# Title: Evaluating Frailty Index Integrity: Insights from an International Network Study

Authors: Brianne Olivieri-Mui (1,2,3), Robert Cavanaugh (2, 4), Chelsea Wong (3,5), Louisa Smith (1,2), Maytal Bivas-Benita (6), Pinchas Akiva (6), Tal El-Hay (6), Ariela Orkaby (3,5,7), Chen Yanover (6)

Author Affiliations:

1. Northeastern University, Boston, MA
2. The Roux Institute, Northeastern University, Portland, ME
3. Hebrew SeniorLife, Harvard Medical School, Roslindale, MA
4. MGH Institute of Health Professions, Boston, MA
5. Beth Israel Deaconess Medical Center, Boston, MA
6. KI Research Institute, Kfar Malal, Israel
7. Veterans Affairs Medical Center, Boston, MA

To Do:

* Figures
* Update all data
* Update table from Chen

Abstract

Deficit-accumulation Frailty indices (FIs) quantify the state of reduced physiologic reserve to recover from a health insult. Common data models (CDMs) have enabled international, interinstitutional applications of FIs. However, the validity of such applications is unknown.

We conducted an international network study comparing estimates of frailty from two EHR-based FIs: United States (US)-based Veterans Affairs FI (VAFI) and United Kingdom (UK)-based electronic FI (eFI) in individuals 40+ years old with at least 1-year of data and across 5 databases: US *All of Us* (n=159,721), US IQVIA Pharmetrics+ (n=5,099,557), UK IMRD-THIN (n=3,036,003), UK IMRD-EMIS (n=832,455), and UK BioBank (n=207,202).

In US databases, VAFI identified higher proportions of frailty ([VAFI; eFI] *All of US*: 10.3%; 2.5%, Pharmetrics+: 9.6%; 2%) while in UK databases, the eFI identified higher proportions of frailty (IMRD-THIN: <0.003%; 0.08%, IQVIA-EMIS: 0.09%; 0.4%, UK BioBank: 0.03%; 0.1%). Additional manipulations (alternative lookback periods, FI variations) are discussed.

Valid FIs would show similar distributions of frailty within each FI and systematic bias between FIs across databases. However, results demonstrate that FIs are dependent on their development context, limiting their international external validity despite CDM harmonization. We caution international application of FIs and recommend further instrument development to support international use.

# **Introduction**

Frailty indices (FIs) quantify the state of reduced physiologic reserve of older adults. They cover multiple domains of health across organ systems including cognition, comorbidity, physical function, and mental health, making them distinct from comorbidity indices (1). FIs are often guided by the deficit accumulation approach whereby each additional morbidity or deficit within each health domain means a higher and worse frailty index score (1–3). Most often each deficit is counted equally, though select claims-based FIs weigh the contribution of each condition (4,5). As such they are viewed as comprehensive measures of patient health. Their benefits span use as a decision aid for clinicians to anticipating prognostic sequelae related to illness, injury, and even surgery(6,7). Therefore they are an important tool for understanding relative health when comparing groups of people that, for example, may have similar chronologic age, but different health status.

Many FIs have been developed using observational and administrative health data including but not limited to the hospital frailty risk score (HFRS) (8), the claims-based frailty index (CFI) (4), the Veterans’ Affairs frailty index (VAFI) (9), and the electronic frailty index (eFI) (10). Each was developed and validated in healthcare data for specific geographic or demographic populations. For example, the HFRS is only validated for hospitalized patients over 75 in the US, the CFI was developed and validated for use in US Medicare claims data, the VAFI was developed and validated among US veterans using the VA healthcare system electronic health records (EHR), and the eFI was developed and validated in the UK using primary care EHR data. With many FIs to choose from, there is now a movement toward understanding frailty agreement across measures. Kim et al. (2024) at Harvard University have developed a tool for understanding differences in FIs and selecting appropriate FIs in clinical settings (11). They have also examined the application of several clinical FIs to one single database suggesting only modest agreement among measures (12,13). The variety of measures, even within only those using observational or administrative health data, speaks to the complexity of frailty and each measure, though valid in its own context, has an appropriate use in time and clinical space.

