

The IDEA *Journal*

International Digestive Experts Articles

Year II - N. 1, 2020

Editorial

Digestive diseases and innovative techniques
during the pandemic

COVID-19 and liver disease

Artificial intelligence in the detection and
characterisation of Barrett's related neoplasia:
a review of recent literature

Big data application in prevention and
control of infectious diseases

The IDEA Journal

International Digestive Experts Articles

QUARTERLY SCIENTIFIC UPDATE IN GASTROENTEROLOGY

Year II - N. 1, 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

Editorial

Digestive diseases and innovative techniques during the pandemic

5

COVID-19 and liver disease

7

Artificial intelligence in the detection and characterisation of Barrett's related neoplasia: a review of recent literature

12

Big data application in prevention and control of infectious diseases

17

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

Digestive diseases and innovative techniques during the pandemic

The International Digestive Experts Articles Journal welcomes three world-class experts as members of the Editorial Board:

- Arjan Bredenoord, Professor of Neurogastroenterology & Motility at Amsterdam Medische Centra,
- Cesare Hassan, Clinical Gastroenterologist at Nuovo Regina Margherita Hospital in Rome, and
- Mette Vesterhus, Associate Professor and Senior Consulting Gastroenterologist & Hepatologist at the University of Bergen and at Haraldsplass Deaconess Hospital

Written during this unique period, the first Number of the second year of IDEA is focused on digestive diseases during the pandemic and on innovative medical techniques.

Many questions on COVID-19 and liver disease are raised and discussed in Professor Vesterhus article. Elevated liver biochemistries observed in COVID-19 reflect a distinct liver disease induced by the SARS-CoV-2 virus or a bystander phenomenon? Is the liver affection in covid-19 due to primary liver infection? Are chronic liver disease patients or patients with liver transplants at risk for severe disease in COVID-19?

The latest evidence on the use of deep learning or

convolutional neural networks for the development of computer aided detection systems for the detection of high-grade dysplasia and early esophageal cancer within Barrett's Esophagus are examined in the short review by Professor Hassan and his group.

This is the first of a series of 4 articles dedicated to artificial intelligence and gastrointestinal endoscopy, that will be published in the following numbers of IDEA.

Big data has attracted more and more attentions from various fields, and has been applied in lots of fields, including health care. Tingting Dai and colleagues summarized the characteristics of these novel data sources and illustrated the application of big data in infectious disease field, with a focus on COVID-19 and also emphasizing the potential of big data in preventing and/or controlling endemic pathologies.

The progress of computer science and the improvement of computing speeds have led to the extensive application of artificial intelligence in many fields and big data provides an unprecedented opportunity making it a useful and low-cost complement to traditional epidemiology. The authors state that there is much work needs to be done to demonstrate the capability and reduce the risks of using relatively novel big data and it is vital to collaborate with computer scientists.

COVID-19 and liver disease

Mette Vesterhus

Associate Professor, Department of Clinical Science, University of Bergen, Bergen, Norway

Senior Consulting Gastroenterologist & Hepatologist, Haralds plass Deaconess Hospital, Bergen, Norway

Elevated liver biochemistries are observed in up to half of the patients admitted to hospitals for COVID-19, and are twice as common in patients with severe COVID-19. Does this reflect a distinct liver disease induced by the SARS-CoV-2 virus or a bystander phenomenon?

Are chronic liver disease patients or patients with liver transplants at risk for severe disease in COVID-19?

These issues as well as current recommendations for the management of liver disease patients are discussed in the following.

ELEVATED LIVER BIOCHEMISTRIES IS COMMON IN PATIENTS WITH COVID-19

Elevated liver biochemistries are reported in 14-53% of patients admitted to hospital with COVID-19 disease (1-3). Transaminase elevations in the range of 1-2 x upper limit of normal (ULN) is most commonly observed, while bilirubin is often normal or slightly elevated and ALP is rarely elevated. Generally, the transaminase elevation is transient and does not require specific treatment. Low albumin at admittance is however identified as a risk factor for severe disease course.

Only rare cases of severe acute liver failure have been described in patients with COVID-19 (2,4).

IS THE LIVER AFFECTION IN COVID-19 DUE TO PRIMARY LIVER INFECTION?

SARS-CoV-2 infects cells via binding to angiotensin converting enzyme 2 (ACE2) (5). This receptor is not confined to the lungs but is also expressed in the liver, by hepatocytes as well as cholangiocytes (3,6).

This has led to the hypothesis that liver affection seen

in COVID-19 may stem from direct cytotoxic effects of a primary liver infection; conceivably, the virus may translocate to the liver from the gut (SARS-CoV-2 is detected in the stools and hence present in the gut), infect hepatic cells and activate proinflammatory signals in the liver, releasing a range of cytokines and chemokines (Figure 1). However, proof of this theory is lacking, as liver biopsies from patients with COVID-19 have so far failed to show virus inclusions.

SEVERAL INDIRECT CAUSES OF LIVER DAMAGE IN COVID-19

Likely, liver affection in the COVID-19 patient population is at least partly due to a mix of several indirect causes of liver damage (7,8) (Figure 2).

Hypoxia is an important hallmark of COVID-19, hence liver ischemia due to hypoxia or congestive liver disease following mechanical ventilation with high end-respiratory pressures and reduced venous return are potential contributory factors for liver damage in some patients; however, the evidence suggests that these are not important factors overall. Drug-induced liver injury is a likely cause of elevated liver enzymes in some of the reported cases; studies from Chinese centers report that up to 84% of the patients were treated with antivirals and to some extent also high doses of paracetamol surpassing commonly used prescription recommendations (9). However, as mild liver test pathology was observed at admittance in many patients, drug-induced liver injury may not be the sole explanation.

It should be kept in mind that the presence of elevated transaminases does not always equal liver disease - ordinary flu may give rise to a myositis with increased transaminases as a consequence. Possibly, a similar

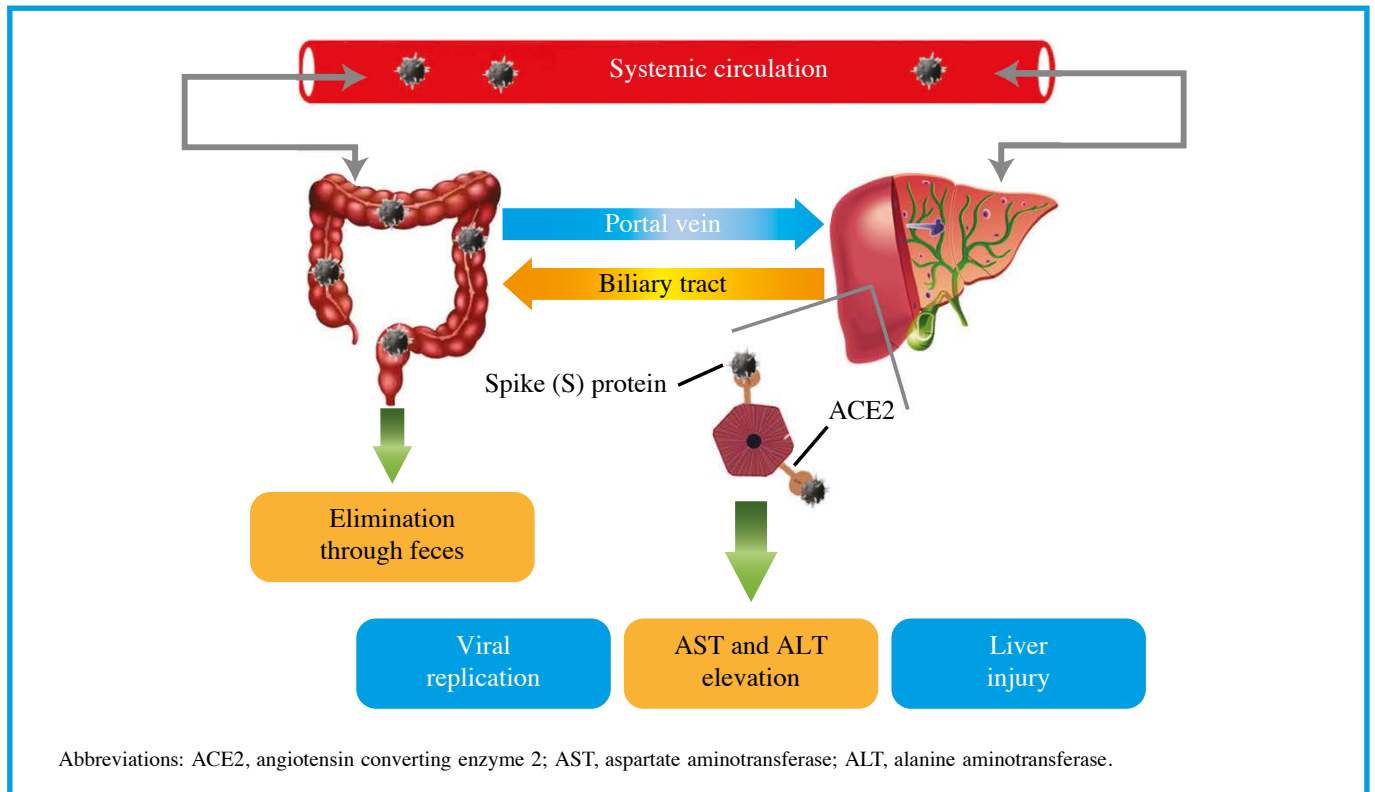


Figure 1. Mechanisms of liver injury induced by SARS-CoV-2: direct liver injury by the virus (mod. from Mendez-Sanchez N et al. J Clin Transl Hepatol 2020)

