Supplementary File S4. Secondary and sensitivity analyses

1. Secondary analysis

The time-to-event from inclusion in the cohort to cirrhosis (defined as date of cirrhosis minus date of inclusion in cohort + 1 day) was assessed based on the Kaplan-Meier product-limit estimator. TTE was right-censored at the end of follow-up (date of end of follow-up minus the date of inclusion in cohort + 1 day) and any subject with a missing event date or event before inclusion in the cohort was excluded. The crude hazard ratios (HR) were estimated, based on a Cox proportional hazards regression model, overall and by patient characteristics. HRs were also calculated by adjusting for calendar year and patient characteristics. Results are shown below.

Table S4_1. Results of the secondary analyses

	Hazard ratio (95% CI)	
Compared cohorts	Unadjusted[a]	Adjusted[b]
Cirrhosis		
T2DM+HBV vs T2DM	11.620 (3.729–36.211)	14.290 (4.546–44.921)
T2DM+HBV vs HBV	1.334 (0.408–4.364)	0.690 (0.208–2.288)
Hepatocellular		
T2DM+HBV vs T2DM	1.732 (0.649–4.617)	2.850 (1.067–7.610)
T2DM+HBV vs HBV	3.599 (1.194–10.843)	1.403 (0.465–4.237)

CI, confidence interval

Note: [a] Unadjusted 95% CI hazard ratio based on Cox PH regression model;

[b] Adjusted 95% CI hazard ratio based on Cox PH regression model adjusted for calendar year, age group, sex, ethnicity, body mass index, tobacco smoking status, dyslipidemia and HBV vaccination.

2. Sensitivity analyses

The cohorts included in each analysis are presented in Table S4_2.

Table S4_2. Definition of cohorts

Cohorts	Definitions	Sensitivity analysis
Cohort 1 (T2DM+HBV cohort)	T2DM (new onset or prevalent) HBV-infection (new onset or prevalent)	3
Cohort 2 (T2DM cohort)	T2DM (new onset or prevalent) HBV-free (throughout CPRD-GOLD and HES APC health records)	1a, 1b
Cohort 3 (HBV cohort)	T2DM-free (throughout CPRD-GOLD health records) HBV-infection (new onset or prevalent)	2
Cohort 4 (T2DM+NAFLD+HBV cohort)	T2DM (new onset or prevalent) HBV-infection (new onset or prevalent) NAFLD diagnosis (anytime during CPRD-GOLD or HES APC health records, any stage)	2
Cohort 5 (T2DM+NAFLD cohort)	T2DM (new onset or prevalent) HBV-free (throughout CPRD-GOLD and HES APC health records) NAFLD diagnosis (anytime during CPRD-GOLD or HES APC health records, any stage)	
Cohort 6 (NAFLD+HBV cohort)	T2DM-free (throughout CPRD-GOLD health records) HBV-infection (new onset or prevalent) NAFLD diagnosis (anytime during CPRD-GOLD and HES APC health records, any stage)	
Cohort 7	T2DM (new onset or prevalent) Chronic HBV-infection (new onset or prevalent)	1a
Cohort 8	T2DM-free (throughout CPRD-GOLD health records) Chronic HBV-infection (new onset or prevalent)	1a
Cohort 9	T2DM (new onset or prevalent) HBV-infection (new onset or prevalent), including HBsAg presence in the definition	1b
Cohort 10	T2DM-free (throughout CPRD-GOLD health records) HBV-infection (new onset or prevalent), including HBsAg presence in the definition	1b
Cohort 11	T2DM (new onset or prevalent) HBV-free (throughout CPRD-GOLD and HES APC health records) or HBV-free before inclusion in cohort 1 (at HBV onset)	3
Cohort 12	T2DM-free (throughout CPRD-GOLD health records) or T2DM-free before inclusion in cohort 1 (at T2DM onset) HBV-infection (new onset or prevalent)	3
Cohort 13	T2DM (new onset or prevalent) HBV-infection (new onset or prevalent) Being matched to subjects with T2DM (new onset or prevalent) and without HBV-infection (throughout CPRD-GOLD and HES APC health records)	4
Cohort 14	T2DM (new onset or prevalent) HBV-free (throughout CPRD-GOLD and HES APC health records)	4

	Having been matched to subjects with T2DM (new onset or prevalent) and HBV-infection (new onset or prevalent)	
Cohort 15	T2DM (new onset or prevalent) HBV-infection (new onset or prevalent) Without applying medical history exclusion criteria	5a
Cohort 16	T2DM (new onset or prevalent) HBV-free (throughout CPRD-GOLD and HES APC health records) Without applying medical history exclusion criteria	5a
Cohort 17	T2DM-free (throughout CPRD-GOLD health records) HBV-infection (new onset or prevalent) Without applying medical history exclusion criteria	5a
Cohort 18	T2DM (new onset or prevalent) HBV-infection (new onset or prevalent) Applying modified medical history exclusion criteria	5b
Cohort 19	T2DM (new onset or prevalent) HBV-free (throughout CPRD-GOLD and HES APC health records) Applying modified medical history exclusion criteria	5b
Cohort 20	T2DM-free (throughout CPRD-GOLD health records) HBV-infection (new onset or prevalent) Applying modified medical history exclusion criteria	5b

1a. This analysis only included **chronic HBV-infection**, rather than any HBV-infection, as HBV exposure and compared the IR of developing cirrhosis by estimating IRRs between cohorts 7 and 2 and 7 and 8.

