The IDEA Journal
International Digestive Experts Articles

Year II - N. 3, 2020



mediAbout

# The IDEA Journal

International Digestive Experts Articles

QUARTERLY SCIENTIFIC UPDATE IN GASTROENTEROLOGY

Year II - N. 3, 2020





**PUBLISHER**
MEDIABOUT S.r.l.
Via Olmetto, 5 - 20123 Milan, Italy
E-mail: info@mediabout.it
www.mediabout.it

**PUBLISHING DIRECTOR**
Mauro Rissa

**EDITORIAL BOARD**
Arjan Bredenoord, Amsterdam
Cesare Hassan, Rome
Mette Vesterhus, Bergen

**EDITORIAL SECRETARIAT**
MEDIABOUT S.r.l.

**GRAPHIC LAYOUT**
MEDIABOUT S.r.l.

**PRESS**
Galli Thierry Stampa S.r.l.
Via Caviglia, 3 - 20139 Milan, Italy

Registration at Court of Milan
n. 260, 09/12/2019



© Copyright 2020 MEDIABOUT S.r.l.
All rights reserved. No part of this publication may be reproduced without the prior written permission of the publisher.

# EDITORIAL STANDARDS

## AIMS AND SCOPE

The International Digestive Experts Articles Journal is a peer-reviewed journal on prevention, diagnosis and management of gastrointestinal disorders. The journal is primarily intended for the update of gastroenterologists, hepatologists, and all medical doctors focusing their practice on digestive pathologies.

The International Digestive Experts Articles Journal encourage the submission of spontaneous manuscripts, when in line with the aims and scope of the journal. Specific topics articles can be covered by invitation of the Board.

## AUTHORS INSTRUCTION

Submitted manuscripts will be reviewed by two or more experts who will be asked to evaluate whether the manuscript is scientifically sound and coherent, and whether or not the manuscript is sufficiently clear for publication.

Copyright of accepted articles will be transferred to the Publisher. Reproduction is permitted for non-commercial purposes, provided that the source is correctly acknowledged.

### √ Preparing the manuscript

The type of articles published by the International Digestive Experts Articles Journal include:
- Congress Reports
- Original Papers
- Hot Topics in Current Clinical Practice
- Case Reports

All manuscripts must be submitted by e-mail to the Publisher [info@mediabout.it]. Only Microsoft Word documents will be accepted.

The **first page** of the manuscript must include authors' full names, authors' institutions, and corresponding author contact information (including phone and e-mail).

The **second page** of the manuscript must contain the abstract (including background and aims, methods, results, conclusion or appropriately structured), 3-5 keywords, abbreviation list, and the conflict of interest declaration (such as, but not limited to, employment, consultancies, stock ownership or options, honoraria, patents, and paid expert testimony) of all the authors.

When appropriate, the manuscript must be structured in sections as background and aims, methods, results, conclusion or using appropriate headings where needed.

The full generic name of pharmaceutical active ingredients must be used.

All units of measurement must follow the International System of Units. Please use a point to separate the fractional part of decimal numbers.

All acronyms and abbreviations must be spelled out when first used.

### √ Bibliography

All references must be numbered consecutively in the order in which they are cited in the text, followed by any in tables or legends. The reference numbers must be finalized and the reference list fully formatted before submission. Reference list must be formatted as follows:

Kinner S, Reeder SB, Yokoo T. Quantitative Imaging Biomarkers of NAFLD. Dig Dis Sci. 2016 May; 61(5):1337-1347.

### √ Tables and Figures

Tables can be included in the main manuscript file, before Bibliography.

Please submit figures as separate file in JPEG, TIFF, EPS, or PDF formats (300 PPI resolution).

# *Editorial*

# Novel advances in Gastroenterology (COVID-19-triggered or not!)

The first paper of this Issue of the IDEA Journal is dedicated to an exhaustive review by Giulio Antonelli and Cesare Hassan of the application of Artificial Intelligence (AI) to the detection and characterization of colorectal polyps and on the unmet needs that AI is (and will be) able to tackle. At the moment, the main reasons for colorectal neoplasia miss rate are both failure in recognising a lesion although fully visible on the endoscopy screen, due to attention or recognition issues, and failure to expose enough colorectal mucosa. While the first depends on the endoscopist examination technique and the eventual use of add-on devices, failure to recognise a polyp when visible on the endoscopy screen can be addressed and improved by the application of AI, or "deep learning" systems.

Liver elastography in cholestatic liver diseases is the topic of another interesting paper. Anders Batman Mjelle presents here the current data demonstrating how liver stiffness measurement is useful for staging and outcome prediction in cholestatic liver disease and to provide a clinical introduction to the use of liver stiffness measurement in cholestatic liver diseases. Liver stiffness measurement is a method used in the follow-up of patients with viral hepatitis and constitutes an increasingly important part of the prognostic armamentarium in the health care for patients with chronic liver diseases of different aetiologies.

Finally, don't miss Arjan Bredenoord comments on the paper by Schuitenmakers et al. on COVID-19-triggered ways of performing research in a virtual way (which can guarantee trials to continue in the COVID-19 era and will be attractive afterwards) and on the complex relationship among achalasia, obstructive motor disorders and eosinophilic esophagitis, based on a recent paper by Ghisa and colleagues.

# Artificial intelligence in the detection and characterisation of colorectal polyps: a review of recent literature

Giulio Antonelli, Cesare Hassan
*Gastroenterology Unit, Nuovo Regina Margherita Hospital, Rome, Italy*

## INTRODUCTION

Colonoscopy and polypectomy are the mainstay in the prevention of colorectal cancer (CRC), and have been shown to reduce its incidence and mortality (1-3). The development of quality improvement programs and performance measures, their measurement with audit and eventual retraining have led to an increase in adenoma detection rate (ADR), directly associated with a decrease in interval cancer (i.e. a cancer arising in the interval time between two screening colonoscopies) (4-6).

Notwithstanding the increasing awareness and the ever-improving quality, a substantial rate of colorectal neoplasia is still missed during colonoscopy, variably reported between 5 and 25%, leading to a interval colorectal cancer rate ranging between 0.5 and 1 per 1000 person-years (7).

The main reasons identified for colorectal neoplasia miss rate are both failure in recognising a lesion although fully visible on the endoscopy screen, due to attention or recognition issues, and failure to expose enough colorectal mucosa. While mucosal exposure depends on the endoscopist examination technique and the eventual use of recently developed add-on devices, failure to recognise a polyp when visible on the endoscopy screen can be addressed and improved by the application of Artificial Intelligence (AI), or "deep learning" systems (8,9).

Contrary to human-programmed computer systems, "deep learning" systems autonomously learn to distinguish the characteristics within the images provided using multiple levels of processing (10). In this way, AI systems are able to recognize discriminatory characteristics between images that differ from those commonly used and elaborated by the human mind. In addition, an AI system developed with deep learning techniques can acquire an image processing speed so fast that it can be used in real time during an endoscopic examination. Consequently, AI systems have been recently developed that are capable of flagging with a box the suspect area during the endoscopic examination, coupled with a sound alarm. These systems have shown a high accuracy when retrospectively applied to still images or stored videos, and more recently have been tested in trials in real time during endoscopic examinations (10).

The other domain in which AI is believed to have a considerable impact on everyday clinical practice is lesion characterization and aid in "optical diagnosis". When considering the magnitude of colonoscopies performed, covering between 1% and 6% of the target general population per year, the financial and economic burden is relevant (11). A relevant contribution of such burden is represented

by the post-polypectomy histology cost, mostly attributed to diminutive polyps that represent over 90% of all the resected lesions (12-14). By predicting in vivo histology, such pathology costs may be averted by the implementation of two separate but synergistic strategies, namely the 'leave-in-situ' strategy for <5 mm hyperplastic lesions in the rectosigmoid tract, and 'resect and discard' for the other diminutive lesions (15,16). Despite the very promising accuracy rates by experts, the accuracy of optical diagnosis in community setting has been suboptimal, preventing the implementation of these cost-saving interventions (17,18). In addition, the clinical relevance of these lesions has been debated, being mostly represented by either non-advanced adenomas or indolent hyperplastic polyps. In addition, the role of pathology as reference standard has been questioned because of possible high interobserver agreement, inadequate orientation or insufficient material (19).

By automatizing the perception phase, AI may overcome both of these pitfalls in detection and characterization. Namely, Computer Aided Detection (CADe) based on deep learning can recognize in real-time lesions that are present on the screen and that may have been missed by the endoscopist. Similarly, Computer Aided Characterization (CADx) can predict the histology of the lesion providing the correct classification to the endoscopist (8).

We present here an overview of recent literature regarding the real-time clinical application of CADe and CADx for colorectal neoplasia.

