

## 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

<b>TITLE (PROVISIONAL)</b>	Linking electronic mental healthcare and benefits records in South London: design, procedure, and descriptive outcomes.
<b>AUTHORS</b>	Stevelink, Sharon; Phillips, Ava; Broadbent, Matthew; Boyd, Andy; Dorrington, Sarah; Jewell, Amelia; Leal, Ray; Bakolis, Ioannis; MADAN, IRA; Hotopf, Matthew; Fear, Nicola; Downs, Johnny

### VERSION 1 – REVIEW

<b>REVIEWER</b>	Brown, Heather Lancaster University, Division of Health Research
<b>REVIEW RETURNED</b>	24-Aug-2022

<b>GENERAL COMMENTS</b>	<p><b>Abstract:</b> As per my comments below and in the discussion, I would clarify that this is a first for data linkage in the UK but it is not the first of this type of data linkage as this is already done in Nordic countries.</p> <p><b>Introduction:</b> As I also mention in the discussion section, it may be worthwhile to put the research into an international context by mentioning some of the linkage of health and benefit data in Nordic countries and how it is used for service planning.</p> <p><b>Methods:</b> Why did you choose a fuzzy matching? As a robustness check did you try any other matching techniques?</p> <p>I also think it would be helpful to include potentially as a supplemental file, a brief description of the different types of benefits so the reader can get a better idea of the context for the results in Tables 4 and 5.</p> <p><b>Results:</b> Overall, I think the results section would benefit from a re-write. It is hard to follow the results and see how all the Tables fit together. It may be helpful to combine Tables 4 and 5 as there is not really much description of Table 4 in the text. On page 12, the description around age and linkage is confusing to follow. So very young adults &lt;21 were more likely to be linked than middle aged adults but older adults were more likely to be linked than both middle aged and younger adults (which I can see is the case from the table, but is not really described in a clear way in the text)</p> <p><b>Discussion:</b> In the discussion it would be worthwhile discussing how the findings relate to some of the Nordic literature from countries such as Finland</p>
-------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

	where such linkage is possible because of a universal identifier and is used for policy making. See for example: Karolaakso T, Autio R, Näppilä T, Leppänen H, Rissanen P, Tuomisto MT, Karvonen S, Pirkola S. Contextual and mental health service factors in mental disorder-based disability pensioning in Finland—a regional comparison. BMC health services research. 2021 Dec;21(1):1-3.
--	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

<b>REVIEWER</b>	Tseliou, Foteini Cardiff University
<b>REVIEW RETURNED</b>	03-Oct-2022

<b>GENERAL COMMENTS</b>	<p>This is a well-conducted study that undertook a novel data-linkage of routine records from UK health and welfare government service providers within a secure research environment. However, I have some comments on the methods implemented and how these could affect the interpretation of the observed results.</p> <ul style="list-style-type: none"> <li>• Were there any inconsistencies or issues when trying to collate information in the format of free clinical text notes with those within structured fields? Knowing that the level of information provided in records can vary significantly, this could be an issue.</li> <li>• Is women's low linkage rate linked to changing name due to marriage? Could that be accounted for via marriage status?</li> <li>• It is mentioned that a higher number of historical postcodes was associated with more successful linkage. Could this be linked to the fact that only up to five addresses could be included? Could this indicate that individuals who have moved more than 5 times during the study period might be missed out?</li> <li>• In terms of the lack of comparison group, could this be resolved with the use of healthcare records on physical conditions?</li> <li>• If there were multiple diagnoses across years was one of them chosen as the primary diagnosis? How was that determined? For example, if one individual had a primary long-term physical condition e.g. heart condition and a secondary mental health condition e.g. depression, would they be included?</li> <li>• Were there any individuals that had data in both SLAM services and IAPT services? Was it possible to control for that so that there is no overlap/ duplication? This was not very clear.</li> <li>• Has there been any feedback by the Advisory group on the preliminary findings presented in this paper e.g. around the choice of a sample with more severe mental health symptomatology? If yes, it should be mentioned in the results or discussion section.</li> <li>• Why were individuals who had resided in Northern Ireland excluded? Was that decided on the basis on a different policy on benefits receipt or receipt of treatment?</li> <li>• It would be interesting to further explore how individuals who engaged with DWP but were not successful in claiming benefits differed from those who successfully claimed benefits especially in terms of their mental health.</li> <li>• Why are the "contact with SLAM" variables categorized into before 2010 and after 2010? Did a policy change take place on that year?</li> <li>• In Table 4, is there an overlap between different types of benefits received e.g. being in receipt of ICA and DLA? This should be acknowledged in text or perhaps create a composite variable for multiple benefits receipt as these individuals would be different from those who only receive one type of benefits.</li> </ul>
-------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

