLOS < 4h		Median TTPC		Proportion of TTPC< 1h		Median PMT		Proportion of LWBS		Rate of admission	
	Control	Evaluation	Control	Evaluation	Control	Evaluation	Control	Evaluation	Control	Evaluation	Control	Evaluation	Control	Evaluation	Control	Evaluation
	Numbers of		*hh:mm*		*Percent*		*hh:mm*		*Percent*		*hh:mm*		*Percent*			
Abd. pain	1421	949	04:46	04:43	38.4	39.4	02:16	02:21	19.1	20.5	01:40	01:40	4.4	3.7	33.8	31.4
Allergy	130	73	03:25	03:05	64.6	64.4	00:51	00:41	56.3	64.2	02:00	01:49	3.1	4.1	7.7	4.1
Chest pain	895	546	03:56	04:00	51.2	49.8	02:38	01:36	31.3	29.3	01:39	01:32	3.0	4.2	36.0	39.4
Dizziness	261	143	05:17	05:22	36.8	30.1	02:13	02:23	25.8	17.5	02:06	02:11	3.8	2.8	42.1	39.9
Dyspnea	672	416	04:15	04:30	47.2	41.6	01:39	01:39	34.9	33.8	02:03	02:12	1.8	2.4	54.8	54.8
Headache	242	145	05:46	05:29	25.6	31.0	**02:32**	**02:01**	16.0	21.8	02:36	02:44	7.4	5.5	15.3	17.9
Head injury	281	189	04:08	03:30	**48.0**	**58.7**	01:19	01:21	38.3	38.5	01:56	01:50	4.3	3.2	**16.7**	**25.4**
Neur. Deficit	348	228	04:01	04:28	49.7	42.5	01:37	01:49	37.5	35.9	01:53	01:56	2.0	1.8	59.2	63.4
TLoC	123	81	04:46	04:23	39.0	48.1	02:04	01:48	29.1	33.3	02:01	01:37	4.1	4.9	**56.1**	**39.5**
TOTAL	4373	2770	04:25	04:25	43.9	43.4	01:52	01:52	28.4	28.2	01:49	01:50	3.6	3.5	37.7	37.9
p-value			0.990		0.677		0.890		0.890		0.800		0.896		0.861	

P-values are shown for the comparison of the total values of the control period and the evaluation period. Bold types indicates values that were statistically significant ($p<0.05$). mLOS = modified length of stay, TTPC= time to physician contact, PMT= physician management time, LWBS= left without being seen. TLoC= transient loss of consciousness. Data is not shown for presenting complaints with $n<20$, since statistical analyses were not feasible.

Table 4. Outcomes for the control period and the cases in which checklists were used

Complaint	Visits		Median mLOS		Proportion of mLOS < 4h		Median TTPC		Proportion of TTPC< 1h		Median PMT		Rate of admission	
	Control	Checklists	Control	Checklists	Control	Checklists	Control	Checklists	Control	Checklists	Control	Checklists	Control	Checklists
	Numbers of		*hh:mm*		*Percent*		*hh:mm*		*Percent*		*hh:mm*		*Percent*	
Abd. pain	1421	85	04:46	04:52	38.4	35.3	02:16	02:48	19.1	15.5	01:40	01:46	33.8	28.2
Allergy	130	2	-	-	-	-	-	-	-	-	-	-	-	-
Chest pain	895	44	03:56	03:43	51.2	52.3	02:38	01:37	31.3	22.7	01:39	01:41	36.0	36.4
Dizziness	261	10	-	-	-	-	-	-	-	-	-	-	-	-
Dyspnea	672	32	**04:15**	**05:41***	**47.2**	**15.6***	01:39	01:56	**34.9**	**15.6**	**02:03**	**03:40**	54.8	37.5
Headache	242	13	-	-	-	-	**-**	**-**	-	-	-	-	-	-
Head injury	281	7	-	-	**-**	**-**	-	-	-	-	-	-	**-**	**-**
Neur. Deficit	348	18	-	-	-	-	-	-	-	-	-	-	-	-
TLoC	123	8	-	-	-	-	-	-	-	-	-	-	**-**	**-**
TOTAL	4373	219	**04:25**	**05:01***	**43.9**	**33.8**	**01:52**	**02:20**	**28.4**	**16.1***	**01:49**	**02:14**	37.7	33.8
p-value			<0.001		0.003		0.005		<0.001		0.001		0.253	

P-values are shown for the comparison of the total values of the control period and the evaluation period. Bold types indicates values that were statistically significant ($p<0.05$), * indicates $p<0.001$. mLOS= modified length of stay, TTPC= time to physician contact, PMT= physician management time, LWBS= left without being seen. TLoC= transient loss of consciousness.

