ISSN 2644-3236 (online)

# Artificial Intelligence in *Gastroenterology*

*Artif Intell Gastroenterol* 2023 June 8; 4(1): 1-27

Published by Baishideng Publishing Group Inc

AIG

Artificial Intelligence in Gastroenterology

Contents

Quarterly Volume 4 Number 1 June 8, 2023

## MINIREVIEWS

1 Big data and variceal rebleeding prediction in cirrhosis patients

*Yuan Q, Zhao WL, Qin B*

## ORIGINAL ARTICLE

### Retrospective Study

10 Risk factor profiles for gastric cancer prediction with respect to *Helicobacter pylori*: A study of a tertiary care hospital in Pakistan

*Aziz S, König S, Umer M, Akhter TS, Iqbal S, Ibrar M, Ur-Rehman T, Ahmad T, Hanafiah A, Zahra R, Rasheed F*

*Artificial Intelligence in Gastroenterology*

Quarterly Volume 4 Number 1 June 8, 2023

# Contents

## ABOUT COVER

Editorial Board Member of *Artificial Intelligence in Gastroenterology*, Haseeb Ahmad Khan, PhD, Full Professor, Department of Biochemistry, King Saud University, Riyadh 11451, Saudi Arabia. khan_haseeb@yahoo.com

## AIMS AND SCOPE

The primary aim of *Artificial Intelligence in Gastroenterology* (*AIG*, *Artif Intell Gastroenterol*) is to provide scholars and readers from various fields of artificial intelligence in gastroenterology with a platform to publish high-quality basic and clinical research articles and communicate their research findings online.

*AIG* mainly publishes articles reporting research results obtained in the field of artificial intelligence in gastroenterology and covering a wide range of topics, including artificial intelligence in gastrointestinal cancer, liver cancer, pancreatic cancer, hepatitis B, hepatitis C, nonalcoholic fatty liver disease, inflammatory bowel disease, irritable bowel syndrome, and *Helicobacter pylori* infection.

## INDEXING/ABSTRACTING

The *AIG* is now abstracted and indexed in Reference Citation Analysis, China Science and Technology Journal Database.

## RESPONSIBLE EDITORS FOR THIS ISSUE



**NAME OF JOURNAL**
*Artificial Intelligence in Gastroenterology*

**ISSN**
ISSN 2644-3236 (online)

**LAUNCH DATE**
July 28, 2020

**FREQUENCY**
Quarterly

**EDITORS-IN-CHIEF**
Rajvinder Singh, Ferruccio Bonino

**EDITORIAL BOARD MEMBERS**
https://www.wjgnet.com/2644-3236/editorialboard.htm

**PUBLICATION DATE**
June 8, 2023



**INSTRUCTIONS TO AUTHORS**
https://www.wjgnet.com/bpg/gerinfo/204

**GUIDELINES FOR ETHICS DOCUMENTS**
https://www.wjgnet.com/bpg/GerInfo/287

**GUIDELINES FOR NON-NATIVE SPEAKERS OF ENGLISH**
https://www.wjgnet.com/bpg/gerinfo/240

**PUBLICATION ETHICS**
https://www.wjgnet.com/bpg/GerInfo/288

**PUBLICATION MISCONDUCT**
https://www.wjgnet.com/bpg/gerinfo/208

**ARTICLE PROCESSING CHARGE**
https://www.wjgnet.com/bpg/gerinfo/242

**STEPS FOR SUBMITTING MANUSCRIPTS**
https://www.wjgnet.com/bpg/GerInfo/239

**ONLINE SUBMISSION**
https://www.f6publishing.com

Artificial Intelligence in Gastroenterology

Submit a Manuscript: https://www.f6publishing.com

*Artif Intell Gastroenterol* 2023 June 8; 4(1): 1-9

DOI: 10.35712/aig.v4.i1.1

ISSN 2644-3236 (online)

*MINIREVIEWS*

# Big data and variceal rebleeding prediction in cirrhosis patients

Quan Yuan, Wen-Long Zhao, Bo Qin

**Specialty type:** Gastroenterology and hepatology

**Provenance and peer review:** Unsolicited article; Externally peer reviewed.

**Peer-review model:** Single blind

**Peer-review report's scientific quality classification**

Grade A (Excellent): 0
Grade B (Very good): B
Grade C (Good): 0
Grade D (Fair): D
Grade E (Poor): 0

**P-Reviewer:** Byeon H, South Korea; Leowattana W, Thailand

**Received:** January 8, 2023
**Peer-review started:** January 8, 2023
**First decision:** January 21, 2023
**Revised:** February 3, 2023
**Accepted:** March 10, 2023
**Article in press:** March 10, 2023
**Published online:** June 8, 2023

**Quan Yuan,** Department of Gastroenterology, The First Affiliated Hospital of Chongqing Medical University, Chongqing 400042, China

**Wen-Long Zhao,** College of Medical Informatics, Chongqing Medical University, Chongqing 400016, China

**Wen-Long Zhao,** Medical Data Science Academy, Chongqing 400016, China

**Wen-Long Zhao,** Chongqing Engineering Research Centre for Clinical Big-data and Drug Evaluation, Chongqing 400016, China

**Bo Qin,** Department of Infectious Diseases, The First Affiliated Hospital of Chongqing Medical University, Chongqing 400042, China

**Corresponding author:** Bo Qin, MD, Professor, Department of Infectious Diseases, The First Affiliated Hospital of Chongqing Medical University, No. 1 Youyi Road, Yuzhong District, Chongqing 400042, China. qinbo@cqmu.edu.cn

## Abstract

Big data has convincing merits in developing risk stratification strategies for diseases. The 6 "V"s of big data, namely, volume, velocity, variety, veracity, value, and variability, have shown promise for real-world scenarios. Big data can be applied to analyze health data and advance research in preclinical biology, medicine, and especially disease initiation, development, and control. A study design comprises data selection, inclusion and exclusion criteria, standard confirmation and cohort establishment, follow-up strategy, and events of interest. The development and efficiency verification of a prognosis model consists of deciding the data source, taking previous models as references while selecting candidate predictors, assessing model performance, choosing appropriate statistical methods, and model optimization. The model should be able to inform disease development and outcomes, such as predicting variceal rebleeding in patients with cirrhosis. Our work has merits beyond those of other colleagues with respect to cirrhosis patient screening and data source regarding variceal bleeding.

