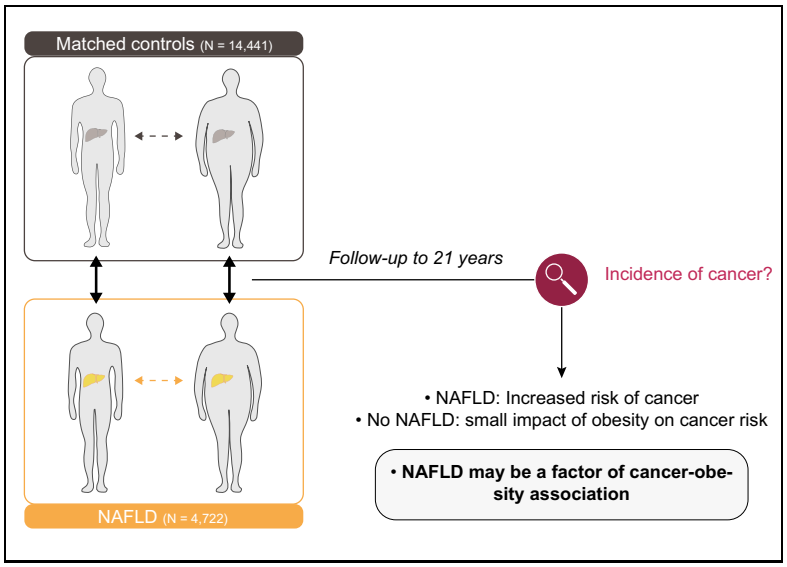
Research Article
Genetic and Metabolic Diseases

JOURNAL
OF HEPATOLOGY

The risk of incident extrahepatic cancers is higher in non-alcoholic fatty liver disease than obesity – A longitudinal cohort study

Graphical abstract



Highlights

- NAFLD is associated with a nearly 2-fold increase in the overall risk of incident cancers.
- The highest risk was noted in liver, uterine, stomach, pancreas and colon cancers.
- Obesity in the absence of NAFLD had minimal impact on malignancy risk.

Authors

Alina M. Allen, Stephen B. Hicks, Kristin C. Mara, Joseph J. Larson, Terry M. Therneau

Correspondence

Allen.alina@mayo.edu
(A.M. Allen)

Lay summary

We studied the incidence of malignancies in a community cohort of adults with non-alcoholic fatty liver disease (NAFLD) in reference to age- and sex-matched adults without NAFLD. After 21 years of longitudinal follow-up, NAFLD was associated with a nearly 2-fold increase in the risk of developing cancers, predominantly of the liver, gastrointestinal tract and uterus. The association with increased cancer risk was stronger in NAFLD than obesity.

http://dx.doi.org/10.1016/j.jhep.2019.08.018
© 2019 European Association for the Study of the Liver. Published by Elsevier B.V. All rights reserved. J. Hepatol. 2019, 71, 1229–1236

Research Article
Genetic and Metabolic Diseases

JOURNAL OF HEPATOLOGY

The risk of incident extrahepatic cancers is higher in non-alcoholic fatty liver disease than obesity – A longitudinal cohort study

Alina M. Allen[1,*], Stephen B. Hicks[2], Kristin C. Mara[3], Joseph J. Larson[3], Terry M. Therneau[3]

[1]*Division of Gastroenterology and Hepatology, Mayo Clinic, Rochester, MN, USA;* [2]*Department of Internal Medicine, Mayo Clinic, Rochester MN, USA;* [3]*Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN, USA*

See Editorial, pages 1073–1075

Background & Aims: Cancer is a major cause of death in patients with non-alcoholic fatty liver disease (NAFLD). Obesity is a risk factor for cancers; however, the role of NAFLD in this association is unknown. We investigated the effect of NAFLD versus obesity on incident cancers.

Methods: We identified all incident cases of NAFLD in a US population between 1997-2016. Individuals with NAFLD were matched by age and sex to referent individuals from the same population (1:3) on the index diagnosis date. We ascertained the incidence of cancer after index date until death, loss to follow-up or study end. NAFLD and cancer were defined using a code-based algorithm with high validity and tested by medical record review. The association between NAFLD or obesity and cancer risk was examined using Poisson regression.

Results: A total of 4,722 individuals with NAFLD (median age 54, 46% male) and 14,441 age- and sex-matched referent individuals were followed for a median of 8 (range 1–21) years, during which 2,224 incident cancers occurred. NAFLD was associated with 90% higher risk of malignancy: incidence rate ratio (IRR) = 1.9 (95% CI 1.3–2.7). The highest risk increase was noted in liver cancer, IRR = 2.8 (95% CI 1.6–5.1), followed by uterine IRR = 2.3 (95% CI 1.4–4.1), stomach IRR = 2.3 (95% CI 1.3–4.1), pancreas IRR = 2.0 (95% CI 1.2–3.3) and colon cancer IRR = 1.8 (95% CI 1.1–2.8). In reference to non-obese controls, NAFLD was associated with a higher risk of incident cancers (IRR = 2.0, 95% CI 1.5–2.9), while obesity alone was not (IRR = 1.0, 95% CI 0.8–1.4).

Conclusions: NAFLD was associated with increased cancer risk, particularity of gastrointestinal types. In the absence of NAFLD, the association between obesity and cancer risk is small, suggesting that NAFLD may be a mediator of the obesity-cancer association.

Lay summary: We studied the incidence of malignancies in a community cohort of adults with non-alcoholic fatty liver disease (NAFLD) in reference to age- and sex-matched adults without NAFLD. After 21 years of longitudinal follow-up, NAFLD was associated with a nearly 2-fold increase in the risk of developing cancers, predominantly of the liver, gastrointestinal tract and uterus. The association with increased cancer risk was stronger in NAFLD than obesity.

© 2019 European Association for the Study of the Liver. Published by Elsevier B.V. All rights reserved.

Introduction

Cancer is a major cause of death in the United States and worldwide.[1,2] Numerous meta-analyses support the link between the risk of malignancy and excess body weight.[3–5] Some associations are flawed because of bias that exaggerates the effect of obesity on cancer incidence, but strong evidence supports this association with 11 cancers, predominantly among digestive organs and hormone-related malignancies in women.[4]

The prevalence of obesity has more than doubled in the last 4 decades[6,7] and as a result, the incidence of non-alcoholic fatty liver disease (NAFLD) has increased substantially.[8–10] Large population studies have clearly established that malignancy is among the top 2 causes of death in NAFLD, vastly surpassing liver-related mortality, which occurs in 1–2% of patients.[11,12] However, the specific types of cancer that patients with NAFLD are at increased risk for, or the magnitude of risk compared to those without NAFLD is not known. Moreover, it is not clear whether there are particular characteristics of malignancy risk among those with NAFLD that are distinct from those with obesity alone. Such data have important implications in patient education, counseling and the application of screening strategies in this high-risk population.