However, whether clinical time and space for a certain FI in one country’s data translates to clinical time and space for the same FI in another is not fully understood. To date, few studies have compared frailty across international clinical settings and even fewer have done it using administrative health data. examiningpopulation-level Studies that do manage such a comparison are limited to a multi-site, prefabricated and harmonized data collection method that ensures all documentation across clinical domains are uniform (14). Alternatively, literature reviews or meta analyses have also compared frailty across international populations, but seem to focus more on phenotypic frailty, frailty measured as a function of physical strength in extremities (15,16).

sacross institutions and specifically provide the infrastructure necessary to calculate and compare FIs across international databases(17)international comparisons of frailty to date: . This s and, maintaining privacy of patient-level data (18)

However, while CDMs enable calculation of frailty across international databases, they do not ensure that the tools used to calculate frailty (FIs) are valid outside of the contexts in which they were developed. Indeed, in some cases adaptations are required to adjust existing FIs to local data; for example, eFRAGICAP is an adaptation of the UK-based eFI in the context of primary healthcare centers in Barcelona (19). However, adaptations and applications outside of the original setting run the risk of misclassification. As such, we sought to examine, first, the differences in frailty prevalence for two FIs across two US and three UK EHR databases formatted for the OMOP CDM and, second, to demonstrate fidelity of the FIs within these databases. Our hypothesis was that UK patients would have lower frailty because of the universal healthcare access and known better health outcomes than the US population (20). Additionally, valid FIs should show similar distributions of frailty within each FI and systematic bias between FIs across databases.

# **Methods**

## *Study design*

This was a multinational retrospective cohort study using routinely collected healthcare data from both the US and UK, standardized to the OMOP CDM (21).

## 

## *Data sources*

1. Five years of IQVIA PharMetrics+® for Academics (2017-2022) includes US commercial claims for inpatient and outpatient healthcare, and prescriptions from private health insurers, select managed care plans and supplements for Medicaid and Medicare for ~110 million people, 16 million of which have one full year of data; these data do not allow for following people switching insurers, limiting longitudinal analysis. These data include only age and binary sex demographic data. PharMetrics+ is licensed to Northeastern University and use of the data received a non-HSR determination from the Northeastern University IRB.
2. *All of Us* (AoU) Research Program data (V7; accessed July 2025) represents a convenience sample of more than 400,000 participants across the US, with an emphasis of oversampling groups historically underrepresented in research. These data include self-report surveys on personal demographics, family and personal health and lifestyle, data from AoU admission physical exams (e.g., height, weight, blood pressure), as well as EHR from contributing regional health centers, federally qualified health centers, Veterans Affairs medical centers, and genomic data from biospecimens. All experimental protocols and data collection involving human participants are covered by the Ethics Committee/Institutional Review Board (IRB) of the *All of Us* Research Program. (22)
3. IQVIA Medical Research Data (IMRD) contains longitudinal non-identified patient EHR collected from two different UK General Practitioner (GP) clinical systems (23):
   * 1. IMRD-THIN (The Health Improvement Network) (version: 2022-09, 17M individuals) is a Cegedim database where UK practices are recruited to contribute longitudinal patient data
     2. IMRD-EMIS contains data from practices using the EMIS practice management software in Great Britain (version: 2022-12, 5.4M individuals).

The use of IMRD for research has been approved by the NHS Health Research Authority (NHS Research Ethics Committee ref 18/LO/0441) for medical and public health research; this study received Scientific Review Committee (SRC) approval Ref 23SRC015. Validity of these transformed data have been demonstrated (24).

1. The UK Biobank (UKBB) is an ongoing prospective cohort study of over 500,000 participants, residents of England, Scotland, and Wales, recruited in 2006-2010 between ages 40-69 years. Participants completed a set of questionnaires (e.g., diet and well-being), underwent a brief interview, and had their physical measurements and biological samples taken. Self-report data has been linked to EHRs for most participants. UK Biobank has received ethical approval from the UK National Health Service's National Research Ethics Service (reference 11/ NW/0382), and this work is part of project number 58770. As the research is based on the UK Biobank data, informed consent and the Declaration of Helsinki were not relevant/necessary.

## *Study Samples*

The index date for entry into the study sample was established in two ways, based on database type. For All of Us and the UKBB, we used the date 365 days after study enrollment date, to ensure sufficient lookback after entering each cohort. Because participants enter these programs at a range of dates, this provided a varied range of index dates. For PharMetrics+, IMRD-THIN, and IMRD-EMIS, a majority of participants are present in the data on the first data of data availability. Therefore, in order to to minimize bias, we selected a random visit date from each unique person-identifier to serve as the index date.

To be included in the study the participants had to be age 40+ years at the index date and had to have at least one-year continuous observation prior to the index date in the US databases and three years – in the UK databases. Three years was chosen because UK practice standards do not include coding all conditions at each encounter and previous research has demonstrated a linear relationship between time and eFI (25). We were unable to test a similar 3-year lookback in the US data, due to a lack of longitudinal data over this long of a time period. Age 40+ was used to be inclusive and was supported by the age ranges analyzed in a 2018 meta-analysis demonstrating the utility of a FI for mortality risk across ages from 40 to 80+ (26) .

