PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Thiazide Diuretics and the Risk of Hip Fracture after Stroke: A Population-based Propensity-matched Cohort Study Using Taiwan's National Health Insurance Research Database
AUTHORS	Lin, Shu-Man; Yang, Shih-Hsien; Cheng, Hung-Yu; Liang, Chung-Chao; Huang, Huei-Kai

VERSION 1 - REVIEW

REVIEWER	Petra Benzinger Robert Bosch Hospital Department of Geriatrics
REVIEW RETURNED	20-Apr-2017

GENERAL COMMENTS	There is information lacking on the observation period: When did observation period end - please give precise information. How many patients died during the observation period? Please provide details about time to fracture. Person years differe between group - please provide information on reasons. Were patients censored after the first fracture? In the discussion you mention that patietns were not censored and might have received thiazide medication AFTER the fracture. This is not accaptable. Patients should have been censored after first fracture. Matching was done for the group of all thiazide recipients - how did the subgroups (according to duration of prescription) differ from the total group (mortality / time to fracture / baseline characteristics). Please provide information in Tab 1 for these three subgroups. Study population: From 200 to 2001: please be more precise "those with a history of stroke before the 2000–2011 study period" - over what time period did you identify previous fractures? Osteoporosis medication could have biased your results. Please provide information about osteoporosis medication after stroke in all groups/subgroups and adjust for it in case there is a relevant proportion of patietns exposed to such medication. Results Table 1 Loop diuretics 328 (8.6%) vs 344 (9.0%) This similarity of both groups is not plausible at first glance. Patients rarely are treated with both loop diuretics and thizide diuretics given the double blockage of the tubulus. Such medication may lead to disturbance in electrolytes. Were all loop diuretic patients switched to thiazide? Please explaine. Chronic kidney disease 360 (9.5%) 349 (9.2%) Again, this similarity is rather inplausible given that in advanced renal insufficiency thiazide diuretics are no longer effective. How did you define CKD?

REVIEWER	Prof Johnny Yu JIANG Peking Union School of Public Health, PUMC & CAMS, China
REVIEW RETURNED	22-Apr-2017

GENERAL COMMENTS	This is an interesting population-based cohort study based on Taiwan's National Health Insurance Research Database. It indicated that long-term use of thiazides is associated with a decreased hip fracture risk after stroke, which may provide evidence-based information for hip fracture prevention of stoke patients. But there are some questions need to be clarified in the paper. 1. There were 4382 patients with thiazide use in the database initially, but the number reduced to 3810 after propensity score matching, were there any differences between the population included in the analysis and the population excluded? Why did not you include all 12820 patients as research subjects since all the potential confounders you balanced by propensity score matching were also adjusted in the cox regression? 2. Please explain the reason you excluded patients aged<20 years in this study. 3. The method you used to calculate propensity score was not given in the paper. Please specify. 4. The proportion of hip fracture due to falls in the two groups respectively can be explained and further analyzed in the paper if possible, since there are concerns that diuretics may increase the risk of fall in the elderly. 5. Was there any possibility that the concomitant drugs were different between the two groups that may influence the results?

REVIEWER	KoKo Aung, MD, MPH Paul L. Foster School of Medicine, Texas Tech University Health Sciences Center at El Paso, U.S.A.
REVIEW RETURNED	30-Apr-2017

GENERAL COMMENTS	Overall, this is a well written manuscript of an observational study that addresses a knowledge gap related to effects of thiazide on hip fracture in patients who had stroke. The research question was clearly defined. The database and the methods used were appropriate. The manuscript was written in a very clear and concise manner. The tables were very clear for the readers to follow, understand, and interpret. Addressing the questions and concerns listed below would further strengthen the already strong manuscript. 1. How was the decision arrived to perform propensity score matching instead of directly adjusting for all covariates that were used to calculate the propensity score? 2. It was reported that propensity score matching was performed in this study to balance covariates and baseline differences, including age, sex, baseline comorbidities, socioeconomic factors, stroke severity proxies and use of medications as listed in Table 1. The patients who had thiazide use after stroke were matched (without replacement) with those who did not, and a nearest-neighbor algorithm was applied to construct matched pairs. After matching, 7620 patients with ischemic stroke without previous hip fracture were included. It appears that matches could be found for 3810

	thiazide users, among patients with ischemic stroke without previous hip fracture. a. How many thiazide users (among patients with ischemic stroke without previous hip fracture) were left unmatched and thus excluded from the study? b. What was the acceptable “neighborhood” of propensity scores for matching purpose? c. What factors were weighed in to balance the trade-off between inexact matching (and residual confounding) and incomplete matching (and loss of statistical power) in this study? The trend of fewer hip fractures among thiazide users in this study without reaching the level of statistical significance (p = 0.059 in table 2) raised the question on whether there was incomplete matching. d. How were missing data related to more than 30 covariates in table 1 handled when calculating propensity scores? 3. Are there any patients with more than one hip fracture within 2 years follow-up period? If so, how was the data handled? Do 76 and 92 hip fractures in table 2 represent 76 and 92 unique patients who sustained hip fractures or merely the number of hip fractures? 4. Please clarify whether propensity score was included as one of the covariates in regression analysis that evaluates effect of thiazides.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Petra Benzinger
Robert Bosch Hospital, Department of Geriatrics
Please state any competing interests or state ‘None declared’: None declared
==========
Please leave your comments for the authors below

There is information lacking on the observation period:
When did observation period end - please give precise information.

