Research article

A systematic review of the diagnostic accuracy of physical examination for the detection of cirrhosis

Guy de Bruyn*[1] and Edward A Graviss[1,2]

Address: [1]Department of Medicine, Baylor College of Medicine, One Baylor Plaza, Houston, TX, 77030, USA and [2]Department of Pathology, Baylor College of Medicine, One Baylor Plaza, Houston, TX, 77030, USA

E-mail: Guy de Bruyn* - gdebruyn@fhcrc.org; Edward A Graviss - egraviss@bcm.tmc.edu

*Corresponding author

Abstract

Background: We conducted a review of the diagnostic accuracy of clinical examination for the diagnosis of cirrhosis. The objectives were: to identify studies assessing the accuracy of clinical examination in the detection of cirrhosis; to summarize the diagnostic accuracy of reported physical examination findings; and to define the effects of study characteristics on estimates of diagnostic accuracy.

Methods: Studies were identified through electronic literature search of MEDLINE (1966 to 2000), search of bibliographic references, and contact with authors. Studies that evaluated indicants from physical examination of patients with known or suspected liver disease undergoing liver biopsy were included. Qualitative data on study characteristics were extracted. Two-by-two tables of presence or absence of physical findings for patients with and without cirrhosis were created from study data. Data for physical findings reported in each study were combined using Summary Receiver Operating Characteristic (SROC) curves or random effects modeling, as appropriate.

Results: Twelve studies met inclusion criteria, including a total of 1895 patients, ranging in age from 3 to 90 years. Most studies were conducted in referral populations with elevated aminotransferase levels. Ten physical signs were reported in three or more studies and ten signs in only a single study. Signs for which there was more study data were associated with high specificity (range 75–98%), but low sensitivity (range 15–68%) for histologically-proven cirrhosis.

Conclusions: Physical findings are generally of low sensitivity for the diagnosis of cirrhosis, and signs with higher specificity represent decompensated disease. Most studies have been undertaken in highly selected populations.

Background

Cirrhosis is a pathologic condition characterized by fibrosis of the liver parenchyma and evidence of regenerative activity [1]. It is a common pathologic end-point for a number of processes, including damage by toxins, metabolic diseases, or virus-induced hepatitis, autoimmune hepatitis, passive congestion secondary to heart failure, or infection with parasites. The epidemiology of cirrhosis is linked to both alcohol consumption [2–5] and the prevalence of hepatitis B and hepatitis C virus infections [5].

Worldwide, cirrhosis is the cause of approximately 850 thousand deaths per year [6]. The majority of these are in developing nations, where viral hepatitis is the main underlying cause of cirrhosis. In developed nations, alcohol is a major cause of cirrhosis, but the early mortality risk attributed to alcohol in younger age groups (15–45 years old) is somewhat offset by the proposed beneficial effect of alcohol on cardiovascular mortality [5].

The diagnosis of cirrhosis has both prognostic and therapeutic implications [7]. The life expectancies for persons diagnosed with cirrhosis are significantly reduced [8–11] compared with age and sex-matched controls, with the exception of certain pathological subtypes [12]. The likelihood of variceal bleeding is reported to be greatest within the first two years of diagnosis [13]. Owing to the high mortality risk associated with variceal hemorrhage, primary prophylaxis against a first bleed becomes an important therapeutic intervention [9]. There are several other therapies that are clinically useful for patients with established cirrhosis [14,15], as well as therapies for specific causes of liver disease, such as viral hepatitis [16].

Cirrhosis may be suspected from physical examination findings [17], with imaging studies frequently supporting the diagnosis. It is known though, that imaging studies are limited in their ability to make a definite assessment in patients with diffuse infiltrative conditions of the liver [18]. Ultrasound, the most widely used modality, is neither sensitive nor specific for the diagnosis of cirrhosis [19]. Estimates of the sensitivity range from 57 to 92% and specificity between 39 and 81% [20,21]. Despite certain limitations, liver biopsy remains the gold standard for diagnosis. This procedure is not without risk. There is a small mortality risk of around 0.015 percent, together with the resulting discomfort and other non-fatal complications [13,22]. The patchy or regional nature of hepatic changes means that percutaneous liver biopsies may miss the diagnosis of cirrhosis in 24 percent of patients with the condition [23].

Given the invasive nature of the diagnostic standard, there has been interest in whether non-invasive diagnosis of cirrhosis might be possible [13,24]. Physical examination is a diagnostic tool available to every clinician. This study was conducted to review the published literature on clinical examination of patients with known or suspected liver disease and identify the validity and generalizability of primary studies that distinguished those with cirrhosis from those without. A summary of diagnostic test performance for elements of the physical examination was derived using meta-analytic methods for diagnostic studies. Where possible, effects of study characteristics on summary estimates were determined by subgroup analysis.

Methods

Study identification

Two online searches of the National Library of Medicine MEDLINE database were performed using the PubMed search engine. The searches covered the years 1966–2000. The last search was completed on August 15, 2000. The first search utilized the Clinical Queries tool, using the term "cirrhosis", and the options "diagnosis" and "specificity". A second search combined medical subject headings (MeSH) "Signs and Symptoms", "Physical Examination", and "Medical History Taking", linked by OR terms, with a text word search for "cirrhosis", through an AND term. The resulting set was further limited by the term "human", and by excluding reviews, editorials, or letters. The titles of the resulting citations were scanned. Abstracts of potentially relevant articles were then retrieved. If the article appeared likely to satisfy inclusion criteria from information available in the abstract, or if no abstract was available, the full text article was evaluated. Additional articles were sought by scanning bibliographies in standard reference textbooks, proceedings of meetings or conferences on cirrhosis or liver disease, the reference sections of selected articles or review articles on liver disease, cirrhosis, or physical examination, or other potentially relevant chapters in books. The authors of primary studies identified through literature searches were contacted by letter or by email or both, seeking data not presented in the published study, and to enquire about knowledge of unpublished or additional studies.

Study selection criteria

Articles were included if they: evaluated patients with known or suspected liver disease or cirrhosis; evaluated items obtained by physical examination; described liver biopsy results; and provided data sufficient to calculate sensitivity and specificity of elements of the clinical examination in the detection of cirrhosis. Studies that evaluated only biochemical or radiological examinations were excluded.

