

Abdominal auscultation does not provide clear clinical diagnoses

Maja Durup-Dickenson¹, Marie Kirk Christensen² & John Gade³

ABSTRACT

INTRODUCTION: Abdominal auscultation is a part of the clinical examination of patients, but the determining factors in bowel sound evaluation are poorly described. The aim of this study was to assess inter- and intra-observer agreement in physicians' evaluation of pitch, intensity and quantity in abdominal auscultation.

MATERIAL AND METHODS: A total of 100 physicians were presented with 20 bowel sound recordings in a blinded set-up. Recordings had been made in a mix of healthy volunteers and emergency patients. They evaluated pitch, intensity and quantity of bowel sounds in a questionnaire with three, three and four categories of answers, respectively. Fleiss' multi-rater kappa (κ) coefficients were calculated for inter-observer agreement; for intra-observer agreement, calculation of probability was performed.

RESULTS: Inter-observer agreement regarding pitch, intensity and quantity yielded κ -values of 0.19 ($p < 0.0001$), 0.30 ($p < 0.0001$) and 0.24 ($p < 0.0001$), respectively, corresponding to slight, fair and fair agreement. Regarding intra-observer agreement, the probability of agreement was 0.55 (95% confidence interval (CI), 0.51-0.59), 0.45 (95% CI: 0.42-0.49) and 0.41 (95% CI: 0.38-0.45) for pitch, intensity and quantity, respectively.

CONCLUSION: Although relatively poor, observer agreement was slight to fair and thus better than expected by chance. Since the diagnostic value of auscultation increases with addition of history and clinics, and may be further improved by systematic training, it should still be used in the examination of patients with acute abdominal pain.

FUNDING: not relevant.

TRIAL REGISTRATION: not relevant.

Abdominal auscultation has been a part of the clinical examination of patients with gastroenterological complaints for more than 150 years. Hooker published an essay in 1849 in which he described variations in frequency and intensity of intestinal gurgling in the course of different diseases of the digestive organs [1].

Bowel sounds arise when peristaltic contractions propel the intraluminal contents of the intestine forward. Different factors are thought to influence the sound characteristics, such as the intraluminal material, varying amounts of air and gas and the bowel diameter [2]. Efforts to objectify bowel sounds have been made

with computerized auscultation in an attempt to develop an analysis similar to electrocardiography and electroencephalography [3]. Still, no reproducible patterns of bowel sounds in healthy persons have been found during these attempts, and no apparatus aside from the stethoscope (acoustic and electronic) is yet available for abdominal auscultation.

It has been shown, however, that physicians can separate pathological from normal bowel sounds with a substantial degree of certainty [4, 5], and abdominal auscultation is therefore still being used in the evaluation of patients with acute abdominal pain. On the other hand, it is not well-defined which specific sound characteristics physicians pay attention to in their evaluation. The aim of this study was to assess inter- and intra-observer agreement among physicians in their evaluation of pitch, intensity and quantity of bowel sounds in abdominal auscultation.

MATERIAL AND METHODS

Technical equipment

The technical set-up has previously been described [4]. Briefly, bowel sounds were recorded with a digital tape recorder (Denon DAT DTR-2000) and a microphone with a rubber cuff enclosing an air volume approximating that of a stethoscope. The physicians used their own customary stethoscope placed on a wooden "abdominal dummy" containing a small loudspeaker. The system frequency response was kept within ± 3 dB bounds in the range of 60-1,200 Hz.