## CADe SYSTEM DEVELOPMENT PROCESS

Briefly, when developing a deep learning system the goal is to build a mathematical model (algorithm) from a set of humanly marked data (images, videos) that will be able to accurately interpret new data that the system has never seen before. Deep learning systems start from libraries of labelled data (in this case, images containing a polyp), from which they autonomously build their own algorithms ("learn") acquiring parameters that will make them capable of recognizing a polyp in an image they have never encountered (10).

Artificial neural networks develop algorithms from increasing amounts of data and can continuously improve their discriminatory abilities. The neural network accumulates multiple layers ("deep") over time and improves its performance proportionally to the depth of the network. Deep learning can be supervised or non-supervised by humans. Currently, most systems are developed with human supervision, meaning that the machine is provided with data that has been labelled by humans. In unsupervised deep learning, which is to be considered the gold standard, the systems are provided with unlabelled data. This is currently more apt to research purposes as it aims mostly at identifying the structure of the data. Convolutional neural networks are formed when an artificial neural network receives external signals on a layer of input nodes (processing units), each of which is connected with numerous internal nodes, organized in several levels. Each node processes the received signals and transmits the result to subsequent nodes working in parallel (10).

When developing an AI system with the purpose of polyp detection, we can summarise the development phases in a training phase, a validation phase and a testing phase. In the training phase, a very large number of images labelled for the regions/features of interest are presented to the system, that learns to recognise the labelled features building its own algorithms. The system is then initially tested on another set of unknown images, the validation set, in which the performance is evaluated, and the system is fine-tuned by the use of "hyperparameters", optional settings that are calibrated by the programmers to optimise the system's performance. Lastly, a third, unseen set of data (the "test" set) is presented to the system, to evaluate its standalone performance. Ideally, the test set should be a library of unseen images that come from different centres or at least completely different patients from those of the training and validation sets.

As previously stated, after the system has obtained a satisfying level of performance on a test set, the last step is to evaluate its performance it in a real life clinical trial, possibly with randomisation and

control, where it will have to face the pitfalls of clinical practice, like suboptimal preparation, patient compliance, operator skills, etc.

## COMMON BIASES IN CADe DEVELOPMENT

Every phase of a CADe system development is susceptible to many possible biases that can undermine the building of an effective system (10,20). Of course, many technical (i.e. engineering) biases can occur even before the training phase that are of little interest (or comprehension) for clinicians. When we set out to train and test a system for a particular task, we should nevertheless be aware of the different kind of development biases, that can be avoided or limited implementing some measures that can reduce the risk.

In a recent meta-analysis, for each phase of CADe systems development, the most common forms of bias that can arise and the methods by which they can be avoided or limited were identified. Furthermore, in this meta-analysis, a modified version of the QUADAS score has been proposed that can be used effectively to assess the presence of bias in studies (20).

In the training phase, three main kinds of bias have been identified, namely selection bias, spectrum bias and operator bias.

- *Selection bias*: arises when the cases selected for training (images, videos) have not been randomly or consecutively selected, generating a selection of the best cases available. For example, only polyps with a "standard", polypoid appearance. This can be avoided or limited by selecting all consecutive cases of polyps seen in the centre, regardless of any other consideration.

- *Spectrum bias*: arises when for each given case, in our case, a polyp, only the best frames have been selected. For example, only frames in which the polyp is at the centre of the screen may lead the system to skip lesions when they are in the periphery of the frame, or if we select only frames in which the polyp is fully visible with no debris or water, that may lead to the system to miss lesions when they are dirty or partially covered in faeces or mucus. To limit or avoid this bias we suggest to include every image of a given polyp regardless of the image quality or of limitations that the human operator might feel as exclusion criteria. It must be kept in mind that it is precisely the use of completely different recognition parameters from those used by the human mind that constitutes the added value of AI systems.

- *Operator bias*: arises when data acquisition is undergone by a small number of operators, or when few operators are involved in data annotation (i.e. underlining a polyp within an image). This could result in an unbalanced selection of cases due to the tendency of an operator to choose certain characteristics and not others, choices that would then be "transferred" and made to learn to the artificial intelligence system.

In the testing phase two main domains have been identified:

- *Overfitting bias*: "Overfitting" is a specific term that refers to the phenomenon that may occur when the model is too "tightly fitted" to the training data and does not generalise to new data (ie, the model only works for the given training examples). Overfitting can be recognised by high training performance combined with low test performance. When an excessively low number of images are used for the training phase it may happen that the system "memorizes" instead of "learning". It can also occur if the system is left for too much time to learn from a given image dataset.

Other "clinical" forms of overfitting may occur if the system is tested on frames that were also included in the training phase, which is to be considered a severe bias and should raise doubts on the accuracy of the system. Another less severe form of the same bias is if the system is tested on different images but coming from the same center and/or operators that produced and marked the training sets. This may limit the external validation of the system.

- *Operator bias*: We identified this clinical bias that is present whenever accuracy data for a system is not compared to accuracy data of independent observers reassessing the ground truth. Whenever a system is tested, the same image testing dataset should be evaluated by independent observers that have never seen the images before, so to have an independent

benchmarking. This can also effectively evaluate the performance of human operators, and can be stratified for experience, or for type of training.

## CADe: EVIDENCE FROM RETROSPECTIVE STUDIES

The initial experiences in programming CADe systems date back to the early 2000s when a first system based on color analysis was proposed and tested (21). As noted previously, this was still a human programmed system that was intrinsically limited by long processing time and consequently impossibility to apply in real time. The real "great leap forward" was made with the successful application of Convolutional Neural Networks (CNN) and deep learning. As already pointed out, these systems can recognize polypoid features without a continuous external input after a proper training.

In 2017, the first algorithm able to outperform expert endoscopists in polyp detection accuracy was published (86% vs 74%), reporting also a 98% sensitivity using a dataset which contained images with polyps mixed to an immense number of images from any natural domain (nature, mountains, etc.) Subsequently, many other CNN for CADe experiences were published: one system, trained from 1.8 million of frames of polyps achieved a sensitivity of 90% with a specificity of 63.3% (22). Another algorithm was trained with a reduced number of carefully pre marked images achieving an accuracy of 96.4% (23). The novel contribution of this study was the elegant design that included 11 videos and 73 polyps deliberately missed by the endoscopist, that were subsequently identified by the AI system in 67/73 cases, with a false positive rate of 5%.

The first European regulatory approved CADe system on the market, GI-Genius (Medtronic) also was initially evaluated with a retrospective design (24), before moving on to a randomized controlled trial (25). Initially, using a dataset of 2684 histologically confirmed neoplastic polyps manually annotated by expert endoscopists and listed in 1.5 million frames, the system showed an overall sensitivity per lesion was 99.7%, and interestingly showed a reaction lag time that was faster than the endoscopists in 82% of cases.

Recently, a novel version of EndoBRAIN-EYE, a Japanese regulatory approved CADe system that was already tested in a randomized trial (26), was perfected by using a large-scale database of publicly available colonoscopy videos to be used for validation purposes (27). This methodology and the choice of developing a publicly available database would ideally permit to evaluate the performance of different CADe systems on the same datasets of images and videos. This dataset, named "SUN-database" contained 49,799 polyp frames annotated with bounding boxes and 102,761 frames without polyps, making a total of 152,560 frames. In this case, the AI algorithm detected 98 of 100 test polyps; thus, per-polyp sensitivity was 98.0% (95% CI, 93.0%-99.8%). Regarding per-frame analysis, the sensitivity and specificity were 90.5% (95% CI, 90.2%-90.7%) and 93.7% (95% CI, 93.5%-93.8%), respectively. Although this is still a retrospective experience, the unique choice of developing and using a publicly accessible database seems to be a step forward towards greater standardization and generalizability of AI studies. Furthermore, they only used images derived from videos, that may be more suitable as learning material as they better resemble real-life experiences.

## CADe: EVIDENCE FROM CLINICAL TRIALS: DETECTION RATE

After many publications focusing on the stand-alone performance of AI systems, mainly retrospectively on datasets of still data, recently many publications, often in high ranking journals, have reported results of the first randomised controlled trials testing the performance of CADe systems in real-life clinical practice, mostly from Chinese groups (25,28-34).

All but one of these trials have chosen the so-called "parallel" design, in which a patient is randomised either to the standard colonoscopy group or to the intervention (standard colonoscopy + CADe) group. Only one study, one of the most recently published,

has chosen a "tandem" design, and it will be discussed separately for its slightly different implications.

Interestingly, no clinical trial showed differences in colonoscopy withdrawal time between groups undergoing CADe examinations and control.

