BMJ Open

Identification of reciprocal causality between non-alcoholic fatty liver disease and metabolic syndrome by a simplified Bayesian network in a Chinese population

Yongyuan Zhang,[1,2] Tao Zhang,[1] Chengqi Zhang,[3] Fang Tang,[3] Nvjuan Zhong,[1] Hongkai Li,[1] Xinhong Song,[3] Haiyan Lin,[3] Yanxun Liu,[1] Fuzhong Xue[1]

For numbered affiliations see end of article.

Correspondence to
Professor Fuzhong Xue;
xuefzh@sdu.edu.cn

ABSTRACT

Objectives: It remains unclear whether non-alcoholic fatty liver disease (NAFLD) is a cause or a consequence of metabolic syndrome (MetS). We proposed a simplified Bayesian network (BN) and attempted to confirm their reciprocal causality.

Setting: Bidirectional longitudinal cohorts (subcohorts A and B) were designed and followed up from 2005 to 2011 based on a large-scale health check-up in a Chinese population.

Participants: Subcohort A (from NAFLD to MetS, n=8426) included the participants with or without NAFLD at baseline to follow-up the incidence of MetS, while subcohort B (from MetS to NAFLD, n=16 110) included the participants with or without MetS at baseline to follow-up the incidence of NAFLD.

Results: Incidence densities were 2.47 and 17.39 per 100 person-years in subcohorts A and B, respectively. Generalised estimating equation analyses demonstrated that NAFLD was a potential causal factor for MetS (relative risk, RR, 95% CI 5.23, 3.50 to 7.81), while MetS was also a factor for NAFLD (2.55, 2.23 to 2.92). A BN with 5 simplification strategies was used for the reciprocal causal inference. The BN's causal inference illustrated that the total effect of NAFLD on MetS (attributable risks, AR%) was 2.49%, while it was 19.92% for MetS on NAFLD. The total effect of NAFLD on MetS components was different, with dyslipidemia having the greatest (AR%, 10.15%), followed by obesity (7.63%), diabetes (3.90%) and hypertension (3.51%). Similar patterns were inferred for MetS components on NAFLD, with obesity having the greatest (16.37%) effect, followed by diabetes (10.85%), dyslipidemia (10.74%) and hypertension (7.36%). Furthermore, the most important causal pathway from NAFLD to MetS was that NAFLD led to elevated GGT, then to MetS components, while the dominant causal pathway from MetS to NAFLD began with dyslipidaemia.

Conclusions: The findings suggest a reciprocal causality between NAFLD and MetS, and the effect of MetS on NAFLD is significantly greater than that of NAFLD on MetS.

Strengths and limitations of this study

- This is the first bidirectional longitudinal study designed to verify the reciprocal causality between NAFLD and MetS in a cohort within the same population.
- Bayesian network with five simplification strategies is proposed for the reciprocal causal inference between NAFLD and MetS.
- This study indicates a reciprocal causality between NAFLD and MetS, and the effect of MetS on NAFLD is significantly greater than that of NAFLD on MetS.
- The presence of NAFLD is assessed by experienced radiologists using abdominal ultrasonography, and we have no information on the intraobserver or interobserver reliability of the ultrasonographic examinations.
- The diagnostic criteria of MetS is based on the Chinese medical association diabetes branch rather than the international standard criteria, owing to the absence of waist circumference measurement in the health check-up programme.

INTRODUCTION

Metabolic syndrome (MetS) is a constellation of metabolic and cardiovascular disease (CVD) risk factors, including obesity, hypertension, hyperglycaemia, dyslipidemia and insulin resistance.[1] Non-alcoholic fatty liver disease (NAFLD) is defined as a disorder with excess fat in the liver due to non-alcoholic causes.[2] In recent years, due to lifestyle and economic changes in Chinese populations, the prevalence of NAFLD and MetS has been rapidly increasing, and has become a major public-health challenge.[3-7] Both disorders predict type 2 diabetes, cardiovascular disease, non-alcoholic steatohepatitis (NASH) and hepatocellular carcinoma.

Insulin resistance (IR) plays a critical role in the development of both NAFLD and MetS.[8 9] Patients with MetS frequently have an increase in fat accumulation in the liver and hepatic insulin resistance. In patients with NAFLD, glucose and triglycerides are overproduced by the fatty liver due to the impaired ability of insulin. Furthermore, a growing number of epidemiological studies support an association between NAFLD and MetS.[10–21] From the conventional viewpoint, NAFLD is regarded as the hepatic manifestation of MetS. Nevertheless, a series of longitudinal studies have reported that NAFLD might be a precursor to MetS, suggesting NAFLD as a risk factor for MetS rather than merely its hepatic manifestation.[15 16 22–29] Meanwhile, other longitudinal studies have also confirmed that MetS precedes the future development of NAFLD.[29–34] Therefore, it remains unclear whether NAFLD is a cause or consequence of MetS, and a 'chicken or egg' scientific debate has arisen recently and gained intense new interest.[35 36]

Previous studies partially confirmed the complicated and bidirectional relationship between NAFLD and MetS in single-directed longitudinal cohorts, by focusing on the temporal sequence of NAFLD to MetS or MetS to NAFLDs separately. Up to now, to the best of our knowledge, there has been no bidirectional longitudinal cohort study in the same population to clarify their reciprocal relationship. In addition, the previous studies usually utilised regression models, such as the Cox and the generalised estimating equation (GEE) models,[37] to analyse the temporal association between NAFLD and MetS. The specified statistical technique for causal inference, such as the Bayesian network (BN),[38 39] has not been used to analyse their reciprocal causality.

In this study, we proposed an assumption of reciprocal causality between NAFLD and MetS. To identify this reciprocal causality, a bidirectional longitudinal cohort study (from NAFLD to MetS, and from MetS to NAFLD) was conducted based on a large-scale health check-up in an urban Han Chinese population. A BN with five simplification strategies was used for reciprocal causal inference. Additionally, the relative importance of the pathogenesis and the public health significance of a specific pathway were evaluated.

MATERIALS AND METHODS
Design of bidirectional subcohort
On the basis of the routine health check-up system at the Center for Health Management of Shandong Provincial Qianfoshan Hospital and Shandong Provincial Hospital, we set up a large-scale longitudinal cohort and conducted a follow-up from 2005 to 2011 in an urban Han Chinese population. Within this large-scale longitudinal cohort, the bidirectional longitudinal cohorts (subcohorts A and B, shown in figure 1)

were designed to identify the reciprocal causality between NAFLD and MetS.

