Serum Alanine Transaminase as a Predictor of Type 2 Diabetes Incidence: The Yuport Prospective Cohort Study

Yu Hatano,[1] **Kazuo Inoue,**[2] **Saori Kashima,**[3] **Masatoshi Matsumoto**[4] **and Kimihiko Akimoto**[5]

[1]Department of Family Medicine and Community Health, Duke University School of Medicine, Durham, NC, USA
[2]Department of Community Medicine, Chiba Medical Center, Teikyo University School of Medicine, Ichihara, Chiba, Japan
[3]Environmental Health Sciences Laboratory, Graduate School of Advanced Science and Engineering, Hiroshima University, Higashi-Hiroshima, Hiroshima, Japan
[4]Department of Community-Based Medical System, Graduate School of Biomedical and Health Sciences, Hiroshima University, Hiroshima, Hiroshima, Japan
[5]Akimoto Occupational Health Consultant Office, Tokyo, Japan

Prior studies have shown an association between the incidence of diabetes with liver enzymes, such as alanine transaminase (ALT). Liver fibrosis scores, such as the Fibrosis-4 index which indicates chronic liver damage, were also associated with diabetes development. However, no literature compared predictive accuracy between ALT and Fibrosis-4 index. Thus, we aimed to determine it, and to assess its association using inverse probability of treatment weighting. This was a non-concurrent prospective cohort study of 9,748 subjects without diabetes receiving Yuport Health Checkup in Japan between 1998 and 2006. ALT was categorized into three groups: the highest ALT group (men ≥ 30 U/L and women ≥ 20 U/L), the middle (men ≥ 20 and < 30 U/L, and women ≥ 14 and < 20 U/L), and the lowest (men < 20 U/L and women < 14 U/L). The primary outcome was the new onset of diabetes. The area under the receiver operating characteristic curves (AUC) of ALT for predicting the diabetes development was higher than that of any other markers of liver damage. The AUC for ALT was 0.71, while that for the Fibrosis-4 index was 0.51 (p < 0.001 for the difference between the AUCs). The highest and middle ALT groups had a significantly higher incidence of diabetes than the lowest group: adjusted relative risk 1.79 [95% confidence interval (CI): 1.29, 2.58], and 1.64 [95% CI: 1.17, 2.38] respectively. Of the various indicators of liver function, ALT is likely to be the most accurate and associated predictor of diabetes development.

Introduction

Diabetes mellitus is a serious growing global public health threat, affecting a number of patients that has nearly quadrupled since 1980, and causing substantial morbidity and mortality, as well as enormous economic burden (World Health Organization 2016; IDF Diabetes Atlas Group 2013). Fatal complications are caused by diabetes: heart disease, strokes, diabetic retinopathy, diabetic nephropathy, and peripheral arterial disease which can lead to amputa-tions. Type 2 diabetes makes up about 90% of cases of dia-betes (Guariguata et al. 2014). It primarily occurs due to insulin resistance as a result of obesity. Identifying predic-tors and causes of type 2 diabetes mellitus is necessary to reduce the harm that it causes.

Prior prospective cohort studies have suggested that liver enzymes indicating current liver damage or inflamma-tion (alanine transaminase (ALT), aspartate aminotransfer-ase (AST), and γ-glutamyltransferase (GGT)) are associated with an increased risk for type 2 diabetes (Nakanishi et al.

2004; Cho et al. 2007; Ford et al. 2008; Goessling et al. 2008; Fraser et al. 2009; Kim et al. 2009; Xu et al. 2015; Liu et al. 2016; Zhang et al. 2018; De Silva et al. 2019; Kaneko et al. 2019; Li et al. 2019). Liver dysfunction can lead to impaired glucose homeostasis and type 2 diabetes development through insulin resistance (Kawaguchi et al. 2011). Liver fibrosis arises from chronic damage to the liver in conjunction with the accumulation of extracellular matrix proteins (scar tissue), which is a characteristic of most types of chronic liver diseases (Friedman 2003). Advanced liver fibrosis leads to liver failure and cirrhosis. A liver biopsy is required to measure the severity of liver fibrosis as a gold standard, but it is difficult and invasive to perform. Alternatively, liver fibrosis scores calculated from blood examinations are more frequently used because of its safety and convenience. Although a past report showed that liver fibrosis scores (Fibrosis-4 index (FIB-4 index), AST to platelet ratio index (APRI), and non-alcoholic fatty liver disease fibrosis score) were also associated with an increased incidence of type 2 diabetes, no literature directly compared predictive accuracy between liver fibrosis scores and liver enzymes (Chang et al. 2013). Thus, we aimed to determine which liver damage indicator is the most accurate predictor of diabetes incidence, which may lead to a better understanding of the mechanism of the association between liver damage and type 2 diabetes incidence, leading to a more specific target for future prevention.

Additionally, we examined the association of the most accurate predictor with the subsequent incidence of diabetes using a marginal statistical approach, inverse probability of treatment weighting (IPTW). To the best of our knowledge, all previous research papers studying the association between liver damage and the incidence of diabetes utilized a conditional statistical approach (Nakanishi et al. 2003a, 2004; Okamoto et al. 2003; Doi et al. 2007; Fujita et al. 2010; Hozawa et al. 2010; Kaneko et al. 2019), parametric models, which requires correct model specification (e.g., generalized linear model) and can potentially bias the results (Williamson and Ravani 2017). Analysis using propensity score does not require setting up a model between an outcome variable and covariates, and can have less bias (Cepeda et al. 2003).

Using data from the prospective cohort study based in the Yuport Heath Checkup Center, we assessed (1) which marker of liver damage, including liver fibrosis score, is best at predicting type 2 diabetes, and (2) the association between the best predictor and the development of type 2 diabetes, using the IPTW.

Material and Methods

Study design

The Yuport Health Checkup Center Study is a non-concurrent prospective single cohort study.