## 

## *Measures and demographics*

## Following Elhussein et al (27), we defined polypharmacy as having a clinical finding of polypharmacy or prescriptions for at least 10 unique ingredients, excluding antibiotics identified as any drug in ATC J01, in 1-year lookback. Demographic information was limited to age and sex for all databases due to the deidentified nature of the data, though US databases also included (limited) race/ethnicity.

## *Frailty Indices*

VAFI: Construct validity for the VAFI has been demonstrated against other frailty assessments and used in non-veteran populations making it a valid choice for this project (28). This FI uses a list of international classification of disease version 10 (ICD-10), current procedural terminology (CPT), and healthcare common procedure coding (HCPCS) codes, available at the VA Boston github (<https://github.com/bostoninformatics/va_frailty_index>), across 31 deficits. The ICD codes from the github were mapped to Systematized Nomenclature of Medicine – Clinical Terms (SNOMED-CT) clinical terminology codes using OHDSI vocabulary relationships (21). Current Precedural Terminology (CPT) codes were mapped directly to OMOP concept IDs. Codes which did not map 1:1 to SNOMED-CT, were independently, manually reviewed by two geriatricians, CW and AO, to ensure that the manually mapped OMOP concept codes and their definitions were appropriate FI deficits. As a check of the accuracy of our transformation of the ICD and CPT, VAFI codes to OMOP standard vocabularies, we compared the prevalence of each deficit for the VAFI based on the OMOP concept IDs to the prevalence based on ICD values from the source data. This ensured that the resulting OMOP FI code-list was in high agreement with the VA Boston github source codes for the VAFI.

Deficits were considered present if the person had evidence of any of the relevant concept code and absent if all concept code was not present. The ratio of the number of deficits out of the total possible deficits resulted in a value from 0 to 1 for the VAFI. Categories of robust or fit (≤0.1), pre-frail (>0.1–0.2), and frail (mild: >0.2–0.3, moderate: >0.3–0.4, severe: >0.4) were determined by published cut-points (29).

eFI: SNOMED-CT codes for calculating eFI were determined by and provided to our study team by Andrew Clegg (30). As several of these SNOMED-CT codes are non-standard OMOP concepts, we expanded the list of codes to include standard concepts that are mapped to the non-standard ones with the following relations: 'Concept poss\_eq from', 'Concept poss\_eq to', 'Concept same\_as from', 'Concept same\_as to'. The eFI can include laboratory measures not available in the US data, including hemoglobin estimation (anamia and hematinic deficient), urine albumin or protein levels (chronic kidney disease) and Thyroid-stimulating hormone (TSH) levels (thyroid disease). Therefore, we did not include these measurements in our base comparison, but did include them for the UK databases for sensitivity analyses. There are 36 possible deficits in the eFI and the discrete eFI is typically categorized as robust or fit ( ≤0.12), mild frailty (>0.12–0.24), and frail (moderate frailty >0.24–0.36, severe frailty; >0.36) based on published cut-points(30). To match the VAFI, we reduced these categories to robust or fit ( ≤0.12), mild frailty (pre-frail) (>0.12–0.24), and frail (frailty >0.24). Reference supplemental table with the components from each FI to visualize the differences in deficits captured by each

## *Data Analysis*

Extraction of frailty concepts from the OMOP CDM and calculation of prevalence of frailty and overall frailty categories was completed using a unified code base written in R (R Core Team) using the R package dbplyr (CITE), which translates R code into the appropriate SQL code for each database (Amazon Redshift, Google Bigquery etc.). Analytical code is avaibable at:[GITHUB LINK]. Pharmetrics+, IMRD-EMIS, IMRD-THIN are all IQVIA data products licensed to their respective institutions. Data within the *Allofus* Research Program and UK BioBank are openly available.

# **Results**

Final sample sizes across databases were *All of Us* n=159,721, US IQVIA Pharmetrics+ n=5,099,557, UK IMRD-THIN n=3,036,003, UK IMRD-EMIS n=832,455, and UK BioBank n=207,202. The effects of inclusion criteria on sample sizes for each database are reported in Supplemental Table 1.