Table 5a. Baseline characteristics for the substudy

	Visits	Median age	Female sex	Prio 2	Prio 3	Prio 4
	Numbers of	*Years*		*Percent*		
Checklists not used	25	64 (43-76)	52	48	52	0
Checklists used	36	71 (42-80)	53	42	58	0
p-values		0.56	1.00	0.79	0.79	-

Median age is presented with interquartile range within parentheses.

Table 5b. Complaint distribution for the substudy

	TOTAL	Abd. pain	Allergy	Chest pain	Dizziness	Dyspnea	Headache	Head injury	Neur. Deficit	TLoC
Checklists not used	25	9	0	7	2	4	4	0	0	1
Checklists used	36	13	0	8	5	4	1	0	6	1

TLoC= transient loss of consciousness.

Table 6. Outcomes for the substudy

	Visits	Median mLOS	Proportion of mLOS < 4h	Median TTPC	Proportion of TTPC< 1h	Median PMT	Rate of admission	Median time spent with patient	DDx assessment	Admission assessment
	Numbers of	*hh:mm*	*Percent*	*hh:mm*	*Percent*	*hh:mm*	*Percent*	*hh:mm*	*Median absolute difference*	
Checklists not used	25	06:09	16.0	02:35	8.0	02:48	40.0	00:30	**2.5**	1
Checklists used	36	05:28	25.0	02:36	20.0	02:43	44.4	00:24	**1**	1
p-values		0.352	0.530	0.770	0.281	0.605	0.796	0.392	**0.018**	0.719

Bold types indicates values that were statistically significant ($p<0.05$). mLOS= modified length of stay, TTPC= time to physician contact, PMT= physician management time, DDx= differential diagnoses