Study subjects

We used data acquired from the health screening pro-

gramme performed by the Yuport Medical Check-up Center in Tokyo (details described in our previous studies (Inoue et al. 2009, 2012)). For the current study we set the four-year baseline period to be between April 1998 and March 2002 and the four-year follow-up period between April 2002 and March 2006. During the baseline period, 21,885 persons underwent a check-up at least once (Fig. 1). If a subject underwent more than one check-up during the baseline period, the initial check-up was used. Of 11,129 persons who had been examined during both baseline and follow-up periods, 129 subjects with known diabetes mellitus at baseline were excluded, leaving 11,000 persons. Next, 410 individuals who had a baseline fasting plasma glucose level ≥ 7.0 mmol/l (126 mg/dl) were excluded (Committee of the Japan Diabetes Society on the Diagnostic Criteria of Diabetes Mellitus et al. 2010), as were 114 who had a baseline glycated haemoglobin (HbA$_{1c}$) ≥ 48 mmol/mol (6.5%) (Committee of the Japan Diabetes Society on the Diagnostic Criteria of Diabetes Mellitus et al. 2010). Of the remaining 10,476 persons, 433 with < 2 years between their baseline and follow-up check-ups were excluded. In addition, 295 subjects with hepatitis B or C virus infection were excluded. After these exclusions, 9,748 subjects formed the study group.
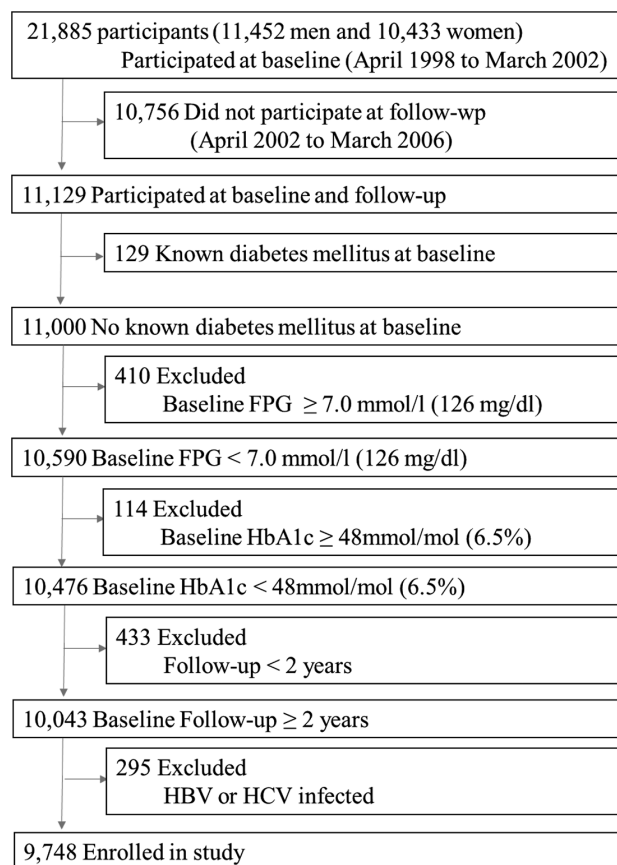

Fig. 1. Flow of enrolled study participants.

Ethical considerations

Informed consent for anonymous participation in epidemiological research was obtained at every check-up. This study was approved by the review boards of the Yuport Medical Checkup Center, Teikyo University School of Medicine (No.15-205) and Hiroshima University (No. E-1241).

Check-up procedures

All the check-up procedures were performed in the same manner, at both the baseline and follow-up visits, including blood tests (details described in our previous studies (Inoue et al. 2009, 2012)).

Laboratory tests

Blood samples were obtained after overnight fasting and analyses were performed at the centre's laboratory. For the measurements of fasting plasma glucose and HbA_{1c}, a Toshiba TBA-40FR Autoanalyser (Toshiba Medical Systems, Tokyo, Japan) was used. Plasma glucose was measured using the hexokinase-G6PD method (Denka Seiken, Niigata, Japan), with an inter-assay coefficient of variation $\leq 3.0\%$. For glucose analysis, sodium fluoride was employed as a preservative. HbA_{1c} was measured using the latex immunoagglutination method (Determiner haemoglobin A_{1c}; Kyowa Medex, Tokyo, Japan), with an inter-assay coefficient of variation of 1.7-2.1%, which was aligned with the Japan Diabetes Society (JDS) assigned values. The reference range was set at 4.3-5.8% by the manufacturer and was in accordance with the JDS guideline for the normal upper limit (Kuzuya et al. 2002). The JDS developed a set of national calibrators, with a recommendation to adjust the calibration of all routine HbA_{1c} methods to these calibrators (Hoelzel et al. 2004). Comparison of the JDS primary standard material with the assay used by the anchor laboratory of the *National Glycohemoglobin Standardization Program* (NGSP) in the USA revealed that the NGSP value (%) = (1.02 × JDS value (%)) + 0.25 (Hoelzel et al. 2004; Kashiwagi et al. 2012). Thus, the JDS values were converted into NGSP values according to this formula in this study.

Liver damage indices

Serum AST and ALT were measured using enzymatic methods (reagents supplied by Denka Seiken), as were the serum GGT levels (Wako Junyaku, Osaka, Japan). Serum alkaline phosphatase (ALP) and total bilirubin were also measured as liver damage indicators.

The following three indexes were calculated as the indicators of liver fibrosis scores. The FIB-4 index was calculated according to the published formula: FIB-4 index = (age [years] × serum AST [U/L]) / (platelet count $[10^9/L]$ × (serum ALT $[U/L]^{1/2}$) (Vallet-Pichard et al. 2007; Shah et al. 2009). The serum AST to ALT ratios (AAR) were calculated by the following formula: AAR = ([serum AST [U/L]] / [serum ALT [U/L]]). The APRIs were calculated: ARPI = ([serum AST [U/L] / upper limit of normal] / platelet count $[10^9/L]$) × 100 (Wai et al. 2003; Cales et al. 2009). The values for the upper limit of normal were defined according to the IFCC: serum AST 35 U/L for men, 30 U/L for women.

Categorized data

Serum ALT was categorized into three groups: the highest group, in which serum ALT was ≥ 30 U/L for men and ≥ 20 U/L for women, based on recent American College of Gastroenterology guidelines (Kwo et al. 2017), and the remaining individuals were divided into two quantile groups (the middle group: serum ALT ≥ 20 and < 30 U/L for men and ≥ 14 and < 20 U/L for women, and the lowest group: serum ALT < 20 U/L for men and < 14 U/L for women). Impaired fasting glucose (IFG) was defined as fasting plasma glucose ≥ 5.6 mmol/L (100 mg/dl). Subjects with IFG were categorized into those with a fasting glucose < 6.1 mmol/L (110 mg/dl) and ≥ 6.1 mmol/L (110 mg/dl).