We aimed to analyze the incidence of the most common cancer types in NAFLD in reference to a control population. Second, we aimed to investigate the association between cancer and NAFLD versus obesity alone. To answer these questions, we used a medical record-linkage system that includes prospectively acquired information on the healthcare of all residents in a well-defined population with extended longitudinal follow-up. Population-based research is a major source of evidence to support medical and public health practices.

Patients and methods

Study population

We constructed a historical cohort of all adults diagnosed with NAFLD in Olmsted County, Minnesota between 1997 and 2016.

Keywords: Natural history; Outcomes; NAFLD; NASH; Epidemiology.
Received 18 March 2019; received in revised form 22 July 2019; accepted 17 August 2019; available online 27 August 2019

* Corresponding author. Address: Division of Gastroenterology and Hepatology, Mayo Clinic, 200 First Street SW, Rochester, MN, USA. Tel.: +1 507 284 3917; fax: +1 507 284 0538.
E-mail address: Allen.alina@mayo.edu (A.M. Allen).

Each patient was individually matched by age (±1 year) and sex to 3 individuals who resided in Olmsted County at the time of the index diagnosis date, who did not carry a diagnosis of NAFLD. To identify these 2 groups, we used the medical record-linkage system of the Rochester Epidemiology Project (REP). REP is a unique research infrastructure which links and indexes the medical records of virtually all individuals who have resided in Olmsted County, Minnesota, regardless of age, sex, ethnicity, disease status, socio-economic or insurance status.[13] The REP links together the medical records of persons from 65 different health care providers, including Mayo Clinic, Olmsted Medical Center, Rochester Family Medicine Clinic and other health care facilities.[14] The data available electronically include demographic characteristics, medical diagnostic codes and services, surgical procedure codes, laboratories, drug prescriptions and death information. In addition, for each resident, the system keeps a complete list of all paper and electronic records and scanned documents that are available in full text for in-depth review and abstraction.[15] Each time an Olmsted County resident visits a health care provider, the information from that clinical visit is automatically integrated into the REP research infrastructure. Of all participants, 93% had at least 1 follow-up visit within 3 years, and only 4% were never seen again after the baseline visit.[13,15] This comprehensive medical record-linkage system provides an optimal sampling framework for epidemiologic studies.

NAFLD was ascertained using a code-based algorithm described in previous epidemiologic studies of NAFLD in this community.[8] Briefly, we used the NAFLD-specific Hospital International Classification of Diseases Adapted (HICDA) codes, a system developed at Mayo Clinic for research diagnosis coding and adapted by REP in 1976: HICDA 05710421 (fatty liver), 05710431 (non-alcoholic steatohepatitis). Additionally, the International Classification of Diseases (ICD) codes ICD 9-CM 571.5 (cirrhosis of the liver without mention of alcohol), 571.8 (other chronic non-alcoholic liver disease), 571.9 (unspecified chronic liver disease without mention of alcohol) and ICD10-CM K75.81 (non-alcoholic steatohepatitis) and K76.0 (fatty liver, NOS) were used. From this initial cohort, we excluded individuals with other etiologies of liver disease identified by codes prior to the index NAFLD diagnosis or during the following year (list of codes in Table S1). All study participants (among both NAFLD and referent groups) with less than 1-year of follow-up were excluded in order to avoid bias. In-depth chart review of a random 10% sample of this cohort showed that this algorithm identified NAFLD with a positive and negative predictive value of 85% and 87%, respectively.

Outcomes and covariates

The NAFLD and matched referent individuals were followed prospectively until death, last medical visit, end of Olmsted County residency or June 2018. Primary outcomes were incident cancers documented after the index NAFLD diagnosis or referent matching date. Cancers documented prior to the diagnosis of NAFLD were not included. The cancers of interest were the most common solid cancers, which were classified into 3 groups: liver and gastrointestinal (colon, esophageal, gastric and pancreatic) cancers; hormone-sensitive cancers (breast, uterine/endometrial, ovarian and prostate); and lung cancer. The cancer ascertainment occurred in 2 steps. First, cancer diagnoses were identified in the medical record-linkage system using the codes shown in Table S2; to minimize spurious diagnoses, for each cancer type, the case was ascertained by the presence of at least 2 codes documented at separate dates, at least 30 days apart. Subsequently, a physician (AMA) reviewed the complete medical records of each individual with gastrointestinal and liver cancer codes and of a 10% random sample of the remaining cancer types to confirm the validity of the outcome ascertainment algorithm. To avoid immortal time bias, for analysis any cancer diagnosis was ascertained if it occurred on the day it was confirmed (after the second code). Only 16 individuals died within 30 days from the first cancer code, in whom a second cancer code was not confirmed due to death, therefore the bias risk is negligible.

Covariates of interest included body mass index (BMI), diabetes mellitus, hypertension, dyslipidemia and smoking status at the time of diagnosis or matching. Comorbidities were defined based on combinations of ICD 9 and 10 or HICDA codes (Table S3), medications (Table S4) and laboratory values, as follows: diabetes mellitus, diagnostic codes plus medications or laboratory values (fasting glucose ≥126 mg/dl or hemoglobin A1c ≥6.5%); dyslipidemia, diagnostic codes plus medications or laboratory values (low-density lipoprotein cholesterol >100 mg/dl or triglycerides >150 mg/dl); hypertension, diagnostic codes plus medications. Cirrhosis was defined as the presence of an ICD 9/10 or HICDA code plus Fib-4 score >2.67.

Statistical analysis

Baseline characteristics were compared between groups using Kruskal-Wallis and chi-square methods for continuous and categorical variables, respectively. The incidence of cancer was assessed using Poisson regression. In order to best capture the dependence of cancer rates on age and sex, the model treated these covariates as multipliers of age/sex specific rates obtained from the Iowa Surveillance, Epidemiology and End Results (SEER) registry, which is in the closest geographic proximity to Olmsted County. Cancer types were identified in SEER using the codes listed in Table S2. A primary advantage of this approach is that it allows data analysis across a wide age range without having to recreate the age/sex shape of each underlying incidence curve. The model fit is equivalent to a Cox model, but with a known baseline hazard. The coefficients of the model are hazard ratios and can be interpreted in the same way as those from a Cox model. Formally, the fit uses Poisson regression with the expected number of cases in each age/sex stratum as the reference, where the expected number of cases is the product of the total years of observed follow-up in an age stratum multiplied by the SEER rate for that age stratum.[16] We report the absolute incidence rates at the arbitrary age of 65, because it is near the median age at diagnosis for most cancer types.

