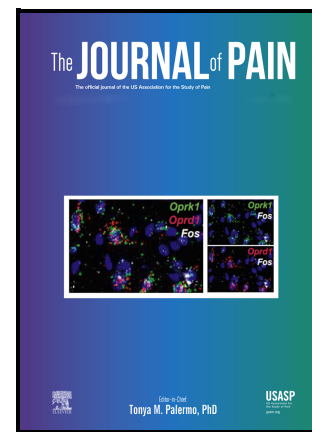


# Journal Pre-proof

How well can we measure chronic pain impact in existing longitudinal cohort studies? Lessons learned  
Running title: Measuring Chronic Pain Impact in Cohort Studies

Diego Vitali, Charlotte S C Woolley, Amanda Ly, Matthew Nunes, Laura Oporto Lisboa, Edmund Keogh, John McBeth, Beate Ehrhardt, Amanda C de C Williams, Christopher Eccleston



PII: S1526-5900(24)00641-2

DOI: <https://doi.org/10.1016/j.jpain.2024.104679>

Reference: YJPAI104679

To appear in: *The Journal of Pain*

Received date: 30 May 2024

Revised date: 20 August 2024

Accepted date: 10 September 2024

Please cite this article as: Diego Vitali, Charlotte S C Woolley, Amanda Ly, Matthew Nunes, Laura Oporto Lisboa, Edmund Keogh, John McBeth, Beate Ehrhardt, Amanda C de C Williams and Christopher Eccleston, How well can we measure chronic pain impact in existing longitudinal cohort studies? Lessons learned  
Running title: Measuring Chronic Pain Impact in Cohort Studies, *The Journal of Pain*, (2024) doi:<https://doi.org/10.1016/j.jpain.2024.104679>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

**Title: How well can we measure chronic pain impact in existing longitudinal cohort studies? Lessons learned**

**Running title:** Measuring Chronic Pain Impact in Cohort Studies

**Authors:** Diego Vitali [a], Charlotte S C Woolley [b] Amanda Ly [c], Matthew Nunes [d], Laura Oporto Lisboa [e], Edmund Keogh [x], John McBeth [b], Beate Ehrhardt [e,x], Amanda C de C Williams [a], Christopher Eccleston [x,y,z],

**Affiliations:** [a] Research Dept of Clinical, Educational & Health Psychology, University College London, London UK. [b] Centre for Epidemiology Versus Arthritis, Division of Musculoskeletal and Dermatological Science, Faculty of Biology, Medicine and Health, School of Biological Sciences, The University of Manchester, Manchester, UK. [c] Centre for Academic Mental Health, Bristol Medical School, University of Bristol, Bristol, UK. [d] School of Mathematical Sciences, University of Bath, UK [e] Institute for Mathematical Innovation, University of Bath, UK. [x] Centre for Pain Research, University of Bath, UK, [y] Department of Experimental-Clinical and Health Psychology, Ghent University, Belgium, [z] Department of Psychology, The University of Helsinki, Finland.

**Corresponding Author:** Diego Vitali, PhD. **Present Address:** [d.vitali@ucl.ac.uk](mailto:d.vitali@ucl.ac.uk); Clinical, Educational & Health Psychology University College London - Faculty of Brain Sciences University College London, Gower St, London WC1E 6BT

Color print: No color is required for print.

**Abstract:** Multiple large longitudinal cohorts provide opportunities to address questions about predictors of pain and pain trajectories, even when not anticipated in design of the historical databases. This focus article uses two empirical examples to illustrate the processes of assessing the measurement properties of data from large cohort studies to answer questions about pain. In both examples, data were screened to select candidate variables that captured the impact of chronic pain on self-care activities, productivity and social activities. We describe a series of steps to select candidate items and evaluate their psychometric characteristics in relation to the measurement of pain impact proposed. In UK Biobank, a general lack of internal consistency of variables selected prevented the identification of a satisfactory measurement model, with lessons for the measurement of chronic pain impact. In the English Longitudinal Study of Ageing, a measurement model for chronic pain impact was identified, albeit limited to capturing the impact of pain on self-care and productivity but lacking coverage related to social participation. In conjunction with its supplementary material, this focus article aims to encourage exploration of these valuable prospectively collected data; to support researchers to make explicit the relationships between items in the databases and constructs of interest in pain research; and to use empirical methods to estimate the possible biases in these variables.

**Perspective:** This focus article outlines a theory-driven approach for fitting new measurement models to data from large cohort studies, and evaluating their psychometric properties. This aims to help researchers develop an empirical understanding of the gains and limitations connected with the process of re-purposing the data stored in these datasets.