Patients

Bowel sounds were recorded for 8-20 minutes in four healthy volunteers and eight emergency patients from a surgical gastroenterological ward (Table 1). They were all more than 18 years old and participated voluntarily. The patients' diagnoses were verified radiologically and perioperatively, with the exception of one case in which the diagnosis was obtained strictly by radiological and clinical findings. The healthy volunteers had no history of gastroenterological disease and received no medication. A selection of bowel sound sequences was obtained from patients and volunteers to represent both typical and atypical cases. For each subject, approximately one minute of bowel sounds was selected for a

ORIGINAL ARTICLE

1) Department of Cardio-Thoracic and Vascular Surgery, Aarhus University Hospital

2) Surgery South, Aalborg University Hospital

3) Department of Surgery, Vejle Hospital

Dan Med J

2013;60(5):A4620

 TABLE 1

Patients in the master tape and their diagnoses.

Patient	Diagnosis
No. 1	Normal ^a
No. 2 (= no. 13)	Duodenal ulcer perforation
No. 3 (= no. 14)	Small-bowel obstruction ileus
No. 4	Normal ^a
No. 5 (= no. 15)	Normal ^a
No. 6	Large-bowel obstruction ileus
No. 7 (= no. 16)	Normal ^a
No. 8	Small-bowel obstruction ileus
No. 9 (= no. 17)	Bacterial peritonitis (rectal perforation)
No. 10 (= no. 18)	Strangulation volvulus of the caecum
No. 11 (= no. 19)	Small-bowel obstruction ileus
No. 12 (= no. 20)	Subileus ^b

a) The normal subjects had no history or signs of gastroenterological disease and received no medication.

b) Diagnosis was verified radiologically and clinically.

 TABLE 2

Physicians' specialization and distribution by departments. The values are n (N = 100).

	Specialists ^a	Non-specialists ^b	Not stated
Surgical gastroenterology	18	32	2
Medical gastroenterology	8	17	1
Urology	11	11	–
Total	37	60	3

a) Consultants and senior registrars.

b) Registrars, trainees and interns.

 TABLE 3

Interpretation of kappa coefficients (κ).

κ	Interpretation
0	Less than chance agreement
0.01-0.20	Slight agreement
0.21-0.40	Fair agreement
0.41-0.60	Moderate agreement
0.61-0.80	Substantial agreement
0.81-0.99	Almost perfect agreement

 TABLE 4

κ-values for inter-observer agreement.

	In general ^a	Healthy volunteers	Patients with obstruction
Pitch	0.19 (p < 0.0001)	0.02 (p < 0.0138)	0.16 (p < 0.0001)
Intensity	0.30 (p < 0.0001)	0.05 (p < 0.0001)	0.31 (p < 0.0001)
Quantity	0.24 (p < 0.0001)	0.26 (p < 0.0001)	0.16 (p < 0.0001)

a) 4 healthy volunteers, 6 patients with intestinal obstruction, 2 patients with peritonitis.

master tape, which was later presented to the physicians. Eight bowel sound recordings were then duplicated in order to examine intra-observer variation.

Physicians

A total of 100 physicians with different specialization and experience were included (**Table 2**). They answered questions regarding their present field of practice, their level of specialization, years since university graduation and, finally, the type of stethoscope used. While listening to the master tape, they completed the questionnaire for each patient's bowel sounds. All participating physicians were informed and aware that the bowel sounds were obtained from a mix of both healthy control subjects and acute patients from a surgical gastroenterological department. They were blinded, however, to information regarding age, sex, history, other clinical findings and ultimate diagnoses. They were also unaware that certain recordings had been duplicated.

Questionnaires

In the questionnaire, the physicians were asked to evaluate the 20 bowel sounds recording pitch, intensity and quantity of sounds. Pitch was defined as the highness or lowness of tones assigned to relative positions on a musical scale and was presented as a scale of 1-11 in the questionnaire. Intensity was defined as the volume of sounds and was presented as three check boxes (normal, increased or decreased). Quantity was defined as the multitude of sounds and was presented as four check boxes (normal, increased, decreased or absent).

