Does a Structured Data Collection Form Improve The Accuracy of Diagnosis of Acute Abdomen in an Urban Private Hospital?

Makanga W[1], Wasike R[1], Saidi H[2]

1 – Department of Surgery, Aga Khan University Hospital
2 – School of Medicine, University of Nairobi

Correspondence to: Dr. Winston Makanga, P.O. Box 168, 20116, Gilgil, Kenya. Email: wmakanga@yahoo.com

Abstract

Background: Accuracy of initial assessment of acute abdominal pain (AAP) is confounded by subjectivity and multiple etiologies for similar presentation. Standardized forms may harmonize the initial assessment, improve accuracy of diagnosis and enhance outcomes.

Objectives: To determine the extent to which use of a structured data collection form (SDCF) affected the diagnostic accuracy of AAP.

Methodology: A before and after study carried out from October 2011 to March 2012 of patients aged 13 years and older presenting with AAP in the emergency department (ED) of Aga Khan University Hospital, Nairobi (AKUH,N). Patients clerked by ED physicians using conventional history taking and examination between October and December 2011 were compared to a second group clerked after the introduction and use of a SDCF (January – March 2012) for proportion of correct diagnosis at initial encounter. This influence of age, gender and disease type on the impact of the form was evaluated as was the impact of the introduction of the structured forms on time to ED disposition, hospital stay, number and cost of investigations. Data were compiled in MS-Excel spreadsheets and analyzed using SPSS v16. P value of <0.05 was significant

Results: 125 participants were included, 60 in Period 1 and 65 in Period 2. The overall mean age for males was 28 and 34 for females. Patients with surgical abdominal conditions were 21% and 49% for medical conditions. The diagnostic accuracy was 58% and 43% before and after the introduction of the SDCF respectively (p=0.088). For surgical patients, diagnostic accuracy was 77% before and 31% after the introduction of the form (p=0.018).

Conclusions: The structured form did not improve the accuracy of diagnosing the causes of acute abdomen. It had a negative impact on the surgical diagnoses.

Introduction

Acute abdominal pain is a common presentation at hospital casualties with an estimated frequency of 5-10%; with one third of cases severe enough to require hospital admission(1). The challenges in identifying the latter group is compounded by multiplicity of possible etiologies, nonspecific nature of the pain, atypical presentation in up one third of patients and a high rate of inter-observer variations in elicitation of signs (2). Predictably, accuracy rates reported in literature have ranged from 42% to 65% (3,4).

Uncertainty in making the initial diagnosis often leads to a 'shot-gun' approach in ordering of laboratory and radiologic investigations in emergency departments, a practice that increases costs. Further, uncertainty or missing the diagnoses altogether may lead to increased complications. Rates of perforated appendicitis are markedly increased after day 3 of onset of symptoms of acute appendicitis which in turn influences hospital stays and hospitalization costs (5).

Attempts to improve accuracy in the evaluation of AAP have included the use of SDCFs (4,6). A SDCF not only organizes the data collection process but also makes sure that no detail in the history or in the clinical examination is left out (7). Their widespread adoption outside centers in the United Kingdom have been hampered by cost of digitization, learning curve and the view that this is an attempt to dehumanize history taking art (1). We purposed to investigate if the SDCF can improve the accuracy in a non-British resource constrained setting as previously shown.

Methodology

This was a quasi-experimental before and after study conducted at the ED of the Aga Khan University Hospital-Nairobi (AKUH, N) which consists of a two bed acute room, six bed observation area and nine

consultation rooms and manned by 14 senior house officers. Patients aged 13 years and older presenting with non-traumatic AAP between 1st October 2011 and 31st March 2012 were consecutively recruited by trained triage nurses. Period 1 lasted for 3 months, 1st October to 31st December while Period 2 from 1st January to 31st March.

Patients with abdominal surgery in the preceding 3 months, advanced pregnancy (more than 20 weeks), known recurrent abdominal pain and patients with prior investigations from the referring facility were excluded. From an earlier pilot study (unpublished) we had computed an underlying diagnostic accuracy of 40%. We wanted to detect a 25% point increase (averaged from previous studies) in accuracy on introduction of the SDCF i.e. to 65% with a power of 80% and a 95% confidence interval. This yielded a sample size of 120 (60 per period).