**Keywords:** "measurement model"; "epidemiology"; "Item Response Theory", "UK Biobank", "UK ELSA"

## 1. Introduction

Large, existing biomedical and health datasets are available for research purposes in many countries <sup>1,2</sup>. These secondary data sources include data collected primarily for research, although also for administrative reasons and clinical documentation, e.g., electronic health records and registry data sources <sup>3</sup>. This paper focusses on the utility of high quality historical studies that pre-date definitions of health related constructs that we may wish to apply in current or future research. These historical datasets are generally population-based, longitudinal cohorts that are largely representative of the source population and contain a wide range of variables collected prospectively across years or decades (e.g. English Longitudinal Study of Ageing <sup>4,5</sup>. The UK in particular has a wealth of data from longitudinal cohort studies that can be made available to researchers; many are discoverable via platforms such as the UK Longitudinal Linkage Collaboration <sup>6</sup>. It is also important to note that many of the processes and lessons learned detailed in this paper are applicable to cross-sectional studies. How can pain researchers best make use of these valuable resources to address new research questions that may differ from those of the scientists who initiated these historical studies, since in the absence of data that directly address variables of interest, researchers must look for alternative variables that best approximate the information required.

The gap between what a researcher wishes to capture and what existing data can provide to build this construct can constitute both a conceptual and an empirical limitation. A good example from our own work is pain impact, where these issues became evident as we tried to put into practice a framework using the dichotomies of duration (acute vs chronic) and impact (high vs low) to define states of chronic pain and transitions between these states over time <sup>7</sup>. What was lacking was any clear guidance on how best to address this issue, so in this paper we use the example of chronic pain impact to describe one pragmatic approach to assessing and documenting the subtle but sometimes critical empirical limitations of using existing data.

Chronic pain is often defined by time, where pain persists for three months or longer <sup>8-10</sup>. A comprehensive synthesis of epidemiological studies estimated the prevalence of chronic pain in the adult population of high income countries at 31% (CI: 8.7%-64.4%) <sup>11</sup>. In assessing chronic pain, there is increasing recognition that single indices of pain (often average intensity) are insufficient to capture important variations in the effects of chronic pain on individuals' lives <sup>12-14</sup>. Qualitative enquiry is rich in description of the ways in which chronic pain affects an individual's life <sup>15,16</sup> and even health economics calculates cost-utilities based on the effects of pain, not its intensity <sup>17</sup>. Further, the US national strategy for pain <sup>18,19</sup> has highlighted the need to identify 'high impact' chronic pain, which it defined by three parallel criteria: pain severity (intensity of the pain experience), activity limitations (difficulty carrying out subjectively relevant activities), and participation restrictions (difficulties in engaging in subjectively relevant life situations or social roles). We accepted this definition of high impact chronic pain

as we felt it provides a general framework to help policy makers and service providers to identify those people with chronic pain most in need of support. It also enables researchers to study the consequences of chronic pain at the population level - e.g.<sup>20,21</sup>.

As we started exploring how this framework could be applied to data captured in large longitudinal studies, we were immediately faced with a practical obstacle: chronic pain impact was not readily available as a variable or as a set of variables. In fact, most existing datasets pre-date the National Institutes of Health initiative<sup>2</sup>, and focus on pain intensity to discriminate the severity of the pain experience, without a direct question about duration. In the best-case scenario, a dataset includes one or more established pain questionnaires that can be used to assess pain intensity and pain impact. However, when such questionnaires are absent, trying to assess a construct such as pain impact can be challenging<sup>13</sup>. In this paper we describe ways of approximating this missing information, and present a workflow for detecting and measuring High Impact Chronic Pain (HICP) in existing longitudinal datasets. We illustrate this using two examples: the English Longitudinal Study of Ageing<sup>5</sup> ([www.elsa-project.ac.uk/](http://www.elsa-project.ac.uk/)) and the UK Biobank (UKB)<sup>22</sup> ([www.ukbiobank.ac.uk](http://www.ukbiobank.ac.uk)).

## **2. Workflow for detecting HICP in longitudinal datasets**

Large datasets can contain thousands of variables with several thousand respondents in each cohort, with data spanning decades. The vast majority of the variables in these datasets contain data generated by the respondent answering a specific item, using a number on a scale or a category. Here we refer to each question asked directly of the respondent that produces a specific datapoint as an “item”. Since new questions

cannot be added to retrospectively collected data, if an existing dataset does not include a standardised questionnaire that samples pain impact, the information that we can extract on this topic is likely to present unknown limitations that need assessing and documenting. Here we suggest it is important to: (i) evaluate if and how items can be used to capture the domains of pain impact as the construct of interest (content validity); (ii) assess the limitations and properties of candidate items aggregated while developing this construct (construct validity); (iii) be open and transparent about the conceptual and empirical implications of such limitations. Figure 1 presents an example of a work flow and outlines the decision-making tree we used to evaluate the assessment of pain impact in ELSA and UKB.