Generally, participants who had a health check-up at least twice between 2005 and 2011 were recruited in this study, with the first health check-up data as baseline and the end of follow-up as end point. Subcohort A (n=8426) included the participants with or without NAFLD at baseline to follow-up the incidence of MetS (shown in figure 1A). The exclusion criteria were: presence of any MetS components (obesity, dyslipidemia, hyperglycaemia or hypertension) at baseline; regular alcohol intake; positive serological marker for hepatitis B surface antigen (HBsAg) or hepatitis C virus antibody (HCVAb) at baseline; and the development of MetS before the development of NAFLD during the follow-up period. The inclusion/exclusion criteria for subcohort B (n=16 110) were similar to subcohort A, except that subcohort B participants were free from NAFLD at baseline and the group excluded those with NAFLD occurring before MetS (shown in figure 1B).

Measurements
The health check-up examinations were performed after an overnight fasting period of at least 12 h, and all the participants underwent routine anthropometric, clinical and laboratory testing. The anthropometric measurements included height, weight and blood pressure.
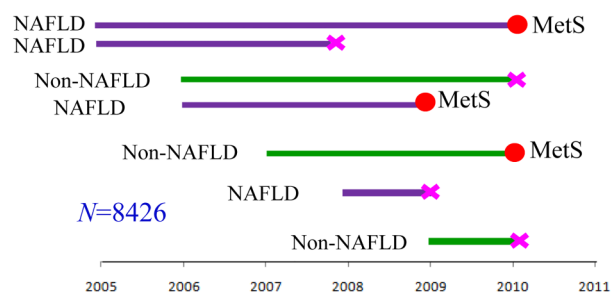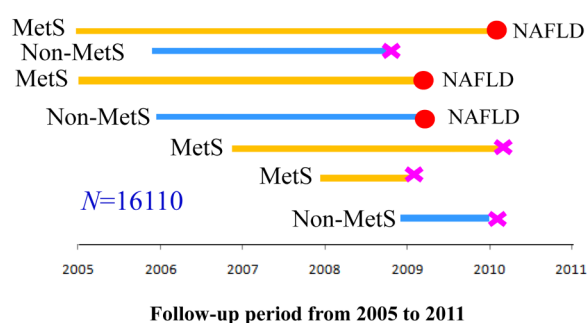

Figure 1 Diagram of bidirectional longitudinal cohorts. (A) Subcohort A (from NAFLD to MetS, n=8426) includes participants with or without NAFLD at baseline to follow-up the incidence of MetS and (B) Subcohort B (from MetS to NAFLD, n=16 110) includes participants with or without MetS at baseline to follow-up the incidence of NAFLD.

Height and weight were measured with participants wearing light clothing and no shoes. Body mass index (BMI) was calculated as weight (kg) divided by the square of height (m), and was used to estimate obesity. Blood pressure, including systolic blood pressure (SBP) and diastolic blood pressure (DBP), was measured from the right arm after 5 min of rest in a sitting position.

Blood biochemical analysis was performed using a fully automatic blood analyser (E9000, Sysmex Corporation, Japan); the abbreviations of variables and value assignments are shown in table 1. All the participants consented to and underwent an abdominal B-ultrasonography examination performed by experienced radiologists using a 3.5 MHz transducer (Logic Q700 MR, GE, Milwaukee, Wisconsin, USA). Additionally, lifestyle behaviours, including diet, smoking, alcohol intake, sleeping quality and physical activity, were surveyed by a general health questionnaire. Questions about alcohol intake included the type of alcohol consumed, the frequency of alcohol consumption per week and the usual amount per day ($\geq$20 g/day). Based on these questions, alcohol intake was coded as an ordered categorical variable as follows: 0, never; 1, seldom; 2, often, wine; 3, often, beer; 4, often, Chinese spirits; and 5, often, mixed/all types. Persons with a value >1 were considered regular alcohol users.

Definitions of NAFLD and MetS

According to the revised definition and treatment guidelines laid down by the Chinese Hepatology Association in February 2006,[40] NAFLD was diagnosed by abdominal ultrasonography based on evidence of liver brightness and a diffusely echogenic change in the liver parenchyma, with exclusion of participants who had a prior diagnosis of NAFLD, hepatitis virus infection (HBsAg or HCVAb positive) or other known causes of steatosis.

The diagnostic criteria for MetS were classified according to the Chinese Medical Association diabetes branch (CDS),[41] which defines MetS as meeting three or more of the following four categories: (1) overweight or obesity (BMI $\geq$25.0 kg/m²); (2) hypertension (SBP $\geq$140 mm Hg, DBP $\geq$90 mm Hg or prior diagnosis); (3) hyperglycaemia (FPG $\geq$6.1 mmol/L or 2 h postprandial glucose (PG) $\geq$7.8 mmol/L, or prior diagnosis); (4) dyslipidemia (TG $\geq$1.7 mmol/L, or HDL $\leq$0.9 mmol/L in males and $\leq$1.0 mmol/L in females).

Missing data imputation

As missing values existed in our longitudinal cohort data, multiple imputation had to be performed before the GEE analysis and causal network construction. Because the imputation method was dependent on the patterns of the missing data and the types of imputed variables, without loss of generality, the Markov chain Monte Carlo (MCMC) method was chosen according to the Multiple Imputation (MI) Procedure of SAS V.9.1.3.[42] Most variables had <2% missing observations before imputation except smoking and exercise, having <10% missing values.