All UK and the Pharmetrics+ databases were approximately half female sex at birth, while the US All of Us database was >60% female (Table 1). A plurality of the Pharmetrics+ sample were between ages 50-65 (40.4%), in All of Us it was between ages 55-70 (44.1%), in IMRD-EMIS it was between ages 40-55 (38.5%), and in IMRD-THIN it was between ages 45-60 (37.7%); IMRD-EMIS was the youngest sample. The UKBB does not have data for ages past 75 and the largest proportion (61.5%) were between 55-70. The UKBB and All of Us were the oldest samples. The US databases appeared more frail than the UK databases regardless of the FI used.

Within the US databases, the VAFI found higher levels of frailty than the eFI ([VAFI; eFI] Pharmetrics+ 9.8%; 2.0%; All of Us 11.9%;3.3%). Within the UK databases the eFI found higher levels of frailty than the VAFI (IMRD-EMIS 0.1%;0.5%; IMRD-THIN 0%; 0.1%; UKBB 0%;0.1%) (Table 1).

Using either FI, the frailest database overall was All of Us which also had the most females and skewed towards older ages. Pharmetrics+ had an absolute difference in frailty that was 1.3% and 2.1% less than All of Us by eFI and VAFI, respectively. The least frail database by either FI measure was IMRD-THIN, though only marginally because this database strongly resembled the frailty distribution of UKBB. Despite skewing younger in age, IMRD-EMIS had the highest frailty among the UK databases. By sex, Figure 1a shows that females had slightly higher frailty and pre-frailty in the US databases, but were indistinguishable from males in the UK databases for the VAFI. For the eFI, males and females were generally indistinguishable across frailty categories (Figure 1b).

In our sensitivity analysis of the UK samples which had three years of data, we show that using a three-year lookback period decreases the proportion of who were robust at all ages. The difference at the oldest age groups was from approximately 5% fewer robust in UKBB to >10% fewer robust in the IMRD-EMIS data. When measurements are included in the eFI calculation for UK databases with 3-year lookback, there is an additional reduction in robust classification. In the oldest age groups, IMRD-EMIS shows 3.2-5.7% fewer robust individuals and IMRD-THIN shows 3.4-7.2% fewer robust when measurements are added to the eFI. The US samples were limited in their lookback period to one year, All of Us because of the recency of the study start and Pharmetrics+ due to the lack of ability to follow people from one insurance plan to the next. However, longer lookback in US data likely would elicit little to no change in frailty, similar to the VA validation studies (31).

Interactive results can be found at: https://roux-ohdsi.observablehq.cloud/interactive-frailty-indices/

# **Discussion**

Using a network study approach, we calculated 2 FIs across 2 US and 3 UK databases harmonized using the OMOP CDM. We aimed to demonstrate differences in frailty across international clinical settings and to understand the fidelity of FIs in data transformed to a CDM. On the surface, the US cohorts appeared more frail than the UK, as we hypothesized. However, variability in the prevalence of frailty between measures and across databases, even after changes were made to the lookback period during which deficits were assessed and the inclusion of measurements, implies that the way diagnoses are documented and then extracted into the CDM from a health care system is a significant factor in understanding if fidelity of FIs is maintained across CDM network studies. Specifically, the finding that VAFI demonstrates greater frailty in the US while the eFI demonstrates greater frailty in the UK suggests that frailty as evidenced by these two instruments may be more dependent on the nature of source data, rather than the prevalence of frailty in each cohort. In the following, we unpack these findings and suggest potential features of the source data that may contribute to this non-ideal pattern of results.

**Topic=Translation:** The transformation of source data to a CDM-compatible vocabulary (i.e., the Extract, Transform, Load or ETL process) requires interpreting disparate health data into a common language and format. Therefore the transformation is only as accurate as the translation, which is typically limited in the ability to make 1:1 translations of certain components of health records. For example, there is often a one-to-many and many-to-one transformation of US-centric ICD-9 or ICD-10 codes to SNOMED-CT codes which are used in the OMOP CDM. Furthermore, not all SNOMED-CT codes are considered standard codes in the OMOP CDM, and can require further transformation. These challenges with the ETL process that harmonizes databases may constitute one source of the discrepancy between our findings and for example what was originally reported for EMIS and THIN during the validation study for the eFI (30). The original study saw 20% moderate to severe frail status in THIN where we saw less than 1%; though the data used in the original study was older, from 2008-2013 (30), than ours.

