香港中文大學
The Chinese University of Hong Kong

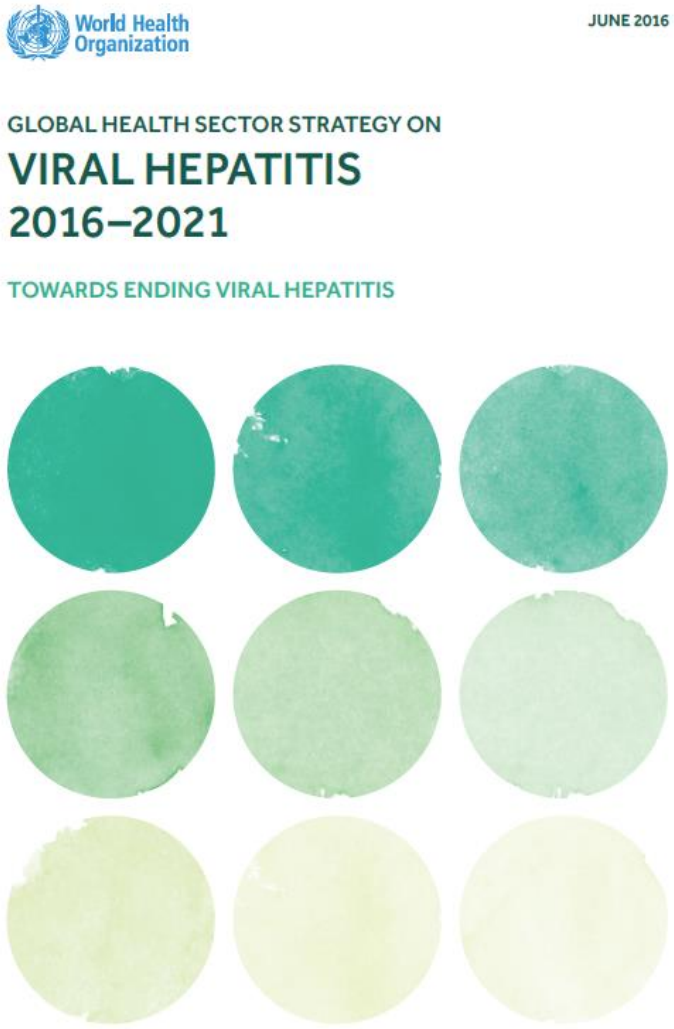
香港中文大學醫學院
Faculty of Medicine
The Chinese University of Hong Kong

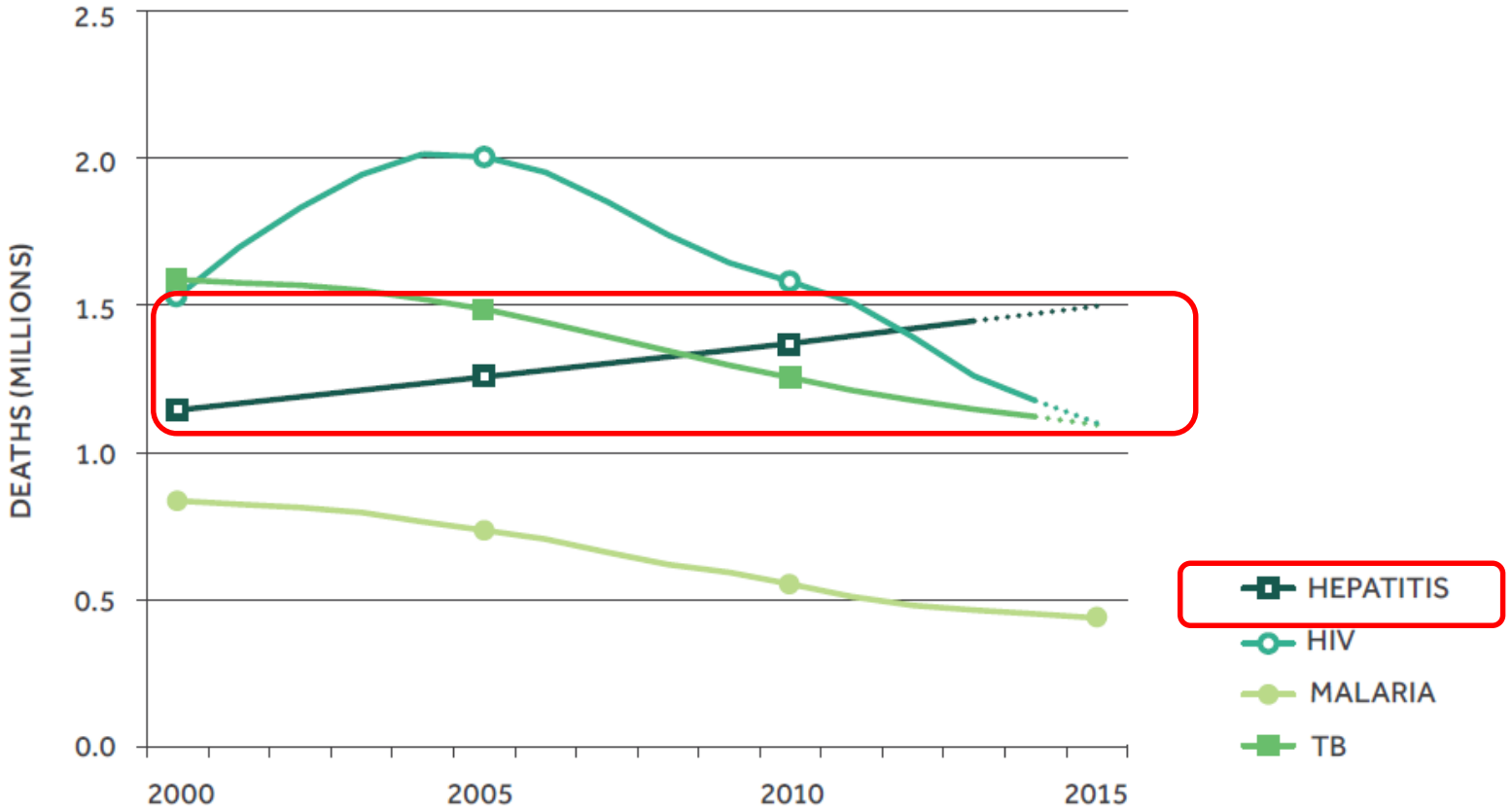
Parallel Session 4: Advanced Technologies