Table 1 Variable abbreviations and assignments

Abbreviation	Variables (value assignments)
NAFLD	Non-alcoholic fatty liver disease (0=without NAFLD, 1=with NAFLD)
MetS and its components	Metabolic syndrome (0=without MetS, 1=with MetS); MetS components (obesity, hypertension, hyperglycaemia, dyslipidemia; 0=no, 1=yes)
SBP	Systolic blood pressure, mm Hg
DBP	Diastolic blood pressure, mm Hg
GGT	Gamma-glutamyltranspeptidase, U/L
TP	Serum total protein, g/L
ALB	Serum albumin, g/L
GLO	Serum globulins, g/L
A/G	The ratio between ALB and GLO
BUN	Blood urea nitrogen, mg/L
CREA	Serum creatinine, mg/dL
CHOL	Total cholesterol, mg/dL
TG	Triglycerides, mmol/L
LDL-C	Low-density lipoprotein cholesterol, mmol/L
HDL-C	High-density lipoprotein cholesterol, mmol/L
FPG	Fasting Plasma Glucose, mg/dL
Hb	Haemoglobin, g/L
MCHC	Mean corpuscular haemoglobin concentration, g/L
HCT	Haematocrit, %
MCV	Mean corpuscular volume, fL
MCH	Mean corpuscular haemoglobin, pg
RDW	Red blood cell distribution width, %
RDW-CV	Variation coefficient of red blood cell distribution width, %
RDW-SD	SD of red blood cell distribution width, fL
WCC	White cell count, 10^9/L
PDW	Platelet distribution width, %
MPV	Mean platelet volume, fL
PCT	Thrombocytocrit, %
Diet	0=Vegetarian, 1=meat based, 2=normal, 3=sea food
Drinking	0=never, 1=seldom, 2=often
Smoking	0=never, 1=seldom, 2=quit, 3=1–4/day, 4=5–15/day, 5$\geq$15/day
Quality of sleep	0=excellent, 1=well, 2=fair, 3=poor, 4=very poor
Exercise	0=never, 1=seldom, 2=often or everyday

Statistical analysis

Quantitative variables were summarised by mean±SD for normal distributed variables, median (25th, 75th centile) for non-normal distributed variables and categorical variables by percentages (%). The p values between the two groups were calculated by t test for normal distributed quantitative variables, with non-parametric test for skew distributed quantitative variables and χ^2 test for categorical variables. The number of

person-years was calculated as the sum of the follow-up times from the baseline to the occurrence of NAFLD (or MetS) or the last health check-up. The potential causality of the temporal sequence from NAFLD to MetS (or from MetS to NAFLD) was detected by GEE models. Simple GEE analyses were first performed to select the potential risk factors for MetS in subcohort A and NAFLD in subcohort B, separately. The variables with p value less than the significance level 0.05 were then included in the multiple GEE models. Statistical analyses were performed using SAS V.9.1.3 (SAS Institute, Inc, Cary, North Carolina, USA). A two-sided p value <0.05 was considered statistically significant.

Causal inference by simplified BN

BN[43–45] was used to construct the reciprocal causality pathway of NAFLD and MetS (see online supplementary text 1 for details). The primary network was usually too complex to identify the causal effect pathways efficiently. Thus, network simplification was essential before causal inference. We proposed the following simplification criteria: (1) keep the direct and indirect effect pathway; (2) keep the confounding pathway; (3) drop the nodes with irrationality on temporal logic; (4) drop the independent causal factors; (5) drop the collider nodes and collider edges (see online supplementary text 2 for details).[38]

To find the relative importance of the pathogenesis of a specific pathway on the simplified causal network, it is necessary to rank its effect by conditional distribution with all the variables set to the highest level. Taking the causal inference of NAFLD on MetS as an example, suppose we have a pathway NAFLD (0, 1), A (0, 1), B (0, 1), C (0, 1), MetS (0, 1): its relative importance in pathogenesis can be assessed by

$$P(\text{MetS} = 1 | \text{NAFLD} = 1, A = 1, B = 1, C = 1) \quad (1)$$

Furthermore, the joint probability distribution was calculated to evaluate public health significance of the specific pathway. As for the above pathway, the joint probability was calculated by

$$\begin{aligned} &P(\text{MetS} = 1, \text{NAFLD} = 1, A = 1, B = 1, C = 1) \\ &= P(\text{MetS} = 1 | \text{NAFLD} = 1, A = 1, B = 1, C = 1) \cdot \quad (2) \\ &\quad P_e(\text{NAFLD} = 1, A = 1, B = 1, C = 1) \end{aligned}$$

where the $P_e(\text{NAFLD} = 1, A = 1, B = 1, C = 1)$ was the co-exposure rate of these risk factors (NAFLD, A, B, C) in the population. Usually, since the joint probability could be quite small due to the quite lower exposure rate, public health significance of the pathway might be limited. The natural causal effect of the specific pathway (including direct and indirect pathway) was calculated by the theorem of causal effects identification (see online supplementary text 3 for further information).[46–49] The BN construction and causal inference were performed on *Hugin* 7.0.[50 51]

RESULTS

Characteristics of subcohorts

The baseline characteristics of participants in subcohorts A and B are shown in table 2 and online supplementary table S1. In subcohort A, among 8426 participants, 1243 (14.75%) participants suffered from NAFLD at baseline. During the follow-up from 2005 to 2011, 93 incidences of MetS were diagnosed in patients with NAFLD, with an incidence density of 2.47 per 100 person-years (93/3767 person-years), while 103 were diagnosed in the non-NAFLD group, with an incidence density of 0.54 per 100 person-years (103/19 040 person-years). In subcohort B, among 16 110 participants, 2170 (13.47%) suffered from MetS at baseline. The incidence density of NAFLD in patients with MetS (17.39 per 100 person-years, 1089/6264) was significantly higher than that in the non-MetS group (6.81 per 100 person-years, 2558/37 572).

Bidirectional associations between NAFLD and MetS analysed by GEE models

The multiple GEE analyses for subcohorts A and B, adjusting for the potential confounding factors selected by simple GEE models, are presented in figure 2, online supplementary tables S2 and S3. They revealed that NAFLD was a strong potential risk factor for MetS (relative risks (RRs) and 95% CI 5.23, 3.50 to 7.81) and its components (obesity, diabetes, hypertension and dyslipidemia), while MetS and its components were also potential predictors for NAFLD, with obesity the largest effect (shown in figure 2).

Reciprocal causal inference by BN

Based on the proposed simplification criteria in the Materials and methods section, the simplified BN from NAFLD to MetS retained 14 nodes, 33 edges and 36 pathways (shown in figure 3) from the primary network (see online supplementary figure S1). The total effects of NAFLD on MetS or its components are summarised in table 3, indicating that the total effect was greatest on dyslipidemia, followed by obesity, diabetes, hypertension and MetS.

The relative importance from the viewpoints of pathogenesis for the pathway from NAFLD to MetS is shown in online supplementary table S4, with their ranking effects by conditional distribution with all the variables set to the highest level. Generally, the most important pathway was that NAFLD led to elevated GGT, then to dyslipidaemia, followed by hypertension and, finally, the incidence of MetS. The second important pathway was that persistent NAFLD led to obesity, then to diabetes, or dyslipidaemia or hypertension and, finally, to MetS. The elevated CHOL level in the pathway would result in a decrease in the incidence of MetS.