**Key Words:** Big data; Disease onset; Prognosis; Modeling; Cirrhosis; Gastrointestinal rebleeding

**Core Tip:** Big data have been applied in many fields including finance, traffic control, logistics, healthcare, and environmental protection. Modeling is an efficient method for completing various tasks, and verification of its validity is vital for ensuring high-quality operation and yielding satisfactory results. Predictor screening guarantees the establishment of a practical, convenient, and favorable model for prognosis prediction. Utilizing a regression model trained with numerous data mined from big data acquired from real-world hospitals is helpful for informing disease or status onset and its prognosis such as in variceal rebleeding, which is one of the leading causes of death in cirrhosis patients.

**Citation:** Yuan Q, Zhao WL, Qin B. Big data and variceal rebleeding prediction in cirrhosis patients. *Artif Intell Gastroenterol* 2023; 4(1): 1-9
**URL:** https://www.wjgnet.com/2644-3236/full/v4/i1/1.htm
**DOI:** https://dx.doi.org/10.35712/aig.v4.i1.1

## INTRODUCTION

Many risk stratification strategies for diseases mainly depend on single-/medium-sized cohort studies or their meta-analysis[1,2], with lead-time bias taken into consideration[3,4]. This type of study method is, by design, well scheduled and well phenotyped but selective for the population sampled, which may not reflect the real-world, pan-subject profile. Real-world patients may have comorbidities, be taking concomitant medications, may be excluded from short-term follow-up, or have poor patient compliance. Direct data acquisition from basic healthcare institutions and cohorts is more representative than limited sampling.

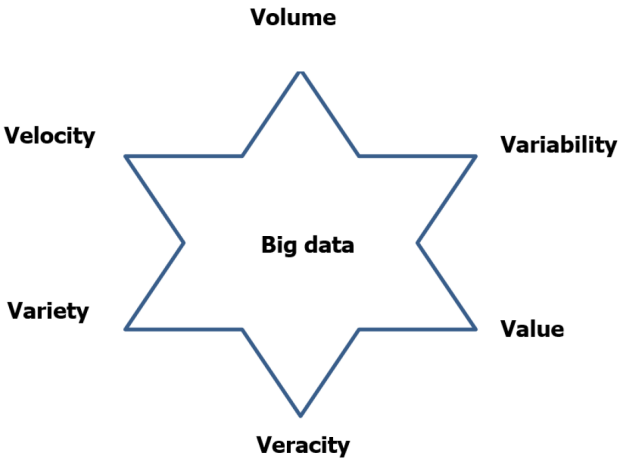
## HISTORY OF BIG DATA

Although the use of piles of data in the medical field has a relatively long history[5-7], the term "big data" appeared only in the 1990s and quickly became popular[8-10]. "Big" is a relative term, especially when it relates to data. Big data usually refers to datasets that exceed the capabilities of commonly used software tools to store, manage, and process that amount of data within a suitable period of time[11]. The term is described by 315 characteristics[12] and fundamentally by the 6 "V"s: volume, velocity, variety, veracity, value, and variability[13-17] (Figure 1).

During the recent decade, methods for collecting, storing, and managing big data have evolved[18-20]. We are now entering an era of monitoring health changes using clinical indicators, such as vital signs, serum sugar, lipids, sweating, and bladder fullness, with wearable devices[11]. These changes can reflect physiological change. Constant variation and altered levels may result in different pathological states. Here, we review the applications of big data in predicting disease onset and prognosis, especially variceal rebleeding prediction in cirrhosis patients.

## APPLICATIONS OF BIG DATA

Applications of big data include its use as a tool to monitor the onset of conditions and diseases. Big data have been used for this purpose in relation to hypertension[21], pediatric oncology[22], oral care [23], general practice[24], rheumatic diseases[25], renal diseases[26], mechanical ventilation management in the intensive care unit[27], and cirrhosis and hepatocellular carcinoma morbidity in the nonalcoholic fatty liver disease/nonalcoholic steatohepatitis population[28]. Situations such as the commencement, development, and control of diseases can be studied and visualized using big data techniques, which is a promising and beneficial approach. With the help of big data, the creation of large, collaborative data can lay a more solid foundation for robust data sharing and scientific discovery in predicting the onset of pediatric oncology. Registry-based research, however, is one of the conventional research methods regarding pediatric cancers. In these studies, a multisite registry for the study of pediatric patients was utilized, including fields of descriptive epidemiology, survivors, genomics, new registry description, data harmonization, palliative and supportive care, radiology, consensus guidelines, hereditary pediatric cancer, electronic health records, and prospective clinical trials. Limitations of registry-based research include the latest publication time range only, a restricted single publication database, and a limited amount of research and registries only if they have yielded publicly-published peer-reviewed papers[22]. With this study strategy, data cannot be automatically mined, cleaned, and integrated to perfect the already existing study. When it comes to new subjects, we need to redo the statistical analysis, while modeling and machine study in the big data scenario can perform the




**DOI**: 10.35712/aig.v4.i1.1 

**Figure 1 Six "V"s of big data.**

whole analysis process.

Healthcare data in some regions are complete and accessible for analysis. Real-world data from primary healthcare facilities in communities in European countries are a good resource, as the primary healthcare service is state-covered and there are few or no co-payments. Therefore, healthcare information and data are collected and stored by state-run big data centers. Most residents are registered at birth and have their complete healthcare information in electronic form, which can be accessed by regional practitioners and analyzed for real-world application scenarios[29]. However, numerous parameters, especially administrative data, mined from patients' inpatient and outpatient Hospital Information System/Electronic Medical Record system *via* various algorithms are at risk of information and privacy leaking. Therefore, preliminary selection of data, especially low-dimensional administrative data, is preferable to decrease information leakage and privacy invasion.