In Period 1, patients were clerked conventionally followed by diagnosis formulation, ordering of investigations and disposition. In Period 2, a SDCF (Fig. 1) was used. This consisted of a series of specific questions and examination maneuvers with definitive responses that were checked. The doctor used the summed up responses to come up with the most likely clinical diagnosis based on those responses. The proposed diagnosis guided further investigations and ED disposition.

The time from initial contact to disposition decision was recorded. The final diagnosis was crafted from the results of the investigations and consultant inputs at follow up after ED disposition. Where the definitive diagnosis could not be verified from the charts, yet the patients improved on medications, the conditions were labeled as non-specific abdominal pain (NSAP). This occurred in 19% (24 out of 125) of the patients. Two interactive training sessions, lasting two hours each, were conducted at the ED to familiarize the physicians and nurses with the use of the SDCF before implementation. The author monitored the initial days of implementation and addressed challenges with accuracy of data entered and completeness of information.

The ED initial diagnosis (without the aid of tests) was evaluated for concordance with the definitive diagnosis. This was determined by a definitive investigation (e.g. CT Scan), or laparotomy. If the definitive diagnosis was not found initially and patient sent home, further evaluation proceeded as an outpatient – at the clinic, and a diagnosis made by outpatient tests. Ambiguous cases were determined by a consultant surgeon's validation of the diagnosis during clinic follow-up.

This diagnostic accuracy was compared for the period before and after introduction of the structured form and stratified for age, gender and disease type. We also analyzed the impact of the introduction of the structured forms on time to disposition, hospital stay, number and cost of investigations. The Statistical Package for the Social Sciences (SPSSTM) Version 16 was used to analyze the data. Proportions and means were compared using the z test, t test as appropriate. Significance of difference was set at $p < 0.05$. This study was undertaken after approval by the Research and Research Ethics Committees of AKUH, N.

Results

From October 2011 to March 2012, a total of 196 patients were eligible for the study, 106 in the Period 1 and 90 in Period 2. In Period 1, 14 participants were excluded; five who were pregnant, two had abdominal malignancies, two were below the age of 13 and one had been assessed initially by a surgeon. In Period 2, 15 participants were excluded; five had pain for more than a week, five presented with predominant symptoms other than abdominal pain e.g. fever, two were known cases of peptic ulcer disease, one had spontaneous abortion, one had recent surgery and one had been seen by a surgeon. In Period 1, 32 participants (30%) had incomplete entry on initial diagnosis or were not followed up to ascertain the final diagnosis. The second period had 10 participants (11%) with incomplete entry on initial and/or final diagnosis. The difference was statistically significant (p=0.006). A total of 125 participants met the inclusion criteria and had complete data, 60 in Period 1 (57%) and 65 in Period 2 (72%).

Of the 60 in Period 1, a definitive diagnosis was ascertained in 47 patients while 13 patients had a diagnosis by consensus. In Period 2, 51 had a definitive diagnosis while 14 of the participant had a diagnosis by consensus. Of the 125 patients, 64 were male (35 in Period 1, 29 in Period 2), and 61 were female (31 in Period 1 and 30 in Period 2). The mean age in Period 1 and 2 was 28.25 (27.6 male and 28.8 female) and 34.2 years (34.9 male, 33.4 female) respectively.

MODIFIED STRUCTURED DATA COLLECTION FORM

DATA

NAME		PATIENT NO	
AGE (yrs)		CLINICIAN	
SEX – M / F		DATE	TIME

HISTORY

DURATION (hrs)						
SITE (ONSET) (+ for mild pain, +++ for severe)		Progression: Worse, Better, Same	Aggravators: Movement, Coughing, Breathing, Food, None, Other _____	Relievers: Lying still, Vomiting, Antacids, Food, None, Other _____	Female: LMP_____, Vaginal discharge Yes No, Per Vaginal bleed Yes No	
SITE (CURRENT)		Nature (circle): Sharp, Burning, Dull, Other_____	Character (circle): Constant, Intermittent, Colicky	Drugs (circle): Yes No	PMHx (circle): Cancer, Gallstones, Kidney stones, Pancreatitis
RADIATION		Other: Back, Shoulder, Groin	Micturition: Frequency, Dysuria, Hematuria	Ongoing medications: Antibiotics _____, Analgesics _____, Other _____	

SEVERITY (visual analog score)	1	2	3	4	5	6	7	8	9	10

Associated symptoms (circle)

Vomiting	Anorexia	Bloating	Jaundice

BOWELS (circle)