## VERSION 1 – AUTHOR RESPONSE

Reviewer 1	Response
<p>Abstract: As per my comments below and in the discussion, I would clarify that this is a first for data linkage in the UK but it is not the first of this type of data linkage as this is already done in Nordic countries.</p>	<p>We have clarified this throughout the manuscript as follows:</p> <p>This data linkage is the first of its kind <u>in the UK</u> to demonstrate the use of routinely collected mental health and benefits data (<i>Abstract, page 2</i>).</p> <p>There are no pre-existing datasets <u>in the UK</u> that can currently address this (<i>Introduction, page 4</i>).</p> <p>Previous population-based research reporting on mental health and benefit receipt <u>in the UK has been limited in its use of self-report survey data, as well as a basic level of detail in relation to benefit receipt</u> (<i>Discussion, page 21</i>).</p> <p>Despite the limitations, this novel data linkage between routinely collected electronic mental healthcare records and benefits records contains extensive time-variant data that allows us to explore the bidirectional and complex relationships between mental health, employment and benefit receipt, something that has not previously been seen <u>in the UK</u> (<i>Discussion, page 22</i>).</p>
<p>Introduction: As I also mention in the discussion section, it may be worthwhile to put the research into an international context by mentioning some of the linkage of health and benefit data in Nordic countries and how it is used for service planning.</p>	<p>We have incorporated the following in the Introduction as suggested:</p> <p><u>However, research into welfare and mental health using data registries have been led by those in Nordic countries where a unique personal identifier is available to all those with a permanent residence record, paving the way for opportunities in linkages between health and welfare registers</u> (21). (<i>Introduction, page 4</i>).</p>
<p>Methods: Why did you choose a fuzzy matching? As a robustness check did you try any other matching techniques?</p>	<p>The matching process was led by the Department for Work and Pensions and their standard matching process only uses deterministic matching for administrative data. An ad hoc fuzzy matching deterministic technique</p>

	<p>was used to allow for agreement on partial identifiers.</p> <p>As the Department for Work and Pensions led on the matching, we were not able to implement any other matching techniques. However, the high uniqueness cut-off threshold applied for the fuzzy matching, and the resulting high linkage rate support the robustness of the deterministic matching process undertaken.</p>
<p>I also think it would be helpful to include potentially as a supplemental file, a brief description of the different types of benefits so the reader can get a better idea of the context for the results in Tables 4 and 5.</p>	<p>We have referenced a <a href="#">link</a> to the UK Government website providing an overview of the benefits in the Methods section of the manuscript (<a href="https://www.gov.uk/browse/benefits">https://www.gov.uk/browse/benefits</a>)</p> <p>As the eligibility criteria for benefits are subject to change, as well as how much people are entitled to, we believe this information will remain most accurate to readers by including this link, instead of providing a supplementary file. A most recent example is the increase in the number of hours people need to work on a weekly basis to retain their full entitlement to Universal Credit.</p>
<p>In the discussion it would be worthwhile discussing how the findings relate to some of the Nordic literature from countries such as Finland where such linkage is possible because of a universal identifier and is used for policy making. See for example: Karolaakso T, Autio R, Näppilä T, Leppänen H, Rissanen P, Tuomisto MT, Karvonen S, Pirkola S. Contextual and mental health service factors in mental disorder-based disability pensioning in Finland—a regional comparison. BMC health services research. 2021 Dec;21(1):1-3.</p>	<p>We have incorporated the following in the Discussion as suggested:</p> <p><u>Though we are yet to explore sickness and disability related benefits among our sample in detail, much research into disability pension (DP) awards has already been conducted in Norway using large population-based cohorts containing mental and physical health data linked to national databases of disability benefits using national identification numbers. For example, one study investigated the impact of anxiety and depression on DPs awarded for mental health and physical health diagnoses. They showed long-term occupational impact of anxiety and depression symptoms and their subsequent independent contribution towards DPs</u></p>