*[Figure 1 about here]*

The first step in Figure 1 focuses on arriving at a definition of pain impact (domain definition), and then declaring the information that is critical to capture this construct using a set of clear “detection criteria” (for which see Table 1). After accepting the definition of high-impact chronic pain suggested by the NIH, we decided that a measure of pain impact must capture the experience of limitations imposed by pain on an individual’s ability to self-care, or to engage in work activities, or to participate in social life. Using these limitations in self-care, work, and social life as criteria, Step 2 is a search of the dataset (item selection) for items that best approximated the detection criteria as defined. This was a tentative process because the selection of items is constrained by availability, and by how well items pertain (or not) to the criteria set in the previous step. Further, it is inevitable that many of the variables selected were not originally designed to be used to indicate pain impact. For example, we might

select and mix items found in a series of questions related to physical health with other items found in a separate, psychosocial, section of the dataset. Consequently, many of the items resulting from our initial search may collect responses influenced by other causes than pain. This immediately poses challenges in terms of the relationship of these responses to the construct of interest. With this in mind, we selected items worded in a way that specified a link between pain and the domain of limitation: e.g. "Pain limits the kind or amount of paid work I could do, should I want to". Where this link was not explicit, or 'anchored in pain', attribution of cause between pain impact and the proposed indicators becomes ambiguous and may dilute the measure's ability to reflect the aspects of pain impact intended. However, when such pain-specific data was not available, we relaxed this criterion to include items anchored in health or disability. Although guided by pragmatism, we recognise that this risks being over-inclusive, and sensitivity analysis may be useful to test the extent to which the decision affects results. Using consensus of expert opinion (from both research and lived experience of pain), we initially assessed the selected items against the detection criteria: did selected items capture all the characteristics of the domain? Did they do so in a balanced way? We then documented which aspects of impact were covered by the proposed set of items, whether there were aspects missing, and what were the conceptual implications of these gaps in terms of the initial domain definition. When any such conceptual deficit was noted, we acknowledged that it was not possible to satisfy the detection criteria.

In the context of this type of dataset, this is not uncommon, and can preclude further analysis. However, it is also possible that relaxing one or more of the initial detection criteria could allow the selection of items to become suitable for analysis. When this is the case, discussion and documentation of changes and their implications for domain



definition generate a transparent redefinition of what the selected items capture about impact.

Step 3 focuses on evaluating construct validity through item analysis. This step begins by defining a measurement model: a formal way to describe the quantitative relationships observed between the selected items (the “observed” variables), and the construct that we want to measure (the “latent” variable). Item analysis uses statistical methods from classical test theory, such as exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). Factor analysis (FA) was chosen over principal components analysis (PCA) as FA assumes that total variance can be partitioned into unique and common variance, whereas PCA assumes all variance is attributed to common variance and aims to identify variables that are composite of the observed data.

As seen in recent contributions to the literature, the use of an item response theory (IRT) model allows for a metric to be compared across multiple waves even when different set of items are scored at each wave <sup>23</sup>. These analyses are important to clarify and critically examine the information conveyed by each of the items in the selection, thereby guiding the creation of a more balanced metric <sup>24</sup>. In summary, this step aims to evaluate empirically how the scoring of items selected in step 2 can be used to define a unidimensional measurement model of pain impact. This process also evaluates the reliability of item responses as indicators of pain impact.

In step 4, findings from step 3 are discussed in relation to the domain definition, reflecting on the alignment between the latent construct (e.g. pain impact) and what was actually measured using the items in the dataset. This step documents any

challenges or limitations in the measurement model, describing the adequacy of selected items as indicators of the latent construct.

### **3. Modelling impact of chronic pain in two UK longitudinal datasets**

We present two working examples, from ELSA and UKB datasets, where we used this workflow to identify items to represent high impact chronic pain in order to address questions about the development, maintenance, and recovery from chronic pain.

*Step 1.* In both our examples, the initial definition of impact adopted was the one proposed by the NIH (2016), as described above. We selected items that referred to restriction of ‘self-care activities’, of ‘participation in work activities’, and of ‘participation in social activities’. For both datasets, we started with the same domain definition of chronic pain impact and explored each dataset using the following baseline detection criteria (see Table 1).

[ table 1 about here]

#### **3.1. ELSA**

*Background:* ELSA, the English Longitudinal Household Survey, is designed to examine the health, economic status, and overall well-being of participants aged 50+. It began in 2002 with 12,099 participants, collecting self-report data every 2 years. The sample considered here is the initial ELSA sample (Cohort 1) which was drawn from households that had previously responded to the Health Survey for England (HSE) <sup>25</sup> between 1998 and 2001. The HSE consists of a nationally representative sample of individuals living in private households <sup>26</sup>.

*Step 2:* After an initial broad selection of ELSA items against the detection criteria (Table 1), we used the expertise of researchers and of people with chronic pain to evaluate the items against the detection criteria. Most candidate items for the categories of self-care, social, and work restrictions failed the strict criterion of anchoring to pain. In order to proceed, we relaxed this criterion and included items concerned with social, work and self-care restriction that the respondent attributed to physical or health limitations (see supplementary materials for items). Similarly, since half the respondents to ELSA were retired, we extended the definition of ‘work-related limitations’ to include unpaid commitments as represented in the dataset, such as productivity from volunteering or housework. The selection could now include items where participants were asked (for instance) whether they had “any difficulty shopping for groceries”, or “getting in and out of bed”, or whether “you have any health problem or disability that limits the kind or amount of paid work you would do, should you want to” (see supplementary material for a full list of items). Relaxing these criteria allowed extension of item representation to cover more, but not all, of the three types of restrictions in the definition of pain impact (Table 2). This discrepancy illustrates the conceptual limitations of the selected items, such that any model of pain impact associated with these items will capture self-care and work limitations, but not social activity limitations.

