Agreement of clinical measurements of liver size with ultrasound when performed by medical students
Selena A Hunter, Justin Brimble, Mark Weatherall, Duncan C Galletly

Abstract

Aim The accuracy of physical examination techniques in detecting liver disease is unclear. We sought to determine the perceived location of the lower liver border via scratch, percussion, palpation and ballottement with novice medical student examiners, compared with ultrasound localisation.

Method Five novice medical students learnt four liver examination techniques and measured the lower liver border in 19 healthy volunteers. The difference between the examination technique and ultrasound (bias) and limits of agreement of each method were estimated by mixed linear models and shown using Bland-Altman-like plots.

Results All techniques had similar bias for the lower liver border, between 1.6 and 1.9 cm superior to the ultrasound measurement. Percussion had the smallest variability with increasing liver size. Limits of agreement were wide for all techniques, smallest for palpation (6.04 cm) and largest for scratch (7.2 cm) compared to the mean liver distance of 8 cm.

Conclusion There was no difference in bias between the scratch, percussion, palpation and ballottement techniques regarding the lower liver border. All techniques had very wide limits of agreement, although palpation had the smallest. Liver size examination by novice medical students in healthy subjects is unreliable. Further research is needed using examiners with a different level of expertise and participants of varying body habitus and confirmed liver disease.

Bedside physical examination techniques are important to diagnose liver pathology. They are used to locate the lower border of the liver and to estimate liver size.1,2 These techniques can be performed rapidly and are an important part of the traditional gastro-intestinal examination.2–7 The accuracy of these techniques is uncertain8–12 and their value in clinical settings has been questioned.1

Limitations of studies investigating hepatic physical examination techniques are the small numbers of study participants11,13,14 and examiners9,11,12 as well as the poorly standardised expertise of the examiners performing the tests.13,15 Estimates of physical examination accuracy may lack statistical power to detect departures from accuracy or may not be generalisable to all examiners who use these techniques.

The best examination technique to localise the lower hepatic border is uncertain. All of palpation12, the scratch technique16 and percussion11 are reported to be best when compared to ‘gold standard’ measurements using ultrasound or hepatic scintiscan. All published studies report that physical examination underestimates liver span and that the lower border is estimated superior to the true liver edge.11,12,17 Resolution of this uncertainty about the best examination technique is relevant to what and how undergraduate students should be taught about liver examination.

The uncertainty about optimal hepatic examination technique is partly explained by variations in teaching and execution of each examination method. Clinical examination texts variably describe liver examination techniques2,4,18–24. Differences include the placement position of stethoscopes,4,18,25 starting position of a technique,5,6,20,25 whether a measurement is in the mid-clavicular or mid axillary line,22,25 the correct placement of the hands throughout examination procedures, and the intensity of percussion.2,4,26
Studies of liver examination have also used a variety of types of examiners, for example physicians or medical students\(^1,10,12,14\) and participants who were examined, healthy people\(^9,17\) or those with known liver disease.\(^11\) In addition most studies report comparison of one or two physical examination techniques to hepatic ultrasound or scintiscan.\(^1,14,15\)

The aims of this study are to describe the agreement of clinical measurement of liver size performed by novice medical students compared to hepatic ultrasound.

**Methods**

**Recruitment of study participants**

**Medical student examiners**—We recruited five junior medical students by social media (Facebook) and student webmail. Inclusion criteria were that students be enrolled in the University of Otago Medical School and be available to attend all of the teaching and examination sessions. Students were not recruited if they had previous clinical experience. All students were yet to begin clinical training, but had received a tutorial during their ‘Early Learning in Medicine’ course which introduced the physical examination techniques of Palpation and Percussion of the Liver. Students had no previous experience in the techniques of Scratch or Ballottement.

**Patient examination participants**—Twenty stable outpatient volunteers were recruited for the two-day testing period by social media notices and through a contact list of ‘Friends of the Wellington School of Medicine’, a patient group who volunteer to participate in student teaching.