Reply:
We apologize for the ambiguous statement in original manuscript. All subjects were followed from the index date until a new diagnosis of hip fracture, death, or 2 years after stroke. The index date was defined as the date of diagnosis of new-onset stroke. Death was defined as the date a patient was withdrawn from the Taiwan NHI program [1, 2]. Because NHI is compulsory, there are very few occasions that a patient, particularly an ill one, can and will drop the insurance coverage, of which death is the most probable cause. The date of withdrawal from the NHI program has been recognized as an accurate and reliable proxy for date of mortality [3, 4].

The statements for the definition of observation period were amended in the Primary Outcome section in METHODS of our revised manuscript (p.7).

References:
1. Wu CY, Chen YJ, Ho HJ, et al. Association between nucleoside analogues and risk of hepatitis B virus-related hepatocellular carcinoma recurrence following liver resection. JAMA.

2012;308(18):1906-14.
2. Su TH, Chang SH, Chen PC, et al. Temporal Trends in Treatment and Outcomes of Acute Myocardial Infarction in Patients With Chronic Obstructive Pulmonary Disease: A Nationwide Population-Based Observational Study. J Am Heart Assoc. 2017;6(3).
3. Lien HM, Chou SY, Liu JT. Hospital ownership and performance: evidence from stroke and cardiac treatment in Taiwan. J Health Econ. 2008;27(5):1208-23.
4. Cheng CL, Chien HC, Lee CH, et al. Validity of in-hospital mortality data among patients with acute myocardial infarction or stroke in National Health Insurance Research Database in Taiwan. Int J Cardiol. 2015;201:96-101.

How many patients died during the observation period?

Reply:
In total, 987 patients died during the observation period in our study population. In thiazide cohort, 369 (9.9%) deaths were observed; in non-thiazide cohort, 618 (16.6%) deaths were observed.

A difference in the proportion of death between the groups was found, and the developing hip fracture may be confounded by mortality. Therefore, in our revised manuscript, we conducted modified Cox proportional hazards models in the presence of competing risk event with adjustment for competing mortality [1, 5].

The use of competing risk analysis was mentioned in the Statistical Analysis section in METHODS in the revised manuscript (p.9). The information about death number was mentioned in the Demographic characteristics of subjects section in RESULTS (p.10).

Besides, when calculating and examining the detailed information about deaths during the follow-up period, we found that some patients died (n = 649) or developed hip fracture (n = 19) during index hospitalization. These may cause bias in our results because the follow-up period was too short. In addition, most of them were allocated into the non-thiazide cohort because they have little opportunity to receive thiazide treatment before censoring in study. It may lead to overestimation of the hip fracture risk in the non-thiazide cohort, thereby overestimating the effect of thiazide in preventing hip fracture. Therefore, these patients were excluded in our revised manuscript.

The revised exclusion criteria was stated in the Study Population section in METHODS (p.7) and the flow diagram of the selection of study subjects (Figure 1) was also modified.

Thank you for your suggestions; they made us recognize the shortcomings and gave us the opportunity to revise our manuscript and improve the study design.

References:
1. Wu CY, Chen YJ, Ho HJ, et al. Association between nucleoside analogues and risk of hepatitis B virus-related hepatocellular carcinoma recurrence following liver resection. JAMA. 2012;308(18):1906-14.
5. Fine JP, Gray RJ. A Proportional Hazards Model for the Subdistribution of a Competing Risk. Journal of the American Statistical Association. 1999;94(446):496-509.

Please provide details about time to fracture.

Reply:
The information about time to fracture for patients who developed hip fracture during follow-up period are as follows:
In the thiazide cohort:
Mean = 1.09 years; median = 1.20 years; first quartile (Q1) = 0.49 years; third quartile (Q3) = 1.63; and interquartile range (IQR) = 1.14 years.
In the non-thiazide cohort:
Mean = 0.89 years; median = 0.85 years; Q1 = 0.35 years; Q3 = 1.32; and IQR = 0.97 years.

Person years differe between group - please provide information on reasons.

Reply:
The main reason for difference of person-years between the groups was possibly because of the different numbers of death during the follow-up period, as mentioned above. In addition, in our original study design, some patients died or developed hip fracture during index hospitalization. Most of these patients were allocated into the non-thiazide cohort because they have less opportunity to receive thiazide treatment before censoring, thereby causing the lower follow-up person-years in the non-thiazide cohort. In the revised version, the patients who died or developed hip fracture during index hospitalization were initially excluded, as mentioned above. Moreover, the hip fracture risk in the non-thiazide cohort was higher than that in the thiazide cohort. When hip fracture occurred, the patient was censored, thereby causing lower follow-up person-years. Therefore, in our study, the person-years was higher in the thiazide cohort than in the non-thiazide cohort.

Were patients censored after the first fracture? In the discussion you mention that patietns were not censored and might have received thiazide medication AFTER the fracture. This is not accaptable. Patients should have been censored after first fracture.

Reply:
Yes, when performing survival analysis and Cox regression model, the patients were censored after the first hip fracture. However, when we allocated patients to the thiazide or non-thiazide cohorts, we defined the thiazide cohort as patients with any prescription of thiazides within 2 years after stroke in our original manuscript. This misclassification may overestimate the risk of hip fracture in the thiazide cohort and underestimate the effect of thiazide in preventing hip fracture. If patients had a hip fracture before they began taking thiazides, they may have been misclassified into the thiazide cohort; this had been mentioned in Limitation section in our original manuscript.

Therefore, in the revised manuscript, only patients receiving thiazides before censoring (from the index date until a new diagnosis of hip fracture, death, or 2 years after stroke) were allocated to the thiazide cohort. If patients received thiazides after developing hip fracture, they were allocated to the non-thiazide cohort.