To study the effect of NAFLD on cancer incidence we used a hierarchical Poisson regression model, treating the NAFLD impact on each cancer type as a random effect. A primary benefit of this approach is that it allows researchers to study not only the variable impact of NAFLD on different types of cancer, while stabilizing the estimates for the infrequent cancers, but it also provides an average estimate of effect on overall malignancy risk. The results are reported as incidence rate ratios (IRR), which are interpreted like the hazard ratios of a Cox model, with 95% CIs. The models are adjusted for age and sex. Because diabetes mellitus can be a confounder due to its association with both NAFLD and malignancy, a sensitivity analysis adjusting for diabetes mellitus was performed to assess its effect on the association. A similar secondary analysis was performed to adjust for cirrhosis as a confounder. The model was fit

using the JAGS software package via the rjags interface to R.[17] More details are described in the supplementary methods.

Next, we examined the effect of NAFLD on malignancy risk compared to that of obesity irrespective of NAFLD presence (defined by BMI ≥30 kg/m^2), by repeating the above analysis in 2 ways. First, we examined the effect of obesity irrespective of NAFLD status, by splitting the total cohort in 2 groups: non-obese and obese. The malignancy risk associated with obesity was reported as the IRR of cancer in obese compared to non-obese individuals. Next, we repeated the analysis using 3 groups: NAFLD, obese without NAFLD and non-obese without NAFLD, using the latter group as a reference. The results were reported as IRR in NAFLD versus obese without NAFLD, NAFLD versus non-obese without NAFLD and obese versus non-obese.

Statistical analyses were performed in SAS v9.4 (SAS Institute; Cary, NC) and R statistical software, version 3.2.0 (R Foundation for Statistical Computing, Vienna). The study was approved by the Institutional Review Boards of Mayo Clinic and Olmsted Medical Center. All study patients provided research authorization.

Results

A total of 7,413 individuals met the inclusion criteria for NAFLD diagnosis (study flow chart in Fig. 1). Of these 2,691 were excluded due to previous or concurrent liver disease diagnoses of other etiology (within 1 year of the NAFLD code), inconclusive Olmsted County residency or less than 1 year of follow-up. The final NAFLD cohort consisted of 4,722 individuals (median age 54, 46% male). The prevalence of diagnosed NAFLD over the study period in this population was 8%. An age- and sex-matched cohort of 14,441 adults from the general population was identified. Compared to controls, individuals with NAFLD had a higher proportion of obesity (66% *vs.* 35%), diabetes mellitus (33% *vs.* 9%), hypertension (46% *vs.* 26%) and dyslipidemia (59% *vs.* 33%) (Table 1).

A total of 2,224 incident cancers (656 in NAFLD and 1,568 in controls) were identified after NAFLD diagnosis/matching during a median follow-up of 8 (range 1–21) years. The top 3 most common types of cancer in NAFLD and controls were breast, prostate and colon (Table 1). Of all malignancies, the proportion of gastrointestinal/liver cancers was higher in the NAFLD group than the referent group (27% *vs.* 18% of all cancers).

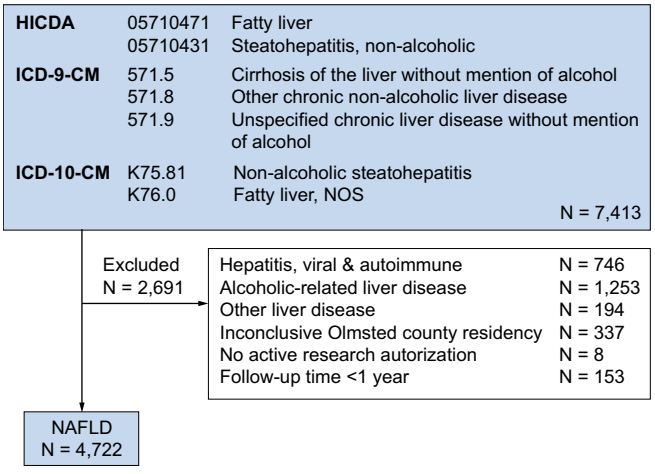


Fig. 1. Flowchart of identification of individuals with NAFLD in the medical record-linkage system. NAFLD, non-alcoholic fatty liver disease.

The incidence of cancers in NAFLD and controls by cancer type at an arbitrary age of 65 is shown in Table 2. There were no statistically significant differences in the rates of malignancy (overall or by cancer type) between the Olmsted County control population and SEER database (Fig. S1).

The effect of NAFLD on the malignancy risk is shown in Fig. 2A. In reference to age- and sex-matched controls, NAFLD was associated with 90% higher overall risk of malignancy: IRR = 1.9 (95% CI 1.3–2.7). Adjustment for cirrhosis status at index or at any point during follow-up did not considerably impact the overall malignancy risk: IRR = 1.8 (95% CI 1.3–2.7). The highest increase in risk was noted in liver cancer, IRR = 2.8 (95% CI 1.6–5.1), followed by uterine IRR = 2.3 (95% CI 1.4–4.1), stomach IRR = 2.3 (95% CI 1.3–4.1), pancreas IRR = 2.0 (95% CI 1.2–3.3) and colon cancer IRR = 1.8 (95% CI 1.1–2.8).

Fig. 2B illustrates that the effect of NAFLD on malignancy risk varied by sex. Most of the differences in rate ratios between women or men with NAFLD and their control counterparts were minor and within the margin of random variation. However, a notable difference in risk was found in colon cancer, which was higher in NAFLD versus referent men IRR = 2.4 (95% CI 1.6–3.9), but not in women IRR = 1.3 (95% CI 0.8–2.1).