All published trials showed an increase in ADR among the CADe groups: Wang et al. (28) reported a significant increase in ADR (p<0.001) from 20% in the control group to 29% in the CADe group. Liu et al. and Su et al. (30,32) both reported a significant increase in ADR, 39% vs 24%, p<0.001 and 29% vs. 17%, p<0.001, respectively. All these trials had the limitation of a somewhat low ADR in the control group, raising concerns about whether AI might compensate, albeit partially, a poor operator technique. However, a trial by Repici et al. (25) showed a very high baseline ADR at 40.4% in the control group, that was outmatched by a 54.8% ADR in the CADe group.

A recent meta-analysis of published randomised control trials (35) has shown that the increase in ADR was consistent across all trials. Among the included 4354 patients, the ADR in the control groups was 25.2%, vs an ADR of 36.6% in the CADe groups, with a RR of 1.44 [95% CI, 1.27-1.62; p<0.01].

When looking at the characteristics of the adenomas detected in the two groups, sub-analysis reveal that the increase in ADR was mainly due to an increase in small adenomas (less than 5 mm), that was detected across all studies included in the meta-analysis. When looking at advanced adenomas (>10mm), no study reported an increase in ADR, while only Repici et al. showed an increase in the detection of adenomas between 6 and 9 mm (12.7% vs 17.2%, p<0.05) (35). Only one study by Gong et al. (31), among available literature, that however was not included in the reported meta-analysis, has shown a role of CADe in increasing advanced adenoma detection rate. Indeed, this study reports an AADR of 3% vs 1% in the control group, that resulted statistically significant. However, this study has a very low overall ADR (8% in the control group, 16% in the CADe group) raising concerns about the consistency of these results.

Regarding Polyp Detection Rate, meta-analysis have shown (35,36) that it is overall significantly improved in CAD groups: (50.3% vs 34.6%; RR 1.43; 95% CI, 1.34-1.53; p<0.01). Furthermore, when looking at the Adenoma per Colonoscopy Rate, CADe showed to improve APC, irrespectively of polyp size: overall APC, 0.58 vs. 0.36, RR: 1.70; 95% CI, 1.53-1.89; p<0.01; RR <5 mm: 1.69; 95% CI, 1.48-1.84 vs RR 6-9 mm: 1.44, 95% CI, 1.19-1.75 vs RR >10 mm: 1.46, 95% CI, 1.04-2.06.

Lastly, the detection of serrated lesions was improved by CADe among randomized studies reporting this value: 0.06 vs. 0.04, RR: 1.52; 95% CI, 1.14-2.02; p<0.01(36).

## CHARACTERIZATION OF COLORECTAL NEOPLASIA: EVIDENCE FROM CLINICAL TRIALS

CADx is the other promising field of clinical application of AI to colonoscopy. While the human operator depends on the application of virtual or physical chromo-endoscopy to improve visualisation of mucosal and vascular patterns and accurately predict lesion histology, an AI system adequately trained on a wide library should be able to predict histology regardless of the optical visualisation modality (36).

Currently, no randomised clinical trial is available evaluating performance of detection systems. However, many systems are under development and their standalone performance has been evaluated.

A recent metanalysis has summarised existing literature, showing how among the only 3 prospective studies on CADx (36), AI showed an impressive 92.3% (95% CI, 88.8%-94.9%) sensitivity on polyp histology prediction and a high specificity: 89.8% (95% CI, 85.3%-93.0%). Among the considerable number of retrospective studies similar pooled results were found. It is important to notice that the majority of these systems are shallow learning systems that are being abandoned in favour of deep learning systems, and that solid data from randomised trials, using real-life images will be needed before a true estimate of CADx performance can be made. The performance during a live colonoscopy is of course the main interest around this kind of system, where

pitfalls such as inadequate bowel preparation or incomplete lesion visualisation and interrogation can be much more common.

An interesting sub-domain of CADx is the comparison between the optical diagnosis performance of expert endoscopists and non-expert endoscopists, using or not using CADx. It is well known that an extensive training is needed for an endoscopist to achieve acceptable results in predicting in vivo histology of encountered lesions, and that this knowledge has to be regularly updated and retrained. According to available evidence, AI performs similarly to expert endoscopists but better than non-expert endoscopists in lesion characterisation (33). Therefore, a significant improvement of non-experts' performance through CADx could be of great interest both for training and for quality assurance.

## FUTURE PERSPECTIVES AND CONTROVERSIES

A possible drawback of CADe is the potential large number of false positive results (37). As discussed previously, CADe systems autonomously learn their own detection algorithms and therefore its outcomes incorporate some unpredictability in the clinical setting that must be interpreted cautiously. Indeed, the system may flag frames that the endoscopists may never have selected as suspicious areas and consequently reduce colonoscopy efficiency. The endoscopist might spend an excessive amount of time to discriminate between an actual false positive and a possible false negative result. Furthermore, although areas flagged by CADe must always be interpreted by a trained endoscopists, it is still possible that a false positive area may result in unnecessary polypectomy with related avoidable adverse events. In a recent study (37) authors underwent a post-hoc analysis of a RCT on CADe performance, and have measured false positive burden and clinical relevance and classified false positives in two broad categories: artefacts from bowel wall and artefacts from bowel content. Overall, they found a mean of 27.3 false positive activations per colonoscopy, with nearly 90% of them due to artefacts from the bowel wall (folds, ileocecal valve, diverticula, appendicular foramen, etc.). Interestingly, according to their measurements, less than 10% of the false positive activations resulted in additional time spent by the endoscopist in examining the flagged area, while the majority were instantly dismissed as not relevant. These results will need to be confirmed with other systems and other settings.

Another domain in which CADe performance has yet to be improved is the detection of non-polypoid lesions (38). It is well known that flat colorectal lesions account for a large portion of missed colorectal neoplasia and may be associated with a more aggressive biological behaviour. A recent review (38) has shown that among the published RCTs on CADe systems, some of them did not report the number of flat lesions included in the training sets, and others did not report sub-analysis on the performance of AI specifically regarding flat lesions. The authors conclude that in future CADe systems development and refinement, additional training and validation for the recognition of the individual subtypes of non-polypoid lesions, especially for LST-NG, is urgently needed. The authors speculate that a joint partnership between Eastern and Western centres should be prioritized to create datasets with a large number of flat lesions.

In the era of colonoscopy quality measurement and improvement, CAD systems that can integrate quality measurement and reporting have been initially evaluated (4,31). The indicators that have been measured with CAD are caecal intubation, withdrawal time, and even slipping of the scope that can leave areas of the colon uninspected.

The cost-effectiveness of CAD systems has yet to be fully analysed. Only one preliminary study has been published so far by Mori et al. (39) focusing on the implementation of AI alongside a "diagnose and leave behind" strategy, and showing that this can lead to substantial cost reductions. In detail, they estimated that the use of AI could save 119 dollars, 52 dollars, 34 dollars, and 125 dollars per colonoscopy, and up to 149.2 million dollars, 12.4 million dollars, 1.1 million dollars, and 85.2 million dollars for the annual reimbursement for colonoscopies conducted

under public health insurances in Japan, England, Norway, and the United States, respectively. Further cost-effectiveness models could further tailor this analysis, for example in the setting of organised screening programs, that in most of the Western world currently account for the greater part of the colonoscopy burden in public health systems.

A considerable improvement in AI-aided colonoscopy should be the implementation of systems that integrate CADe and CADx in the same machine (40). This could reduce costs and increase the practical considerations regarding everyday clinical use. Randomised trials and cost-effectiveness models combining the additional detection provided by CADe to the optical diagnosis improvement provided by CADx could pave the way to a swift implementation of these systems in clinical practice.