Normal	Constipation	Diarrhea	Blood	Mucus

EXAMINATION

VITALS	T	P	R	BP	
General appearance (circle): Calm, Writhing /restless, Lying still		COLOUR (circle): Jaundice, Cyanosis, Pallor		Bowel sounds (circle): Present, Absent, Reduced	Clinical Diagnosis
Surgical Scar (circle): Yes No		Location of tenderness (+ for mild, +++ for severe)		Recto-vaginal tenderness (circle)**: Left, Right, Generalized, Mass, Other_____	Investigations
ASSOCIATED SIGNS (circle): Rebound tenderness, Guarding, Rigidity, Distension → Generalized / localized, Mass		Describe mass: Size _____x_____cm, Diffuse / well defined, Consistency - S F H C*, Mobile Yes / No			Diagnosis and plan after investigations Time_____
					Discharge Diagnosis

The predominant diagnosis was medical in 61 patients, 49%, (28 in Period 1 and 33 in Period 2). Surgical diagnoses constituted 26 patients, 21% (13 in Period 1 and Period 2) and gynaecological conditions were 7% (5 in Period 1 and 4 in Period 2). The category of 'other' was 29% and constituted the conditions which couldn't be classified clearly e.g. non-specific AAP and conversion disorder.

A total of 29 (23% of the study) participants were admitted (17 in Period 1 and 12 in Period 2). Of these, 26 underwent surgery (6 and 2 laparotomies in Period 1 and 2 respectively, 11 and 6 appendectomies in Period 1 and 2 respectively). There were 2 negative appendectomies (all in the first period and no negative laparotomy in both periods. One cholecystectomy was done in Period 2. Three of the admitted patients in the Period 2 did not have surgical intervention.

The mean emergency disposition time was two hours thirty six minutes, (two and a half hours in Period 1 and two hours, forty two minutes in Period 2, p 0.452). The mean number of investigations (MNI) done was 3.38, (3.25 in Period 1 and 3.5 in Period 2, p 0.542). The median cost of investigations in Period 1 was 4315 shillings and that of Period 2 was 6530 shillings (p 0.368).

Table 1; Baseline group characteristics.

		Period 1	Period 2	P value*
Number of patients	Male	29	35	
	Female	31	30	
	Surgical	13	13	
	Medical	28	33	
	Gynaecological	5	4	
	Others	14	15	
Mean age	Overall	28.3	34.2	*0.001*
	Male	27.6	34.9	*0.012*
	Female	28.8	33.4	*0.028*
Mean Pain score	Overall	7.3	7.32	*0.927*
	Male	7.41	7.69	*0.468*
	Female	7.19	6.9	*0.376*

*p value <0.05

Primary outcome

The diagnostic accuracy in Period 1 was 58% and 43% in Period 2. The difference was not statistically significant ($p = 0.08$). Sub-analyses by disease type, age, sex and time of assessment (i.e. day vs. night) did not reveal statistical significance in all analyses except in male patients who had a drop in accuracy of 62% vs. 37% in Period 1 and 2 (p 0.04) and in patients with surgical diseases with a similar drop in accuracy of 77% vs. 31% (p 0.018) respectively (Table 2).

The overall diagnostic accuracy for both periods was 51.2%. The sex specific diagnostic accuracy was 50% for male and 49% for female participants. When stratified for age sets, the diagnostic accuracy for those aged 13 – 20 years old had 47%, 21 – 40 years old had 53% and those aged 40 – 60 years old had accuracy of 44%. Surgical cases had an overall accuracy of 58%, medical cases had 72%, while the gynaecological cases were 22% accurate.

Table 2: Comparative diagnostic accuracies

Category		Period 1	Period 2	P value*
	Overall	58 (35/60)	43 (28/65)	*0.08863*
Age	Under 20	50	43	*0.77*
	21-40	61	45	*0.14*
	41-60	50	43	*0.8*
Sex	Male	62 (18/29)	37 (13/35)	*0.047*
	Female	55 (17/31)	50 (15/30)	*0.705*
Disease type	Surgical	77 (10/13)	31 (4/13)	*0.018*
	Medical	75 (21/28)	70 (23/33)	*0.645*
	Gynecologic	20 (1/5)	25 (1/4)	*0.858*
	Other	21 (3/14)	0	*0.058*

* $p < 0.05$

Secondary outcomes

The mean emergency disposition time (EDt) was compared in patients with complete documentation of the EDt. There was a modest increase in EDt on introducing the SDCF from 2 hours 30 minutes to 2 hours 42 minutes (p 0.452).