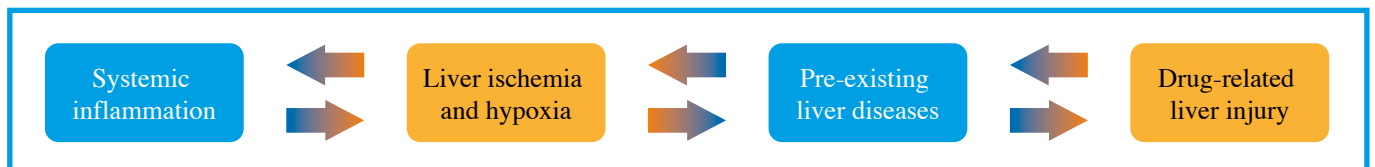


Figure 2. Mechanisms of liver injury induced by SARS-CoV-2 Indirect causes of liver injury associated with COVID-19 infection (mod. from Mendez-Sanchez N et al. J Clin Transl Hepatol 2020)

mechanism in COVID-19 might partly explain the frequent occurrence of transaminase elevation in this disease, and this hypothesis is supported by several studies reporting an association of elevated CK, LD or myoglobin with COVID-19 severity.

LIKELY LINKS TO IMMUNITY: IMMUNE FAILURE OR CYTOKINE STORM?

There is clear evidence that COVID-19 activates the innate and adaptive immune systems in the infected host. Patients with COVID-19 and progressive pneumonia, ARDS and multiorgan failure frequently display signs of hyperinflammation and «cytokine storm syndrome».

Such development is particularly dangerous in the elderly, who are commonly affected by so-called T-cell senescence (halted cell division) and in others with similar low-grade, proinflammatory conditions (10). Common flu and other systemic viral infections are known to be associated with transiently elevated transaminases reflecting general immune activation. Severe COVID-19 with pathological liver biochemistries is associated with concomitant activation of coagulation and fibrinolytic factors, slight thrombocytopenia, neutrophilia and elevated ferritin, all of which are often interpreted as unspecific expressions of a systemic inflammatory response; alternatively, however, these findings might reflect

innate immune dysfunction. Being increasingly prevalent in high age, this might explain why the elderly constitute such an important risk group.

Liver cirrhosis is a condition with low-grade systemic inflammation and reduced or imbalanced immune function. In consequence, if any of the above hypotheses apply, patients with cirrhosis are at risk for a severe disease course in COVID-19.

CHRONIC VIRAL OR AUTOIMMUNE LIVER DISEASES ARE NOT ASSOCIATED WITH SEVERE COVID-19

Reports so far do not support an increased risk of attracting COVID-19 disease in need of hospital care in patients with chronic viral or autoimmune liver diseases, as the prevalence of pre-existing disease in these categories is relatively low in patients with COVID-19 admitted to hospital (11,12).

Furthermore, unlike the experience of the previous SARS pandemic in which SARS-CoV was shown to induce increased replication of hepatitis virus and severe hepatitis, there are no clear indications of increased risk of severe disease course in COVID-19 for patients with chronic viral hepatitis.

In autoimmune liver diseases, the risk of a severe disease course may be affected by stage (presence of cirrhosis) and potentially immune suppression. Contrary to initial concerns, however, observations do not support any increased risk of severe COVID-19 in patients on immunosuppressive therapy for either autoimmune liver disease or other chronic diseases (e.g. inflammatory bowel disease) (13).

Emerging international recommendations advice that immunosuppressive therapy in general should be continued as the consequences of discontinuation may be severe, and this should be stressed in the communication with patients; however, in selected cases dose adjustments or discontinuation should be considered in collaboration with a hepatologist; see Figure 3 and section on liver transplant recipients (14).

CIRRHOSIS

The main challenge for patients with chronic liver diseases may turn out to be the risk of decompensation subsequent to the COVID-19 infection, which may lead to development of severe liver failure with its complications. Reduced out-patient clinic activity is recommended in many countries to prevent spreading

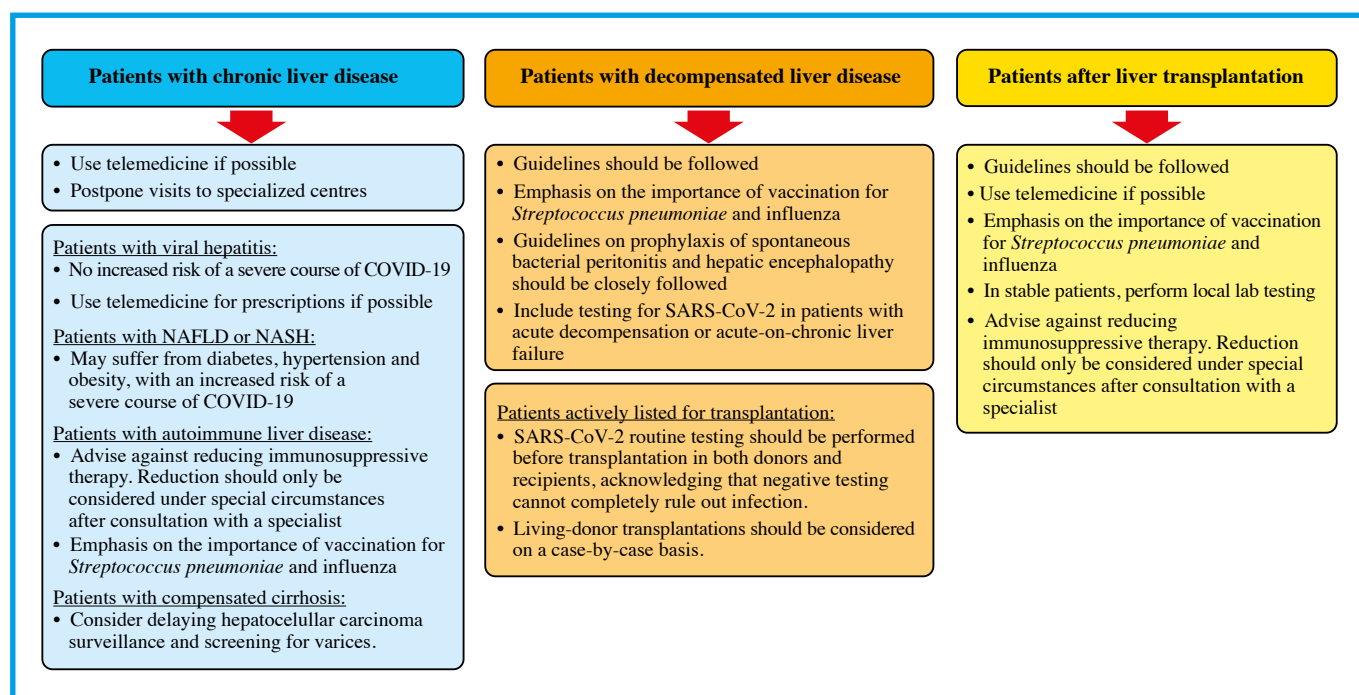


Figure 3. Management of patients with chronic liver diseases during the COVID-19 pandemic (mod. from the EASL-ESCMID position paper by Boettler et al., JHEP reports 2020)

of the virus and protect at-risk patients, such as patients with cirrhosis. While this is an important measure, one of the implications might be reduced quality of care in the management of this patient group, as intervals for surveillance procedures (gastroscopy for varices; imaging for HCC) and clinical assessment (hepatic encephalopathy, ascites) are increased. It is important to maintain focus on this group of patients who need regular follow-up to avoid “collateral damage” from the protective measures in the form of increased morbidity and mortality due to complications of cirrhosis (Figure 3).