Statistics

Fleiss' multi-rater kappa coefficients were calculated in the statistical software R for strength of inter-observer agreement. This method calculates the degree of agreement better than expected by chance, expressed as a figure between zero and one. Zero represents no agreement, 0.01-0.20 slight agreement, 0.21-0.40 fair agreement, 0.41-0.60 moderate agreement, 0.61-0.80 substantial agreement and 0.81-0.99 almost perfect agreement (**Table 3**) [6, 7]. Patient data were removed from analysis if one or more physicians had not submitted an assessment for a particular patient's bowel sounds. In order to assess intra-observer agreement, calculation of probability was performed to determine the probability that the physicians would agree with their previous evaluation of the same sound recording. The probability of agreement expected by chance alone was 33%, 33% and 25%, for the three sound qualities, respectively. The reason for this is that physicians had to choose between three or four possible answers in the questionnaire.

Evaluation of pitch in the questionnaire was de-

signed as a visual analogue scale based on the assumption that this would correspond to the continuous nature of sound frequencies. However, data analysis showed that evaluation of inter- and intra-observer agreement in the resulting 11 categories was neither possible nor meaningful. Instead, calculations were performed after graduating the scale into three categories: 1-3, 4-8 and 9-11 representing low, medium and high pitches, respectively.

Ethics

The investigation was approved by the regional ethics committee and conducted in accordance with the Declaration of Helsinki II. Informed consent was obtained from all subjects participating in the study.

Trial registration: not relevant.

RESULTS

Inter-observer agreement

In the full material, the question of pitch was answered by 55% of physicians, resulting in $\kappa = 0.19$ ($p < 0.0001$). Correspondingly, intestinal sound intensity was answered by 83% of physicians with $\kappa = 0.30$ ($p < 0.0001$) and quantity by 91% of physicians with $\kappa = 0.24$ ($p < 0.0001$). Interpretation of κ -values is shown in Table 3. Patients were divided into groups of healthy volunteers, patients with peritonitis, and patients with obstruction. It was not possible to obtain results for patients with peritonitis since too few patients were included. The results of inter-observer agreement are shown in Table 4.

Intra-observer agreement

One hundred physicians evaluated the sound recordings of eight patients twice. The probability that a physician agreed with a previous assessment regarding pitch was 0.55 (95% confidence interval (CI): 0.51-0.59), or slightly more than every other time. For intensity, the probability was 0.45 (95% CI: 0.42-0.49); while for quantity, the probability was 0.41 (95% CI: 0.38-0.45) or between every third and every other time. Due to the three or four answering possibilities of each category, all results are higher than would be expected by chance (see Statistics).

DISCUSSION

It is not well-defined which specific sound characteristics physicians pay attention to in their evaluation of bowel sounds. We present data on agreement among physicians in their evaluation of pitch, intensity and quantity of bowel sounds in abdominal auscultation.

We found slight inter-observer agreement in the evaluation of pitch ($\kappa = 0.19$, $p < 0.0001$) and fair agreement in the evaluation of sound intensity ($\kappa = 0.30$, $p <$

0.0001) and quantity ($\kappa = 0.24$, $p < 0.0001$). The results of intra-observer agreement are higher than would be expected by chance alone, since physicians agreed with their previous assessment of pitch, intensity and quantity in 55%, 45% and 41% of the times, respectively, in comparison to the 33%, 33% and 25% expected by chance alone.

Inter-observer agreement of sound qualities in abdominal auscultation has been evaluated in only one small study by Bjerregaard et al [8]. Four physicians were asked to categorize bowel sounds as normal, increased, reduced, metallic or "other" upon examination of 40 patients admitted to a surgical ward with acute abdominal pain. The inter-observer agreement was fair ($\kappa = 0.29$), which approximately corresponds to our results.