1b. This analysis included HBV infection as defined by **HBsAg seropositivity**, as HBV exposure and compared the IR of developing cirrhosis by estimating IRRs between cohorts 9 and 2 and 9 and 10.

2. This analysis was based on the secondary analysis in sub-population with NAFLD diagnosis, but used a **T2DM-free cohort** from the primary analysis (not requiring NAFLD) as comparator and compared the IR of developing cirrhosis by estimating IRRs between cohorts 4 and 3.

3. This analysis allowed patients to contribute person-time to either the T2DM or the HBV cohort, before being included in the T2DM+HBV cohort, allowing for **time-varying T2DM and HBV-infection**, and compared the IR of developing cirrhosis by estimating IRRs between cohorts 1 and 11 and 1 and 12.

4. This analysis included **matched sub-cohorts**, obtained from cohorts 1 and 2 by matching on calendar year, age group, sex, ethnicity, T2DM exposure status (new-onset vs. prevalent) and diabetes complications severity index complications and disease progression stage. Matching ratios 1:1, 1:2 and 1:3 were considered, i.e. matching one, two or three subjects to each index case was attempted, after which the optimal matching ratio of 1:3 was chosen primarily based on the proportion matched and the analysis performed for this single ratio only. The analysis was performed to assess the potential impact of the T2DM progression stage, complications and severity, in particular in prevalent T2DM subjects at the time of study inclusion, as a confounder of the effect of HBV on the progression of NAFLD to cirrhosis. It compared the IR of developing cirrhosis by estimating IRRs between cohorts 13 and 14.

5. Two sensitivity analyses were performed to assess the potential impact of the strict exclusion criteria applied anytime during the subjects follow up to exclude potential alternative aetiology of liver disease that would confound the associated of NAFLD with cirrhosis, namely alcoholic liver disease. A variation of sensitivity analysis 5 (5b) was applied where the exclusion criteria related to alcoholic liver disease were applied at cohort inclusion date instead of never (5a) or throughout the entire follow up period (primary aim analysis).

Sensitivity analysis 5a- assessed the impact of the exclusion criteria on the primary aim analyses, by **not applying** the exclusion criteria related to medical-history of liver disease (see Supplementary File S2, Table 1_14_a and 1_14_b, Variable: ClassType). In practice, the exclusion criteria were *not* applied to Cohorts 1, 2 and 3 resulting in the extended cohorts 15, 16 and 17, including in addition:

- Subjects with the following liver diseases ***prior to*** *cohort inclusion date or with missing event date*: liver transplantation, liver cancer, liver failure.

- Subjects with other liver diseases (see Supplementary File S2, Table 1_14_a and 1_14_b, Variable: ClassType), ***anytime*** *during their entire medical record* including: alcoholic liver disease, autoimmune liver disease, any viral hepatitis other than HBV, hemochromatosis.

Sensitivity analysis 5b- assessed the impact of the exclusion criteria on the primary aim analyses, by **applying modified** cohort-specific exclusion criteria related to medical-history of liver disease (see Supplementary File S2, Table 1_14_a and 1_14_b, Variable: ClassTypeAlt). In practice, the exclusion criteria were applied in a modified version to Cohorts 1, 2 and 3 resulting in the extended cohorts 18, 19 and 20, namely:

- Subjects with the following liver diseases ***prior to*** *cohort inclusion date or with missing event date*: liver transplantation, liver cancer, alcoholic liver disease, liver failure.

- Subjects with other liver diseases, ***anytime*** *during their entire medical record* including: autoimmune liver disease, any viral hepatitis other than HBV, hemochromatosis.

Table S4_3. Results of sensitivity analyses estimating risk of cirrhosis

		Incidence rate ratio (95% CI)	
Sensitivity analysis	Compared cohorts	Unadjusted[a]	Adjusted[b]
1a	7 vs 2	34.98 (7.18–103.00)	50.24 (15.79–159.84)
	7 vs 8	1.63 (0.31–5.44)	0.91 (0.27–3.09)
1b	9 vs 2	5.42 (1.11–15.97)	9.69 (3.08–30.51)
	9 vs 10	1.85 (0.36–5.92)	0.78 (0.23–2.58)
2	4 vs 3	0.00 (0.00–18.47)[a]	-
3	1 vs 11	10.83 (2.22–31.90)	13.97 (4.44–43.91)
	1 vs 12	1.34 (0.26–4.29)	0.71 (0.22–2.37)
4	13 vs 14	405184515947.528[b]	-
5a	15 vs 16	18.43 (11.77–27.54)	19.72 (13.09–29.71)
	15 vs 17	2.91 (1.78–4.57)	1.89 (1.20–2.96)
5b	18 vs 19	12.22 (4.47–26.74)	14.83 (6.59–33.38)
	18 vs 20	1.81 (0.63–4.25)	1.02 (0.43–2.42)

Note: [a] No cases were reported in cohort 4.

[b] No cases were reported in cohort 14.