**References**

1. Atkin W, Wooldrage K, Brenner A, Martin J, Shah U, Perera S, et al. Adenoma surveillance and colorectal cancer incidence: a retrospective, multicentre, cohort study. The Lancet Oncology. 2017 Jun;18(6):823-34.
2. Brenner H, Arndt V, Stürmer T, Stegmaier C, Ziegler H, Dhom G. Long-lasting reduction of risk of colorectal cancer following screening endoscopy. Br J Cancer. 2001 Sep;85(7):972-6.
3. Brenner H, Stock C, Hoffmeister M. Effect of screening sigmoidoscopy and screening colonoscopy on colorectal cancer incidence and mortality: systematic review and meta-analysis of randomised controlled trials and observational studies. BMJ. 2014 Apr 9;348:g2467.
4. Kaminski MF, Thomas-Gibson S, Bugajski M, Bretthauer M, Rees CJ, Dekker E, et al. Performance measures for lower gastrointestinal endoscopy: a European Society of Gastrointestinal Endoscopy (ESGE) Quality Improvement Initiative. Endoscopy. 2017 Apr;49(4):378-97.
5. Kaminski MF, Regula J, Kraszewska E, Polkowski M, Wojciechowska U, Didkowska J, et al. Quality Indicators for Colonoscopy and the Risk of Interval Cancer. New England Journal of Medicine. 2010 May 13;362(19):1795-803.
6. Corley DA, Jensen CD, Marks AR, Zhao WK, Lee JK, Doubeni CA, et al. Adenoma Detection Rate and Risk of Colorectal Cancer and Death. New England Journal of Medicine. 2014 Apr 3;370(14):1298-306.
7. Zhao S, Wang S, Pan P, Xia T, Chang X, Yang X, et al. Magnitude, Risk Factors, and Factors Associated With Adenoma Miss Rate of Tandem Colonoscopy: A Systematic Review and Meta-analysis. Gastroenterology. 2019 May;156(6):1661-1674.e11.
8. Bisschops R, East JE, Hassan C, Hazewinkel Y, Kamiński MF, Neumann H, et al. Advanced imaging for detection and differentiation of colorectal neoplasia: European Society of Gastrointestinal Endoscopy (ESGE) Guideline - Update 2019. Endoscopy. 2019;51(12):1155-79.
9. de Groen PC. Using artificial intelligence to improve adequacy of inspection in gastrointestinal endoscopy. Techniques in Gastrointestinal Endoscopy. 2019 Oct; 150640.
10. van der Sommen F, de Groof J, Struyvenberg M, van der Putten J, Boers T, Fockens K, et al. Machine learning in GI endoscopy: practical guidance in how to interpret a novel field. Gut. 2020 May 11;gutjnl-2019-320466.
11. Greuter MJE, de Klerk CM, Meijer GA, Dekker E, Coupé VMH. Screening for Colorectal Cancer With Fecal Immunochemical Testing With and Without Postpolypectomy Surveillance Colonoscopy: A Cost-Effectiveness Analysis. Ann Intern Med. 2017 Oct 17;167(8):544-54.
12. Gupta N, Bansal A, Rao D, Early DS, Jonnalagadda S, Wani SB, et al. Prevalence of advanced histological features in diminutive and small colon polyps. Gastrointest Endosc. 2012 May;75(5):1022-30.
13. Laish I, Sergeev I, Stein A, Naftali T, Konikoff FM. Risk of metachronous advanced lesions after resection of diminutive and small, non-advanced adenomas. Clin Res Hepatol Gastroenterol. 2019;43(2):201-7.
14. Schachschal G, Sehner S, Choschzick M, Aust D, Brandl L, Vieth M, et al. Impact of reassessment of colonic hyperplastic polyps by expert GI pathologists. Int J Colorectal Dis. 2016 Mar;31(3):675-83.
15. Hassan C, Pickhardt PJ, Rex DK. A Resect and Discard Strategy Would Improve Cost-Effectiveness of Colorectal Cancer Screening. Clinical Gastroenterology and Hepatology. 2010 Oct;8(10):865-869.e3.

16. Rex DK, Kahi C, O'Brien M, Levin TR, Pohl H, Rastogi A, et al. The American Society for Gastrointestinal Endoscopy PIVI (Preservation and Incorporation of Valuable Endoscopic Innovations) on real-time endoscopic assessment of the histology of diminutive colorectal polyps. Gastrointest Endosc. 2011 Mar;73(3):419-22.
17. Rex DK. Can we do resect and discard with artificial intelligence-assisted colon polyp "optical biopsy?" Techniques in Gastrointestinal Endoscopy. 2019 Oct; 150638.
18. Vu HT, Sayuk GS, Hollander TG, Clebanoff J, Edmundowicz SA, Gyawali CP, et al. Resect and discard approach to colon polyps: real-world applicability among academic and community gastroenterologists. Dig Dis Sci. 2015 Feb;60(2):502-8.
19. Vennelaganti S, Cuatrecasas M, Vennalaganti P, Kennedy KF, Srinivasan S, Patil DT, et al. Moderate Inter-Observer Agreement Among Pathologists in Differentiation of Sessile Serrated From Hyperplastic Polyps. Gastroenterology. 2020 Sep 17.
20. Arribas J, Antonelli G, Frazzoni L, Fuccio L, Ebigbo A, van der Sommen F, et al. Standalone performance of artificial intelligence for upper GI neoplasia: a meta-analysis. Gut. 2020 Oct 30.
21. Karkanis SA, Iakovidis DK, Maroulis DE, Karras DA, Tzivras M. Computer-aided tumor detection in endoscopic video using color wavelet features. IEEE Trans Inf Technol Biomed. 2003 Sep;7(3):141-52.
22. Misawa M, Kudo S-E, Mori Y, Cho T, Kataoka S, Yamauchi A, et al. Artificial Intelligence-Assisted Polyp Detection for Colonoscopy: Initial Experience. Gastroenterology. 2018;154(8):2027-2029.e3.
23. Urban G, Tripathi P, Alkayali T, Mittal M, Jalali F, Karnes W, et al. Deep Learning Localizes and Identifies Polyps in Real Time With 96% Accuracy in Screening Colonoscopy. Gastroenterology. 2018;155(4):1069-1078.e8.
24. Hassan C, Wallace MB, Sharma P, Maselli R, Craviotto V, Spadaccini M, et al. New artificial intelligence system: first validation study versus experienced endoscopists for colorectal polyp detection. Gut. 2020 May;69(5):799-800.
25. Repici A, Badalamenti M, Maselli R, Correale L, Radaelli F, Rondonotti E, et al. Efficacy of Real-Time Computer-Aided Detection of Colorectal Neoplasia in a Randomized Trial. Gastroenterology. 2020 May; S0016508520305837.
26. Mori Y, Kudo S-E, Misawa M, Saito Y, Ikematsu H, Hotta K, et al. Real-Time Use of Artificial Intelligence in Identification of Diminutive Polyps During Colonoscopy: A Prospective Study. Ann Intern Med. 2018 18;169(6):357-66.
27. Misawa M, Kudo S-E, Mori Y, Hotta K, Ohtsuka K, Matsuda T, et al. Development of a computer-aided detection system for colonoscopy and a publicly accessible large colonoscopy video database (with video). Gastrointest Endosc. 2020 Jul 31.
28. Wang P, Berzin TM, Glissen Brown JR, Bharadwaj S, Becq A, Xiao X, et al. Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study. Gut. 2019;68(10):1813-9.
29. Wang P, Liu X, Berzin TM, Glissen Brown JR, Liu P, Zhou C, et al. Effect of a deep-learning computer-aided detection system on adenoma detection during colonoscopy (CADe-DB trial): a double-blind randomised study. The Lancet Gastroenterology & Hepatology. 2020 Apr;5(4):343-51.
30. Su J-R, Li Z, Shao X-J, Ji C-R, Ji R, Zhou R-C, et al. Impact of a real-time automatic quality control system on colorectal polyp and adenoma detection: a prospective randomized controlled study (with videos). Gastrointestinal Endoscopy. 2020 Feb;91(2):415-424.e4.
31. Gong D, Wu L, Zhang J, Mu G, Shen L, Liu J, et al. Detection of colorectal adenomas with a real-time computer-aided system (ENDOANGEL): a randomised controlled study. The Lancet Gastroenterology & Hepatology. 2020 Apr;5(4):352-61.
32. Liu W-N, Zhang Y-Y, Bian X-Q, Wang L-J, Yang Q, Zhang X-D, et al. Study on detection rate of polyps and adenomas in artificial-intelligence-aided colonoscopy. Saudi J Gastroenterol. 2020;26(1):13.
33. Wang P, Liu P, Glissen Brown JR, Berzin TM, Zhou G, Lei S, et al. Lower Adenoma Miss Rate of Computer-Aided Detection-Assisted Colonoscopy vs Routine White-Light Colonoscopy in a Prospective Tandem Study. Gastroenterology. 2020 Jun 17.
34. Luo Y, Zhang Y, Liu M, Lai Y, Liu P, Wang Z, et al. Artificial Intelligence-Assisted Colonoscopy for Detection of Colon Polyps: a Prospective, Randomized Cohort Study. J Gastrointest Surg [Internet]. 2020 Sep 23 [cited 2020 Nov 9]; Available from: http://link.springer.com/10.1007/s11605-020-04802-4

35. Hassan C, Spadaccini M, Iannone A, Maselli R, Jovani M, Chandrasekar VT, et al. Performance of artificial intelligence for colonoscopy regarding adenoma and polyp detection: a meta-analysis. Gastrointest Endosc. 2020 Jun 26.
36. Lui TKL, Guo C-G, Leung WK. Accuracy of artificial intelligence on histology prediction and detection of colorectal polyps: a systematic review and meta-analysis. Gastrointest Endosc. 2020 Jul;92(1):11-22.e6.
37. Hassan C, Badalamenti M, Maselli R, Correale L, Iannone A, Radaelli F, et al. Computer-aided detection-assisted colonoscopy: classification and relevance of false positives. Gastrointest Endosc. 2020 Jun 16.
38. Hassan C, Bhandari P, Antonelli G, Repici A. Artificial Intelligence for non-polypoid colorectal neoplasms. Digestive Endoscopy: Official Journal of the Japan Gastroenterological Endoscopy Society. 2020 Aug 7.
39. Mori Y, Kudo S-E, East JE, Rastogi A, Bretthauer M, Misawa M, et al. Cost savings in colonoscopy with artificial intelligence-aided polyp diagnosis: an add-on analysis of a clinical trial (with video). Gastrointestinal Endoscopy. 2020 Oct;92(4):905-911.e1.
40. Kudo S-E, Mori Y, Misawa M, Takeda K, Kudo T, Itoh H, et al. Artificial intelligence and colonoscopy: Current status and future perspectives. Dig Endosc. 2019 Jul;31(4):363-71.