Big data boosts the depth and breadth of research in fundamental biology and clinical medicine. There is already impressive progress due to this, including in exome sequencing[30], genomics, and proteomics. Taking the coronavirus disease 2019 pandemic as an example, primary research, clinical practices regarding treatment, and even trends in media campaigns of whether or not executing lockdown and a positive policy of nucleic acid testing can be swiftly analyzed with big data tools to assist epidemic control[31].

## STUDY DESIGN

Study design comprises data source selection, inclusion and exclusion criteria, standard confirmation and cohort establishment, follow-up strategy, and events of interest. A multicountry European real-world study acquired patient data within a set research period mined from central transcription, laboratories, pharmacy offices, medical insurance departments, administrative departments, and other departmental databases *via* an electronic health record data repository along with molecular typing from molecular biology laboratories for preventing outbreaks of hospital infections[32]. Chart presentations can be used to analyze and interpret descriptive data. The Fib-4 score (age, aspartate aminotransferase, alanine aminotransferase, and platelets), which is composed of entirely non-invasive parameters, has been used to detect early liver fibrosis[28].

## MODEL DEVELOPMENT AND EFFICACY VERIFICATION

With respect to development and efficiency verification of disease onset and prognosis models, researchers have performed extensive work. Model development is the process of collecting vital parameters (risk factors) of consequence and weighted with varied weight coefficients to form a weighted function. This requires the identification of predominant predictors from a large amount of preselected candidate predictors, assigning proper weights to each predictor to obtain a combined risk score, and assessing the model's predictive performance with statistical methods such as a calibration plot. The latter includes calibration, discrimination, and (re)classification properties, assessing its potential for generalization using internal validation techniques and if necessary optimizing the model to avoid overfitting. Data sources should preferably be prospective cohort(s) with a randomized controlled trial design or real-world medical record data. Preferred outcome choices are those that are

related to patients or individuals such as remission time and follow-up period. Methods for outcome verification should be included, and the blind method is preferred.

Regarding the selection of candidate predictors, a surplus should be defined and analyzed before finally including a subset in the final model. Incorporation bias should be avoided by blinding. Data quality control, missing data processing, continuous predictor modeling, final model development, relative weight assignment for each predictor, and internal validation are essential in the process of creating a final prediction model[33].

Choosing appropriate statistical methods during model establishment is vital to guarantee reliability and validity. Regression analysis, including univariate and multivariate regression, is the most commonly used statistical method, especially Cox regression[34] and LASSO[35]. The hazard ratio is used to differentiate cohorts across different conditions and coefficients. Featured with net benefit and threshold probability for more convenient yet trusty clinical decision making, decision curve analysis has been used to evaluate whether or not to use a certain prediction model[36]. In this approach, the theoretical relationship between the threshold probabilities of a disease (that a disease will take place) and the relative frequency of false positives and false negatives are examined to ensure the validity of a prediction model.

The benefits of applying decision curve analysis can be quantified as whether a model can be easily and effectively applied in clinical situations. Its ability to help compare several different models regarding one issue is another advantage[37]. The parameter indicating risk threshold "T value" has been used to study treatment decisions in risk models. The harm-to-benefit ratio is related to the T value, which is in line with the former. Balancing all benefits and harms in different scenarios is key to determining which T value is reasonable[38]. The net benefit (NB) value, which is a combined "net" effect of the true positives and false positives, was introduced to evaluate the potential clinical application of an estimating tool or a risk-predicting model. Setting the decisive threshold range in modeling is important, which is the boundary to determine whether a patient is judged as positive for a disease or not[39]. However, NB does not directly make up the harms and costs in acquiring the predictors for the chosen model. The focus of NB is to derive the best tradeoff between sufficient indicators and convenience in clinical application[40].

Model optimization should be conducted in order to reduce the number of predictors and avoid an unmanageable dataset or workload. AMSGrad ("far from the minimum"), a putative optimal method for optimizing models, is commonly used for low-cost cause. By switching to the direct linear method near the end of the optimization, AMSGrad can do its magic as it has long convergence tails[41]. As for multiobjective racing algorithms with fixed confidence, SPRINT-Race is the first algorithm developed and uses a nonparametric, ternary-decision, dual-sequential probability ratio test to infer a pairwise dominance or nondominance relationship. In order to minimize the computational effort, the probability of mistakenly erasing any Pareto-optimal models or returning any clearly dominating models is restricted, which can achieve a pre-estimated confidence level to ensure the quality of the models generated[42], by sequentially applying a Holm's step-down family-wise error rate control method. The quantification of model-to-data correspondence is pivotal to measure a model's performance and future application for the problem at hand. The *Drosophila melanogaster* gap gene system model demonstrated the importance of error quantification, and it is applicable to a wide array of developmental modeling studies[43]. The support vector machine, GLM-Net, generalized linear model, partial least squares, neural network, k-nearest neighbors, random forest, and boosted tree are useful tools for establishing the model to predict prognosis in patients with breast cancer[44]. Comparing their differences in performance and necessary model optimization can lead to better and more efficient application in practice.

## PREDICTOR SCREENING FOR PROGNOSIS

Researchers have proposed methods for predictor screening with regard to disease prognosis, such as the Model for End-stage Live Disease (MELD) for cirrhosis-related mortality prediction and the APACHE model for critically ill patients. The clinical data of cirrhosis patients who had early admission, including clinical and socioeconomic factors, were mined from electronic medical records and classified for risk stratification in order to predict readmission within 30 d[45]. The European Organization for Research and Treatment of Cancer (EORTC) risk tables[46], which include six clinical and pathological factors (number of tumors, tumor size, prior recurrence rate, T category, carcinoma *in situ*, and grade), were recommended by the European Association of Urology and used to separately predict the short-term and long-term risks of progression and recurrence in an individual patient with a non-muscular invasive bladder tumor. It divided patients into four groups with individual recurrent and progression scores. However, as EORTC risk tables overestimated recurrence in all risk groups and progression in the high-risk group, the Club Urológico Español de Tratamiento Oncológico scoring model[47] was developed. The well-known new EORTC model[48], or European Association of Urology risk groups, was popular in recurrence and progression prediction, in which tumor diameter and extent were key predictors for progression prediction in multistate analyses. The health belief model has been

used for risk factors identifying aged Jordanian adults for prostate cancer screening[49]. Development and validation of a prediction model, including internal and external, temporal and geographical, domain validation, and their revision, are all crucial to identify predictors of prognosis[50].