**Topic = documentation** Second, the ETL does adjust for differences in the source data such as practice, billing, and cultural norms around healthcare use. For example, trasnformation to the OMOP CDM cannot overcome source data quality challenges such as missing data, misclassification, or misdiagnosis (32). In the UK, the NHS is a large publicly funded universal health care system; all UK residents have compulsory membership that makes healthcare accessible and affordable. The US, however, relies on voluntary engagement with insurance coverage, which has resulted in significant numbers whom are uninsured or underinsured without consistent access to affordable healthcare. Moreover, there is no single system in which US residents are engaged because there are options such as Medicare, Medicaid, private insurance, or the insurance marketplaces for each state, depending on age, income, or employment. The US is a mix of fee-for-service and bundled payment models in a multi-payer system where not everyone has insurance coverage or access to care (33,34), while the UK is a single payer tax-funded system where all people have equal access to care (35). In the current payment model, US providers are often incentivized to document all conditions at each encounter, and EHR systems often default to “carrying forward” diagnoses from past visits to the present (36). In the UK this is not the case. Standard practice in the UK is to maintain a problem list for active conditions precluding condition documentation at each visit (37). Even looking as far back as 3 years in UK data did not capture as many deficits as were captured in a single year of US EHR or claims. Therefore, it is likely that the discrepancies in prevalence of conditions in a one year lookback period are due to differences in healthcare documentation practices rather than true differences in condition prevalence and subsequent frailty prevalence. However, this explanation does not fully account for our findings: even when extending the UK lookback period to 3-years we found that the overall pattern of results holds between FIs.

**Topic=sampling bias** Documentation is also dependent on who is seeking healthcare. The sample captured by each dataset represented an array of convenience samples from both countries. Even though the UKBB had a sampling frame, their efforts at a representative sample were diminished by a 5.5% response rate (38). The *All of Us* research program in the US seeks to oversample individuals from groups underrepresented in biomendical research but does not report a sampling frame(39). Subsequently both UKBB and *All of Us* display volunteer bias; the UKBB in particular (40); volunteers are generally less frail, and female (40,41). In comparing sexes, male to female, we saw little difference in their levels of frailty, which is contrary to known statistics that show females experience frailty more often than males as they age (42). In the UK EHR, a most interesting difference was observed where EMIS is a medical record program widely used in the UK and THIN is a research oriented dataset from providers that are all trained for data collection and entry, who volunteer to contribute their practices’ data and to use the Vision medical record program (43–45). THIN was more often robust in our study and is considered more broadly representative of the UK than EMIS, which is thought to reflect more afluent groups. This suggests some amount of selection bias and underlying cohort difference play a role in our findings in each database, a critical consideration for contextualizing our findings and future network studies.

**Topic=impacts on frailty estimates** As a result of these potential differences and explanations, we saw extreme variability in frailty prevalences in all databases from both the eFI and VAFI prompting us to question the validity of using FIs developed in single contexts (such as the eFI and VAFI) in international studies, whether comparison of frailty is the primary goal, or frailty is used as a covariate. Severe frailty in the US is around 10% of an older population and frailty prevalence tends to increase as age advances (46). However, we saw increases in frailty toward 40% in Pharmetrics+ and a maximum of 25% frail in AoU by age 80+ using the VAFI. This increased proportion is higher than expectations and based on validation studies in non-veteran populations, unlikely due to differences between the veteran population and these samples (28). The eFI returns results closer to expected in oldest US populations, around 10% frail.

In the UK, established estimates suggest frailty prevalence is around 8% however there is significant variation by rurality and proximity to coastal locations (47). At less than 5%, our results imply much lower prevalence of frailty in the UK than other studies, particularly using the VAFI. Using the eFI the highest frailty was 3.25% in IMRD-EMIS, a surprise because the IMRD-THIN data, in its original non-CDM format, served as the external validation data for the eFI. That study demonstrated 43% robust, 37% mild frailty (pre-frailty in our study), and 20% frail in THIN. This discrepancy could be related to the use of a problem list in UK practice and the exclusion of that from the ETL for these data. However further research is necessary to understand the impact of CDMs on validity of FIs, particularly because each ETL is data and site specific.

Though we have discussed several limitations as the goal of our discussion, there were select limitations that prevented certain additional comparisons. Requiring more than one year for US databases created significant selection bias because of the inability to track individuals through switching of private plans in Pharmetrics+ and the recency with which All of Us began (2018). However, one year is the most commonly used lookback for frailty calculations in US data given the practice of coding all conditions at each visit in the US health system. Moreover, the validation studies of the VAFI used one and three year lookback periods demonstrating little difference extending beyond one year and the eFI has no predefined lookback period. There was no measurement table available in three of the datasets. Only IMRD data contained measurements, use of which changed the prevalence of frailty in each dataset as demonstrated on our website.

