

10-26-2016

# Preparing Nursing Home Data from Multiple Sites for Clinical Research – A Case Study Using Observational Health Data Sciences and Informatics

Richard D. Boyce PhD

*University of Pittsburgh - Main Campus, rdb20@pitt.edu*

Steven M. Handler MD, PhD

*University of Pittsburgh, handler@pitt.edu*

Jordan F. Karp MD

*University of Pittsburgh, karpjf@upmc.edu*

Subashan Perera PhD

*University of Pittsburgh, ksp9@pitt.edu*

*See next pages for additional authors*

Follow this and additional works at: <http://repository.edm-forum.org/egems>



Part of the [Epidemiology Commons](#), and the [Geriatrics Commons](#)

## Recommended Citation

Boyce, Richard D. PhD; Handler, Steven M. MD, PhD; Karp, Jordan F. MD; Perera, Subashan PhD; and Reynolds, Charles F. III, MD (2016) "Preparing Nursing Home Data from Multiple Sites for Clinical Research – A Case Study Using Observational Health Data Sciences and Informatics," *eGEMs (Generating Evidence & Methods to improve patient outcomes)*: Vol. 4: Iss. 1, Article 21.

DOI: <http://dx.doi.org/10.13063/2327-9214.1252>

Available at: <http://repository.edm-forum.org/egems/vol4/iss1/21>

This Methods Case Study is brought to you for free and open access by the the Publish at EDM Forum Community. It has been peer-reviewed and accepted for publication in eGEMs (Generating Evidence & Methods to improve patient outcomes).

The Electronic Data Methods (EDM) Forum is supported by the Agency for Healthcare Research and Quality (AHRQ), Grant 1U18HS022789-01. eGEMs publications do not reflect the official views of AHRQ or the United States Department of Health and Human Services.

---

# Preparing Nursing Home Data from Multiple Sites for Clinical Research – A Case Study Using Observational Health Data Sciences and Informatics

## Abstract

**Introduction:** A potential barrier to nursing home research is the limited availability of research quality data in electronic form. We describe a case study of converting electronic health data from five skilled nursing facilities to a research quality longitudinal dataset by means of open-source tools produced by the Observational Health Data Sciences and Informatics (OHDSI) collaborative.

**Methods:** The Long-Term Care Minimum Data Set (MDS), drug dispensing, and fall incident data from five SNFs were extracted, translated, and loaded into version 4 of the OHDSI common data model. Quality assurance involved identifying errors using the Achilles data characterization tool and comparing both quality measures and drug exposures in the new database for concordance with externally available sources.

**Findings:** Records for a total 4,519 patients (95.1%) made it into the final database. Achilles identified 10 different types of errors that were addressed in the final dataset. Drug exposures based on dispensing were generally accurate when compared with medication administration data from the pharmacy services provider. Quality measures were generally concordant between the new database and Nursing Home Compare for measures with a prevalence  $\geq 10\%$ . Fall data recorded in MDS was found to be more complete than data from fall incident reports.

**Conclusions:** The new dataset is ready to support observational research on topics of clinical importance in the NH including patient-level prediction of falls. The extraction, translation, and loading process enabled the use of OHDSI data characterization tools that improved the quality of the final dataset.

## Acknowledgements

This research was funded in part by the US National Institute on Aging (K01AG044433), the National Library of Medicine (R01LM011838), NIMH P30 MH90333, and the UPMC Endowment in Geriatric Psychiatry.

## Keywords

elderly, individuals who need chronic care, common data model, informatics

## Disciplines

Epidemiology | Geriatrics

## Creative Commons License



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 License](https://creativecommons.org/licenses/by-nc-nd/3.0/).

---

**Authors**

Richard D Boyce, *University of Pittsburgh - Main Campus*; Steven M Handler, *University of Pittsburgh*; Jordan F Karp, *University of Pittsburgh*; Subashan Perera, *University of Pittsburgh*; Charles F Reynolds, *University of Pittsburgh*.



# Preparing Nursing Home Data from Multiple Sites for Clinical Research – A Case Study Using Observational Health Data Sciences and Informatics

Richard D. Boyce, PhD; Steven M. Handler, MD, PhD; Jordan F. Karp, MD; Subashan Perera, PhD; Charles F. Reynolds III, MD<sup>1</sup>

## ABSTRACT

**Introduction:** A potential barrier to nursing home research is the limited availability of research quality data in electronic form. We describe a case study of converting electronic health data from five skilled nursing facilities to a research quality longitudinal dataset by means of open-source tools produced by the Observational Health Data Sciences and Informatics (OHDSI) collaborative.

**Methods:** The Long-Term Care Minimum Data Set (MDS), drug dispensing, and fall incident data from five SNFs were extracted, translated, and loaded into version 4 of the OHDSI common data model. Quality assurance involved identifying errors using the Achilles data characterization tool and comparing both quality measures and drug exposures in the new database for concordance with externally available sources.

**Findings:** Records for a total 4,519 patients (95.1%) made it into the final database. Achilles identified 10 different types of errors that were addressed in the final dataset. Drug exposures based on dispensing were generally accurate when compared with medication administration data from the pharmacy services provider. Quality measures were generally concordant between the new database and Nursing Home Compare for measures with a prevalence  $\geq 10\%$ . Fall data recorded in MDS was found to be more complete than data from fall incident reports.