INCREASED RISK FOR PATIENTS WITH NAFLD AND COVID-19

The majority of patients with non-alcoholic fatty liver disease (NAFLD) have comorbidities such as obesity, diabetes and hypertension and published data demonstrate increased risk of a severe disease course for COVID-19 in this patient population (3, 14).

Whether part of this increased risk is conferred by the liver disease per se or associated with the comorbidities (which are well-known, independent risk factors) is unclear, and hence the risk status of patients with NAFLD without any comorbidity is less certain.

LIVER TRANSPLANTATION AND COVID-19

The COVID-19 pandemic has led to a significant reduction in solid organ transplantations (SOT) internationally. The USA and France, two of the countries which have been hit hard by COVID-19, have experienced reductions of 50 and 91%, respectively, in the number of transplanted organs

from deceased donor (15). Reduced kidney transplantation activity is the main factor behind this fall, but liver transplantation rates contribute. As COVID-19 is transmissible from donor to recipient, all potential donors and recipients should be thoroughly tested for SARS-CoV-2 and the possibility of false negative tests is of concern.

The data regarding COVID-19 in liver transplanted patients is limited and based on only a few reported cases. The largest series of transplant recipients to date reported that out of 13 liver recipients, 31% had severe disease (15).

In the previous SARS (2003-2004) and MERS (2012) epidemics, immunosuppression following transplantation was not associated with increased mortality. Immunosuppression might even be protective, according to the hypothesis regarding innate immunity as an important driver of severe COVID-19 disease development (3). In line with this, data from Italy suggest that minimally immunosuppressed patients with remote liver transplantation were paradoxically at higher risk than other liver recipients (16).

While experts disagree in the wake of more solid evidence, the general notion is that there is currently not evidence to support dose reduction of immunosuppressants in healthy liver recipients. In liver recipients who are infected with SARS-CoV-2, dose reduction should be discussed and individually weighed against the risk of rejection, particularly in cases of high-dose steroid therapy, MMF-induced leukopenia or systemic fungal or bacterial co-infections (3,13) (Figure 3). As a note of caution, some data suggest prolonged viremia and hence risk of disease transmission in patients receiving immunosuppressive therapy (3).

References

1. Zhang C, Shi L, Wang FS. Liver injury in COVID-19: management and challenges. *Lancet Gastroenterol Hepatol* 2020;5:428-430.
2. Chen N, Zhou M, Dong X, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet* 2020;395:507-513.
3. AASLD. Clinical best practice advice for hepatology and liver transplant providers during the COVID-19 Pandemic: AASLD Expert Panel Consensus Statement.. Released: May 14, 2020 <https://www.aasld.org/sites/default/files/2020-05/AASLD-COVID19-ClinicalInsights-May142020-FINAL.pdf> 2020.

4. Richardson S, Hirsch JS, Narasimhan M, et al. Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area. *JAMA* 2020.
5. Lan J, Ge J, Yu J, et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* 2020;581:215-220.
6. Jia HP, Look DC, Shi L, et al. ACE2 receptor expression and severe acute respiratory syndrome coronavirus infection depend on differentiation of human airway epithelia. *J Virol* 2005;79:14614-21.
7. Mendez-Sanchez N, Valencia-Rodriguez A, Qi X, et al. What Has the COVID-19 Pandemic Taught Us so Far? Addressing the Problem from a Hepatologist's Perspective. *J Clin Transl Hepatol* 2020;8:0024.
8. Sun J, Aghemo A, Forner A, et al. COVID-19 and liver disease. *Liver Int* 2020.
9. Cai Q, Huang D, Yu H, et al. COVID-19: Abnormal liver function tests. *J Hepatol* 2020.
10. Neurath MF. Covid-19 and immunomodulation in IBD. *Gut* 2020.
11. Team CC-R. Preliminary Estimates of the Prevalence of Selected Underlying Health Conditions Among Patients with Coronavirus Disease 2019 - United States, February 12-March 28, 2020. *MMWR Morb Mortal Wkly Rep* 2020;69:382-386.
12. Guan WJ, Ni ZY, Hu Y, et al. Clinical Characteristics of Coronavirus Disease 2019 in China. *N Engl J Med* 2020;382:1708-1720.
13. D'Antiga L. Coronaviruses and Immunosuppressed Patients: The Facts During the Third Epidemic. *Liver Transpl* 2020.
14. Boettler T, Newsome PN, Mondelli MU, et al. Care of patients with liver disease during the COVID-19 pandemic: EASL-ESCMID position paper. *JHEP Rep* 2020;2:100113.
15. Bhoori S, Rossi RE, Citterio D, et al. COVID-19 in long-term liver transplant patients: preliminary experience from an Italian transplant centre in Lombardy. *Lancet Gastroenterol Hepatol* 2020;5:532-533.
16. Pereira MR, Mohan S, Cohen DJ, et al. COVID-19 in solid organ transplant recipients: Initial report from the US epicenter. *Am J Transplant* 2020.

Artificial intelligence in the detection and characterisation of Barrett's related neoplasia: a review of recent literature

Giulio Antonelli¹, Alanna Ebigbo², Cesare Hassan¹

¹ Gastroenterology Unit, Nuovo Regina Margherita Hospital, Rome, Italy

² III Medizinische Klinik, Universitätsklinikum Augsburg, Augsburg, Germany

Esophageal adenocarcinoma, which is steadily rising in the last decades in most Western countries, is related to pre-neoplastic conditions like Barrett's esophagus (BE). Early BE associated neoplasia is still very often missed at surveillance endoscopy, when this is undergone by non-BE-expert endoscopists.

This is because majority of community endoscopists are rarely confronted with BE and even less with early BE associated neoplasia and are unfamiliar with the subtle changes associated. A recent attractive option is Artificial Intelligence (AI). In contrast to human-derived methods, deep learning systems autonomously learn to distinguish characteristic features within the images using multiple processing layers. In this way, computer aided detection (CAD) systems can recognise discriminative features for neoplastic images that differ from those commonly used and known by the human operator and, most importantly, can acquire such a fast image processing rate to be able to be used in real time during an endoscopical examination. In the setting of BE related neoplasia, this can mean both identifying possible dysplastic tissue as well as target the operators' biopsies to the spots where dysplasia can be present with the higher probability. A synthetic review of the latest evidence reporting the use of deep learning/convolutional neural networks for the development of CAD systems for the detection of high-grade dysplasia/early esophageal cancer within BE is presented here.

BARRETT'S ESOPHAGUS

Esophageal adenocarcinoma (EAC) has been steadily rising in the last decades in most Western countries (1).

This has been related to an increase in risk factors such as obesity, as well as pre-neoplastic conditions like Barrett's esophagus (BE).

BE is the condition in which a metaplastic columnar mucosa replaces an esophageal squamous mucosa damaged by gastroesophageal reflux disease (GERD) (2,3) (Figure 1).

The presence of intestinal metaplasia in the distal

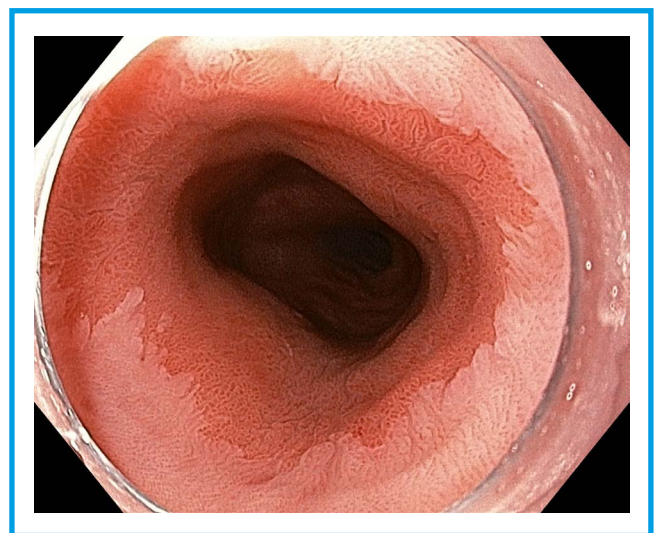


Figure 1. Long-Segment Barrett's Esophagus

esophagus (BE) is completely asymptomatic and has been strongly associated with a higher predisposition to the development of EAC (4). Recent, high-quality studies have shown that the risk of EAC in the general population of patients with nondysplastic BE ranges from 0.1 to 0.3% per year (5) (Figure 2).