# Liver elastography in cholestatic liver diseases - why and how you should do it

Anders Batman Mjelle[1,2,3]

[1] *Department of Clinical Medicine, University of Bergen, Bergen, Norway*
[2] *National Centre for Ultrasound in Gastroenterology, Department of Medicine, Haukeland University Hospital, Bergen, Norway*
[3] *Department of Pediatric and Adolescent Medicine, Haukeland University Hospital, Bergen, Norway*

Liver stiffness measurement is a well-established method in the follow-up of patients with viral hepatitis and constitutes an increasingly important part of the prognostic armamentarium in the health care for patients with chronic liver diseases of other aetiologies. The scope of this paper is to summarize current data demonstrating how liver stiffness measurement is useful for staging and outcome prediction in cholestatic liver disease and to provide a clinical introduction to the use of liver stiffness measurement in cholestatic liver diseases.

**KEYWORDS**
Cholestasis, Elastography, Primary biliary cholangitis, Primary sclerosing cholangitis, Ultrasound

**CONFLICT OF INTEREST DECLARATION**
No conflicts of interest to declare.

## INTRODUCTION

In chronic cholestatic liver disease, e.g., primary biliary cholangitis (PBC) and primary sclerosing cholangitis (PSC), active surveillance of fibrosis development constitutes an integral part of patient care. This is because fibrosis progression is an inherent part of the pathogenesis, with strong associations to clinical outcome. Cholestasis causes progressive hepatic fibrosis through bile-acid induced inflammation unless relieved (1). Correspondingly, periductal inflammation may give rise to cholestasis through stricture formation, revealing a complex relationship between inflammation and cholestasis (1,2).

Liver elastography, commonly referred to as liver stiffness measurement (LSM), is based on non-invasive ultrasound technology and represents an attractive option for repeated fibrosis evaluations. Allowing bedside estimation of liver fibrosis, LSM may easily be integrated into the out-patient clinic and provide immediate results. In contrast to liver biopsy, elastography causes no discomfort, demands little resources, requires a limited amount of training for the investigator, and offers quick, bedside assessment of the entire liver. The latter is advantageous, as a single liver biopsy captures only 1/50'000 of the liver; at the same time, significant sample variation has been demonstrated in the case of cholestatic liver diseases with patchy disease distribution (3,4).

LSM is well established in the follow-up of liver fibrosis in viral hepatitis (5) (Fig. 1) but has been less explored in cholestatic liver diseases. Case series illustrating the confounding effects of cholestasis

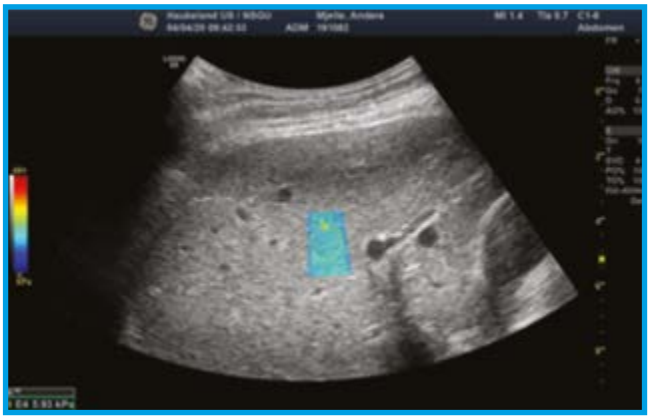

*Figure 1. Ultrasonographic liver evaluation using two-dimensional shear wave elastography (2D-SWE).*
*A color map is created, allowing visualization of liver stiffness. A region of interest (ROI) (yellow circle) is placed in a homogenous part of the color map (blue square), measuring liver stiffness*

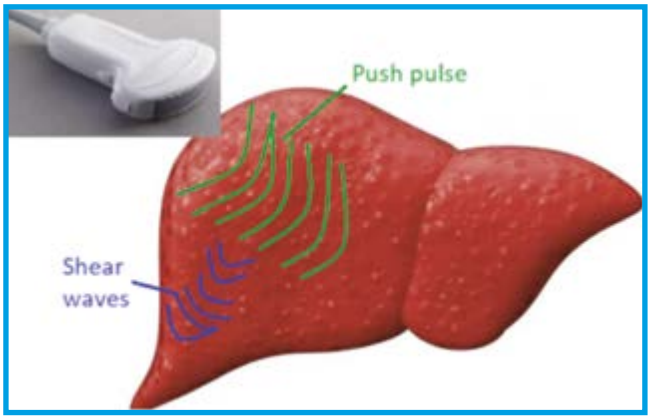


*Figure 2. Shear wave elastography of the liver: a push pulse is transmitted from the transducer and through the adjacent tissue, causing perpendicular shear waves to arise. Shear waves travel very slowly through the tissue and are easy to measure. The speed of shear waves corresponds with the tissue stiffness: stiffer tissue will facilitate quicker waves, i.e., a high speed indicates a stiff and fibrotic liver. Measurements are given in meters/second (m/s) or converted into kilopascals (kPa)*

(causing falsely elevated LSM resolving upon stricture treatment) led to initial concern regarding the use of liver stiffness measurement in cholestatic liver disease (2,6-8). However, the association of LSM with disease stage and clinical outcome has been demonstrated in PSC, PBC, and biliary atresia (BA), and the results have been confirmed in independent studies (9-11). Furthermore, the Baveno-VI criteria, using LSM to exclude the risk of varices needing treatment, thereby avoiding unnecessary upper endoscopies in patients with compensated cirrhosis, has been shown to apply in PBC and PSC patients (12). Elastography is now recommended as part of follow-up in both PBC and PSC (13-15).

## HOW DOES LIVER STIFFNESS MEASUREMENT BY ELASTOGRAPHY WORK?

Elastography is essentially a modern version of the ancient art of palpation, using ultrasound to assess the stiffness of a given tissue: When comparing a hard tissue and a soft tissue, the former is more difficult to displace, and waves will travel through at a higher speed (16). These properties enable estimation of the degree of tissue stiffness: When either a mechanical or an acoustic push pulse is propagated through the tissue, relatively slow shear waves arise, the speed of which is then measured, reflecting the relative tissue stiffness (Fig. 2). Measurements are either given directly as shear wave speed (SWS) in m/s or converted into kilopascals (kPa), using the equation $kPa = 3(ms^{-1})^2$.

There is a wide variety of elastography methods on the market. For LSM, the methods mainly conform to one of three principally different technologies: Transient elastography (TE), point shear wave elastography (pSWE), and two-dimensional shear wave elastography (2D-SWE). Acoustic radiation force impulse (ARFI) quantification may be grouped with pSWE. The most validated and widely used method is transient elastography (TE; e.g., Fibroscan). TE uses a brief mechanical push pulse from a special transducer (piston) to induce shear waves, measured and used to calculate liver stiffness. There is no B-mode imaging involved. A substantial body of literature has emerged supporting the use of pSWE, ARFI, and 2D-SWE, which are elastography methods integrated into high-end ultrasound platforms from a range of producers, allowing clinicians to measure LSM while scanning the liver.

## HOW TO & PITFALLS

LSM is performed with the subject in a supine position, with the right arm abducted and placed under the head. International guidelines (5) recommend performing measurements while the patient is fasting and shortly suspending breath, after a minimum of ten minutes of rest, to avoid confounding effects of respiration and increased hepatic circulation following a meal or exercise (i.e., elevated LSM, not due to fibrosis). The need for rest is probably most relevant after relatively strenuous activity (17).