Overall, after improving the study design, such as matching and adjusting the osteoporosis medication, modifying the exclusion criteria, improving the classification of patients, and adjusting the competing mortality, the association of thiazide use and decreased hip fracture risk became more significant (Table 2 and Table 3). The statements about the changes of study design are mentioned in METHODS in the revised manuscript (p.7).

Matching was done for the group of all thiazide recipients - how did the subgroups (according to duration of prescription) differ from the total group (mortality / time to fracture / baseline characteristics). Please provide information in Tab 1 for these three subgroups.

Reply:
The baseline characteristics of the subgroups according to the duration of thiazide use have been described in the supplementary material (Table S1) because the data are too large (more than 2 pages) and should be published as online only supplementary material according to the formatting guidelines of BMJ Open. We found that most of the baseline characteristics were still balanced between subgroups; however, there were minor differences regarding the distribution of age, income level, urbanization level, stroke severity index score and the baseline prevalence of hypertension, and use of ACEI/ARB and calcium channel blockers. Next, we performed multivariate Cox regression model with adjustment for all baseline characteristics to estimate the HR of each subgroups compared with non-thiazide user. Therefore, the residual confounding effects caused by the minor difference in baseline characteristics could be controlled when performing subgroup comparisons.

To balance the case numbers in each subgroups, the duration categories were changed from 1–180 days, 181–365 days, and >365 days in the original manuscript to 1–90 days, 91–365 days, and >365 days in the revised manuscript. We have confirmed that both the ways of duration categories have the same results: only long-term use of thiazide (>365 days) is associated with significantly decreased risk of hip fracture.

The mortality rates differed between the subgroups according to the duration of thiazide use. In the duration of 1–90 days, 226 (15.2 %) deaths were observed; in the duration of 91–365 days, 117 (9.0 %) deaths were observed; and in the duration of >365 days, 26 (2.7 %) deaths were observed. In the revised manuscript, the adjustment for competing risk (mortality) was done when performing survival analysis and Cox regression model [1, 5].

The time of fracture in each subgroups according to duration of thiazide use are as follows:
Duration of 1–90 days:
Mean = 0.99 years; median = 0.98 years; Q1 = 0.45 years; Q3 = 1.54; and IQR = 1.09 years.
Duration of 91–365 days:
Mean = 1.08 years; median = 1.23 years; Q1 = 0.44 years; Q3 = 1.58; and IQR = 1.14 years.
Duration of >365 days:
Mean = 1.39 years; median = 1.55 years; Q1 = 1.17 years; Q3 = 1.76; and IQR = 0.59 years.

In addition, the detailed results of hip fracture risk according to the duration of thiazide use are added in Table 3 in the revised manuscript.

Please note that we could not create a table to present the data in the rebuttal letter because of the format provided by BMJ open; therefore, so we have presented the data in statement form, which is slightly difficult to read. We apologize for this inconvenience.

References:
1. Wu CY, Chen YJ, Ho HJ, et al. Association between nucleoside analogues and risk of hepatitis B virus-related hepatocellular carcinoma recurrence following liver resection. JAMA. 2012;308(18):1906-14.
5. Fine JP, Gray RJ. A Proportional Hazards Model for the Subdistribution of a Competing Risk. Journal of the American Statistical Association. 1999;94(446):496-509.

Study population: From 200 to 2001: please be more precise
"those with a history of stroke before the 2000–2011 study period" - over what time period did you identify previous fractures?

Reply:
We apologize for the ambiguous statement. Our longitudinal database included the ambulatory claims, inpatient care claims, and the registry of beneficiaries since 1999. Our study population was defined as patients with new-onset ischemic stroke during 2000–2011, and any patients with previous diagnosis of stroke before the year 2000 (that means in 1999) were excluded. After identifying the index date of each patient, we excluded patients with any previous hip fracture before index date. Therefore, the time period to identify previous hip fracture was at least 1 year (from Jan 1, 1999 to index date)
We have corrected the statements in the Study Population section in METHODS (p.7).

Osteoporosis medication could have biased your results. Please provide information about osteoporosis medication after stroke in all groups/subgroups and adjust for it in case there is a relevant proportion of patietns exposed to such medication.

Reply:
Yes, in our revised manuscript, the osteoporosis medication was added as one of the baseline characteristics and was used for propensity score matching and was adjusted in multivariate Cox regression model. The use of osteoporosis medication was defined as any prescription of medication including bisphosphonates, hormone replacement therapy, calcitonin, or vitamin D supplementation available in Taiwan [6] during the year preceding the index date.

In addition, as per your suggestion, we tested the analysis results that defined the use of medication as any prescription from the date of one year before index date to the end-follow-up point (after stroke), and the results remained the same. To unify the methods for defining the baseline use of medications, we used the definition for osteoporosis medications as a prescription during the year preceding the index date, which was similar to the definition for other medications in the revised manuscript.

The proportion of patients receiving osteoporosis medication in each group and subgroup has been included in Table 1 and Table S1 in the revised manuscript.

Reference:
6. Hwang JS, Chan DC, Chen JF, et al. Clinical practice guidelines for the prevention and treatment of osteoporosis in Taiwan: summary. J Bone Miner Metab. 2014;32(1):10-6.

Results
Table 1
Loop diuretics 328 (8.6%) vs 344 (9.0%)
This similarity of both groups is not plausible at first glance. Patients rarely are treated with both loop diuretics and thizide diuretics given the double blockage of the tubulus. Such medication may lead to disturbance in electrolytes. Were all loop diuretic patients switched to thiazide? Please explaine.

