

# BMJ Open Bayesian network modelling study to identify factors influencing the risk of cardiovascular disease in Canadian adults with hepatitis C virus infection

Alaa Badawi <sup>1,2</sup>, Giancarlo Di Giuseppe <sup>3,4</sup>, Alind Gupta,<sup>5</sup> Abbey Poirier,<sup>6</sup> Paul Arora<sup>3</sup>

**To cite:** Badawi A, Di Giuseppe G, Gupta A, *et al.* Bayesian network modelling study to identify factors influencing the risk of cardiovascular disease in Canadian adults with hepatitis C virus infection. *BMJ Open* 2020;**10**:e035867. doi:10.1136/bmjopen-2019-035867

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2019-035867>).

Received 29 November 2019  
Revised 18 March 2020  
Accepted 19 March 2020



© Author(s) (or their employer(s)) 2020. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

## Correspondence to

Dr Alaa Badawi;  
[alaa.badawi@canada.ca](mailto:alaa.badawi@canada.ca)

## ABSTRACT

**Objectives** The present study evaluates the extent of association between hepatitis C virus (HCV) infection and cardiovascular disease (CVD) risk and identifies factors mediating this relationship using Bayesian network (BN) analysis.

**Design and setting** A population-based cross-sectional survey in Canada.

**Participants** Adults from the Canadian Health Measures Survey ( $n=10\ 115$ ) aged 30 to 74 years.

**Primary and secondary outcome measures** The 10-year risk of CVD was determined using the Framingham Risk Score in HCV-positive and HCV-negative subjects. Using BN analysis, variables were modelled to calculate the probability of CVD risk in HCV infection.

**Results** When the BN is compiled, and no variable has been instantiated, 73%, 17% and 11% of the subjects had low, moderate and high 10-year CVD risk, respectively. The conditional probability of high CVD risk increased to  $13.9\% \pm 1.6\%$  ( $p < 2.2 \times 10^{-16}$ ) when the HCV variable is instantiated to 'Present' state and decreased to  $8.6\% \pm 0.2\%$  when HCV was instantiated to 'Absent' ( $p < 2.2 \times 10^{-16}$ ). HCV cases had 1.6-fold higher prevalence of high-CVD risk compared with non-infected individuals ( $p=0.038$ ). Analysis of the effect modification of the HCV-CVD relationship (using median Kullback-Leibler divergence;  $D_{KL}$ ) showed diabetes as a major effect modifier on the joint probability distribution of HCV infection and CVD risk ( $D_{KL}=0.27$ , IQR: 0.26 to 0.27), followed by hypertension (0.24, IQR: 0.23 to 0.25), age (0.21, IQR: 0.10 to 0.38) and injection drug use (0.19, IQR: 0.06 to 0.59).

**Conclusions** Exploring the relationship between HCV infection and CVD risk using BN modelling analysis revealed that the infection is associated with elevated CVD risk. A number of risk modifiers were identified to play a role in this relationship. Targeting these factors during the course of infection to reduce CVD risk should be studied further.

## INTRODUCTION

The prevalence of hepatitis C virus (HCV) infection is estimated to be around 3% worldwide and 1.3% in North America.<sup>1</sup> The infection causes chronic liver diseases in 170 to

## Strengths and limitations of this study

- This is the first study to introduce Bayesian network (BN) analysis to characterise the aetiological role of hepatitis C virus (HCV) infection in cardiovascular disease (CVD) risk in a Canadian adult population.
- The study examined 10 115 subjects aged 30 to 74 years from the Canadian Health Measures Survey, a cross-sectional survey of the non-institutionalised civilian residents.
- Assessment for the 10-year risk of CVD in asymptomatic individuals was carried by estimating the Framingham Risk Score algorithm in both HCV-positive and HCV-negative subjects.
- Machine learning algorithms using BNs were used to determine the relationship between HCV and CVD risk in the study population.
- The study did not capture the status of HIV or the use of direct-acting antiviral therapy, factors that may have a significant effect on the relationship between HCV infection and CVD risk.

200 million people worldwide rendering it a leading factor in the aetiology of progressive liver fibrosis that results in cirrhosis, liver cancer, liver failure and death.<sup>2-3</sup> Presently, the rate of mortality among persons living with HIV infection falls behind those living with HCV,<sup>3</sup> ranking the latter among the top causes of death globally<sup>4</sup> and the leading infection-related cause of mortality in North America.<sup>5</sup> While HCV exerts its main effects in the liver, over the past decade several lines of evidence emerged to suggest it as a causative factor in a number of extrahepatic manifestations<sup>6</sup> including abnormal endocrine, haematological, neurological and renal functions.<sup>7-9</sup> Although yet to be fully substantiated,<sup>10-11</sup> evidence suggest that patients with chronic HCV infection exhibit elevated rates of cause-specific mortality from cardiovascular disease (CVD).<sup>12-15</sup> The role of HCV in the manifestation of CVD was thought to be related to the

interference of infection with glucose and lipid metabolism and the subsequent risk of type 2 diabetes and CVD.<sup>16</sup> Recently, in addition to this altered profile of cardiometabolic risk factors, we reported a distinctive pattern of acute phase reactants (such as C-reactive protein, fibrinogen and homocysteine) in the HCV infected patients compared with non-infected subjects, implicating the status of chronic inflammation in the elevated CVD risk associated with HCV infection.<sup>17</sup> Furthermore, vascular endothelial dysfunction resulting in carotid atherosclerotic plaques has been proposed to play a role in CVD risk among HCV-infected individuals.<sup>18–20</sup>

Despite proposing a number of mechanisms and an array of causative factors, the aetiological relationship between HCV infection and risk of CVD is still speculative as described above. Together with experimental work, and clinical studies, the potential causal links between HCV infection and CVD risk can be further explored, predicted and ascertained using population-based data sets. This direction represents the first step towards identification of disease-causing mechanisms, treatment and control.<sup>21</sup> A fundamental step in analysing big data or large population-based data sets is to identify relations among a large number of variables in a data set.<sup>22</sup> Bayesian networks (BN) are a widely-used class of probabilistic graphical models that represent a set of variables and their conditional dependencies, permitting the identification of relationships that may highlight causality.<sup>22</sup> These relationships are represented by a graphical structure, whereas the quantitative dependencies between individual variables is expressed by conditional probability.<sup>23</sup>

A BN is defined by a graph structure of nodes (variables) connected by directed edges and a set of conditional probabilities. The generated model is dynamic where the probability of all variables changes by changing the state of one variable and reveals the influences and interactions between the depicted variables.<sup>23</sup> The model has been used for intelligent information processing in many scientific fields, ranging from biomedical to social sciences, for example, risk prediction, diagnosis and assessment,<sup>24–25</sup> clinical decision-making,<sup>26</sup> inference of causation<sup>27</sup> and social network analysis.<sup>28</sup> Over the past few years, BNs have been extensively used to model clinical problems in CVD for the purposes of diagnosis, risk assessment and disease prediction.<sup>29–33</sup> In the present study, we introduced a BN analysis to evaluate the aetiological role of HCV infection in CVD risk. Our objective is to characterise the multivariable probabilistic connection between the two diseases and identify factors that mediate and influence this relationship in a population of Canadian adults.

## METHODS

### Study population

Data were collected from the Canadian Health Measures Survey (CHMS), a cross-sectional survey of the non-institutionalised civilian Canadian residents aged 3 to

79 years, designed to collect information on the health and wellness as well as nutrition status of the population. The CHMS covers approximately 96.3% of the Canadian population at the selected age group. The current study uses a cohort that we previously described.<sup>17</sup> Briefly, participants in this study were obtained from CHMS cycles 1 to 4, collected between March 2007 and December 2015 (n=5604, 6395, 5785 and 5794 for cycles 1, 2, 3 and 4, respectively, n=23 578). The multistage sampling design, participation and data collection methods have been comprehensively mentioned elsewhere.<sup>34–36</sup> Participation was voluntary and was through a household interview that included general sociodemographic questions as well as an in-depth health questionnaire. Participants subsequently visited a mobile examination centre (MEC) where physical and biological measurements were collected.<sup>37</sup>

### Assessment for the 10-year risk of cardiovascular disease

The 10-year CVD risk was estimated in subjects who were asymptomatic at baseline and aged 30 to 74 years, using the Framingham Risk Score (FRS) algorithm in both HCV-positive (HCV+) and HCV-negative (HCV-) cases as previously described.<sup>17 38 39</sup> As per the conditions of the FRS algorithm, we excluded subjects under the age of 30 years or over the age of 74 years (n=12 348). History of CVD was assessed via self-report as the participants were asked if they have or had heart disease and if they have ever been told by health professional that they have or had heart attack. Patients with confirmed or a history of heart diseases (n=888) were excluded. We also excluded participants with missing data for any of the components used in the FRS algorithm that included age, gender, blood pressure, total cholesterol (T-Chol), high-density lipoprotein cholesterol (HDL-C), smoking status, diabetes and anti-hypertensive medication (n=51). Furthermore, we did not include participants missing information on HCV status or with an indeterminate HCV infection (n=176). The total number of participants included in the present study was 10 115 subjects (male:female ratio of 1:1.12). This group was further divided to HCV-positive (HCV+, n=87) and HCV-negative (HCV-, n=10 028) subgroups. In both subgroups, we calculated the FRS score for each participant. The points assigned for each risk factor are based on the value for the  $\beta$ -coefficient of the proportional hazard regressions.<sup>38 39</sup> The 10-year risk score was then derived as a percentage. As previously described,<sup>39</sup> the 10-year CVD risk is categorised into low (FRS <10%), moderate (10% to 19%) or high ( $\geq$ 20%) based on the derived estimate.