*[ Table 2 about here ]*

*Step 3:* We initially scrutinised the items using EFA and CFA, to assess the fit of the hypothesized measurement model, including data from all participants enrolled in ELSA in 2002 (supplementary material ELSA\_a). We then used an IRT model to a)

analyse item characteristics, and b) estimate the probability of each observed response pattern as a function of the latent trait and of the item characteristics <sup>27</sup>. Briefly, the IRT model was used to study the response patterns of participants to each item in order to estimate a parameter  $\theta$ : the unobserved level of impact.

Just over half the respondents reported no restrictions on the selected items. However, the distribution of latent trait scores differed between subgroups reporting 'chronic pain', 'no pain' or 'acute pain'. In Figure 2 we can see the distribution of  $\theta$  scores (pain impact scores) between respondents with no report of pain, respondents with acute pain, and respondents with chronic pain. To facilitate interpretation, the model estimates of  $\theta$  were centred around the mean and scaled so that one unit of  $\theta$  represents one standard deviation. The sample used for this model was based on the respondents to ELSA wave 2 (n = 9032) after the exclusion of respondents younger than 50 (n = 261), and those with responses given by proxy (n = 102). As in Table 3, the vast majority of respondents (97.8%) included in this analysis were white. This is consistent with the proportion reported by ELSA technical report in 2004 <sup>5</sup> and represents a slightly greater proportion than that reported by 2004 census which reported as "white" 96% of English residents over 50 years old <sup>28</sup>. Deprivation indices, and detailed information about ethnicity are sensitive data and they are archived separately to the data considered in this paper.

*[ Table 3 about here]*

*[Figure 2 about here]*

The lowest value of  $\theta$  corresponds to the particular scoring pattern in which respondents reported no difficulties or limitations described by the items included in the measurement model. In the subgroup of respondents with no pain,  $\theta$  scores were right-skewed and presenting evident zero-inflation with the majority of respondents presenting the lowest estimates of  $\theta$ . The acute pain subgroup had a mean  $\theta$  score above zero ( $\theta = 0.09$ , bootstrap 95%CI 0.05-0.13); the chronic pain subgroup presented a more symmetric distribution with mean  $\theta = 0.7$  (bootstrap 95%CI 0.66-0.72). In comparison, the respondents in the chronic pain subgroup were far more likely to present  $\theta$  scores above one. In relation to reported levels of pain severity, 80% of the 7652 respondents with  $\theta$  scores within 1 standard deviation of the sample mean reported no pain or mild pain. For those with  $\theta$  scores at one or more standard deviations above the mean ( $n = 1387$ ), 69.6% reported moderate or severe levels of pain. Details to replicate this analysis are reported in supplementary material (ELSA\_b).

### 3.2. UK Biobank

*Background:* UK Biobank is a large-scale longitudinal study established in 2006 that collects genetic, biological, environmental, lifestyle and health information from over 500,000 participants ([www.ukbiobank.ac.uk](http://www.ukbiobank.ac.uk)).

*Step 1* is as outlined above.

*Step 2:* We evaluated UKB items collected between 2006 and 2010 to map them to the detection criteria for pain impact (see Table 1). 20 items derived from 12 UKB

fields were suitable for the categories of self-care, social, and work restrictions, but failed the criterion of anchoring to pain. For example, items included UKB field 1031, “How often do you visit friends or family or have them visit you?” and UKB field 6142 about employment status, “Which of the following describes your current situation?”, with the answer “Unable to work because of sickness or disability” used as the outcome of interest. As with ELSA, when we relaxed our criterion to include any self-reports of social, work and self-care restrictions, we obtained improved item representation across the three categories of restriction (see Table 3), but items were not anchored in pain, nor did they identify a restriction that lasted more than 3 months (see Table 2 and supplementary materials “Suppl.UKB”).

*[ Table 4 about here ]*

*Step 3:* This study used baseline assessment data from 502,154 participants, which was downloaded on the 3<sup>rd</sup> of July 2024. Of these participants, 218,507 reported chronic pain, defined as pain (at specific locations or all over the body) that interfered with usual activities for over three months. Of participants with chronic pain, 20,067 (9.18%) had missing data in one of the 20 items in step 2. Therefore, 198,440 participants formed the cohort to be analysed: the summary demographics of these participants are shown in Table 5. The mean Townsend deprivation index was  $-1.06$  ( $SD \pm 3.21$ ), indicating lower deprivation levels than the national average of 0. For full demographic tables of the UKB cohort stratified by pain status, please refer to the supplementary materials “Suppl.UKB”.