Reply:

The similarity in the proportion of loop diuretic use between the groups was because of propensity score matching. As your concern, previous studies have suggested that combination use of thiazides and loop diuretics leads to disturbance in electrolytes, causing hypokalemia and hyponatremia [7]. However, current studies revealed that some patients may benefit from receiving both thiazides and loop diuretics simultaneously, particularly in patients with resistant hypertension, chronic kidney disease, or congestive heart failure [7-10]. In clinical practice, some patients so receive these two kinds of diuretics simultaneously. The possible reasons are described below:

As recommended by NICE and Kidney Disease: Improving Global Outcomes (KDIGO) guidelines, loop diuretics should be prescribed when GFR is less than 30 mL/min in patients with chronic kidney disease [7, 11]. Loop diuretics are also indicated in the presence of edema or volume overload due to nephrotic syndrome or heart failure [12]. However, counter-regulatory rebound sodium retention could eliminate the efficacy of loop diuretics in patients with chronic kidney disease. Therefore, sequential nephron blockade using a combination of loop diuretics and thiazides may be needed in patients with resistant hypertension [7].

One previous study also found that the combination of thiazide diuretics and loop diuretics improves BP levels, decreases proteinuria even in advanced stage type 2 diabetic kidney disease patients with severe edema, and protects the kidneys [8]. Moreover, fluid overload refractory to loop diuretics can complicate acute or chronic heart failure management. Addition of thiazide-type diuretics can induce diuresis in patients refractory to loop diuretic doses [9], enhance sodium excretion, and prevent post-diuretic sodium retention after cessation of loop diuretic activity because thiazides have a longer half-life than loop diuretics [10].

Therefore, in clinical practice, we found that some patients received a combination of loop diuretics and thiazides. Careful monitoring of renal function, serum electrolytes, and fluid status to detect dehydration, hypokalemia, hyponatremia, or renal dysfunction is recommended [7].

References:
7. Rossignol P, Massy ZA, Azizi M, et al. The double challenge of resistant hypertension and chronic kidney disease. Lancet. 2015;386(10003):1588-98.
8. Hoshino T, Ookawara S, Miyazawa H, et al. Renoprotective effects of thiazides combined with loop diuretics in patients with type 2 diabetic kidney disease. Clin Exp Nephrol. 2015;19(2):247-53.
9. Jentzer JC, DeWald TA, Hernandez AF. Combination of loop diuretics with thiazide-type diuretics in heart failure. J Am Coll Cardiol. 2010;56(19):1527-34.
10. ter Maaten JM, Valente MA, Damman K, et al. Diuretic response in acute heart failure-pathophysiology, evaluation, and therapy. Nat Rev Cardiol. 2015;12(3):184-92.
11. Tamargo J, Segura J, Ruilope LM. Diuretics in the treatment of hypertension. Part 1: thiazide and thiazide-like diuretics. Expert Opin Pharmacother. 2014;15(4):527-47.
12. Tamargo J, Segura J, Ruilope LM. Diuretics in the treatment of hypertension. Part 2: loop diuretics and potassium-sparing agents. Expert Opin Pharmacother. 2014;15(5):605-21.

Chronic kidney disease 360 (9.5%) 349 (9.2%)
Again, this similarity is rather inplausible given that in advanced renal insufficiency thiazide diuretics are no longer effective. How did you define CKD?

Reply:
Similarity in the prevalence of CKD between groups was because of propensity score matching. A preexisting comorbidity such as CKD was defined as patients with diagnosis during at least one hospital admission or two outpatient visits in the year preceding the index date. The diagnosis of CKD

was according to ICD-9-CM code (403.X, 404.X, 582.X, 585.X) [13].

Chronic kidney disease is a general term for heterogeneous disorders affecting the structure and function of the kidney [14]. The Taiwan Society of Nephrology adopted the simplified Modification of Diet in Renal Disease equation to calculate the estimated glomerular filtration rate (eGFR) [15]. Unfortunately, the exact eGFR and CKD stage were not available from our claims-based dataset. However, the 2013 ESH/ESC guidelines recommend thiazide diuretics in hypertensives with eGFR ≥ 30 ml/min/1.73 m2 [16]. Moreover, current studies revealed that patients with advanced renal insufficiency and hypertension may still benefit from the combination of thiazides and loop diuretics, as mentioned above [7, 8]. Therefore, a part of CKD patients were allocated to the thiazide cohort.

References:
7. Rossignol P, Massy ZA, Azizi M, et al. The double challenge of resistant hypertension and chronic kidney disease. Lancet. 2015;386(10003):1588-98.
8. Hoshino T, Ookawara S, Miyazawa H, et al. Renoprotective effects of thiazides combined with loop diuretics in patients with type 2 diabetic kidney disease. Clin Exp Nephrol.
13. Chen J-S, Lu C-L, Huang L-C, et al. Chronic Kidney Disease is Associated With Upper Tract Urothelial Carcinoma: A Nationwide Population-Based Cohort Study in Taiwan. Medicine. 2016;95(14):e3255.
14. Webster AC, Nagler EV, Morton RL, et al. Chronic Kidney Disease. Lancet. 2017;389(10075):1238-52.
15. Hsu CC, Hwang SJ, Wen CP, et al. High prevalence and low awareness of CKD in Taiwan: a study on the relationship between serum creatinine and awareness from a nationally representative survey. Am J Kidney Dis. 2006;48(5):727-38.
16. Mancia G, Fagard R, Narkiewicz K, et al. 2013 ESH/ESC guidelines for the management of arterial hypertension: the Task Force for the Management of Arterial Hypertension of the European Society of Hypertension (ESH) and of the European Society of Cardiology (ESC). Eur Heart J. 2013;34(28):2159-219.