### Assessment of hepatitis C infection

Infection with HCV was determined using serum specimens screened for anti-HCV antibodies or HCV-RNA. HCV-positive status was considered if HCV-RNA was present in the serum and/or a confirmed antibody test yielded positive results. Participants with indeterminate anti-HCV and without RNA testing were considered 'unknown' and were excluded as HCV status could not

be determined with certainty. HCV infection was detected using VITROS Anti-HCV Assay (Ortho Clinical Diagnostics, Raritan, New Jersey, USA). All samples positive on initial screening for anti-HCV were confirmed by INNO-LIA HCV Score immunoblot assay (Innogenetics, Fujirebio Inc, Georgia, USA). All HCV-positive patients (n=87) had anti-HCV testing performed. Among those, 21 subjects were also examined for HCV-RNA, 14 of which were positive for HCV-RNA. Data on genotypes and sub-genotypes was previously described.<sup>17</sup> Information on the use of the recent direct-acting antivirals or the length of stay in prison in HCV positive or negative groups (if any) were not captured.

### Metabolic markers, sociodemographic factors and other covariates

A number of metabolic markers were measured in the study including cardiometabolic disease markers (apolipoprotein (Apo) A1 (g/L), ApoB (g/L), low-density lipoprotein cholesterol (LDL-C) (mmol/L), HDL-C (mmol/L), T-Chol (mmol/L), T-Chol:HDL-C ratio, triglycerides (mmol/L) and glycosylated haemoglobin (HbA1c) (%)); inflammatory biomarkers (C-reactive protein (mg/L), fibrinogen (g/L) and homocysteine (μmol/L)); systolic and diastolic blood pressure and plasma 25-hydroxyvitamin D (25(OH)D) (nmol/L) as well as factors related to obesity such as body mass index (kg/m<sup>2</sup>), waist circumference (cm) and waist-to-hip ratio were all assessed as described previously.<sup>40</sup> Diabetes status was defined as a self-reported or HbA1c ≥6.5%.<sup>41</sup> Assessing metabolic syndrome was carried out as previously described.<sup>42</sup> Individuals who have been diagnosed as hypertensive, diabetic or those who were using antihypertensive drugs were included.<sup>40 43 44</sup> Insulin resistance (IR) was approximated using the homeostatic model assessment (HOMA-IR) formula ((glucose (mmol/L) x insulin (μIU/mL)) ÷ 22.5) as described.<sup>17</sup> Liver functions were evaluated in the study subjects using serum enzyme markers of alanine aminotransferase, alkaline phosphatase, aspartate aminotransferase, lactate dehydrogenase and γ-glutamyl transferase (U/L). Furthermore, a number of serum micronutrients including vitamins B12 (pmol/L) and D (nmol/L) were measured and captured in the present study. Sociodemographic information such as age, gender, ethnicity, education, marital status, history of injection drug use and household income was assessed through responses to questionnaires given during the structured interview portion of the survey. Ethnicity was categorised into four main subgroups: White, African-Americans, Asian (ie, Korean, Filipino, Japanese, Chinese, South Asian, Southeast Asian, Arab and West Asian) and Other (ie, Latin American or mixed racial origins). Household income was dichotomised according to self-reported income below or above \$20 000 per annum. Self-reported smoking status was categorised into smokers (daily/occasional) and non-smokers. A comprehensive list of all metabolic markers, sociodemographic

factors and other covariates assessed in the present study are listed in online supplementary table S1.

### Bayesian network analysis and statistical methods

Machine learning algorithms using BNs were used to determine the associations between HCV infection and CVD risk in the study population.

#### Bayesian network formalism

According to the chain rule of probability, for any set of  $N$  categorical variables  $\{X_1, X_2, \dots, X_N\}$ , the joint probability distribution (JPD) was:

$$P(X_1, X_2, \dots, X_n) = \prod_{i=1}^N P(X_i | X_{i1}, \dots, X_{iN})$$

Given a directed acyclic graph (DAG) over the variables, the parents of any variable  $X_i$  denoted by  $Pa(X_i)$  is the set of variables in the DAG that have a directed edge ( $\rightarrow$ ) to  $X_i$ . For a BN induced by a DAG, the JPD factorises as follows:

$$P(X_1, X_2, \dots, X_n) = \prod_{i=1}^N P(X_i | X_{i1}, \dots, X_{iN}) = \prod_{i=1}^N P(X_i | Pa(X_i))$$

For example, given the two node BN with the configuration  $X_1 \rightarrow X_2$ , the JPD is  $P(X_1, X_2) = P(X_2 | X_1) P(X_1)$ . In a BN, any variable  $X_i$  is independent of all other variables conditioned on its Markov blanket  $MB(X_i)$ , which is the set containing  $Pa(X_i)$ , the children of  $X_i$  (ie, have a directed edge from  $X_i$ ) and parents of children of  $X_i$ .  $MB(X_i)$ , if fully observed, provides all the information about  $X_i$  in a BN. Detailed statistical description of BNs can be found elsewhere.<sup>45 46</sup> The property stating that any node  $X$  is conditionally independent of any other node, given its Markov blanket, is the global Markov property. Another property that has been taken into account is the local Markov condition, that is, a node  $X$  is conditionally independent of those of all its non-descendants given the set of all its parents. The flow of influence has been taken into consideration, when two variables (nodes) are d-separated for different types of connections.<sup>45 46</sup>

#### Learning Bayesian network structure from missing data

We excluded variables containing >20% missing values from our analysis and assumed that missingness was at random. To learn BN structures from missing data, we used structural expectation maximisation (EM). EM is an algorithm for finding maximum likelihood estimates of models with latent parameters by iteratively calculating the expectation of the model with respect to the parameter estimates (usually initialised randomly) at the current EM step, and then finding parameters that maximise this expectation.

For initialising EM, we sampled 20 DAGs from a uniform distribution over the space of connected DAGs with a maximum degree of 3 using the method by Ide and Cozman<sup>47</sup> implemented in the R package *bnlearn*.<sup>48</sup> These DAGs were parametrised on the data set (see Bayesian network parametrisation) and used to initialise 20 independent EM runs. Convergence for EM was assessed using



the structural Hamming distance (SHD)<sup>49</sup> of the DAG at the current EM step relative to the previous EM step. An SHD, which is the number of edge additions, deletions or reversals required to transform one graph to another, of <20 was assumed to signify convergence because SHD did not reliably decrease if additional EM steps were allowed. Following convergence, the resulting 20 structures were averaged to retain oriented edges that occurred in at least 50% of the structures. This was the highest decile that did not render the three outcomes of interest independent.

For the maximisation step of EM, we used the score-based hill climbing structure learning algorithm<sup>46</sup> to search the space of DAGs that maximise the Bayesian information criterion (BIC), a commonly used penalised maximum likelihood-based model score. Hill climbing is a greedy local search algorithm that performs sequential edge modifications (edge additions, reversals and deletions) to find a DAG that maximises a given network score. No constraints on edge orientations or presence/absence were placed during training. The validity of the recovered final structure was assessed visually by comparing it to known/expected relationships, such as between correlated variables (eg, diabetes and obesity) and causally related variables (eg, having hypertension and taking blood pressure medicine). Model averaging over 20 structures from EM was done to reduce uncertainty in model learning to select a robust model structure.

#### Learning Bayesian network parameters

Network parameters were learnt using Bayesian estimation, which calculates the expected value of the posterior distribution over the parameters. A Dirichlet prior with an equivalent sample size of 1 was used as a weak prior over the parameters.

#### Approximate inference using likelihood weighting

To calculate the probability of a given query from the BN, we used a form of importance sampling called likelihood weighting implemented in *bnlearn*. Briefly, given a BN and an ordering over variables induced by the DAG underlying the BN, we can sample from the network (with or without conditioning) from a proposal distribution (usually the uniform distribution) along with weights that represents the probability of acceptance of the sample from the target distribution (represented by the trained BN). The probability of the query is computed by dividing the sum of the weights of all samples that were equal to the query of interest by the total weight of all samples.