	<p>awarded (23). Another study linking mental health cohort data and the National Insurance Administration database containing DP award data showed that anxiety and depression at baseline were strongly associated with receiving a DP award at follow-up (22). A Finnish study found that there was evidence of regional variation in mental disorder DP and mental health service factors, a critical finding when considering service provision (21). (Discussion, page 22).</p>
<p>Overall, I think the results section would benefit from a re-write. It is hard to follow the results and see how all the Tables fit together. It may be helpful to combine Tables 4 and 5 as there is not really much description of Table 4 in the text.</p>	<p>We have moved Table 4 to the supplementary materials to improve the flow of the results. We have also carefully re-read the Results section and made several amendments to improve readability.</p>
<p>On page 12, the description around age and linkage is confusing to follow. So very young adults &lt;21 were more likely to be linked than middle aged adults but older adults were more likely to be linked than both middle aged and younger adults (which I can see is the case from the table, but is not really described in a clear way in the text),</p>	<p>Thank you for highlighting this. We now write:</p> <p>Results from adjusted logistic regression analyses indicated that the following groups of patients were less likely to be linked (an OR greater than 1 denotes greater chance of successful linkage compared with the reference): female patients vs. male patients, ethnic minority groups vs. patients from a white ethnic background, and middle-aged patients vs. younger patients (&lt;21 years) and patients with only one postcode available vs. two or more postcodes. Compared to younger patients (&lt;21 years), middle-aged patients (21-60 years) were less likely to be successfully linked, whereas older patients (&gt;60 years) were more likely to be linked compared to all other age groups (Table 1). Further, the linkage rate was also higher among patients who had a higher number of historical postcodes available. On the other hand, older patients (&gt;60 years) were more likely to be linked than younger patients. (Results, page 10).</p>
<p><b>Reviewer 2</b></p>	
<p>This is a well-conducted study that undertook a novel data-linkage of routine records from UK health and welfare government service providers within a secure research environment. However, I have some comments on the</p>	<p>Thank you.</p>

methods implemented and how these could affect the interpretation of the observed results.	
Were there any inconsistencies or issues when trying to collate information in the format of free clinical text notes with those within structured fields? Knowing that the level of information provided in records can vary significantly, this could be an issue.	<p>For the current data linkage paper, no data was used from the clinical text notes. Only data from structured fields as part of the patient's electronic mental healthcare record was used. We agree that this is something we need to explore carefully in future papers.</p> <p>We now write:  <u>For the current paper, only data from structured fields were used.</u> (Methods, page 5).</p>
Is women's low linkage rate linked to changing name due to marriage? Could that be accounted for via marriage status?	As outlined in the Discussion of the manuscript, the lower linkage rate is indeed likely to be due to changes in relationship status (see page 21). Unfortunately, marital status is not consistently recorded in the electronic mental healthcare records (and absent in benefits records).
It is mentioned that a higher number of historical postcodes was associated with more successful linkage. Could this be linked to the fact that only up to five addresses could be included? Could this indicate that individuals who have moved more than 5 times during the study period might be missed out?	<p>If patients had more than five historical postcodes, they were still included in the sample that was sent for linkage to the Department for Work and Pensions. In this case, only the five postcodes closest to when a referral to SLaM was accepted or when an event was recorded were included. This is now detailed under Table 1:</p> <p>Number of home/residential postcodes available-:</p> <p><u>-based on five closest postcodes to the earliest accepted referral into SLaM or event within the time window (Table 1, page 12).</u></p>
In terms of the lack of comparison group, could this be resolved with the use of healthcare records on physical conditions?	Welfare records are only linked with the electronic mental healthcare records, and this was only achieved after approximately 5 years of negotiations with the Department for Work and Pensions. We are keen to extend the linkage of welfare records with other data sources, however this is not a current priority. We want to capitalise first on this novel data linkage.
If there were multiple diagnoses across years was one of them chosen as the primary diagnosis? How was that determined? For example, if one individual had a primary long-term physical condition e.g. heart condition and a secondary mental health	As outlined in the Materials section of the Methods, the psychiatric primary diagnosis recorded in a patient's

