

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)	The analyses and identification of ICD codes for dementias in the research based on the NHIRD: A scoping review protocol.
AUTHORS	Shih, Ying-Jyun; Wang, Jiun-Yi; Wang, Ya-Hui; Shih, Rong-Rong; Yang, Yung-Jen

VERSION 1 – REVIEW

REVIEWER	Mateos-Alvarez, Raimundo University of Santiago de Compostela, Department of psychiatry
REVIEW RETURNED	04-Apr-2022

GENERAL COMMENTS	<p>This is a Scoping Review of research work conducted in Taiwan on dementia using clinical data from the Taiwan National Health Insurance Research Database (NHIRD). The rationale for conducting this review is clearly stated, as well as a brief introduction to the concept of scoping reviews. Likewise, the content of the research protocol is clearly stated, and its sections are justified.</p> <p>We do not appreciate any major flaw in the data collection protocol of the works that will be reviewed. All variables seem relevant to us and we do not consider any of them are superfluous.</p> <p>Overall, we believe that this is an interesting article, focused on a field of research that is growing rapidly (Health claim data), that it constitutes an example of the type of revisions that will be increasingly necessary, and that it is didactic from an academic scientific point of view.</p> <p>We only make minor suggestions to the authors that could enrich the protocol. But we understand that its viability may depend on the sociological and cultural circumstances of the country in which this research will be carried out (Taiwan).</p> <p>MINOR SUGGESTIONS TO AUTHORS</p> <p>1. Exclusion criteria. Consider DO NOT exclude: study protocols, grey literature and texts that are not peer-reviewed. We suggest including any work that clearly includes indications of the variables that have been used or are intended to be used to investigate dementia in Health claim data in Taiwan. Rationale: All these potential articles are of interest to know the "state of opinion" of researchers in Taiwan, which can only be improved if it is fully known. If the authors consider it practical, they can analyze the grey literature in a separate group.</p> <p>2. Consider expanding the concept of "dementia": including also "cognitive impairment". Rationale: In the clinical context of Western countries, "cognitive impairment" is often used to refer to dementia (syndrome). We</p>
-------------------------	---

	<p>understand that the authors should assess this suggestion based on the language barrier and the health context of Taiwan.</p> <p>3. In the Methods section, consider including a new variable/s. Whichever code was used in the articles reviewed, has any empirical validation work been done on that code? Rationale: We consider that this variable is important in order to know if the diagnostic codes are being used rigorously in the national clinical data base. This information is important to know the degree of precision with which these codes are used, at least what is their sensitivity and specificity to designate dementias in a clinical context. This will be critical in epidemiological studies of the prevalence and incidence of dementias.</p>
--	--

REVIEWER	Wilkinson, Tim The University of Edinburgh
REVIEW RETURNED	15-Apr-2022

GENERAL COMMENTS	<p>This is a protocol for a scoping review with the intention of identifying and summarising all studies of dementia using the Taiwan National Health Insurance Research Database. The protocol appropriately outlines the background, aims and search strategy. The PRISMA-ScR checklist is included.</p> <p>Often, scoping reviews are done non-systematically and informally. The authors should be congratulated on formalising this process and making their methods transparent and replicable. A good breadth of databases will be searched, both in English and Chinese.</p> <p>I note the BMJ Open reviewer guidance, which states that for studies that are ongoing, very few changes can be made to the methodology. With this in mind, I have focussed this review mainly on the other aspects of the protocol.</p> <p>I have a few minor points of feedback:</p> <p>1. Background: 1.1 A bit of background on Taiwan's healthcare system and the NIHRD database would be helpful for those unfamiliar with this. For example, does it include primary (general practice) and secondary care (hospital)? Are mortality data included? What coding systems are used? Do physicians code for the purposes of billing, or does a third party produce the codes?</p> <p>2. Data extraction table 2.1 One of the variables being collected is 'main discipline of the lead author' - how will this be defined/identified? 2.2 For 'the number of dementias identified', I presume this means the number of people in the study who had dementia, rather than the number of different dementia subtypes (e.g. Alzheimer's disease, vascular dementia) measured? 2.3 Related to 2.2, although 'the number of dementias identified' is listed under 'Outcomes', some of the studies may only include people with dementia, so dementia would not be an outcome in these studies (e.g. studies of prognosis in people with dementia, studies of prescribing trends in people with dementia). Perhaps renaming the 'Outcomes' section as 'Results', may help avoid this confusion.</p>
-------------------------	--