Outpatients were included if their medical conditions were stable, were between 18 and 80 years old, and able to attend all of the required sessions. Participants were not recruited if they had any known hepatic or abdominal pathology, unstable medical conditions, were pregnant, or had recent surgery. Ultrasound measurements were lost for one participant so that analysis was carried out on the remaining 19 participants.

**Teaching session**

The five examiners attended a two hour teaching session the week before the testing period, where they were introduced to the four examination procedures. The examiners then practiced each technique under supervision.

Following the teaching session students completed a short survey about their level of confidence and preferred physical examination method using a Likert scale. The palpation, percussion, scratch and ballottement methods used in this study to locate the lower liver edge are shown in the appendix.

**Testing period**

Testing took place over two days in 1.5 hour block periods. The mid-clavicular line was measured by a test observer. The test observers were the study’s authors (JB, SH, DG). After the mid-clavicular line (MCL) was determined, a pen mark was made on the MCL, at a random point above the lower costal margin. This was the reference point for the examiners and test observers. Examination subjects lay supine with their knees slightly flexed, a pillow underneath their legs, and their arms by their sides.

The order of examination technique was randomized for each examiner. The four examination techniques performed on each examination participant were scratch, palpation, ballottement, and percussion.

Percussion, ballottement and scratch were performed during normal inspiration. Palpation was performed during deep inspiration. The student examiners were given ten to fifteen minutes to complete all four techniques in each examination participant to localise the lower border of the liver. Further details of each technique are in the appendix. Written instructions for the techniques were provided if the examiners required prompting. A test observer was present in each room. Students were not given feedback as to the accuracy of their perceived liver edge localisation.

For each technique if the examiner was confident that they had detected the lower liver border they measured the distance in centimetres from the marked reference point to the estimated lower border of the liver along the mid-clavicular line with a tape measure as the “Examiners’ Measurement” (EM). If the student examiner didn’t detect a lower liver border this was recorded as such.

After the physical examination, an ultrasound was performed on each examination participant to determine the ultrasound location of the lower border of the liver, the ‘Ultrasound measurement’ (UM). The ultrasound probe
was placed on the mid-clavicular line and the location of the lower border of the liver was measured relative to
the reference point marked on the mid-clavicular line. UM was done during both normal and deep inspiration.

After the two-day test period the student examiners repeated the survey about their level of confidence in
performing these examination techniques and stated which method they preferred.

The main assessment was the difference between the EM and UM. If a student examiner didn’t detect the
position of the lower liver edge then that particular EM was set to be the distance from the reference chest mark
to the costal margin. If a patient’s lower liver edge was also above the costal margin when ultrasound was used,
then the UM for this patient was also the distance from the reference mark to the costal margin. Thus if a student
examiner couldn’t detect a liver edge and the liver edge as determined by ultrasound was above the costal
margin then the difference between EM and UM was set to zero.

**Statistical analysis**

Simple data descriptions are shown for the position of the liver by ultrasound measurement and for the
differences from this measurement by medical student and examination technique.

A mixed linear model with examiners as random effects was used to estimate the mean bias, defined as the
difference between the examiner’s nominated difference and that of the ultrasound measurement, for each
measurement technique. In these models examination technique was treated as a fixed effect and medical
student examiner and the patients who were examined were treated as random effects. A large value of the
medical student random effect in relation to the patient random effect means that there is large variability
induced by different examiners. The variance components were estimated by Restricted Maximum Likelihood
(REML) which gives unbiased estimates of random effects compared to method of moments estimates derived
from conventional Analysis of Variance.

Bland-Altman-like plots are shown for each measurement versus ultrasound. In the plots the symbols reflect the
five different student examiners. A random jitter has been introduced to these plots to distinguish the individual
measurements. In the figures the solid line is the line of no bias compared to ultrasound. The long dashed line is
a mixed linear model estimate of the mean difference between the particular measurement technique and
ultrasound measurement (bias), and the shorter dash lines are the 95% confidence limits for a single prediction
from the mixed linear model representing the limits of agreement.