Disease burden of chronic viral hepatitis in Hong Kong – towards eliminating viral hepatitis by year 2030 according to World Health Organization (WHO) targets

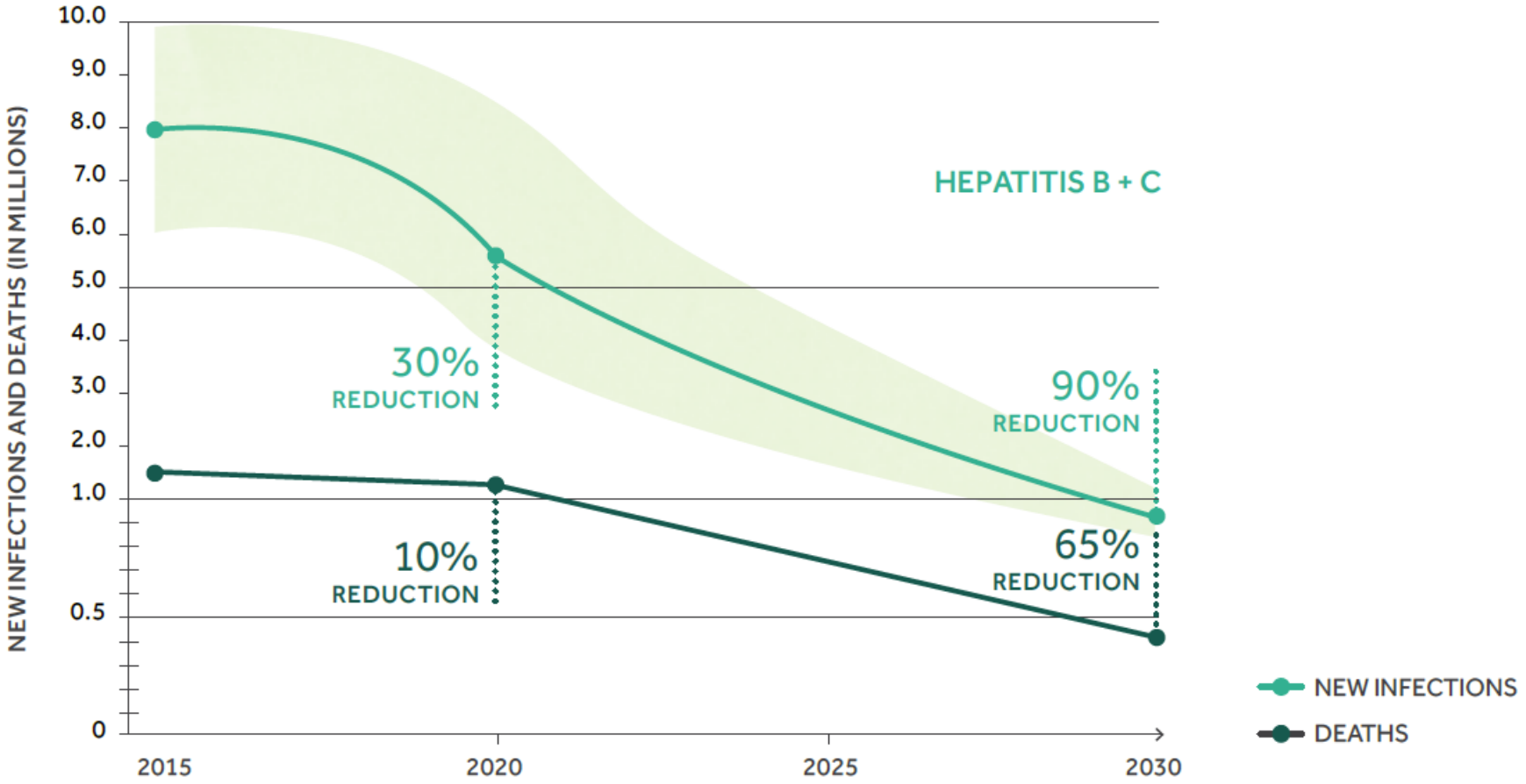
Applicants: *Wong Lai Hung Grace (PA), Wong Wai Sun Vincent (Co-PA), Tse Yee-Kit (Co-PA), Department of Medicine and Therapeutics, The Chinese University of Hong Kong* (HMRF 07180216)

World Health Organization Advocacy issued in June 2016 Elimination of Hepatitis B and C by 2030





Targets for reducing new cases of and deaths from chronic viral hepatitis B and C infection



Global Vision, Goal and Targets

Impact targets

TARGET AREA	BASELINE 2015	2020 TARGETS	2030 TARGETS
Incidence: New cases of chronic viral hepatitis B and C infections	6 -10 million infections reduced to 0.9 million infections by 2030 (95% decline in HBV, 80% decline in HCV)	30% reduction (equivalent to 1% prevalence of HBsAg among children)	90% reduction (equivalent to 0.1% prevalence of HBsAg among children)
Mortality: Viral hepatitis B and C deaths	1.4 million deaths reduced to <500 000 by 2030 (65% for both HBV/HCV)	10% reduction	65% reduction