The transducer is placed perpendicularly in an intercostal space overlying the liver. For TE, the sample area's size and depth are predefined, definitive quality criteria are clearly stated, and ten acquisitions with limited variability (IQR/M ≤30%) are required for a valid result. For other methods, measurements are performed in a region of interest selected by the investigator and placed in a homogenous area, free of vessels and visible bile ducts, and at least 10 mm below the liver capsule, with the most stable results 20-50 mm below the capsule (5). Obesity increases the chances of measurement failure (5,18,19). Some of the elastography platforms offer specific probes for obese patients, and these may improve the success rate. Three to ten acquisitions should be performed depending on each equipment's recommendations, with LSM reported using the median value. This inter-system discrepancy of recommended amount of acquisitions is challenged (20), and recommendations of a low number of acquisitions should be regarded as a minimum amount. The goal is to obtain a stable median value with as little variation as possible, reflecting true liver stiffness.

International guidelines recommend using IQR/M ≤30% as a LSM quality criterion for both TE and pSWE while stating that there are no current recommendations for 2D-SWE, though the criterion has been adopted in several studies on 2D-SWE (5). A higher IQR/M reflects a higher degree of measurement value variation; hence, measurements with a high IQR/M are less reliable (20,21). It has been proposed to label measurements as either very reliable, reliable, or poorly reliable, based on the IQR/M level, which is adopted by the TE system producer (21). It is good practice to report IQR/M alongside the median value (5), irrespective of system, which may be compared to reporting mean values with confidence intervals. In a clinical setting where IQR/M ends above 30%, it is possible to erase acquisitions and start over, or to continue acquisitions to see if variability decreases, rather than dismissing the measurement as failed.

Technical issues, such as excess mechanical stress upon the liver, either due to a deep breath or too much force upon the ultrasound probe, may lead to falsely increased LSM values. It is also essential to know that LSM values for a given level of liver fibrosis may vary between elastography methods, between producers, and even between different platforms from the same producer. Consequently, it is crucial to be aware that cutoff values for fibrosis stages may differ based on equipment, and differences in LSM measured by two different methods may not reflect a true change in fibrosis. Moreover, although boasting a low intra- and inter-operator variability, variability exists, and in mild to moderate liver disease, the main clinical benefit often lies in the longitudinal follow-up with LSM, revealing stiffness development over time.

## CONFOUNDING FACTORS

Several confounding factors (i.e., other factors besides liver fibrosis causing increased liver stiffness) have been described. For the sake of simplicity, these may be divided into either A) acute inflammation or B) increased pressure due to fluid accumulation (cholestasis and increased blood volume within the liver). Current guidelines recommend that acute hepatitis (including viral, toxic, or other origins) or elevated transaminase levels > 5x upper limit of normal, cholestasis (elevated bilirubin), and hepatic congestion (hepatic vein dilatation upon ultrasound; congestive cardiac failure) should be excluded before performing LSM. Caution is warranted in the interpretation of LSM values obtained in the presence of any of these factors (5). LSM has been shown to fall significantly following endoscopic retrograde cholangiopancreatography (ERCP) in cholestatic

patients (7,8), though these patients had very high bilirubin levels. The confounder effect of bilirubin in PSC patients with low to moderate bilirubin levels is less explored. However, a significant correlation with LSM levels has been shown for both TE, pSWE, and 2D-SWE (9,22).

In cases of alcoholic hepatitis, abstinence for 1-4 weeks prior to LSM is recommended (5). Current alcohol abuse induces inflammatory activity, and studies have demonstrated reductions in LSM following periods of abstinence in this patient group. Recent alcohol exposure in healthy subjects does not seem to interfere with LSM (17), but there is a lack of similar publications on patients with chronic liver disease.

## LIVER STIFFNESS MEASUREMENTS IN PRIMARY BILIARY CHOLANGITIS

LSM was first described in PBC (and PSC) in 2006, demonstrating a significant correlation between LSM values by TE and histopathological fibrosis assessment (23). In a cohort of 103 PBC patients comparing liver biopsy with TE, the diagnostic threshold for ≥F1, ≥F2, ≥F3, and ≥F4 were 7.1, 8.8, 10.7, and 16.9 kPa, respectively, outperforming various biochemical markers (10). The authors found that the risk of liver decompensation, liver transplantation, or death increased more than 8-fold in patients with an LSM increase above 2.1 kPa/year and 5-fold in patients with a baseline LSM above a 9.6 kPa threshold. In 1 in 5 patients, valid LSM results were not achieved, associated with a greater fibrosis stage. A recent Japanese study compared TE with liver histology, using the Nakanuma classification, found median LSM to be 5.1, 5.9, 8.9, and 23.7 kPa for stages 1-4, respectively (24). LSM is commonly observed to be less able to discriminate between early fibrosis stages. The area under the curve (AUC) for LSM diagnosing F4 fibrosis in PBC is very high, ranging from 0.91 to 0.99 (10,23-26). Although these studies report LSM to outperform other non-invasive markers, the combination of LSM and serological markers has been described to increase accuracy (24). Similar results have been found in patients with autoimmune hepatitis (AIH)-PBC overlap syndrome (25). International guidelines now recommend liver elastography at the time of diagnosis as well as during follow-up of PBC (15).

## LIVER STIFFNESS MEASUREMENTS IN PRIMARY SCLEROSING CHOLANGITIS

There are no established prognostic markers to accurately predict PSC development's notoriously variable course at the individual level, but LSM is one of the leading candidates for risk stratification and prognostication. LSM by TE as well as pSWE is associated with clinical outcome in PSC (9,27,28). A landmark paper from 2014, followed by an independent validation study demonstrated a high degree of correlation between LSM and histological fibrosis stage, particularly for the advanced stages of fibrosis. The threshold values for the different fibrosis stages were similar: 6.6-7.4, 8.6-8,8, 9.6, and 14.4 kPa for ≥F1, ≥F2, ≥F3, and ≥F4, respectively (9,27). The AUC for diagnosing F4 fibrosis in these two studies was similar to studies in PBC: 0.95 and 0.98, respectively. Notably, in this study, LSM was correlated with fibrosis stage even in patients with signs of cholestasis. Both baseline LSM value and LSM change over time were associated with clinical outcomes, with optimal cutoffs for predicting adverse outcomes of 18.5 kPa for baseline LSM and 1.3 kPa/year for LSM change. This very high baseline LSM cutoff was calculated using Youden's index and should not stop the clinician from keenly observing whether LSM indicates ≥F2 fibrosis, with values above 9-10 kPa as especially concerning. Baseline LSM was closely linked to LSM change with time, and the study demonstrated that patients with no or slight fibrosis (F0-1) hardly progressed during the 3.9±2.1 years follow-up. In contrast, patients with significant to severe fibrosis (F2-4) at baseline showed an almost exponential progression of fibrosis. Other studies have shown that liver elastography is correlated with both B-mode findings indicative of liver fibrosis (28) and with the ELF test, a test with predictive value regarding transplant-free survival in PSC (29,30).

## LIVER STIFFNESS MEASUREMENTS IN NEONATAL CHOLESTASIS AND BILIARY ATRESIA (BA)

Elastography has been applied in neonatal cholestasis, finding higher LSM in BA than non-BA cholestatic infants, irrespective of serum bilirubin or bile acid level, correlating with the degree of liver fibrosis (11,31,32). LSM in BA has been shown to rise in patients in need of liver transplantation (33). Different studies report variable cutoff values as useful for diagnosing BA, ranging from 1.35-1.84 m/s (~5.5-10.2 kPa). In children with BA after surgical correction with hepatoportoenterostomy (Kasai procedure), LSM one week after surgery was significantly correlated with the presence of esophageal varices six months after surgery and the need for liver transplantation 1.5 years later (34). LSM has also been shown to be elevated in otherwise healthy neonates with transient neonatal cholestasis, a common clinical situation in neonatology (35).

## DISCUSSION AND CONCLUSION

In cholestatic liver diseases, elastography is useful for staging and outcome prediction during follow-up, with values corresponding to the histological degree of fibrosis. In mild to moderate disease, its strength lies in the assessment of disease progression over time. In more severe disease, LSM discriminates cirrhosis from milder stages with high accuracy and differentiates stages within the spectrum of cirrhosis (e.g., cirrhosis with or without clinically significant portal hypertension), giving a more immediate and accurate assessment, e.g., when evaluating for liver transplantation. It has been argued that it is particularly useful in patients with advanced liver disease where liver biopsy may be contraindicated (24). It should thus be part of both the diagnosis and follow-up of all cholestatic liver diseases.