## RISK INDICATORS OF VARICEAL REBLEEDING IN CIRRHOSIS

Studies have reported several prediction models that predict variceal rebleeding in patients with cirrhosis. Risk indicators are components of prediction models. Invariably, studies in spotting possible risk indicators of variceal rebleeding among cirrhosis patients require a long study period. Child-Pugh score and hepatic-venous pressure gradient are the most significant prognostic factors in stratifying the probability of variceal rebleeding[51]. Antiviral treatment significantly reduced rebleeding in patients with hepatitis B virus (HBV)-related cirrhosis. In-time prophylactic endoscopic treatment of upper gastrointestinal varices after first-time bleeding, including endoscopic varix ligation (EVL) and gastric fundus varix gluing, is important in postponing variceal rebleeding[52]. Tachycardia, high creatinine level, and low albumin level are independent factors associated with rebleeding, suggesting a potential predictive role. The transverse of these variables into predictive scores may provide improved prognosis for patients with variceal bleeding[53]. Pre-emptive transjugular intrahepatic portosystemic shunt was independently related to a lower rebleeding rate[54]. Albumin transfusion in patients with low albumin levels was positively associated with a decreased rebleeding rate[55]. Five studies showed a lower rebleeding rate after EVL or drug therapy (non-selective β-blockers ± isosorbide mononitrate), and four trials found decreased variceal rebleeding with combined therapy (EVL+ non-selective β-blockers+ isosorbide mononitrate)[56].

However, some indicators have a negative function in preventing rebleeding. A multicenter, double-blind, parallel study of 158 patients indicated that taking simvastatin besides standard prophylaxis (rest, fluid restriction, preventing infection, regular endoscopic examination, anti-HBV therapy, non-selective β-blocker, *etc*.) did not decrease the rebleeding rate[57]. The rate of variceal rebleeding was not reduced after anticoagulation according to a single-center, prospective cohort study[58]. Worsened liver function or insensitive hemodynamic response to non-selective β-blockers indicated an elevated rebleeding rate [51]. A Chinese study of 3289 hospitalized patients who underwent EVL indicated that male sex, Child-Pugh score > 7.2, and volume of blood vomited before EVL were independent risk indicators of early rebleeding, while albumin concentration > 31.5 g/L was a protective indicator[59]. Bacterial infection in patients with variceal bleeding was strongly positively related to early rebleeding[60]. Acute-on-chronic liver failure is an independent risk factor of variceal rebleeding[54]. The presence of ascites or hepatic encephalopathy, MELD score > 12, or hepatic-venous pressure gradient > 20 mmHg indicated an elevated early (less than 6 wk) rebleeding rate[61].

The above indicators were then filtered and optimized by statistical methods, such as Cox regression or LASSO, and systemically integrated into a function with the help of programming or statistical software such as R, Python, SPSS, or SAS. This function was actually a preliminary prediction model.

## SIGNIFICANCE OF PREDICTION MODELS

Models predicting disease onset and prognosis play an essential and sometimes surprising role as convenient assistants in planning prophylactic, therapeutic, and follow-up strategies. Traditionally, medical data such as medical history, results of physical examination, laboratory tests, imaging and endoscopic information, *etc*. were integrated by doctors' clinical comprehension or into patients' timelines drafted on a paper to identify how disease progressed and predicted the possible prognosis according to the trend in medical indicators. Prediction models free doctors from numerous medical data of patients with different diseases, complications, physical, psychological, and socioeconomic situations. All they need to do is to type prescribed parameters into the model and click! The results of the onset and prognosis of a given disease are then provided.

Prediction models are currently extensively applied in the medical field to inform individuals and healthcare providers on the risks of developing a particular disease, its outcome, and to guide doctors to make better decisions in mitigating these risks. A recent Chinese study indicated that the MELD score and MELD-Na score, including the R score, were useful in predicting variceal rebleeding[62]. Another study indicated that the MELD-Na score model, which indicates liver function, was more efficient than the MELD model and Child-Pugh score model in predicting rebleeding among cirrhosis patients who underwent EVL.

## SAFETY AND PRIVACY CONCERNS

Last but not least, it is worth noting that models using low-dimensional administrative data outper-

formed in big data analysis with respect to decreasing information safety and privacy invasion. According to several studies, the models did not improve when high-resolution, privacy-invasive behavioral data were included[63]. De-ID software (De-ID Data) has been used to assign a study identification number to every enrolled patient. Therefore, criteria, included in the informed consent established by the research review board, for exemption from enrollment were met[32]. The *Drosophila melanogaster* gap gene system gives a good example of demonstrating the significance of error quantification, in which model parameters were optimized against *in situ* immunofluorescence intensities. It can be applied to other studies in various fields with regard to model development.

## DISCUSSION

Gastrointestinal (GI) rebleeding is a leading cause of mortality in patients with cirrhosis, as massive GI bleeding can induce hemorrhagic shock, disseminated intravascular coagulation, and opportunistic infections, especially pulmonary infection and spontaneous bacterial peritonitis. Thus, reducing or postponing GI rebleeding is significant. A handy tool for clinicians that can be operated on smart phones or other mobile intelligent devices within seconds to evaluate the GI rebleeding rate is interesting and useful for risk grading. Just type in several common laboratory test indicators, click on "go," and the rebleeding rate and prognosis of a specific patient are provided.