Validity assessment

The validity of included articles was evaluated for the following four criteria [25,26] by one author (GdB): independent, blind comparison against a reference standard of diagnosis; evaluation in an appropriate spectrum of patients; the reference standard was applied, regardless of diagnostic test (in this case, physical examination) result; and the reference standard was measured prior to starting any interventions based on the examination findings.

The reference standard for diagnosis of cirrhosis was histological evidence of irreversible chronic injury to liver parenchyma with extensive fibrosis and formation of regenerative nodules. This definition conforms to the definition used in the study from which the sensitivity of per-

cutaneous biopsy for diagnosis of cirrhosis was obtained [23].

Data extraction
Primary data abstracted from included studies by one author (GdB). The data was arranged in 2×2 contingency tables with participants classified according to the presence or absence of a physical sign, given the presence or absence of cirrhosis as defined by liver biopsy. Qualitative and quantitative study characteristics regarding validity, clinical description of the study population, description of the examination methods, and applicability, were extracted onto data forms.

Calculation of summary statistics
Primary study data for each physical finding was analyzed using a computer program, Meta-Test version 0.6 (J Lau, New England Medical Center, Boston 1997). The meta-analytic approach followed the recommendations of Midgette *et al*[27] and are similar to methods outlined in a recent review [28]. The correlation between sensitivity and (1-specificity) was tested using Spearman's rank correlation coefficient. Where sensitivity and (1-specificity) were significantly correlated (Spearman's $\rho > 0.5$), a Summary Receiver Operating Characteristic (SROC) curve was used to summarize the data. Where the correlation between sensitivity and (1-specificity) was <0.5 or negative, homogeneity of sensitivity and specificity was tested by chi-square tests of association. If sensitivity and specificity were both homogeneous, a summary estimate of sensitivity and specificity were independently derived using a random-effects model (REM). If the sensitivity and specificity were heterogeneous, no summary estimate was derived. The SROC curves were created using the method described by Moses *et al*[29,30]. The SROC curve depicts the trade-off between sensitivity and specificity across different studies with explicit or implicit variation in test threshold. Unweighted regression analyses were examined to determine the SROC curve. Symmetry of the SROC curve was confirmed by determining whether the slope of the fitted regression line differed significantly from zero. The point of maximum joint sensitivity and specificity was determined from the curve, as an overall summary measure. In some instances, the SROC curve did not extend to a point where sensitivity is equal to specificity. In those instances, the REM summary estimates were used to describe the data. Statistical significance was assigned to a *P*-value less than 0.05.

Subgroup analyses
Where a minimum of three studies per subgroup were present, subgroup analyses were planned to determine the effects of covariates on estimates of diagnostic accuracy. Covariates postulated to be of possible significance were: study size (<100 vs. $\geq$ 100), independence of examination

and reference standard (4 point scale – 1 = test measured independent of reference standard, and reference standard measured independent of test, 2 = test measured independent of reference but not *vice versa*, 3 = reference standard measured independent of test, but not *vice versa*, 4 = neither test nor reference standard measured independently of each other), study design (prospective, retrospective, or experimental), and primary etiology of cirrhosis in the study population (alcohol, viral hepatitis, or other). Residuals of the unweighted regression line describing the SROC curve were grouped by the covariate of interest. Student's t-test was used to compare the subgroups [29,31].

Results
Results of literature search
The trial flow is presented in Figure 1. The literature searches identified 3451 citations. Of these, 44 titles appeared relevant and were retrieved. Thirteen articles [32–44] of potential significance were selected for review. A further 16 articles [13,22,24,45–57] were found through hand searching bibliographic references. Twelve articles [37,38,40,41,44–47,49,51,53,57] met inclusion criteria. Of the 17 studies excluded, eight studies did not report data for physical signs, two enrolled only cirrhotic patients, five studies did not include sufficient data to calculate summary measures of diagnostic accuracy, and one was a review article. Two studies grouped data for patients with cirrhosis together with data for patients with other histological lesions.

Study characteristics
Details of clinical characteristics of the study population for each included study are presented in Additional Table (see Additional file 1: Additional Table). Included studies enrolled a total of 1895 patients. Patients ranged in age from 3 to 90 years. The ten studies that reported gender composition included 1168 males (65%) and 628 females (37%). The median study size was 101 patients, with an interquartile range of 59 to 229. Most of the included studies took place in referral populations (*n* = 10). Detection of persistent elevation of aminotransferase levels was the most frequent eligibility criterion for study entry (*n* = 5). Five studies reported an estimated duration of illness, which varied from a total duration of illness of less than six months to a mean of 30 months. Comorbid illness was mentioned in one study [38] that enrolled patients with inherited coagulation disorders, usually hemophilia. Fifteen of these patients (44%) were also infected with the Human Immunodeficiency Virus (HIV).

The principal causes of liver disease, as reported in each included study, are presented in Table 1. Three studies mainly enrolled patients with chronic viral hepatitis [38,40,44], four mainly with alcohol- related liver disease
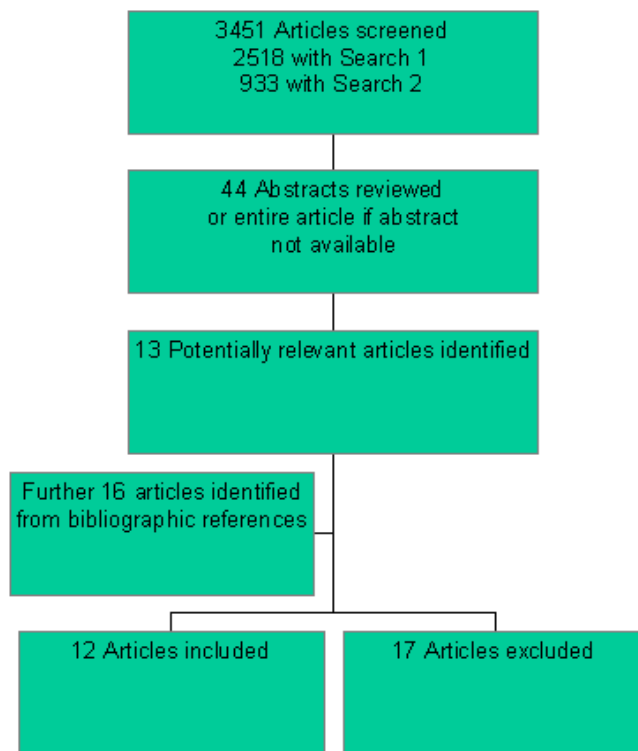
Figure 1
Trial Flow Of Studies.