	<p>2.4 Related to 2.3 above, some of the fields in 'Outcomes' could probably benefit from being moved to 'Methods'. For example, the authors would first extract information on what the study's main and secondary outcomes were (e.g. whether dementia diagnosis, mortality, care home admission...), as these are part of that study's methods. Then each study's results would go in the 'Results/Outcomes' section.</p> <p>2.5 What will go under 'research method used'?</p> <p>3. General</p> <p>3.1 The use of the term 'researches' is a bit confusing, as at first I thought it was a typo for 'researchers'. I think 'research' would be sufficient wherever the word 'researches' has been used (including the title).</p> <p>4. References</p> <p>4.1 The end of reference 5 is missing - it should read: 2022;51(1):afab231.</p>
--	--

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1. Prof. Raimundo Mateos-Alvarez, University of Santiago de Compostela

Comments to the Author:

This is a Scoping Review of research work conducted in Taiwan on dementia using clinical data from the Taiwan National Health Insurance Research Database (NHIRD).

The rationale for conducting this review is clearly stated, as well as a brief introduction to the concept of scoping reviews. Likewise, the content of the research protocol is clearly stated, and its sections are justified.

We do not appreciate any major flaw in the data collection protocol of the works that will be reviewed.

All variables seem relevant to us and we do not consider any of them are superfluous.

Overall, we believe that this is an interesting article, focused on a field of research that is growing rapidly (Health claim data), that it constitutes an example of the type of revisions that will be increasingly necessary, and that it is didactic from an academic scientific point of view.

We only make minor suggestions to the authors that could enrich the protocol. But we understand that its viability may depend on the sociological and cultural circumstances of the country in which this research will be carried out (Taiwan).

MINOR SUGGESTIONS TO AUTHORS

1. Exclusion criteria. Consider DO NOT exclude: study protocols, grey literature and texts that are not peer-reviewed. We suggest including any work that clearly includes indications of the variables that have been used or are intended to be used to investigate dementia in Health claim data in Taiwan. Rationale: All these potential articles are of interest to know the "state of opinion" of researchers in Taiwan, which can only be improved if it is fully known. If the authors consider it practical, they can analyze the grey literature in a separate group.

Response:

We thank review's suggestions about not excluding some types of literatures and we have made amendments of our exclusion criteria by adding conditions on those literatures as: [...] review articles, study protocols, grey literatures and texts that are not peer-reviewed or fail to provide detailed information that is in line with our study are excluded. (Line 4, Page 9) We also agree with the reviewer's opinion on analysing the grey literature in the subgroup analysis. We have added a variable labelled as "type of the literature" in the electronic data extraction form to signify the individual article that can be further analysed (refer to Table 2 on Page 11, Line 9 on Table 2) and explained the subgroup analysis as: "We will also perform a subgroup analysis based on the type of literature to

compare whether there are differences between the formal and grey literatures.”. (Line 15 to 17, Page 12)

We still keep the condition that such types of literatures should be peer-reviewed for the following considerations:

- (1) Difficulties in information access: Empirically, it is not a requirement to submit or even open the study protocol when performing studies using the NHIRD in Taiwan. Even if there to be protocols, they are not open to the researchers and often not accessible.
- (2) The reliability issues: It is not uncommon that there are deviations between protocols and the final execution/publications. To ensure the quality of the literature enrolled, we believe that the condition of “peer-reviewed” is still necessary.

2. Consider expanding the concept of “dementia”: including also “cognitive impairment”.

Rationale: In the clinical context of Western countries, “cognitive impairment” is often used to refer to dementia (syndrome). We understand that the authors should assess this suggestion based on the language barrier and the health context of Taiwan.

Response:

We appreciate the suggestions to consider including “cognitive impairment” in the concept of “dementia”, but, however, we argue for not doing so for the following reasons. First, as the reviewer mentioned, there are contextual issues for rhetoric choices (e.g., language barrier, while cognitive impairment is not an exact term for dementia in local language). Secondly, as the NHIRD is a claim database, the codes for dementia are usually used for national health insurance service reimbursements and the diagnosis of dementia is usually a key hinge for the following medical and social services. In contrast to dementia, there is no specific intervention indicated for the condition of “cognitive impairment”, and we assume that most physicians will use dementia (and its relevant codes) to mean for dementia. Finally, as the results of our study can be served for informing policymaking or healthcare administration where the decision-makers are usually familiar with the dichotomous choices of “dementia” or “non-dementia”, we are thus worried that expending “dementia” by including “cognitive impairment” in the study will blur the focus.