Our work has merits beyond those of other colleagues. According to our literature retrieval on PubMed, there are no other studies on the prediction and prognosis analysis of GI rebleeding except for one article published last year indicating that the degree of liver stiffness is consistent with GI rebleeding rate in cirrhosis patients[64]. However, the above mentioned exclusive study has limitations. First, it was a prospective cohort study with only 289 patients enrolled in the final analysis, although PASS 15 was applied to calculate the statistically minimum sample size. In our ongoing study applying big data platform to evaluation the rebleeding rate of cirrhosis patients, we obtained real-world data from a big data platform collecting many more indicators from six hospitals, which were automatically collected. Second, our study included patients with esophageal and gastric fundus varices rebleeding, which were the most common varices presented in cirrhosis patients, and the other study only included esophageal varix rebleeding. Finally, the previous study only included patients with HBV-related decompensated cirrhosis, while our data were collected from cirrhosis patients with alcohol-related cirrhosis, autoimmune-related cirrhosis, primary biliary cirrhosis, and lipogenic cirrhosis in addition to HBV-related cirrhosis. Following parameter filtering and modeling, our study used a visual nomogram to demonstrate correlations among risk indicators, occurrence, and prognosis of GI rebleeding, which provides clinicians with a more explicit demonstration of all indicators and their effects on one page to easily and rapidly evaluate a patient to establish a strategy for further management and follow-up.

## CONCLUSION

Modeling is popular using regression analysis and has vast applications in predicting disease occurrence and prognosis. However, modeling and its validation are not the ultimate objective in terms of healthcare provider's clinical participation and patients' health outcomes. They need to be applied and provide convenience for clinical practice. Studies on the application and optimization of these models should be designed and conducted, focusing on the utilization of existing and updated models and their impact on behavior and (self-) management of physicians, healthcare providers, and general individuals[65,66], especially in patients with decompensated cirrhosis at high risk of variceal rebleeding and mortality. For diagnostic and prognostic modeling with higher consistency and efficiency in predicting, treating, and following up decompensated cirrhosis, more comprehensive data and a clearer display mode are needed.

## FOOTNOTES

**Author contributions:** Yuan Q selected the topic and performed the majority of conception, writing, and revision of the manuscript; Zhao WL provided think tank, platform with regard to big data, site for academic discussion, and revision suggestions for the manuscript; Qin B provided administrative help and was the instigator and coordinator of the study; All authors have read and approved the final manuscript.

**Conflict-of-interest statement:** All the authors declare that they have no conflicts of interest.

**Country/Territory of origin:** China

**ORCID number:** Quan Yuan 0000-0001-7761-4113; Bo Qin 0000-0002-7802-2854.

**S-Editor:** Liu JH
**L-Editor:** Filipodia
**P-Editor:** Liu JH

## REFERENCES

1 **Koehler EM**, Schouten JN, Hansen BE, van Rooij FJ, Hofman A, Stricker BH, Janssen HL. Prevalence and risk factors of non-alcoholic fatty liver disease in the elderly: results from the Rotterdam study. *J Hepatol* 2012; **57**: 1305-1311 [PMID: 22871499 DOI: 10.1016/j.jhep.2012.07.028]
2 **Younossi ZM**, Koenig AB, Abdelatif D, Fazel Y, Henry L, Wymer M. Global epidemiology of nonalcoholic fatty liver disease-Meta-analytic assessment of prevalence, incidence, and outcomes. *Hepatology* 2016; **64**: 73-84 [PMID: 26707365 DOI: 10.1002/hep.28431]
3 **Facciorusso A**, Ferrusquía J, Muscatiello N. Lead time bias in estimating survival outcomes. *Gut* 2016; **65**: 538-539 [PMID: 26163490 DOI: 10.1136/gutjnl-2015-310199]
4 **Jansen RJ**, Alexander BH, Anderson KE, Church TR. Quantifying lead-time bias in risk factor studies of cancer through simulation. *Ann Epidemiol* 2013; **23**: 735-741 [PMID: 23988688 DOI: 10.1016/j.annepidem.2013.07.021]
5 **Graunt J**. Mathematical Demography. Berlin, Heidelberg: Springer, 1975: 11-20
6 **Dumbill E**. A Revolution That Will Transform How We Live, Work, and Think: An Interview with the Authors of Big Data. *Big Data* 2013; **1**: 73-77 [PMID: 27442060 DOI: 10.1089/big.2013.0016]
7 **Rothman KJ**. Lessons from John Graunt. *Lancet* 1996; **347**: 37-39 [PMID: 8531550 DOI: 10.1016/S0140-6736(96)91376-8]
8 **de Mauro A**, Greco M, Grimaldi M. A formal definition of big data based on its essential features. Lib Rev. 2016 Apr 4; 65: 122-135 [DOI: 10.1108/LR-06-2015-0061]
9 **John R**. Mashey. Big Data and the Next Wave of Infra Stress. USENIX: The Advanced Computing Systems Association. 1998. Available from: https://www.usenix.org/Legacy/event/usenix99/invited_talks/mashey.pdf
10 **Lohr S**. The Origins of 'Big Data': An Etymological Detective Story. The New York Times. B4. 2013. Available from: https://www.mendeley.com/catalogue/45aafddc-f02a-37bd-b201-c5edbcd31e82/
11 **Mirchev M**, Mircheva I, Kerekovska A. The Academic Viewpoint on Patient Data Ownership in the Context of Big Data: Scoping Review. *J Med Internet Res* 2020; **22**: e22214 [PMID: 32808934 DOI: 10.2196/22214]
12 **Kapil G**, Agrawal A, Khan RA. A Study of Big Data Characteristics. International Conference on Communication and Electronics Systems; ICCES'16; 2016 October 21-22, Coimbatore, India [DOI: 10.1109/CESYS.2016.7889917]
13 **Nobanee H**. A Bibliometric Review of Big Data in Finance. *Big Data* 2021; **9**: 73-78 [PMID: 33861644 DOI: 10.1089/big.2021.29044.edi]
14 **Beyer MA**, Laney D. The Importance of 'Big Data': A Definition. Gartner Inc. June 21, 2021. Available from: https://www.gartner.com/en/documents/2057415/the-importance-of-big-data-a-definition
15 **Tseng IL**. Big data: related technologies, challenges and future prospects. Computing reviews, 2015, 56: 476-477. Available from: https://www.zhangqiaokeyan.com/academic-journal-foreign_other_thesis/020411385974.html
16 **Dobre C**, Xhafa F. Intelligent services for big data science. *Future Gener Comput Syst* 2014; **37**: 267-281 [DOI: 10.1016/j.future.2013.07.014]
17 **Owais SS**, Hussein NS. Extract five categories CPIVW from the 9V's characteristics of the big data. *Int J Adv Comput Sci Appl* 2016; **7**: 254-258 [DOI: 10.14569/IJACSA.2016.070337]
18 **O'Driscoll A**, Daugelaite J, Sleator RD. 'Big data', Hadoop and cloud computing in genomics. *J Biomed Inform* 2013; **46**: 774-781 [PMID: 23872175 DOI: 10.1016/j.jbi.2013.07.001]
19 **Costa FF**. Big data in biomedicine. *Drug Discov Today* 2014; **19**: 433-440 [PMID: 24183925 DOI: 10.1016/j.drudis.2013.10.012]
20 **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 [PMID: 26843812 DOI: 10.4137/BII.S31559]
21 **Okada M**. Big data and real-world data-based medicine in the management of hypertension. *Hypertens Res* 2021; **44**: 147-153 [PMID: 33250517 DOI: 10.1038/s41440-020-00580-3]
22 **Major A**, Cox SM, Volchenboum SL. Using big data in pediatric oncology: Current applications and future directions. *Semin Oncol* 2020; **47**: 56-64 [PMID: 32229032 DOI: 10.1053/j.seminoncol.2020.02.006]
23 **Finkelstein J**, Zhang F, Levitin SA, Cappelli D. Using big data to promote precision oral health in the context of a learning healthcare system. *J Public Health Dent* 2020; **80** Suppl 1: S43-S58 [PMID: 31905246 DOI: 10.1111/jphd.12354]
24 **Waschkau A**, Wilfling D, Steinhäuser J. Are big data analytics helpful in caring for multimorbid patients in general practice? *BMC Fam Pract* 2019; **20**: 37 [PMID: 30813904 DOI: 10.1186/s12875-019-0928-5]
25 **Manrique de Lara A**, Peláez-Ballestas I. Big data and data processing in rheumatology: bioethical perspectives. *Clin Rheumatol* 2020; **39**: 1007-1014 [PMID: 32062767 DOI: 10.1007/s10067-020-04969-w]
26 **Yang C**, Kong G, Wang L, Zhang L, Zhao MH. Big data in nephrology: Are we ready for the change? *Nephrology (Carlton)* 2019; **24**: 1097-1102 [PMID: 31314170 DOI: 10.1111/nep.13636]