The confounding effects of cholestasis should be kept in mind. Cholestasis may lead to falsely elevated LSM, which must be considered when interpreting results. This justifies the recommendation to combine LSM and B-mode ultrasonography to rule out significant cholestasis, which may be segmental in PSC. In the diagnosis of PBC and PSC, the main target is to have a baseline LSM, which correlates with prognosis and forms a basis for analyzing progression in future follow-ups. In cholestatic newborns, LSM may have more directly diagnostic properties, indicating BA in neonates with high liver stiffness. However, since there is an absolute need for a sensitivity of 100% in the diagnosis of BA, LSM is more promising in the follow-up after surgical correction.

In conclusion, liver elastography represents an attractive aid in cholestatic liver diseases, allowing a non-invasive and quick estimation of liver fibrosis development and progression. Further studies on elastography in cholestatic liver diseases are warranted to evaluate how baseline LSM and LSM change over time can be used to guide treatment and follow-up.

### ABBREVIATION LIST

| | | | |
|---|---|---|---|
| 2D-SWE | Two-dimensional shear wave elastography | LSM | Liver stiffness measurement |
| APRI | Aspartate aminotransferase to platelet ratio index | PBC | Primary biliary cholangitis |
| ARFI | Acoustic radiation force impulse, | PFIC | Progressive familial intrahepatic cholestasis |
| BA | Biliary atresia | PSC | Primary sclerosing cholangitis |
| CLD | Cholestatic liver disease | pSWE | Point shear wave elastography |
| ELF | Enhanced liver fibrosis test | SWS | Shear wave speed |
| kPa | Kilopascals | TE | Transient elastography |

**References**


1. Woolbright BL. Inflammation: Cause or consequence of chronic cholestatic liver injury. Food Chem Toxicol 2020;137.
2. Karlsen TH, Folseraas T, Thorburn D et al. Primary sclerosing cholangitis - a comprehensive review. Journal of Hepatology 2017;67:1298-1323.
3. Scheuer PJ. Ludwig Symposium on biliary disorders--part II. Pathologic features and evolution of primary biliary cirrhosis and primary sclerosing cholangitis. Mayo Clin Proc 1998;73:179-183.
4. Garrido MC, Hubscher SG. Accuracy of staging in primary biliary cirrhosis. J Clin Pathol 1996; 49:556-559.
5. Dietrich CF, Bamber J, Berzigotti A. EFSUMB Guidelines and Recommendations on the Clinical Use of Liver Ultrasound Elastography, Update 2017 (Long Version) (vol 38, pg e16, 2017). Ultraschall Med 2017; 38: E52-E52.
6. Ehlken H, Lohse AW, Schramm C. Transient Elastography in Primary Sclerosing Cholangitis-the Value as a Prognostic Factor and Limitations. Gastroenterology 2014;147:542-543.
7. Attia D, Pischke S, Negm AA et al. Changes in liver stiffness using acoustic radiation force impulse imaging in patients with obstructive cholestasis and cholangitis. Digest Liver Dis 2014;46:625-631.
8. Millonig G, Reimann FM, Friedrich S et al. Extrahepatic cholestasis increases liver stiffness (FibroScan) irrespective of fibrosis. Hepatology 2008;48:1718-1723.
9. Corpechot C, Gaouar F, El Naggar A et al. Baseline Values and Changes in Liver Stiffness Measured by Transient Elastography Are Associated With Severity of Fibrosis and Outcomes of Patients With Primary Sclerosing Cholangitis. Gastroenterology 2014;146: 970-U511.
10. Corpechot C, Carrat F, Poujol-Robert A et al. Noninvasive elastography-based assessment of liver fibrosis progression and prognosis in primary biliary cirrhosis. Hepatology 2012;56:198-208.
11. Hanquinet S, Courvoisier DS, Rougemont AL et al. Contribution of acoustic radiation force impulse (ARFI) elastography to the ultrasound diagnosis of biliary atresia. Pediatr Radiol 2015;45:1489-1495.
12. Moctezuma-Velazquez C, Saffioti F, Tasayco-Huaman S et al. Non-Invasive Prediction of High-Risk Varices in Patients with Primary Biliary Cholangitis and Primary Sclerosing Cholangitis. American Journal of Gastroenterology 2019;114:446-452.
13. Hirschfield GM, Dyson JK, Alexander GJM et al. The British Society of Gastroenterology/UK-PBC primary biliary cholangitis treatment and management guidelines. Gut 2018;67:1568-1594.
14. Chapman MH, Thorburn D, Hirschfield GM et al. British Society of Gastroenterology and UK-PSC guidelines for the diagnosis and management of primary sclerosing cholangitis. Gut 2019;68:1356-1378.
15. European Association for the Study of the Liver. Electronic address eee, European Association for the Study of the L. EASL Clinical Practice Guidelines: The diagnosis and management of patients with primary biliary cholangitis. J Hepatol 2017;67:145-172.
16. Bamber J, Cosgrove D, Dietrich CF et al. EFSUMB Guidelines and Recommendations on the Clinical Use of Ultrasound Elastography. Part 1: Basic Principles and Technology. Ultraschall Med 2013;34:169-184.
17. Taraldsen V, Tomasgard S, Rudlang MT, Gilja OH, Vesterhus M, Mjelle AB. Point shear wave elastography and the effect of physical exercise, alcohol consumption, and respiration in healthy adults (In press). Ultrasound Int Open 2020, DOI:1-8.
18. Thiele M, Detlefsen S, Moller LS et al. Transient and 2-Dimensional Shear-Wave Elastography Provide Comparable Assessment of Alcoholic Liver Fibrosis and Cirrhosis. Gastroenterology 2016;150:123-133.
19. Cassinotto C, Boursier J, de Ledinghen V et al. Liver stiffness in nonalcoholic fatty liver disease: A comparison of supersonic shear imaging, FibroScan, and ARFI with liver biopsy. Hepatology 2016;63:1817-1827.
20. Mjelle AB, Mulabecirovic A, Havre RF, Olafsdottir EJ, Gilja OH, Vesterhus, M. Liver elastography in healthy children using three different systems - how many measurements are necessary? (In press). Ultraschall Med 2020, DOI:1-10.
21. Boursier J, Zarski JP, de Ledinghen V et al. Determination of reliability criteria for liver stiffness evaluation by transient elastography. Hepatology 2013; 57: 1182-1191

22. Mjelle AB, Fossdal G, Gilja OH et al. Liver Elastography in Primary Sclerosing Cholangitis Patients Using Three Different Scanner Systems. Ultrasound Med Biol 2020; 46:1854-1864.
23. Corpechot C, El Naggar A, Poujol-Robert A et al. Assessment of biliary fibrosis by transient elastography in patients with PBC and PSC. Hepatology 2006;43: 1118-1124.
24. Joshita S, Yamashita Y, Sugiura A et al. Clinical utility of FibroScan as a non-invasive diagnostic test for primary biliary cholangitis. J Gastroen Hepatol 2019, DOI: 10.1111/jgh.14929.
25. Wu HM, Sheng L, Wang Q et al. Performance of transient elastography in assessing liver fibrosis in patients with autoimmune hepatitis-primary biliary cholangitis overlap syndrome. World J Gastroenterol 2018; 24: 737-743.
26. Floreani A, Cazzagon N, Martines D et al. Performance and utility of transient elastography and noninvasive markers of liver fibrosis in primary biliary cirrhosis. Dig Liver Dis 2011;43:887-892.
27. Ehlken H, Wroblewski R, Corpechot C et al. Validation of Transient Elastography and Comparison with Spleen Length Measurement for Staging of Fibrosis and Clinical Prognosis in Primary Sclerosing Cholangitis. Plos One 2016;11.
28. Mjelle AB, Mulabecirovic A, Hausken T et al. Ultrasound and Point Shear Wave Elastography in Livers of Patients with Primary Sclerosing Cholangitis. Ultrasound in Medicine and Biology 2016;42:2146-2155.
29. Vesterhus M, Hov JR, Holm A et al. Enhanced liver fibrosis score predicts transplant-free survival in primary sclerosing cholangitis. Hepatology 2015;62:188-197.
30. de Vries EM, Farkkila M, Milkiewicz P et al. Enhanced liver fibrosis test predicts transplant-free survival in primary sclerosing cholangitis, a multi-centre study. Liver International 2017;37:1554-1561.
31. Chen Y, Zhao D, Gu S et al. Three-color risk stratification for improving the diagnostic accuracy for biliary atresia. Eur Radiol 2020, DOI: 10.1007/s00330-020-06751-7.
32. Dillman JR, DiPaola FW, Smith SJ et al. Prospective Assessment of Ultrasound Shear Wave Elastography for Discriminating Biliary Atresia from other Causes of Neonatal Cholestasis. J Pediatr-Us 2019;212:60-+.
33. Hanquinet S, Courvoisier DS, Rougemont AL et al. Acoustic radiation force impulse sonography in assessing children with biliary atresia for liver transplantation. Pediatr Radiol 2016;46:1011-1016.
34. Wu JF, Lee CS, Lin WH et al. Transient elastography is useful in diagnosing biliary atresia and predicting prognosis after hepatoportoenterostomy. Hepatology 2018;68:616-624.
35. Alison M, Biran V, Tanase A et al. Quantitative Shear-Wave Elastography of the Liver in Preterm Neonates with Intra-Uterine Growth Restriction. PLoS One 2015; 10:e0143220.