Global Vision, Goal and Targets

Impact targets

TARGET AREA	BASELINE 2015	2020 TARGETS	2030 TARGETS
Incidence: New cases of chronic viral hepatitis B and C infections	6 -10 million infections reduced to 0.9 million infections by 2030 (95% decline in HBV, 80% decline in HCV)	30% reduction (equivalent to 1% prevalence of HBsAg among children)	90% reduction (equivalent to 0.1% prevalence of HBsAg among children)
Mortality: Viral hepatitis B and C deaths	1.4 million deaths reduced to <500 000 by 2030 (65% for both HBV/HCV)	10% reduction	65% reduction


12
1
2
10
11
9
8
7
6
Jan
10

GLOBAL HEPATITIS SUMMIT 2020/2021
The 17th International Symposium on Viral Hepatitis and Liver Disease (ISVHLD)

June 18-20
EVOLUTION AND REVOLUTION OF VIRAL HEPATITIS AND LIVER DISEASE

CU Medicine HONG KONG

Concurrent Session - June 19, 2021 (Abstract no. 234)

Secular trend of disease burden in patients with chronic viral hepatitis – a territory-wide study of 143,701 patients with data from HADCL in 2000-2018

Grace Wong, Vicki Hui, Yee-Kit Tse, Terry Yip, PC Yuen, Vincent Wong

Supported by HMRF 07180216

Presidential Awards (Public Health), Global Hepatitis Summit 2021


GLOBAL HEPATITIS SUMMIT 2020/2021
The 17th International Symposium on Viral Hepatitis and Liver Disease (ISVHLD)

Certificate of Award

The Presidential Award – Oral Presentation is presented to

Grace Wong

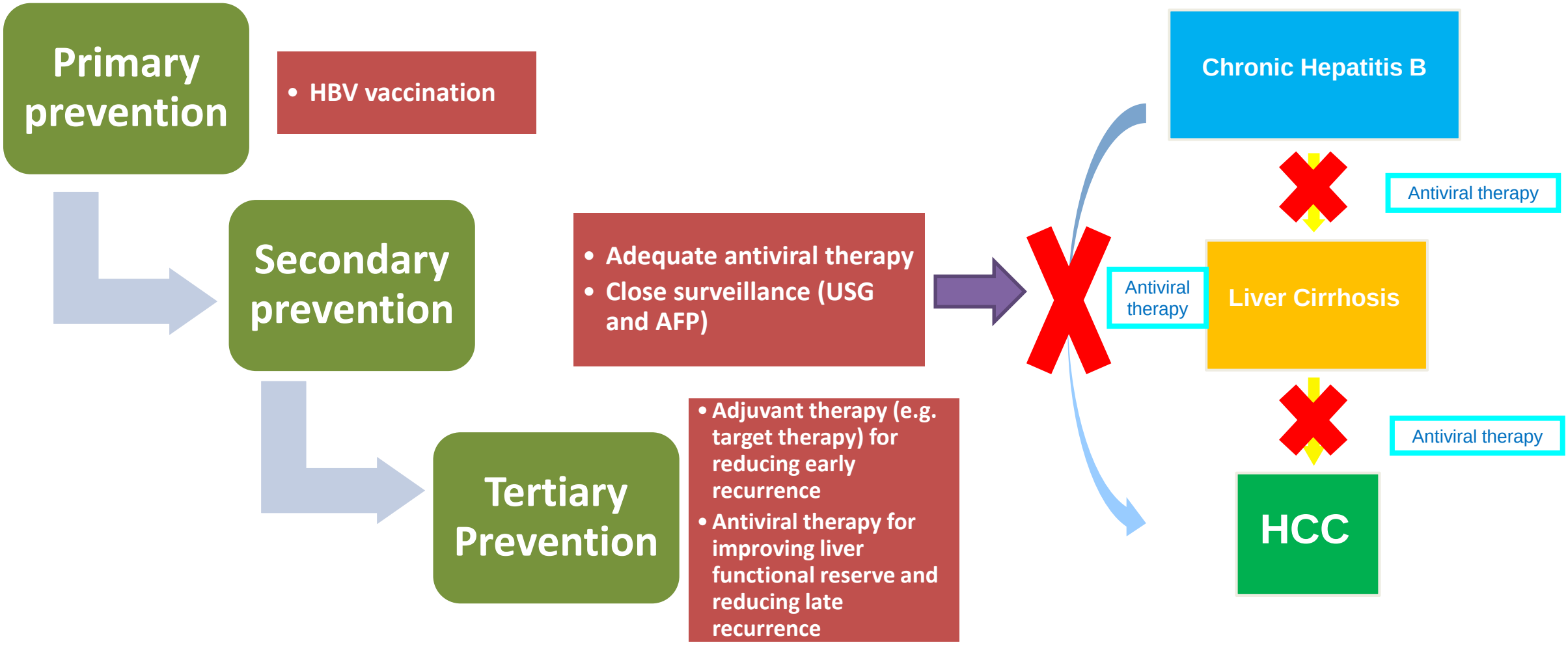
In the recognition of an outstanding presentation of the paper

Secular trend of disease burden in patients with chronic viral hepatitis – a territory-wide study of 143,701 patients with data from HADCL in 2000-2018

at ISVHLD GHS 2020 / 2021

JUNE 18-20, 2021, Virtual Conference

Antiviral treatment – secondary prevention for HCC to reduce disease burden



Wong GL, et al. Hepatology 2013;58:1537–1547
Wong GL, et al. Gastroenterology 2013;144:933-44

HADCL: Hospital Authority Data Collaboration Lab, Hong Kong

HADCL: Enabling HK Healthcare Innovation

HKSAR CE's 2017 Policy Address

- HA to establish a Big Data Analytics Platform
- Support formulation of healthcare policies
- Facilitate biotechnological research
- Improve clinical and healthcare services
- Promote innovation in healthcare services

IT.INNOVATION DATA.COLLABORATION.LAB

香港中文大學 The Chinese University of Hong Kong

香港中文大學醫學院 Faculty of Medicine The Chinese University of Hong Kong

Source: HADCL

Aim

- To determine the secular trend of disease burden in the territory-wide CVH cohort in Hong Kong from year 2000 to 2018 with data retrieved from HADCL.