*[ Table 5 about here ]*

The dataset was split into two randomly selected sub-sets for EFA and CFA respectively. Out of 190 possible pairs of items, 188 had weak ( $<0.4$ ) intercorrelations and only two had moderate ( $>0.4$   $<0.6$ ) intercorrelations. Parallel analysis of the number of factors suggested that the underlying data structure contained seven factors. A series of exploratory analyses were repeated, iterating different item selections and converting some item scoring (e.g. ordinal into binary format). After examining several theoretically plausible sub-selections of the items, EFA and CFA continued to demonstrate poor model fit with most items presenting large proportions of unexplained unique variance (the sum of both error variance and variance specific to each item). Therefore, in this case it was not possible to assume that the scoring of considered items represents the effect of a common underlying factor.

#### **4. Discussion**

We have tried to explore the conceptual and methodological concerns that arise from asking research questions using longitudinal datasets that were not designed to address those questions. Various large-scale datasets are already available and others underway, with a scope much wider than many clinical studies, and with the power to address questions to the large sets or to subsets of the full sample. Here we used the example of chronic pain impact, the consequences of pain on individuals' lives, to illustrate the approach with two longitudinal cohorts, ELSA and UKB. The former provided items to model impact of pain on self-care and work activities, but without adequate coverage of the impact of pain on social activities. Given the growing

interest in social aspects of pain, and rich data on a variety of social activities unanchored to pain or health in ELSA, this is a significant shortcoming. Using UKB, we provided an example of when the latent construct cannot be adequately modelled using available items, and the exercise must be abandoned rather than proceeding with untenable assumptions. Although the UKB items themselves gave good coverage of the three areas of potential pain impact according to the NIH definition, restrictions in work, self-care and social life were not anchored in pain, or even in physical illness, or disability. This raised possibilities of unobserved factors, including various sociodemographic factors, exerting substantial influence on activity restrictions, undermining attempts to answer pain-related questions. However, it was not our aim to produce several measures of pain impact that differed depending on sociodemographic status, but rather a single unidimensional measure that could be applied to the entire population. In addition, all but one item in ELSA referred to restrictions lasting (or expected to last) for a period of 3 months, whereas in UKB the persistence of restrictions varied across items, thereby including some transitory and brief restrictions that would undermine attempts to model chronic pain impact.

## **5. Conclusion**

There are substantial advantages to being able to use existing large longitudinal datasets to retrofit recently developed constructs. Such datasets represent substantial investment of time and money and were intended for repeated use by many researchers. However, simply selecting items that seem to fit research questions can seriously mislead. Rather, establishing the soundness (or not) of proposed measurement models adds value to the datasets for future researchers who aim to explore cognate areas. It is also an opportunity to draw on the expertise of people with



chronic pain in defining and interpreting latent constructs, as illustrated here, in a broader collaboration of clinicians and specialist academics, promoting constructive discussions about the empirical measurement of pain-relevant constructs. This multidisciplinary cooperation can then be used to produce better and more interpretable measurement models based on consensus, transparency, and empirical evidence.

## **6. Acknowledgements**

The authors gratefully acknowledge the public contributors who shared their time, knowledge, and experiences enabling mutual learning and collaboration.

## **7. Data statement**

This study makes use of data from the UK Biobank (project ID 98481), which was approved by the National Information Governance Board for Health and Social Care and the National Health Service North West Multicentre Research Ethics Committee (Ref: 06/MRE08/65). All participants gave informed consent, and the study was approved by the Research Ethics Committee (No: 11/NW/0382). The English Longitudinal Study of Ageing was developed by a team of researchers based at University College London, NatCen Social Research, the Institute for Fiscal Studies, the University of Manchester and the University of East Anglia. The data were collected by NatCen Social Research. The funding is currently provided by the National Institute on Aging in the US, and a consortium of UK government departments coordinated by the National Institute for Health Research. Funding has also been received by the Economic and Social Research Council. ELSA has received ethical approval from different institutional review boards for its waves (<https://www.elsa-project.ac.uk/ethical-approval>). Informed consent was sought from all the ELSA

participants. Both UKB and ELSA datasets have established data sharing processes. ELSA anonymised datasets with corresponding documentation can be downloaded for use by researchers from the UK Data Service. UKB is globally accessible to approved researchers who are undertaking health-related research that's in the public interest. We have detailed the processes for each dataset in the supplementary file "Suppl.UKB.pdf", "Suppl.ELSA\_a.pdf", and "Suppl.ELSA\_b.pdf", and in a public github repository (<https://github.com/UCL/Pain.Impact.Measures>)