Compared to women with NAFLD, men with NAFLD were 90% more likely to develop colon cancer (IRR = 1.9, 95% CI 1.3–2.8). Therefore, the cancer risk hierarchy in NAFLD varies by sex. The highest risk in NAFLD men occurs in colon cancer (IRR = 2.4, 95% CI 1.6–3.9), followed by liver (IRR = 2.3, 95% CI 1.4–4.1), stomach (IRR = 2.0, 95% CI 1.2–3.6) and pancreas cancer (IRR = 1.9, 95% CI 1.1–3.3), whereas in women with NAFLD the highest risk increase occurs in liver (IRR = 2.5, 95% CI 1.4– 4.8), stomach (IRR = 2.2, 95% CI 1.3–4.3), uterus (IRR = 2.2, 95% CI 1.4–3.8) and pancreas (IRR = 2.0, 95% CI 1.2–3.4) cancer.

We also examined whether the effect of NAFLD on cancer risk was greater at younger or older ages. Fig. 3 illustrates the cumulative incidence of cancers in NAFLD versus referent cohort on an age scale. The most notable differences were in pancreas, colon and ovarian cancer, which occurred more commonly in NAFLD at a young age. We further analyzed the age effect using the Poisson regression model, which showed that the risk of incident cancer in NAFLD versus controls decreased with age, and was therefore higher at younger ages, in pancreas (IRR = 0.85, 95% CI 0.74–0.98), colon (IRR = 0.93, 95% CI 0.87–1.00) and ovarian (IRR = 0.86, 95% CI 0.75–0.98) cancer (Fig. S2).

The effect of NAFLD versus obesity on malignancy risk

To examine the effect of obesity on malignancy risk, we first analyzed the cancer rates in the community among 2 groups: obese referenced to non-obese, defined based on BMI ≥30 and <30 kg/m^2, respectively. The distribution of BMI groups among individuals with NAFLD was 10% normal BMI, 27% overweight and 63% obese. The distribution of BMI among controls was 30%, 35% and 34%, respectively. Fig. 4A shows that obesity is associated with a trend towards increased malignancy risk: IRR = 1.2 (95% CI 0.9–1.6). Next, we selected those with NAFLD from the community (66% of them derived from the obese group, while 34% of them were derived from the non-obese group) and performed the same analysis among 3 groups: NAFLD, obese controls and non-obese controls. In Fig. 4B, in reference to non-obese controls, NAFLD was associated with a

Table 1. Baseline characteristics and number of incident cancers in NAFLD and referent individuals.

	NAFLD (n = 4,722)	Referent cohort (n = 14,441)
Characteristics at baseline		
Age–median	54	53
IQR	42.3–64.0	43.0–64.0
Male	46%	47%
Body mass index–median	32	29
IQR	28.6–37.6	24.4–31.9
Obese	3,001 (66%)	4,616 (35%)
Diabetes mellitus	1,547 (33%)	1,348 (9%)
Hypertension	2,181 (46%)	3,682 (26%)
Dyslipidemia	2,769 (59%)	4,829 (33%)
Smoking	477 (10%)	1,469 (10%)
Number of incident cancers after NAFLD diagnosis or matching		
Gastrointestinal cancers, n (% of total cancers)	176 (27%)	282 (18%)
Colon	95	181
Liver	28	23
Pancreas	29	43
Stomach/cardia	16	14
Esophagus	8	21
Hormone-sensitive cancers, n (% of total cancers)	410 (62%)	1,118 (71%)
Breast	181	495
Prostate	134	447
Uterus	76	126
Ovary	19	50
Lung/bronchus, n (% of total cancers)	70 (11%)	168 (11%)

NAFLD, non-alcoholic fatty liver disease.

Table 2. The incidence rate of cancers in NAFLD individuals and the referent cohort.

Type of cancer	Incidence* per 100,000-person years	
	NAFLD	Referent cohort
Gastrointestinal/liver cancers		
Liver	56.0 (29.3–82.7)	18.1 (14.7–21.8)
Colon	297.6 (245.1–350.1)	141.6 (130.7–152.3)
Pancreas	81.4 (36.9–125.9)	37.7 (28.7–46.5)
Stomach/cardia	41.8 (20.0–64.0)	15.3 (12.4–18.2)
Esophagus	36.1 (27.2–44.9)	20.7 (18.2–23.2)
Hormone-sensitive cancers		
Breast	923.9 (789.5–1,057.5)	692.0 (630.7–753.3)
Prostate	1,355.9 (1,115.7–1,596.1)	1,243.6 (1,127.0–1,360.2)
Uterus/endometrium	439.8 (344.5–555.1)	217.3 (178.9–255.7)
Ovary	89.8 (48.4–131.2)	70.3 (50.0–90.7)
Lung/bronchus	261.1 (184.2–331.2)	161.9 (142.7–181.0)

NAFLD, non-alcoholic fatty liver disease.
* Incidence shown at age 65.

higher risk of incident cancers (overall malignancy IRR = 2.0, 95% CI 1.5–2.9), while obesity alone was not (IRR = 1.0, 95% CI 0.8–1.4). These data suggest that the increased risk of malignancy associated with obesity is largely attributed to the presence of NAFLD, and when individuals with NAFLD are removed from the obese group the obesity-cancer association diminishes significantly. In reference to obese controls, those with NAFLD had a 2-fold overall increase in incident malignancies (IRR = 2.0, 95% CI 1.5–2.7). Among cancer groups, the largest effect of NAFLD in reference to both obese and non-obese controls was highest in liver and gastrointestinal cancers, where the risk increase varied between 2- and 3-fold, while the effect of NAFLD compared to obesity was not as high in uterine and ovarian cancer (individual IRRs by cancer type in Table 3).

Sensitivity analysis adjusting for diabetes mellitus did not change these findings, although the association of NAFLD with malignancy risk was slightly decreased: NAFLD versus non-obese controls IRR = 1.8, 95% CI 1.4–2.5; NAFLD versus obese controls IRR = 1.9, 95% CI 1.4–2.6 (Fig. S3 and Table S5). Secondary analysis of the effect of overweight status on malignancy risk using normal weight as a reference did not show a significant association IRR = 0.74 (0.32–1.69). The number of incident cancers by BMI subgroups is shown in Table S6. We did not adjust for smoking given similar prevalence among NAFLD and controls. Similarly, we did not perform secondary analysis exploring the potential impact of alcohol use on the increased malignancy risk in NAFLD because a very small proportion of NAFLD individuals (4%) were subsequently diagnosed with alcohol use disorder, at a median of 3.5 years after index date (*vs.* 10% of referents). Therefore, disparities in alcohol use among NAFLD and referents are unlikely to explain the increase in malignancy risk.