However, a single-causal pathway had less public health significance due to the relatively lower exposure rate of the risk factors through each pathway in the population. Take the typical pathway of NAFLD, GGT1,

Table 2 Baseline characteristics of participants in subcohorts A and B

	Subcohort A (n=8426)			Subcohort B (n=16 110)		
	NAFLD	non-NAFLD	p Value	MetS	non-MetS	p Value
Sample size	1243	7183		2170	13 940	
Age (years)	43.93±11.90	37.87±11.89	<0.001	54.69±15.66	41.43±14.45	<0.001
BMI (kg/m²)	23.42±1.42	21.82±2.30	<0.001	27.51±2.38	23.23±2.99	<0.001
SBP (mm Hg)	118.15±10.51	112.14±12.72	<0.001	146.90±17.36	119.00±17.63	<0.001
DBP (mm Hg)	71.83±8.39	67.43±9.05	<0.001	83.22±11.50	70.46±10.41	<0.001
GGT (U/L)*	19.00 (15.00, 26.00)	13.00 (11.00, 18.00)	<0.001	22.00 (17.00, 32.00)	15.00 (11.00, 21.00)	<0.001
TP (g/L)	74.12±4.20	73.65±4.30	0.034	74.68±4.49	73.77±4.33	<0.001
ALB (g/L)	46.36±2.59	46.39±2.53	0.628	46.03±2.66	46.34±2.57	0.0002
GLO (g/L)	27.76±4.03	27.24±3.88	0.058	28.65±4.25	27.43±4.00	<0.001
A/G	1.71±0.29	1.74±0.29	0.074	1.65±0.29	1.73±0.29	<0.001
BUN (mg/L)	5.08±1.18	4.67±1.17	<0.001	5.39±1.39	4.84±1.25	<0.001
CREA (mg/dL)*	82.90 (73.45, 91.32)	73.65 (66.50, 84.00)	<0.001	85.10 (77.01, 92.80)	77.30 (68.20, 87.90)	<0.001
FPG (mmol/L)	4.96±0.58	4.75±0.61	<0.001	6.15±1.65	4.93±0.85	<0.001
CHOL (mg/dL)	5.01±0.88	4.74±0.87	<0.001	5.37±1.09	4.86±0.92	<0.001
TG (mmol/L)*	1.10 (0.77, 1.38)	0.76 (0.55, 1.06)	<0.001	1.98 (1.47, 2.60)	0.95 (0.64, 1.41)	<0.001
HDL-C (mmol/L)	1.34±0.27	1.46±0.29	<0.001	1.19±0.38	1.38±0.33	<0.001
LDL-C (mmol/L)	2.92±0.68	2.60±0.69	<0.001	3.19±0.77	2.74±0.71	<0.001
Hb (g/L)	149.66±13.02	140.46±14.73	<0.001	150.80±13.87	143.30±15.06	<0.001
HCT (%)	44.53±3.52	42.24±3.86	<0.001	44.85±3.66	42.98±3.94	<0.001
MCV (fL)	89.66±4.02	89.86±4.80	0.085	89.69±4.36	89.86±4.70	0.2373
MCH (pg)	30.13±1.65	29.87±2.05	0.054	30.15±1.73	29.94±1.98	0.0001
MCHC (g/L)	336.01±10.24	332.21±11.46	<0.001	336.20±11.50	333.00±11.42	<0.001
RDW-CV (%)	12.79±0.79	12.80±1.06	0.15	12.87±0.85	12.81±1.02	0.0206
RDW-SD (fL)	41.29±2.55	41.26±2.54	0.881	41.51±2.60	41.33±2.59	0.03
WCC (10⁹/L)	6.71±1.59	6.12±1.46	<0.001	7.04±1.61	6.31±1.52	<0.001
PDW (%)	12.10±1.60	12.34±1.71	0.017	12.22±1.74	12.31±1.94	0.1026
MPV (fL)	10.30±0.73	10.46±0.81	0.001	10.31±0.82	10.42±0.82	<0.001
PCT (%)	0.26±0.07	0.25±0.10	0.009	0.24±0.07	0.25±0.09	0.0367

*Non-normal distributed variables were presented as median (25th, 75th centile), and the p values were calculated using non-parametric test.
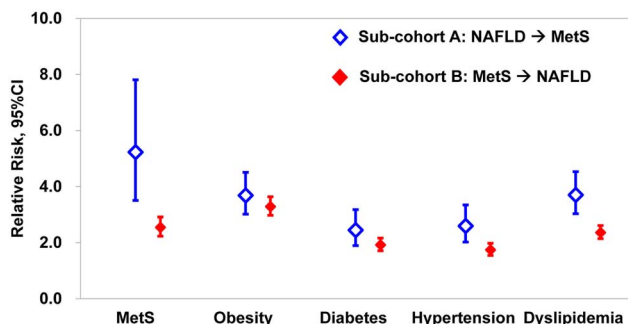
Figure 2 Relative risks (RRs) and 95% CIs of developing MetS or its components having NAFLD at baseline (hollow diamond, subcohort A), and developing NAFLD having MetS or its components at baseline (solid diamond, subcohort B). The RRs were calculated from the multiple generalised estimating equation (GEE) analyses, adjusting for the potential confounding factors selected by simple GEE model.

GGT2, dyslipidemia, hypertension and MetS as an example, which was a local structure extracted from a simplified network. The conditional probability was calculated to arrive at the indirect effect of this specific pathway (as shown in figure 4). In this pathway, GGT and dyslipidemia were key factors for the development of MetS in pathogenesis, but its effect was very small (0.025%) in the population, with less public health significance.

The simplified causal BN from MetS to NAFLD retained 17 nodes and 98 pathways (shown in figure 5).

The total effect of MetS and its component on NAFLD are shown in table 3; it revealed that MetS had the largest effect, followed by obesity, diabetes, dyslipidemia and hypertension. The relative importance of these pathways is shown in online supplementary table S5, and indicates that the dominant causal pathway is that dyslipidaemia leads to other MetS components and finally results in NAFLD.

DISCUSSION

To the best of our knowledge, this has been the first large-scale bidirectional longitudinal cohort study to clarify the reciprocal causality between NAFLD and MetS within the same study population. We confirmed that NAFLD could be both a cause and consequence of MetS in this bidirectional longitudinal cohort from a section of the Chinese population. As for the results of the longitudinal association between NAFLD and MetS, similar results have been found in other national and regional populations for the temporal sequence of NAFLD to MetS[15 16 22–29] and MetS to NAFLD.[29–34] Furthermore, we found that the effect of MetS on NAFLD was higher than that of NAFLD on MetS in reciprocal causality between NAFLD and MetS.