# Investigator initiated research in times of COVID-19: let's go digital!

Comment to: Schuitenmaker JM, Oude Nijhuis RAB, Bredenoord AL, Fockens P, Bredenoord AJ. Investigator initiated research in times of COVID-19: Let's go digital! Neurogastroenterol Motil. 2020 Nov;32(11):e14011.

by A.J. Bredenoord
*Department of Gastroenterology and Hepatology, Academic Medical Center, Amsterdam, The Netherlands*

The COVID-19 pandemic has accelerated digital transformation in many sectors; business meetings are now being done virtually, education at universities is performed with Zoom and internet shopping has seen a huge increase. Medical research has been behind for a while, in part due to regulatory hurdles but maybe even more importantly because of a lack of experience and a hesitant approach of both investigators and sponsors.

In clinical trials, handheld devices and apps are used to register symptoms and CRFs are now often web-based, but recruitment, informed consent procedures, IMPD distribution, and follow up visits are nearly always performed at the research site. Investigator-initiated studies are even less virtual, and even questionnaires and CRFs are usually done using paper hard-copies.

Lockdowns, travel restrictions and non-critical hospital visit restrictions have led to a search for alternatives. In the journal *Neurogastroenterology & Motility*, Schuitenmakers et al describe several virtual solutions that can be used in the execution of investigator-initiated trials and allow continuation even during the most strict lockdowns (1)(2).

The study to which the manuscript refers to is a randomized controlled trial of the effect of a wearable sleep position device that helps patients to sleep on the left lateral side during the night and in such a way, prevents nocturnal reflux symptoms as it is known these mainly occur in the right lateral side, this is also the reason endoscopy is always performed on the left side (3). The wearable device vibrates in an annoying way when it senses a subjects rolls towards the right lateral side, causing the subject to roll back to the left. After a few nights, the subject has learned to avoid the right lateral sleeping position. Subjects with nocturnal reflux symptoms are randomized to the device with the vibration algorithm or to a sham device with a random vibration algorithm.

Recruitment of subjects is performed via a Google Adwords solution, in which subjects that enter *"reflux during sleep"*, *"how to prevent acid during the night"*, receive a targeted advertisement of the hospital research team. If they leave their contact details at the landing page, they are contacted by the investigator and via video calling they are further informed about the study. Eligibility checks and informed consent is also performed via video calling and if they participate, randomization is performed using a web-based CRF program and the device of sham-device is being shipped to their home address. Symptom questionnaires are being emailed daily by the CRF system and responses appear automatically in the CRF. The entire trial can be performed while the patient never visited the hospital.

The authors address the possibilities of their approach and also discuss when their virtual solutions are

not possible. Physical examinations, required when new drugs and devices are being used, can not be performed distantly. Endoscopy also requires a patient to visit an endoscopy department. There are also some privacy and ethical issues, in particular around the Google recruitment (4). Traceable data to the Google Adwords learns Google that a specific individual has certain medical complaints while the reputation of tech/internet platforms regarding the respect for privacy is controversial (5).

In summary, the article by Schuitenmakers illustrates new ways of performing research in a virtual way. This allows to perform clinical trials with less or no physical contact between patient and investigator. This approach can guarantee trials to continue in the COVID-19 era, but will also be attractive afterwards.

**References**

1. J.M. Schuitenmaker, R.A.B. Oude Nijhuis, A.L. Bredenoord, P. Fockens, A.J. Bredenoord. Investigator initiated research in times of COVID-19: let's go digital! Neurogastroenterology & Motility 2020 Nov;32(11):e14011. doi: 10.1111/nmo.14011.
2. FDA Guidance on Conduct of Clinical Trials of Medical Products during COVID-19 Public Health Emergency 2020 07-07-2020. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/fda-guidance-conduct-clinical-trials-medical-products-during-covid-19-public-health-emergency.
3. Khoury RM, Camacho-Lobato L, Katz PO, Mohiuddin MA, Castell DO. Influence of spontaneous sleep positions on nighttime recumbent reflux in patients with gastroesophageal reflux disease. The American journal of gastroenterology. 1999;94(8):2069-73.
4. Sharon T. The Googlization of health research: from disruptive innovation to disruptive ethics. Per Med. 2016;13(6):563-74.
5. Curtis BL. Social networking and online recruiting for HIV research: ethical challenges. J Empir Res Hum Res Ethics. 2014;9(1):58-70.

# Achalasia and Obstructive Motor Disorders Are Not Uncommon in Patients With Eosinophilic Esophagitis

## Clinical Gastroenterology and Hepatology in press

Comment to: Ghisa M, Laserra G, Marabotto E, Ziola S, Tolone S, de Bortoli N, Frazzoni M, Mauro A, Penagini R, Savarino V, Barberio B, Giannini EG, Zentilin P, Gyawali CP, Savarino E. Achalasia and Obstructive Motor Disorders Are Not Uncommon in Patients With Eosinophilic Esophagitis. Clin Gastroenterol Hepatol. 2020 Aug 4:S1542-3565(20)31066-1.

by A.J. Bredenoord
*Department of Gastroenterology and Hepatology, Academic Medical Center, Amsterdam, The Netherlands*

Eosinophilic esophagitis (EoE) is a chronic immune-mediated disorder characterized by esophageal symptoms and eosinophilic inflammation, in which antigens, mainly of food origin, stimulate a Th2-mediated immune response that recruits eosinophils into the esophagus (1). Manometric abnormalities are often described in these patients, ranging from compartmentalized pressurizations, to premature contractions, ineffective esophageal motility and hypertensive contractions (2)(3). It is not entirely clear what the relationship is between tissue eosinophilia in EoE and these esophageal motility abnormalities.

Ghisa and coworkers have systematically performed high-resolution manometry (HRM) in a large series of patients with EoE and describe their findings in a paper in Clinical Gastroenterology and Hepatology (4). Esophageal motility abnormalities were found in 38 of all patients, of which 59% had a minor motor disorders but 41% had a major motor disorder, including several patients with a manometric pattern compatible with achalasia, Jackhammer esophagus and distal esophageal spasm. Clinical features of EoE patients with achalasia and obstructive motor disorders were almost identical to those without these motility disorders, indicating that dysphagia induced by motor disorders can mimic that related to EoE. Some patients with a major motor disorder responded to medical therapy including proton pump inhibitors and other to topical corticosteroids. The EoE patients with achalasia responded much less to topical corticosteroids than the other EoE patients. In some patients, pneumatic dilation was required.

The authors conclude that dedicated dysphagia evaluation with both esophagogram and HRM is needed when symptoms persist despite EoE management because major motility disorders can be present and responsible.

Several explanations for the high co-occurrence of motility disorders and esophageal eosinophilia are possible. Stasis of food in the esophagus of patients with severe motility disorders can cause eosinophilia, what pleads against this explanation is that after treatment with resolution of the stasis, eosinophilia does not always disappear. The second possible explanation is that mediators released by the inflammatory calls cause motility abnormalities by a direct effect on muscle fibers.

A third explanation is a direct cytotoxic effect of eosinophils on esophageal nerve cells. A fourth, not mentioned, explanation could be that fibrosis results in manometric abnornmalities. The current study does allow differentiations between these explanations and more research is clearly required.

**References**

1. Dellon et al. Updated International Consensus Diagnostic Criteria for Eosinophilic Esophagitis: Proceedings of the AGREE Conference. Gastroenterology 2018 Oct;155(4):1022-1033.e10. doi: 10.1053/j.gastro.2018.07.009
2. van Rhijn BD et al. Prevalence of esophageal motility abnormalities increases with longer disease duration in adult patients with eosinophilic esophagitis. Neurogastroenterol Motil 2014;26:1349-1355.
3. Roman S, Hirano I, Kwiatek MA, et al. Manometric features of eosinophilic esophagitis in esophageal pressure topography.Neurogastroenterol Motil 2011; 23:208-e111.
4. Ghisa et al. Achalasia and Obstructive Motor Disorders Are Not Uncommon in Patients With Eosinophilic Esophagitis Clinical Gastroenterology and Hepatology in press