**Conclusions:** The new dataset is ready to support observational research on topics of clinical importance in the nursing home including patient-level prediction of falls. The extraction, translation, and loading process enabled the use of OHDSI data characterization tools that improved the quality of the final dataset.

<sup>1</sup>University of Pittsburgh

## Introduction

The nursing home is a highly utilized, heavily regulated, and understudied care setting. There are approximately 16,000 certified nursing home facilities that provide care for more than 1.4 million residents,<sup>1</sup> and 10 percent of all persons over 85 receive care in that setting.<sup>2</sup> Clinical researchers have noted that much more research within the nursing home setting is needed to obtain improvements in the quality and effectiveness of care received by residents.<sup>3</sup> Compared to community-dwelling patients, residents in the nursing home setting are more likely to be older and have a greater burden of medical comorbidity. Nearly half of the nursing home population suffers from Alzheimer's disease or a related dementia,<sup>4</sup> compared to one out of every eight persons in the general population of persons over the age of 65.<sup>5</sup> Nursing home patients also tend to be prescribed more medications and to be more functionally impaired than elderly persons in the community.

Potential barriers to research in the nursing home setting include the unique characteristics of the patient population, as well as the complexity of the clinical environment. The population of any given nursing home is generally a combination of heterogeneous patient types. A significant proportion of patients might be in the home for only a short period to receive targeted physical or occupational therapy. Another group of patients might be long-term residents who require skilled nursing to accomplish activities of daily living. There are also patients receiving care for advanced dementia, conditions requiring intubation, severe psychiatric or addiction disorders, or hospice care as they approach the end of life. The complex care setting includes physicians (both primary care and specialist), nurses of various levels of training, occupational and physical therapists, nurse practitioners, pharmacists, dieticians, and social workers.

Another potential barrier to nursing home research is the limited availability of research quality data in electronic form. Here we describe a case study of converting electronic health data that are readily available in many nursing homes into a research-quality, longitudinal data set for skilled nursing facilities (SNFs) by means of open-source tools produced by the Observational Health Data Sciences and Informatics (OHDSI) collaborative.<sup>6</sup> OHDSI provides advanced, open-source clinical research tools including a common data model (CDM), standard vocabulary of clinical terminologies, and various software programs to assist with clinical research. We used these resources to link electronic health data created during SNF patient care from five sites in Pennsylvania for the initial purpose of studying the safety of psychotropic-drug therapy and fall adverse events, tracking quality measures (QMs), generating population-level analytics, and triggering patient-specific clinical interventions. After providing context for this work, we describe how we loaded data from multiple nursing home sites and validated the new nursing home database as useful for clinical research. We then discuss lessons learned and some implications of the results for our future clinical research.

## Background

There are a few data sources that can be used to conduct nursing home research. Population-level data includes the National Nursing Home Survey,<sup>7</sup> the National Long-Term Care Survey,<sup>8</sup> Online Survey and Certification Reporting System,<sup>9</sup> and the Long-Term Care Minimum Data Set (MDS).<sup>10</sup> Of these data sets, the first two are both cross-sectional surveys conducted more than 10 years ago. The Online Survey and Certification Reporting System provides operational characteristics of specific nursing facilities and aggregated patient characteristics. Only the MDS contains longitudinal data collected at regular intervals during the course of patient care and for that reason is the focus of this report.



## Clinical Research with Nursing Home Minimum Data Set (MDS) Data

Specially trained assessment coordinators collect MDS data for all skilled-nursing patients in any facility certified for Medicare and Medicaid reimbursement. The collection of MDS data is an administrative procedure that involves completing a relatively complex survey-like form.<sup>11</sup> Once completed, the data collected in an MDS form are transmitted to the Centers for Medicare and Medicaid Services (CMS) where it is used to identify the resource utilization group that a patient belongs to.<sup>11</sup> The data might also be sent to insurers besides Medicare (e.g., a state Medicaid agency) for a similar purpose of facilitating reimbursement. The fact that the MDS is used for regulatory and billing purposes, together with the fact that validated measures are used to collect much of the data (see below), means that the data may be able address certain clinical research questions. Assuming that ethical considerations are addressed properly, it is possible for researchers having Institutional Review Board (IRB) approval either to request a sample of MDS data from CMS, or to work directly with nursing home facilities to use MDS data for clinical research.

## Other Nursing Home Data Potentially Useful for Clinical Research

While the MDS is of special interest because of its ubiquitous nature, a growing number of nursing homes (or organizations they contract with) are collecting health care related data electronically.<sup>12</sup> Many facilities are implementing electronic medical records, and the majority already generate laboratory and pharmacy data in electronic format.<sup>12</sup> This means that many nursing homes are already collecting data that are potentially useful for generating both dynamic analytic reports of a given population and interventions that actively monitor patients for potential risks or currently active

adverse events.<sup>13,14</sup> However, there is little information to guide organizations on how best to assemble a research-quality, longitudinal nursing home data set that combines MDS and other electronic sources of data. The remaining sections of this paper present the methods used by the authors to accomplish this task, the results, and lessons learned.