Diagram 1. Title distribution of checklist evaluators

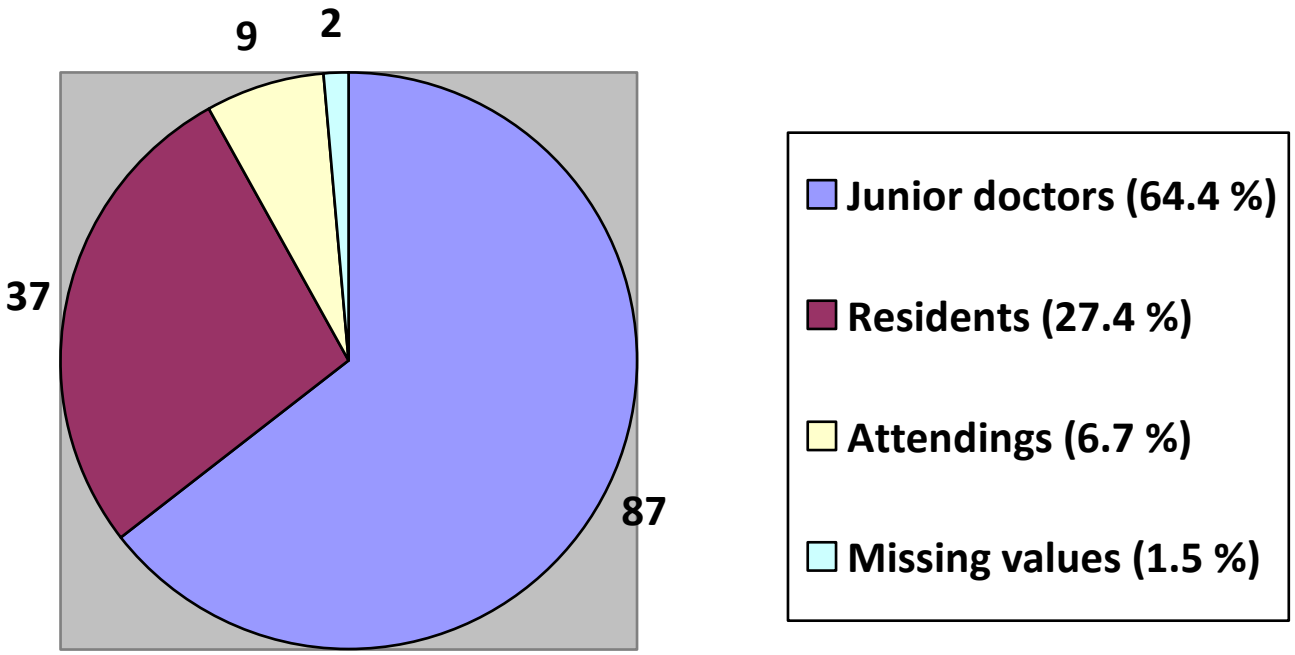


Diagram 2. "Do you want the checklists to remain after the evaluation period?"