#### Calculating predictiveness of variable for an outcome

We used conditional entropy  $H(Y|X)$  as the measure of predictiveness of an outcome  $Y$  by a predictor  $X$ ,  $H(Y|X)$  quantifies the averaged reduction in uncertainty about the outcome  $Y$  conditioned on having observed the variable  $X$ . Conditional entropy is defined as follows:

$$H(y|X) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \cdot \log_2 \left( \frac{p(x)}{p(x, y)} \right)$$

If  $X$  is perfectly predictive of  $Y$ , the conditional entropy  $H(Y|X)=0$  since there is no uncertainty about  $Y$  given  $X$ . The maximum value of conditional entropy is achieved when  $X$  and  $Y$  are independent and is equal to the entropy of  $Y$  which is:

$$H(y) = \sum_{y \in Y} p(y) \cdot \log p(y)$$

Therefore, conditional entropy is bounded by  $0 \leq H(Y|X) \leq H(Y)$ . For the predictive index (PI) reported in the paper, we normalised  $H(Y|X)$  by dividing it by  $H(Y)$  and inverted the values such that a value of 0 represents independence and 1 representing perfect prediction using the following transformation:

$$PI(Y|X) = 1 - \left( \frac{H(Y|X)}{H(Y)} \right)$$

#### Effect modification

A variable  $X$  was defined to be an effect modifier of outcomes  $Y$  and  $Z$  if the joint probability distribution of  $Y$  and  $Z$  conditioned on  $X$ , that is,  $P(Y, Z|X=x)$ , was not constant for  $x \in X$ . The strength of effect modification was calculated as the median Kullback-Leibler (KL) divergence ( $D_{KL}$ )—also known as relative entropy<sup>50</sup>—for  $Q_x = P(Y, Z|X=x) \forall x \in X$ . For example, for a variable  $X$ , we calculated  $D_{KL}(Q_i || Q_j)$  where  $i \neq j$  for all  $i, j \in X$ . The KL divergence is an asymmetric measure of how similar a candidate probability distribution  $B(x)$  is to some reference distribution  $A(x)$ , and is equal to:

$$D_{KL}(A || B) = \sum_{x \in X} A(x) \cdot \log \left( \frac{A(x)}{B(x)} \right)$$

In all cases,  $D_{KL}(A||B) \geq 0$ , with it being 0 if the two distributions are identical. Effect modification refers to the scaled and inverted KL divergence calculated using the following transformation:

$$EffMod(Y, Z | X) = 1 - \exp(-\text{medhum}(\{D_{KL}(Q_i || Q_j) \forall i, j \in X | i \neq j\}))$$

A scaled KL divergence of zero for a variable  $X$  implies that the joint distribution of the outcomes was identical for different values of  $X$ , that is,  $X$  was not an effect modifier. A scaled KL divergence of 1 implies the highest effect modification observed among survey variable.

Guidelines and principles published by CHMS were used to combine survey data over multiple cycles.<sup>51</sup> Data from each cycle were treated as a completely random population sample and survey weights were excluded from all analyses. All analyses were performed on untransformed, unadjusted and unweighted data. To describe baseline characteristics, frequency distributions and proportions were reported for categorical data whereas means ( $\pm$ SD) were reported for continuous data.

#### Patient and public involvement

No patient involved.

## RESULTS

Respondents who were eligible for this study (n=10 115) had an average age of  $49.2\pm 12.5$  years and an approximately 1:1 male:female ratio (table 1). Among study participants, HCV infection was prevalent in 1% of the population. As shown in table 1, approximately 73% of the study subjects had a low 10-year CVD risk (ie, FRS <10%), 17% had a moderate risk (10% to 19%) and 11% were at high risk (FRS  $\geq 20\%$ ). HCV-positive (HCV+) cases, had a FRS of  $10.5\%\pm 8.8\%$ , which was significantly higher (p=0.008) than their HCV-negative (HCV-) counterparts ( $8.0\%\pm 6.6\%$ ). Furthermore, subjects with HCV infection had a 1.6-fold higher prevalence of high-CVD risk (ie, FRS  $\geq 20\%$ ) compared with non-infected individuals (p=0.038). Marital status, level of education, household income, use of intravenous drug and smoking status were all significantly different between HCV- and HCV+ cases (table 1). Diabetes (assessed as self-reported or with HbA1c  $\geq 6.5\%$ ) was found in 8% of the participants with a higher prevalence in the HCV+ cases. Total cholesterol levels and total cholesterol:HDL-C ratio were significantly higher in HCV+ cases whereas fasting triglycerides, LDL-C and apolipoprotein B levels were significantly lower in this group compared with their levels in non-HCV-infected individuals. Furthermore, liver enzyme markers were—as expected—all significantly increased in those who were HCV+.

The BN structures of the Markov blankets for HCV and CVD is shown in figure 1. The Markov blanket of CVD risk network structure included traditional CVD risk variables such as age, sex, obesity, diabetes, smoking status, hypertension and high levels of cholesterol (figure 1A). However, the HCV infection network structure primarily included factors such as smoking status and past illicit drug injections (figure 1B). Full BN structure for the entire set of variables examined here is provided in online supplementary figure S1. It clearly demonstrates diabetes as a primary risk modifier that influences the interaction between HCV infection and CVD risk.

As shown in figure 2, age and blood pressure were the most predictive variables for CVD risk (figure 2A) whereas injection drug use, alcohol use, smoking status and a number of diet-related factors were most predictive of HCV infection (figure 2B). The top 30 survey variables that were used to generate figure 2 to predict CVD risk in HCV infection are provided in online supplementary table S2. The conditional probabilities of CVD risk in the presence and absence of HCV infection, as estimated from the BN model, are shown in table 2. The presence of HCV infection resulted in an increased probability of CVD risk. The probabilities of medium ( $16.4\%\pm 0.28\%$  vs  $21.5\pm 1.5\%$ ) and high ( $8.6\%\pm 0.16\%$  vs  $13.9\pm 1.55\%$ ) risk of CVD were significantly elevated in individuals infected with HCV. In contrast, the probability of low-CVD risk was decreased ( $75.0\%\pm 0.29\%$  vs  $64.2\pm 1.78\%$ ) in the presence of HCV infection. Since age was a significant variable that had an effect modifier on the joint probability distribution of HCV and CVD, we estimated the variation

in the conditional probabilities through the different age intervals (table 3). The conditional probability of being infected with HCV was increased from  $0.66\pm 0.11$  at age group 30 to 40 years to  $0.87\pm 0.21$  at  $>70$  whereas that of CVD was increased from  $0.07\pm 0.05$  to  $33.8\pm 1.37$  within the same age groups.

The extent to which the joint distribution of HCV infection and CVD risk varies in the presence of a given effect modifier was quantified using the KL divergence ( $D_{KL}$ ). Figure 3 shows the full set of modifiers of the joint probability distributions between HCV infection and CVD risk. The insert in figure 3 highlights the top factors modifying the HCV-CVD association. Diabetes (any type) had the largest divergence, that is, the top effect modifier of the pairwise joint probability distribution between HCV infection and CVD risk (median scaled  $D_{KL}$  of 0.27; IQR, 0.26 - 0.27). This was followed by hypertension ( $D_{KL}$ =0.24; IQR, 0.23 - 0.25), age ( $D_{KL}$ =0.21; IQR, 0.10 - 0.38) and injection drug use ( $D_{KL}$ =0.19; IQR, 0.06 - 0.59). Median scaled  $D_{KL}$  values ranging from 0.03 to 0.06 were observed for other factors such as HbA1C, smoking, history of cancer and metabolic syndrome.

## DISCUSSION

The present study applies a BN analysis and machine learning techniques to explore the probabilistic relationship between HCV infection and CVD risk in Canadian adults. HCV infection was associated with an increased probability of CVD risk of 5.3% (table 2). Diabetes was the most prominent effect modifier of this relationship (figure 3) in our analysis, suggesting the disruption of pathways involved in glucose metabolism, insulin resistance and liver function as the leading factors in CVD comorbidities related to HCV infection.

BN analysis was both informative (figure 1A) and predictive (figure 2A) in implicating a number of conventional risk factors, for example, age, sex, obesity, diabetes, smoking status, hypertension and serum lipid profiles<sup>52</sup> in CVD risk. In addition, lifestyle factors, for example, illicit drug use, smoking and alcohol consumption, were predictive of CVD risk in HCV infection (figures 1B and 2B). The findings demonstrating that injecting drug use is highly associated with HCV infection are well supported by previous studies showing that injection drug use may result in more cases of HCV than HIV.<sup>53</sup> Furthermore, a recent study using data from the National Health and Nutrition Examination Survey cohort reported that, among smokers, those with HCV infection are more likely to be daily smokers and had smoked for longer.<sup>54</sup> Additionally, alcohol abuse was found to hinder the processes of hepatic spontaneous clearance of HCV.<sup>55</sup> Overall, conventional risk factors for CVD and HCV infection were highlighted by the Markov blankets and BN models demonstrating the abilities of the graphical elements in BN—irrespective of expert opinion—to generate and build convenient models structuring and

**Table 1** Sociodemographic and clinical characteristics in the study populations stratified by the hepatitis C virus infection