Table 3: Emergency department time to disposition

	Period 1	Period 2	P value
Overall	2hrs 30min	2hrs 42min	*0.452*
Male	2hrs 24min	2hrs 54min	*0.255*
Female	2hrs 36min	2hrs 30min	*0.848*

- Mann Whitney U test

Subgroup analysis did not show any statistically significant difference but a modest reduction on EDt in females. The overall EDt was 2 hours 38 minutes. The mean number and cost of investigations (MNI) were computed; as a sum of all investigation (and their cost) over all patients per period, and compared across both study groups. The charges were based on the hospital prices at the time of the study. The resulting figures were compared across the two study periods. The MNI in Period 1 was 3.25 as compared to 3.49 seen in Period 2 (p 0.542). The overall MNI was 3.4.

Table 4: Number and mean cost of investigations

		Period 1	Period 2	P value
Mean no. of investigation	Overall	3.25	3.49	*0.542*
	Male	3.25	3.85	*0.345*
	Female	3.35	3.13	*0.687*
	Surgical	5	3.85	*0.294*
	Medical	2.86	3.21	*0.48*
Median cost of investigations (cost in KES)	Overall	4315	6530	*0.368*
	Male	3490	5390	*0.475*
	Female	5390	4045	*0.839*
	Surgical	15000	13600	*0.839*
	Medical	3450	3550	*0.487*

- Mann Whitney U test

There was no significant difference in the cost of tests. There were less investigations and lower cost of tests in female (p 0.839) and surgical patients (p 0.839) but these were not significant. Sub-analyses showed an insignificant increase in cost of lab tests on introduction of the SDCF.

23 patients were admitted in the entire study (18.4% of the study population), 17 Period 1 and 8 in Period 2. Eight laparotomies were performed, 6 in Period 1 and 2 in Period 2. 15 appendectomies were done during the study period, 11 in Period 1 and 4 in Period 2. The negative appendectomy rate

was the proportion of normal appendices found at surgery and confirmed by the pathology report. There were 2 negative appendectomies (18.2% negative appendectomy rate) in Period 1 and none in Period 2. The mean hospital stay in days was computed in the two study periods and compared. In Period 1, mean hospital stay was 3.92 while that of Period 2 was 3.2 days (p 0.921).

Discussion

The results presented do not show a significant difference yet the aim was to improve clinical diagnostic accuracy by use of the SDCF. There was a reduction in diagnostic accuracy by a factor of 15% on the primary outcome (p 0.08). This was clinically significant especially on the anticipation that the study would show improvement in diagnostic accuracy. The study suggests that the ED clinicians fared worse with the use of SDCF.

A significant difference was noted in the proportion of patients who were excluded due to inadequate entry of initial and/or final diagnosis. Period 1 had a 30% exclusion against 11% in Period 2 (p=0.006). The ED clinicians were not as keen in filling in the form properly in Period 1. This changed in Period 2, due to increasing awareness. Our two study periods were comparable. Period 2 patients however were on average older by six years (p = 0.001). This was found not to be significant since disease conditions, especially abdominal, in this age-set do not differ much.

Surgical cases with the most concordance were appendicitis; ten out of the 15 cases. This was in keeping with Korner's paper in which he showed high baseline accuracy in the diagnosis of acute appendicitis of 75% (4). Of the remaining surgical cases, there were three with perforated peptic ulcer and two with urolithiasis.

Despite numerous studies showing positive outcomes with use of diagnostic aids, some authors have also shown minimal impact. Sutton et al showed a drop in accuracy from 65% to 47 – 58% in CADs but lacked clarity in participant selection and external validation(8). Kikerby showed a similar drop from 65% to 53% on CAD usage citing differences in disease pattern and delayed referral as reasons for the drop(9). Ohmann et al showed a modest drop from 59% to 51% when computers were introduced. His ED clinicians were a heterogeneous group (10). Our study attempted to overcome limitations cited in previous studies by ensuring proper use of the SDCF by pre-introduction seminar and continuous

monitoring of the filled forms, prospective feedback to clinicians on appropriate use of the form, excluding cases with a referral diagnosis and by strictly using a homogenous group of ED clinicians without surgical training.