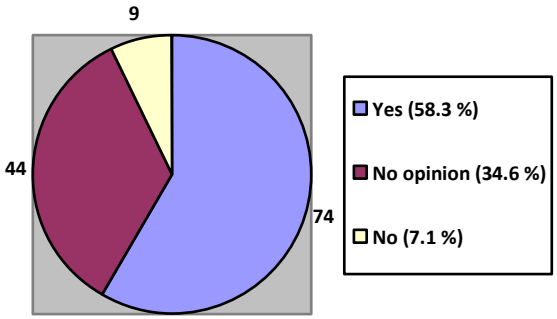


Diagram 2a. Total

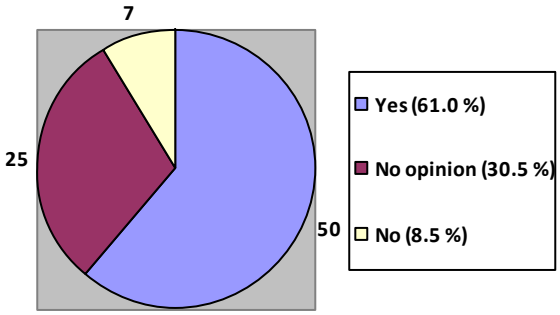


Diagram 2b. Junior physicians

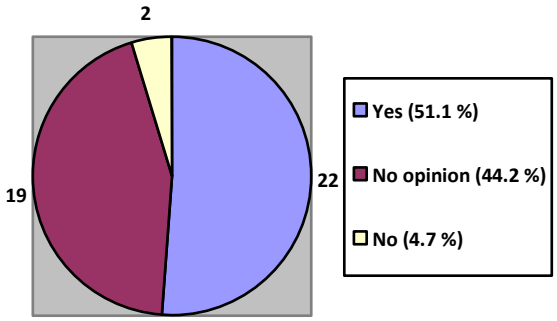
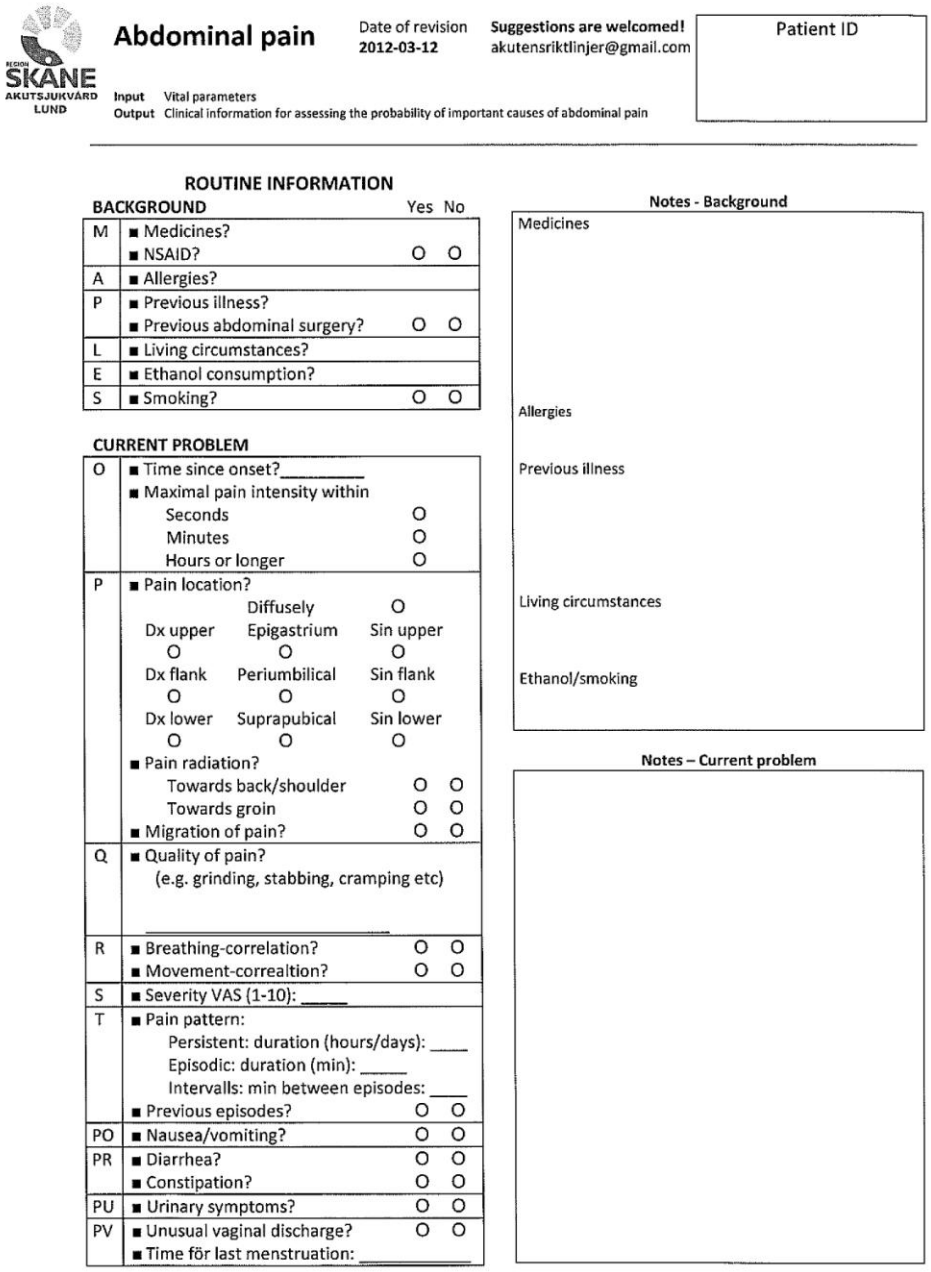


Diagram 2c. Licensed physicians

Appendix 1a. Example of the abdominal pain checklist. Page 1



Abdominal pain

Date of revision 2012-03-12

Suggestions are welcomed! akutensriktlinjer@gmail.com

Patient ID

Input Vital parameters

Output Clinical information for assessing the probability of important causes of abdominal pain

ROUTINE INFORMATION

	BACKGROUND	Yes	No
M	■ Medicines?		
	■ NSAID?	O	O
A	■ Allergies?		
P	■ Previous illness?		
	■ Previous abdominal surgery?	O	O
L	■ Living circumstances?		
E	■ Ethanol consumption?		
S	■ Smoking?	O	O

CURRENT PROBLEM

O	■ Time since onset?_____			
	■ Maximal pain intensity within			
	Seconds		O	
	Minutes		O	
	Hours or longer		O	
P	■ Pain location?			
		Diffusely	O	
	Dx upper	Epigastrium	Sin upper	
	O	O	O	
	Dx flank	Periumbilical	Sin flank	
	O	O	O	
	Dx lower	Suprapubical	Sin lower	
	O	O	O	
	■ Pain radiation?			
	Towards back/shoulder		O	O
	Towards groin		O	O
	■ Migration of pain?		O	O
Q	■ Quality of pain? (e.g. grinding, stabbing, cramping etc) _____			
R	■ Breathing-correlation?		O	O
	■ Movement-correaltion?		O	O
S	■ Severity VAS (1-10): _____			
T	■ Pain pattern:			
	Persistent: duration (hours/days): ____			
	Episodic: duration (min): _____			
	Intervalls: min between episodes: ____			
	■ Previous episodes?		O	O
PO	■ Nausea/vomiting?		O	O
PR	■ Diarrhea?		O	O
	■ Constipation?		O	O
PU	■ Urinary symptoms?		O	O
PV	■ Unusual vaginal discharge?		O	O
	■ Time för last menstruation: _____			