Diagnosis of type 2 diabetes mellitus

Type 2 diabetes was diagnosed when the subject met at least one of the following criteria (Committee of the Japan Diabetes Society on the Diagnostic Criteria of Diabetes Mellitus et al. 2010; American Diabetes Association 2017):
・fasting plasma glucose level ≥ 7.0 mmol/L (126 mg/dl);
・$HbA_{1c} \geq 48$ mmol/mol (6.5%);
・diagnosis of type 2 diabetes by a physician

Statistical analysis

The area under the receiver operating characteristic (ROC) curves (AUC) for each indicator of liver damage and the difference in AUC for each indicator of liver damage compared to one with the highest AUC were utilized to compare their ability to predict the development of type 2 diabetes. For 95% confidence intervals (CIs) of AUC and the difference in AUC, we used a non-parametric bootstrap approach with 1,000 iterations.

Inverse probability of treatment weighting (IPTW) based on the propensity score was used in order to adjust for potential confounding due to baseline imbalances in study covariates while preserving sample size (Robins et al. 2000; Xu et al. 2010; Austin 2011). The propensity score that predicted the probability of being in each main exposure variable category given baseline characteristics was used to generate patient-specific stabilized weights which control for imbalances of covariates (Robins et al. 2000; Xu et al. 2010) : age, sex, BMI, hypertension, total cholesterol, high-density lipoprotein (HDL) cholesterol, log transformed triglycerides, C-reactive protein (CRP), IFG and HbA_{1c}. A multivariable ordinal logistic regression model was utilized to estimate generalized propensity scores. Adjusted risk differences (RDs) and relative risks (RRs) were calculated by the formula below.

Adjusted RD = (Risk of the highest or middle ALT group) – (Risk of the lowest ALT group)

$$= \frac{\frac{1}{n}\sum_{i=1}^{n}\frac{\mathrm{I\!I}\,(A_i=\text{the middle or the highest ALT group})}{\hat{g}(A_i|W_i)}Y_i}{\frac{1}{n}\sum_{i=1}^{n}\frac{\mathrm{I\!I}\,(A_i=\text{the middle or the highest ALT group})}{\hat{g}(A_i|W_i)}}$$

$$-\frac{\frac{1}{n}\sum_{i=1}^{n}\frac{\mathrm{I\!I}\,(A_i=\text{the lowest group})}{\hat{g}(A_i|W_i)}Y_i}{\frac{1}{n}\sum_{i=1}^{n}\frac{\mathrm{I\!I}\,(A_i=\text{the lowest group})}{\hat{g}(A_i|W_i)}}$$

Adjusted RR = (Risk of the highest or middle ALT group) / (Risk of the lowest ALT group)

$$= \frac{\frac{1}{n}\sum_{i=1}^{n}\frac{\mathrm{I\!I}\,(A_i=\text{the middle or the highest ALT group})}{\hat{g}(A_i|W_i)}Y_i}{\frac{1}{n}\sum_{i=1}^{n}\frac{\mathrm{I\!I}\,(A_i=\text{the middle or the highest ALT group})}{\hat{g}(A_i|W_i)}}$$

$$/\frac{\frac{1}{n}\sum_{i=1}^{n}\frac{\mathrm{I\!I}\,(A_i=\text{the lowest group})}{\hat{g}(A_i|W_i)}Y_i}{\frac{1}{n}\sum_{i=1}^{n}\frac{\mathrm{I\!I}\,(A_i=\text{the lowest group})}{\hat{g}(A_i|W_i)}}$$

W: covariates: age, sex, BMI, hypertension, total cholesterol, HDL-C, triglycerides, CRP, IFG and HbA_{1c}.

A: ALT category.

Y: type 2 diabetes incidence.

The balance of covariates between the weighted cohorts was assessed, using standardized mean differences (SMD) (Austin 2011). SMD less than 0.1 was recognized as indicating good balance (Austin 2011). For 95% CIs of measure of association, we used a non-parametric bootstrap approach with 1,000 iterations.

Sensitivity analysis was conducted in order to estimate how much association unmeasured confounders need to have with main exposures (ALT category) and outcome (the incidence of type 2 diabetes) so that only unmeasured confounders could explain our observed statistical association (Ding and VanderWeele 2016). E-value was calculated with the formula below, using our adjusted RR. If the association of unmeasured confounders with the main exposure and outcome are equal to or more than E-value, the observed statistical estimate, RR, can be explained by unmeasured confounders. On the other hand, if not, the observed statistical estimate cannot be explained only by unmeasured confounders, which means that the estimate remains significant.

E-value = RR + √[RR*(RR-1)]

Because of their skewed distributions, serum levels of hepatic enzymes and triglycerides were log transformed for statistical analysis when they were recognized as continuous variables.

Statistical analyses were done using R for Windows, version 3.5.1 (R Foundation for Statistical Computing, Vienna, Austria). All analyses were two-tailed. P < 0.05 was considered statistically significant.

Results

Of the 9,748 subjects, 2,656 (27.2%) subjects had a serum ALT value above the upper limit of normal at baseline (Table 1). Sex, hypertension, BMI, CRP, cholesterols, fasting plasma glucose and HbA_{1c} were not balanced between ALT categories. After weighting by inversed probability with stabilized weight, the distribution of baseline characteristics was balanced because all SMDs were less than 0.1.

Fig. 2 and Table 2 show the ROC curves and the AUC for each indicator of liver damage in comparison of their ability to predict type 2 diabetes development. The AUC for ALT was 0.71 [95% CI: 0.68, 0.74], which was significantly higher than the AUC for any other indicator except for AAR. The AUC for AAR was 0.69 [95% CI: 0.66, 0.72] and that for GGT was 0.67 [95% CI: 0.64, 0.69]. The AUCs for liver fibrosis scores (FIB-4 index and APRI) were significantly less than that of ALT (P value < 0.001 and < 0.001, respectively).