## Case Description

### Setting

We obtained data from five skilled nursing facilities (SNFs) affiliated with a single nonprofit, academically affiliated, health system located in Pennsylvania. The facilities had a combined total of approximately 709 skilled and long-term care beds (range 80–174). Two facilities were located in an urban setting, two in suburban settings, and one was rural. All facilities provided skilled nursing services such as occupational and physical therapy.

### Data Sources and Study Period

Data included de-identified resident assessment data (MDS 3.0), drug dispensing data, and fall incident reports submitted by nursing staff within the five facilities for all patients who had a stay in the nursing home during the period spanning October 1, 2010 to January 31, 2014.

### Extraction Process

We had previously worked with a clinical research informatics service affiliated with the health system to create a digital archive of HL7 data transmitted from the computer systems used in the facilities. The archive stored data from each source in a relatively unprocessed text format. We then requested an extract of data from each source as text files with the following data elements:

- MDS 3.0: assessment type, the facility where assessment was completed, patient age, race,

marital status, cognitive status (Brief Interview of Mental Status<sup>15</sup>), functional status (Katz Activities of Daily Living instrument<sup>16</sup>), depression rating (Patient Health Questionnaire<sup>17</sup>), delirium status (Confusion Assessment Method<sup>18</sup>), behavioral status, wandering status, pain status, chronic condition diagnoses, acute condition diagnoses, history of falls, history of injurious falls, and history of exposure to various drugs including antipsychotics;

- Drug dispensing: prescription start and stop dates, drug identifiers, dosage, form, quantity dispensed, days supplied, schedule, written directions (i.e., “sig” line), a code for the ordering clinician, and the facility that the resident was staying in at the time of the order;
- Fall incidents: the date of the fall and the location where the fall occurred; and
- Population census: a daily account of all patients admitted to any of the five SNF facilities including the location (site and care wing within the facility), date of admission, and length of stay.

### The OHDSI-Based Clinical Research Framework

The open-source, clinical-research framework developed by the OHDSI collaborative was chosen as the target environment. One reason for this choice is that the set of clinical research tools developed by the collaborative is available to any interested organization at no cost. Another reason is that the collaborative includes a clinical research network that could support the growth of large-scale, multisite SNF research. In line with this decision, the target schema for the clinical research database was the OHDSI common data model, version 4 (version 5 came later).<sup>19</sup> All source data on drugs, observations, and conditions were translated to concepts provided by version 4.5 of the OHDSI standard vocabulary (April 2014).<sup>20</sup> The schema and loading scripts were downloaded from the OHDSI GitHub repository on April 3, 2014.<sup>19</sup>

### Translation Process

Data were received as separate text files that were each checked for consistent date and row formatting and were scanned for unusual characters. The drug dispensing data were further processed to remove duplicate orders that were present because the dispensing data often provided two records for each order, one indicating the start of a prescription (“start” record), and the other indicating the end of a prescription (“stop” record).

Clinical entities and concepts mentioned in the source data were translated to the OHDSI standard vocabulary as follows:

- Dispensing drug orders were originally coded in the using Medi-Span Generic Product Identifiers (GPI). GPIs were mapped to Level 1 RxNorm<sup>21</sup> clinical drug concept identifiers in the OHDSI standard vocabulary by identifying GPIs in the OHDSI SOURCE\_TO\_CONCEPT\_MAP table.
- MDS 3 data were mapped to the OHDSI standard vocabulary by first identifying mappings to Logical Observation Identifiers Names and Codes (LOINC) and Systematized Nomenclature of Medicine (SNOMED) Clinical Terms (CT) provided by a white paper written by the United States Department of Health and Human Services.<sup>22</sup> The LOINC and SNOMED CT codes were then identified in the OHDSI Standard Vocabulary to arrive at the OHDSI concept identifiers. A script written in the Python programming language<sup>23</sup> implemented the conversion of the original MDS 3 data to a more human-readable representation.
- Fall incident data were mapped directly to the OHDSI concept identifier for the Medical Dictionary for Regulatory Activities (MedDRA) preferred term for “fall.”



## Loading Process

Data from all five SNF locations were inserted into the CDM Location table using a programming script written in Java:

- Data were loaded into the Person table by iterating through all MDS 3 records to identify each patient's research identifier, gender, year of birth, race, and code for their location based on the MDS data. We assumed that patients did not transition between the nursing home facilities.
- Observation periods were determined from MDS 3 data using the business rules provided by the MDS 3 Quality Measures user manual.<sup>24</sup> No observation periods were created for residents whose records indicated that they entered and left the SNF facility on the same day. If we could not locate a discharge date in the resident's MDS 3 data, we set the observation end date to be January 1, 2014.
- The Procedure table records the date, procedure, and procedure type for procedures done for a resident during the stay. In the current case, each MDS record is considered an administrative procedure that has several subtypes (admission, discharge, yearly, death, change in status). Because none of the procedure types have direct codes in the standard vocabulary, we mapped them to two general SNOMED CT concepts ("Evaluation AND/OR management - new patient" and "Evaluation AND/OR management - established patient") and retained the custom codes in the PROCEDURE\_SOURCE\_VALUE row of the table, in order to support querying by specific MDS procedure types.
- Observations were loaded from the human-readable version of the MDS 3 data that had also been mapped to OHDSI concept identifiers (see above).
- MDS condition data were loaded into the "Condition Era" (a time span when the Person is assumed to have a given condition) table. We

made note that some conditions are generally temporary (e.g., thrombosis, malnutrition, and most infections), while other conditions (e.g., Alzheimer's disease, hypertension, and cirrhosis) effectively follow residents throughout their entire stay in the SNF. Thus, for those conditions that are generally treatable in the nursing home, we set the end dates of condition era to be the expected date of the next MDS report. We set the condition era end dates for the remaining conditions to be the end of the observation period.