Characteristic	All*	HCV negative	HCV positive	P value
	(n=10 115)	(n=10 028)	(n=87)	
Males	47%†	47%	53%	
Age (years)	49.2±12.5	49±13	51±10	0.0333‡
Ethnicity				
White	80%	80%	80%	0.0023‡
Black	2%	2%	2%	
Asian	11%	11%	3%	
Other	6%	6%	14%	
Marital status				
Married	70%	71%	34%	<0.0013‡
Divorced	17%	17%	28%	
Single	12%	12%	38%	
Highest level of education				
Less than grade 12	11%	11%	23%	<0.0013‡
High-school graduate	20%	19%	32%	
Post-secondary graduate	69%	68%	45%	
Household income				
< \$20 000 per annum	8%	7%	41%	<0.0013‡
History of intravenous drug use	1%	1%	53%	<0.0013‡
Smoking status				
Daily/occasional	19%	18%	57%	<0.0013‡
Non-smoker	81%	82%	30%	
Diabetes (self-reported or HbA1c ≥6.5%)¶	8%	8%	13%	
Cardiometabolic markers				
Systolic blood pressure (mm Hg)	114±16	114±15	115±21	
Diastolic blood pressure (mm Hg)	73±9	73±9	75±10	0.0314§
Triglycerides, fasting (mmol/L)	1.43±0.89	1.43±0.89	1.15±0.52	0.0474§
Total cholesterol (mmol/L)	5.06±1.02	3.94±1.27	4.50±0.97	<0.0014§
LDL-C, fasting (mmol/L)	3.02±0.92	3.03±0.92	2.48±0.89	<0.0014§
HDL-C (mmol/L)	1.38±0.40	1.38±0.40	1.42±0.45	
Total cholesterol:HDL-C ratio	3.93±1.27	3.94±1.27	3.37±1.01	<0.0014§
Insulin, fasting (pmol/L)	77±67	77±67	94±62	
Glucose, fasting (mmol/L)	5.3±1.27	5.3±1.26	5.63±2.23	
HOMA-IR	3.5±3.9	3.5±3.9	4.4±3.5	
HbA1c (%)	5.61±0.77	5.61±0.77	5.67±0.88	
Apolipoprotein A1, fasting (g/L)	1.46±0.27	1.46±0.27	1.49±0.28	
Apolipoprotein B, fasting (g/L)	0.97±0.26	0.97±0.26	0.75±0.20	<0.0014§
Obesity				
Body mass index (kg/m <sup>2</sup> )	27.8±5.7	27.9±5.7	26.9±5.7	
Waist circumference (cm)	94.6±15.3	94.6±15.3	94.5±15.7	
Waist-to-hip ratio	0.90±0.11	0.90±0.10	0.93±0.1	0.0044§
Inflammatory markers				
C-reactive protein (mg/L)	2.5±2.9	2.5±2.9	2.5±3.2	
Fibrinogen (mmol/L)	3.1±0.6	3.1±0.6	2.9±0.6	0.0424§
Homocysteine (umol/L)	8.2±3.0	8.2±3.0	9.0±3.3	

Continued

Table 1 Continued

Characteristic	All* (n=10 115)	HCV negative (n=10 028)	HCV positive (n=87)	P value
<b>Micronutrients</b>				
Vitamin B12 (pmol/L)	340±193	340±193	362±145	
Vitamin D (nmol/L)	64±26	64±26	64±27	
<b>Liver enzyme markers</b>				
Alanine aminotransferase (U/L)	34±17	33±16	63±54	<0.0014§
Alkaline phosphatase (U/L)	78±24	78±24	90±36	0.0044§
Aspartate aminotransferase (U/L)	28±13	28±12	55±47	<0.0014§
γ-Glutamyl transferase (U/L)	32±40	32±37	92±176	0.0024§
Lactate dehydrogenase (U/L)	381±86	381±86	427±67	0.0284§
<b>Framingham risk score (FRS, %)</b>				
Low (FRS <10%)	73%	73%	62%	
Medium (FRS=10% to 19%)	17%	17%	20%	
High (FRS >20%)	11%	11%	18%	0.0383‡

\*Numbers represent percentage of the corresponding group or mean±SD.

†Percentages are for unweighted frequency.

‡ $\chi^2$  test for the difference between hepatitis C virus-positive (HCV+) and hepatitis C virus-negative (HCV-) cases.

§t-test for the difference between hepatitis C virus-positive (HCV+) and hepatitis C virus-negative (HCV-) case. Only significant differences are shown.

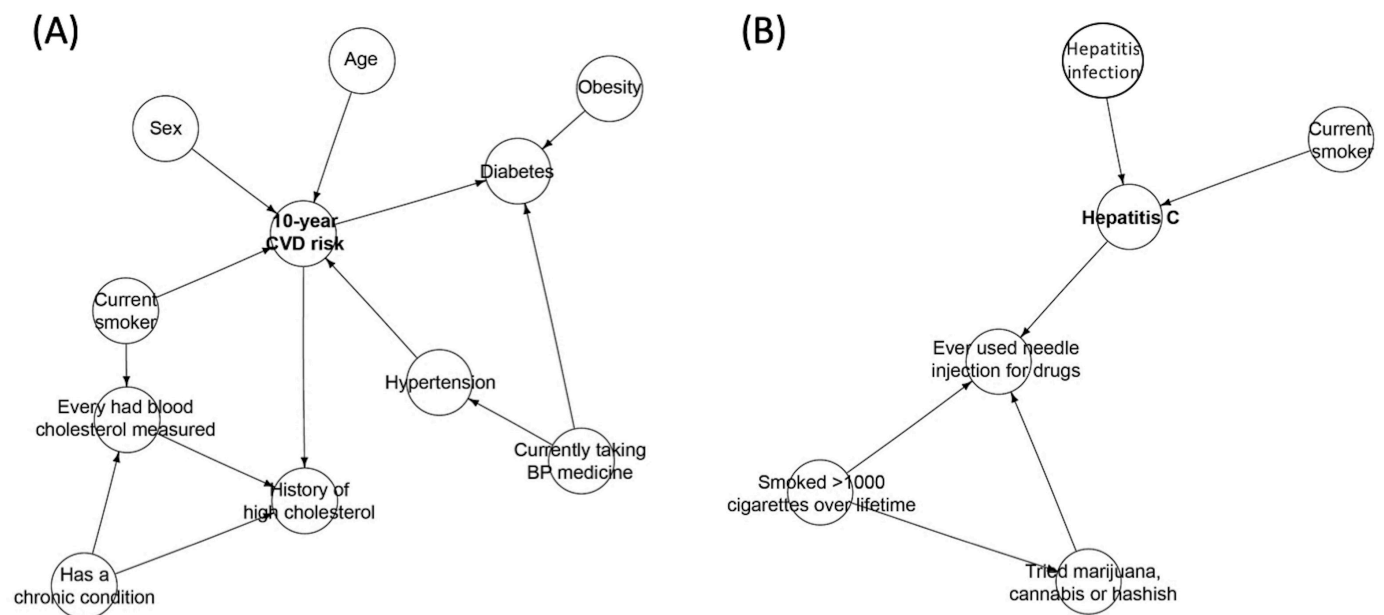
¶Diabetes defined as Hb1Ac ≥6.5% or self-reported cases as per survey questionnaire.

HbA1c, glycosylated haemoglobin; HCV, hepatitis C virus; HDL-C, high-density lipoprotein cholesterol; HOMA-IR, homoeostatic model assessment insulin resistance; LDL-C, low-density lipoprotein cholesterol.

detecting probabilistic relationships between the examined variables.

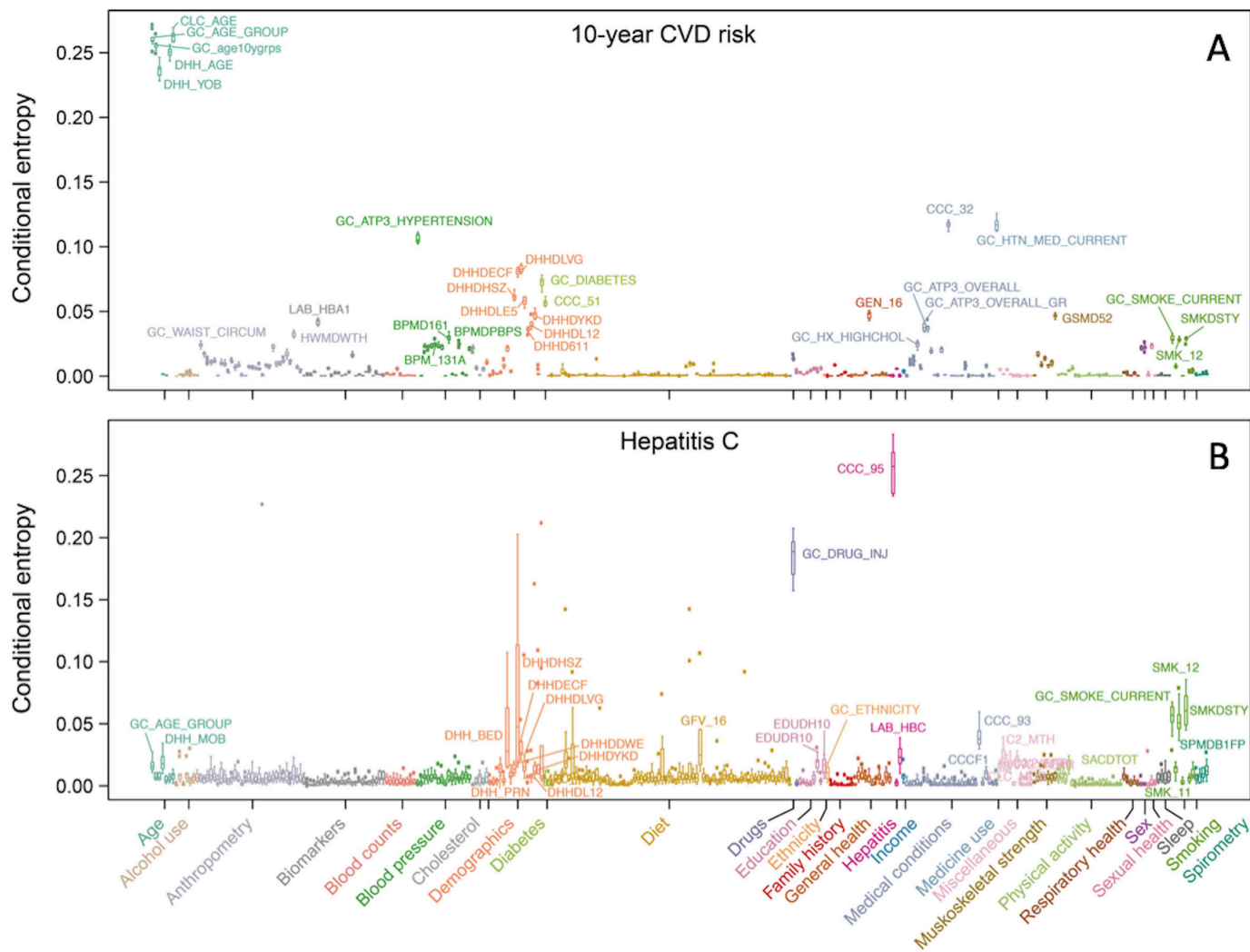
In line with observations from a number of earlier reports,<sup>4 17 56</sup> we noted a significant increase in the conditional probability of CVD risk (from 8.6% to 13.9%) in

presence of HCV infection (table 2). As we reported previously,<sup>17</sup> a number of hepatic and extrahepatic processes related to HCV infection may explain the observed increased CVD risk. For example, the development of hepatic steatosis and the later manifestation of