SAS (version 9.3) software was used to analyse the results.

**Ethics approval**

Category A Ethical Approval was obtained from the University of Otago Human Ethics Committee. Written
informed consent was obtained from both the outpatient examination participants and student examiners after
the provision of study information sheets. There was no financial remuneration for student examiners or
examination participants.

**Results**

The mean (SD) liver position below the reference point was 8.8 (3.3) cm; median (interquartile range) 8 (7 to 12) cm and minimum to maximum 3.5 to 15.5 cm.

Table 1 shows the summary measurements by examiner and technique for all 19 patients.
Table 1. Examiners’ measurement minus ultrasound measurement

<table>
<thead>
<tr>
<th>Scratch</th>
<th>Examiner</th>
<th>Mean (SD)</th>
<th>Median (IQR)</th>
<th>Min to Max</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>0.11 (3.80)</td>
<td>-0.5 (-1.5 to 3.5)</td>
<td>-7.5 to 8.4</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-3.13 (3.40)</td>
<td>-3.0 (-5.5 to 0.0)</td>
<td>-10.5 to 5.0</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>-2.47 (3.36)</td>
<td>-2.5 (-4.5 to -1.0)</td>
<td>-8.5 to 5.5</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>-2.74 (3.85)</td>
<td>-3.0 (-5.0 to 0.0)</td>
<td>-10.5 to 4.0</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>-0.03 (2.94)</td>
<td>0.5 (-2.5 to 1.0)</td>
<td>-5.5 to 5.0</td>
</tr>
<tr>
<td></td>
<td>All</td>
<td>-1.65 (3.78)</td>
<td>-1.4 (-4.0 to 0.5)</td>
<td>-10.5 to 8.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Palpate</th>
<th>Examiner</th>
<th>Mean (SD)</th>
<th>Median (IQR)</th>
<th>Min to Max</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>-0.84 (2.90)</td>
<td>-1.0 (-2.1 to 0.0)</td>
<td>-7.0 to 5.9</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-2.38 (2.93)</td>
<td>-1.0 (-4.2 to 0.0)</td>
<td>-10.5 to 1.0</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>-2.01 (3.03)</td>
<td>-1.0 (-4.2 to 0.0)</td>
<td>-10.5 to 2.0</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>-2.38 (2.95)</td>
<td>-1.0 (-4.5 to 0.0)</td>
<td>-10.5 to 1.0</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>-1.98 (3.34)</td>
<td>-0.5 (-4.5 to 0.0)</td>
<td>-10.5 to 4.0</td>
</tr>
<tr>
<td></td>
<td>All</td>
<td>-1.92 (3.02)</td>
<td>-1.0 (-4.2 to 0.0)</td>
<td>-10.5 to 5.9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Percussion</th>
<th>Examiner</th>
<th>Mean (SD)</th>
<th>Median (IQR)</th>
<th>Min to Max</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>0.11 (3.51)</td>
<td>0.2 (-1.7 to 2.5)</td>
<td>-10.5 to 4.8</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-3.93 (2.77)</td>
<td>-3.5 (-6.0 to -1.5)</td>
<td>-10.5 to 0.5</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>-2.63 (3.34)</td>
<td>-2.5 (-5.0 to -0.2)</td>
<td>-11.5 to 2.5</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>-3.18 (3.62)</td>
<td>-3.0 (-5.5 to 0.0)</td>
<td>-11.5 to 2.5</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0 (3.55)</td>
<td>0 (-1.5 to 2.0)</td>
<td>-10.5 to 6.0</td>
</tr>
<tr>
<td></td>
<td>All</td>
<td>-1.93 (3.70)</td>
<td>-1.5 (-4.2 to 0.5)</td>
<td>-11.5 to 6.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Balloting</th>
<th>Examiner</th>
<th>Mean (SD)</th>
<th>Median (IQR)</th>
<th>Min to Max</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>0.66 (3.61)</td>
<td>0.5 (-1.0 to 4.0)</td>
<td>-8.0 to 6.0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-3.79 (2.84)</td>
<td>-3.5 (-6.0 to -1.5)</td>
<td>-10.5 to 0.0</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>-2.06 (3.06)</td>
<td>-1.0 (-4.2 to 0.0)</td>
<td>-10.5 to 1.5</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>-2.38 (2.95)</td>
<td>-1.0 (-4.5 to 0.0)</td>
<td>-10.5 to 1.0</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>-1.54 (4.06)</td>
<td>-1.0 (-4.5 to 0.0)</td>
<td>-10.5 to 7.0</td>
</tr>
<tr>
<td></td>
<td>All</td>
<td>-1.82 (3.57)</td>
<td>-1.0 (-4.5 to 0.0)</td>
<td>-10.5 to 7.0</td>
</tr>
</tbody>
</table>