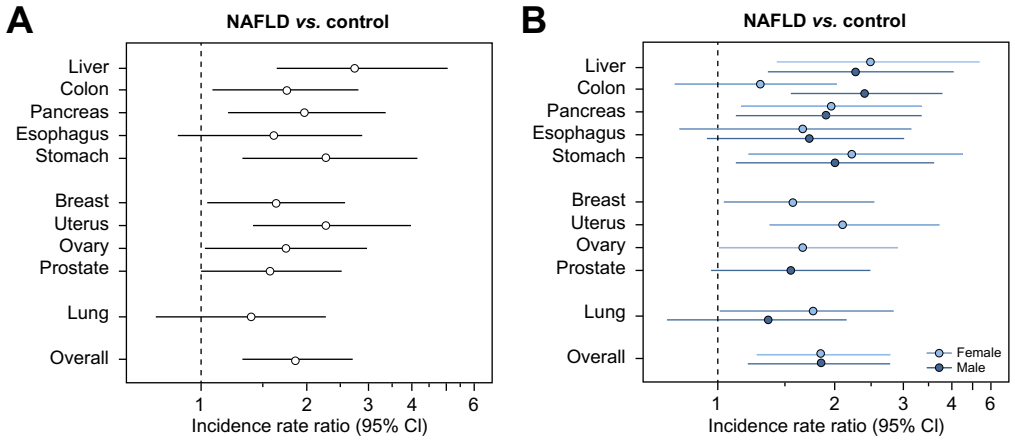


Fig. 2. Forrest plot of risk of incident cancer among individuals with NAFLD compared to age- and sex-matched referent individuals without NAFLD (controls) from the same population. Plot shows the incidence rate ratios and 95% CIs. Incidence rate ratios >1 indicate increased cancer risk in obese compared to non-obese. NAFLD, non-alcoholic fatty liver disease.

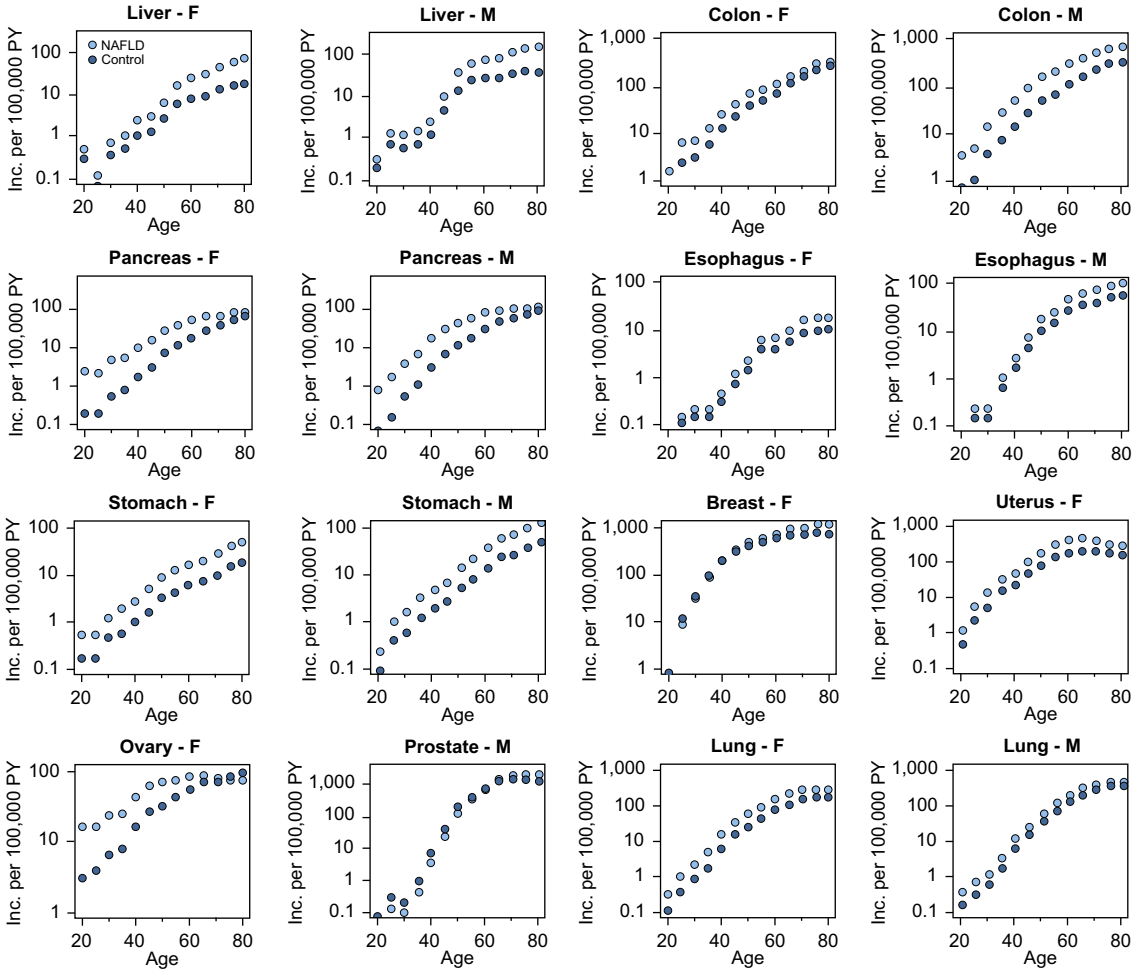


Fig. 3. Incidence of cancer by age in NAFLD (light blue) and referent individuals (dark blue). Smooth curve illustrating the results of Poisson regression, performed using SEER registry rates at each decade of age, as the reference category. NAFLD, non-alcoholic fatty liver disease; Surveillance, Epidemiology and End Results.