**Figure 1** Markov blankets of the Bayesian network structures learnt in Canadian adults. Markov blanket of hepatitis C virus infection (A) and Markov blanket for cardiovascular disease (B) in the Bayesian network. BP, blood pressure; CVD, cardiovascular disease.





**Figure 2** Outcome prediction of 10-year cardiovascular disease risk and hepatitis C infection in Canadian adults. BMI, body mass index; BP, blood pressure; BPM\_131A, diastolic blood pressure - first set (3); BPM161, average systolic blood pressure (mm Hg) first; BPMDBP5, final previous average systolic blood pressure; C2\_YEAR, year of household interview; CCC\_32, past medication for high blood pressure; CCC\_51, has diabetes; CCC\_93, has liver disease or gallbladder problem; CCC\_95, has hepatitis; CCCF1, has a chronic condition; CLC\_AGE, age at clinic visit; CLC\_MOB, month of birth; CVD, cardiovascular disease; DHH\_AGE, age at household interview; DHH\_BED, dwelling - number of bedrooms; DHH\_MOB, month of birth; DHH\_PRN, province of residence; DHH\_YOB, year of birth; DHHDL11, number of persons 6 to 11 years in household; DHHDDWE, type of dwelling; DHHDECF, household type; DHHDHSZ, household size; DHHDL12, number of persons <12 years old in household; DHHDL5, number of persons ≤5 years old in household; DHHDLVG, living arrangements for selected respondents; DHHDYKD, number of persons ≤15 years old in household; EDUDH10, highest level of education; FMH\_19, immediate family had diabetes; GC\_AGE\_GROUP, age groups (10 year age groupings); GC\_age10ygrps, 10 year age groups 30 to 74 years; GC\_ATP3\_HYPERTENSION, high blood pressure (BP ≥130/85 or HTN medication); GC\_ATP3\_OVERALL, number of ATPIII syndromes; GC\_ATP3\_OVERALL\_GR, ATP-III group (groups of 0, 1, 2 and 3+); GC\_BMI, body mass index; GC\_BMI\_OBESE, obesity (BMI ≥30 kg/m<sup>2</sup>); GC\_CVDRISKGRP\_P, 10-year CVD risk group (Framingham risk score); GC\_DIABETES, self-reported diabetes (any type) based on fasting glucose; GC\_DRUG\_INJ, ever used needle injections for drugs; GC\_ETHNICITY, ethnicity; GC\_HTN\_MED\_CURRENT, current medication for high blood pressure (past month); GC\_HX\_HIGHCHOL, history of high cholesterol; GC\_SMOKE\_CURRENT, current smoker; GC\_WAIST\_CIRCUM, waist circumference (cm); GEN\_16, employment; GFV\_16, number of eats of instant/seasoned/wild rice; GSMD52, grip strength norms for respondents; HTN, hypertension; HWM\_15 CM, hip circumference (cm); HWMD14IN, waist circumference (in); HWMD15IN, hip circumference (in); HWMDBMI, body mass index; HWMDWSTA, waist circumference norm for respondents 15 to 69 years; HWMDWTH, waist-to-hip ratio; HWTDBMI, body mass index, self-reported; LAB\_HBA1, hepatitis B virus surface antigen; LAB\_HBC, hepatitis B virus core antibody; SACDTOT, total hours of sedentary activity over the last 3 months; SMK\_11, smoked 100 or more cigarettes in lifetime; SMK\_12, type of smoker; SMKDSTY, type of smoker in dwelling; SPMDB1FP, predicted body mass index; V2\_YEAR, year of clinic visit.



**Table 2** Conditional probabilities (%) of low, medium and high 10-year CVD risk in the presence and absence of hepatitis C virus infection in adult Canadians

Hepatitis C Virus Infection	Conditional Probability (%)		
	CVD risk Low (FRS<10%) (72.7%)*	Medium (FRS 10%–19%) (16.6%)	High (FRS≥20%) (10.7%)
Absent (99.1%)	75.0±0.29	16.4±0.28	8.6±0.16
Present (0.9%)	64.2±1.78	21.5±1.50	13.9±1.55
P	2.2×10 <sup>-16</sup>	2.2×10 <sup>-16</sup>	2.2×10 <sup>-16</sup>

\*Percentage in parenthesis represents the estimated conditional probability in the different states when the BN has been compiled without any variable instantiated.

BN, Bayesian network; CVD, cardiovascular disease; FRS, Framingham Risk Score.

insulin sensitivity, metabolic abnormalities and visceral obesity<sup>19 57</sup> together with altered LDL-C,<sup>58 59</sup> triglycerides and ApoB<sup>60</sup> were all noted at progressive stages of chronic HCV infection and deteriorating liver functions. This altered profile of serum lipids indicates that HCV infection may result in the imbalance of serum lipoproteins/apolipoproteins, a known risk factor in atherosclerosis and the later development of CVD.<sup>61</sup> Additional evidence for the role of viraemia on cardiometabolic risk markers and CVD risk is that HCV cases positive for the viral DNA had lower total cholesterol, LDL-C and total cholesterol:HDL-C ratio than the RNA-negative HCV infected cases.<sup>17</sup> Another explanation for the role of HCV infection in CVD risk might be due to the alteration in the levels of proinflammatory cytokines and the downstream acute phase reactants (APR), for example, fibrinogen, in chronic infection as liver cirrhosis worsened.<sup>17 62</sup> It is possible, therefore, that the severity of cirrhosis and alteration in the haemostatic regulation of the fibrinolytic system (that is linked to control of inflammation) are interrelated to play a role in CVD risk.<sup>8</sup> As such, activation of inflammatory processes, disruption of cytokine

**Table 3** Variation in the conditional probabilities for hepatitis C virus infection and cardiovascular disease risk at different age intervals in Canadian adults

Age groups	Conditional probability (%)	
	Probability of having HCV	Probability of high CVD risk
30–40	0.66±0.11	0.07±0.05
40–50	0.65±0.11	1.12±0.14
50–60	0.74±0.14	8.18±0.54
60–70	0.80±0.15	22.7±0.57
>70	0.87±0.21	33.8±1.37

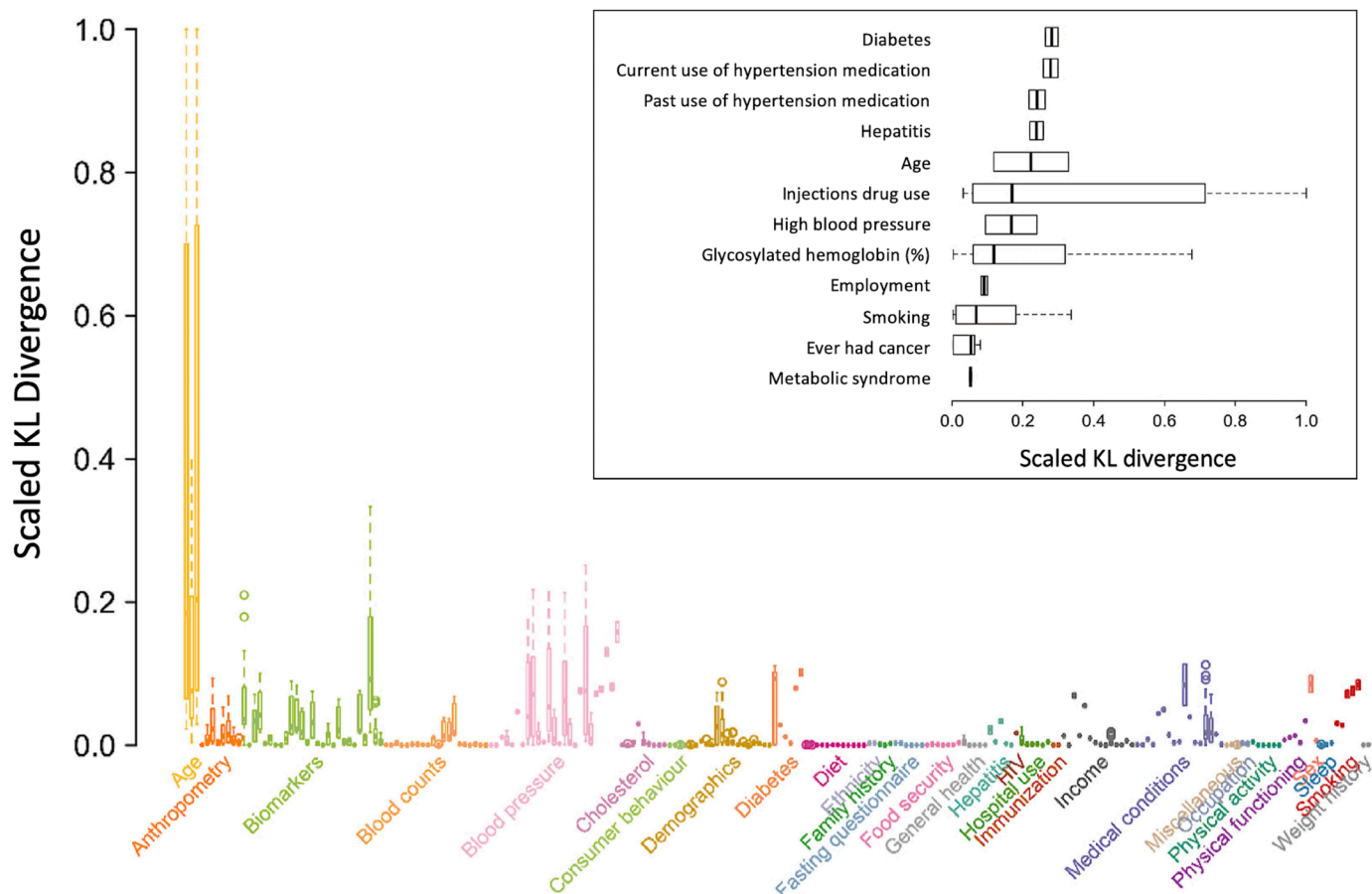
CVD, cardiovascular disease; HCV, hepatitis C virus.

