Reviewer #1:

This study has well constructed analyses to support the conclusion.
However, readers would probably become curious about an analysis of type 2 diabetes mellitus.
The term 'diabetes' is used because type 1 and type 2 diabetes patients are not separated. There is a claim that there are insufficient numbers of type 1 diabetes patients for meaningful outcomes. As types 1 and 2 diabetes are generally believed to have different causes it is unwise to present only a mix of data from both types. Thus, the data for types 1 and 2 need to be analysed separately in addition to the presented analyses. If there are insufficient numbers of type 1 diabetes patients, then an analysis of type 1 by itself should be omitted and an analysis of the data from type 1 alone should be added.
Minor: correction needed in abstract: 'explained by with increased'

Our reply: We would like to acknowledge the reviewer 1 for very thorough review and highly constructive comments.

In the NHANES, no information was collected regarding type of diabetes and autoantibodies of type 1 diabetes were not assessed. Thus, we are not able to separate type 1 diabetes out in this cohort. Due to a low prevalence of type 1 diabetes, more than 90% of diabetes can be assumed to have type 2 diabetes. To address this issue, we added the follow statement in the revised manuscript.

"Another potential limitation is that in the NHANES database, no information was collected regarding the type of diabetes, nor were auto-antibodies for type 1 diabetes assessed. Thus, patients with type 1 diabetes cannot be separated out unambiguously. In a National Health Interview Survey in 2016, type 2 diabetes accounted for 90.9% of diabetes and type 1 diabetes accounted for 5.8% of diabetes [22], defined based on self-reported type and/or current use of insulin. In the 2016 Survey, the reported prevalence of type 1 diabetes was 0.55% of the adult population in the United Stated [22], which is much higher than the reported 0.18% from the Hispanic Community Health Study/Study of Latinos, which is a community-based epidemiologic study in Hispanic/Latino adults residing in four US communities [23]. The overestimation in the 2016 Survey [22] could be due to an assumption that current use of insulin is indicative of type 1 diabetes. However, given that type 2 diabetes develops when beta cell function is reduced to 55%, and that beta cell function declines by 5% per year regardless of treatment modalities [24], insulin treatment is expected to be required sometime during the course of type 2 diabetes. Indeed, it has been demonstrated that almost two-thirds of patients with type 2 diabetes received insulin treatment [25]. Thus, defining type 1 diabetes based on current use of insulin is unreliable and could overestimate the prevalence of type 1 diabetes. Given that type 1 and type 2 diabetes are the two most common forms of diabetes, and that less than 10% of patients with diabetes have type 1 diabetes, we expect that the vast majority (≥90%) of diabetic patients in the NHANES cohort have type 2 diabetes. Thus, we are able to confidently conclude an association between HAV infection and type 2 diabetes.

Current evidence does not rule out an association between HAV infection and type 1 diabetes. Hyperglycemia is associated with an increased risk of infection [26]. Infectious burden increases by 2.14 times in diabetes [27]. Furthermore, hospitalization rates from infection are almost 3.8 times higher in adults with diabetes than without diabetes, especially in young adults with diabetes [28] who are more likely to have type 1 diabetes. Thus, patients with type 1 diabetes are also likely to have an increased risk of HAV infection. However, due to the relatively low rate of type 1 diabetes and our inability to separate out subjects with type 1 diabetes unequivocally based on the NHANES study design, we are not able to generalize our observed results to type 1 diabetes. Nevertheless, our

observation of an increased risk of HAV infection in subjects with diabetes is in line with the reported increased risk of infection in diabetes, regardless of the type of diabetes."

Correction was made as suggested by Reviewer 1 in abstract.

Science editor:
1. Scientific quality. The manuscript is analysis of data from observational study collected between 2005-2012. The topic is within the scope of the WJD.

a. Classification Grade B

b. Summary of peer review report: Manuscript concern the association between HAV infection / vaccination, and diabetes, but authors did notdistinguish persons with diabetes type 1 and 2. Authors carefully analysed many important factors in different waysbut did not eliminate group of type 1 diabetic patients. The questions raised by the reviewer should be answered.

c.Format- there are 7 tables and 1 diagram.

d. References – a total of 20 references are cited, included only 1 reference published in the last 3 years.
Our reply: In the revised manuscript, we increased references to a total of 29 with 9 references dated after 2018.

e. Self-cited references- there is only one self-cited references.
f. Reference recommendations- the authors should seek more recent references.
Our reply: In the revised manuscript, we increased references to a total of 29 with 9 references dated after 2018.