27 **Smallwood CD**. Monitoring Big Data During Mechanical Ventilation in the ICU. *Respir Care* 2020; **65**: 894-910 [PMID: 32457178 DOI: 10.4187/respcare.07500]

28 **Alexander M**, Loomis AK, van der Lei J, Duarte-Salles T, Prieto-Alhambra D, Ansell D, Pasqua A, Lapi F, Rijnbeek P, Mosseveld M, Waterworth DM, Kendrick S, Sattar N, Alazawi W. Risks and clinical predictors of cirrhosis and hepatocellular carcinoma diagnoses in adults with diagnosed NAFLD: real-world study of 18 million patients in four European cohorts. *BMC Med* 2019; **17**: 95 [PMID: 31104631 DOI: 10.1186/s12916-019-1321-x]

29 **Kringos D**, Boerma W, Bourgueil Y, Cartier T, Dedeu T, Hasvold T, Hutchinson A, Lember M, Oleszczyk M, Rotar Pavlic D, Svab I, Tedeschi P, Wilm S, Wilson A, Windak A, Van der Zee J, Groenewegen P. The strength of primary care in Europe: an international comparative study. *Br J Gen Pract* 2013; **63**: e742-e750 [PMID: 24267857 DOI: 10.3399/bjgp13X674422]

30 **Suwinski P**, Ong C, Ling MHT, Poh YM, Khan AM, Ong HS. Advancing Personalized Medicine Through the Application of Whole Exome Sequencing and Big Data Analytics. *Front Genet* 2019; **10**: 49 [PMID: 30809243 DOI: 10.3389/fgene.2019.00049]

31 **Jung JH**, Shin JI. Big Data Analysis of Media Reports Related to COVID-19. *Int J Environ Res Public Health* 2020; **17** [PMID: 32781727 DOI: 10.3390/ijerph17165688]

32 **Sundermann AJ**, Miller JK, Marsh JW, Saul MI, Shutt KA, Pacey M, Mustapha MM, Ayres A, Pasculle AW, Chen J, Snyder GM, Dubrawski AW, Harrison LH. Automated data mining of the electronic health record for investigation of healthcare-associated outbreaks. *Infect Control Hosp Epidemiol* 2019; **40**: 314-319 [PMID: 30773168 DOI: 10.1017/ice.2018.343]

33 **Moons KG**, Kengne AP, Woodward M, Royston P, Vergouwe Y, Altman DG, Grobbee DE. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. *Heart* 2012; **98**: 683-690 [PMID: 22397945 DOI: 10.1136/heartjnl-2011-301246]

34 **In J**, Lee DK. Survival analysis: part II - applied clinical data analysis. *Korean J Anesthesiol* 2019; **72**: 441-457 [PMID: 31096731 DOI: 10.4097/kja.19183]

35 **Tibshirani R**. The lasso method for variable selection in the Cox model. *Stat Med* 1997; **16**: 385-395 [PMID: 9044528 DOI: 10.1002/(SICI)1097-0258(19970228)16:4<385::AID-SIM380>3.0.CO;2-3]

36 **Van Calster B**, Wynants L, Verbeek JFM, Verbakel JY, Christodoulou E, Vickers AJ, Roobol MJ, Steyerberg EW. Reporting and Interpreting Decision Curve Analysis: A Guide for Investigators. *Eur Urol* 2018; **74**: 796-804 [PMID: 30241973 DOI: 10.1016/j.eururo.2018.08.038]