[37,49,51,53], and one with predominantly autoimmune hepatitis [47]. The remainder reported a presumed etiology for less than 50% of the study population.

Aspects of study design and validity are presented in Table 2. Details regarding the expertise of the clinician performing the biopsy were not provided in any of the studies. However, it is reasonable to infer that the biopsy was performed by expert clinicians in at least ten of the studies, namely those that originated in referral centers. Eleven of twelve studies reported that either the clinical examination was performed prior to the biopsy *and* that blinded pathologists read the biopsy, or that at least the examination was performed prior to the biopsy. It was generally not stated how many examiners reviewed the biopsies (*n* = 7). The reliability of the pathological diagnosis was assessed in a single study [57]. These investigators reported that biopsies were examined by two experienced pathologists. Good agreement beyond chance (κ = 0.67) was present for the pathological diagnoses in this study. Three other studies stated that a single examiner reviewed the slides [46,47,51] and one stated that experience pathologists reviewed the slides [45].

In nine studies, the biopsy was performed irrespective of the findings on clinical examination. In the case-control

study [44], the control group did not undergo liver biopsy. In two studies [38,49], it was not clear whether the decision to perform the biopsy was taken independently of clinical data. Five studies reported on exclusions [37,38,40,53,57]. These ranged from 12.7% (44 of 347) to 39.3% (22 of 56). Common reasons for exclusion were: the presence of contraindications to biopsy, the patient did not undergo biopsy, or inadequate biopsy.

In all studies, examiners were not blinded to other clinical information. In some cases, the entire history and examination was performed together. In other cases, multiple physical signs were evaluated. Only two studies gave any details regarding the specific clinical methods used [44,46], and four mentioned the explicit threshold for abnormality of a clinical finding [44,46,51,57]. Five studies did not state how many examiners assessed each patient, four studies reported that a single examiner was used, and three studies used two examiners to evaluate physical signs. Reliability was assessed in four studies [41,44,46,57], although results were only provided in one study [57], and for only three clinical indicants. These investigators found good agreement beyond chance for liver firmness (κ = 0.72), spleen size (κ = 0.66), and stated that agreement for liver enlargement was similar (κ not stated).

Accuracy of history and physical examination

Ten physical signs or findings were reported in three of more studies. SROC curves were constructed for splenomegaly and hepatomegaly (Figures 2 and 3). The data for ascites were unsuitable for SROC analysis, and so were combined using summary measures from a REM. The maximum joint sensitivity and specificity of the finding of splenomegaly was 0.64, and 0.75 for hepatomegaly.

A sensitivity analysis was performed to evaluate the effect of inclusion of those studies most liable to various types of bias. When the case-control study [44] was excluded from the analysis of the hepatomegaly data, the SROC curve was essentially unchanged. A joint maximum sensitivity and specificity for the revised hepatomegaly SROC curve was 0.75 (data not shown). By comparison, exclusion of the study in which clinical and pathological assessments were not independent [45] from the splenomegaly dataset resulted in a dramatic change. The SROC curve for this data was greatly influenced by the presence of this outlying point. Random-effects estimates for the revised splenomegaly data of sensitivity were 0.82 (95% CI: 0.52 – 0.95) and specificity 0.60 (95% CI: 0.21 – 0.89).

Data on seven physical signs were combined using summary measures derived from a REM, as shown in Table 3. The data for palmar erythema was heterogeneous and unsuitable for meta-analysis. Reported sensitivity ranged

Table 1: Principal etiology of liver disease for patients enrolled to included studies, according to study classification

FIRST AUTHOR	ALCOHOL (%)	CHRONIC VIRAL HEPATITIS (%)	AUTOIMMUNE (%)	CRYPTOGENIC (%)
Schenker	64	-	-	-
Nakamura	100	-	-	-
Rankin	100	-	-	-
Cozzolino	-	20†	-	-
Hay CRM	-	71‡	-	-
Lashner	-	58	-	42
Hay JE	-	-	53	47*
Tinè	9	15†	-	-
Marmo	25	23†	-	-
Zoli	21	53	12	11
Hamberg	100	1	-	-

*21% possibly non-alcoholic steatohepatitis (NASH) † Hepatitis B surface antigen positive ‡ Classified as non-A, non-B hepatitis

Table 2: Design and validity criteria of twelve studies that met entry criteria

FIRST AUTHOR	YEARS OF STUDY ENROLLMENT	N	DESIGN	INDEPENDANCE OF MEASUREMENTS*	APPLICATION OF BIOPSY
Schenker	1958–1959	61	Prospective	2	All
Nakamura	1951–1961	52	Retrospective	2	All
Rankin	NS	100	Prospective	1	All
Czaja	NS	101	Prospective	1	All
Cozzolino	NS	213	Retrospective	4	All
Hay CRM	1977–1986	35	Retrospective	2	All
Lashner	1977–1981	94	Retrospective	1	All
Hay JE	1984–1987	47	Prospective	1	All
Tinè	1984–1985	277	Prospective	1	All
Marmo	1986–1990	412	Retrospective	2	All
Zoli	NS	200	Case-control	2	Control patients not biopsied
Hamberg	1968–1971	303	Prospective	2	All

*1, test measured independant of reference standard, and reference standard measured independant of test, 2, test measured independant of reference but not vice versa, 3, reference standard measured independant of test, but not vice versa, 4, neither test nor reference standard measured independantly of each other. NS, not stated.

from 0.12 to 0.63 and specificity ranged from 0.49 to 0.98. A further ten physical signs were evaluated in single studies. The diagnostic accuracy for these findings is represented in Table 4.