The simplified BN was constructed to infer the reciprocal causality between NAFLD and MetS. The total effect of NAFLD on MetS was 2.49%, while it was 19.92% for MetS on NAFLD, in the framework of causal network, indicating that the effect of MetS on NAFLD
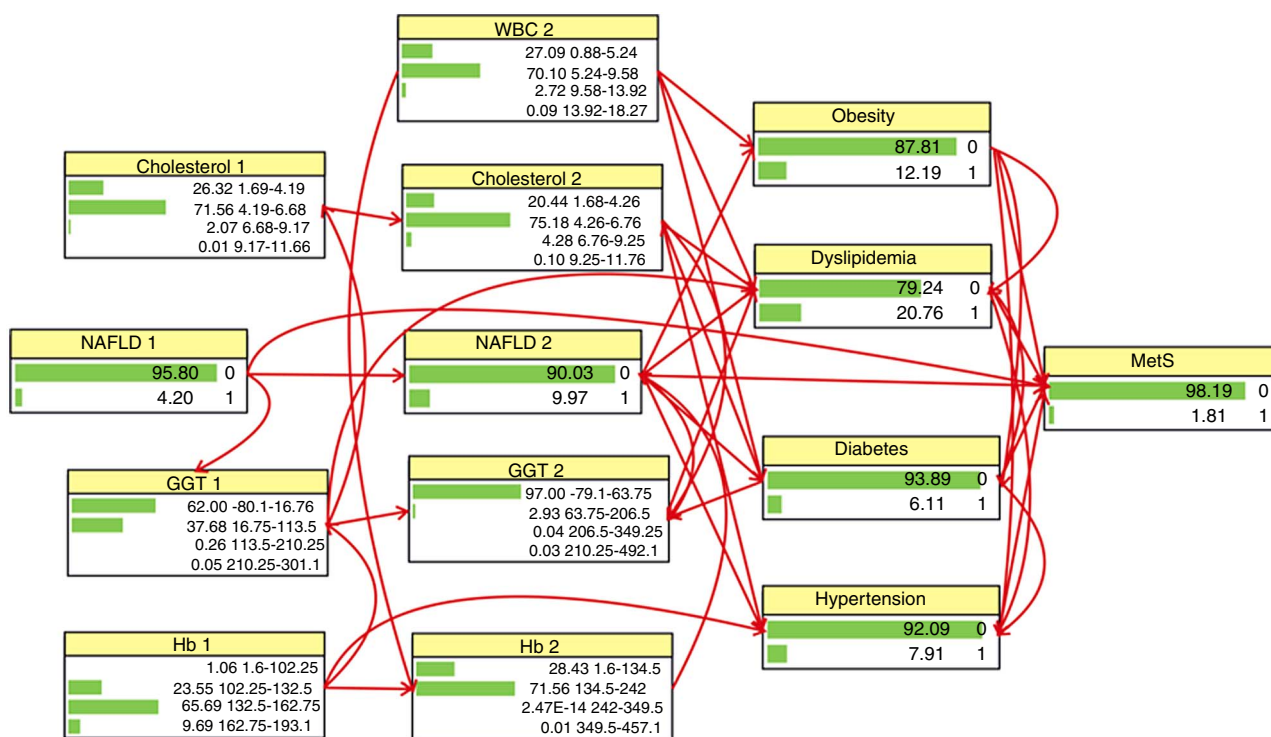

Figure 3 Simplified Bayesian network from NAFLD to MetS retained 14 nodes, 33 edges and 36 pathways. The numbers '1' and '2' associated with the variables denote the status at baseline and at the end of follow-up, respectively.

Table 3 Total effects of NAFLD to MetS or MetS to NAFLD

From NAFLD to MetS			From MetS to NAFLD		
Effects*	P(M∣NAFLD=1)%†	AR (%)‡	Causes*	P(NAFLD∣M=1)%§	AR (%)¶
Dyslipidemia	30.5	10.15	MetS	36.35	19.92
Obesity	19.5	7.63	Obesity	28.59	16.37
Diabetes	9.85	3.90	Diabetes	27.29	10.85
Hypertension	11.27	3.51	Dyslipidemia	25.27	10.74
MetS	4.19	2.49	Hypertension	23.63	7.36

*The rows of the table were ranked by AR%.
†$P(M|NAFLD=1)\%$ denoted the conditional probability of MetS and its components (M) given the presence of NAFLD.
‡Attributable risks, AR(%), were calculated as $P(M|NAFLD=1)-P(M|NAFLD=0)$.
§$P(NAFLD|M=1)\%$ denoted the conditional probability of NAFLD given the presence of MetS or its components (M).
¶Attributable risks, AR(%), were calculated as $P(NAFLD|M=1)-P(NAFLD|M=0)$.

was higher than that of NAFLD on MetS. This unbalanced causal effect was consistent with unbalanced incidence densities observed in the bidirectional subcohort. However, the effect of NAFLD on MetS and its components was different, with the effect on dyslipidemia the largest (AR%=10.15%), followed by that on obesity (7.63%), diabetes (3.90%) and hypertension (3.51%). Similar patterns were inferred for the effect of MetS components on NAFLD. The result of the BN was similar to the above GEE analysis results. The above results demonstrated that obesity and dyslipidemia were key factors linking NAFLD and MetS. Several studies suggested that obesity was associated with an increased risk of NAFLD[52][53] and might also be important in determining the development of MetS.[54] These viewpoints were confirmed in this study.

Among the 36 causal pathways from NAFLD to MetS, the most important was that NAFLD led to elevated GGT, then to dyslipidaemia, hypertension and, finally, to MetS. GGT hosted the key node in this causal pathway, and participants with elevated GGT levels would have an increased risk for MetS by increasing oxidative stress, insulin resistance and hepatic steatosis.[55–60] The second
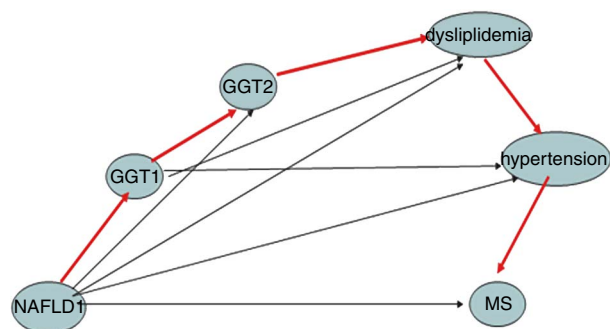