Notes - Background

Medicines

Allergies

Previous illness

Living circumstances

Ethanol/smoking

Notes – Current problem

MAPLES: Medicines, Allergies, Previous illness, Living circumstances, Ethanol, Smoking

OPQRST: Onset, Position, Quality, Relieving/aggravating factors, Severity, Time

PO= per os, PR= per rectum, PU= per urethra, PV= per vagina

Appendix 1b. Example of the abdominal pain checklist. Page 2

REGION SKÅNE AKUTSJUKVÅRD LUND

Abdominal pain

Date of revision
2012-03-12

Suggestions are welcomed!
akutensriktlinjer@gmail.com

Input Vital parameters
Output Clinical information for assessing the probability of important causes of abdominal pain

PHYSICAL EXAMINATION

	Yes	No
■ Heart auscultation; regular rhythm?	O	O
■ Pulmonary auscultation: crepitations?	O	O
■ Abdominal distension?	O	O
■ Scars from previous surgery?	O	O
■ Bowel sounds:		
Normal?	O	O
Absent?	O	O
Metallic?	O	O
■ Generalised peritonitis?	O	O
■ Localised peritonitis?	O	O

TESTS

- ■ WBC, CRP, blood gas
  Consider liver function test + amylase
- ■ U-stix
- ■ Pregnancy test on fertile women
- ■ ECG on patients > 50 yrs
- ■ Ultrasound abdominal aorta on patients > 60 yrs

HYPOTHESIS-DRIVEN INFORMATION

Additional info only gathered on clinical suspicion

ACS/PULMONARY EMBOLISM		
■ ECG and Troponin		
AORTIC DISSECTION		
■ Syncope/presyncope?	O	O
■ Neurological deficits?	O	O
■ Equal blood pressure in both arms?	O	O
■ Equal femoral/carotid pulsations?	O	O
GYNECOLOGICAL DIAGNOSES		
■ Known ovarial cysts?	O	O
■ Previous salpingitis?	O	O
■ US when suspecting ectopic pregnancy: free fluid?		
BOWEL OBSTRUCTION		
■ Inspection/palpation of potential herniation sites		
■ Rectal exam: fecaloma? tumor?		
TESTICULAR TORSION		
■ Scrotal examination		
URINE RETENTION		
■ Bladderscan		

Notes – Physical examination and tests

Notes - Other

CHECKLIST EVALUATION

- Did the checklist help you in the management of your patient?

Not at all O O O O O O Much

- Du you want them to remain after the evaluation period?

No O | No opinion O | Yes O

Medical title

UL	O
AT	O
UL-leg	O
ST-akut	O
ST-annan	O
Spec-akut	O
Spec-annan	O

Appendix 2. Evaluation form for the substudy

Preconsultation assessment of TLoC

What is the probability for the following conditions?

	1 Excluded	2 Low probability	3 Compatible with	4 Probable	5 Very probable
Cardiogenic syncope	⊗	o	o	o	o
Seizure	o	o	⊗	o	o
Ortostatic hypotension	⊗	o	o	o	o
Reflex syncope (e.g. vasovagal reaction)	o	⊗	o	o	o

Do you think the patient needs to be admitted?

No	Probably not	Maybe	Probably	Yes
o	o	⊗	o	o
1	2	3	4	5

Postconsultation assessment of TLoC

What is the probability for the following conditions?

	1 Excluded	2 Low probability	3 Compatible with	4 Probable	5 Very probable
Cardiogenic syncope	o	⊗	o	o	o
Seizure	o	o	o	⊗	o
Ortostatic hypotension	⊗	o	o	o	o
Reflex syncope (e.g. vasovagal reaction)	o	⊗	o	o	o

Was the patient admitted?

No	*Probably not*	*Maybe*	*Probably*	**Yes**
o				⊗
1				5

Absolute difference

Diagnoses	Admission
2	2

TLoC= Transient loss of consciousness

This evaluation form corresponds to the (fictitious) example on page 7.