Methods

- A territory-wide retrospective observational cohort study in Hong Kong

- CVH patients through HADCL, based on laboratory data of viral markers, diagnosis codes and medication record of antiviral treatment for chronic hepatitis B and/or C.

- Advanced liver fibrosis was defined by serum fibrosis scores (APRI, Forns index and FIB-4 scores).

Advanced liver fibrosis over the four periods

Period	2000 – 2004	2005 – 2009	2010 – 2013	2014 – 2018
APRI ≥ 2	552(19.59%)	1017(9.87%)	948(7.92%)	1307(8.99%)
FIB-4 ≥ 3.25	98(3.48%)	176(1.71%)	246(2.05%)	447(3.08%)
Forns index ≥ 8.4	163(22.42%)	609(14.82%)	725(13.63%)	1211(16.58%)

Prevalence of advanced liver fibrosis decreased over the last two decades.

Anti-HBV treatment over the four periods

Period	2000 – 2004	2005 – 2009	2010 – 2013	2014 – 2018
Any antiviral treatment	2248(11.79%)	5889(12.05%)	15249(17.76%)	51572(40.64%)
Lamivudine	2059(10.80%)	4934(10.10%)	7985(9.30%)	9464(7.46%)
Adefovir Dipivoxil	38(0.20%)	1180(2.41%)	2611(3.04%)	2954(2.33%)
Entecavir	0(0.00%)	1214(2.48%)	7709(8.98%)	42494(33.49%)
Telbivudine	0(0.00%)	229(0.47%)	1443(1.68%)	2741(2.16%)
Tenofovir*	<5(<0.05%)	102(0.21%)	1007(1.17%)	6036(4.76%)
Any NA	2077(10.90%)	6642(13.59%)	16512(19.23%)	51191(40.34%)
Conventional interferon	99(0.52%)	53(0.11%)	63(0.07%)	72(0.06%)
Peginterferon	11(0.06%)	375(0.77%)	690(0.80%)	909(0.72%)

Anti-HBV treatment increased over the last two decades.

Anti-HCV treatment over the four periods

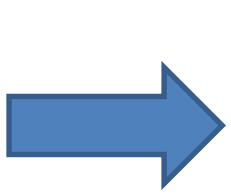
Period	2000 – 2004	2005 – 2009	2010 – 2013	2014 – 2018
Any antiviral treatment	1593(29.71%)	1162(12.83%)	1173(9.51%)	1733(10.31%)
Sofosbuvir+/-epclusa	0(0.00%)	0(0.00%)	0(0.00%)	273(1.62%)
Sofosbuvir+/-ledipasvir	0(0.00%)	0(0.00%)	0(0.00%)	12(0.07%)
Dasabuvir/Ombitasvir/ paritaprevir combination therapy (Viekira Pak)	0(0.00%)	0(0.00%)	0(0.00%)	119(0.71%)
Elbasvir/Grazoprevir (Zepatier)	0(0.00%)	0(0.00%)	0(0.00%)	0(0.00%)
Asunaprevir+/-Daclatasvir	0(0.00%)	0(0.00%)	0(0.00%)	107(0.64%)
Glecaprevir/Pibrentasvir (Mavyret)	0(0.00%)	0(0.00%)	0(0.00%)	6(0.04%)
Boceprevir	0(0.00%)	0(0.00%)	<5(<0.05%)	<5(<0.05%)
PegIFN + Ribavirin	1593(29.71%)	1161(12.82%)	1172(9.50%)	1293(7.69%)

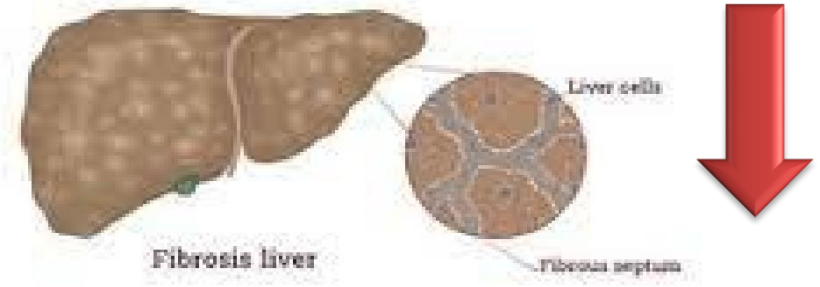
DAA treatment increased over the last few years.

Conclusions

- Prevalence of advanced liver fibrosis decreased over the last two decades

- This was likely related to the increasing uptake rate of antiviral treatment.