==========

Reviewer: 2
Prof Johnny Yu JIANG
Peking Union School of Public Health, PUMC & CAMS, China
Please state any competing interests or state ‘None declared’: None declared

==========

Please leave your comments for the authors below

This is an interesting population-based cohort study based on Taiwan's National Health Insurance Research Database. It indicated that long-term use of thiazides is associated with a decreased hip fracture risk after stroke, which may provide evidence-based information for hip fracture prevention of stoke patients. But there are some questions need to be clarified in the paper.

Reply:
Thank you for considering our manuscript for further revision.

---

1. There were 4382 patients with thiazide use in the database initially, but the number reduced to 3810 after propensity score matching, were there any differences between the population included in the analysis and the population excluded? Why did not you include all 12820 patients as research subjects since all the potential confounders you balanced by propensity score matching were also adjusted in the cox regression?

Reply:
In this revised manuscript with improved study design (such as matching and adjusting the osteoporosis medication, modifying the exclusion criteria, improving the classification of patients, and adjusting the competing mortality), 4303 patients with thiazide use were initially identified. After propensity score matching, 3,735 patients were included for analysis and 568 patients were excluded because of not pairing. To answer your question, we evaluated the baseline characteristics and found differences in sex distribution, proportion of hypertension, congestive heart failure, chronic kidney disease, osteoporosis, dementia, depression, rheumatoid arthritis, and some of baseline medications in the 3,735 patients included for the analysis and the 568 patients excluded because of not pairing. The purpose of propensity score matching was to balance the baseline characteristics between the thiazide and non-thiazide cohorts; therefore, it was not surprising that there are some difference between the included 3,735 patients and the excluded 568 patients when performing propensity score matching with nearest-neighbor algorithm. Actually, if there were no differences between the included and excluded patients, the matching procedure would lose its meaning. Regarding the details describing how unmatched treated subject were excluded from the resultant matched sample during propensity score matching, kindly refer to the article published by Professor Peter Austin [17].

To eliminate the possibility of different study results between performing propensity score matching or not, we conducted analyses that included all patients (4303 patients in the thiazide cohort and 7849 patients in the non-thiazide cohort) without matching and only performed multivariate Cox regression model with adjustment of baseline characteristics. Between these two methods, similar results and the same conclusion was obtained. The main outcomes of the analysis without propensity score matching is described here. Patients with thiazide use also had a lower hip fracture risk (adjusted HR = 0.64, 95%CI = 0.48–0.87, p = 0.004). Analyses based on the duration of thiazide use revealed that the effect was significant in patients with long-term use of thiazides (adjusted HR = 0.47, 95% CI = 0.26–0.84, p = 0.011). Analyses with or without performing propensity score matching revealed the same conclusions, and the consistency in these two different methods further strengthens the confidence in our findings.

However, we believe that performing the propensity score matching to balance all the baseline characteristics and then using a multivariate Cox regression model to control residual confounding effects is more appropriate. Our rationale for this approach is mentioned below.

According to Professor Peter Austin, the propensity score allows one to design and analyze an observational (nonrandomized) study so that it mimics some of the particular characteristics of a randomized controlled trial [17]. In an article entitled, “An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies,” he discussed the differences between regression-based methods and propensity score-based methods for the analysis of observational data.

There are several practical reasons for preferring the use of propensity score-based methods over regression-based methods when estimating treatment effects using observational data. The comments below were directly extracted from the article published by Professor Peter Austin [17].
“First, it is simpler to determine whether the propensity score model has been adequately specified than to assess whether the regression model relating treatment assignment and baseline covariates

to the outcome has been correctly specified."
"Second, propensity score methods allow one to separate the design of the study from the analysis of the study. This is similar to an RCT, in which the study is designed first; only after the study has been completed is the effect of treatment on the outcome estimated. However, when using regression adjustment, the outcome is always in sight, and the researcher is faced with the subtle temptation to continually modify the regression model until the desired association has been achieved [18]."
"Third, there may be increased flexibility when outcomes (when binary or time-to-event in nature) are rare and treatment is common [19]."
"Fourth, one can explicitly examine the degree of overlap in the distribution of baseline covariates between the two treatment groups."

Moreover, Heinze et al. had concluded that propensity analyses may help in evaluating the comparability of patients in observational studies and may account for more potential confounding factors than conventional covariate adjustment approaches [20].

Therefore, we preferred conducting propensity score matching rather than only using Cox regression model to adjust the baseline characteristics, although these two methods revealed the same conclusions, as mentioned above.

References:
17. Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivariate Behav Res. 2011;46(3):399-424.
18. Rubin DB. Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation. Health Services and Outcomes Research Methodology. 2001;2(3):169-88.
19. Braitman LE, Rosenbaum PR. Rare outcomes, common treatments: analytic strategies using propensity scores. Ann Intern Med. 2002;137(8):693-5.
20. Heinze G, Juni P. An overview of the objectives of and the approaches to propensity score analyses. Eur Heart J. 2011;32(14):1704-8.

---

2. Please explain the reason you excluded patients aged<20 years in this study.

Reply:
We aimed to evaluate the adult patients with stroke. In clinical practice, most of the stroke and hypertension patients with thiazide use are adults. According to Taiwan's civil law, the age of adult is defined as at least 20 years old. Therefore, the patients with age < 20 years were excluded.

---

3. The method you used to calculate propensity score was not given in the paper. Please specify.

Reply:
We apologize for the unclear statement about the method of calculating propensity score and matching.

For each thiazide user and nonuser, a propensity score was calculated to estimate the probability of thiazide use by logistic regression model with all baseline covariates listed in Table 1. We matched each thiazide user with a thiazide nonuser with a similar propensity score based on nearest-neighbor matching without replacement using a caliper width equal to 0.2 of the standard deviation of the logit of the propensity score [17, 21, 22].

The statements about the propensity score method were added in the Covariates and propensity score matching section in METHODS (p.9).