Figure 4 Conditional probability and local structure extracted from the simplified network, for calculating the indirect effect of this specific pathway (NAFLD, GGT1, GGT2, dyslipidemia, hypertension and MetS). The numbers '1' and '2' associated with the variables denote the status at baseline and at the end of follow-up, respectively.

important causal pathway was that NAFLD led to obesity, then to the other components, and finally resulted in MetS. This was concordant with previous reports, which considered NAFLD and MetS may be linked by fat ectopic accumulation and insulin resistance.[8 9 20 54]

Among the 98 causal pathways from MetS or its components to NAFLD, the dominant causal pathways begin by leading to dyslipidaemia, and finally resulted in NAFLD. In these pathways, dyslipidaemia might cause the increased triglyceride synthesis in liver cells and triglyceride accumulation in the liver, and then block the low-density lipoprotein synthesis, finally resulting in NAFLD.[61] Although the cause–effect pathogenesis still needs to be clarified in further investigation, the association between haematocrit and NAFLD has been detected in another Chinese population,[62] and haemoglobin has been identified as a biomarker of NAFLD in some studies.[63–65]

The association between cholesterol and NAFLD (or the metabolic syndrome) has been fairly well established through long-term studies of high levels of serum cholesterol and the incidence of NAFLD, MetS and coronary heart diseases. Surprisingly, in this study, we found that the elevated total cholesterol appearing in the above pathways would result in a lower probability of MetS and NAFLD. This may not be in accordance with the conventional viewpoint. Nevertheless, a meta-analysis reported that serum total cholesterol levels were significantly lower in non-alcoholic steatohepatitis (NASH) than in simple steatosis.[66] This study concluded that lower cholesterol levels were independently associated with NASH, in addition to the well-known association with MS and IR. However, the mechanistic explanations linking a lower cholesterol level with NAFLD and MetS still need further investigation.

Our study has several limitations. First, the presence of NAFLD was assessed by experienced radiologists using abdominal ultrasonography, and we have no information on the intraobserver or interobserver reliability of ultrasonographic examinations. The diagnosis of NAFLD was not subjected to any semiquantitative indices.[67 68] Second, owing to the absence of waist circumference
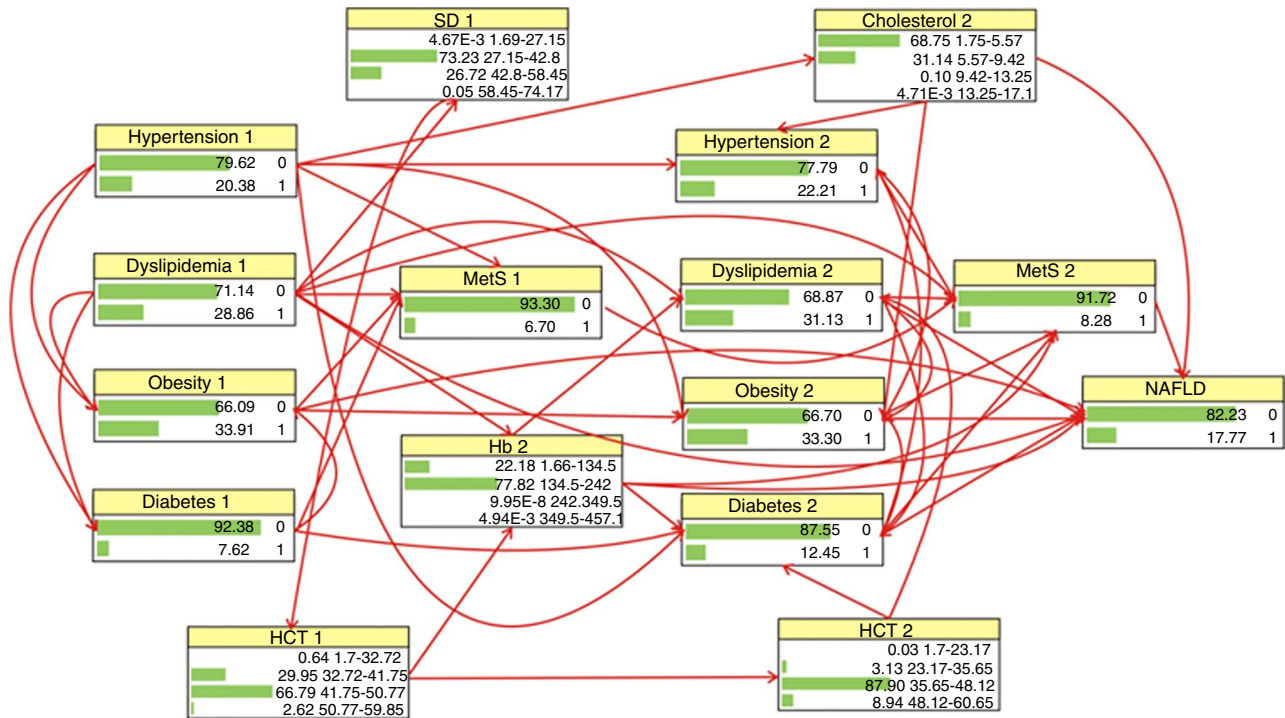
Figure 5 Simplified causal Bayesian network from MetS to NAFLD (17 nodes and 98 pathways). The numbers '1' and '2' associated with the variables denote the status at baseline and at the end of follow-up, respectively.

measurement in the health check-up programme, the diagnostic criteria of MetS were based on the Chinese medical association diabetes branch, rather than the international standard criteria. Third, because the present study was based on a routine health check-up system in an urban Han Chinese population of Shandong province, generalisability to the general population was uncertain. Further investigation needs to be carried out to confirm the reciprocal causality between NAFLD and MetS in a larger sample of the general population.

Author affiliations
[1]Department of Biostatistics, School of Public Health, Shandong University, Jinan, Shandong, China
[2]Medical Department, Qilu Hospital of Shandong University, Jinan, Shandong, China
[3]Health Management Center, Shandong Provincial QianFoShan Hospital, Jinan, Shandong, China

Patient consent Obtained.