- Constructing proper "drug eras" (a time span when the Person is assumed to be exposed to a particular active ingredient; not the same as a Drug Exposure) required additional processing because, for various reasons, it is possible that a drug order for a given SNF resident that is dispensed from the pharmacy will be not be administered to the resident (i.e., placed on hold). A very common situation involves the patient leaving the SNF facility temporarily, perhaps to visit family or go to the hospital. In the case of the current study, the HL7 data we received did not capture the "hold" cases. Drug eras created from only the dispensing data would therefore incorrectly include dates when the resident was not actually in the facility.

We thought that more accurate drug exposure periods could be created by using both the dispensing data and data on the resident's observation periods that were derived as described above. We wrote an algorithm that implemented the simple heuristic that a drug era ends during the start of a temporary leave and a new drug era begins on the readmission date, so long as the dispensing order does not indicate that the medication was stopped. This algorithm was applied to both regular and *pro re nata*, "as needed," (PRN) dispensing orders for data loaded into the Drug Era table and Drug Exposure tables.



## Quality Assurance

It is important to check that the database resulting from the extract-transform-load (ETL) protocol is as free as possible from systematic error or bias in order to minimize the risk of erroneous research results. Our quality assurance efforts involved two activities. We first used the OHDSI Achilles and Achilles Heel data characterization tools<sup>25</sup> to identify errors in the database that could be attributed to the ETL process. After addressing each issue, we queried the database for data on drug exposures and QMs, then compared the results for concordance with externally available sources. Specifically, counts of patient exposure to specific drugs associated with fall adverse events were compared between the CDM and data provided by the pharmacy services provider for the five facilities.

Second, for further quality assurance, we created CDM queries for seven QMs that CMS derives from MDS 3 data. We then executed the queries on the database and compared the results from the CDM data set with data reported publically by the CMS Nursing Home Compare program<sup>26</sup> and, where relevant, those from the Pharmacy Services Provider. Chi-square goodness of fit tests were used to determine agreement between expected and CDM-observed values within each facility for each quarter for which sufficient data were available. The procedure was run on consecutive quarterly periods for which data were available from the facilities starting from the second quarter of 2011, the earliest quarter that we could locate publically available Nursing Home Compare data. When cell counts were not sufficient for the goodness of fit test, we descriptively compared expected and observed rates. To compare the counts of falls reported in MDS data with fall incident reporting, positive agreement was examined using 2-by-2 contingency tables for each of the consecutive quarters.

## Findings

### Loading Process

The original data had records for 4,750 individual patients based on the study identifier count. Records for a total 4,519 patients (95.1 percent) made it into the final CDM database (Figure 1). The records for 48 patients were dropped because their study identifiers were not present in the population census file, indicating that the patients might not have had a billable stay. Another 36 patients appeared to have been erroneously assigned more than one study identifier during the de-identification process. The study identifiers for these patients were reassigned to be the same as patients that had exactly the same location, birthdate, gender, race, and marital status. The records for 147 patients were not loaded because their MDS or drug dispensing records had irreconcilable errors including the following:

- No valid observation periods (106 patients);
- MDS records indicating that entry into the SNF facility occurred after a discharge (9 patients);
- No gender indicated in the MDS data (7 patients); and
- Issues with their drug dispensing data (25 patients)
  - No drug dispensing data available or drug orders that fell outside of an observation period (14 patients); and
  - >3 drug records for which the end date was before the start date (11 patients).

Focusing on the dispensing orders, the original drug dispensing data indicated 430,910 dispensing occurrences. Of these, 40,950 (9.5 percent) were dropped for not providing a source drug code (i.e., GPI) that could be mapped to a standard vocabulary concept identifier. A visual inspection of the dropped records indicated that the great majority were for gastrointestinal aids (e.g., suppositories and