Discussion

This large community cohort study with 21 years of longitudinal follow-up adds several important observations to our knowledge of the natural history of NAFLD and its association with subsequent malignancies. First, people with NAFLD had a nearly 2-fold increase in the overall risk of incident cancers when compared to an age- and sex-matched general population cohort. Second, this study provides a hierarchical overview of

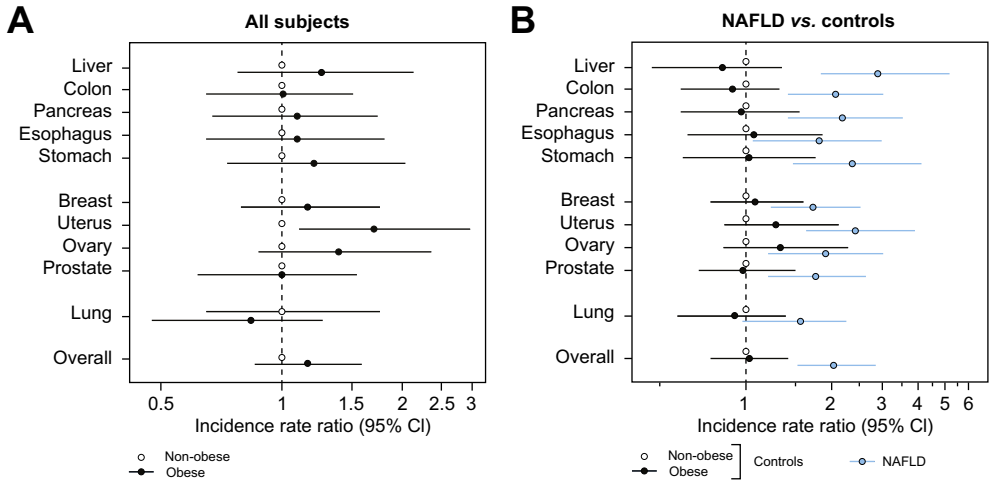


Fig. 4. Forrest plot of risk of incident cancers, adjusted by age and sex. (A) Obese *vs.* non-obese participants, irrespective of NAFLD status. (B) NAFLD *vs.* non-obese controls (red), obese controls *vs.* non-obese controls (blue). Plot shows the incidence rate ratios and 95% CI. Incidence rate ratios >1 indicate increased cancer risk. NAFLD, non-alcoholic fatty liver disease.

Table 3. The effect of obesity and NAFLD on malignancy risk (adjusted for age and sex).

Cancer type	Incidence rate ratio – 95% CI			
	Random effect: BMI ≥30 kg/m^2	Random effect: BMI ≥30 kg/m^2 and NAFLD		
	Obese *vs.* non-obese (irrespective of NAFLD presence)	NAFLD *vs.* non-obese controls	NAFLD *vs.* obese controls	Non-NAFLD obese *vs.* non-obese
All cancers	1.2 (0.9–1.6)	**2.0 (1.5–2.9)**	**2.0 (1.5–2.7)**	1.0 (0.8–1.4)
GI cancers				
Liver	1.3 (0.8–2.2)	**2.9 (1.8–5.3)**	**3.6 (2.0–7.5)**	0.8 (0.5–1.4)
Colon	1.0 (0.7–1.5)	**2.1 (1.4–3.1)**	**2.3 (1.7–3.2)**	0.9 (0.6–1.3)
Pancreas	1.1 (0.7–1.8)	**2.2 (1.4–3.5)**	**2.3 (1.4–3.9)**	1.0 (0.6–1.6)
Esophagus	1.1 (0.7–1.9)	**1.8 (1.0–3.0)**	1.7 (0.9–3.1)	1.1 (0.6–1.9)
Stomach	1.2 (0.7–2.1)	**2.4 (1.5–4.1)**	**2.3 (1.3–4.5)**	1.0 (0.6–1.8)
Hormone-sensitive cancers				
Breast	1.2 (0.8–1.8)	**1.7 (1.2–2.5)**	**1.6 (1.3–2.0)**	1.1 (0.8–1.6)
Uterus	**1.7 (1.1–3.0)**	**2.4 (1.6–4.0)**	**1.9 (1.4–2.7)**	1.3 (0.9–2.2)
Ovary	1.4 (0.9–2.4)	**1.9 (1.2–3.1)**	1.4 (0.8–2.4)	1.3 (0.8–2.3)
Prostate	1.0 (0.7–1.6)	**1.8 (1.2–2.7)**	**1.8 (1.4–2.3)**	1.0 (0.7–1.5)
Lung cancer	0.8 (0.5–1.3)	1.5 (0.9–2.3)	**1.7 (1.2–2.3)**	0.9 (0.6–1.4)

Bold values represent statistically significant results. BMI, body mass index; GI, gastrointestinal; NAFLD, non-alcoholic fatty liver disease.

the cancer types that are most likely to increase in NAFLD, namely liver and gastrointestinal cancers. Lastly, we show that NAFLD may be a more important intermediary biomarker of cancer risk. In this cohort, the obesity-related risk was largely driven by NAFLD, while obesity in the absence of NAFLD had minimal association with malignancy risk. These findings serve as hypothesis-generators for future studies into the biological mechanisms underpinning this link, which will need to assess whether NAFLD is a *predictor* by association or a *mediator* of causal pathways in the development of cancer.

A large volume of epidemiologic data has established that excess adiposity, measured by BMI, is a risk factor for several, but not all, common cancers.[4,18] The proposed candidate mechanisms for the adiposity-cancer link include altered sex hormone metabolism, increased insulin levels and bioavailability of insulin-like growth factor 1, adipokine pathophysiology and systemic inflammation.[19,20] On the other hand, it has been recognized that excess body fat can have distinct consequences despite similar BMI.[21] One such instance has been observed in those with 'metabolically healthy obesity', a phenotype which is not associated with cardiovascular, metabolic, or even malignancy risk.[22,23] Variations in fat distribution may potentially explain the risk differential. Visceral adipose tissue and ectopic hepatic fat may contribute to local and systemic inflammation, insulin resistance and metabolic disease. Emerging translational and epidemiologic data support the importance of local ectopic fat as a paracrine mechanism for cancer development in the liver, pancreas[24,25] and breast, where the local adipose tissue microenvironment impacts tumor progression.[26] It is therefore biologically plausible that NAFLD is a risk factor for cancer, not only of the liver, but also of close proximity organs, such as the gastrointestinal tract.

Whether non-alcoholic steatohepatitis versus simple steatosis have distinct associations with extrahepatic cancer risk is difficult to establish in a large population, due to the lack of universal non-invasive diagnostic methods and the unreliability of liver enzymes as serum NASH biomarkers. Nevertheless, it was clear that diabetes mellitus was not an important confounding variable for the NAFLD-malignancy association, which suggests that insulin resistance is not the dominant common link between cancer and NAFLD. Another potential explanation of these findings is that NAFLD has no direct causality in cancer

biology, but is a better predictor than BMI in reflecting an obesity phenotype with higher malignancy potential, as it is closely associated with central adiposity and insulin resistance. BMI might be too crude a measure of body fatness to accurately quantify the relationship between adiposity and cancer. Unfortunately, measures such as waist-hip ratio or waist circumference are not routinely collected during medical encounters. Regardless of the mechanism of association, to the extent that the presence of hepatic fat is indeed relevant to cancer development, the challenging task of applying reliable and cost-effective non-invasive modalities of NAFLD diagnosis to the community becomes even more imperative.