香港中文大學
The Chinese University of Hong Kong

AASLD Nov. 12-15, 2021
The Liver Meeting®
DIGITAL EXPERIENCE

CU Medicine HONG KONG
香港中文大學醫學院
Faculty of Medicine
The Chinese University of Hong Kong

Oral Presentation

Machine learning model is more accurate than conventional risk scores to predict hepatocellular carcinoma in patients with chronic viral hepatitis

Grace Wong, Vicki Hui, Qingxiong Tan, Terry Yip, Yee-Kit Tse, Chong Yin, Vincent Wong, Pong-Chi Yuen

Medical Data Analytics Centre (MDAC), Department of Medicine and Therapeutics, Institute of Digestive Disease, The Chinese University of Hong Kong; Department of Computer Science, Hong Kong Baptist University

Supported by HMRF 07180216

AASLD

HA Data Collaboration Lab (HADCL)

HA Data Collaboration Lab (HADCL)

Anonymized data
Physical access control
Standalone network
No data goes outside the lab

Artificial Intelligence, Machine Learning

Data Technology Platform for Innovation

IT.INNOVATION
DATA.COLLABORATION.LAB

香港中文大學
The Chinese University of Hong Kong

香港中文大學醫學院
Faculty of Medicine
The Chinese University of Hong Kong

Source: HADCL

HADCL Data Catalogue

Source: HADCL

Data, Tools & Platforms Requested

Library and Driver	Version		Package Name	Version	Package Name	Version	Other unlisted libraries	
CUDA	9.0.176		h5py	2.8.0	Pillow	5.2.0	tqdm	4.11.2
NVidia Driver	384.111		html5lib	1.0.1	plainbox	0.25	tflearn	0.3
			ipython	6.5.0	prompt-toolkit	1.0.15	nose	1.3.0
			ipython-genutils	0.2.0	protobuf	3.6.1	leveldb	0.191
			jupyter	1.0.0	Python	3.5.2	Cython	0.19.2
			jupyter-client	5.2.3	python-apt	1.1.0b1 ubuntu0.16.4.1	mxnet	1.3.0
			jupyter-console	5.2.0	python-dateutil	2.7.3	pyvlfeat	0.1.1a3
			jupyter-core	4.4.0	python-debian	0.1.27	keras-svm	1.0.0b10
			Keras	2.2.2	python-systemd	231	colorlog	2.10.0
			Keras-Applications	1.0.4	PyYAML	3.13	click	6.7
			Keras-Preprocessing	1.0.2	requests	2.9.1	chainer	1.17.0
			matplotlib	2.2.3	scikit-image	0.14.0	R for windows	version 3.5.1
			networkx	2.1	scikit-learn	0.19.2		
			numpy	1.14.5	scipy	1.1.0		
			pandas	0.23.4	six	1.10.0		

Aim

- To compare the accuracy of novel machine learning models with existing risk scores to predict HCC in the territory-wide cohort of chronic viral hepatitis in Hong Kong

- Patients diagnosed in year 2000-2018 with anonymized and de-identified data accessed in HADCL.

Wong GL, et al. Hepatology 2021;74(Supp):92A.
(Oral presentation in The Liver Meeting, AASLD, 2021)

Methods

- A territory-wide retrospective observational cohort study

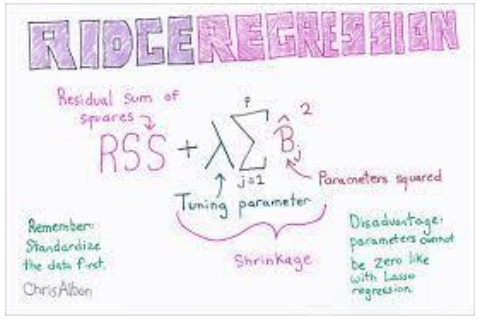
- Patients with chronic viral hepatitis identified through HADCL based on:
  - Laboratory data of viral markers
  - Diagnosis codes
  - Drug record of antiviral treatment for chronic hepatitis B/C

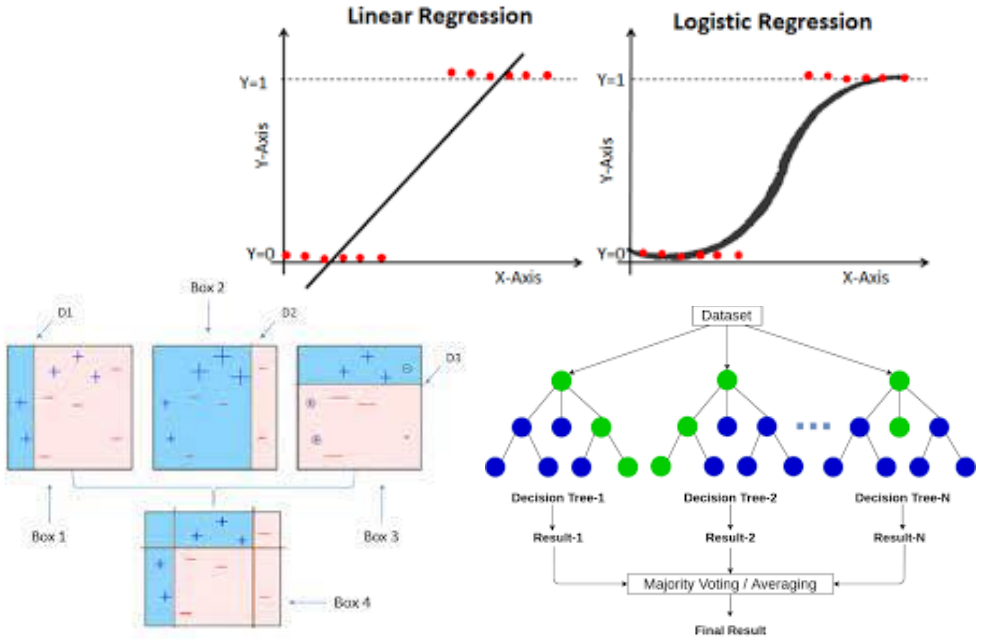
- Cohort was randomly split into training and validation cohorts in 7:3 ratio

Wong GL, et al. Hepatology 2021;74(Supp):92A.
(Oral presentation in The Liver Meeting, AASLD, 2021)

Machine learning models to predict HCC

- Five popular machine learning models applied:
  - Logistic regression
  - Ridge regression
  - AdaBoost
  - Decision tree
  - Random Forest





- Accuracy of the models was assessed by the area under the receiver operating characteristic curve (AUROC)