synthesis and imbalance in APR homeostasis all are well-known risk factors in the increased CVD risk.<sup>63–65</sup>

Diabetes status (any) was the largest effect modifier of the joint probability distribution between HCV infection and CVD risk with 0.274 median  $D_{KL}$ . Chronic HCV impairs hepatocyte insulin signalling pathways leading to insulin resistance and the subsequent risk of diabetes via increasing the synthesis of tumour necrosis factor alpha, mediating the phosphorylation of insulin receptors and causing an over expression of suppressor cytokines.<sup>66 67</sup> Several previous reports have demonstrated that patients with HCV infection exhibit elevated HOMA-IR,<sup>68</sup> higher burden of cardiometabolic comorbidities<sup>68</sup> and increased diabetes-related mortality rates<sup>68</sup> compared to healthy controls. Our results confirm previous reports demonstrating lower levels ApoB<sup>60</sup> in chronic HCV infection. This lower level of lipoprotein, together with the lower levels of triglycerides and LDL-C (table 2) reflect a disrupted pattern of serum lipids and can lead to the deterioration of liver functions during HCV infection. It may also indicate that the virus itself can be involved in the imbalance of serum lipoproteins/apolipoproteins.<sup>60</sup> Alterations in hepatic lipid metabolism is known to be responsible for the acceleration of atherogenesis, a significant risk factor in the development of cardiovascular diseases.<sup>69</sup>

Given the role of insulin resistance in diabetes and the subsequent risk of CVD, it is reasonable to suggest diabetes as a modifier for the HCV-associated CVD risk through extrahepatic insulin resistance related pathways (online supplementary table S2).<sup>69</sup> In support to this assumption, an increased risk of diabetes was observed in the HCV-infected subjects examined here (data not shown). Although not well-defined in some studies,<sup>70</sup> several reports and meta-analyses revealed an association between HCV infection and diabetes risk.<sup>71–75</sup> Furthermore, liver cirrhosis, a common manifestation of HCV infection<sup>2 3</sup> is a known risk factor in diabetes.<sup>76–78</sup> The results of our machine-learned BNs have identified age, ethnicity and gender as risk modifiers of the joint probability distribution between HCV infection and CVD risk. These factors are known to increase the risk of diabetes in HCV cases at high risk of cirrhosis due to concurrent non-alcoholic fatty liver.<sup>78</sup> As expected, the conditional probability of having CVD has been increased with age together with that of being infected with HCV (although to a much lesser extent, table 3). Such differences in the prevalence of HCV and CVD risk in relation to age were previously noted.<sup>17</sup>

Modelling the statistical relationships between covariates without the assumption of independence of variables that is involved in linear regression approaches is a key strength of applying a machine-learned BNs.<sup>22 23</sup> Linking multiple covariates and outcomes into a single transparent modelling assumption, generating an intuitive graphical representation of a complex system and the ability to handle missing data and multicollinearity has led to the increasing use of BNs in medicine and healthcare systems.<sup>25–27</sup> Our introduction of BN analysis into a



**Figure 3** Median Kullback-Leibler (KL) divergence effect modification on the joint probability distributional relationship of hepatitis C infection and cardiovascular disease risk. The survey variables with the largest effect modification for Markov blanket of hepatitis C virus infection and Markov blanket for cardiovascular disease in the Bayesian network. The insert highlights the top factors modifying the association.

nationally-representative and population-based surveys has provided—without prior assumptions—a number of sociodemographic and biological variables that may influence the effect of HCV infection of CVD risk. The significance of this approach can be highlighted by the fact that several of the variants characterised here were previously recognised as risk factors for HCV and CVD risk. Additionally, novel aetiologies for the HCV-related CVD risk can be substantiated from a set of risk modifiers that were identified in the present study, such as diabetes, hypertension, age, injection drug use, smoking status, levels of glycosylated haemoglobin, history of cancer and metabolic syndrome.

The present study has several limitations. The status of HIV was not captured in the study participants as this factor was self-reported and deemed unreliable. We realise that coinfection of HCV and HIV can have a synergistic effect on increasing the risk of CVD as reported elsewhere.<sup>79</sup> Furthermore, detailed clinical features of participants, such as the degree of liver fibrosis, cirrhosis or non-alcoholic fatty liver disease were not evaluated in the study population. These factors would have been valuable for classifying the extent of liver functions and damage and could have further emphasised on other modifiers

for the relationship between HCV infection and CVD risk. For example, information on fatty liver would have added more weight to the role of diabetes in modifying the risk of CVD in HCV infection as it is a known risk factor for both diabetes and CVD.<sup>78 80</sup> Furthermore, only approximately 25% of the anti-HCV positive patients had been tested for HCV-RNA and only ~60% of the tested patients were HCV-RNA positive. Although previously reported,<sup>17</sup> data on genotypes and sub-genotypes were not considered in this study. Additionally, the study did not consider the use of direct-acting antiviral (DAA) therapy. Viral eradication by DAA therapy has been associated with decreased risk of CVD events among cirrhotic HCV individuals.<sup>81 82</sup> Also, DAA treatment in diabetic patients has resulted in a significant improvement of the related CVD events.<sup>83</sup> Information on the treatment for HCV infection would have shed insight into the sustained viraemia of the HCV-infected subjects assessed here and may have provided a detailed clinical response to the infection and the affected metabolic pathways. These limitations warrant further studies to consider the status of HIV, a detailed liver function profile and the use of DAA into understanding the relationship between HCV infection and CVD risk.

In conclusion, we have introduced a novel approach—applying BN modelling and machine learning algorithms—to investigate the relationship between HCV infection and CVD risk in a population-based survey data set. Our results provide evidence that increased CVD risk is likely to arise from HCV infection. Several factors, to varying degrees, were identified to play a role in the increased HCV-related CVD risk including diabetes, hypertension, age, injection drug use, smoking, levels of glycosylated haemoglobin, history of cancer and metabolic syndrome. It may be reasonable, therefore, to suggest that if these factors are targeted during the course of HCV infection, the risk of CVD could be reduced. However, before any such interventions are proposed, further research needs to be conducted to study the direct effect of DAA therapy on HCV infected individuals with and without chronic sequelae of infection to enable a reliable evaluation for the CVD risk in this population.

#### Author affiliations

<sup>1</sup>Public Health Risk Sciences Division, Public Health Agency of Canada, Toronto, Ontario, Canada

<sup>2</sup>Department of Nutritional Sciences, Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada

<sup>3</sup>Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada

<sup>4</sup>Pediatric Oncology Group of Ontario, Toronto, Ontario, Canada

<sup>5</sup>Lighthouse Outcomes, Toronto, Ontario, Canada

<sup>6</sup>Department of Cancer Epidemiology and Prevention Research, Alberta Health Services, Calgary, Alberta, Canada

**Acknowledgements** We thank Statistics Canada and the Canadian Institute for Health Information (CIHI) for granting the permission to access the Canadian Health Measures Survey data.

**Contributors** AB conceived the study idea and design and contributed to manuscript writing. GDG contributed to data analysis. AG performed the data analysis and interpretation and contributed to manuscript writing. AP assisted in data analysis and manuscript writing. PA contributed to results interpretation and manuscript writing. All authors critically reviewed the manuscript, approved the final draft and agreed to be accountable for all aspects of the work.