N=19 for each examiner and N=95 for All.

Figures 1a to 1d show Bland-Altman-like plots of the differences between measurement techniques compared to ultrasound measurements.
Figure 1. Bland-Altman-like plots for agreement of different examination techniques with ultrasound measurement

Fig 1a: Scratch technique

Fig 1b: Palpation technique
Fig 1c: Percussion technique

![Percussion technique graph]

Fig 1d: Ballottement technique

![Ballottement technique graph]

Note: Symbols represent different medical student examiners, solid line the zero reference line, long-dotted line estimated mean difference from ultrasound measurement, and short-dotted lines the limits of agreement.

Table 2 shows the estimated mean difference between particular measurement techniques and ultrasound measurements and the limits of agreement. There was no evidence that any measurement technique had a different mean difference compared to any other, P=0.94.
Table 2. Estimates of the difference between different measurement techniques and ultrasound and limits of agreement

<table>
<thead>
<tr>
<th>Examination</th>
<th>Estimated mean difference from ultrasound measurement (cm) (95% CI)</th>
<th>Limits of agreement (cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scratch</td>
<td>-1.65 (-3.59 to 0.29)</td>
<td>-8.83 to 5.53</td>
</tr>
<tr>
<td>Palpation</td>
<td>-1.91 (-2.53 to -1.3)</td>
<td>-7.95 to 4.13</td>
</tr>
<tr>
<td>Percussion</td>
<td>-1.93 (-4.25 to 0.39)</td>
<td>-8.67 to 4.81</td>
</tr>
<tr>
<td>Ballottement</td>
<td>-1.82 (-3.83 to 0.19)</td>
<td>-8.5 to 4.86</td>
</tr>
</tbody>
</table>

Table 3 shows the variance components for each measurement technique. For the method of palpation neither the REML, maximum likelihood, or minimum variance quadratic unbiased estimation, could converge on a solution for variance components.

Table 3. Variance components estimates for different physical examination techniques

<table>
<thead>
<tr>
<th>Examination</th>
<th>Patient variance component</th>
<th>Medical students variance component</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scratch</td>
<td>12.9</td>
<td>1.8</td>
</tr>
<tr>
<td>Palpation</td>
<td>9.1</td>
<td>Unable to estimate</td>
</tr>
<tr>
<td>Percussion</td>
<td>11.4</td>
<td>2.9</td>
</tr>
<tr>
<td>Ballottement</td>
<td>11.1</td>
<td>2.0</td>
</tr>
</tbody>
</table>

Discussion

In summary this study finds that irrespective of technique, novice medical students perceive the location of the lower liver border superiorly compared with ultrasound examination. This may be due to the “wedge” shape of the lower liver edge, which could interfere with the initial localisation by physical examination methods at the liver/abdominal interface.

The mean estimate of difference between each examination method and ultrasound measurement was between 1.6 and 1.9 cm superior to the liver ultrasound measurements. This finding that the liver span is underestimated by student examiners in healthy patients is consistent with previous studies, although these studies report that difficulty in determination of the position of the upper border of the liver is the primary factor accounting for underestimation of liver size. The results obtained by the student examiners in our study suggest that inaccuracy in locating the lower liver border can also contribute to underestimation of liver size.