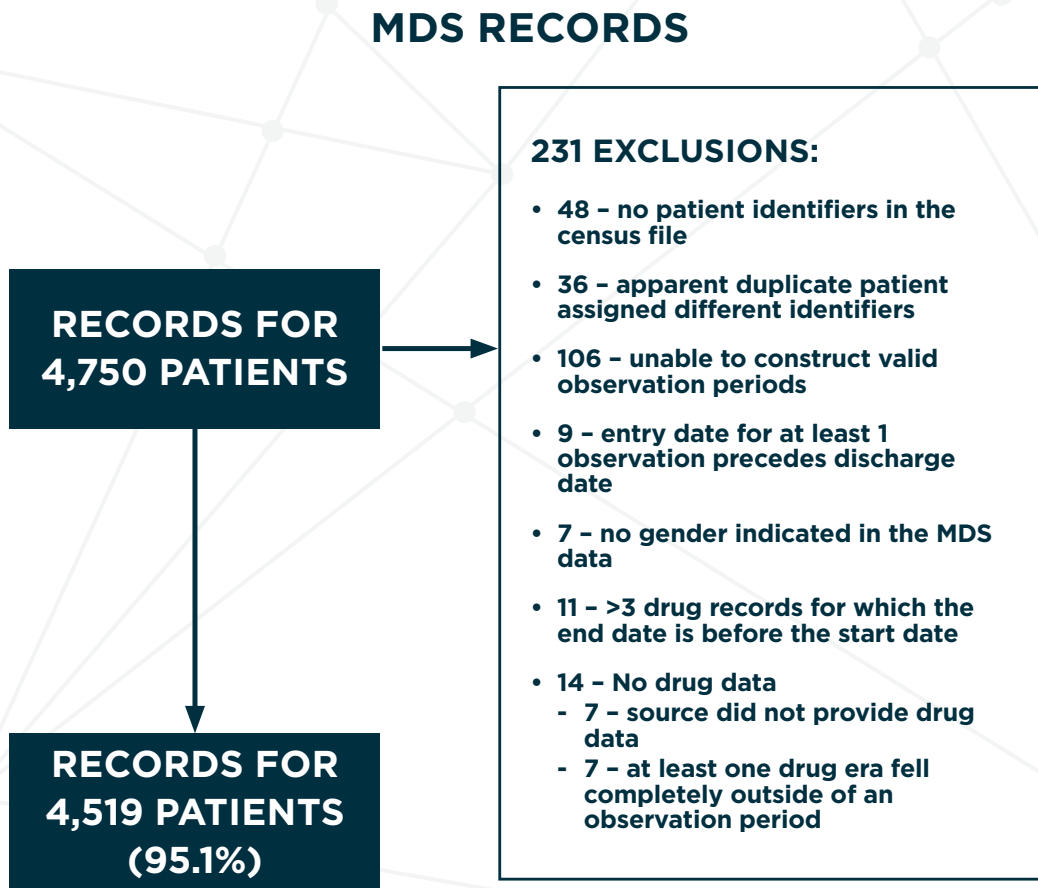


enema products), skin treatments (e.g., powders to prevent rash, spray bandages, etc.), and vitamins. The remaining dispensing records provided drug identifiers for 2,406 distinct drugs products that were mapped to 2,290 distinct drug concept identifiers in the standard vocabulary.

The source MDS 3 data contained 558 columns representing the full range of data fields specified by the standard. Limitations on the scope and resources for the project meant that not every available MDS data item could be translated to the CDM. Thus, those elements reported in the literature

as potentially relevant to studying medication safety and falls (e.g., mobility status, cognitive and functional status, and exposure to sedating and psychoactive medications) were given priority.<sup>27</sup> The translation process reorganized these columns into metadata (patient study ID, MDS report type, date of MDS report), demographics (age, gender, marital status, race, and SNF facility), 57 patient conditions, and 41 clinical observations. Table A1 (Appendix) provides the population characteristics for selected variables from all five SNF facilities included in the study.

**Figure 1. Flow Diagram Showing the Count and Reasons that Minimum Data Set 3.0 Records Were Dropped During the Loading Process**



## Quality Assurance

A summary of the ETL errors identified by Achilles Heel is shown in Table A2 (Appendix) along with an explanation of the steps we took to address them. The table shows that 10 types of errors were identified. These included the following: (1) issues with how codes from the standard vocabulary were used—two error types; (2) date patient data are collected falls outside a valid observation period—four error types; (3) drug exposure periods with invalid date values—three error types; and observation records with invalid values—one error type. Most errors were addressed by improvements to the ETL procedures. The issues that remained were not thought to affect future analyses.

## Drug Dispensing Comparisons

Table A3 (Appendix) shows a comparison between CDM drug dispensing data and medication administration records (MAR) with respect to the prevalence of exposure to six drug classes during the first week of each quarter: anticoagulants, antidepressants, antipsychotics, benzodiazepines, sedative hypnotics, and HMG-CoA reductase inhibitors (statins). An additional grouping includes drugs known to be associated with falls based in a recent systematic review.<sup>28</sup> All comparisons were for regularly scheduled drugs. We did not include PRN orders in the analysis because our initial testing found that dispensing data did not accurately capture such exposures. The number of quarters for which the comparison was done depended on the availability of electronic medication administration data from each facility, and ranged from 4 to 11 quarters—from the second quarter of 2011 through the fourth quarter of 2013. With the exception of sedatives, sufficient data were available to generate goodness of fit statistics across all facilities for most quarters.

In general, there were few statistically significant differences between CDM and MAR data in the prevalence of exposure to drugs within each class. Those differences that were identified were facility specific and did not exhibit a consistent pattern (Facility A for fall associated drugs, Facility C for benzodiazepines Facility D for antipsychotics, Facility E for antidepressants and statins). A visual comparison of the median absolute difference in percent prevalence with the median percent prevalence of exposure according MAR data suggests that a large difference in prevalence for one or two quarters underlies the differences in antipsychotic exposure for Facility D. No such evidence is apparent for the other identified differences.