Subgroup analyses of SROC analyses
Subgroup analyses were possible for splenomegaly and ascites. None of the covariates examined, namely: study size, study design, etiology of cirrhosis, and independence of measurements, resulted in a significant difference in diagnostic performance of the physical sign in question. There was significant heterogeneity of TPR and TNR in the subgroup of studies reporting splenomegaly that had as-

sessed biopsy findings with knowledge of the clinical findings.

Discussion
The present study attempted to summarize data from primary sources in the published medical literature on the diagnostic accuracy of clinical findings for the diagnosis of cirrhosis. The studies identified represent a rather select population. Studies were primarily conducted in referral centers, by clinicians experienced in dealing with liver disease, seeing patients with increased clinical suspicion of the disorder of interest. The underlying etiology of liver disease was not comprehensively reported, and in several
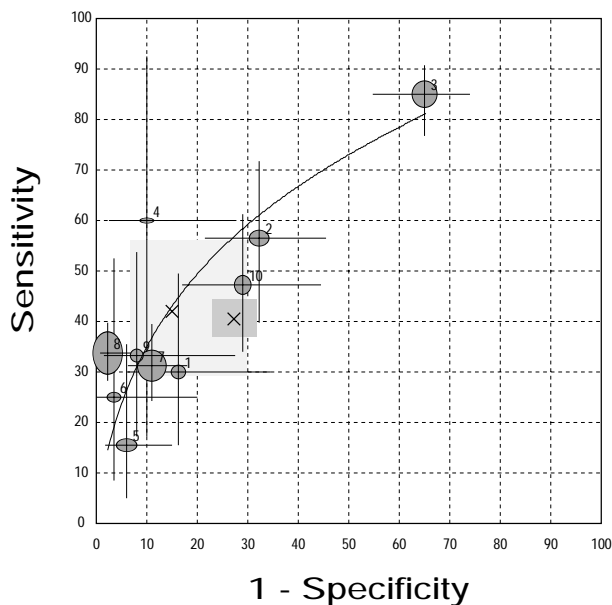
Figure 2
Summary receiver operating characteristic curve of the diagnostic accuracy of splenomegaly in the diagnosis of histologically-proven cirrhosis. Numbers refer to studies: 1 = Schenker [53], 2 = Czaja [46], 3 = Cozzolino [45], 4 = Hay CRM [38], 5= Lashner [40], 6 = Hay JE [47], 7 = Tinè [57], 8 = Marmo [41], 9 = Nakamura [49], 10 = Rankin [51]. Light gray box depicts point estimate (cross) with 95% confidence limits for random effects model estimates of sensitivity and specificity. Dark gray box depicts fixed effects model point estimates and confidence limits of estimates of sensitivity and specificity.

studies, no description of the patient population was given. Few studies described their clinical methods, where several exist for eliciting a finding [58–60]. These deficiencies are deficiencies that should be carefully addressed in designing future studies that evaluate aspects of clinical practice, such as the history or physical examination. These deficiencies are a phenomenon that is not restricted to studies of clinical evaluation of liver disease, adding support to the call for more rigorous studies of the clinical examination [61].

The summary measures derived from the data presented here demonstrate that the physical findings elicited in the primary studies have high specificity, though generally low sensitivity, for detecting cirrhosis. As expected, physical signs consistent with advance portal hypertension or severe liver dysfunction (decompensated cirrhosis), like ascites, abdominal wall veins (collateral circulation), or encephalopathy, had high specificity of over 90%. There was significant heterogeneity in the reported diagnostic accuracy of palmar erythema, precluding meta-analysis.

In the case of the physical signs splenomegaly and hepatomegaly, studies indicated the presence of a variation in test threshold, as an explanation for the variation in test accuracy. In the case of splenomegaly, however, this finding was sensitive to the inclusion of one study with potential for bias of the diagnostic accuracy of clinical evaluation.

Subgroup analyses were not able to detect variation in accuracy as a consequence of study characteristics. The small number of studies available conferred low power to detect such differences. Data were insufficient to examine other study characteristics that may have influenced study outcome, such as age of participants, cause of liver disease, or symptom duration. Increases in age or duration of illness may bias studies towards higher prevalence of cirrhosis.

The present study employed a comprehensive search strategy. Formal criteria for study inclusion were defined prior to analysis of the search results. We were unable to find previous attempts to summarize the accuracy of clinical examination for diagnosis of cirrhosis in the medical literature. Our results agree however, with comparisons spanning five decades of the antemortum clinical diagnosis with subsequent autopsy studies [62].

This study has several limitations. Firstly, a question remains on the choice of diagnostic reference standard. Liver biopsy is likely to underestimate the actual prevalence of cirrhosis [50]. Methods to overcome this limitation include examining multiple biopsies. One study required three successive negative biopsies to exclude the diagnosis of cirrhosis [46]. Although this approach addresses the concern of sampling variability, it is questionable whether it could be considered in future studies using biopsy as the gold standard, as it may not be ethically defensible. This remains an area for improvement in future studies, and consensus on standardization.

Misclassification of patients by an imperfect reference test will lead to bias in the assessment of a diagnostic test [63,64]. In general, an imperfect reference test will underestimate the performance of a diagnostic test. Sample calculations to show how such errors in a reference test lead to alterations in the apparent sensitivity and specificity of a diagnostic test are presented in Appendix 1 and 2 (see Additional file 2: Appendices). These calculations are based on methods developed by Gart and Buck [65]. As can be seen, a diagnostic test with true sensitivity of 0.9 would appear to have a sensitivity of only 0.612 if the reference standard against which it was measured had an imperfect sensitivity and specificity of 0.8 and 0.95, respectively. Meta-analytic methods have been described to adjust for imperfections in the reference standard, al-
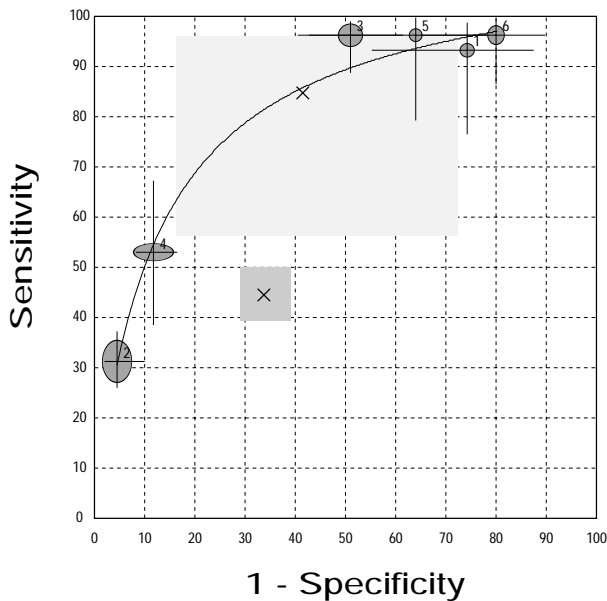
Figure 3
Summary receiver operating characteristic curve of the diagnostic accuracy of hepatomegaly in the diagnosis of histologically-proven cirrhosis. Numbers refer to studies: I = Schenker [53], 2 = Marmo [41], 3 = Zoli [44], 4 = Hamberg [37], 5 = Nakamura [49], 6 = Rankin [51]. Light gray box depicts point estimate (cross) with 95% confidence limits for random effects model estimates of sensitivity and specificity. Dark gray box depicts fixed effects model point estimates and confidence limits of estimates of sensitivity and specificity.