Wong GL, et al. Hepatology 2021;74(Supp):92A.
(Oral presentation in The Liver Meeting, AASLD, 2021)

46 parameters used to develop the machine learning models

Male gender	Gamma glutamyl transferase	Urinary/renal malignancies	ACEI/ARB	Other oral hypoglycemic agents
Age	Total cholesterol	Cervical cancer	Antiplatelet agents	Proton pump inhibitor
Platelet	HbA1c	Breast cancer	Beta blockers	Potassium sparing diuretics
Albumin	Fasting glucose	Lymphoma	Histamine-2 receptor antagonist	Statins
Total bilirubin	HBV DNA	Chronic kidney disease	Insulin	Sulphonylurea
Alanine aminotransferase	Positive HBeAg	Osteopenia	Immunosuppressant	Thiazides
Aspartate aminotransferase	Cirrhosis	Osteoporosis	Loop diuretics	
Alpha-fetoprotein	Cardiovascular disease	Diabetes mellitus	Metformin	
International normalized ratio	Colorectal cancer	Hypertension	NSAID	
Creatinine	Lung cancers	Anticoagulants	Other lipid lowering agents	

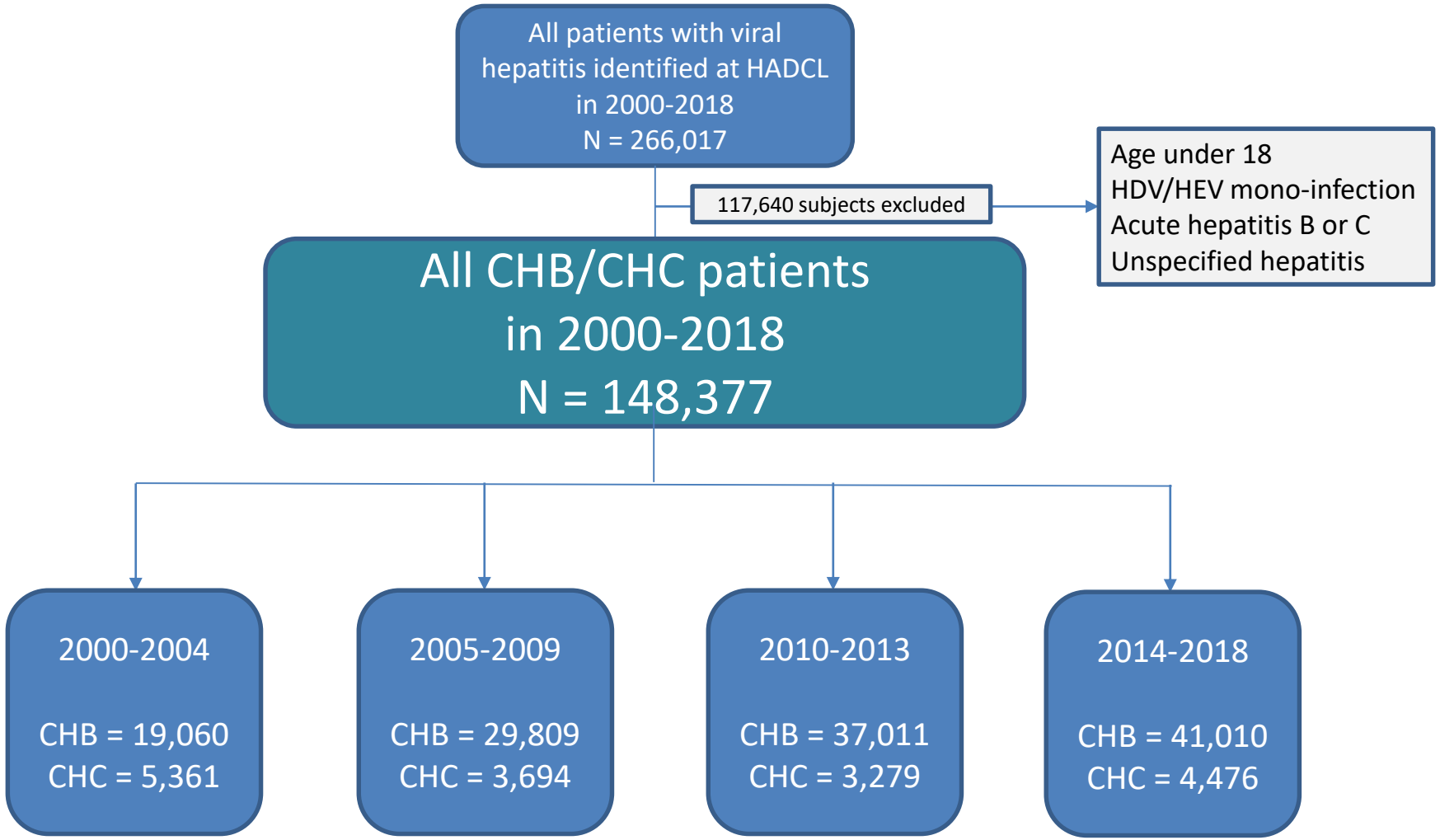
Wong GL, et al. Hepatology 2021;74(Supp):92A.
(Oral presentation in The Liver Meeting, AASLD, 2021)

26 selected parameters

Male gender	Gamma glutamyl transferase	Urinary/renal malignancies	ACEI/ARB	Other oral hypoglycemic agents
Age	Total cholesterol	Cervical cancer	Antiplatelet agents	Proton pump inhibitor
Platelet	HbA1c	Breast cancer	Beta blockers	Potassium sparing diuretics
Albumin	Fasting glucose	Lymphoma	Histamine-2 receptor antagonist	Statins
Total bilirubin	HBV DNA	Chronic kidney disease	Insulin	Sulphonylurea
Alanine aminotransferase	Positive HBeAg	Osteopenia	Immunosuppressant	Thiazides
Aspartate aminotransferase	Cirrhosis	Osteoporosis	Loop diuretics	
Alpha-fetoprotein	Cardiovascular disease	Diabetes mellitus	Metformin	
International normalized ratio	Colorectal cancer	Hypertension	NSAID	
Creatinine	Lung cancers	Anticoagulants	Other lipid lowering agents	