3. In the Methods section, consider including a new variable/s. Whichever code was used in the articles reviewed, has any empirical validation work been done on that code?

Rationale: We consider that this variable is important in order to know if the diagnostic codes are being used rigorously in the national clinical data base. This information is important to know the degree of precision with which these codes are used, at least what is their sensitivity and specificity to designate dementias in a clinical context. This will be critical in epidemiological studies of the prevalence and incidence of dementias.

Response:

We thank the suggestion of the reviewer, and we have added a variable of “Whether and what methods used to validate the selection of diagnostic codes to define dementia” in the data extraction form and will examine whether the enrolled individual report had tested the validation of the codes used (refer to Table 2 on Page 11, Line 21 to 22 on Table 2).

Reviewer: 2. Dr. Tim Wilkinson, The University of Edinburgh

Comments to the Author:

This is a protocol for a scoping review with the intention of identifying and summarising all studies of dementia using the Taiwan National Health Insurance Research Database.

The protocol appropriately outlines the background, aims and search strategy. The PRISMA-ScR checklist is included.

Often, scoping reviews are done non-systematically and informally. The authors should be congratulated on formalising this process and making their methods transparent and replicable. A good breadth of databases will be searched, both in English and Chinese.

I note the BMJ Open reviewer guidance, which states that for studies that are ongoing, very few changes can be made to the methodology. With this in mind, I have focussed this review mainly on the other aspects of the protocol.

I have a few minor points of feedback:

1. Background:

1.1 A bit of background on Taiwan's healthcare system and the NHIRD database would be helpful for those unfamiliar with this. For example, does it include primary (general practice) and secondary care (hospital)? Are mortality data included? What coding systems are used? Do physicians code for the purposes of billing, or does a third party produce the codes?

Response:

We thank the precious suggestion of the reviewer, and we have revised our "Introduction" section by adding 3 paragraphs to provide more information about the NHIRD database. The revision is shown as the 3 added paragraphs from Line 17 on Page 6 to Line 13 on Page 7 in the "Introduction" section. For easier to identify the revision, the added descriptions are also excerpted here: "Since 1995, Taiwan has launched the Taiwan National Health Insurance (THNI) with the coverage of 99.9% of the whole 23 million population and established an HCD database which cumulates the health-related records of the users in the national health insurance system.[26, 27] In practice, although increasing hospitals have begun to employ the clinical coders to help the task of coding in recent years, physicians across the levels in the health system are still the main persons responsible for coding and inputting the diagnostic codes as well as the interventional codes into the administrative systems for reimbursement in THNI. From 1995 to 2016, the diagnostic codes were based on the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM), which was replaced by ICD-10-CM after 2016.[27]

Thanks to its abundant information, the National Health Insurance Administration has built up the National Health Insurance Research Database (NHIRD) and released the purchasable datasets which included de-identifiable and encrypted sampling of the health records for the researchers in the academic organisations since 2000. At present time, there are three forms of the NHIRD datasets released at different chronological time, and they are general dataset from 2000, disease-specific dataset, and the latest full population dataset which was released since 2016. These datasets consist both inpatient and outpatient claim data, and the sets include demographic profiles as well as clinical data of the codes for diagnoses, prescriptions and interventions. From 2016, the NHIRD data were authorised to link with other government databases at the Data Science Centre, including some national census data, disease registries, health surveys, social service data, death cause data, and welfare registries. The linkage with other large national databases expands the applicability especially in research and health policymaking.[26, 27]

In recent years, the NHIRD has provided the researchers with abundant resource for secondary database medical research, and hundreds of studies have been published, including the research for dementia. Many of them have been used as references for healthcare practice guidance and public policymaking. Despite of its strengths, however, the NHIRD still bears the same inherent weakness of HCD and there have been inconsistencies in selecting the diagnostic codes in defining dementias in the research using the NHIRD. As a result, it is imperative to investigate the characteristics of dementia researches based on the NHIRD and how the diagnostic codes are selected as well as used in such studies. This will aid in identifying the potential research gap and reduce the research waste."

2. Data extraction table

2.1 One of the variables being collected is 'main discipline of the lead author' - how will this be defined/identified?

Response:

We appreciate that the reviewer noticed this interesting variable of “the main discipline of the lead author”. It is a locally hot issue which is possibly regarding publication ethics because it is quite common that many studies using the NHIRD have been completed by the researchers who are suspected not familiar with the topic they performed as they are not trained for the specialities. For example, a nuclear medicine physician has published 149 research articles in a single year using the NHIRD [1], and such situation raises the concerns about the quality (e.g., breadth and depth) of the study [2]. In our study, we simply intend to know how the situation would be in the field of dementia research using the NHIRD. For the example above, please refer to the following references.