References:
17. Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivariate Behav Res. 2011;46(3):399-424.
21. Chang CC, Chen YT, Hsu CY, et al. Dipeptidyl Peptidase-4 Inhibitors, Peripheral Arterial Disease, and Lower Extremity Amputation Risk in Diabetic Patients. Am J Med. 2017;130(3):348-55.
22. Leuven E, Sianesi B. PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing 2003 [cited 2017 May 5]. version 4.0.11:[Available from: http://ideas.repec.org/c/boc/bocode/s432001.html.

---

4. The proportion of hip fracture due to falls in the two groups respectively can be explained and further analyzed in the paper if possible, since there are concerns that diuretics may increase the risk of fall in the elderly.

Reply:
We apologize that we could not obtain the proportion of hip fractures due to falls in the two respective groups as we could not determine the actual mechanism of hip fractures according to the claim-based data.

---

5. Was there any possibility that the concomitant drugs were different between the two groups that may influence the results?

Reply:
Yes, in our revised manuscript, the osteoporosis medication was added as one of the baseline characteristics and was used for propensity score matching and adjusted in multivariate Cox regression model. The use of osteoporosis medication was defined as any prescription of medication including bisphosphonates, hormone replacement therapy, calcitonin, or vitamin D supplementation available in Taiwan [6].

In summary, after matching and adjusting the osteoporosis medication as well as improving the study design (such as modifying the exclusion criteria, improving the classification of patients, and adjusting the competing mortality), the relationship between thiazide use and decreased hip fracture risk became more significant, as reported in our revised manuscript (Table 2 and Table 3). In addition, the difference in hip fracture risk between thiazide user and nonuser became significant (Table 2) in the revised manuscript; however, the conclusion remained the same that only long-term use of thiazides is associated with a decreased risk of hip fracture after stroke (Table 3).

Reference:
6. Hwang JS, Chan DC, Chen JF, et al. Clinical practice guidelines for the prevention and treatment of osteoporosis in Taiwan: summary. J Bone Miner Metab. 2014;32(1):10-6.

==========

Reviewer: 3

KoKo Aung, MD, MPH
Paul L. Foster School of Medicine, Texas Tech University Health Sciences Center at El Paso, U.S.A.
Please state any competing interests or state 'None declared': None declared
==========
Please leave your comments for the authors below

Overall, this is a well written manuscript of an observational study that addresses a knowledge gap related to effects of thiazide on hip fracture in patients who had stroke. The research question was clearly defined. The database and the methods used were appropriate. The manuscript was written in a very clear and concise manner. The tables were very clear for the readers to follow, understand, and interpret. Addressing the questions and concerns listed below would further strengthen the already strong manuscript.

Reply:
Thank you for your appreciation and pertinent opinion and for considering our manuscript for further revision.

---

1. How was the decision arrived to perform propensity score matching instead of directly adjusting for all covariates that were used to calculate the propensity score?

Reply:
We believe that performing the propensity score matching to balance all the baseline characteristics and then using a multivariate Cox regression model to control for residual confounding effects was more appropriate than only adjusting for covariates by Cox regression model. Our rationale for this approach is mentioned below.

According to Professor Peter Austin, the propensity score allows one to design and analyze an observational (nonrandomized) study so that it mimics some of the particular characteristics of a randomized controlled trial [17]. In an article entitled, "An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies," he discussed the differences between regression-based methods and propensity score-based methods for the analysis of observational data.

There are several practical reasons for preferring the use of propensity score-based methods over regression-based methods when estimating treatment effects using observational data. The comments below were directly extracted from the article published by Professor Peter Austin [17].
"First, it is simpler to determine whether the propensity score model has been adequately specified than to assess whether the regression model relating treatment assignment and baseline covariates to the outcome has been correctly specified."
"Second, propensity score methods allow one to separate the design of the study from the analysis of the study. This is similar to an RCT, in which the study is designed first; only after the study has been completed is the effect of treatment on the outcome estimated. However, when using regression adjustment, the outcome is always in sight, and the researcher is faced with the subtle temptation to continually modify the regression model until the desired association has been achieved [18]."
"Third, there may be increased flexibility when outcomes (when binary or time-to-event in nature) are rare and treatment is common [19]."
"Fourth, one can explicitly examine the degree of overlap in the distribution of baseline covariates

between the two treatment groups."
Moreover, Heinze et al. had concluded that propensity analyses may help in evaluating the comparability of patients in observational studies and may account for more potential confounding factors than conventional covariate adjustment approaches [20].
Therefore, we preferred conducting propensity score matching rather than only using Cox regression model to adjust the baseline characteristics.

To eliminate the possibility of different study results between performing propensity score matching or not, we conducted analyses that included all patients (4303 patients in the thiazide cohort and 7849 patients in the non-thiazide cohort) without matching and only performed multivariate Cox regression model with adjustment of baseline characteristics. The results and conclusion between these two methods were similar. The main outcomes of the analysis without propensity score matching is described here. Patients with thiazide use also had a lower hip fracture risk (adjusted HR = 0.64, 95%CI = 0.48–0.87, p = 0.004). Analyses based on the duration of thiazide use revealed that the effect was significant in patients with long-term use of thiazides (adjusted HR = 0.47, 95% CI = 0.26–0.84, p = 0.011). Analyses with or without performing propensity score matching revealed the same conclusions, and the consistency in these two different methods further strengthens the confidence in our findings.

References:
17. Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivariate Behav Res. 2011;46(3):399-424.
18. Rubin DB. Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation. Health Services and Outcomes Research Methodology. 2001;2(3):169-88.
19. Braitman LE, Rosenbaum PR. Rare outcomes, common treatments: analytic strategies using propensity scores. Ann Intern Med. 2002;137(8):693-5.
20. Heinze G, Juni P. An overview of the objectives of and the approaches to propensity score analyses. Eur Heart J. 2011;32(14):1704-8.