Our negative results may be a reflection of a failure of acceptance, poor use or inadequate time for assimilation of the SDCF rather than a failure of the tool. We occasionally noticed in Period 2 that clinicians clerked conventionally and later filled the SDCF. A post-study informal survey was conducted to assess the attitude on use of the form. Six out of the twelve ED clinicians did not fully understand how to use the form. Three clinicians found the form to be tedious. These findings reflect those of Guerlain et al who reported 35% compliance on use of SDCF where the reasons for not using it were forgetfulness and difficulty in patients with multiple complaints(6). This in turn led to selection on use of the form for only those patients with a single complaint. Wellwood et al documented 65% compliance to SDCF and a 50% compliance to CAD because of a robust pre-study training(7). Other instances of robust training on use of diagnostic aids include Guerlain who dedicated two weeks of crossover within which the clinicians were trained and had feedback sessions before full adoption of the SDCF(6).

There wasn't a significant difference in the EDt on introduction of the SDCF. The MNI across the two study periods also did not show significant difference, albeit a tendency towards doing fewer investigations in the surgical cases in Period 2. The overall cost of investigations was not significantly different in the two groups, although there was a trend towards an increase in the cost of investigations on the introduction of the form. Only the surgical cases showed a drop in the cost during Period 2.

Significant limitations of the study included incomplete data whereby 25% of patients who met the inclusion criteria did not have the initial diagnosis documented. In addition, there was a tendency by clinicians to use the SDCF only in patients who had higher pain scores. Conversely, patients who presented with AAP as part of a systemic medical condition may have been included in the study if there pain score was high. This may explain the large number of cases of medical cases. A significant number of patients were lost to follow-up because of rapid resolution of symptoms and failure to attend subsequent clinics. Failure of this follow-up led to non-entry of diagnosis and subsequent exclusion from the study. This may have led to a selection bias

for patients who had severe enough pain to comply with follow up. The assumption that clinical diagnosis by the consulting clinician was the correct one has not been validated. In this study, the main interest was in the ED clinicians' ability to formulate a clinical as opposed to a histological diagnosis. However, when the definitive diagnosis was available, it was used as validation. In situations where the definitive diagnosis could not be ascertained, the attending clinician's diagnosis was found fit for our purposes. The design of the study did not cater for the time required for the ED staff to adequately familiarize with the form and build enough confidence to use it. This being a new concept for our hospital, a learning phase of at least 3 months would have been required to show effect.

Conclusion

The SDCF did not improve diagnostic accuracy as proposed. It is probable that adoption of such a tool in an inadequately prepared system may result in suboptimal results. More studies are required to show the true impact of the structured form. This study offers baseline information on issues of concern in our ED e.g. multiple non-contributory investigations and prolonged time spent in the ED, some of which have been raised before.

Refferences

1. Brewer BJ, Golden GT, Hitch DC et al. Abdominal pain . An analysis of 1000 consecutive cases in a University Hospital emergency room . Am J Surg. 1976;131(2):219–23.
2. Hickey M, Kiernan G, Weaver K et al. Evaluation of abdominal pain. Emerg Med Clin North Am. 1989;7(3):437.
3. Lawrence PC, Clifford P. Acute abdominal pain: Computer aided diagnosis by non-medically qualified staff. Ann Roy Coll Surg Engl. 1987;69(5):233–4.
4. Körner H, Söndenaa K, Söreide JA, et al. Structured data collection improves the diagnosis of acute appendicitis. Br J Surg. 1998;85(3):341–4.
5. Claire L, Temple BA. The natural history of appendicitis in adults. Ann Surg. 1995;221(3):278–81.
6. Guerlain S, Lebeau K, Thompson M, et al. The effect of a standardized data collection form on diagnostic accuracy of acute abdominal pain. Human factors. 2001;45:1284–8.
7. Wellwood J, Johannessen S. How does computer-aided diagnosis improve the management of acute abdominal pain ? Ann Roy Coll Surg Engl. 1992;74:40–6.
8. Sutton GC. How accurate is computer-aided diagnosis? Lancet. 1989;2(8671):1102–3.
9. Kirkeby OJ, Ris C. Use of a computer system for diagnosing acute abdominal pain in a small hospital. Scan J Gastroenterol. 1987;22:174–6.
10. Ohmann C. Acute abdominal pain--standardized findings as diagnostic support. Results of a prospective multicenter intervention study and testing of a computer-assisted diagnosis system. Chirurg. 1992;63(2):113–22.