Reference:

1. Chien TW, Chang Y, Wang HY. Understanding the productive author who published papers in medicine using National Health Insurance Database: A systematic review and meta-analysis. *Medicine*. 2018;97(8):e9967.
2. Hsieh WT, Chien TW, Kuo SC, Lin HJ. Whether productive authors using the national health insurance database also achieve higher individual research metrics: A bibliometric study. *Medicine (Baltimore)*. 2020;99(2):e18631.

2.2 For 'the number of dementias identified', I presume this means the number of people in the study who had dementia, rather than the number of different dementia subtypes (e.g. Alzheimer's disease, vascular dementia) measured?

Response:

We appreciate the reviewer's suggestion for the ambiguity of wording. We have revised the wording as: “the number of people with dementia identified” on Table 2 (Line 26 on Table 2, Page 11).

2.3 Related to 2.2, although 'the number of dementias identified' is listed under 'Outcomes', some of the studies may only include people with dementia, so dementia would not be an outcome in these studies (e.g. studies of prognosis in people with dementia, studies of prescribing trends in people with dementia). Perhaps renaming the 'Outcomes' section as 'Results', may help avoid this confusion.

2.4 Related to 2.3 above, some of the fields in 'Outcomes' could probably benefit from being moved to 'Methods'. For example, the authors would first extract information on what the study's main and secondary outcomes were (e.g. whether dementia diagnosis, mortality, care home admission...), as these are part of that study's methods. Then each study's results would go in the 'Results/Outcomes' section.

Response:

For the feedback 2.3 and 2.4, we thank and agree with the reviewer's suggestions that it is sometimes difficult to classify a variable into the field of “Method” or “Results”. In this way, we thus have revised the fields of “Method” and “Outcomes” to be a single field of “Method/Results” to avoid ambiguities caused by different perspectives. (Please refer to Table 2 on Page 11 for the revision.)

2.5 What will go under 'research method used'?

3. General

3.1 The use of the team 'researches' is a bit confusing, as at first I thought it was a typo for 'researchers'. I think 'research' would be sufficient wherever the word 'researches' has been used (including the title).

4. References

4.1 The end of reference 5 is missing - it should read: 2022;51(1):afab231.

Response:

For point 2.5, 3 and 4, we thank the reviewer's meticulous inspection and pointing out our editing and grammar errors. We have deleted the item of “research method used” on Table 2 which was caused by editing errors and corrected the inappropriate spelling as well the missing reference detail.

For both reviewers:

1. After discussing with the librarians and performing the pilot search during the period waiting for the reviewers' comments, we have amended the Table 1 and move the search terms of "National Health Insurance Research Database (NHIRD)", "National Health Programme/Service", "National data (or nationwide or population-based) research" and "Taiwan Longitudinal (or follow-up) Study" from the Concept 3 to Concept 2. (Please refer to Table 1, on Page 10) Such changes would be better in line with the Participant-Concept-Context (PCC) framework for scoping review and could increase the sensitivity of the search strategies as supported by the increased number of records at our pilot search.
2. Because of the decreased time for the researchers as a result of recent wave of COVID-19 Omicron variant burst in Taiwan and the time needed for the reviewing process, we have to extend the endpoint of the study from August to 31 December 2022 to allow the research team ample time to execute the task. (Line 15-16, Page 2 (the Abstract), and Line 12, Page 8) We have simultaneously updated this change of our protocol on the protocol registry site at the Open Science Framework (OSF: osf.io/fc65g).
3. We also noticed that the original description of third research aim was not clearly stated and could possibly bring about confusions or misunderstandings for the readers. For better clarification, we have modified the descriptions of the third research aim as "What differences in terms of the additional approaches other than diagnostic codes used in the inclusion or exclusion criteria to identify the individuals with dementias in the databases and the time length of database were adopted across the studies?". (Line 35 to Line 38 on Page 7)

VERSION 2 – REVIEW

REVIEWER	Mateos-Alvarez, Raimundo University of Santiago de Compostela, Department of psychiatry
REVIEW RETURNED	09-Jul-2022
GENERAL COMMENTS	The minor changes suggested in the paper have been made by authors.
REVIEWER	Wilkinson, Tim The University of Edinburgh
REVIEW RETURNED	23-Jun-2022
GENERAL COMMENTS	Many thanks for addressing the points I raised - congratulations on an excellent protocol.