---

2. It was reported that propensity score matching was performed in this study to balance covariates and baseline differences, including age, sex, baseline comorbidities, socioeconomic factors, stroke severity proxies and use of medications as listed in Table 1. The patients who had thiazide use after stroke were matched (without replacement) with those who did not, and a nearest-neighbor algorithm was applied to construct matched pairs. After matching, 7620 patients with ischemic stroke without previous hip fracture were included. It appears that matches could be found for 3810 thiazide users, among patients with ischemic stroke without previous hip fracture.
a. How many thiazide users (among patients with ischemic stroke without previous hip fracture) were left unmatched and thus excluded from the study?

Reply:
In the revised manuscript with improvement of study design (such as matching and adjusting the osteoporosis medication, modifying the exclusion criteria, improving the classification of patients, and adjusting the competing mortality), 4303 patients with thiazide use were initially identified. After propensity score matching, 3,735 patients were included for analysis and 568 patients were excluded because of not pairing. This has been included in Figure 1 in our revised manuscript.

Regarding the details describing how unmatched treated subject were excluded from the resultant matched sample during propensity score matching, kindly refer to the explanations below or the article published by Professor Peter Austin [17].

Reference:
17. Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivariate Behav Res. 2011;46(3):399-424.

---

b. What was the acceptable “neighborhood” of propensity scores for matching purpose?

Reply:
For each thiazide user and nonuser, a propensity score was calculated to estimate the probability of thiazide use by logistic regression model with all baseline covariates listed in Table 1. We matched each thiazide user with a thiazide nonuser with a similar propensity score based on nearest-neighbor matching without replacement using a caliper width equal to 0.2 of the standard deviation of the logit of the propensity score [17, 21, 22]

Regarding the method of nearest-neighbor matching, kindly refer to the explanations below, which is directly extracted from the article published by Professor Peter Austin [17].
“Nearest-neighbor matching within a specified caliper distance make restriction that the absolute difference in the propensity scores of matched subjects must be below some prespecified threshold (the caliper distance). Thus, for a given treated subject, one would identify all the untreated subjects whose propensity score lay within a specified distance of that of the treated subject. From this restricted set of untreated subjects, the untreated subject whose propensity score was closest to that of the treated subject would be selected for matching to this treated subject. If no untreated subjects had propensity scores that lay within the specified caliper distance of the propensity score of the treated subject, that treated subject would not be matched with any untreated subject. The unmatched treated subject would then be excluded from the resultant matched sample.”

As per previous studies, a caliper width equal to 0.2 of the standard deviation was suggested. For further details regarding the nearest-neighbor matching, kindly refer to the article published by Professor Peter Austin [17].

References
17. Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivariate Behav Res. 2011;46(3):399-424.
21. Chang CC, Chen YT, Hsu CY, et al. Dipeptidyl Peptidase-4 Inhibitors, Peripheral Arterial Disease, and Lower Extremity Amputation Risk in Diabetic Patients. Am J Med. 2017;130(3):348-55.
22. Leuven E, Sianesi B. PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing 2003 [cited 2017 May 5]. version 4.0.11:[Available from: http://ideas.repec.org/c/boc/bocode/s432001.html.

---

c. What factors were weighed in to balance the trade-off between inexact matching (and residual confounding) and incomplete matching (and loss of statistical power) in this study? The trend of fewer hip fractures among thiazide users in this study without reaching the level of statistical significance (p = 0.059 in table 2) raised the question on whether there was incomplete matching.

Reply:
According to the textbook “Methods in Social Epidemiology” published by J. Michael Oakes and Jay S. Kaufman [23], estimated propensity scores can be used in three general ways: (i) matching, (ii) subclassification, or (iii) as a regression covariate or sampling weight; each has trade-offs [24]. If we use the propensity score for direct matching, we can achieve better covariate balance; however, it is

possible that not all subjects will be matched, so we throw away some information/data. Using the propensity score for subclassification permits retention of all data, but minor covariate imbalances as well as some support problems may be found. Using a propensity score as a regression covariate or weight in social epidemiology can yield troubling off-support inference; therefore, it was discouraged for using [23].

Overall, to balance the trade-off between inexact matching and incomplete matching, the propensity score matching with algorithm of the nearest-neighbor within calipers was suggested [17, 23]. However, there is also a trade-off between specifying tight calipers and loose calipers, which is directly related to the trade-off between incomplete and inexact matching. A caliper width equal to 0.2 of the standard deviation (SD) was optimal and suggested by Professor Peter Austin according to previous studies [17]; therefore, it was chosen in our present study.
Moreover, to eliminate the possibility that different values of caliper chosen influenced our results, we tested the propensity score matching with various calipers (including 0.1 of SD, 0.2 of SD, and 0.4 of SD), and all of them revealed similar results and the same conclusion.

In addition, after improving study design in our revised manuscript (such as matching and adjusting the osteoporosis medication, modifying the exclusion criteria, improving the classification of patients, and adjusting the competing mortality), the relationship between thiazide use and decreased hip fracture risk became more significant (Table 2 and Table 3). The difference in hip fracture risk between thiazide user and nonuser became significant ($p = 0.007$, Table 2) in the revised manuscript; however, the conclusion was the same that only long-term use of thiazides is associated with a decreased risk of hip fracture after stroke (Table 3).

References
17. Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivariate Behav Res. 2011;46(3):399-424.
23. Oakes JM, Kaufman JS. Methods in Social Epidemiology. second edition ed: Wiley; 2017.
24. Stuart EA. Matching methods for causal inference: A review and a look forward. Stat Sci. 2010;25(1):1-21.