## CMS-Reported “Nursing Home Compare” QMs Versus the CDM Data Set

Table A4 (Appendix) shows a comparison of the seven Nursing Home Compare QMs—between the CDM data set and data publically reported by CMS. For most measures, the comparison was done over 11 quarters, starting with the 2011 second quarter and ending with the 2013 fourth quarter. One exception was the two antipsychotic medication usage measures (i.e., exposure within seven days for short stay and long stay residents), for which data were available for only 7 quarters across all facilities. This was because there was an official change made by CMS to the MDS 3 in 2012 in how one of the MDS data fields for this measure was entered by nursing staff that was not accounted for when the data were pulled from the health system archive for our study. Also, data for Facility E were archived by the health system at a later date than that of the other facilities, which limited the analysis of all other QMs to only 8 quarters.

Sufficient data were available to generate goodness of fit statistics across all facilities for four QMs



(“Percent of [Long Stay] Residents Who Received an Antipsychotic Medication,” “Percent of [Long Stay] Residents Who Self-Report Moderate to Severe Pain,” “Percent of [Short Stay] Residents Who Self-Report Moderate to Severe Pain,” and “Percent of [Long Stay] Residents With a Urinary Tract Infection”). No statistically significant difference was found for any facility between CDM data and Nursing Home Compare data for three of these QMs. A statistically significant difference was identified for one facility for the QM “Percent of [Short Stay] Residents Who Self-Report Moderate to Severe Pain.”

There were three QMs for which there were insufficient data at least one facility (“The Percentage of [Short Stay] Residents Who Newly Received an Antipsychotic Medication,” “Percent of [Long Stay] Residents Who Have Depressive Symptoms,” and “Percent of [Long Stay] Residents Experiencing One or More Falls with Major Injury”). The health events monitored by these three QMs had a relatively low median prevalence ( $\leq 3$  percent) in those facilities for which the goodness of fit test could not be applied. The median absolute difference in percent prevalence for these QMs and facilities was generally close to the median percent prevalence of the health outcome monitored by the event. In the facilities where the goodness of fit test could be applied, the test generally indicated that the CDM data were significantly different from Nursing Home Compare data. The exception was for “Percent of [Long Stay] Residents Who Have Depressive Symptoms,” for which two of the three tested facilities did not appear to be significantly different from Nursing Home Compare.

### Comparison of MDS 3 Falls Data and Incident Reports

MDS fall data were available for all five facilities, while fall incident reports were available for four of the five

facilities. In general, many more falls were reported in MDS 3 data than in the incident report data. When comparing the percent of residents for which a fall was recorded each quarter, the MDS data recorded a median of 8.0–11.1 percent more short-stay residents, and 29.4–38.7 percent more long-stay residents. The 2-by-2 contingency tables showed that the majority of the falls recorded in incident reports were also recorded by MDS data (mean 78.4 percent, median 81.8 percent). This means that about 20 percent of falls recorded as incident reports might be events not captured by the MDS. However, the majority of falls recorded in MDS data were not identified in the fall incident data (mean 24.7 percent, median 26.5 percent).