though those techniques were not applied to the data presented here [66].

Secondly, it is likely that other types of bias were present in some of these studies. Empirical observation of the quantitative effects of study design flaws on the findings of diagnostic studies has shown that case-control designs, studies that use different reference tests for positive and negative results of the diagnostic test under study, and lack of blinding led to overestimation of diagnostic test accuracy [67]. One case-control study was retrieved in the searches [44]. No important difference was noted in the hepatomegaly SROC curve, whether this particular study was included or not. One retrospective study appeared to have a significant risk of bias from lack of blinding [45]. The exclusion of this study from the splenomegaly SROC significantly altered the analysis of this data, showing the vulnerability of the meta-analysis to this form of study bias.

Several other forms of bias have not been shown to be important predictors of variation in assessment of diagnostic

accuracy [67]. Selection bias refers to the bias that may result if not all patients with the condition under study are consecutively included. Such a bias may have influenced the results of at least one study (reference 42). Verification bias refers to the bias that may occur if the reference test is applied based on the results of the diagnostic test being studied. Several studies were described as having exclusions, which raised the possibility of selective application of the gold standard. Statistical methods for evaluating diagnostic test performance in the presence of such partial verification have been described, but require the application of a second initial diagnostic test [68]. Consequently, such an approach was not used in this paper.

Thirdly, the data obtained were insufficient to assess the effect of study location on diagnostic accuracy. The decrease in specificity of physical signs as patients are referred from primary to tertiary care settings, in which most of these studies were performed, has been described [61]. As most of the studies utilized persistently elevated aminotransferase levels as eligibility criteria, rather than the physical signs under study, no data are available to evaluate whether over- or underrecognition by the referring clinicians led to an altered specificity in results of the cited article.

Fourth, the available data did not allow an analysis of how individual patient characteristics affect diagnostic accuracy of physical findings. It has been observed that cirrhosis undetected during life was a finding typical of elderly patients [69]. Similarly, the prevalence of certain physical findings, or alternatively the ability to detect those signs, may vary between patients of different race [70].

Fifth, this study cannot assess the independence of isolated clinical findings. The role and diagnostic accuracy of items from the clinical history were not examined. It is likely that some of the study clinicians were aware of indicants from the clinical history. Whether these elements of the evaluation made clinicians examine patients more intensively than those with a negative history is unknown. The possibility also exists that finding one physical sign may have heightened attempts by examining clinicians to elicit further signs. If clinicians utilized additional maneuvers, hence deviating from their usual course of clinical examination, on the strength of finding other more easily elicited signs, the study results may have become biased. The appropriateness of examining an isolated clinical sign may also be challenged [61].

Sixth, data from the primary studies are limited on the specific clinical maneuvers used to derive the estimates of diagnostic accuracy. Thus, the findings cannot be related to the accuracy of a specific clinical approach.

Table 3: Summary diagnostic accuracy derived from random-effects models

FINDING	STUDIES	SENSITIVITY	95%CI	SPECIFICITY	95% CI
Ascites	7	0.34	0.22–0.49.	0.95	0.89–0.98
Collateral circulation	3	0.42	0.26–0.61	0.94	0.71–0.99
Encephalopathy	3	0.15	0.06–0.33	0.98	0.97–0.99
Firm liver	3	0.68	0.55–0.79	0.75	0.62–0.85
Jaundice	3	0.36	0.25–0.48	0.85	0.80–0.89
Spider angiomata	8	0.50	0.39–0.61	0.88	0.75–0.95

Table 4: Sensitivity and specificity of selected physical findings reported in a single study.

FINDING	FIRST AUTHOR	SENSITIVITY	95%CI	SPECIFICITY	95%CI
Facial telangectasia	Hamberg	0.82	0.68–0.91	0.92	0.88–0.95
Gynaecomastia	Hamberg	0.18	0.09–0.32	0.97	0.94–0.99
Irregular liver surface	Marmo	0.52	0.46–0.58	0.93	0.87–0.97
Parotidomegaly	Schenker	0.13	0.04–0.32	1.00	0.89–1.00
Peripheral edema	Hamberg	0.24	0.14–0.39	0.91	0.87–0.94
Sharp liver edge	Zoli	0.49	0.37–0.60	0.21	0.11–0.36
Tender liver	Schenker	0.33	0.18–0.53	0.65	0.45–0.80
Testicular atrophy	Hamberg	0.18	0.09–0.32	0.97	0.94–0.99
Thick liver edge	Zoli	0.51	0.40–0.63	0.79	0.64–0.89
White nails	Hamberg	0.43	0.29–0.58	0.98	0.95–0.99

This study may have implications for how clinicians perform and interpret their clinical examination findings. Perloff recently drew attention to the ongoing importance of skillful clinical evaluation in cardiovascular disease [71]. These data show that in patients with cirrhosis, individual clinical signs usually have low sensitivity, with high specificity. Thus, physical signs cannot be used to exclude the diagnosis of cirrhosis, given these performance characteristics. Signs are likely to be more useful to clinicians in the discrimination of patients with disease by identifying patients with a higher likelihood of cirrhosis, thereby "ruling-in" the diagnosis [25]. The main obstacle to the timely use of such an approach is the finding that signs with the highest specificity are those present in decompensated disease.