CONCLUSION: Using existing FIs in network studies of frailty may not be a valid approach to understanding international trends in frailty. We speculate that potential threats to the exeternal validty of international applications of frailty include differences in documentation, translation, and data tables populated in the OMOP CDM and impact of ETL processes from the source data into a harmonized common data model vocabulary. We caution on use of existing FI measures in contexts in which they have not yet been validated and suggest that it may be necessary to create a FI that is not tied to any one database, but rather is validated for use amidst the complexities of translated multi-source health data, such as the OMOP CDM.

Key take aways:

* FIs seem to be dependent on their development context, limiting their external validity in international applications
* CDMs may alter the validity of the measures due to details lost in translation from source data to harmonized data.

Table 1. Baseline characteristics of samples from each database that have at least one year of data.

[ Table is autogenerated by R. In folder. ]

Note: Race/ethnicity is not included in table 1 because these data are not available from the IQVIA products due to the deidentified nature of those data. UK BioBank does not collect race/ethnicity data. eFI frailty categories cut-points for robust (0–0.12), pre-frail (>0.12–0.24), and frail (>0.24); VAFI frailty category cut-points for robust (≤0.1), pre-frail (>0.1–0.2), and frail (>0.2)

*A graph of a graph

AI-generated content may be incorrect.*

*A graph of different colored lines

AI-generated content may be incorrect.*

References

1. Searle SD, Mitnitski A, Gahbauer EA, et al. A standard procedure for creating a frailty index. *BMC Geriatr.* 2008;8(1):24–24.

2. Rockwood K, Mitnitski A. Frailty in Relation to the Accumulation of Deficits. *J. Gerontol. Ser. A*. 2007;62(7):722–727.

3. Mitnitski AB, Mogilner AJ, MacKnight C, et al. The mortality rate as a function of accumulated deficits in a frailty index. *Mech. Ageing Dev.* 2002;123(11):1457–1460.

4. Kim DH, Schneeweiss S, Glynn RJ, et al. Measuring Frailty in Medicare Data: Development and Validation of a Claims-Based Frailty Index. *J. Gerontol. A. Biol. Sci. Med. Sci.* 2018;73(7):980–987.

5. Best K, Shuweihdi F, Alvarez JCB, et al. Development and external validation of the electronic frailty index 2 using routine primary care electronic health record data. *Age Ageing*. 2025;54(4):afaf077.

6. Kim DH, Rockwood K. Frailty in Older Adults. *N. Engl. J. Med.* 2024;391(6):538–548.

7. Cheng D, Dumontier C, Sheikh AR, et al. Prognostic value of the veterans affairs frailty index in older patients with non-small cell lung cancer. *Cancer Med.* 2022;11(15):3009–3022.

8. Gilbert T, Neuburger J, Kraindler J, et al. Development and validation of a Hospital Frailty Risk Score focusing on older people in acute care settings using electronic hospital records: an observational study. *Lancet Lond. Engl.* 2018;391(10132):1775–1782.

9. Cheng D, DuMontier C, Yildirim C, et al. Updating and Validating the U.S. Veterans Affairs Frailty Index: Transitioning From ICD-9 to ICD-10. *J. Gerontol. A. Biol. Sci. Med. Sci.* 2021;76(7):1318–1325.

10. Clegg A, Bates C, Young J, et al. Development and validation of an electronic frailty index using routine primary care electronic health record data. *Age Ageing*. 2016;45(3):353–360.

11. Kim DH, Cheslock M, Sison SM, et al. eFrailty: Making frailty assessment accessible to clinicians and researchers. *J. Am. Geriatr. Soc.* 2024;

12. Sison SDM, Shi SM, Kim KM, et al. A crosswalk of commonly used frailty scales. *J. Am. Geriatr. Soc.* 2023;71(10):3189–3198.

13. Orkaby AR, Hshieh TT, Gaziano JM, et al. Comparison of two frailty indices in the physicians’ health study. *Arch. Gerontol. Geriatr.* 2017;71:21–27.

14. Santos-Eggimann B, Cuénoud P, Spagnoli J, et al. Prevalence of frailty in middle-aged and older community-dwelling Europeans living in 10 countries. *J. Gerontol. A. Biol. Sci. Med. Sci.* 2009;64(6):675–681.

15. Ofori-Asenso R, Chin KL, Mazidi M, et al. Global Incidence of Frailty and Prefrailty Among Community-Dwelling Older Adults: A Systematic Review and Meta-analysis. *JAMA Netw. Open*. 2019;2(8):e198398.

16. Nguyen T, Cumming RG, Hilmer SN. A review of frailty in developing countries. *J. Nutr. Health Aging*. 2015;19(9):941–946.