Of the 9,748 subjects, 322 (3.30%) participants developed type 2 diabetes. Crude and adjusted RDs and RRs of type 2 diabetes incidence according to ALT categories are shown in Table 3. Baseline ALT level was significantly associated with the incidence of type 2 diabetes in a dose-response manner. The adjusted RRs were 1.79 [95% CI: 1.29, 2.58] and 1.64 [95% CI: 1.17, 2.38] for the highest and middle ALT groups, respectively, in reference to the lowest ALT group. The adjusted RDs were 1.60 % [95% CI: 0.73, 2.42] and 1.29 % [95% CI: 0.41, 2.17], respectively.

Fig. 3 shows the results of sensitivity analysis. The E-value of RR was 2.98 when the bounding factor of RR was 1.79. The RR between unknown confounders and the incidence of type 2 diabetes needs to be at least 2.0 in order to explain the observed statistical RR with 1.79.

Discussion

Of the indicators of liver damage that we examined, ALT was likely to be the most accurate predictor of diabetes incidence, performing even better than most fibrosis scores among health check-up participants. We found a significant positive relationship between the ALT level and the incidence of diabetes, using IPTW. Health care providers may be able to use ALT level to predict future type 2 diabetes, and preventive interventions might be necessary for subjects with high ALT levels.

To the best of our knowledge, this is the first cohort study comparing the performance of various indicators of liver damage, including fibrosis scores, to predict the devel-

Table 1. Demographic characteristics in unweighted (baseline) and weighted (adjusted) cohort members according to ALT category.

Patient Characteristics	Unweighted cohorts ALT level				Weighted cohorts ALT level			
	The Lowest (n = 3,672)	The Middle (n = 3,420)	The Highest (n = 2,656)	SMD	The Lowest (n = 3,689)	The Middle (n = 3,432)	The Highest (n = 2,635)	SMD
Female (%)	44.3	52.2	50.0	0.105	48.4	51.4	50.8	0.040
Mean age, years	52.2 (12.5)	53.7 (11.0)	52.8 (11.0)	0.085	52.6 (12.0)	53.5 (11.1)	52.5 (11.3)	0.057
Body mass index, kg/m²	21.9 (2.5)	22.9 (2.7)	24.3 (3.2)	0.548	23.0 (3.2)	22.8 (2.9)	22.9 (3.2)	0.049
Obesity (%)	11.0	20.2	39.3	0.457	21.2	20.7	23.7	0.048
Hypertension (%)	15.8	20.2	25.6	0.164	21.3	20.3	19.6	0.028
CRP, mg/dl	0.11 (0.32)	0.11 (0.30)	0.15 (0.36)	0.084	0.13 (0.39)	0.11 (0.31)	0.13 (0.32)	0.041
Total cholesterol, mg/dl	196 (33.2)	206 (34.0)	211 (35.6)	0.302	203 (34.2)	205 (34.5)	202 (35.0)	0.045
HDL-cholesterol, mg/dl	60.5 (14.6)	59.5 (15.5)	56.1 (15.2)	0.191	58.9 (14.6)	59.6 (15.6)	59.0 (16.0)	0.029
Triglycerides, mg/dl	84 (63-116)	97 (70-138)	118 (83-172)	0.389	113 (66.9)	115 (80.5)	117 (83.8)	0.034
Fasting glucose, mg/dl	93.4 (8.6)	94.8 (9.0)	97.0 (9.4)	0.268	94.8 (9.2)	94.8 (9.0)	94.8 (9.1)	< 0.001
IFG (%)				0.457				0.030
>= 100, < 110 mg/dl	16.8	19.9	26.9		20.8	19.2	20.4	
>= 110, < 126 mg/dl	3.9	6.8	10.2		6.7	7	6.5	
HbA1c, %	5.26 (0.40)	5.35 (0.41)	5.43 (0.42)	0.279	5.33 (0.42)	5.34 (0.42)	5.33 (0.41)	0.012

Data are means (s.d.), medians (interquartile range), or percentages.

Lowest: ALT < 20 U/L for men and ALT < 14 U/L for women; Middle: ALT ≥ 20, < 30 U/L for men and ALT ≥ 14, < 20 U/L for women; Highest: ALT ≥ 30 U/L for men and ALT ≥ 20 U/L for women.

ALT, alanine aminotransferase; CRP, C-reactive protein; HbA1c, glycated hemoglobin; HDL-cholesterol, high-density lipoprotein-cholesterol; IFG, impaired fasting glucose; SMD, standardized mean difference.
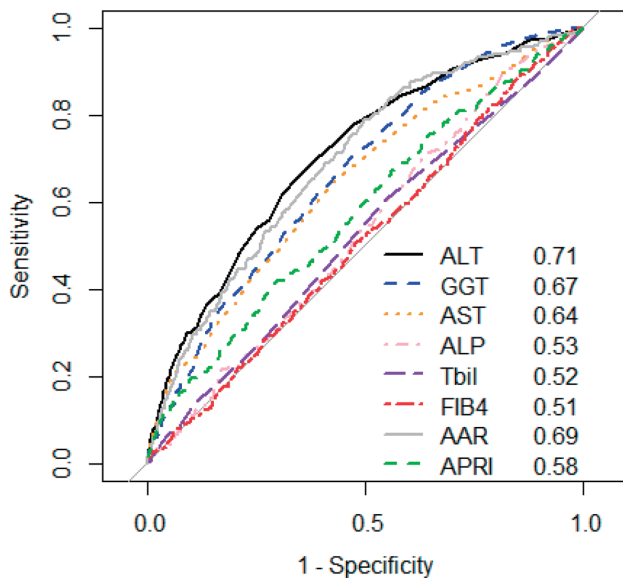

Fig. 2. Receiver operating characteristic (ROC) curves for each liver function indicator as a predictor of diabetes mellitus onset.

The AUC for ALT is higher than the AUCs for any other liver damage indicators, even higher than the AUCs for liver fibrosis scores: FIB4 index, AAR and APRI.