The general knowledge that patients with NAFLD have a higher propensity to develop extrahepatic cancers due to concurrent obesity has had, thus far, limited applicability in clinical practice beyond raising general awareness. This study offers a more detailed synopsis of the specific high-risk cancers in this population, and the magnitude of risk in reference to an individually matched population free of NAFLD. Of the extrahepatic cancers, stomach, pancreas and colon have an over 2-fold increase in incidence in those with NAFLD, with a trend towards younger age at diagnosis in the latter 2. These findings have great applicability in clinical practice, where they can help individualize risk-counseling in NAFLD. Furthermore, they establish a framework that can be used in future large-scale studies of the effectiveness of screening policies in obesity in general and NAFLD in particular.

We found an important interaction between sex and the risk of colon cancer. Whereas the overall risk was significantly higher in NAFLD versus controls, stratification by sex showed that the effect was entirely present in men while insignificant in women, and this was confirmed by formal testing of the sex interaction. The reason remains elusive (although noted in previous studies on obesity-related colorectal cancer[20]), yet the findings suggest that the counseling on risk should be individualized by sex. In women, the risk of uterine cancer was considerably higher, while that of breast cancer was not significant. This is consistent with findings from obesity-related studies, in which the increased risk of breast cancer was inconsistently found. Also, like previous obesity studies, the risk of lung and prostate cancer was not associated with obesity or NAFLD.

As with any observational epidemiologic studies, it is particularly important to note potential sources of bias. Studies of populations such as Olmsted County are likely to have lower disease prevalence than those estimated from NHANES data which used ultrasound screening, especially given the lack of symptoms and reliable biomarkers that would prompt screening of everyone at risk. Therefore, the prevalence of diagnosed NAFLD is expected to be significantly lower than that of hepatic steatosis incidentally noted on imaging. However, natural history data obtained from patients who were diagnosed with NAFLD remain important and are closest to "real-world" scenarios because they allow longitudinal follow-up with complete ascertainment of outcomes such as malignancy. In the absence of systematic screening, it is possible that a proportion of controls had undiagnosed NAFLD. If NAFLD is associated with malignancy risk, this sampling bias would lead to a higher estimated incidence of cancers among controls. The impact that this bias would have had on the results is a lower relative incidence rate ratio between patients with NAFLD and controls, and underestimation of relative risk due to an artificial increase in denominator. Thus, it is possible that after careful removal of undiagnosed NAFLD from the reference population, the relative risk of malignancy in NAFLD would be even higher than our estimates. Survival bias, resulting from the association of NAFLD with mortality from causes unrelated to cancer, would have a less clear impact on the validity of our results. Shorter lifespan would result in shorter person-year follow-up in NAFLD (denominator) but also a lower cumulative incidence of cancers (numerator), thus an uncertain impact on the incidence rate ratio. Medical surveillance bias, resulting from more rigorous cancer screening of those with NAFLD during more frequent medical evaluations is possible, but more likely to affect studies of subclinical outcomes or stage at diagnosis rather than the overall diagnosis of cancer in a person's lifetime, especially given that these cancers are likely to become symptomatic and lead to medical evaluation eventually. Moreover, the referent individuals with no follow-up or active medical visits represent a very small proportion of the population. Previous analyses of REP studies showed that 93% of Olmsted County residents have at least 1 medical visit every 3 years and only 4% of the population is never seen again after a baseline visit.

The strengths of this study include the large sample size, the use of a reference population individually matched by age and sex and the historical depth provided by long-term follow-up. We used routinely collected and linked medical data to provide essential information about the natural history of the disease in the community, which limits the risk of selection bias which registries or referral centers are prone to. Although disease was defined using electronic indices, we reinforced the ascertainment validity by in-depth chart review using the medical record-linkage system, for both NAFLD and each cancer type.

The size of the Olmsted County population limits robust conclusions on rare cancers. This may explain why we did not find a higher risk of esophageal cancer in NAFLD or obesity, despite strong evidence that this is one of the several cancers strongly associated with obesity. The age, sex, ethnic and socioeconomic characteristics of the Olmsted County population are similar to other populations in the upper Midwest region of the United States, but some racial and ethnic groups are under-represented; these characteristics should be considered when attempting to generalize to other populations. However, no single community in the United States is completely representative of the entire country and results from cancer epidemiology studies in Olmsted County have been consistent with national data.[27–31]

These limitations notwithstanding, these unique epidemiologic observations reframe our understanding of the association between obesity and cancer risk. There is a continued need for better characterization of excess adiposity, because current measures of obesity, such as BMI, are insufficient and may overlook other potential key contributors to outcomes, based on ectopic fat distribution. Our findings provide a platform for future mechanistic studies of NAFLD as the concealed driver or intermediary biomarker of cancer risk in obesity.

Financial support

Alina M. Allen: National Institute of Diabetes and Digestive and Kidney Diseases (K23DK115594); American College of Gastroenterology Junior Faculty Development Grant. This study was made possible using the resources of the Rochester Epidemiology Project, which is supported by the National Institute on Aging of the National Institutes of Health under Award Num-

ber R01AG034676. The funding sources did not have any role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Conflict of interest

The authors declare no conflicts of interest that pertain to this work.

Please refer to the accompanying ICMJE disclosure forms for further details.

Authors' contributions

AMA: study concept and design; acquisition of data; analysis and interpretation of data; drafting of the manuscript; obtained funding. SBH: acquisition of data. KCM: statistical analysis; critical revision of the manuscript for important intellectual content. JJL: statistical analysis; critical revision of the manuscript for important intellectual content. TMT: study concept and design; statistical analysis; analysis and interpretation of data; critical revision of the manuscript for important intellectual content.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhep.2019.08.018.