17. Voss EA, Makadia R, Matcho A, et al. Feasibility and utility of applications of the common data model to multiple, disparate observational health databases. *J. Am. Med. Inform. Assoc.* 2015;22(3):553–564.

18. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *J. Am. Med. Inform. Assoc. JAMIA*. 2013;20(1):144–151.

19. Orfila F, Carrasco-Ribelles LA, Abellana R, et al. Validation of an electronic frailty index with electronic health records: eFRAGICAP index. *BMC Geriatr.* 2022;22(1):404.

20. Desai M, Rachet B, Coleman MP, et al. Two countries divided by a common language: health systems in the UK and USA. *J. R. Soc. Med.* 2010;103(7):283–287.

21. OHDSI. The Book of OHDSI. OHDSI; 2019.(https://books.google.co.il/books?id=JxpnzQEACAAJ)

22. All of Us Research Program Investigators, Denny JC, Rutter JL, et al. The “All of Us” Research Program. *N. Engl. J. Med.* 2019;381(7):668–676.

23. Edwards L, Pickett J, Ashcroft DM, et al. UK research data resources based on primary care electronic health records: review and summary for potential users. *BJGP Open*. 2023;7(3):BJGPO.2023.0057.

24. Candore G, Hedenmalm K, Slattery J, et al. Can We Rely on Results From IQVIA Medical Research Data UK Converted to the Observational Medical Outcome Partnership Common Data Model?: A Validation Study Based on Prescribing Codeine in Children. *Clin. Pharmacol. Ther.* 2020;107(4):915–925.

25. Ambagtsheer RC, Beilby J, Dabravolskaj J, et al. Application of an electronic Frailty Index in Australian primary care: data quality and feasibility assessment. *Aging Clin. Exp. Res.* 2019;31(5):653–660.

26. Kojima G, Iliffe S, Walters K. Frailty index as a predictor of mortality: a systematic review and meta-analysis. *Age Ageing*. 2018;47(2):193–200.

27. Elhussein L, Burn E, Delmestri A, et al. Quantifying polypharmacy in elderly people: a comparison between source and mapped data in the UK Clinical Practice Research Datalink GOLD. In: *Abstract presentations*. Online: 2021 (Accessed September 5, 2024)(https://www.ohdsi.org/2021-global-symposium-showcase-36/). (Accessed September 5, 2024)

28. DuMontier C, Hennis R, Yilidirim C, et al. Construct validity of the electronic Veterans Affairs Frailty Index against clinician frailty assessment. *J. Am. Geriatr. Soc.* 2023;71(12):3857–3864.

29. Orkaby AR, Nussbaum L, Ho Y-L, et al. The Burden of Frailty Among US Veterans and Its Association With Mortality, 2002-2012. *J. Gerontol. A. Biol. Sci. Med. Sci.* 2019;74(8):1257–1264.

30. Clegg A, Bates C, Young J, et al. Development and validation of an electronic frailty index using routine primary care electronic health record data. *Age Ageing*. 2016;45(3):353–360.

31. Cheng D, DuMontier C, Yildirim C, et al. Updating and Validating the U.S. Veterans Affairs Frailty Index: Transitioning From ICD-9 to ICD-10. *J. Gerontol. A. Biol. Sci. Med. Sci.* 2021;76(7):1318–1325.

32. Kent S, Burn E, Dawoud D, et al. Common Problems, Common Data Model Solutions: Evidence Generation for Health Technology Assessment. *Pharmacoeconomics*. 2021;39(3):275–285.

33. Jia L, Meng Q, Scott A, et al. Payment methods for healthcare providers working in outpatient healthcare settings. *Cochrane Database Syst. Rev.* 2021;1(1):CD011865.

34. Petrou P, Samoutis G, Lionis C. Single-payer or a multipayer health system: a systematic literature review. *Public Health*. 2018;163:141–152.

35. Jacob Z. A Comparative Analysis of the US and UK Health Care Systems – Michigan Journal of Economics. *Mich. J. Econ.* [electronic article]. 2023;Online. (https://sites.lsa.umich.edu/mje/2023/05/26/a-comparative-analysis-of-the-us-and-uk-health-care-systems/). (Accessed August 13, 2025)

36. Tsou AY, Lehmann CU, Michel J, et al. Safe Practices for Copy and Paste in the EHR. *Appl. Clin. Inform.* 2017;8(1):12–34.