DIAGNOSIS

Diagnosis of BE is based on findings during endoscopy that columnar mucosa extends above the gastroesophageal junction and into the distal esophagus, coupled with histopathological confirmation of presence of columnar metaplasia (3). Endoscopically, the gastroesophageal junction is recognised as the most proximal extension of gastric folds, while columnar mucosa is usually salmon pink coloured, differently from the pale looking esophageal squamous mucosa. The extent of esophageal columnar metaplasia determines whether a “long-segment” (≥ 3 cm) or “short-segment” (< 3 cm) BE is present (6). International guidelines recommend different surveillance intervals based on the extent of BE (6). In detail, patients with an irregular junction or columnar lined esophagus of less than 1 cm, no routine biopsies or endoscopic surveillance are recommended. For “short segment BE”, surveillance is recommended every 5 years, while for “long-segment” BE, the recommended interval for endoscopic surveillance is 3 years (6).

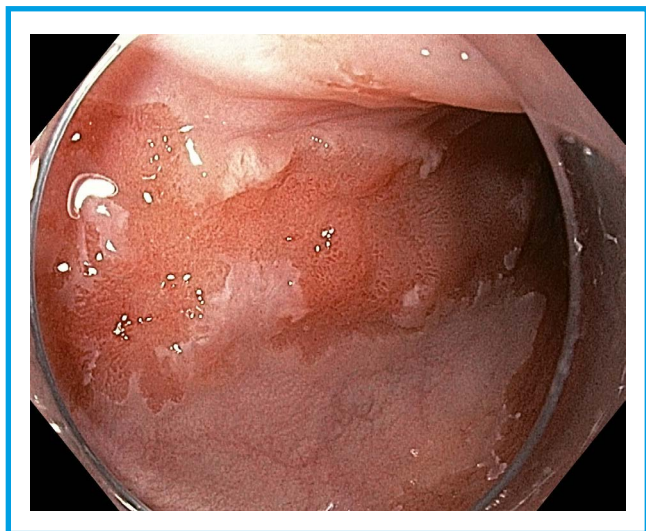


Figure 2. Barrett's Esophagus with Early Esophageal Carcinoma

SCREENING

Screening for BE and of EAC in patients with BE has yet to be associated robustly with reduced death from EAC, but has been associated to earlier-stage tumors and improved survival (7).

When detected early, dysplastic BE can be treated endoscopically, resulting in improved outcomes and quality of life and sparing invasive surgical treatments.

The standard-of-care for detecting dysplastic tissue has been for a long time the well-established “Seattle Protocol”, that recommends taking “four-quadrant, 2-cm endoscopic biopsies performed at closely timed intervals” in the whole esophageal segment that appears to be metaplastic (8).

This paradigm has shifted in recent years towards a more “targeted” approach with the use of high definition endoscopy and advanced endoscopic imaging in the thorough interrogation of the whole metaplastic segment (6,9). These recent technological advancements permit expert endoscopists to reach very high early neoplasia detection rates.

HOW TO IMPROVE DETECTION?

However, early BE associated neoplasia is still very often missed at surveillance endoscopy, when this is undergone by non-BE-expert endoscopists.

This is because majority of community endoscopists are rarely confronted with BE and even less with early BE associated neoplasia and are unfamiliar with the subtle changes associated. In this setting, both classroom-based and online training modules have been proposed, to a certain extent of success, to improve diagnostic capabilities of both expert and non-expert endoscopists (10).

A more recently developed and attractive option is Artificial Intelligence (AI) or machine learning, especially after the successful application of Convolutional Neural Networks.

In contrast to human-derived methods, deep learning systems autonomously learn to distinguish characteristic features within the images using multiple processing layers. In this way, computer-aided detection (CAD) systems can recognise

discriminative features for neoplastic images that differ from those commonly used and known by the human operator. In addition, a CAD system developed with deep learning techniques can acquire such a fast image processing rate to be able to be used in real time during an endoscopic examination. In the setting of BE related neoplasia, this can mean both identifying possible dysplastic tissue as well as target the operators' biopsies to the spots where dysplasia can be present with the higher probability.

AI SYSTEMS FOR DETECTION OF BE

Recently, a number of systems have been developed to aid detection (CADE) of BE related neoplasia. Among AI systems for the detection and characterisation of upper-GI neoplasia, the setting of diagnosis and surveillance of BE is one of the most studied in Western countries, differently from AI systems for the detection of gastric pre-cancerous conditions and early gastric cancer, that are more studied in Asian settings for the higher prevalence of the disease.

Probably the most studied AI system for detection of BE related neoplasia is the one developed by the "ARGOS" consortium, which has published four papers (11-14) following the development phases of a state-of-the-art CAD system for the detection of BE related neoplasia.

In particular, the most recent version of the system has been recently developed, tested and validated first in an experimental setting (12) and then in a clinical, live, setting (13).

This system is designed to give firstly a diagnosis of neoplastic or non-neoplastic image. After this binary classification, the system automatically delineates a "heatmap" around the region identified as neoplastic. Subsequently, the system automatically flags the most abnormal site in the delineated region, indicating the primary location to perform biopsies. In light of the limitations of optical detection of BE related neoplasia discussed above, this system appears to be clinically valuable for the surveillance of BE.

The training and experimental validation of

this system is the perfect example of a correct development of an AI system for endoscopy.

In brief, the system was trained and tested through the stepwise use of five different image datasets. Subsequently, it was tested against the diagnostic performance of 53 community, non-BE-expert endoscopists. This benchmarking by community endoscopists is the most efficacious way to test and measure the potential diagnostic improvement provided by the CAD system.

On the external validation image set, the CADE system showed an accuracy of 87.5%, a sensitivity of 92.5%, and a specificity of 82.5%, in correctly classifying endoscopic BE images as neoplastic or non-neoplastic.

In comparison, these values are superior to the classification performance of non-expert endoscopist human assessors, both senior and junior. In detail, overall performance of the benchmarking endoscopists showed an accuracy of 73%, a sensitivity of 71.8%, and a specificity of 74.3%, in correctly classifying endoscopic BE images as neoplastic or non-neoplastic.

Following the testing of this CAD system, it was applied for the first time to a live, real-time scenario (13). The system was tested on a total of 20 patients, 10 with confirmed non dysplastic BE and 10 with confirmed dysplastic BE.

To test this system in vivo, first the operating endoscopist washed and cleaned the esophageal mucosa. Then, starting from the GE junction, with the endoscope in a centralised position, a series of three images was taken for every 2 cm of metaplastic mucosa. These white-light endoscopy images were automatically transferred to the CAD system via the output of the endoscopy system. Every 2 cm, before CAD analysis, the endoscopist would indicate if a visible lesion was present. The regions identified by the CAD system and/or by the expert endoscopists as containing a lesion were resected or biopsied. If neither the CAD nor the endoscopists identified a lesion, biopsies were taken random as per BE surveillance protocol.

In this study, the CADE system largely confirmed the results of the offline validation studies, as it showed an image-based accuracy, sensitivity, and

specificity of 84%, 76%, and 86%, respectively. In addition, the system would grade the degree of confidence of its predictions. Diagnostic accuracy of the high confidence predictions was around 91% for accuracy, sensitivity, and specificity. Only one other CADe system for the same purpose has been reported that followed a similar development, undergoing first an offline, ex-vivo validation, and subsequently a live validation on patients during endoscopy (15,16).

In the first paper (15), the training and testing phase was reported, using two separate databases of optimal endoscopic images previously delineated by expert endoscopists.

This system was trained both on white light images and on narrow band imaging (NBI) imaging. In this offline setting, the CAD system showed a sensitivity of 97% and a specificity of 88% for WL images, and a sensitivity and specificity of 94% and 80% for NBI images, in diagnosing early esophageal carcinoma in the context of a BE. In contrast, 13 expert endoscopists were asked to evaluate the same images evaluated by the CAD system, and showed a mean sensitivity and specificity of 76% and 80% on one image dataset and 99% and 78% on the other image dataset.

The same group then tested the same system during live endoscopy (16), after improving speed and capacity of imaging processing. In this system, the CAD can be turned on at any time, and when activated, will continuously provide a prediction of probability of neoplasia in real time with a blue bar. In this system, an image is classified as neoplastic if the probability of neoplasia exceeds 90%. In real time endoscopy, the system showed a sensitivity and specificity of 83.7% and 100.0%, respectively, and an overall accuracy of 89.9%.