## 8. References

1. Palmer LJ. UK Biobank: bank on it. *The Lancet*. 2007;369(9578):1980-1982.
2. NIH NIOH. National Pain Strategy - a comprehensive population health-level strategy for pain. Accessed 2024/07/20, [https://www.iprcc.nih.gov/sites/default/files/documents/NationalPainStrategy\\_508C.pdf](https://www.iprcc.nih.gov/sites/default/files/documents/NationalPainStrategy_508C.pdf)
3. Ahlbom A. Modern Epidemiology, TL Lash, TJ VanderWeele, S Haneuse, KJ Rothman. Wolters Kluwer, 2021. *European Journal of Epidemiology*. 2021;36(8):767-768. doi:10.1136/bmj-2023-076036
4. Banks J, Marmot M, Blundell R, Lessof C, Nazroo J. *Health, wealth and lifestyles of the older population in England: the 2002 English Longitudinal Study of Ageing*. Institute for Fiscal Studies; 2003.
5. Banks J, Batty G, Breedvelt J, et al. Data from: English Longitudinal Study of Ageing: Waves 0-10, 1998-2023. 2024. Deposited February 2024. Accessed 2024/07/20, doi:10.5255/UKDA-SN-5050-27
6. Boyd A, Flaig R, Oakley J, et al. The UK Longitudinal Linkage Collaboration: a trusted research environment for the longitudinal research community. *International Journal of Population Data Science*. 2022;7(3):2046. doi:10.23889/ijpds.v7i3.2046
7. Eccleston C, Begley E, Birkinshaw H, et al. The establishment, maintenance, and adaptation of high-and low-impact chronic pain: a framework for biopsychosocial pain research. *Pain*. 2023;164(10):2143-2147. doi:10.1097/j.pain.0000000000002951
8. Chronic pain (primary and secondary) in over 16s: assessment of all chronic pain and management of chronic primary pain. London: National Institute for Health and Care Excellence (NICE); April 7, 2021. Accessed 2024/07/20, <https://www.nice.org.uk/guidance/ng193/>
9. Kang Y, Trewern L, Jackman J, McCartney D, Soni A. Chronic pain: definitions and diagnosis. *BMJ*. 2023;381:e076036. doi:10.1136/bmj-2023-076036
10. Nicholas M, Vlaeyen JW, Rief W, et al. The IASP classification of chronic pain for ICD-11: chronic primary pain. *Pain*. 2019;160(1):28-37.

11. Steingrimsdóttir ÓA, Landmark T, Macfarlane GJ, Nielsen CS. Defining chronic pain in epidemiological studies: a systematic review and meta-analysis. *Pain*. 2017;158(11):2092-2107.
12. Von Korff M, Ormel J, Keefe FJ, Dworkin SF. Grading the severity of chronic pain. *Pain*. 1992;50(2):133-149.
13. Turk DC, Fillingim RB, Ohrbach R, Patel KV. Assessment of Psychosocial and Functional Impact of Chronic Pain. *J Pain*. 2016;17(9 Suppl):T21-49. doi:10.1016/j.jpain.2016.02.006
14. Gershoni T, Pud D, Aviram J, Eisenberg E. Wellness of patients with chronic pain is not only about pain intensity. *Pain Practice*. 2023;23(2):145-154.
15. Crowe M, Whitehead L, Seaton P, et al. Qualitative meta-synthesis: The experience of chronic pain across conditions. *Journal of Advanced Nursing*. 2017;73(5):1004-1016.
16. Pagé MG, Lacasse A, Dassieu L, et al. A cross-sectional study of pain status and psychological distress among individuals living with chronic pain: the Chronic Pain & COVID-19 Pan-Canadian Study. *Health Promot Chronic Dis Prev Can*. 2021;41(5):141-152. doi:10.24095/hpcdp.41.5.01
17. Phillips CJ. The cost and burden of chronic pain. *Reviews in pain*. 2009;3(1):2-5.
18. Von Korff M, Scher AI, Helmick C, et al. United States national pain strategy for population research: concepts, definitions, and pilot data. *The Journal of Pain*. 2016;17(10):1068-1080.
19. Li H, Flynn DM, Highland KB, et al. Pattern of cumulative treatment hours on pain impact and PROMIS outcomes. *Military Medicine*. 2023;188(1-2):e278-e285.
20. Dahlhamer J, Lucas J, Zelaya C, et al. Prevalence of Chronic Pain and High-Impact Chronic Pain Among Adults - United States, 2016. *MMWR Morb Mortal Wkly Rep*. 2018;67(36):1001-1006. doi:10.15585/mmwr.mm6736a2
21. Yong RJ, Mullins PM, Bhattacharyya N. Prevalence of chronic pain among adults in the United States. *Pain*. 2022;163(2):e328-e332. doi:10.1097/j.pain.0000000000002291
22. Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS medicine*. 2015;12(3):e1001779.
23. Caballero FF, Soulis G, Engchuan W, et al. Advanced analytical methodologies for measuring healthy ageing and its determinants, using factor analysis and machine learning techniques: the ATHLOS project. *Scientific reports*. 2017;7(1):43955.
24. Bean GJ, Bowen NK. Item response theory and confirmatory factor analysis: Complementary approaches for scale development. *Journal of Evidence-Based Social Work*. 2021;18(6):597-618.
25. Prior G, Primatesta P, Bajekal M. *Health Survey for England: 2001. Methodology and Documentation*. Stationery Office; 2003.
26. Mindell J, Biddulph JP, Hirani V, et al. Cohort Profile: The Health Survey for England. *International Journal of Epidemiology*. 2012;41(6):1585-1593. doi:10.1093/ije/dyr199
27. Paek I, Cole K. *Using R for item response theory model applications*. Routledge; 2019.
28. Scotland GROf, Agency NISaR, Office for National Statistics CD. Data from: 2001 Census: Aggregate Data. 2020. Accessed 2024/07/20, doi:10.5257/census/aggregate-2001-2