This study has implications for clinical research. This study has highlighted the selected nature of existing data on diagnostic accuracy of clinical examination and measurement parameters for cirrhosis. Attempts to further define diagnostic accuracy of clinical examination will have to consider how to define and measure independently the contribution of individual covariates or physical signs,

and which combination of cofactors has the highest sensitivity, specificity, or positive predictive value. A potentially useful consideration would be the gain in sensitivity and specificity that might result from using a combination of signs. Further attention should be devoted to finding ways to identify patients with cirrhosis. Non-invasive laboratory testing such as serologic or molecular markers, imaging studies, or improved biopsy techniques appear to be the principal avenues open to innovation in this area.

Conclusions

1. Most studies examining the diagnostic accuracy of clinical examination have been undertaken in highly selected populations.

2. Physical signs are generally of low sensitivity for the diagnosis of cirrhosis, and signs with higher specificity are associated with clinically decompensated disease, thus no rules of generality can be deduced.

3. Physical signs are not useful for excluding the presence of cirrhosis, but may be useful to indicate the presence of cirrhosis in a person with moderate to high pretest suspi-

cion of disease. Physical signs need to be utilized with additional clinical criteria to augment the probability of identifying patients with cirrhosis.

List Of Abbreviations Used

AT Aminotransferase

HIV Human immunodeficiency Virus

MeSH Medical Subject Heading

NA Not available

NASH Non-alcoholic steatohepatitis

NS Not stated

REM Random effects model

SGOT Serum glutamic-oxalacetic transaminase

SD Standard deviation

Sn Sensitivity

Sp Specificity

SROC Summary Receiver Operating Characteristic

TRC Tertiary referral center

ULN Upper limit of normal

Competing interests

None declared.

Additional material

Additional file 1
Containing Appendix 1 and Appendix 2 with sample calculations of effect on apparent diagnostic test sensitivity and Specificity of imperfections in the reference standard.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1472-6947-1-6-S1.doc]

Additional file 2
Containing Table: clinical characteristics of patients enrolled in included studies.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1472-6947-1-6-S2.doc]

References
1. Podolsky DK, Isselbacher KJ: **Cirrhosis and alcoholic liver disease.** In: Harrison's Principles of Internal Medicine Edited by Longo DL, 14th ed. pp. 1704–1710. New York: McGraw-Hill; 19981704-1710
2. Corrao G, Ferrari P, Zambon A, Torchio P, Arico S, Decarli A: **Trends of liver cirrhosis mortality in Europe, 1970–1989: age-period-cohort analysis and changing alcohol consumption.** Int J Epidemiol 1997, **26**:100-109
3. Halliday ML, Coates RA, Rankin JG: **Changing trends of cirrhosis mortality in Ontario, Canada, 1911–1986.** Int J Epidemiol 1991, **20**:199-208
4. Roizen R, Kerr WC, Fillmore KM: **Cirrhosis mortality and per capita consumption of distilled spirits, United States, 1949–94: trend analysis.** Bmj 1999, **319**:666-670
5. Murray CJ, Lopez AD: **Global mortality, disability, and the contribution of risk factors: Global Burden of Disease Study.** Lancet 1997, **349**:1436-1442
6. Murray CJ, Lopez AD: **Mortality by cause for eight regions of the world: Global Burden of Disease Study.** Lancet 1997, **349**:1269-1276
7. Schalm SW: **The diagnosis of cirrhosis: clinical relevance and methodology.** J Hepatol 1997, **27**:1118-1119
8. Gentilini P, Laffi G, La Villa G, Romanelli RG, Buzzelli G, Casini-Raggi V, Melani L, Mazzanti R, Riccardi D, Pinzani M, Zignego AL: **Long course and prognostic factors of virus-induced cirrhosis of the liver.** Am J Gastroenterol 1997, **92**:66-72
9. Nevens F, Broeckaert L, Rutgeerts P, Van Steenbergen W, Fevery J: **The long-term morbidity and mortality rate in a cohort of patients with liver cirrhosis and oesophageal varices.** Hepatogastroenterology 1995, **42**:979-984
10. Propst A, Propst T, Zangerl G, Ofner D, Judmaier G, Vogel W: **Prognosis and life expectancy in chronic liver disease.** Dig Dis Sci 1995, **40**:1805-1815
11. Realdi G, Fattovich G, Hadziyannis S, Schalm SW, Almasio P, Sanchez-Tapias J, Christensen E, Giustina G, Noventa F: **Survival and prognostic factors in 366 patients with compensated cirrhosis type B: a multicenter study. The Investigators of the European Concerted Action on Viral Hepatitis (EUROHEP).** J Hepatol 1994, **21**:656-666
12. Roberts SK, Therneau TM, Czaja AJ: **Prognosis of histological cirrhosis in type 1 autoimmune hepatitis.** Gastroenterology 1996, **110**:848-857
13. Oberti F, Valsesia E, Pilette C, Rousselet MC, Bedossa P, Aube C, Gallois Y, Rifflet H, Maiga MY, Penneau-Fontbonne D, Cales P: **Noninvasive diagnosis of hepatic fibrosis or cirrhosis.** Gastroenterology 1997, **113**:1609-1616
14. Miglio F, Valpiani D, Rossellini SR, Ferrieri A: **Rifaximin, a non-absorbable rifamycin, for the treatment of hepatic encephalopathy. A double-blind, randomised trial.** Curr Med Res Opin 1997, **13**:593-601
15. Watanabe A, Sakai T, Sato S, Imai F, Ohto M, Arakawa Y, Toda G, Kobayashi K, Muto Y, Tsujii T, Kawasaki H, Okita K, Tanikawa K, Fujiyama S, Shimada S: **Clinical efficacy of lactulose in cirrhotic patients with and without subclinical hepatic encephalopathy.** Hepatology 1997, **26**:1410-1414
16. Camma C, Giunta M, Pinzello G, Morabito A, Verderio P, Pagliaro L: **Chronic hepatitis C and interferon alpha: conventional and cumulative meta-analyses of randomized controlled trials.** Am J Gastroenterol 1999, **94**:581-595
17. McIntyre N: **Symptoms and signs of liver disease.** In: Oxford Textbook of Clinical Hepatology Edited by Rodes J, vol. 1. pp. 479–499. Oxford: Oxford Medical Publishers; 1999479-499
18. Mergo PJ, Ros PR: **Imaging of diffuse liver disease.** Radiol Clin North Am 1998, **36**:365-375
19. Albrecht T, Blomley MJ, Cosgrove DO, Taylor-Robinson SD, Jayaram V, Eckersley R, Urbank A, Butler-Barnes J, Patel N: **Non-invasive diagnosis of hepatic cirrhosis by transit-time analysis of an ultrasound contrast agent.** Lancet 1999, **353**:1579-1583
20. Gosink BB, Lemon SK, Scheible W, Leopold GR: **Accuracy of ultrasonography in diagnosis of hepatocellular disease.** AJR Am J Roentgenol 1979, **133**:19-23