AAR, aspartate aminotransferase/alanine aminotransferase ratio; ALP, alkaline phosphatase; ALT, alanine aminotransferase; APRI, aspartate aminotransferase platelet ratio index; AST, aspartate aminotransferase; AUC, area under the receiver operating characteristic curves; FIB-4, fibrosis-4 index score; GGT, γ-glutamyltransferase; Tbil, total bilirubin.

opment of type 2 diabetes. According to our results, current liver inflammation can possibly be more important than liver fibrosis indicating chronic liver damage in predicting type 2 diabetes. The ROC curves showed that the AUC for ALT was the highest, even compared with the AUCs for fibrosis scores (FIB-4 index and APRI). Although one previous study examined the accuracy of liver enzyme levels in predicting type 2 diabetes incidence, it did not account for fibrosis scores (Doi et al. 2007). Chang et al. (2013) found that fibrosis scores were associated with an increased incidence of type 2 diabetes in subjects with non-alcoholic fatty liver disease diagnosed by ultrasound: the non-alcoholic fatty liver disease fibrosis score, FIB-4 index and APRI. However, they did not compare the accuracy of liver fibrosis scores with that of liver enzymes to predict the development of type 2 diabetes.

Conditional statistical approaches, parametric models, were used in previous research papers studying the association between liver damage and the incidence of diabetes (Nakanishi et al. 2003a, 2004; Okamoto et al. 2003; Doi et al. 2007; Fujita et al. 2010; Hozawa et al. 2010; Kaneko et al. 2019), which can potentially bias the results (Williamson and Ravani 2017). We utilized IPTW analysis, which does not require setting up a specific model (e.g., generalized linear model) between an outcome variable and covariates, which can have less bias (Cepeda et al. 2003). Thus, our result may support the causal effect of ALT on type 2 diabetes incidence in Asians, which was shown in Europeans by the recent Mendelian randomization study (Liu et al. 2016;

Table 2. Area under the receiver operating characteristic curve for predicting diabetes mellitus onset according to each indicator of liver damage.

	AUC (95% CI)	The difference in AUC (95% CI)	P value
ALT	0.71 (0.68 to 0.74)	Reference	-
AST	0.64 (0.61 to 0.67)	–0.06 (–0.08 to –0.04)	< 0.001
GGT	0.67 (0.64 to 0.69)	–0.04 (–0.07 to –0.01)	0.003
ALP	0.53 (0.50 to 0.56)	–0.17 (–0.22 to –0.13)	< 0.001
Total bilirubin	0.52 (0.49 to 0.56)	–0.18 (–0.23 to –0.14)	< 0.001
FIB-4 index	0.51 (0.48 to 0.54)	–0.19 (–0.23 to –0.15)	< 0.001
AAR	0.69 (0.66 to 0.72)	–0.02 (–0.03 to 0.00)	0.097
APRI	0.58 (0.55 to 0.61)	–0.13 (–0.16 to –0.10)	< 0.001

AAR, aspartate aminotransferase/alanine aminotransferase ratio; ALP, alkaline phosphatase; ALT, alanine aminotransferase; APRI, aspartate aminotransferase platelet Ratio Index; AST, aspartate aminotransferase; AUC, area under the receiver operating characteristic curve; CI, confidence interval; FIB-4 index, fibrosis-4 index score; GGT, γ-glutamyl transferase.

Table 3. Crude (unadjusted) and adjusted measures of association between the incidence of diabetes mellitus and ALT categories comparing inverse probability of treatment-weighted cohort members.

	Persons (#)	Incident cases	Cumulative incidence (%)	Crude RD (95% CI) (%)	Adjusted RD (95% CI) (%)	Crude RR (95% CI)	Adjusted RR (95% CI)
ALT level							
Lowest	3,672	55	1.50	reference	reference	reference	reference
Middle	3,420	111	3.25	1.75 (0.92, 2.58)	1.29 (0.41, 2.17)	2.17 (1.58, 3.01)	1.64 (1.17, 2.38)
Highest	2,656	156	5.87	4.38 (3.49, 5.26)	1.60 (0.73, 2.42)	3.92 (2.90, 5.38)	1.79 (1.29, 2.58)

Propensity score for inverse probability of treatment weighting included sex, age, body mass index, hypertension, total-cholesterol, high-density lipoprotein-cholesterol, log transformed triglycerides, C-reactive protein, impaired fasting glucose and glycated hemoglobin level.
ALT, alanine aminotransferase; CI, confidence interval; RD, risk difference; RR, relative risk.

De Silva et al. 2019). The positive association between liver enzyme levels and type 2 diabetes incidence can be partly explained by obesity-related diabetes mellitus. However, even after removing the effect of obesity and other confounders by the IPTW method, the association remained. ALT can be an indicator of liver inflammation. Perhaps the potential causal effect of liver damage in type 2 diabetes incidence can be explained by liver inflammation, which may impair insulin signalling in the liver and other organs (Bonnet et al. 2011). Inflammatory cytokines and oxidative stress are possible factors leading to worsening insulin resistance (Qatanani and Lazar 2007; Dowman et al. 2010). Our ROC curves and AUCs also support the importance of current liver inflammation over liver fibrosis in type 2 diabetes development, as liver enzyme levels had higher AUCs than fibrosis scores. Liver inflammation may be a more direct causal factor than liver fibrosis. Moreover, inflammation indicators were keener in our current study as liver fibrosis scores might not vary among the relatively healthy individuals, most of whom did not have liver fibrosis. Further research is required to discover the mechanism connecting liver damage to type 2 diabetes development.

Our results did not contradict the current trend of liver enzyme cut-off level definitions. The cut-off level tends to decrease because more emphasis is placed on the potential health hazards of lower liver enzymes than previously (Kwo et al. 2017). Even our middle ALT group (over 20 U/L and less than 30 U/L for men, and over 14 U/L and less than 20 U/L for women) had a significantly higher incidence of type 2 diabetes. Medical providers may need to be cautious about ALT level even though it does not exceed the normal range.