## 9. Disclosures

This work was supported by a joint and equal investment from UKRI [grant numbers MR/W004151/1 and MR/W026872/1] and the charity Versus Arthritis [grant number 22891] as part of the UKRI Strategic Priorities Fund (SPF) Advanced Pain Discovery Platform (APDP), a co-funded initiative by UKRI (MRC, BBSRC, ESRC), and Versus Arthritis. Edmund Keogh reports unrelated consultancy services via the University of Bath to Reckitt Benckiser Health Limited. No authors reported any conflicts of interest associated with this manuscript.

## Figure legends

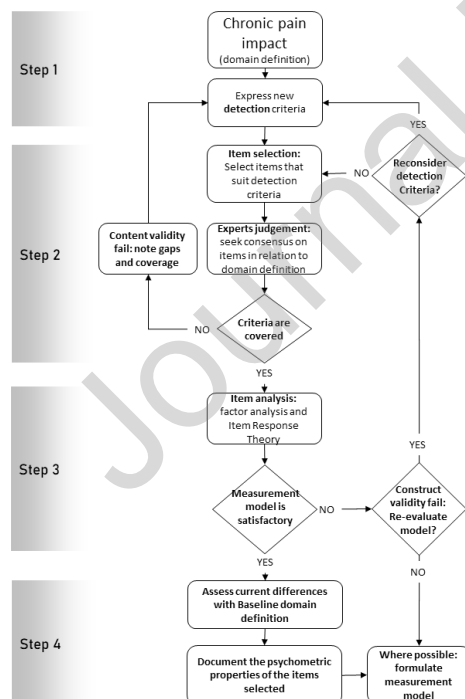


Figure 1. Decision steps from definition of concept – here ‘pain impact’ – to defining the measurement model in the longitudinal database selected, but in which there is no suitable variable for the concept