**Funding** This work was supported by the Public Health Agency of Canada (AB).

**Competing interests** None declared.

**Patient consent for publication** Not required.

**Ethics approval** The study was reviewed and approved by the Health Canada Research Ethics Board.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** All data relevant to the study are included in the article or uploaded as supplementary information. To obtain the Canadian Health Measures Survey (CHMS) data, the Principal Investigator (AB) had to sign a Data Use Agreement (DUA) to be able to access data files containing protected information. This DUA is signed between the Public Health Agency of Canada (PHAC) and the Data Providers (Statistics Canada and the Canadian Institute of Health Information) and is disseminated by the Data Coordination and Access Program (DCAP). This DUA obligates the data user to protect the confidentiality of the information, not to disclose it or distribute it improperly and to comply with all relevant laws, federal government policies and practices. Any information received under this DUA is to be treated as 'Protected Data'. This includes, among others, personal information (ie, as defined by the Privacy Act). Furthermore, according to the DUA, the Data User is required to comply with the Agency's specific contractual obligations, in particular those related to access and confidentiality. Also, the DUA states that 'Protected Data' may not be shared with individuals or organisations within or outside the signed Agency. Individual permissions to access the CHMS data may, however, be requested from Statistics Canada (<https://www.statcan.gc.ca>) or the Canadian Institute of Health Information (<https://www.cihi.ca>) through DCAP (contact information, email: [ac.cg.cpsa-cahp@dcap-pacd](mailto:ac.cg.cpsa-cahp@dcap-pacd)).

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

#### ORCID iDs

Alaa Badawi <http://orcid.org/0000-0002-9115-0025>

Giancarlo Di Giuseppe <http://orcid.org/0000-0001-5112-7560>

#### REFERENCES

- 1 Thrift AP, El-Serag HB, Kanwal F. Global epidemiology and burden of HCV infection and HCV-related disease. *Nat Rev Gastroenterol Hepatol* 2017;14:122–32.
- 2 Chen SL, Morgan TR. The natural history of hepatitis C virus (HCV) infection. *Int J Med Sci* 2006;3:47–52.
- 3 Stanaway JD, Flaxman AD, Naghavi M, et al. The global burden of viral hepatitis from 1990 to 2013: findings from the global burden of disease study 2013. *The Lancet* 2016;388:1081–8.
- 4 Petta S, Maida M, Macaluso FS, et al. Hepatitis C Virus Infection Is Associated With Increased Cardiovascular Mortality: A Meta-Analysis of Observational Studies. *Gastroenterology* 2016;150:145–55.
- 5 Ly KN, Hughes EM, Jiles RB, et al. Rising mortality associated with hepatitis C virus in the United States, 2003–2013. *Clin Infect Dis* 2016;62:1287–8.
- 6 Cacoub P, Gragnani L, Comarmond C, et al. Extrahepatic manifestations of chronic hepatitis C virus infection. *Dig Liver Dis* 2014;46 Suppl 5:S165–73.
- 7 White DL, Ratziu V, El-Serag HB. Hepatitis C infection and risk of diabetes: a systematic review and meta-analysis. *J Hepatol* 2008;49:831–44.
- 8 Babiker A, Jeudy J, Kligerman S, et al. Risk of cardiovascular disease due to chronic hepatitis C infection: a review. *J Clin Transl Hepatol* 2017;5:1–20.
- 9 Khattab MA, Eslam M, Alavian SM. Hepatitis C virus as a multifaceted disease: a simple and updated approach for extrahepatic manifestations of hepatitis C virus infection. *Hepat Mon* 2010;10:258–69.
- 10 Pothineni NV, Rochlani Y, Vallurupalli S, et al. Comparison of angiographic burden of coronary artery disease in patients with versus without hepatitis C infection. *Am J Cardiol* 2015;116:1041–4.
- 11 Wong RJ, Kanwal F, Younossi ZM, et al. Hepatitis C virus infection and coronary artery disease risk: a systematic review of the literature. *Dig Dis Sci* 2014;59:1586–93.
- 12 Kristiansen MG, Løchen M-L, Gutteberg TJ, et al. Total and cause-specific mortality rates in a prospective study of community-acquired hepatitis C virus infection in northern Norway. *J Viral Hepat* 2011;18:237–44.
- 13 Lee M-H, Yang H-I, Wang C-H, et al. Hepatitis C virus infection and increased risk of cerebrovascular disease. *Stroke* 2010;41:2894–900.
- 14 Vajdic CM, Marashi Pour S, Pour SM, et al. The impact of blood-borne viruses on cause-specific mortality among opioid dependent people: an Australian population-based cohort study. *Drug Alcohol Depend* 2015;152:264–71.
- 15 Lee M-H, Yang H-I, Lu S-N, et al. Chronic hepatitis C virus infection increases mortality from hepatic and extrahepatic diseases: a community-based long-term prospective study. *J Infect Dis* 2012;206:469–77.
- 16 Negro F. Facts and fictions of HCV and comorbidities: steatosis, diabetes mellitus, and cardiovascular diseases. *J Hepatol* 2014;61:S69–78.
- 17 Badawi A, Di Giuseppe G, Arora P. Cardiovascular disease risk in patients with hepatitis C infection: results from two general population health surveys in Canada and the United States (2007–2017). *PLoS One* 2018;13:e0208839.
- 18 Oliveira CPMS, Kappel CR, Siqueira ER, et al. Effects of hepatitis C virus on cardiovascular risk in infected patients: a comparative study. *Int J Cardiol* 2013;164:221–6.
- 19 Adinolfi LE, Restivo L, Zampino R, et al. Chronic HCV infection is a risk of atherosclerosis. Role of HCV and HCV-related steatosis. *Atherosclerosis* 2012;221:496–502.
- 20 Petta S, Torres D, Fazio G, et al. Carotid atherosclerosis and chronic hepatitis C: a prospective study of risk associations. *Hepatology* 2012;55:1317–23.
- 21 Dolley S. Big data's role in precision public health. *Front Public Health* 2018;6.