Other two CAD systems have been reported for the purpose of detecting early esophageal carcinoma in the context of BE, but currently only the testing offline phase has been reported. However, the results are encouraging. One system from the US (17) correctly detected early neoplasia with sensitivity of 96.4%, specificity of 94.2%, and accuracy of 95.4%, on 458 test images.

The other system, developed in a US/Egypt

collaboration (18), sensitivity of 88%, specificity of 92%, and accuracy of 90%, on 100 test images.

These two studies, although promising, show a higher risk of bias than the studies reported earlier. In detail, a reduced number of training images and sets seem to have been used, generating the risk of selection bias, and only selected images from referral centers were used, generating risk of spectrum bias.

We have presented a synthetic review of the latest evidence reporting the use of deep learning/convolutional neural networks for the development of CAD systems for the detection of high-grade dysplasia/early esophageal cancer within BE.

The importance of tools to aid endoscopists in this task cannot be stressed sufficiently. Indeed, it is known that the low prevalence of BE in the community setting results in suboptimal rates of early neoplasia detection. Miss rates of early neoplasia arising in BE have been reported as high as 40% (19,20).

CONCLUSIONS

The CAD systems that have been reviewed in this paper have shown very satisfying results in increasing BE-related neoplasia detection rates.

This can be especially relevant in the community-based endoscopy setting, where training in optical diagnosis can be cumbersome and could be overlooked by the daily activities, often already a great burden for operators.

In addition, it is known that also a high diagnostic ability must be continuously updated and retrained, generating additional practical issues.

In this context, the ready availability of a system that can help the endoscopist towards the correct diagnosis, highlighting suspect lesions and guiding biopsies, would result in a higher rate of treatable early neoplasia, ultimately improving patient outcomes and healthcare costs.

It appears that there is urgent need for a widespread implementation of such tools, prompting the design of large, multicenter randomised trials that can further validate existing systems and promote their use in clinical practice.

References

1. Lagergren J, Lagergren P. Oesophageal cancer. *BMJ*. 2010 Nov 26;341:c6280.
2. Sharma P. Clinical practice. Barrett's esophagus. *N Engl J Med*. 2009 Dec 24;361(26):2548-56.
3. Spechler SJ, Souza RF. Barrett's esophagus. *N Engl J Med*. 2014 Aug 28;371(9):836-45.
4. Spechler SJ. Barrett esophagus and risk of esophageal cancer: a clinical review. *JAMA*. 2013 Aug 14;310(6):627-36.
5. Desai TK, Krishnan K, Samala N, Singh J, Cluley J, Perla S, et al. The incidence of oesophageal adenocarcinoma in non-dysplastic Barrett's oesophagus: a meta-analysis. *Gut*. 2012 Jul;61(7):970-6.
6. Weusten B, Bisschops R, Coron E, Dinis-Ribeiro M, Dumonceau J-M, Esteban J-M, et al. Endoscopic management of Barrett's esophagus: European Society of Gastrointestinal Endoscopy (ESGE) Position Statement. *Endoscopy*. 2017 Feb;49(2):191-8.
7. Corley DA, Levin TR, Habel LA, Weiss NS, Buffler PA. Surveillance and survival in Barrett's adenocarcinomas: a population-based study. *Gastroenterology*. 2002 Mar;122(3):633-40.
8. Reid BJ, Blount PL, Feng Z, Levine DS. Optimizing endoscopic biopsy detection of early cancers in Barrett's high-grade dysplasia. *Am J Gastroenterol*. 2000 Nov;95(11):3089-96.
9. 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.
10. Bergman JJGHM, de Groof AJ, Pech O, Ragunath K, Armstrong D, Mostafavi N, et al. An Interactive Web-Based Educational Tool Improves Detection and Delineation of Barrett's Esophagus-Related Neoplasia. *Gastroenterology*. 2019 Apr;156(5):1299-1308.e3.
11. van der Sommen F, Zinger S, Curvers W, Bisschops R, Pech O, Weusten B, et al. Computer-aided detection of early neoplastic lesions in Barrett's esophagus. *Endoscopy*. 2016 Apr 21;48(07):617-24.
12. de Groof AJ, Struyvenberg MR, van der Putten J, van der Sommen F, Fockens KN, Curvers WL, et al. Deep-Learning System Detects Neoplasia in Patients With Barrett's Esophagus With Higher Accuracy Than Endoscopists in a Multistep Training and Validation Study With Benchmarking. *Gastroenterology*. 2020 Mar;158(4):915-929.e4.
13. de Groof AJ, Struyvenberg MR, Fockens KN, van der Putten J, van der Sommen F, Boers TG, et al. Deep learning algorithm detection of Barrett's neoplasia with high accuracy during live endoscopic procedures: a pilot study (with video). *Gastrointestinal Endoscopy*. 2020 Jan;S0016510720300146.
14. Groof J de, van der Sommen F, van der Putten J, Struyvenberg MR, Zinger S, Curvers WL, et al. The Argos project: The development of a computer-aided detection system to improve detection of Barrett's neoplasia on white light endoscopy. *United European Gastroenterology Journal*. 2019 May;7(4):538-47.
15. Ebigbo A, Mendel R, Probst A, Manzeneder J, Souza Jr LA de, Papa JP, et al. Computer-aided diagnosis using deep learning in the evaluation of early oesophageal adenocarcinoma. *Gut*. 2019 Jul;68(7):1143-5.
16. Ebigbo A, Mendel R, Probst A, Manzeneder J, Prinz F, de Souza Jr. LA, et al. Real-time use of artificial intelligence in the evaluation of cancer in Barrett's esophagus. *Gut*. 2020 Apr;69(4):615-6.
17. Hashimoto R, Requa J, Dao T, Ninh A, Tran E, Mai D, et al. Artificial intelligence using convolutional neural networks for real-time detection of early esophageal neoplasia in Barrett's esophagus (with video). *Gastrointestinal Endoscopy*. 2020 Jan;S0016510720300262.
18. Ghatwary N, Zolgharni M, Ye X. Early esophageal adenocarcinoma detection using deep learning methods. *Int J CARS*. 2019 Apr;14(4):611-21.
19. Schölvinck DW, van der Meulen K, Bergman JJGHM, Weusten BLAM. Detection of lesions in dysplastic Barrett's esophagus by community and expert endoscopists. *Endoscopy*. 2017 Feb;49(2):113-20.
20. Rodríguez de Santiago E, Hernanz N, Marcos-Prieto HM, De-Jorge-Turrión MÁ, Barreiro-Alonso E, Rodríguez-Escaja C, et al. Rate of missed oesophageal cancer at routine endoscopy and survival outcomes: A multicentric cohort study. *United European Gastroenterol J*. 2019;7(2):189-98.

Big data application in prevention and control of infectious diseases

Tingting Dai^{1,2}, Jiali Ni^{1,2}, Richai Chen¹, Yu Liu³, Yelin Zhu⁴, Jie Zheng⁴, Yi Fang³, Lingling Tang^{1,2}, Guoping Sheng^{1,2}

¹ Department of Infectious Disease, Shulan Hangzhou Hospital Affiliated to Zhejiang Shuren University Shulan International Medical College, Hangzhou 310022, China.

² State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, National Clinical Research Center for Infectious Diseases, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, The First Affiliated Hospital, College of Medicine, Zhejiang University, Hangzhou 310011, China

³ Zhejiang Merit Interactive Network Technology Co., Ltd, Hangzhou 310000, China

⁴ Zhejiang Institute of Medical-care Information Technology, Hangzhou 310000, China

Big data has attracted more and more attentions from various fields, and has been applied in lots of fields, including health care. In addition, with the rapid development of computer science and technology as well as the Internet, novel data sources, such as mobile phone call data, social media data, search engine query data and news articles, have shown great potential to provide health information of individuals. We summarized the characteristics of these novel data sources and illustrated the application of big data in infectious disease field, hoping that the full potential of big data would be realized to prevent and control emerging and reemerging infectious diseases.

INTRODUCTION

Big data is defined as accumulations of data so large and highly complex that its manipulation and management require the application of a series of computing techniques and big data analytic offers the capacity to collect, rationalize and make use of big data for other useful information (1). Big data has attracted more and more attentions from various fields for its volume, variety and velocity (1,2). Positive examples have been reported in lots of fields, such as marketing, climatology, earth sciences and health care (3). In biomedical research big data has created novel methods to improve the

quality of health care (2). In multiple biomedical subdisciplines, bioinformatics is the main field in which big data analytics are currently being applied, largely due to the massive amount and complexity of bioinformatics data. By using sophisticated platforms and tools like gene sequencing mapping tools, making the application of big data in bioinformatics relatively mature. Nevertheless, there is enormous and untapped potential for big data applications in public health informatics (2).