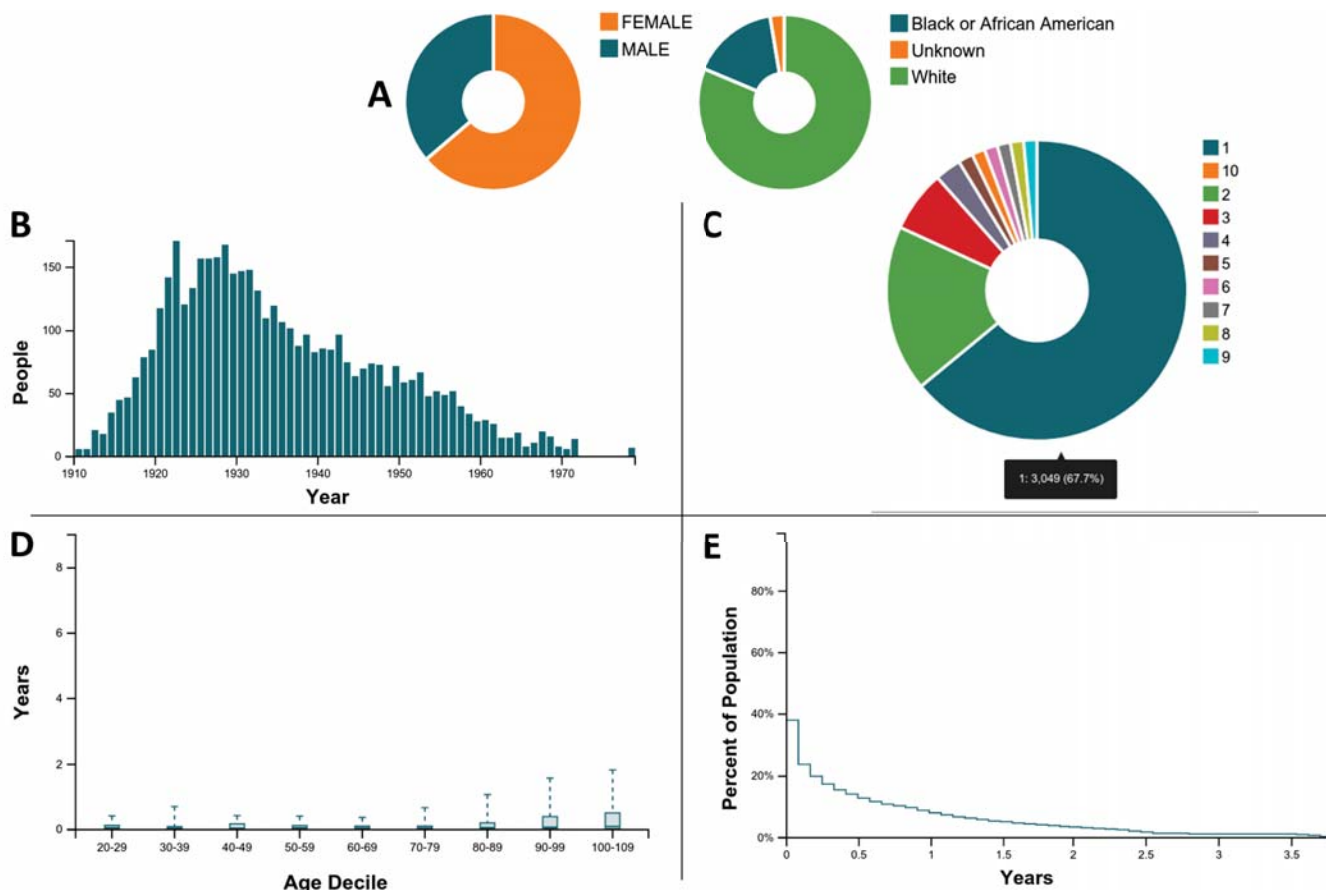
### The Final Data Set

Figure 2 is an overview of the CDM data provided by a subset of descriptive reports created using the OHDSI Achilles data characterization program. The panel of figures shows that the majority of the population is female, white, and born prior to 1940; more than half the population had only a single stay in a facility during the period covered by the data set indicated by the number of observation periods; older patients tended to have a longer observation period; and the majority of observation periods are less than a year.

### Lessons Learned and Implications for Research with the Data

The process of translating and loading data to a CDM format requires a substantial amount of work and creates the potential for errors to be introduced into the data. However, the OHDSI CDM provided the ability to use the Achilles and Achilles Heel data characterization tools, which were very helpful for identifying and diagnosing issues. We also found the process of comparing the QMs and medication exposure rates from the CDM data set with the Nursing Home Compare

**Figure 2. Overview of the CDM Data Provided by a Subset of Reports Created Using the OHDSI Achilles Data Characterization Program**



Notes: A: prevalence of gender and race\*; B: year of birth at first observation; C: number of observations per person; D: distribution of duration of observation by age decile; E: percent of population by cumulative observation period.

\*The "Unknown" category contains race/ethnic categories such as "Asian" and "Hispanic or Latino" that were coded in the source MDS data but were not translated to codes in the CDM. This will be corrected in future work.

program and the Pharmacy Services Provider very useful for exposing errors in the translation and loading process. While the number of drug classes and QMs we could validate against was small, their variety in terms of prevalence and difficulty help provide broader insight into the quality of the CDM data set. For example, early iterations found a significant difference in exposure rates between the CDM and MAR data with respect to stating an antidepressant exposure, while the same was not true for the anticoagulants and antipsychotics. This

was traced to some 17,000 dispensing orders that were not being loaded into the data set because of a programming error. The error significantly affected the statins and antidepressants because the rate of chronic exposure is significantly higher than for the other drug classes included in the analysis.

Drug exposures based on dispensing are generally accurate within a period of one week when compared with MAR data for both low and high prevalence exposures. This is good news for our



future studies examining drug safety where accurate capture of drug exposure is critical. However, some statistically significant differences will need to be accounted for. We plan to apply this knowledge to our research looking at opportunities to reduce fall risks by identifying and alerting about risky drug exposures by not including facility E in the analysis due to uncertainty about the accuracy of CDM data on antidepressant and statin exposure in the facility. Also, the study may exclude certain quarters from facilities A, C, and D where there is large variation between the CDM and MAR data.

The QM results suggest that the accuracy of observation and condition data in the CDM data set varies across both facilities and quarters. Since there is less certainty about the quality of the CDM data for lower prevalence QMs, any future studies focusing on low prevalence conditions and observations should be considered exploratory. Greater confidence seems warranted for future studies focusing on higher prevalence conditions and observations because the QMs with greater prevalence were generally accurate. The QM analysis also helped us identify that the MDS data we received from the clinical research informatics service were missing several drug exposure observations for all but seven quarters (see subsection “CMS-Reported ‘Nursing Home Compare’ QMs Versus the CDM Data Set”). We will account for this in the design of future studies while determining if the error can be corrected.

The comparison of MDS and fall incident data suggests that MDS data by itself captures the majority of fall events. However, a limitation of the MDS data is that the exact date and context of the fall is not available. Fall incident data may provide that additional information, and seem to be a potentially useful complement to MDS data by identifying some fall incidents that might not be recorded in the MDS.