21. Sanford NL, Walsh P, Matis C, Baddeley H, Powell LW: **Is ultra-sonography useful in the assessment of diffuse parenchymal liver disease?** *Gastroenterology* 1985, **89**:186-191

22. Van Ness MM, Diehl AM: **Is liver biopsy useful in the evaluation of patients with chronically elevated liver enzymes?** *Ann Intern Med* 1989, **111**:473-478

23. Nord HJ: **Biopsy diagnosis of cirrhosis: blind percutaneous versus guided direct vision techniques – a review.** *Gastrointest Endosc* 1982, **28**:102-104

24. Poynard T, Aubert A, Naveau S, Bedossa P, Abella A, Chaput JC: **A computer protocol to predict liver cirrhosis in drinkers [abstract].** *J Hepatol* 1988, **7** supple:S70

25. Sackett DL, Richardson WS, Rosenberg W, Haynes RB: *Evidence-based Medicine: how to practice and teach EBM. New York: Churchill Livingstone;* 1997

26. Irwig L, Tosteson AN, Gatsonis C, Lau J, Colditz G, Chalmers TC, Mosteller F: **Guidelines for meta-analyses evaluating diagnostic tests.** *Ann Intern Med* 1994, **120**:667-676

27. Midgette AS, Stukel TA, Littenberg B: **A meta-analytic method for summarizing diagnostic test performances: receiver-operating-characteristic-summary point estimates.** *Med Decis Making* 1993, **13**:253-257

28. Deeks JJ: **Systematic reviews in health care: Systematic reviews of evaluations of diagnostic and screening tests.** *Bmj* 2001, **323**:157-162

29. Moses LE, Shapiro D, Littenberg B: **Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations.** *Stat Med* 1993, **12**:1293-1316

30. Littenberg B, Moses LE: **Estimating diagnostic accuracy from multiple conflicting reports: a new meta-analytic method.** *Med Decis Making* 1993, **13**:313-321

31. Irwig L, Macaskill P, Glasziou P, Fahey M: **Meta-analytic methods for diagnostic test accuracy.** *J Clin Epidemiol* 1995, **48**:119-130

32. Amarapurkar DN, Parikh SS, Shankaran K, Chopra K, Dhawan P, Kalro RH, Desai HG: **Correlation between splenomegaly and oesophageal varices in patients with liver cirrhosis.** *Endoscopy* 1994, **26**:563

33. Angeli P, Albino G, Carraro P, Dalla Pria M, Merkel C, Caregaro L, De Bei E, Bortoluzzi A, Plebani M, Gatta A: **Cirrhosis and muscle cramps: evidence of a causal relationship.** *Hepatology* 1996, **23**:264-273

34. Angulo P, Keach JC, Batts KP, Lindor KD: **Independent predictors of liver fibrosis in patients with nonalcoholic steatohepatitis.** *Hepatology* 1999, **30**:1356-1362

35. Cioni G, Tincani E, D'Alimonte P, Cristani A, Ventura P, Abbati G, Vignoli A, Romagnoli R, Ventura E: **Relevance of reduced portal flow velocity, low platelet count and enlarged spleen diameter in the non-invasive diagnosis of compensated liver cirrhosis.** *Eur J Med* 1993, **2**:408-410

36. Espinoza P, Ducot B, Pelletier G, Attali P, Buffet C, David B, Labayle D, Etienne JP: **Interobserver agreement in the physical diagnosis of alcoholic liver disease.** *Dig Dis Sci* 1987, **32**:244-247

37. Hamberg KJ, Carstensen B, Sorensen TI, Eghoje K: **Accuracy of clinical diagnosis of cirrhosis among alcohol-abusing men.** *J Clin Epidemiol* 1996, **49**:1295-1301

38. Hay CR, Preston FE, Triger DR, Greaves M, Underwood JC, Westlake L: **Predictive markers of chronic liver disease in hemophilia.** *Blood* 1987, **69**:1595-1599

39. Iturriaga H, Hirsch S, Bunout D, Diaz M, Kelly M, Silva G, de la Maza MP, Petermann M, Ugarte G: **[Prediction of histological liver damage in asymptomatic alcoholic patients by means of clinical and laboratory data].** *Rev Med Chil* 1993, **121**:369-378

40. Lashner BA, Jonas RB, Tang HS, Evans AA, Ozeran SE, Baker AL: **Chronic hepatitis: disease factors at diagnosis predictive of mortality.** *Am J Med* 1988, **85**:609-614

41. Marmo R, Romano M, de Sio I, Peduto A, Caporaso N, Persico M, Coltorti M: **Decision-making model for a non-invasive diagnosis of compensated liver cirrhosis.** *Ital J Gastroenterol* 1993, **25**:1-8

42. Pilette C, Croquet V, Vuillemin E, Oberti F, Cales P: **[Non-invasive and early diagnosis of liver cirrhosis].** *Gastroenterol Clin Biol* 1999, **23**:557-565

43. Ratziu V, Giral P, Charlotte F, Bruckert E, Thibault V, Theodorou I, Khalil L, Turpin G, Opolon P, Poynard T: **Liver fibrosis in overweight patients.** *Gastroenterology* 2000, **118**:1117-1123