Wong GL, et al. Hepatology 2021;74(Supp):92A.
(Oral presentation in The Liver Meeting, AASLD, 2021)

Large number of patients are included via HADCL



Wong GL, et al. Hepatology 2021;74(Supp):92A.
(Oral presentation in The Liver Meeting, AASLD, 2021)

AUROC of the machine-learning models

	*Training cohort: N = 86,804, HCC = 6,821		^Validation cohort: N = 37,202, HCC = 2,875	
	All 46 parameters	*Selected 26 parameters*	*All 46 parameters*	*Selected 26 parameters*
Logistic regression	0.825±0.006	0.829±0.006	0.829±0.009	0.832±0.009
+- Ridge regression	0.842±0.006	0.839±0.006	0.844±0.009	0.840±0.009
AdaBoost	0.828±0.006	0.828±0.006	0.832±0.009	0.833±0.009
+ Decision tree	0.881±0.005	0.884±0.005	0.818±0.010	0.819±0.010
+- Random Forest	0.992±0.002	0.991±0.002	0.821±0.010	0.821±0.010

* AUROC of the four machine learning algorithms were overall difference in training group, P<0.05
^ AUROC of the four machine learning algorithms were overall difference in validation group, P<0.05
+ AUROC higher than AdaBoost, P<0.05
- AUROC higher than decision tree, P<0.05

Wong GL, et al. Hepatology 2021;74(Supp):92A.
(Oral presentation in The Liver Meeting, AASLD, 2021)

AUROC of the machine-learning models

	*Training cohort: N = 86,804, HCC = 6,821		^Validation cohort: N = 37,202, HCC = 2,875	
	All 46 parameters	*Selected 26 parameters*	*All 46 parameters*	*Selected 26 parameters*
Logistic regression	0.825±0.006	0.829±0.006	0.829±0.009	0.832±0.009
+- Ridge regression	0.842±0.006	0.839±0.006	0.844±0.009	0.840±0.009
AdaBoost	0.828±0.006	0.828±0.006	0.832±0.009	0.833±0.009
+ Decision tree	0.881±0.005	0.884±0.005	0.818±0.010	0.819±0.010
+- Random Forest	0.992±0.002	0.991±0.002	0.821±0.010	0.821±0.010

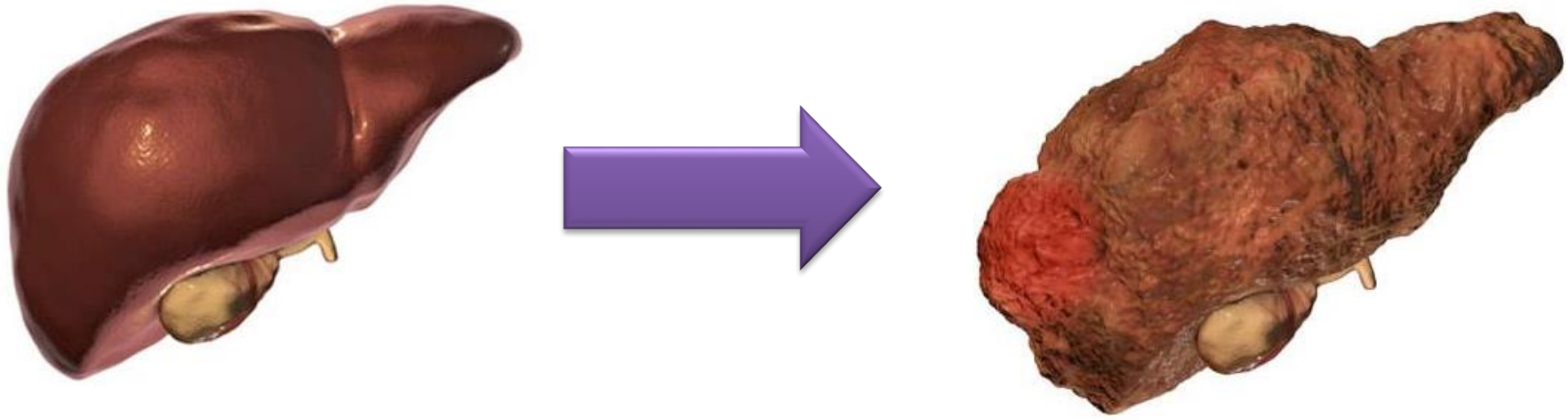
* AUROC of the four machine learning algorithms were overall difference in training group, P<0.05
^ AUROC of the four machine learning algorithms were overall difference in validation group, P<0.05
+ AUROC higher than AdaBoost, P<0.05
- AUROC higher than decision tree, P<0.05

Wong GL, et al. Hepatology 2021;74(Supp):92A.
(Oral presentation in The Liver Meeting, AASLD, 2021)

HCC ridge regression score (HCC-RS)

Consistently good performance in both training and validation cohort

AUROC = 0.84



Wong GL, et al. Hepatology 2021;74(Supp):92A.
(Oral presentation in The Liver Meeting, AASLD, 2021)