Ethics approval Ethics Committee of School of Public Health, Shandong University.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

REFERENCES
1. Alberti KG, Zimmet P, Shaw J. The metabolic syndrome—a new worldwide definition. *Lancet* 2005;366:1059–62.
2. Angulo P. Nonalcoholic fatty liver disease. *N Engl J Med* 2002;346:1221–31.
3. Cornier MA, Dabelea D, Hernandez TL, *et al.* The metabolic syndrome. *Endocr Rev* 2008;29:777–822.
4. Eckel RH, Grundy SM, Zimmet PZ. The metabolic syndrome. *Lancet* 2005;365:1415–28.
5. Adams LA, Lindor KD. Nonalcoholic fatty liver disease. *Ann Epidemiol* 2007;17:863–9.
6. Fan JG, Zhu J, Li XJ, *et al.* Prevalence of and risk factors for fatty liver in a general population of Shanghai, China. *J Hepatol* 2005;43:508–14.
7. Browning JD, Szczepaniak LS, Dobbins R, *et al.* Prevalence of hepatic steatosis in an urban population in the United States: impact of ethnicity. *Hepatology* 2004;40:1387–95.
8. Bloomgarden ZT. Second World Congress on the Insulin Resistance Syndrome: insulin resistance syndrome and nonalcoholic fatty liver disease. *Diabetes Care* 2005;28:1518–23.
9. Angelico F, Del Ben M, Conti R, *et al.* Insulin resistance, the metabolic syndrome, and nonalcoholic fatty liver disease. *J Clin Endocrinol Metab* 2005;90:1578–82.
10. Cortez-Pinto H, Camilo ME, Baptista A, *et al.* Non-alcoholic fatty liver: another feature of the metabolic syndrome? *Clin Nutr* 1999;18:353–8.

11. Bedogni G, Miglioli L, Masutti F, et al. Prevalence of and risk factors for nonalcoholic fatty liver disease: the Dionysos nutrition and liver study. *Hepatology* 2005;42:44–52.

12. Bellentani S, Saccoccio G, Masutti F, et al. Prevalence of and risk factors for hepatic steatosis in Northern Italy. *Ann Intern Med* 2000;132:112–17.

13. Akbar DH, Kawther AH. Nonalcoholic fatty liver disease in Saudi type 2 diabetic subjects attending a medical outpatient clinic: prevalence and general characteristics. *Diabetes Care* 2003;26:3351–2.

14. Donati G, Stagni B, Piscaglia F, et al. Increased prevalence of fatty liver in arterial hypertensive patients with normal liver enzymes: role of insulin resistance. *Gut* 2004;53:1020–3.

15. Adams LA, Waters OR, Knuiman MW, et al. NAFLD as a risk factor for the development of diabetes and the metabolic syndrome: an eleven-year follow-up study. *Am J Gastroenterol* 2009;104:861–7.

16. Fan JG, Li F, Cai XB, et al. Effects of nonalcoholic fatty liver disease on the development of metabolic disorders. *J Gastroenterol Hepatol* 2007;22:1086–91.

17. Marchesini G, Bugianesi E, Forlani G, et al. Nonalcoholic fatty liver, steatohepatitis, and the metabolic syndrome. *Hepatology* 2003;37:917–23.

18. Pacifico L, Anania C, Martino F, et al. Functional and morphological vascular changes in pediatric nonalcoholic fatty liver disease. *Hepatology* 2010;52:1643–51.

19. Chavez-Tapia NC, Mendez-Sanchez N, Uribe M. The metabolic syndrome as a predictor of nonalcoholic fatty liver disease. *Ann Intern Med* 2006;144:379; author reply 80.

20. Hamaguchi M, Kojima T, Takeda N, et al. The metabolic syndrome as a predictor of nonalcoholic fatty liver disease. *Ann Intern Med* 2005;143:722–8.

21. Souza MR, Diniz Mde F, Medeiros-Filho JE, et al. Metabolic syndrome and risk factors for non-alcoholic fatty liver disease. *Arq Gastroenterol* 2012;49:89–96.

22. Zhang T, Zhang Y, Zhang C, et al. Prediction of metabolic syndrome by non-alcoholic fatty liver disease in northern urban Han Chinese population: a prospective cohort study. *PLoS ONE* 2014;9: e96651.

23. Smits MM, Ioannou GN, Boyko EJ, et al. Non-alcoholic fatty liver disease as an independent manifestation of the metabolic syndrome: results of a US national survey in three ethnic groups. *J Gastroen Hepatol* 2013;28:664–70.

24. Rector RS, Thyfault JP, Wei YZ, et al. Non-alcoholic fatty liver disease and the metabolic syndrome: an update. *World J Gastroentero* 2008;14:185–92.

25. Fan JG, Zhu JN, Li XJ, et al. Fatty liver and the metabolic syndrome among Shanghai adults. *J Gastroen Hepatol* 2005;20:1825–32.

26. Ryoo JH, Choi JM, Moon SY, et al. The clinical availability of non alcoholic fatty liver disease as an early predictor of the metabolic syndrome in Korean men: 5-year's prospective cohort study. *Atherosclerosis* 2013;227:398–403.

27. Hamaguchi M, Kojima T, Itoh Y, et al. The severity of ultrasonographic findings in nonalcoholic fatty liver disease reflects the metabolic syndrome and visceral fat accumulation. *Am J Gastroenterol* 2007;102:2708–15.

28. Lonardo A, Ballestri S, Marchesini G, et al. Nonalcoholic fatty liver disease: a precursor of the metabolic syndrome. *Dig Liver Dis* 2015;47:181–90.

29. Mohan V, Farooq S, Deepa M, et al. Prevalence of non-alcoholic fatty liver disease in urban south Indians in relation to different grades of glucose intolerance and metabolic syndrome. *Diabetes Res Clin Pr* 2009;84:84–91.

30. Zhang T, Zhang C, Zhang Y, et al. Metabolic syndrome and its components as predictors of nonalcoholic fatty liver disease in a northern urban Han Chinese population: a prospective cohort study. *Atherosclerosis* 2015;240:144–8.

31. Tsai CH, Li TC, Lin CC. Metabolic syndrome as a risk factor for nonalcoholic fatty liver disease. *South Med J* 2008;101:900–5.

32. Hamaguchi M, Takeda N, Kojima T, et al. Identification of individuals with non-alcoholic fatty liver disease by the diagnostic criteria for the metabolic syndrome. *World J Gastroentero* 2012;18:1508–16.

33. Graham RC, Burke A, Stettler N. Ethnic and sex differences in the association between metabolic syndrome and suspected nonalcoholic fatty liver disease in a nationally representative sample of US adolescents. *J Pediatr Gastr Nutr* 2009;49:442–9.

34. Castro-Martinez MG, Banderas-Lares DZ, Ramirez-Martinez JC, et al. Prevalence of nonalcoholic fatty liver disease in subjects with metabolic syndrome. *Cir Cir* 2012;80:128–33.