One of the limitations of this study is unmeasured potential confounders, especially alcohol consumption. As a result, our inferences may be biased by residual confounders. However, this possibility cannot be high for the following three reasons. First, the association between alcohol consumption and the risk of type 2 diabetes is U-shaped among Japanese people, and is negative in one recent study (Nakanishi et al. 2003b; He et al. 2019). Using no alcohol consumption as a base, the RR for drinkers of ≥ 69.0 g ethanol/day was 0.95 (Nakanishi et al. 2003b). Thus, alcohol may have a relatively negative effect on the incidence of type 2 diabetes for Japanese adults. Thus, alcohol as a confounder causes the statistical result to tend toward the null hypothesis, which would not change our significant associa-
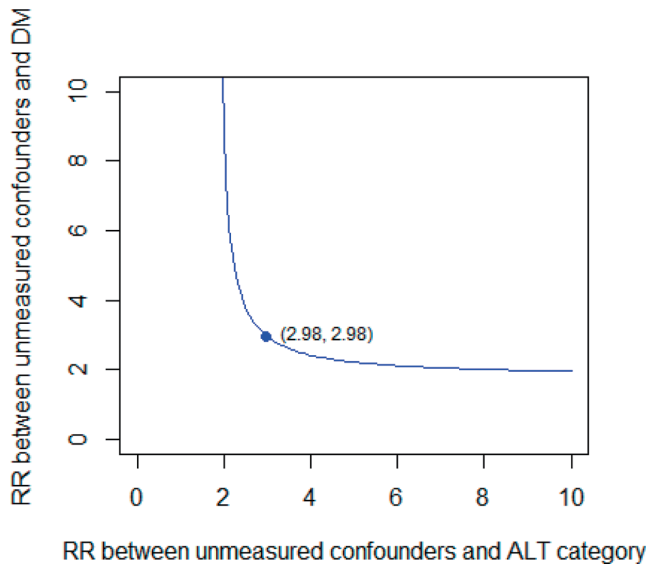
Fig. 3. Sensitivity analysis showing relationship of relative risk of unmeasured confounders with ALT category and relative risk with type 2 diabetes.

The E-value of RR shown as the blue dot was 2.98 when the bounding factor of RR was 1.79. The RR between unknown confounders and the incidence of type 2 diabetes needs to be at least 2.0 in order to explain the observed statistical RR with 1.79.

ALT, alanine aminotransferase; DM, diabetes mellitus; RR, relative risk.

tion. Second, our sensitivity analysis showed that it is unlikely that unmeasured alcohol factors can explain our statistical results. Previous studies showed that the odds ratio for alcoholic liver disease due to alcohol with more than 23g/d was 3.7-7.5 (Brandish and Sheron 2010). Even if the RR of alcoholic liver disease due to alcohol is 7.5, the RR of type 2 diabetes due to alcohol must be more than 2.0 according to our sensitivity analysis, which is a conservative value (Ding and VanderWeele 2016). However, the possibility that alcohol has 2.0 RR is relatively unlikely based on previous research (Nakanishi et al. 2003b; Waki et al. 2005; He et al. 2019). Lastly, direct association between alcohol and the development of type 2 diabetes is thought to be limited. The primary mechanism between them should be through liver damage (a mediator). Thus, we need not take alcohol into account as a confounder when investigating the association between liver damage and the incidence of type 2 diabetes.

There are other possible limitations. While this was a prospective cohort study, the data set does not have data concerning loss to follow up, which could cause selection bias in an unpredictable direction. Some risk factors for type 2 diabetes were not evaluated, such as alcohol use, family history, diet, and physical exercise. The lack of information about alcohol use makes it impossible to distinguish between alcoholic hepatitis and non-alcoholic fatty liver disease. Thus, non-alcoholic fatty liver disease fibrosis score was not used as a fibrosis score in the current study. On the other hand, although a lack of information

about diet and exercise could possibly lead to overestimation of the association between liver damage indicators and type 2 diabetes incidence, these unmeasured potential confounders can be taken into account because their associations to liver damage were conditioned by obesity and other explanatory variables. There were subjects who progressed to type 2 diabetes at subsequent check-ups during the baseline period, who therefore were not eligible to participate during the follow-up period, which would tend to lead to an underestimation of type 2 diabetes incidence during the follow-up period. Lastly, our results may not be broadly applicable as the current study was performed among relatively healthy individuals participating in health check-ups. Fibrosis scores can be an accurate predictor among individuals with liver diseases.

In conclusion, of the indicators of liver damage examined, ALT was likely to be the most accurate at predicting the development of type 2 diabetes and was even better than fibrosis scores. ALT was significantly associated with future incidence of type 2 diabetes, based on IPTW. Health care providers should probably consider intervention for patients with high ALT when monitoring for type 2 diabetes.

Author Contributions

Yu Hatano and Kazuo Inoue conceived of the study. Yu Hatano and Saori Kashima performed the statistical analysis. Kazuo Inoue, Kimihiko Akimoto and Masatoshi Matsumoto helped to draft the manuscript and to interpret the data results. All authors have read and approved the final manuscript.

Conflict of Interest

The authors declare no conflict of interest.

References

American Diabetes Association (2017) 2. Classification and diagnosis of diabetes. *Diabetes Care*, **40**, S11-S24.

Austin, P.C. (2011) An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behav. Res.*, **46**, 399-424.

Bonnet, F., Ducluzeau, P.H., Gastaldelli, A., Laville, M., Anderwald, C.H., Konrad, T., Mari, A. & Balkau, B. ; the RISC Study Group (2011) Liver enzymes are associated with hepatic insulin resistance, insulin secretion, and glucagon concentration in healthy men and women. *Diabetes*, **60**, 1660-1667.

Brandish, E. & Sheron, N. (2010) Drinking patterns and the risk of serious liver disease. *Expert Rev. Gastroenterol. Hepatol.*, **4**, 249-252.

Cales, P., Laine, F., Boursier, J., Deugnier, Y., Moal, V., Oberti, F., Hunault, G., Rousselet, M.C., Hubert, I., Laafi, J., Ducluzeaux, P.H. & Lunel, F. (2009) Comparison of blood tests for liver

fibrosis specific or not to NAFLD. *J. Hepatol.*, **50**, 165-173.

Cepeda, M.S., Boston, R., Farrar, J.T. & Strom, B.L. (2003) Comparison of logistic regression versus propensity score when the number of events is low and there are multiple confounders. *Am. J. Epidemiol.*, **158**, 280-287.

Chang, Y., Jung, H.S., Yun, K.E., Cho, J., Cho, Y.K. & Ryu, S. (2013) Cohort study of non-alcoholic fatty liver disease, NAFLD fibrosis score, and the risk of incident diabetes in a Korean population. *Am. J. Gastroenterol.*, **108**, 1861-1868.