With the epidemic and pandemic potential of new diseases, the risks of infection has increased (4). For infectious diseases, big data has been using in real-time surveillance of disease outbreaks, developing dynamic risk maps and assessing the outcome of strategy (5-7). With the fast developing of computer science, technology, internet, as well as the increasingly using of novel data source, such as mobile phone data, web search query data and social media data, making it possible to gather data faster for population fluxes monitoring and diseases tracking (8,9). We summarized the characteristics of these novel data sources and illustrated the application of big data in prevention and control of infectious diseases by providing real-time Infectious disease surveillance, modeling of disease transmission and simulating the outcome of health interventions, hoping that the full potential of big data can be used to improve situational awareness and response for old and new infectious disease outbreaks.

CHARACTERISTICS OF COMMONLY USED NOVEL DATA SOURCES

Characteristics of mobile phone call data

Traditional data sources for measuring human activities mainly from population, housing censuses, travel history surveys, cross-border and traffic surveys, which can often be lacking and outdated in resource-poor settings (10). Call detail records (CDRs) are routinely collected by mobile phone operators (10,11). Each CDR contains the location of the routing tower and timing of each call and SMS communication between subscribers (11). When an individual makes a call or sends a message, it normally will be routed through the closest tower in the network. Combined with geographic coordinates of relevant towers, the tower level location of each communication can be identified, and the movement of individual mobile users between different calls can be derived further (10). Therefore, data derived from CDRs can paint a dynamic picture of mobility and changing patterns across spatial and temporal scales.

As mobile phones have become an integral part of modern life, with a penetration rate of 66% of the global population in 2017, CDRs data has become a novel and powerful source of information on quantifying human mobility (10). Many recent studies have emphasized that quantitative analysis of location data from large numbers of mobile phone users can significantly improve our understanding of human mobility (6,8,11). A retrospective analysis of mobile phone call data in Sierra Leone showed it can, at minimum cost, provide key information needed for measuring the mobility impact of travel restrictions, lockdowns and regional quarantines in near real-time during the Ebola epidemic (8). However, the problem in mobile phone call data is that it cannot measure spatial movements better than tower-level spatial resolution. That is to say, the spatial resolution of CDR data is limited by the density of cell towers, which is limit-used in some undeveloped regions (8). And estimates are limited to domestic movements, as it is more difficult to obtain CDRs from operators in different countries to trace international travels (10). And previous work found that the penetration of mobile phone was heterogeneous and distinct regional,

gender-related, and socioeconomic variations exist, with particularly low ownership among rural communities and poor people (12). More efforts are needed to establish the most promising uses of mobile phone data and technologies for public health.

Characteristics of social media data

Social media is defined as services that are based on Web 2.0 technologies and largely rely on user-generated content, by which individuals create profiles and develop social networks online (13). In 2016, there were nearly 2.34 billion social media users globally (13). According to a recent report of China Internet Network Information Center (CNNIC), as of June 2019, China had 854 million netizens and their Internet penetration had reached 61.2% (14). Social media sites have been proposed as a data source because they have the potential to offer real-time access to millions of short and geographically localized messages which can be effectively used in a number of ways, including tracking users' interest, concerns as well as measuring public awareness and reactions (15,16).

Because location is identified using a combination of the phone's internal GPS and connected WiFi devices and cell towers, these data are as spatially refined as GPS tracker data. Therefore, from social media big data, we can capture rich features of mobility movements within and between cities and even between countries (10). Elkin et al. have gotten the conclusion that network based model of social media big data can estimate how diseases diffuse across geographical locations and provide forecasting for one week in advance, while earlier approaches that used airline traffic to study contagious disease spread was unsuccessful in predicting ILI activity across geographies (17).

A systematic review of literature indicated that Twitter is the most often studied social media (13). Researchers studied Twitter data to track levels of disease activates and public concerns during the influenza A H1N1 pandemic, predicting that the diffusion of flu and measure the sentiments about H1N1 influenza vaccination (15,17,18). Despite these promising results, there are several limitations in using social medias to make inferences. One is

media coverage, as the penetration of smart-phone is still limited in less-developed countries and the groups that are most at risk of being infected, such as aged or young children, tend to be underrepresented to social media (10,19). Among those who are using Twitter, only a small percentage has GPS location on. The second limitation is the accuracy of social media surveillance systems declines with media attention because media attention increases “chatter” that masks signs of true prevalence (19). Third, social media cannot be used to monitor or track infectious diseases in areas with no access to social networks such as Twitter and Facebook (20).

Characteristics of other Internet streams such as search engine query data and news articles

Public is becoming more dependent on the Internet and millions of users tend to search online for health-related information especially the early stages of illness, which makes web search queries a valuable source of information on collective health trends (21). For example, in the United States, Google web search queries was used to detect regional outbreaks of influenza consistently 1-2 weeks ahead of CDC ILI surveillance reports, because of quick access to search queries (21). Subsequently, Google Flu Trends was developed and got its generic services, Google Trends, serves as a convenient, easy, and accessible source of search data (5). Google search queries data was also used to monitor Dengue epidemics and predict Zika cases before the outbreak (9,22). Baidu, the most popular search engine in China, has been used to detect seasonal influenza activity and monitor spatial and temporal trends of avian influenza A(H7N9) in China (23,24). Furthermore, in Chowell’s work, publicly available online reports were collated and analyzed to gain information on exposure patterns and transmission chains for emerging infections, in the near absence of detailed epidemiological data (25). However, the main disadvantage for social medias is that rural areas and developing nations are likely to have limited Internet access currently, while web-query based surveillance depends on adequate web search volume (9). In addition, the outcome of search queries can be influenced by panic or curiosity when an emerging disease outbreaks (5). Despite

these limitations, internet search query data is being explored as a low-cost approach to estimating disease activity in near real-time.

THE APPLICATION OF BIG DATA IN PREVENTION AND CONTROL OF INFECTIOUS DISEASES

Big data for real-time Infectious disease surveillance

In public health, traditional surveillance systems remain primarily based on manually collected and coded data, they are costly and slow to accumulate, difficult to share for analysis at the same time (26). Furthermore, reports from these systems are tend to be notorious for time lagging and missing spatial information (21). Hay et al. discussed the opportunities for using big data on global infectious disease surveillance (27). Drawing on a greater diversity of data sources and types is one way to improve information collection and strengthen surveillance systems (26,28). Although there is a strong possibility using large volumes of electronic health medical records to monitor infectious diseases, it relies on case reporting by physicians (3). Mobile phone data and Internet data streams have been proved to be beneficial to proactive reporting and real-time surveillance, which are better for public health than a reactive approach (29). Real-time surveillance would inform public health care practitioners of an outbreak or epidemic, enabling them to implement preventive interventions and prepare adequate medical resources, thereby reduce morbidity and mortality. Studies have shown that internet data streams, such as web search query data and social media data, could be useful to monitor disease in real time and predict the level of disease activity (5,9,15). Faced with a global health crisis such as the epidemic of COVID-19, it has been proved to be feasible for repurposing of existing data collection method to collect real-time population-scale data. By leveraging the digital backbone of an existing application, Drew et al. developed the COVID Symptom Tracker mobile app, which could help to monitor the dynamics of COVID incidence (30). Predicting and understanding how future outbreaks are spreading from location to location can help to coordinate limited resources, stratifying and

segmenting populations at risks and give hospitals a strong warning about when outbreaks could occur in particular areas (17).

It became clear that current demographic transitions have fundamental effects on the dynamics of infectious diseases by introducing pathogens into susceptible populations or by changing the frequency of contacts between infected and susceptible individuals, which may bring about an outbreak or a pandemic (31,32). The recent studies of mobile phone calling data and social media location data have shown that it is possible to trace fine scale movement across large numbers of individuals (11,33). Unlike risk maps generated from clinical case counts alone, an integration of epidemiological surveillance data with human mobility data can help to distinguish areas of frequent importation as well as high transmission (6). In the USA, Elkin et al. presented a novel network-based model to analyze social media big data and data on human mobility to make predictions about the spread of influenza-like illnesses (ILI) across states in based on geographical surroundings at 76% accuracy (17). However, modeling disease diffusion is a complex concept, and there is still much room for development in this area.