37. Documenting in patient clinical records : CPPE. (https://www.cppe.ac.uk/programmes/l/clinical-e-02). (Accessed August 13, 2025)

38. Allen N, Sudlow C, Downey P, et al. UK Biobank: Current status and what it means for epidemiology. *Health Policy Technol.* 2012;1(3):123–126.

39. All of Us Research Program Investigators, Denny JC, Rutter JL, et al. The “All of Us” Research Program. *N. Engl. J. Med.* 2019;381(7):668–676.

40. van Alten S, Domingue BW, Faul J, et al. Reweighting UK Biobank corrects for pervasive selection bias due to volunteering. *Int. J. Epidemiol.* 2024;53(3):dyae054.

41. Kambara MS, Sharma S, Spouge JL, et al. Increasing Representativeness in the All of Us Cohort Using Inverse Probability Weighting. *medRxiv*. 2024;2024.10.02.24314774.

42. Gordon EH, Peel NM, Samanta M, et al. Sex differences in frailty: A systematic review and meta-analysis. *Exp. Gerontol.* 2017;89:30–40.

43. Haynes K, Bilker WB, TenHave TR, et al. Temporal and Within Practice Variability in The Health Improvement Network. *Pharmacoepidemiol. Drug Saf.* 2011;20(9):948–955.

44. Denburg MR, Haynes K, Shults J, et al. Validation of The Health Improvement Network (THIN) database for epidemiologic studies of chronic kidney disease. *Pharmacoepidemiol. Drug Saf.* 2011;20(11):1138–1149.

45. Lewis JD, Schinnar R, Bilker WB, et al. Validation studies of the health improvement network (THIN) database for pharmacoepidemiology research. *Pharmacoepidemiol. Drug Saf.* 2007;16(4):393–401.

46. Shi SM, Olivieri-Mui B, Park CM, et al. Frailty in Medicare Advantage Beneficiaries and Traditional Medicare Beneficiaries. *JAMA Netw. Open*. 2024;7(8):e2431067.

47. Sinclair DR, Maharani A, Chandola T, et al. Frailty among Older Adults and Its Distribution in England. *J. Frailty Aging*. 2022;11(2):163–168.

Supplemental Materials

*A graph with a line graph

Description automatically generated*

Supplemental Figure 1. One-way agreement between VAFI deficit prevalence using the OMOP CDM concept codes and the source data ICD-10 codes from Pharmetrics+ to assess the quality of data transformation.

*Note: the correlation coefficient demonstrated high correlation (ICC=0.995) between the OMOP CDM transformation and the source coding for deficits for each FI. This implies validity of the data transformation.*

Supplemntal table 1. Consort table with inclusion criteria for all datasets

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Pharmetrics | All of Us | IMRD-EMIS | IMRD-THIN | UKBB |
| People in data | 34,808,145 | 413,360 | 5,187,327 | 13,209,954 | 502,363 |
| Visit in 2010 or later | 28,076,893 | 287,012 | 2,969,868 | 8,038,288 | 501,472 |
| Sufficient lookback\* | 9,016,272 | 214,375 | 1,480,491 | 5,091,067 | 207,202 |
| Age >=40 years | 5,099,557 | 159,721 | 832,455 | 3,036,003 | 207,202 |
| Final sample | 5,099,557 | 159,721 | 832,455 | 3,036,003 | 207,202 |

\* At least 1 year in US databases or 3 years in UK databases

Extra information

Useful citations:

**Ambagtsheer RC, Beilby J, Dabravolskaj J, Abbasi M, Archibald MM, Dent E. Application of an electronic Frailty Index in Australian primary care: data quality and feasibility assessment. Aging Clin Exp Res. 2019 May;31(5):653-660. doi: 10.1007/s40520-018-1023-9. Epub 2018 Aug 21. PMID: 30132204.** = *There was significant variation with respect to time taken to extract the required data items for the eFI. There was a moderate positive correlation between time (minutes) and the eFI score, (r = 0.54, P = < 0.01), indicating that the higher the frailty score, the longer it took to extract the data from the record. The median extraction time per record was 8 min (range 5–20 min). Almost half of all records (n = 27, 45.0%) were deemed easy to extract, with a further 40% (n = 24) perceived to be of neutral difficulty. In contrast, 15% (n = 9) were judged difficult to extract. Most commonly, the eFI data were perceived as difficult to extract when there were either many problems listed within the patient record, and/or where the majority of items were located other than on the summary problem list for the patient, as indicated by free text commentary entered by the nurse.*

another example of eFI having no real look back was this one: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6814149/>

they relied on up to 9 years of data prior to their followup period but never specifically define an eFI look back.