Cho, N.H., Jang, H.C., Choi, S.H., Kim, H.R., Lee, H.K., Chan, J.C. & Lim, S. (2007) Abnormal liver function test predicts type 2 diabetes: a community-based prospective study. *Diabetes Care*, **30**, 2566-2568.

Committee of the Japan Diabetes Society on the Diagnostic Criteria of Diabetes Mellitus ; Seino, Y., Nanjo, K., Tajima, N., Kadowaki, T., Kashiwagi, A., Araki, E., Ito, C., Inagaki, N., Iwamoto, Y., Kasuga, M., Hanafusa, T., Haneda, M. & Ueki, K. (2010) Report of the committee on the classification and diagnostic criteria of diabetes mellitus. *J. Diabetes Investig.*, **1**, 212-228.

De Silva, N.M.G., Borges, M.C., Hingorani, A.D., Engmann, J., Shah, T., Zhang, X., Luan, J., Langenberg, C., Wong, A., Kuh, D., Chambers, J.C., Zhang, W., Jarvelin, M.R., Sebert, S., Auvinen, J., et al. (2019) Liver function and risk of type 2 diabetes: bidirectional Mendelian randomization study. *Diabetes*, **68**, 1681-1691.

Ding, P. & VanderWeele, T.J. (2016) Sensitivity analysis without assumptions. *Epidemiology*, **27**, 368-377.

Doi, Y., Kubo, M., Yonemoto, K., Ninomiya, T., Iwase, M., Tanizaki, Y., Shikata, K., Iida, M. & Kiyohara, Y. (2007) Liver enzymes as a predictor for incident diabetes in a Japanese population: the Hisayama study. *Obesity* (*Silver Spring*), **15**, 1841-1850.

Dowman, J.K., Tomlinson, J.W. & Newsome, P.N. (2010) Pathogenesis of non-alcoholic fatty liver disease. *QJM*, **103**, 71-83.

Ford, E.S., Schulze, M.B., Bergmann, M.M., Thamer, C., Joost, H.G. & Boeing, H. (2008) Liver enzymes and incident diabetes: findings from the European Prospective Investigation into Cancer and Nutrition (EPIC)-Potsdam Study. *Diabetes Care*, **31**, 1138-1143.

Fraser, A., Harris, R., Sattar, N., Ebrahim, S., Davey Smith, G. & Lawlor, D.A. (2009) Alanine aminotransferase, gamma-glutamyltransferase, and incident diabetes: the British Women's Heart and Health Study and meta-analysis. *Diabetes Care*, **32**, 741-750.

Friedman, S.L. (2003) Liver fibrosis : from bench to bedside. *J. Hepatol.*, **38** Suppl 1, S38-53.

Fujita, M., Ueno, K. & Hata, A. (2010) Association of gamma-glutamyltransferase with incidence of type 2 diabetes in Japan. *Exp. Biol. Med. (Maywood)*, **235**, 335-341.

Goessling, W., Massaro, J.M., Vasan, R.S., D'Agostino, R.B. Sr., Ellison, R.C. & Fox, C.S. (2008) Aminotransferase levels and 20-year risk of metabolic syndrome, diabetes, and cardiovascular disease. *Gastroenterology*, **135**, 1935-1944. e1.

Guariguata, L., Whiting, D.R., Hambleton, I., Beagley, J., Linnenkamp, U. & Shaw, J.E. (2014) Global estimates of diabetes prevalence for 2013 and projections for 2035. *Diabetes Clin. Pract.*, **103**, 137-149.

He, X., Rebholz, C.M., Daya, N., Lazo, M. & Selvin, E. (2019) Alcohol consumption and incident diabetes: the Atherosclerosis Risk in Communities (ARIC) study. *Diabetologia*, **62**, 770-778.

Hoelzel, W., Weykamp, C., Jeppsson, J.O., Miedema, K., Barr, J.R., Goodall, I., Hoshino, T., John, W.G., Kobold, U., Little, R., Mosca, A., Mauri, P., Paroni, R., Susanto, F., Takei, I., et al. (2004) IFCC reference system for measurement of hemoglobin A1c in human blood and the national standardization schemes in the United States, Japan, and Sweden: a method-comparison study. *Clin. Chem.*, **50**, 166-174.

Hozawa, A., Okamura, T., Tanaka, T., Miura, K., Kikuchi, Y., Kadowaki, T., Yoshita, K., Takebayashi, T., Tamaki, J., Minai, J., Tada, T., Chiba, N., Okayama, A. & Ueshima, H. (2010) Relation of Gamma-glutamyltransferase and alcohol drinking with incident diabetes: the HIPOP-OHP study. *J. Atheroscler. Thromb.*, **17**, 195-202.

IDF Diabetes Atlas Group (2013) Update of mortality attributable to diabetes for the IDF Diabetes Atlas: estimates for the year 2011. *Diabetes Res. Clin. Pract.*, **100**, 277-279.

Inoue, K., Inoue, M., Matsumoto, M. & Akimoto, K. (2012) Persistent fasting hyperglycaemia is more predictive of type 2 diabetes than transient fasting hyperglycaemia. *Diabet. Med.*, **29**, e75-81.

Inoue, K., Matsumoto, M. & Akimoto, K. (2009) The threshold for definition of impaired fasting glucose in a Japanese population. *Diabet. Med.*, **26**, 1175-1178.

Kaneko, K., Yatsuya, H., Li, Y., Uemura, M., Chiang, C., Hirakawa, Y., Ota, A., Tamakoshi, K. & Aoyama, A. (2019) Association of gamma-glutamyl transferase and alanine aminotransferase with type 2 diabetes mellitus incidence in middle-aged Japanese men: 12-year follow up. *J. Diabetes Investig.*, **10**, 837-845.

Kashiwagi, A., Kasuga, M., Araki, E., Oka, Y., Hanafusa, T., Ito, H., Tominaga, M., Oikawa, S., Noda, M., Kawamura, T., Sanke, T., Namba, M., Hashiramoto, M., Sasahara, T., Nishio, Y., et al. (2012) International clinical harmonization of glycated hemoglobin in Japan: from Japan Diabetes Society to National Glycohemoglobin Standardization Program values. *J. Diabetes Investig.*, **3**, 39-40.