35. Yki-Järvinen H. Non-alcoholic fatty liver disease as a cause and a consequence of metabolic syndrome. *Lancet Diabetes Endocrinol* 2014;2:901–10.

36. Loria P, Lonardo A, Anania F. Liver and diabetes. A vicious circle. *Hepatol Res* 2013;43:51–64.

37. Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* 1986;42:121–30.

38. Pearl J. *Causality: models, reasoning and inference*. Cambridge University Press, 2009.

39. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology* 1999;10:37–48.

40. Fatty Liver and Alcoholic Liver Disease Study Group of the Chinese Liver Disease Association. [Guidelines for diagnosis and treatment of nonalcoholic fatty liver diseases]. *Zhonghua Gan Zang Bing Za Zhi* 2006;14:161–3.

41. Lu YH, Lu JM, Wang SY, et al. [Comparison of the diagnostic criteria of metabolic syndrome by International Diabetes Federation and that by Chinese Medical Association Diabetes Branch]. *Zhonghua Yi Xue Za Zhi* 2006;86:386–9.

42. Schafer JL. *Analysis of incomplete multivariate data*. Taylor & Francis, 1997.

43. Jensen FV. *Bayesian networks and decision graphs*. Springer, 2001.

44. Kjaerulff UB, Madsen AL. *Bayesian networks and influence diagrams: a guide to construction and analysis*. Springer, 2007.

45. Lauritzen SL, Datalogi AUIfESAfMo. *The EM algorithm for graphical association models with missing data*. University of Aalborg, Institute for Electronic Systems, Department of Mathematics and Computer Science, 1991.

46. Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. *Epidemiology* 1992;3:143–55.

47. Pearl J. Direct and indirect effects. *Proceedings of the Seventeeth Conference on Uncertainty in Artificial Intelligence*; San Francisco: Morgan Kaufmann, 2001:411–20.

48. van der Laan MJ, Petersen ML. *Estimation of direct and indirect causal effects in longitudinal studies*. UC Berkley, Division of Biostaistics Working Paper Series 155. 2004.

49. Petersen ML, Sinisi SE, van der Laan MJ. Estimation of direct causal effects. *Epidemiology* 2006;17:276–84.

50. Andersen SK, Olesen KG, Jensen FV, et al. Hugin-a Shell for Building Bayesian Belief Universes for Expert Systems. *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence*; 1989:1080–5.

51. Anders L, Madsen ML, Uffe B, et al. The Hugin Tool for Learning Bayesian Networks. *Proceedings of 7th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty*. 2003;Vol. 2711:594–605.

52. Burgert TS, Taksali SE, Dziura J, et al. Alanine aminotransferase levels and fatty liver in childhood obesity: associations with insulin resistance, adiponectin, and visceral fat. *J Clin Endocrinol Metab* 2006;91:4287–94.

53. Fabbrini E, Sullivan S, Klein S. Obesity and nonalcoholic fatty liver disease: biochemical, metabolic, and clinical implications. *Hepatology* 2010;51:679–89.

54. Goodpaster BH, Krishnaswami S, Harris TB, et al. Obesity, regional body fat distribution, and the metabolic syndrome in older men and women. *Arch Intern Med* 2005;165:777–83.

55. Andre P, Balkau B, Vol S, et al. Gamma-glutamyltransferase activity and development of the metabolic syndrome (International Diabetes Federation Definition) in middle-aged men and women: Data from the Epidemiological Study on the Insulin Resistance Syndrome (DESIR) cohort. *Diabetes Care* 2007;30:2355–61.

56. Giral P, Jacob N, Dourmap C, et al. Elevated gamma-glutamyltransferase activity and perturbed thiol profile are associated with features of metabolic syndrome. *Arterioscler Thromb Vasc Biol* 2008;28:587–93.

57. Kozakova M, Palombo C, Eng MP, et al. Fatty liver index, gamma-glutamyltransferase, and early carotid plaques. *Hepatology* 2012;55:1406–15.

58. Lee DS, Evans JC, Robins SJ, et al. Gamma glutamyl transferase and metabolic syndrome, cardiovascular disease, and mortality risk: the Framingham Heart Study. *Arterioscler Thromb Vasc Biol* 2007;27:127–33.

59. Oda E, Kawai R, Watanabe K, et al. Prevalence of metabolic syndrome increases with the increase in blood levels of gamma glutamyltransferase and alanine aminotransferase in Japanese men and women. *Intern Med* 2009;48:1343–50.

60. Onat A, Can G, Ornek E, et al. Serum γ-glutamyltransferase: independent predictor of risk of diabetes, hypertension, metabolic syndrome, and coronary disease. *Obesity (Silver Spring)* 2012;20:842–8.

61. Chatrath H, Vuppalanchi R, Chalasani N. Dyslipidemia in patients with nonalcoholic fatty liver disease. *Semin Liver Dis* 2012;32:22–9.

62. Li G, Hu H, Shi W, *et al*. Elevated hematocrit in nonalcoholic fatty liver disease: a potential cause for the increased risk of cardiovascular disease? *Clin Hemorheol Microcirc* 2012;51:59–68.

63. Liu W, Baker SS, Baker RD, *et al*. Upregulation of hemoglobin expression by oxidative stress in hepatocytes and its implication in nonalcoholic steatohepatitis. *PLoS ONE* 2011;6:e24363.

64. Yilmaz Y. Serum proteomics for biomarker discovery in nonalcoholic fatty liver disease. *Clin Chim Acta* 2012;413:1190–3.

65. Yu C, Xu C, Xu L, *et al*. Serum proteomic analysis revealed diagnostic value of hemoglobin for nonalcoholic fatty liver disease. *J Hepatol* 2012;56:241–7.

66. Carulli L, Ballestri S, Lonardo A, *et al*. Is nonalcoholic steatohepatitis associated with a high-though-normal thyroid stimulating hormone level and lower cholesterol levels? *Intern Emerg Med* 2013;8:297–305.

67. Ballestri S, Romagnoli D, Nascimbeni F, *et al*. Role of ultrasound in the diagnosis and treatment of nonalcoholic fatty liver disease and its complications. *Expert Rev Gastroenterol Hepatol* 2015;9:603–27.

68. Ballestri S, Lonardo A, Romagnoli D, *et al*. Ultrasonographic fatty liver indicator, a novel score which rules out NASH and is correlated with metabolic parameters in NAFLD. *Liver Int* 2012;32:1242–52.