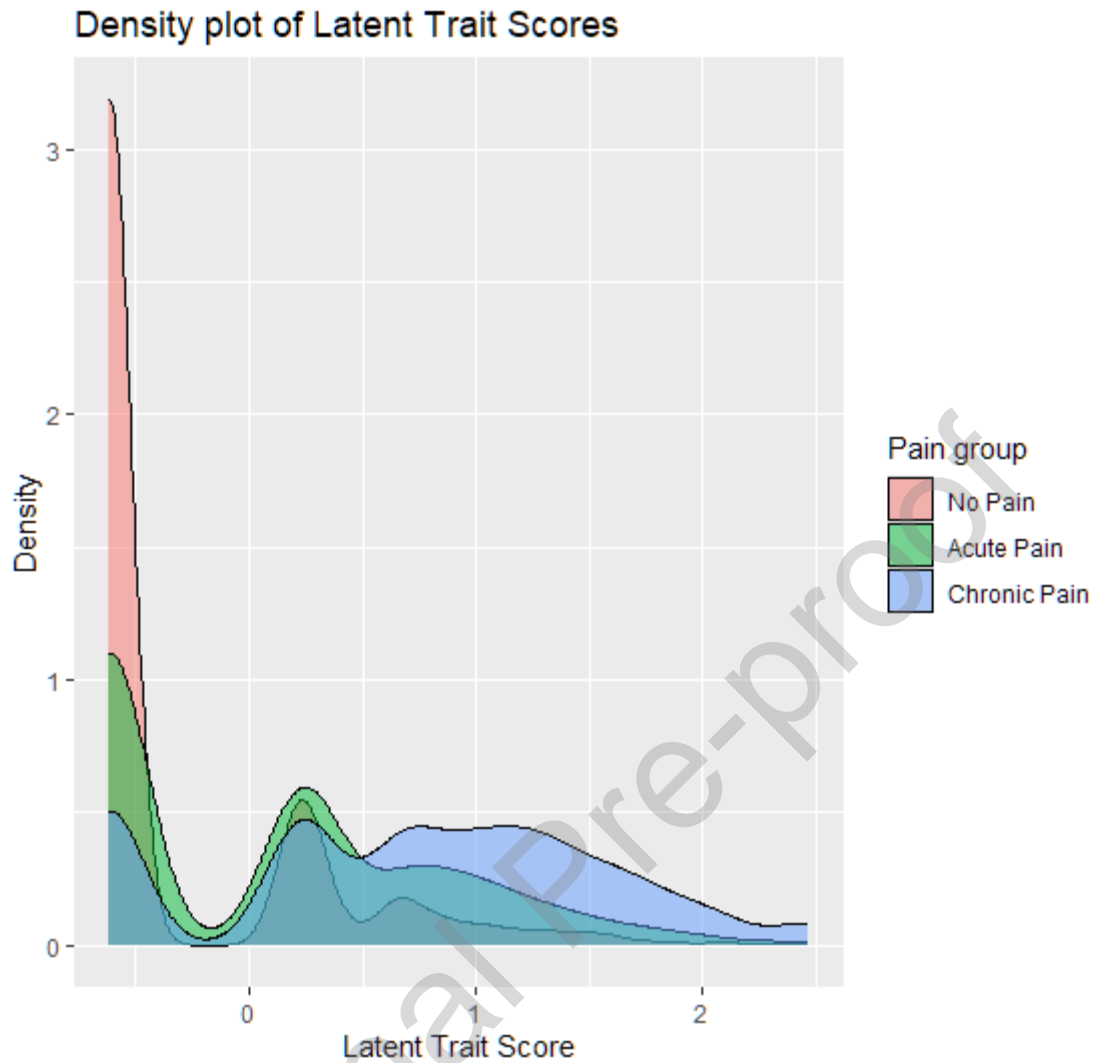


Figure 2. Distribution in ELSA of latent variable  $\theta$  (pain impact) scores between respondents with no pain, with acute pain, and with chronic pain. IRT models estimate the most likely level of the latent variable  $\theta$ . For convenience, the latent trait  $\theta$  was centred around the mean and scaled so that one unit represents one standard deviation.

## Tables

**Table 1.** First set of detection criteria for the domain definition of high impact chronic pain

Criterion Validity	Items should identify respondents with pain or chronic pain.
Anchor	Items should be anchored in pain (e.g. “because of pain ...”).
Self-care	Items should capture the experience of restrictions to self-care activities (e.g. dressing, bathing, or feeding oneself).
Work	Items should capture the experience of restrictions to usual productivity (including paid work, work in the home, volunteer work).
Social	Items should capture the experience of restrictions to valued social and recreational activities (e.g. visiting friends, going to the movies, attending clubs or religious activities).
Persistence	Items should capture a restriction which is explicitly described as persistent (lasting $\geq 3$ months).

**Table 2.** *Limitations in ELSA of the model in relation to the criteria for high impact chronic pain*

Criterion validity	Extent to which addressed in ELSA.
Anchor	Limited coverage. Most items describing limitations were not explicitly linked to a specific cause, but some were anchored in “health”, “physical problem”, or “disability”.
Self-care	Covered.
Work	Limited coverage. Items include one item on paid work, and other items on work in the home or garden, and daily activities such as shopping for groceries.
Social	Not covered as no anchor was present for these items.
Persistence	Covered with the exception of one item.

**Table 3.** *Demographic characteristics of the ELSA sample selected for this analysis*

Variables	Number of participants
Sex	

Female	5020 (44.42%)
Male	4012 (55.58%)
Age category	
50-59	2896(32.06%)
60-69	2889 (31.99%)
70-79	2167 (23.99%)
80-89	985 (10.91%)
90+	95(1.05%)
Ethnic background	
Non-white	198(2.19%)
White	8832(97.79%)
missing	<10(<0.1%)

**Table 4.** Limitations in UKB of the model in relation to the criteria for high impact chronic pain

Criterion validity	Extent to which addressed in UKB
Anchor	Minimal coverage. Most items were not anchored to pain, but some items were anchored to “health”, “illness”, “sickness” or “disability”.
Self-care	Very limited coverage restricted to diet, disability, getting up in the morning, falls, and overall health.
Work	Limited coverage restricted to employment status and hobbies
Social	Covered.
Persistence	Partial: not all items described limitations that lasted or were thought to last 3 months or more.

**Table 5.** Demographic characteristics of the UKB sample selected for this analysis

Variable	Number of participants
Sex	
Female	125,628 (57.49%)
Male	92,879 (42.51%)

## Age category

35 to 44	20,067 (9.18%)
45 to 54	61,229 (28.02%)
55 to 64	92,678 (42.41%)
65 to 74	44,533 (20.38%)

## Ethnic background

Asian or Asian British	4802 (2.20%)
Black or Black British	3818 (1.75%)
Chinese	570 (0.26%)
Missing	798 (0.37%)
Mixed	1406 (0.64%)
Other ethnic group	2176 (1.00%)
White	204,937 (93.79%)

---

## Highlights

- A workflow to develop and assess theoretically sound measurement models from large cohort study data
- Two examples show the challenges and opportunities of using pre-existing data to capture pain impact
- A codebase in R with pipelines and documentation is provided on GitHub and as supplementary material