Kawaguchi, T., Taniguchi, E., Itou, M., Sakata, M., Sumie, S. & Sata, M. (2011) Insulin resistance and chronic liver disease. *World J. Hepatol.*, **3**, 99-107.

Kim, C.H., Park, J.Y., Lee, K.U., Kim, J.H. & Kim, H.K. (2009) Association of serum gamma-glutamyltransferase and alanine aminotransferase activities with risk of type 2 diabetes mellitus independent of fatty liver. *Diabetes Metab. Res. Rev.*, **25**, 64-69.

Kuzuya, T., Nakagawa, S., Satoh, J., Kanazawa, Y., Iwamoto, Y., Kobayashi, M., Nanjo, K., Sasaki, A., Seino, Y., Ito, C., Shima, K., Nonaka, K. & Kadowaki, T. ; Committee of the Japan Diabetes Society on the diagnostic criteria of diabetes mellitus (2002) Report of the Committee on the classification and diagnostic criteria of diabetes mellitus. *Diabetes Res. Clin. Pract.*, **55**, 65-85.

Kwo, P.Y., Cohen, S.M. & Lim, J.K. (2017) ACG Clinical Guideline: evaluation of abnormal liver chemistries. *Am. J. Gastroenterol.*, **112**, 18-35.

Li, Y., Wang, J., Han, X., Hu, H., Wang, F., Yu, C., Yuan, J., Yao, P., Li, X., Yang, K., Miao, X., Wei, S., Wang, Y., Chen, W., Liang, Y., et al. (2019) Serum alanine transaminase levels predict type 2 diabetes risk among a middle-aged and elderly Chinese population. *Ann. Hepatol.*, **18**, 298-303.

Liu, J., Au Yeung, S.L., Lin, S.L., Leung, G.M. & Schooling, C.M. (2016) Liver enzymes and risk of ischemic heart disease and type 2 diabetes mellitus: a Mendelian randomization study. *Sci. Rep.*, **6**, 38813.

Nakanishi, N., Nishina, K., Li, W., Sato, M., Suzuki, K. & Tatara, K. (2003a) Serum gamma-glutamyltransferase and development of impaired fasting glucose or type 2 diabetes in middle-aged Japanese men. *J. Intern. Med.*, **254**, 287-295.

Nakanishi, N., Suzuki, K. & Tatara, K. (2003b) Alcohol consumption and risk for development of impaired fasting glucose or type 2 diabetes in middle-aged Japanese men. *Diabetes Care*, **26**, 48-54.

Nakanishi, N., Suzuki, K. & Tatara, K. (2004) Serum gamma-glutamyltransferase and risk of metabolic syndrome and type 2 diabetes in middle-aged Japanese men. *Diabetes Care*, **27**, 1427-1432.

Okamoto, M., Takeda, Y., Yoda, Y., Kobayashi, K., Fujino, M.A. &

Yamagata, Z. (2003) The association of fatty liver and diabetes risk. *J. Epidemiol.*, **13**, 15-21.

Qatanani, M. & Lazar, M.A. (2007) Mechanisms of obesity-associated insulin resistance: many choices on the menu. *Genes Dev.*, **21**, 1443-1455.

Robins, J.M., Hernan, M.A. & Brumback, B. (2000) Marginal structural models and causal inference in epidemiology. *Epidemiology*, **11**, 550-560.

Shah, A.G., Lydecker, A., Murray, K., Tetri, B.N., Contos, M.J. & Sanyal, A.J. ; Nash Clinical Research Network (2009) Comparison of noninvasive markers of fibrosis in patients with nonalcoholic fatty liver disease. *Clin. Gastroenterol. Hepatol.*, **7**, 1104-1112.

Vallet-Pichard, A., Mallet, V., Nalpas, B., Verkarre, V., Nalpas, A., Dhalluin-Venier, V., Fontaine, H. & Pol, S. (2007) FIB-4: an inexpensive and accurate marker of fibrosis in HCV infection. comparison with liver biopsy and fibrotest. *Hepatology*, **46**, 32-36.

Wai, C.T., Greenson, J.K., Fontana, R.J., Kalbfleisch, J.D., Marrero, J.A., Conjeevaram, H.S. & Lok, A.S. (2003) A simple noninvasive index can predict both significant fibrosis and cirrhosis in patients with chronic hepatitis C. *Hepatology*, **38**, 518-526.

Waki, K., Noda, M., Sasaki, S., Matsumura, Y., Takahashi, Y., Isogawa, A., Ohashi, Y., Kadowaki, T. & Tsugane, S. ; JPHC Study Group (2005) Alcohol consumption and other risk factors for self-reported diabetes among middle-aged Japanese: a population-based prospective study in the JPHC study cohort I. *Diabet. Med.*, **22**, 323-331.

Williamson, T. & Ravani, P. (2017) Marginal structural models in clinical research: when and how to use them? *Nephrol. Dial. Transplant.*, **32**, ii84-ii90.

World Health Organization (2016) Global Report on Diabetes. http://apps.who.int/iris/bitstream/10665/204871/1/978924156 5257_eng.pdf [*Accessed*: October 10, 2019].

Xu, L., Jiang, C.Q., Schooling, C.M., Zhang, W.S., Cheng, K.K. & Lam, T.H. (2015) Liver enzymes and incident diabetes in China: a prospective analysis of 10 764 participants in the Guangzhou Biobank Cohort Study. *J. Epidemiol. Community Health*, **69**, 1040-1044.

Xu, S., Ross, C., Raebel, M.A., Shetterly, S., Blanchette, C. & Smith, D. (2010) Use of stabilized inverse propensity scores as weights to directly estimate relative risk and its confidence intervals. *Value Health*, **13**, 273-277.

Zhang, J., Cheng, N., Ma, Y., Li, H., Cheng, Z., Yang, Y., He, C., Li, J., Pu, H., Shen, X., Ren, X., Shi, D., Pu, R., Gan, T., Ding, J., et al. (2018) Liver enzymes, fatty liver and type 2 diabetes mellitus in a Jinchang Cohort: a prospective study in adults. *Can. J. Diabetes*, **42**, 652-658.