37 **Vickers AJ**, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making* 2006; **26**: 565-574 [PMID: 17099194 DOI: 10.1177/0272989X06295361]

38 **Pauker SG**, Kassirer JP. Therapeutic decision making: a cost-benefit analysis. *N Engl J Med* 1975; **293**: 229-234 [PMID: 1143303 DOI: 10.1056/NEJM197507312930505]

39 **Vickers AJ**, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. *BMJ* 2016; **352**: i6 [PMID: 26810254 DOI: 10.1136/bmj.i6]

40 **Baker SG**, Kramer BS. Evaluating Prognostic Markers Using Relative Utility Curves and Test Tradeoffs. *J Clin Oncol* 2015; **33**: 2578-2580 [PMID: 26124476 DOI: 10.1200/JCO.2014.58.0092]

41 **Sabzevari I**, Mahajan A, Sharma S. An accelerated linear method for optimizing non-linear wavefunctions in variational Monte Carlo. *J Chem Phys* 2020; **152**: 024111 [PMID: 31941334 DOI: 10.1063/1.5125803]

42 **Zhang T**, Georgiopoulos M, Anagnostopoulos GC. Pareto-Optimal Model Selection *via* SPRINT-Race. *IEEE Trans Cybern* 2018; **48**: 596-610 [PMID: 28166512 DOI: 10.1109/TCYB.2017.2647821]

43 **Hengenius JB**, Gribskov M, Rundell AE, Umulis DM. Making models match measurements: model optimization for morphogen patterning networks. *Semin Cell Dev Biol* 2014; **35**: 109-123 [PMID: 25016297 DOI: 10.1016/j.semcdb.2014.06.017]

44 **Boughorbel S**, Al-Ali R, Elkum N. Model Comparison for Breast Cancer Prognosis Based on Clinical Data. *PLoS One* 2016; **11**: e0146413 [PMID: 26771838 DOI: 10.1371/journal.pone.0146413]

45 **Singal AG**, Rahimi RS, Clark C, Ma Y, Cuthbert JA, Rockey DC, Amarasingham R. An automated model using electronic medical record data identifies patients with cirrhosis at high risk for readmission. *Clin Gastroenterol Hepatol* 2013; **11**: 1335-1341.e1 [PMID: 23591286 DOI: 10.1016/j.cgh.2013.03.022]

46 **Sylvester RJ**, van der Meijden AP, Oosterlinck W, Witjes JA, Bouffioux C, Denis L, Newling DW, Kurth K. Predicting recurrence and progression in individual patients with stage Ta T1 bladder cancer using EORTC risk tables: a combined analysis of 2596 patients from seven EORTC trials. *Eur Urol* 2006; **49**: 466-5; discussion 475 [PMID: 16442208 DOI: 10.1016/j.eururo.2005.12.031]

47 **Fernandez-Gomez J**, Madero R, Solsona E, Unda M, Martinez-Piñeiro L, Gonzalez M, Portillo J, Ojea A, Pertusa C, Rodriguez-Molina J, Camacho JE, Rabadan M, Astobieta A, Montesinos M, Isorna S, Muntañola P, Gimeno A, Blas M, Martinez-Piñeiro JA. Predicting nonmuscle invasive bladder cancer recurrence and progression in patients treated with bacillus Calmette-Guerin: the CUETO scoring model. *J Urol* 2009; **182**: 2195-2203 [PMID: 19758621 DOI: 10.1016/j.juro.2009.07.016]

48 **Cambier S**, Sylvester RJ, Collette L, Gontero P, Brausi MA, van Andel G, Kirkels WJ, Silva FC, Oosterlinck W, Prescott S, Kirkali Z, Powell PH, de Reijke TM, Turkeri L, Collette S, Oddens J. EORTC Nomograms and Risk Groups for Predicting Recurrence, Progression, and Disease-specific and Overall Survival in Non-Muscle-invasive Stage Ta-T1 Urothelial Bladder Cancer Patients Treated with 1-3 Years of Maintenance Bacillus Calmette-Guérin. *Eur Urol* 2016; **69**: 60-69 [PMID: 26210894 DOI: 10.1016/j.eururo.2015.06.045]

49 **Abuadas MH**, Petro-Nustas W, Albikawi ZF. Predictors of Participation in Prostate Cancer Screening among Older Men in Jordan. *Asian Pac J Cancer Prev* 2015; **16**: 5377-5383 [PMID: 26225681 DOI: 10.7314/apjcp.2015.16.13.5377]

50 **Moons KG**, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG, Woodward M. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart* 2012; **98**: 691-698 [PMID: 22397946 DOI: 10.1136/heartjnl-2011-301247]

51 **Magaz M**, Baiges A, Hernández-Gea V. Precision medicine in variceal bleeding: Are we there yet? *J Hepatol* 2020; **72**:

774-784 [PMID: 31981725 DOI: 10.1016/j.jhep.2020.01.008]

52 **He L**, Ye X, Ma J, Li P, Jiang Y, Hu J, Yang J, Zhou Y, Liang X, Lin Y, Wei H. Antiviral therapy reduces rebleeding rate in patients with hepatitis B-related cirrhosis with acute variceal bleeding after endotherapy. *BMC Gastroenterol* 2019; **19**: 101 [PMID: 31226942 DOI: 10.1186/s12876-019-1020-2]

53 **Jiménez Rosales R**, Martínez-Cara JG, Vadillo-Calles F, Ortega-Suazo EJ, Abellán-Alfocea P, Redondo-Cerezo E. Analysis of rebleeding in cases of an upper gastrointestinal bleed in a single center series. *Rev Esp Enferm Dig* 2019; **111**: 189-192 [PMID: 30569727 DOI: 10.17235/reed.2018.5702/2018]