References

Author names in bold designate shared co-first authorship

[1] Hastings KG, Boothroyd DB, Kapphahn K, Hu J, Rehkopf DH, Cullen MR, et al. Socioeconomic differences in the epidemiologic transition from heart disease to cancer as the leading cause of death in the United States, 2003 to 2015: an observational study. Ann Intern Med 2018;169:836–844.
[2] Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int J Cancer 2010;127:2893–2917.
[3] Renehan AG, Tyson M, Egger M, Heller RF, Zwahlen M. Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. Lancet 2008;371:569–578.
[4] Kyrgiou M, Kalliala I, Markozannes G, Gunter MJ, Paraskevaidis E, Gabra H, et al. Adiposity and cancer at major anatomical sites: umbrella review of the literature. BMJ 2017;356 j477.
[5] **Arnold M**, **Pandeya N**, Byrnes G, Renehan PAG, Stevens GA, Ezzati PM, et al. Global burden of cancer attributable to high body-mass index in 2012: a population-based study. Lancet Oncol 2015;16:36–46.
[6] Trends in adult body-mass index in 200 countries from 1975 to 2014: a pooled analysis of 1698 population-based measurement studies with 19.2 million participants. Lancet 2016;387:1377–1396.
[7] Ogden CL, Carroll MD, Flegal KM. Epidemiologic trends in overweight and obesity. Endocrinol Metab Clin North Am 2003;32:741–760, vii.
[8] Allen AM, Therneau TM, Larson JJ, Coward A, Somers VK, Kamath PS. Nonalcoholic fatty liver disease incidence and impact on metabolic burden and death: a 20 year-community study. Hepatology 2018;67:1726–1736.
[9] Younossi ZM. Non-alcoholic fatty liver disease – A global public health perspective. J Hepatol 2019;70:531–544.
[10] Younossi ZM, Koenig AB, Abdelatif D, Fazel Y, Henry L, Wymer M. Global epidemiology of nonalcoholic fatty liver disease-Meta-analytic assessment of prevalence, incidence, and outcomes. Hepatology 2016;64:73–84.
[11] Adams LA, Lymp JF, St Sauver J, Sanderson SO, Lindor KD, Feldstein A, et al. The natural history of nonalcoholic fatty liver disease: a population-based cohort study. Gastroenterology 2005;129:113–121.
[12] Kim D, Kim WR, Kim HJ, Therneau TM. Association between noninvasive fibrosis markers and mortality among adults with nonalcoholic fatty liver disease in the United States. Hepatology 2013;57:1357–1365.
[13] Rocca WA, Yawn BP, St Sauver JL, Grossardt BR, Melton 3rd LJ. History of the Rochester Epidemiology Project: half a century of medical records linkage in a US population. Mayo Clin Proc 2012;87:1202–1213.
[14] St Sauver JL, Grossardt BR, Yawn BP, Melton 3rd LJ, Rocca WA. Use of a medical records linkage system to enumerate a dynamic population over time: the Rochester epidemiology project. Am J Epidemiol 2011;173:1059–1068.
[15] St Sauver JL, Grossardt BR, Yawn BP, Melton 3rd LJ, Pankratz JJ, Brue SM, et al. Data resource profile: the Rochester Epidemiology Project (REP) medical records-linkage system. Int J Epidemiol 2012;41:1614–1624.
[16] Laird N, Olivier D. Covariance analysis of censored survival data using log-linear analysis techniques. J Am Stat Assoc 1981;76:231–240.
[17] Plummer M. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. Proceedings of the 3rd International Workshop on Distributed Statistical Computing; 2003: Vienna, Austria, 2003.
[18] Lauby-Secretan B, Scoccianti C, Loomis D, Grosse Y, Bianchini F, Straif K. Body fatness and cancer-viewpoint of the IARC working group. N Engl J Med 2016;375:794–798.
[19] Calle EE, Kaaks R. Overweight, obesity and cancer: epidemiological evidence and proposed mechanisms. Nat Rev Cancer 2004;4:579–591.
[20] Renehan AG, Zwahlen M, Egger M. Adiposity and cancer risk: new mechanistic insights from epidemiology. Nat Rev Cancer 2015;15:484–498.
[21] Canoy D, Boekholdt SM, Wareham N, Luben R, Welch A, Bingham S, et al. Body fat distribution and risk of coronary heart disease in men and women in the European Prospective Investigation Into Cancer and Nutrition in Norfolk cohort: a population-based prospective study. Circulation 2007;116:2933–2943.
[22] Stefan N, Haring HU, Hu FB, Schulze MB. Metabolically healthy obesity: epidemiology, mechanisms, and clinical implications. Lancet Diabetes Endocrinol 2013;1:152–162.
[23] Moore LL, Chadid S, Singer MR, Kreger BE, Denis GV. Metabolic health reduces risk of obesity-related cancer in Framingham study adults. Cancer Epidemiol Biomarkers Prev 2014;23:2057–2065.
[24] Hori M, Takahashi M, Hiraoka N, Yamaji T, Mutoh M, Ishigamori R, et al. Association of pancreatic Fatty infiltration with pancreatic ductal adenocarcinoma. Clin Transl Gastroenterol 2014;5 e53.
[25] Lashinger LM, Malone LM, McArthur MJ, Goldberg JA, Daniels EA, Pavone A, et al. Genetic reduction of insulin-like growth factor-1 mimics the anticancer effects of calorie restriction on cyclooxygenase-2-driven pancreatic neoplasia. Cancer Prev Res (Phila) 2011;4:1030–1040.
[26] **Park J**, **Morley TS**, Kim M, Clegg DJ, Scherer PE. Obesity and cancer–mechanisms underlying tumour progression and recurrence. Nat Rev Endocrinol 2014;10:455–465.
[27] Yang JD, Kim B, Sanderson SO, St Sauver JL, Yawn BP, Pedersen RA, et al. Hepatocellular carcinoma in olmsted county, Minnesota, 1976–2008. Mayo Clin Proc 2012;87:9–16.
[28] Beard CM, Hartmann LC, Atkinson EJ, O'Brien PC, Malkasian GD, Keeney GL, et al. The epidemiology of ovarian cancer: a population-based study in Olmsted County, Minnesota, 1935–1991. Ann Epidemiol 2000;10:14–23.
[29] Yang JD, Ahmed Mohammed H, Harmsen WS, Enders F, Gores GJ, Roberts LR. Recent trends in the epidemiology of hepatocellular carcinoma in Olmsted County, Minnesota: a US population-based study. J Clin Gastroenterol 2017;51:742–748.
[30] Conio M, Cameron AJ, Romero Y, Branch CD, Schleck CD, Burgart LJ, et al. Secular trends in the epidemiology and outcome of Barrett's oesophagus in Olmsted County, Minnesota. Gut 2001;48:304–309.
[31] Beard CM, Hartmann LC, Keeney GL, Crowson CS, Malkasian GD, O'Brien PC, et al. Endometrial cancer in Olmsted County, MN: trends in incidence, risk factors and survival. Ann Epidemiol 2000;10:97–105.