Since abdominal auscultation is part of the general patient assessment and, therefore, routinely performed, it is relevant to evaluate whether physicians can distinguish between normal and pathological bowel sounds in patients admitted with acute abdominal pain. In the same material as ours, Gade et al [4] had 100 physicians identify 12 bowel sound recordings as pathological or non-pathological. Recordings from normal subjects were correctly identified as non-pathological in 72% of the cases. For patients with obstructive ileus or peritonitis, bowel sounds were correctly identified as pathological in 64% and 43% of the times, respectively. The diagnostic value of abdominal auscultation has been further evaluated. Gu et al [5] included 20 physicians who were presented with 43 recordings in a blinded fashion and were asked whether each was from a normal subject or from a subject with bowel obstruction or paralytic ileus. Physicians arrived at the correct diagnosis a median of 30 times out of 43 (accuracy = 69.8%), obtaining a κ -value of 0.57, representing moderate agreement. Further, they found a substantial intra-observer agreement ($\kappa = 0.72$). In a prospective study by Pines et al [9], 122 pairs of residents and attending physicians evaluated 122 patients admitted with acute abdomen consecutively. Inter-observer agreement regarding normal bowel sounds was fair ($\kappa = 0.36$).

It is interesting that in the evaluation of whether the diagnosis was normal, obstruction or paralytic ileus, the strength of agreement obtained by Gu et al [5] was higher than the agreement we found regarding pitch, intensity, and quantity. This suggests that perhaps physicians do not rely solely on the three parameters presently evaluated. Pines et al found inter-observer agreement rates on abdominal auscultation only a little higher than ours. However, it is difficult to compare our study to this, since the set-ups are different. Both the study of Gu et al and our study possess a high degree of external validity, as participating physicians came from various backgrounds in terms of both areas of speciality

Abdominal auscultation is part of general patient assessment and, therefore, routinely performed by physicians.



and years of experience. Another possible explanation for differences in strength of inter-observer agreement could be the greater number of observers in our study.

As demonstrated by Böhner et al [10] and Eskelinen et al [11], abdominal auscultation can contribute to establish the correct diagnosis in patients with bowel obstruction. It is more arguable regarding peritonitis [12-14]. The diagnostic value increases when history and other clinical manifestations are added, such as abdominal distension, guarding and vomiting. Based on the studies above [4, 5, 10-13], abdominal auscultation can be considered relevant in the physical examination of patients with acute abdominal pain.

Studies of inter-observer agreement of auscultation of heart and lungs have reported agreement rates from slight to moderate in the form of κ -values below 0.5, which is similar to those observed in abdominal auscultation [15, 16]. Methods to improve cardiac auscultation, such as recordings of heart sounds with digital stethoscopes and group-wise discussion of these, have been described [17]. These could arguably be well applied in teaching methods of abdominal auscultation and could improve the auscultatory skills acquired by medical students and interns at an early stage in their careers.

A particular strength of our study is the large number of physicians included. Moreover, the study was conducted in artificial and controlled settings, preventing the physicians from incorporating other observations into their evaluation. The physicians' use of their own stethoscopes permitted variability in individual auscultatory techniques and yet mimicked everyday practice. By listening to the same recordings in an isolated setting, a possible bias from non-simultaneous patient assessment was eliminated. This may, however, also be considered a weakness, as it does not reflect typical conditions of the clinical setting. For example, it can be difficult to obtain a quiet environment on busy wards with real patients. Further study limitations include the restricted selection of only 20 sound recordings (20 minutes). One consequence of this was that only two patients with peritoni-

tis were included. A limited number of recordings were decided upon in order to avoid discomfort to the physicians due to ear pressure from the stethoscope over an extended period of time. Furthermore, a longer listening time could have reduced physicians' concentration, thus introducing a new bias.

In conclusion, inter-observer agreement among physicians in their evaluation of pitch, intensity and quantity of bowel sounds in abdominal auscultation is slight to fair. The relatively poor observer agreement obtained in our study suggests that physicians cannot rely on abdominal auscultation alone in patient assessment. However, it is interesting that accuracy determining normal versus pathological sounds was higher on the same material, and that the diagnostic value of auscultation increases when history and other clinical manifestations are added. We therefore believe that auscultation should still be used in the examination of patients with acute abdominal pain.