2. Language evaluation: Classification Grade B-minor language polishing.
Our reply: The manuscript was critically reviewed and edited by our institutional scientific writers: Sarah T. Wilkinson, PhD and Henry Lin, PhD. Sarah T. Wilkinson, PhD re-edited and polished the revised manuscript again before resubmission.

3. Academic norms and rules; The authors provided Conflict of interest Disclosure form.
4.Supplementary comments: Invited manuscript, supported by grant from Jie Chan Chen Foundation. Manuscript was not previously published.
5. Issues raised:
a. The most problematic is the problem of time dependence between HAV infection and diabetes. If we seek the associations between these diseases, we have to know that HAV infection was before the diabetes diagnosis, in the other situation all analyses have no sense.
Our reply: We totally agree with Science Editor. To the nature of the NHANES, frequent subclinical HAV infection and no information on prior HAV infection and when, we took a stepwise hypothesis testing approach to dissect these issues as stated in the manuscript:

"As this is a cross-sectional study, no causal relationship can be proved directly. However, using a stepwise hypothesis testing approach, we were able to deduce the results and to infer diabetes as the cause of the increased susceptibility to HAV infection. Alternatively, one could examine the effect of aggressive prevention of diabetes on the incidence of HAV infection. Given the established benefit of diabetes prevention [21], it would not be ethical to conduct an interventional trial to examine the

impact of diabetes prevention (versus its absence) on the development of HAV infection. Thus, stepwise hypothesis testing is a useful alternative approach to infer a causal relationship when an interventional trial is not feasible to confirm the observation."

b. It should be reanalysed after eliminating data of patients with type 1 diabetes.
Our reply: In the NHANES, no information was collected regarding type of diabetes and autoantibodies of type 1 diabetes were not assessed. Thus, we are not able to separate type 1 diabetes out in this cohort. Due to a low prevalence of type 1 diabetes, more than 90% of diabetes can be assumed to have type 2 diabetes. To address this issue, we added the follow statement in the revised manuscript as stated in our reply to the reviewer #1.

c. Despite the big number of participants, it is necessary to check the normality of data distribution, and variancy, because this is requiring of T- student tests. (Data could be for example two peak). Maybe for several data mean and SD are not appropriate.
Our reply: We did check the data and confirmed the single peak before analyses.

d. Through all manuscript mistake"HVA" abbreviation should be replaced by HAV.
Our reply: All HVA were replaced by HAV.

e. It is very questionable diabetes diagnosis in this group of patients, because HbA1c from whole group was 5,7%- what result is complete norm for healthy persons. In such large group is unlikely to have all extremely good metabolic controlled patients.
Our reply: Since diabetic patients only accounted for 17.37%, the mean HbA1c is heavily weighted by the non-diabetic participants. The highest HbA1c in diabetic patients was 17.8%

f. It is questionable that in group of HAV vaccinated persons is lack of infected despite large number of patients without significant presence of antibodies.
Our reply: The purpose of HAV vaccination is to prevent HAV infection. The anti-HAV antibody was presented in 53.89% of vaccinated subjects which is most likely from progressive loss of HAV antibody as time goes by. Anti HAV antibody would reappear if HAV infection occurred in those previously received HAV vaccine without anti-HAV antibody at the time of survey. Due to low incidence rate of HAV infection (0.4 cases per 100,000), we estimated 0.008 cases of HAV infection in this vaccinated (n=4,229) cohort without anti-HAV antibody (46.11%, n=1,950). Thus, it is negligible.

g. Pointing that between group existed significant differences in ALAT (25 vs 26), AspAT (25 vs 25), and bilirubin is rather not important and should not be described.
Our reply: We believe it has some merit to discuss liver function while testing the hypothesis whether HAV infection is a risk factor of diabetes, since the liver is a particularly important organ in glucose homeostasis. Some reviewers and readers may ask for these data. However, in compliance with Science Editor's comment, we deleted that the results and discussion as recommended.

h. The article" Highlight section" is missing. Please add the Article Highlight section at the end of the main text.
Our reply: Article Highlight was added as instruction in the Guidelines and Requirements for Manuscript Revision.

6. Re-review: Required
7. Recommendation: Conditional acceptance.