44. Zoli M, Magalotti D, Grimaldi M, Gueli C, Marchesini G, Pisi E: **Physical examination of the liver: is it still worth it?** *Am J Gastroenterol* 1995, **90**:1428-1432

45. Cozzolino G, Lonardo A, Francica G, Amendola F, Cacciatore L: **Differential diagnosis between hepatic cirrhosis and chronic active hepatitis: specificity and sensitivity of physical and laboratory findings in a series from the Mediterranean area.** *Am J Gastroenterol* 1983, **78**:442-445

46. Czaja AJ, Wolf AM, Baggenstoss AH: **Clinical assessment of cirrhosis in severe chronic active liver disease: specificity and sensitivity of physical and laboratory findings.** *Mayo Clin Proc* 1980, **55**:360-364

47. Hay JE, Czaja AJ, Rakela J, Ludwig J: **The nature of unexplained chronic aminotransferase elevations of a mild to moderate degree in asymptomatic patients.** *Hepatology* 1989, **9**:193-197

48. Martin JB, Apostolakos PC, Roazen H: **Clinical versus actuarial prediction in the differential diagnosis of jaundice.** *Am J Med Sci* 1960, **240**:571-578

49. Nakamura T, Nakamura S, Suzuki O, Aikawa T, Onodera A, Karoji N: **Clinical studies of alcoholic hepatic diseases.** *Tohoku J Exp Med* 1967, **93**:179-189

50. Pagliaro L, Rinaldi F, Craxi A, Di Franco C, Di Piazza S, Fillippazzo G, Magrin S, Maringhini A, Palazzo U, Vinci M: **Predictive value of clinical, laparoscopic and histologic indicators of cirrhosis.** *Ital J Gastroenterol* 1982, **14**:166-168

51. Rankin JG, Orrego-Matte H, Deschenes J, Medline A, Findlay JE, Armstrong AI: **Alcoholic liver disease: the problem of diagnosis.** *Alcohol Clin Exp Res* 1978, **2**:327-338

52. Sanchez GC, Baunsgaard P, Lundborg CJ: **A comparison between clinical diagnosis and histopathological findings in liver biopsies.** *Scand J Gastroenterol* 1980, **15**:985-991

53. Schenker S, Balint J, Schiff L: **Differential diagnosis of jaundice: report of a prospective study of 61 proved cases.** *Am J Dig Dis* 1962, **7**:449-463

54. Teran JC, Bashour FN, Sondhi SS, Mullet KD: **Validation of a discriminant score for the diagnosis of cirrhosis in patients with chronic hepatitis [abstract].** *Gastoenterology* 1995, **108** suppl:1185

55. Theodossi A, Spiegelhalter D, Portmann B, Eddleston AL, Williams R: **The value of clinical, biochemical, ultrasound and liver biopsy data in assessing patients with liver disease.** *Liver* 1983, **3**:315-326

56. Theodossi A: **The value of symptoms and signs in the assessment of jaundiced patients.** *Clin Gastroenterol* 1985, **14**:545-557

57. Tine F, Caltagirone M, Camma C, Cottone M, Craxi A, Filippazzo MG, Malizia G, Palazzo U, Pinzello GB, Pisa R, Vinci M, Vizzini GB, Pagliaro L: **Clinical indicants of compensated cirrhosis: a prospective study.** *In: Chronic liver damage Edited by Gentilini P. pp. 187–198. Amsterdam: Elsevier Science Publishers;* 1990187-198

58. Naylor CD: **Physical examination of the liver.** *Jama* 1994, **271**:1859-1865

59. Grover SA, Barkun AN, Sackett DL: **Does this patient have splenomegaly?** *Jama* 1993, **270**:2218-2221

60. Williams JW Jr, Simel DL: **Does this patient have ascites? How to divine fluid in the abdomen.** *Jama* 1992, **267**:2645-2648

61. McAlister FA, Straus SE, Sackett DL: **Why we need large, simple studies of the clinical examination: the problem and a proposed solution. CARE-COAD1 group. Clinical Assessment of the Reliability of the Examination-Chronic Obstructive Airways Disease Group.** *Lancet* 1999, **354**:1721-1724

62. Anderson RE, Hill RB, Key CR: **The sensitivity and specificity of clinical diagnostics during five decades. Toward an understanding of necessary fallibility.** *Jama* 1989, **261**:1610-1617

63. Boyko EJ, Alderman BW, Baron AE: **Reference test errors bias the evaluation of diagnostic tests for ischemic heart disease.** *J Gen Intern Med* 1988, **3**:476-481

64. Valenstein PN: **Evaluating diagnostic tests with imperfect standards.** *Am J Clin Pathol* 1990, **93**:252-258

65. Gart JJ, Buck AA: **Comparison of a screening test and a reference test in epidemiologic studies. II. A probabilistic model for the comparison of diagnostic tests.** *Am J Epidemiol* 1966, **83**:593-602

66. Walter SD, Irwig L, Glasziou PP: **Meta-analysis of diagnostic tests with imperfect reference standards.** *J Clin Epidemiol* 1999, **52**:943-951

67. Lijmer JG, Mol BW, Heisterkamp S, Bonsel GJ, Prins MH, van der Meulen JH, Bossuyt PM: **Empirical evidence of design-related bias in studies of diagnostic tests.** *Jama* 1999, **282**:1061-1066

68. Walter SD: **Estimation of test sensitivity and specificity when disease confirmation is limited to positive results.** *Epidemiology* 1999, **10**:67-72

69. Graudal N, Leth P, Marbjerg L, Galloe AM: **Characteristics of cirrhosis undiagnosed during life: a comparative analysis of 73 undiagnosed cases and 149 diagnosed cases of cirrhosis, detected in 4929 consecutive autopsies.** *J Intern Med* 1991, **230**:165-171

70. O'Keefe SJ, Simjee AE, Seedat YK: **Clinical presentation and biochemical abnormalities in black (Zulu) patients with cirrhosis in Durban.** *S Afr Med J* 1982, **61**:775-778

71. Perloff JK: **The jugular venous pulse and third heart sound in patients with heart failure.** *N Engl J Med* 2001, **345**:612-614