54 **Trebicka J**, Gu W, Ibáñez-Samaniego L, Hernández-Gea V, Pitarch C, Garcia E, Procopet B, Giráldez Á, Amitrano L, Villanueva C, Thabut D, Silva-Junior G, Martinez J, Genescà J, Bureau C, Llop E, Laleman W, Palazon JM, Castellote J, Rodrigues S, Gluud L, Ferreira CN, Barcelo R, Cañete N, Rodríguez M, Ferlitsch A, Mundi JL, Gronbaek H, Hernández-Guerra M, Sassatelli R, Dell'Era A, Senzolo M, Abraldes JG, Romero-Gómez M, Zipprich A, Casas M, Masnou H, Primignani M, Weiss E, Catalina MV, Erasmus HP, Uschner FE, Schulz M, Brol MJ, Praktiknjo M, Chang J, Krag A, Nevens F, Calleja JL, Robic MA, Conejo I, Albillos A, Rudler M, Alvarado E, Guardascione MA, Tantau M, Bosch J, Torres F, Pavesi M, Garcia-Pagán JC, Jansen C, Bañares R; International Variceal Bleeding Observational Study Group and Baveno Cooperation. Rebleeding and mortality risk are increased by ACLF but reduced by pre-emptive TIPS. *J Hepatol* 2020; **73**: 1082-1091 [PMID: 32339602 DOI: 10.1016/j.jhep.2020.04.024]

55 **Wang Z**, Xie YW, Lu Q, Yan HL, Liu XB, Long Y, Zhang X, Yang JL. The impact of albumin infusion on the risk of rebleeding and in-hospital mortality in cirrhotic patients admitted for acute gastrointestinal bleeding: a retrospective study of a single institute. *BMC Gastroenterol* 2020; **20**: 198 [PMID: 32576140 DOI: 10.1186/s12876-020-01337-5]

56 **Puente A**, Hernández-Gea V, Graupera I, Roque M, Colomo A, Poca M, Aracil C, Gich I, Guarner C, Villanueva C. Drugs plus ligation to prevent rebleeding in cirrhosis: an updated systematic review. *Liver Int* 2014; **34**: 823-833 [PMID: 24373180 DOI: 10.1111/liv.12452]

57 **Abraldes JG**, Villanueva C, Aracil C, Turnes J, Hernandez-Guerra M, Genesca J, Rodriguez M, Castellote J, García-Pagán JC, Torres F, Calleja JL, Albillos A, Bosch J; BLEPS Study Group. Addition of Simvastatin to Standard Therapy for the Prevention of Variceal Rebleeding Does Not Reduce Rebleeding but Increases Survival in Patients With Cirrhosis. *Gastroenterology* 2016; **150**: 1160-1170.e3 [PMID: 26774179 DOI: 10.1053/j.gastro.2016.01.004]

58 **Amitrano L**, Guardascione MA, Scaglione M, Menchise A, Martino R, Manguso F, Lanza AG, Lampasi F. Splanchnic vein thrombosis and variceal rebleeding in patients with cirrhosis. *Eur J Gastroenterol Hepatol* 2012; **24**: 1381-1385 [PMID: 23114742 DOI: 10.1097/MEG.0b013e328357d5d4]

59 **Zhou JN**, Wei Z, Sun ZQ. [Risk factors for early rebleeding after esophageal variceal ligation in patients with liver cirrhosis]. *Zhonghua Gan Zang Bing Za Zhi* 2016; **24**: 486-492 [PMID: 27784425 DOI: 10.3760/cma.j.issn.1007-3418.2016.07.002]

60 **Boursier J**, Asfar P, Joly-Guillou ML, Calès P. [Infection and variceal bleeding in cirrhosis]. *Gastroenterol Clin Biol* 2007; **31**: 27-38 [PMID: 17273129 DOI: 10.1016/s0399-8320(07)89324-9]

61 **Ardevol A**, Alvarado-Tapias E, Garcia-Guix M, Brujats A, Gonzalez L, Hernández-Gea V, Aracil C, Pavel O, Cuyas B, Graupera I, Colomo A, Poca M, Torras X, Concepción M, Villanueva C. Early rebleeding increases mortality of variecal bleeders on secondary prophylaxis with β-blockers and ligation. *Dig Liver Dis* 2020; **52**: 1017-1025 [PMID: 32653417 DOI: 10.1016/j.dld.2020.06.005]

62 **Ma JL**, Chen X, He LL, Wei HS, Li P. Predictive value of child Pugh score, MELD score, MELD-Na score, APASAL score and R score in rebleeding and death of liver cirrhosis with esophagogastric varices. JCTH 2020; 36: 1278-1283. Available from: https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFDLAST2020&filename=LCGD202006022&uniplatform=NZKPT&v=oJBqDh7hJVjoGmr584nNKS-V3t9sAkvjdOfHnm_2cgW89jdFm-acWVMwM3_eZulC

63 **Bjerre-Nielsen A**, Kassarnig V, Lassen DD, Lehmann S. Task-specific information outperforms surveillance-style big data in predictive analytics. *Proc Natl Acad Sci USA* 2021; **118** [PMID: 33790010 DOI: 10.1073/pnas.2020258118]

64 **Liu L**, Liu Q, Xiao N, Zhang Y, Nie Y, Zhu X. A Liver Stiffness Measurement-Based Nomogram Predicts Variceal Rebleeding in Hepatitis B-Related Cirrhosis. *Dis Markers* 2022; **2022**: 4107877 [PMID: 35692881 DOI: 10.1155/2022/4107877]

65 **Sourabh D**. Clinical Epidemiology: Principles, Methods and Applications for Clinical Research. D E Grobbee and A W Hoes. *Int J Epidemiol* 2010; **39**: 318-319 [DOI: 10.1093/ije/dyn349]

66 **Reilly BM**, Evans AT. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. *Ann Intern Med* 2006; **144**: 201-209 [PMID: 16461965 DOI: 10.7326/0003-4819-144-3-200602070-00009]

Published by **Baishideng Publishing Group Inc**
7041 Koll Center Parkway, Suite 160, Pleasanton, CA 94566, USA
**Telephone:** +1-925-3991568
**E-mail:** bpgoffice@wjgnet.com
**Help Desk:** https://www.f6publishing.com/helpdesk
https://www.wjgnet.com