---

d. How were missing data related to more than 30 covariates in table 1 handled when calculating propensity scores?

Reply:
The missing data were only found in sex (n = 3) and urbanization level (n = 69) when we started to approach the stroke patient initially; therefore, these patients, accounting for less than 1% of study population, were initially excluded. Because our study was based on National Health Insurance Research Database, not surprisingly, there were no missing data found in age and income level, which were important in calculating the insurance premiums. Comorbidities, stroke severity proxies, and information about use of medication listed in Table 1 were identified by whether or not the corresponding diagnosis code, procedure code, prescription code, and drug code were found. Therefore, there were no missing data in these baseline covariates.

---

3. Are there any patients with more than one hip fracture within 2 years follow-up period? If so, how was the data handled? Do 76 and 92 hip fractures in table 2 represent 76 and 92 unique patients who sustained hip fractures or merely the number of hip fractures?

Reply:
In our study, all subjects were followed from the index date until a new diagnosis of hip fracture, death, or 2 years after stroke. Therefore, when a new hip fracture occurred during the follow-up period, the patient was censored. We did not identify whether more than one hip fracture occurred. Therefore, the number of fracture in Table 2 represents the number of patients who sustained hip fractures.

---

4. Please clarify whether propensity score was included as one of the covariates in regression analysis that evaluates effect of thiazides.

Reply:
As mentioned above, estimated propensity scores can be used in three general ways: (i) matching, (ii) subclassification (stratification), or (iii) as a regression covariate or sampling weight [23]. Studies have demonstrated that propensity score matching eliminates a greater proportion of the systematic differences in baseline characteristics between treated and untreated subjects than does stratification on the propensity score or covariate adjustment using the propensity score [17, 25, 26]. Therefore, we used propensity score matching method rather than adjusting propensity score in regression analysis. The propensity score was not included as one of the covariates in regression analysis.

References
17. Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivariate Behav Res. 2011;46(3):399-424.
23. Oakes JM, Kaufman JS. Methods in Social Epidemiology. second edition ed: Wiley; 2017.
25. Austin PC. Type I error rates, coverage of confidence intervals, and variance estimation in propensity-score matched analyses. Int J Biostat. 2009;5(1):Article 13.
26. Austin PC, Mamdani MM. A comparison of propensity score methods: a case-study estimating the effectiveness of post-AMI statin use. Stat Med. 2006;25(12):2084-106.

VERSION 2 – REVIEW

REVIEWER	KoKo Aung Paul L. Foster School of Medicine Texas Tech University Health Sciences Center at El Paso U.S.A.
REVIEW RETURNED	19-Jun-2017

GENERAL COMMENTS	The revised manuscript with modified design has substantially stronger than its original form. Please consider the following minor revisions if space permits: (1) Include a brief information on time to fracture for patients who sustained hip fracture during follow-up period, grouping them into thiazide users (all users and subgroup according to duration of use) and non-thiazide users. (2) Include comparison of main outcome of the analysis without propensity score matching and only with adjustment of multiple covariates used to calculate propensity score. Since the results obtained by these two different methods were strikingly similar, it will serve as a form of sensitivity analysis and will give the readers the information to strengthen the confidence in the findings and conclusions of this study. The authors have already conducted this analysis so it should not be a huge burden to include this information

	in the manuscript.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 3
KoKo Aung
Paul L. Foster School of Medicine
Texas Tech University Health Sciences Center at El Paso. U.S.A.
Please state any competing interests or state 'None declared': None declared
==========
Please leave your comments for the authors below

The revised manuscript with modified design has substantially stronger than its original form. Please consider the following minor revisions if space permits:

Reply:
Thank you for considering our manuscript and suggesting minor revisions.

---

(1) Include a brief information on time to fracture for patients who sustained hip fracture during follow-up period, grouping them into thiazide users (all users and subgroup according to duration of use) and non-thiazide users.

Reply:
Thank you for your suggestion. In our revised manuscript, this information has been included in RESULTS (p.13,14).
The average time from index date to hip fracture for patients who developed hip fracture was 1.09 years in thiazide cohort and 0.89 years in non-thiazide cohort. For each subgroup in the thiazide cohort, the average time from index date to hip fracture was 0.99 years in patients using thiazides for 1–90 days, 1.08 years in those using thiazides for 91–365 days, and 1.39 years in those using thiazides for >365 days.

---

(2) Include comparison of main outcome of the analysis without propensity score matching and only with adjustment of multiple covariates used to calculate propensity score. Since the results obtained by these two different methods were strikingly similar, it will serve as a form of sensitivity analysis and will give the readers the information to strengthen the confidence in the findings and conclusions of this study. The authors have already conducted this analysis so it should not be a huge burden to include this information in the manuscript.

Reply:
Yes, in our revised manuscript, we have included the sensitivity analyses that included all patients without matching, and only multivariate Cox regression model was applied with the adjustment of baseline characteristics. The results between these two methods (with and without propensity score matching) were similar. Detailed outcomes were shown in the Supplementary material (Table S1 for analysis depending on thiazide use; Table S2 for analysis based on the duration of thiazide use). We have mentioned this information in METHOD (p.9) and RESULTS (p.13,14) in our revised manuscript. Thank you for your suggestion, which will help to further strengthen confidence in our findings.

VERSION 3 – REVIEW

REVIEWER	KoKo Aung Paul L. Foster School of Medicine Texas Tech University Health Sciences Center at El Paso United States of America
REVIEW RETURNED	02-Jul-2017

The reviewer completed the checklist but made no further comments.