Ridge regression in a nutshell

To minimize the objective function

f: objective function

error

penalty

$$f(\beta) = \|X\beta - Y\|_2^2 + \lambda\|\beta\|_2^2$$

β: HCC risks

X: training data

Y: observations

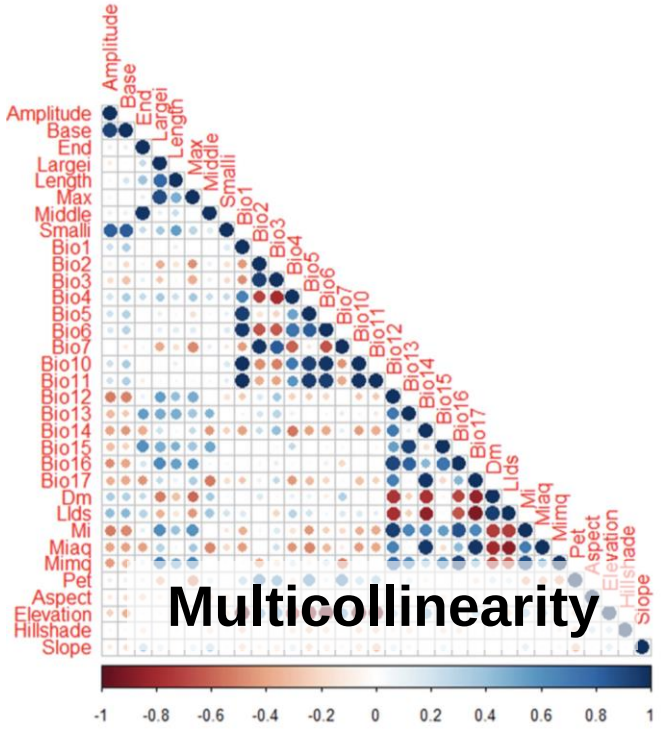
λ: Regularization parameter

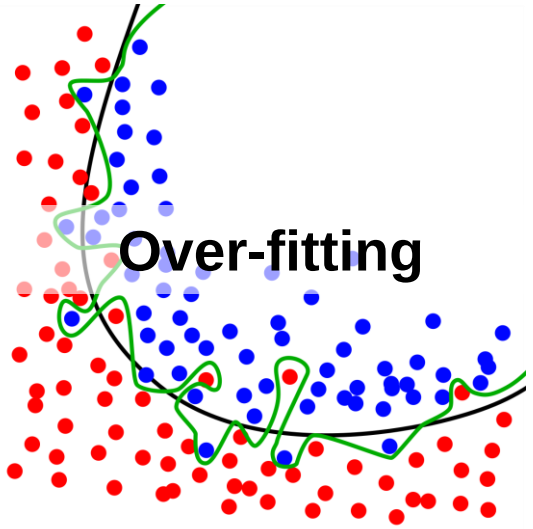
Wong GL, et al. Hepatology 2021;74(Supp):92A.
(Oral presentation in The Liver Meeting, AASLD, 2021)

Why ridge regression?

- A technique for analyzing multiple regression data that suffer from **multicollinearity**
  - works particularly well in clinical medicine
  - many parameters included in the models are closely related so that multicollinearity commonly occurs

- Ridge regression penalty to avoid **over-fitting**





Wong GL, et al. Hepatology 2021;74(Supp):92A.
(Oral presentation in The Liver Meeting, AASLD, 2021)

Machine-learning models outperform conventional HCC risk scores

Risk scores	AUROC	Dual Cut-offs	% (< lower cut-off / ≥ upper cut-off)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
HCC-RS	0.840	<10	56.0	81.2	79.0	30.1	97.4
		≥20	13.3	44.1	92.3	38.8	93.7
CU-HCC score	0.672	<5	72.8	46.4	74.0	10.3	95.6
		≥20	21.0	32.1	79.7	9.1	94.8
GAG-HCC score	0.745	<80	69.3	64.3	71.5	12.3	97.0
		≥101	7.9	28.6	93.4	21.1	95.5
REACH-B score	0.671	<8	50.0	72.7	52.8	16.2	94.1
		≥14	1.5	4.5	98.9	33.3	89.2
PAGE-B score	0.748	<10	27.4	95.7	29.4	10.7	98.7
		≥13	48.3	81.1	54.6	13.7	97.0
REAL-B score	0.712	<4	17.6	96.0	19.2	12.0	97.7
		≥8	11.5	27.0	90.3	24.2	91.5

Machine-learning models outperform conventional HCC risk scores

Risk scores	AUROC	Dual Cut-offs	% (< lower cut-off / ≥ upper cut-off)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
HCC-RS	**0.840**	**<10**	**56.0**	**81.2**	**79.0**	**30.1**	**97.4**
		≥20	**13.3**	**44.1**	**92.3**	**38.8**	**93.7**
CU-HCC score	0.672	<5	72.8	46.4	74.0	10.3	95.6
		≥20	21.0	32.1	79.7	9.1	94.8
GAG-HCC score	0.745	<80	69.3	64.3	71.5	12.3	97.0
		≥101	7.9	28.6	93.4	21.1	95.5
REACH-B score	0.671	<8	50.0	72.7	52.8	16.2	94.1
		≥14	1.5	4.5	98.9	33.3	89.2
PAGE-B score	0.748	<10	27.4	95.7	29.4	10.7	98.7
		≥13	48.3	81.1	54.6	13.7	97.0
REAL-B score	0.712	<4	17.6	96.0	19.2	12.0	97.7
		≥8	11.5	27.0	90.3	24.2	91.5

Conclusions

- HCC-RS developed with ridge regression of machine learning approach outperforms the conventional risk scores to predict HCC in patients with chronic viral hepatitis.

- This machine learning model may become built-in functional keys or calculators in the electronic health systems to facilitate hepatitis elimination.

Wong GL, et al. Hepatology 2021;74(Supp):92A.
(Oral presentation in The Liver Meeting, AASLD, 2021)

Era of machine-learning for HCC risk prediction is coming