- 22 Scutari M, Vitolo C T. A learning Bayesian networks from big data with greedy search: computational complexity and efficient implementation. *Stat Comp* 2019.
- 23 Pearl J. Probabilistic reasoning in intelligent systems. In: *Networks of plausible inference*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc, 1988.
- 24 Loghmanpour NA, Kanwar MK, Druzdel MJ, et al. A new Bayesian network-based risk stratification model for prediction of short-term and long-term LVAD mortality. *Asaio J* 2015;61:313–23.
- 25 Fuster-Parra P, Tauler P, Bennasar-Veny M, et al. Bayesian network modeling: a case study of an epidemiologic system analysis of cardiovascular risk. *Comput Methods Programs Biomed* 2016;126:128–42.
- 26 Seixas FL, Zadrozny B, Laks J, et al. A Bayesian network decision model for supporting the diagnosis of dementia, Alzheimer's disease and mild cognitive impairment. *Comput Biol Med* 2014;51:140–58.
- 27 Zeng Z, Jiang X, Neapolitan R. Discovering causal interactions using Bayesian network scoring and information gain. *BMC Bioinformatics* 2016;17:221.
- 28 Whitney P, White A, Walsh S, et al. Bayesian Networks for Social Modeling. In: Salerno J, Yang SJ, Nau D, et al, eds. *Social computing, behavioral-cultural modeling and prediction*. SBP 2011. *Lecture notes in computer science*. Berlin, Heidelberg: Springer, 2011: 6589. 227–35.
- 29 Ebana Y, Furukawa T. Networking analysis on superior vena cava arrhythmogenicity in atrial fibrillation. *Int J Cardiol Heart Vasc* 2019;22:150–3.
- 30 Tylman W, Waszyrowski T, Napieralski A, et al. Real-Time prediction of acute cardiovascular events using hardware-implemented Bayesian networks. *Comput Biol Med* 2016;69:245–53.
- 31 Orphanou K, Stassopoulou A, Keravnou E, et al. Risk assessment for primary coronary heart disease event using dynamic Bayesian networks. In: Holmes J, Bellazzi R, Sacchi L, et al, eds. *Artificial intelligence in medicine AIME lecture notes in computer science 2016*. Springer Cham, 2015: 9105. 161–5.
- 32 Elsayad A, Fakhr M. Diagnosis of cardiovascular diseases with Bayesian classifiers. *J Comp Sci* 2015;11:274–82.
- 33 Gomathi K, Priyaa DS. An efficient coronary heart disease prediction by semi parametric extended dynamic Bayesian network with optimized cut points. *ARPN J Engin Appl Sci* 2018;13:1539–44.
- 34 Tremblay MS, Connor Gorber S, Gorber SC. Canadian health measures survey: brief overview. *Can J Public Health* 2007;98:453–6.
- 35 Tremblay M, Wolfson M, Connor Gorber S. Canadian health measures survey: rationale, background and overview. *Health Rep* 2007;18 Suppl:7–20.
- 36 Giroux S. Canadian health measures survey: sampling strategy overview. *Health Rep* 2007;18 Suppl:31–6.
- 37 Day B, Langlois R, Tremblay M, et al. Canadian health measures survey: ethical, legal and social issues. *Health Rep* 2007;18 Suppl:37–51.
- 38 D'Agostino RB, Vasan RS, Pencina MJ, et al. General cardiovascular risk profile for use in primary care: the Framingham heart study. *Circulation* 2008;117:743–53.
- 39 Bosomworth NJ. Practical use of the Framingham risk score in primary prevention: Canadian perspective. *Can Fam Physician* 2011;57:417–23.
- 40 Da Costa LA, Arora P, Garcia-Bailo B, et al. The association between obesity, cardiometabolic disease biomarkers, and innate immunity-related inflammation in Canadian adults. *Diabetes Metab Syndr Obes* 2012;5:347–55.
- 41 American Diabetes Association. Diagnosis and classification of diabetes mellitus. *Diabetes Care* 2010;33 Suppl 1:S62–9.
- 42 Grundy SM, Cleeman Jr, Daniels SR, et al. Diagnosis and management of the metabolic syndrome: an American heart Association/National heart, lung, and blood Institute scientific statement. *Circulation* 2005;112:2735–52.
- 43 Brenner DR, Arora P, Garcia-Bailo B, et al. Plasma vitamin D levels and risk of metabolic syndrome in Canadians. *Clin Invest Med* 2011;34:377–84.
- 44 Setayeshgar S, Whiting SJ, Vatanparast H. Prevalence of 10-year risk of cardiovascular diseases and associated risks in Canadian adults: the contribution of cardiometabolic risk assessment introduction. *Int J Hypertens* 2013;2013:1–8.
- 45 Kjærulff UB, Madsen A, Networks B. *Influence diagrams: a guide to construction and analysis*. 2<sup>nd</sup> Edn. New York, NY: Springer-Verlag, 2013.
- 46 Koller D, Friedman N, Bach F. *Probabilistic graphical models: principles and techniques - Adaptive Computation and Machine Learning*. Cambridge, MA: The MIT press, 2009.
- 47 pp. Ide JS, Cozman FG. Random Generation of Bayesian Networks. In: *Proceedings of the 16th Brazilian Symposium on artificial intelligence*. New York, NY: Springer-Verlag, 2002: 366–75.
- 48 Scutari M. Learning Bayesian Networks with the bnlearn R Package. *J Stat Softw* 2010;35.
- 49 Tsamardinos I, Brown LE, Aliferis CF. The max-min hill-climbing Bayesian network structure learning algorithm. *Mach Learn* 2006;65:31–78.
- 50 Kullback S, Leibler RA. On information and sufficiency. *Ann. Math. Statist.* 1951;22:79–86.
- 51 Statistics C. Instructions for combining multiple cycles of Canadian health measures survey (CHMS) data, 2017. Available: <https://www.statcan.gc.ca> [Accessed 8 Aug 2019].
- 52 Smith SC. Multiple risk factors for cardiovascular disease and diabetes mellitus. *Am J Med* 2007;120:S3–11.
- 53 Nelson PK, Mathers BM, Cowie B, et al. Global epidemiology of hepatitis B and hepatitis C in people who inject drugs: results of systematic reviews. *Lancet* 2011;378:571–83.
- 54 Kim RS, Weinberger AH, Chander G, et al. Cigarette smoking in persons living with hepatitis C: the National health and nutrition examination survey (NHANES), 1999–2014. *Am J Med* 2018;131:669–75.
- 55 Pott H, Bricks G, Senise JF, et al. Chronic alcohol abuse and spontaneous clearance of hepatitis C virus. *IDCases* 2019;17:e00534.
- 56 Butt AA, Yan P, Chew KW, et al. Risk of acute myocardial infarction among hepatitis C virus (HCV)-positive and HCV-negative men at various lipid levels: Results from ERCHIVES. *Clin Infect Dis* 2017;65:557–65.
- 57 Adinolfi LE, Gambardella M, Andreana A, et al. Steatosis accelerates the progression of liver damage of chronic hepatitis C patients and correlates with specific HCV genotype and visceral obesity. *Hepatology* 2001;33:1358–64.
- 58 Hofer H, Bankl HC, Wrba F, et al. Hepatocellular fat accumulation and low serum cholesterol in patients infected with HCV-3a. *Am J Gastroenterol* 2002;97:2880–5.
- 59 Nishimura M, Yamamoto H, Yoshida T, et al. Decreases in the serum VLDL-TG/non-VLDL-TG ratio from early stages of chronic hepatitis C: alterations in TG-rich lipoprotein levels. *PLoS One* 2011;6:e1730910.
- 60 Moriya K, Shintani Y, Fujie H, et al. Serum lipid profile of patients with genotype 1b hepatitis C viral infection in Japan. *Hepatol Res* 2003;25:371–6.
- 61 Malaguarnera M, Vacante M, Russo C, et al. Lipoprotein(a) in Cardiovascular Diseases. *Biomed Res Int* 2013;2013:1–9.
- 62 Shao Z, Zhao Y, Feng L, et al. Association between plasma fibrinogen levels and mortality in acute-on-chronic hepatitis B liver failure. *Dis Markers* 2015;2015:468596:1–6.
- 63 Ambrosino P, Lupoli R, Di Minno A, et al. The risk of coronary artery disease and cerebrovascular disease in patients with hepatitis C: a systematic review and meta-analysis. *Int J Cardiol* 2016;221:746–54.
- 64 Ruparelina N, Chai JT, Fisher EA, et al. Inflammatory processes in cardiovascular disease: a route to targeted therapies. *Nat Rev Cardiol* 2017;14:133–44.
- 65 Ahmed MS, Jadhav AB, Hassan A, et al. Acute phase reactants as novel predictors of cardiovascular disease. *ISRN Inflamm* 2012;2012:1–18.
- 66 Kaddai V, Negro F. Current understanding of insulin resistance in hepatitis C. *Expert Rev Gastroenterol Hepatol* 2011;5:503–16.
- 67 Hui JM, Sud A, Farrell GC, et al. Insulin resistance is associated with chronic hepatitis C virus infection and fibrosis progression [corrected]. *Gastroenterology* 2003;125:1695–704.
- 68 Lazo M, Nwankwo C, Daya NR, et al. Confluence of epidemics of hepatitis C, diabetes, obesity, and chronic kidney disease in the United States population. *Clin Gastroenterol Hepatol* 2017;15:1957–64.
- 69 Drazilova S, Gazda J, Janicko M, et al. Chronic hepatitis C association with diabetes mellitus and cardiovascular risk in the era of DAA therapy. *Can J Gastroenterol Hepatol* 2018;2018:1–11.
- 70 Stepanova M, Lam B, Younossi Y, et al. Association of hepatitis C with insulin resistance and type 2 diabetes in US general population: the impact of the epidemic of obesity. *J Viral Hepat* 2012;19:341–5.
- 71 Younossi ZM, Stepanova M, Nader F, et al. Associations of chronic hepatitis C with metabolic and cardiac outcomes. *Aliment Pharmacol Ther* 2013;37:647–52.
- 72 Fabiani S, Fallahi P, Ferrari SM, et al. Hepatitis C virus infection and development of type 2 diabetes mellitus: systematic review and meta-analysis of the literature. *Rev Endocr Metab Disord* 2018;19:405–20.



- 73 Desbois A-C, Cacoub P, mellitus D. Diabetes mellitus, insulin resistance and hepatitis C virus infection: a contemporary review. *World J Gastroenterol* 2017;23:1697–711.
- 74 García-Compeán D, González-González JA, Lavallo-González FJ, et al. Current concepts in diabetes mellitus and chronic liver disease: clinical outcomes, hepatitis C virus association, and therapy. *Dig Dis Sci* 2016;61:371–80.
- 75 Giordanino C, Ceretto S, Bo S, et al. Type 2 diabetes mellitus and chronic hepatitis C: which is worse? results of a long-term retrospective cohort study. *Dig Liver Dis* 2012;44:406–12.
- 76 García-Compeán D, Jáquez-Quintana JO, Lavallo-González FJ, et al. The prevalence and clinical characteristics of glucose metabolism disorders in patients with liver cirrhosis. A prospective study. *Ann Hepatol* 2012;11:240–8.
- 77 Garcia-Compean D, Jaquez-Quintana JO, Gonzalez-Gonzalez JA, et al. Liver cirrhosis and diabetes: risk factors, pathophysiology, clinical implications and management. *World J Gastroenterol* 2009;15:280–8.
- 78 Banks DE, Bogler Y, Bhuket T, et al. Significant disparities in risks of diabetes mellitus and metabolic syndrome among chronic hepatitis C virus patients in the U.S. *Diabetes Metab Syndr* 2017;11 Suppl 1:S153–8.
- 79 Osibogun O, Ogunmoroti O, Michos ED, et al. HIV/HCV coinfection and the risk of cardiovascular disease: a meta-analysis. *J Viral Hepat* 2017;24:998–1004.
- 80 Lonardo A, Ballestri S, Guaraldi G, et al. Fatty liver is associated with an increased risk of diabetes and cardiovascular disease - Evidence from three different disease models: NAFLD, HCV and HIV. *World J Gastroenterol* 2016;22:9674–93.
- 81 Cacoub P, Nahon P, Layese R, et al. Prognostic value of viral eradication for major adverse cardiovascular events in hepatitis C cirrhotic patients. *Am Heart J* 2018;198:4–17.
- 82 Butt AA, Yan P, Shuaib A, et al. Direct-Acting antiviral therapy for HCV infection is associated with a reduced risk of cardiovascular disease events. *Gastroenterology* 2019;156:987–96.
- 83 Hsu Y-C, Lin J-T, Ho HJ, et al. Antiviral treatment for hepatitis C virus infection is associated with improved renal and cardiovascular outcomes in diabetic patients. *Hepatology* 2014;59:1293–302.