Our results show no difference in bias between the methods with respect to the ultrasound method. Previous research reports mixed results without a consensus as to the most reliable technique. In our study percussion had the least evidence for bias, with the largest P-value, compared to the other methods and also the least evidence, based on the Bland-Altman-like plots, for an increased variability with increasing liver size compared to the scratch, palpation and ballottement methods.

The limits of agreement for all of the methods were wide with the smallest value for palpation (plus or minus 6.04 cm) and the widest for scratch (plus or minus 7.18 cm). Thus from our study it appears the most reproducible technique is palpation, in terms of obtaining similar results from different examiners. The least reproducible is the scratch technique. However the limits of agreement for all measurements were very wide in relation to the mean ultrasound measurement of 8 cm so that all
techniques had quite poor performance. Previous studies have shown mixed results regarding inter-rater reliability. Several studies demonstrate significant variation, especially with the percussion method and the intensity with which it is performed. Gupta et al show high inter-rater agreement when investigating the scratch technique. In our study, where it could be estimated, the variance component due to students was relatively low compared to that due to different patients.

After the initial instruction session and prior to the testing period, the novice student examiners were most confident in the percussion technique; they were least confident with the ballottement technique. There was a significant change following the two day testing period, with student examiners most confident in the scratch technique. They had the least amount of confidence in the palpation technique. When palpating, students frequently commented that they did not know what they were trying to feel, with little preceding clinical exposure and “not having felt a liver before.”

The strengths of study are recruitment of only novice medical students who performed four standard techniques to determine the lower border of the liver. Their lack of experience in hepatic examination was useful in reducing the risk of bias due to previous exposure to the different examination methods. Randomisation of technique sequence was also used to reduce bias. As far as we know this is the first contemporary report of a comparison of these techniques.

Limitations of the study are that the student examiners were not completely naïve to hepatic examination. They had limited exposure to palpation and percussion in clinical teaching although they had never previously performed either technique in a clinical setting involving patients.

This report applies to research participants with no known liver disease. The results may be different in patients with liver disease. Peternal and colleagues report that physician palpation of the liver in patients with confirmed liver disease was less biased and more reliable. The authors attribute these findings to increased hepatic texture in liver disease allowing for easier delineation of the organ. Thus a comparison of hepatic examination techniques in patients with abnormal livers may have different findings. The influence of different body size on liver examination may also be important. Sullivan and colleagues suggest that physical build may influence accuracy in physical examination, particularly in obese patients. We could not examine this in our study.

In summary in novice medical students each of the four examination techniques for the lower liver edge was similarly biased compared to ultrasound measurements but with very wide limits of agreement in relation to the actual position of the lower liver edge. This study might be usefully repeated in more experienced examiners and in patients with diseased livers.

**Competing interests:** Nil.

**Author information:** Selena A Hunter, Joint Clinical Consultant Anaesthetist, Department of Anaesthesia and Pain Management, Wellington Hospital & Senior Lecturer, Department of Surgery and Anaesthesia, School of Medicine and Health Sciences, University of Otago, Wellington; Justin Brimble, Trainee Intern, School of Medicine and Health Sciences, Wellington; Mark Weatherall, Consultant Geriatrician, Rehabilitation Teaching and Research Unit, Fourth-Year Convener for Undergraduate Teaching, Department of Medicine, School of Medicine and Health Sciences, University of Otago, Wellington; Duncan C Galletly, Consultant Anaesthetist, Head of Department, Department of Surgery and Anaesthesia, School of Medicine and Health Sciences, University of Otago, Wellington

**Correspondence:** Dr Selena Hunter, c/o Department of Surgery and Anaesthesia, University of Otago Wellington, 23A Mein Street, Newtown, PO Box 7343, Wellington 6242, New Zealand.

selena.hunter@otago.ac.nz

**References**

17. Sapira JD, Williamson D. How Big is the Normal Liver? Archives of Internal Medicine. 1979;139(9):971-3.