A normal range for pitch, intensity and quantity of intestinal sounds has never been established in larger materials, and systematic training including instructions and recordings is entirely absent. We believe that knowledge of a scientifically well-defined normal range combined with training would considerably improve abdominal auscultation as a diagnostic tool.

CORRESPONDENCE: Maja Durup-Dickenson, Hjerter-, lunge- karkirurgisk Afdeling, Aarhus Universitetshospital, 8200 Aarhus N, Denmark.
E-mail: majadickenson@gmail.com

ACCEPTED: 28 February 2013

CONFLICTS OF INTEREST: Disclosure forms provided by the authors are available with the full text of this article at www.danmedj.dk.

ACKNOWLEDGMENTS: We thank the Institute of Biostatistics, Aarhus University, for support in data analysis, and the Department of Surgical Gastroenterology, Glostrup Hospital, for the material which originates for the former Department D, Glostrup Hospital. The department was later closed, and the employees are presently working at other departments.

LITERATURE

- Hooker C. An essay on intestinal auscultation. *Boston Med Surg J* 1849;XL:409-42.
- Craine BL, Silpa ML, O'Toole CJ. Two-dimensional positional mapping of gastrointestinal sounds in control and functional bowel syndrome patients. *Dig Dis Sci* 2002;47:1290-6.
- Dalle D, Devroede G, Thibault R et al. Computer analysis of bowel sounds. *Comput Biol Med* 1975;4:247-56.
- Gade J, Kruse P, Andersen OT et al. Physicians' abdominal auscultation. A multi-rater agreement study. *Scand J Gastroenterol* 1998;33:773-7.
- Gu Y, Lim HJ, Moser MA. How useful are bowel sounds in assessing the abdomen? *Dig Surg* 2010;27:422-6.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-74.
- Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. *Fam Med* 2005;37:360-3.
- Bjerregaard B, Brynitz S, Holst-Christensen J et al. The reliability of medical history and physical examination in patients with acute abdominal pain. *Methods Inf Med* 1983;22:15-8.
- Pines J, Uscherpines L, Hall A et al. The interrater variation of ED abdominal examination findings in patients with acute abdominal pain. *Am J Emerg Med* 2005;23:483-7.
- Bohner H, Yang Q, Franke C et al. Simple data from history and physical examination help to exclude bowel obstruction and to avoid radiographic studies in patients with acute abdominal pain. *Eur J Surg* 1998;164:777-84.
- Eskelinen M, Ikonen J, Lipponen P. Contributions of history-taking, physical examination, and computer assistance to diagnosis of acute small-bowel obstruction. A prospective study of 1333 patients with acute abdominal pain. *Scand J Gastroenterol* 1994;29:715-21.
- Williams NM, Johnstone JM, Everson NW. The diagnostic value of

- symptoms and signs in childhood abdominal pain. *J R Coll Surg Edinb* 1998;43:390-2.
13. Brewer BJ, Golden GT, Hitch DC et al. Abdominal pain. An analysis of 1,000 consecutive cases in a University Hospital emergency room. *Am J Surg* 1976;131:219-23.
 14. Arnbjörnsson E, Bengmark S. Auscultation of bowel sounds in patients with suspected acute appendicitis – an aid in the diagnosis? Preliminary results. *Eur Surg Res* 1983;15:24-7.
 15. Iversen K, Greibe R, Timm HB et al. A randomized trial comparing electronic and conventional stethoscopes. *Am J Med* 2005;118:1289.
 16. Lok CE, Morgan CD, Ranganathan N. The accuracy and interobserver agreement in detecting the ‘gallop sounds’ by cardiac auscultation. *Chest* 1998;114:1283-8.
 17. Tavel ME. Cardiac auscultation: a new method for instruction. *Med Educ* 2010;44:1132.