Big data for modeling of disease transmission and simulating the outcome of health interventions

Disease simulation models aim to represent reality in a simplified form so that the behavior of a disease system can be better understood (28). An interdisciplinary team learned key lessons for emerging infectious disease preparedness and response from the Ebola outbreak, including expanding prediction modeling and intervention studies (29). The wealth of information promised by big data, combined with the development of mathematical modeling of infection transmission could provide frameworks for quantifying the trend of epidemic and for simulating the outcome of prevention and control interventions (26,29). Population movements and network interactions informed from the mobile phone data and internet data streams have been widely used in epidemic modeling (7,17,34). Such models - if developed correctly and appropriately validated - can then be used to assess strategy and guide policy

development with a view to reducing the impact of a disease event, particularly during an epidemic of an emerging infectious disease (28).

Take the COVID-19 as an example, since the start of the COVID-19 epidemic in late 2019, a number of prediction models have been established by scholars and organizations from various countries to predict the epidemic trend (7,34). Wu et al. estimated that 75 815 individuals (95% CrI 37 304-130 330) individuals had been infected in Greater Wuhan as of Jan 25, 2020 (34). One of the limitations was that their epidemic forecast was based on inter-city mobility data from 2019, which took no account of the measures of strict control in 2020. In another modeling, the effects of public health intervention measures restricting migration was modeled (7). Yang et al. integrated population migration data before and after public health interventions was implemented to restrict travel in Hubei province and most updated COVID-19 epidemiological data into the Susceptible-Exposed-Infectious-Removed (SEIR) model to derive the epidemic curve, which was proved to fit well with the actual epidemic trend subsequently (7). The modified SEIR model predicted a single epidemic peak on February 20 with 42 792 (95% CI: 30 149-52 941) cases in Hubei province, close to the number reported by the National Health Commission of the People's Republic of China (35). In their work, the effect of initiating interventions five days before and after the actual intervention time was also model and the outcome further illustrated that the implementation of control measures was indispensable in reducing the eventual COVID-19 epidemic size and the policy of strict interventions should be adopted sooner rather than later.

Big data has played an active role in China's fighting with the epidemic of COVID-19

The World Health Organization (WHO) have made the assessment that COVID-19 was characterized as a pandemic. But it would be the first pandemic in history that could be controlled (36). The epidemic in China was significantly declining during March. Besides the containment strategies adopted in the earliest stage of the outbreak to limit population mobility, including case isolation, lockdown epidemic areas and close

the high way out of the epidemic areas, big data has played a particularly active role in fighting with the epidemic in China in general (37).

Health QR code system is an innovative result of Hangzhou's efforts to control the epidemic by using big data and mobile internet technologies (38). It was promoted in Zhejiang province and other provinces later to prevent the spread of COVID-19 (39). This system allows citizens to report their health status and travel history in an effective way, making it faster to detect suspicious infections and isolate infectious cases as soon as they get diagnosed. And the colors of health QR code based on the information reported by citizens, either a green, red, or yellow, provide a convenient and efficient way to guide travel and facilitate the resumption of work (38,40,41).

Close contact measurement instrument is a big data platform, on which citizens can query whether they are close contacts after inputting their name and id number (42). It also helps to reduce the possibility of infection when travelling.

Diagnosed Cases in Community is an application that provides an approach for people to obtain the information of the COVID-19 cases distribution in local communities on the map, which shows case number and location (41). To some extent, this application can remind people to take protective measures and make travel plan carefully.

At the Munich Security Conference in mid-February, Director-General of the WHO, Tedros Adhanom Ghebreyesus, has termed the Coronavirus as infodemic, which means an epidemic of information, including misinformation, rumors, etc. Therefore, obtaining correct and timely information is of the utmost importance in combatting Coronavirus (40). Apart from long-distance travel restrictions and strict local control measures, wise and timely use of big data and programs was one of the key factors of China's success in containing COVID-19(40, 43, 44).

CHALLENGES AND OPPORTUNITIES

While big data holds significant promise for prevention and control of infectious diseases, three major challenges related to big data application in the infectious disease field are not to be neglected:

diverse data challenges, ethical challenges and methodological challenges (1,45).

Data sets from various sources are massive in size and the quality varies widely. That is to say, the majority of data is irrelevant to the specific certain disease (45). Therefore, it is difficult for researchers to control, manage, and process the multiple data sources efficient and timely. With the development of information and computing science as well as internet technology, steady streaming data is produced through multiple channels. Search engine query data, Twitter data, internet news articles and human mobility information, web-based data, and genome databases and electronic health records are all used to resolve the risk of infection (5,25). Although novel data sources can bring about a considerable wealth of information at low cost, several problems still exist. One of the limitations is the biases generated by unbalanced mobile ownership and mobile internet access, which lead to lack of information among those individuals (26). Despite the highly detailed positions from mobile users can provide powerful, visualized and real-time data sets on human activities at multiple temporal and spatial scales, the interpretation of mobility estimates must take biases into consideration (10,11). And access to these data is still limited. Because of the confidentiality and ethical issues, mobile positioning data automatically generated from individuals is limited by the telecom, internet and data-protection regulations in many countries. Although mobile phone operators and technology companies tend to be more receptive to provide anonymous data for research as the usefulness of these data has been demonstrated in public health, agreements between operators and research groups should be negotiated to access these data, which is a nontrivial and time-consuming task (10,32). By contrast, data from internet tools is more accessible. For example, some researchers utilized Twitter data because of its public free availability and ease of access (17,19,46). Twitter provides historical Twitter data and tweets posted in real time through application programming interfaces (APIs) (17). And geolocated Twitter data can be also provided through the API (46). And publicly available online news reports and investigative reports published by authoritative media outlets, the WHO and other health

care authorities found they are useful in identifying key information on exposure and transmission patterns during epidemic emergencies (25).

During the current COVID-19 pandemic, it's urgent for technology companies and researchers to collaborate on sharing data in a legal, ethical and privacy-protecting way (47).

However, the application of big data in infectious diseases is coupled with some ethical impacts (1). Because of availability of information through various channels, needs for privacy and security are increasing for fear that individual health information will be exploited by insurance companies or unregulated commercial enterprises (48,49). The issues of how mobile phone operators and technology companies should provide access to these data, and to whom, remain to be figured out (8). Strong guidelines for legislative and regulatory, as well as secure platforms with better communication standards and protocols are greatly needed to safeguard confidentiality of the information and ensure the ethical use of data (2,8,10). And it's a potential way to provide legally enforceable commitments that the data used for prevention and control of infectious diseases will be deleted after the epidemic, which may help to gain consent and trust of many people (50). In addition, at the policy level, increasing transparency of data governance are paramount for balancing access to data with individual privacy (49).

It is a challenge to apply novel big data such as internet data to overfit the model, which was lack of accuracy (13,48). The progress of computer science

and the improvement of computing speeds have led to the extensive application of artificial intelligence in many fields (31). There are two categories of AI and machine learning methods applicable to infectious based on two different learning strategies: supervised learning and unsupervised learning. Supervised learning, such as Support Vector Machine (SVM), Artificial Neural Network (ANN) and Bootstrap Aggregating, can handle the problem of classification and regression efficiently (45). In addition, deep learning architectures have been broadly applied into prediction and classification, social network filtering and bioinformatics and showed great potential in infectious diseases (45). For instance, Hay et al. believed that the use of machine learning and crowdsourcing had made it possible to develop a continually updated atlas of infectious diseases (27). Big data provides an unprecedented opportunity in infectious diseases, it is not intended to be a replacement for traditional epidemiology, but by taking advantage of readily available data essentially provided by millions of individuals, making it a useful and low-cost complement.

There is much work needs to be done to demonstrate the capability and reduce the risks of using relatively novel big data. In the era of big data, it is vital method to collaborate with experts in other fields. Computer scientists can use machine learning to valid information in big data sources efficiently (27). Future developments in machine learning and artificial intelligence will facilitate applying big data in infectious disease research, surveillance, and prevention.