<p>condition e.g. depression, would they be included?</p>	<p>record <u>closest</u> and <u>before</u> the SLaM window end date was taken. We have included this description under the Tables where appropriate. If patients received another diagnosis at the same date, this would have been recorded by the clinician as a psychiatric secondary diagnosis.</p> <p>It is unlikely that patients were diagnosed with a long-term physical condition as SLaM is a mental health service provider. Even if that was the case, the secondary mental health condition would still be included as we specified our interest in psychiatric diagnoses during the clinical data extraction phase.</p>
<p>Were there any individuals that had data in both SLaM services and IAPT services? Was it possible to control for that so that there is no overlap/ duplication? This was not very clear.</p>	<p>Some patients accessed both SLaM specialist services and IAPT. However, data are recorded in two different electronic mental health systems, namely ePJS (SLaM) and IAPTus (IAPT). Therefore, there was no duplication or overlap in the data used for this data linkage paper as only data extracted from ePJS was used.</p>
<p>Has there been any feedback by the Advisory group on the preliminary findings presented in this paper e.g. around the choice of a sample with more severe mental health symptomatology? If yes, it should be mentioned in the results or discussion section.</p>	<p>The Advisory Group has been informed of the findings presented in this paper via a newsletter. No feedback was received.</p>
<p>Why were individuals who had resided in Northern Ireland excluded? Was that decided on the basis on a different policy on benefits receipt or receipt of treatment?</p>	<p>The Health Research Authority approval under Regulation 5 of the Health Service (Control of Patient Information) Regulations 2002 to process patient identifiable information without consent only applies to England and Wales. In addition, the devolved legislature of Northern Ireland is responsible for administering benefits to patients who resided in Northern Ireland at the time of their benefit receipt. Therefore, the DWP is not allowed to share this data.</p> <p>We did not have the approvals to maintain this small group of patients (n=795 out of 221,243 unique records) in our linked dataset.</p> <p>We now added the as a footnote (<i>Results, page 9</i>).</p>
<p>It would be interesting to further explore how individuals who</p>	<p>This is an interesting avenue</p>

<p>engaged with DWP but were not successful in claiming benefits differed from those who successfully claimed benefits especially in terms of their mental health.</p>	<p>for further research and something we hope to explore in the future.</p>
<p>Why are the “contact with SLaM” variables categorized into before 2010 and after 2010? Did a policy change take place on that year?</p>	<p>We have now changed this to 2013 as in that year Personal Independence Payment was introduced, thereby replacing Disability Living Allowance.</p> <p>We now write: <u>This cut off was chosen as Personal Independence Payment was introduced in 2013 (Methods, page 8).</u></p> <p>We have also updated this in Table 2 accordingly (Table 2, page 13).</p>
<p>In Table 4, is there an overlap between different types of benefits received e.g. being in receipt of ICA and DLA? This should be acknowledged in text or perhaps create a composite variable for multiple benefits receipt as these individuals would be different from those who only receive one type of benefits.</p>	<p>We have moved Table 4 to the supplementary material based on feedback from Reviewer 1. Some patients received different benefits over the 15-year time period. We now write: <u>Of those receiving benefits, 85.1% received 2 or more different benefits. (Results, page 17).</u></p> <p>Exploration of multiple benefit receipt is out of scope for the current paper.</p>
<p><b>Editor comments</b></p>	
<p>Please revise the ‘Strengths and limitations of this study’ section of your manuscript (after the abstract). This section should contain up to five short bullet points, no longer than one sentence each, that relate specifically to the methods. The novelty, aims, results or expected impact of the study should not be summarised here.</p>	<p>We have shortened the bullet points as requested. We now write:</p> <ul style="list-style-type: none"> <li>• <del>This is a novel data linkage between electronic mental healthcare records and benefits records in the UK, providing the opportunity to answer important questions relating to mental health, work and benefit receipt.</del></li> <li>• <u>A strength of this data linkage is the a high linkage rate of 92.3%. was achieved.</u></li> <li>• The sample does not include a comparison group (e.g., people who did not access</li> </ul>



	<p>secondary mental healthcare services).</p> <ul style="list-style-type: none"> <li>• Although there are indicators of people being in and out of work depending on what type of benefits are being received (unemployment related benefits). There is no reliable employment variable within the data stating whether someone is currently in or out of work (except for Universal Credit).</li> <li>• There is a potential for linkage bias as a result of the method used (ad hoc deterministic fuzzy matching) and having no unique identifier between data sets (<i>Strengths and limitations section, page 3</i>).</li> </ul> <p>We have carefully proofread the revised manuscript and made a few amendments accordingly. These are highlighted in Track Changes for clarity.</p>
--	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

**VERSION 2 – REVIEW**

<b>REVIEWER</b>	Brown, Heather Lancaster University, Division of Health Research
<b>REVIEW RETURNED</b>	01-Dec-2022
<b>GENERAL COMMENTS</b>	Thank you for addressing my comments.