References

1. Garattini C, Raffle J, Aisyah DN, Sartain F. Big Data Analytics, Infectious Diseases and Associated Ethical Impacts. *Philos Technol* 2019; 32:69-85.
2. Luo J, Wu M, Gopukumar D, Zhao Y. Big Data Application in Biomedical Research and Health Care: A Literature Review. *Biomed Inform Insights* 2016; 8:1-10.
3. Simonsen L, Gog JR, Olson D, Viboud C. Infectious Disease Surveillance in the Big Data Era: Towards Faster and Locally Relevant Systems. *J Infect Dis* 2016; 214:380-5.
4. Heymann D, Rodier G. Hot spots in a wired world: WHO surveillance of emerging and re-emerging infectious diseases. *Lancet Infect Dis* 2001; 1:345-53.
5. Carneiro HA, Mylonakis E. Google trends: a web-based tool for real-time surveillance of disease outbreaks. *Clin Infect Dis* 2009; 49:1557-64.
6. Chang HH, Wesolowski A, Sinha I, Jacob CG. Mapping imported malaria in Bangladesh using parasite genetic and human mobility data. *Elife* 2019; 8:e43481.
7. Yang Z, Zeng Z, Wang K, et al. Modified SEIR and AI

- prediction of the epidemics trend of COVID-19 in China under public health interventions. *Journal of thoracic disease* 2020; 12:165-74.
8. Peak CM, Wesolowski A, Zu Erbach-Schoenberg E, et al. Population mobility reductions associated with travel restrictions during the Ebola epidemic in Sierra Leone: use of mobile phone data. *Int J Epidemiol* 2018; 47:1562-70.
9. Chan EH, Sahai V, Conrad C, et al. Using web search query data to monitor dengue epidemics: a new model for neglected tropical disease surveillance. *PLoS Negl Trop Dis* 2011; 5:e1206.
10. Lai S, Farnham A, Ruktanonchai NW, et al. Measuring mobility, disease connectivity and individual risk: a review of using mobile phone data and mHealth for travel medicine. *J Travel Med* 2019; 26:taz019.
11. Wesolowski A, Metcalf CJ, Eagle N, et al. Quantifying seasonal population fluxes driving rubella transmission dynamics using mobile phone data. *Proc Natl Acad Sci U S A* 2015; 112:11114-9.
12. Wesolowski A, Eagle N, Noor AM, et al. Heterogeneous mobile phone ownership and usage patterns in Kenya. *PLoS One* 2012; 7:e35319.
13. Tang L, Bie B, Park SE, et al. Social media and outbreaks of emerging infectious diseases: A systematic review of literature. *Am J Infect Control* 2018; 46:962-72.
14. China Internet Network Information Center. The 44th Statistical Report on Internet Development in China. Available at: http://www.cnnic.net.cn/hlwfzyj/hlwxzbg/hlwtjbg/201908/t20190830_70800.htm. Accessed 11 March 2020.
15. Signorini A, Segre AM, Polgreen PM. The use of Twitter to track levels of disease activity and public concern in the U.S. during the influenza A H1N1 pandemic. *PLoS One* 2011; 6:e19467.
16. Fung IC, Fu KW, Ying Y, et al. Chinese social media reaction to the MERS-CoV and avian influenza A(H7N9) outbreaks. *Infect Dis Poverty* 2013; 2:31.
17. Elkin LS, Topal K, Bebek G. Network based model of social media big data predicts contagious disease diffusion. *Inf Discov Deliv* 2017; 45:110-20.
18. Salathé M, Khandelwal S. Assessing vaccination sentiments with online social media: implications for infectious disease dynamics and control. *PLoS Comput Biol* 2011; 7:e1002199.
19. Broniatowski DA, Paul MJ, Dredze M. National and local influenza surveillance through Twitter: an analysis of the 2012-2013 influenza epidemic. *PLoS One* 2013; 8:e83672.
20. Jahanbin K, Rahmanian F, Rahmanian V. Application of Twitter and web news mining in infectious disease surveillance systems and prospects for public health. *GMS Hyg Infect Control* 2019; 14:Doc19.
21. Ginsberg J, Mohebbi MH, Patel RS, et al. Detecting influenza epidemics using search engine query data. *Nature* 2009; 457:1012-4.
22. Morsy S, Dang TN, Kamel MG, et al. Prediction of Zika-confirmed cases in Brazil and Colombia using Google Trends. *Epidemiol Infect* 2018; 146:1625-7.
23. Chen Y, Zhang Y, Xu Z, et al. Avian Influenza A (H7N9) and related Internet search query data in China. *Sci Rep* 2019; 9:10434.
24. Yuan Q, Nsoesie EO, Lv B, et al. Monitoring influenza epidemics in china with search query from baidu. *PLoS One* 2013; 8:e64323.
25. Chowell G, Cleaton JM, Viboud C. Elucidating Transmission Patterns From Internet Reports: Ebola and Middle East Respiratory Syndrome as Case Studies. *J Infect Dis* 2016; 214:421-6.
26. Bansal S, Chowell G, Simonsen L, et al. Big Data for Infectious Disease Surveillance and Modeling. *J Infect Dis* 2016; 214:S375-S9.
27. Hay SI, George DB, Moyes CL, et al. Big data opportunities for global infectious disease surveillance. *PLoS Med* 2013; 10:e1001413.
28. Brookes VJ, Hernandez-Jover M, Black PF, et al. Preparedness for emerging infectious diseases: pathways from anticipation to action. *Epidemiol Infect* 2015; 143:2043-58.
29. Jacobsen KH, Aguirre AA, Bailey CL, et al. Lessons from the Ebola Outbreak: Action Items for Emerging Infectious Disease Preparedness and Response. *Ecohealth* 2016; 13:200-12.
30. Drew DA, Nguyen LH, Steves CJ, et al. Rapid implementation of mobile technology for real-time epidemiology of COVID-19. *Science (New York, NY)* 2020.
31. Bedford J, Farrar J, Ihekweazu C. A new twenty-first century science for effective epidemic response. *Nature* 2019; 575:130-6.
32. Wesolowski A, Buckee CO, Engo-Monsen K, et al. Connecting Mobility to Infectious Diseases: The

- Promise and Limits of Mobile Phone Data. *J Infect Dis* 2016; 214:S414-S20.
33. Rocklov J, Tozan Y, Ramadona A, et al. Using Big Data to Monitor the Introduction and Spread of Chikungunya, Europe, 2017. *Emerg Infect Dis* 2019; 25:1041-9.
34. Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *The Lancet* 2020; 395:689-97.
35. National Health Commission of the People's Republic of China. Available at: http://www.nhc.gov.cn/xcs/yqtb/list_gzbd_5.shtml. Accessed 10 March 2020.
36. World Health Organization. WHO Director-General's opening remarks at the media briefing on COVID-19 - 9 March 2020. Available at: <https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---9-march-2020>. Accessed 13 March 2020.
37. Chen W, Wang Q, Li YQ, et al. Early containment strategies and core measures for prevention and control of novel coronavirus pneumonia in China *Zhonghua Yu Fang Yi Xue Za Zhi* 2020; 54:1-6.
38. Huang Y, Wu Q, Wang P, et al. Measures Undertaken in China to Avoid COVID-19 Infection: Internet-Based, Cross-Sectional Survey Study. *Journal of medical Internet research* 2020; 22:e18718.
39. Zhejiang China. Zhejiang to promote health QR code system. Available at: http://www.ezhejiang.gov.cn/2020-02/13/c_452063.htm. Accessed 12 March 2020.
40. Hua J, Shaw R. Corona Virus (COVID-19) "Infodemic" and Emerging Issues through a Data Lens: The Case of China. *International journal of environmental research and public health* 2020; 17.
41. Pan XB. Application of personal-oriented digital technology in preventing transmission of COVID-19, China. *Irish journal of medical science* 2020:1-2.
42. Close contact measurement instrument. Available at: <http://2019ncov.cetcloud.com/#/>. Accessed 18 March 2020.
43. Bragazzi NL, Dai H, Damiani G, Behzadifar M, Martini M, Wu J. How Big Data and Artificial Intelligence Can Help Better Manage the COVID-19 Pandemic. *International journal of environmental research and public health* 2020; 17.
44. Kraemer MUG, Yang CH, Gutierrez B, et al. The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science (New York, NY)* 2020; 368:493-7.
45. Wong ZSY, Zhou J, Zhang Q. Artificial Intelligence for infectious disease Big Data Analytics. *Infect Dis Health* 2019; 24:44-8.
46. Young SD, Rivers C, Lewis B. Methods of using real-time social media technologies for detection and remote monitoring of HIV outcomes. *Prev Med* 2014; 63:112-5.
47. McKendry RA, Rees G, Cox IJ, et al. Share mobile and social-media data to curb COVID-19. *Nature* 2020; 580:29.
48. Dolley S. Big Data's Role in Precision Public Health. *Front Public Health* 2018; 6:68.
49. Kostkova P, Brewer H, de Lusignan S, et al. Who Owns the Data? Open Data for Healthcare. *Front Public Health* 2016; 4:7.
50. Parker MJ, Fraser C, Abeler-Dörner L, Bonsall D. Ethics of instantaneous contact tracing using mobile phone apps in the control of the COVID-19 pandemic. *Journal of medical ethics* 2020.