### Implications for a Specific Study

The first study that we plan to use the new data set for is to develop a patient-level predictive model that predicts the probability of a fall for a nursing home patient who is prescribed a psychotropic and exposed to a potential drug-drug interaction. In our original protocol, we anticipated that data for approximately 832 residents will be used for model development based on the characteristics of the nursing homes to be included in the study and an expected 50 percent rate of psychotropic use in the nursing home. We can now more accurately determine that the minimum number of qualifying residents would be 2,109 because the rate of antidepressant use based on drug dispensing data loaded into the CDM is slightly more than 50 percent across the four facilities for which it was validated (Table 3). Our original estimate of fall events was 18.5 percent based on a sample of fall incident data for all residents across the facilities. We now know, because MDS fall data are more complete, that we underestimated the rate. Table 1 shows prevalence for other known predictors of falls (mobility status, cognitive and functional status, and exposure to sedating and psychoactive medications). These results will inform how to best divide the data into a training and test set and also how well the model will be statistically powered to detect promising candidate predictors.

### Limitations

While the drug exposure data seem to be of high quality in the new data set, the data set cannot capture medications that residents might take while outside the facility. Also, the drug exposure data validation did not include drugs that are prescribed “as needed.” The new nursing home data set is small compared with other observational data sets that might contain data from older adults such as large-scale claims data sets (e.g., Truven, Medicaid,

Marketscan). Depending on the outcome being studied, those data sets have the advantage of providing very large longitudinal population that allow for statistical adjustment on a large number of potential confounding factors. However, it is not clear how much coverage of the nursing home population is available in those data sets.

## Conclusions

We consider the new data set sufficiently validated to support a number of studies involving topics of clinical importance in the nursing home, provided that we account for the known limitations of the data set in the study designs. Our immediate focus will be on patient-level predictive modeling for falls risk for patients exposed to psychotropic drugs. Now that the data set is in the OHDSI CDM, we plan to explore the use of several other potentially useful tools and methods provided by the OHDSI community (see <http://www.ohdsi.org/analytic-tools/> and <https://github.com/ohdsi>). Another immediate benefit is the ability to participate in network research studies that originate from the collaborative. This should provide the ability to more easily examine differences between the nursing home population and other populations in clinically important topics such as treatment pathways and risks associated with drug exposures.

## Acknowledgments

This research was funded in part by the US National Institute on Aging (K01AG044433), the National Library of Medicine (R01LM011838), NIMH P30 MH90333, and the UPMC Endowment in Geriatric Psychiatry.

## References

- Dornin J, Ferguson-Rome JC, Castle NG. Nursing Facilities. (Chapter 11). In: *Long-Term Care In an Aging Society*. (in press). Morris Plains, NJ: Springer; 2015:295-313.
- Administration on Aging Administration for Community Living. *A Profile of Older Americans: 2014*. Washington DC: U.S. Department of Health and Human Services; 2014.
- Hardy S. Nursing Home Research. In: Newman AB, Cauley J, eds. *The Epidemiology of Aging*. 2012th ed. ; 2012.
- Magaziner J, German P, Zimmerman SI, et al. The prevalence of dementia in a statewide sample of new nursing home admissions aged 65 and older: diagnosis by expert panel. *Epidemiology of Dementia in Nursing Homes Research Group. Gerontologist*. 2000;40(6):663-672.
- Alzheimer's Association. *2010 Alzheimer's Disease Facts and Figures*. Alzheimer's Association; 2010. [http://www.alz.org/documents\\_custom/report\\_alzfactsfigures2010.pdf](http://www.alz.org/documents_custom/report_alzfactsfigures2010.pdf). Accessed March 7, 2011.
- OHDSI. Observational Health Data Sciences and Informatics. <http://www.ohdsi.org/>. Accessed May 15, 2015.
- CDC. *National Nursing Home Survey (NNHS)*. Centers for Disease Control and Prevention; 2004. <http://www.cdc.gov/nchs/nnhs.htm>.
- Duke University. National Long Term Care Survey (NLTC) Home Page. <http://www.nlctcs.aas.duke.edu/index.htm>. Published 2015. Accessed February 6, 2015.
- AHCA. What is OSCAR Data? [http://www.ahcancal.org/research\\_data/oscar\\_data/pages/whatisoscardata.aspx](http://www.ahcancal.org/research_data/oscar_data/pages/whatisoscardata.aspx). Published 2015. Accessed February 6, 2015.
- Saliiba D, Buchanan J. Development and validation of a revised nursing home assessment tool: MDS 3.0. *Rand Corporation Health, Santa Monica*. 2008.
- Centers for Medicare & Medicaid Services. MDS 3.0 RAI Manual. <http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/NursingHomeQualityInits/MDS30RAIManual.html>. Published 2015. Accessed May 15, 2015.
- Abramson EL, McGinnis S, Moore J, Kaushal R, HITEC investigators. A statewide assessment of electronic health record adoption and health information exchange among nursing homes. *Health Serv Res*. 2014;49(1 Pt 2):361-372. doi:10.1111/1475-6773.12137.
- Lang NM. The promise of simultaneous transformation of practice and research with the use of clinical information systems. *Nursing outlook*. 2008;56(5):232-236.
- Hook ML, Devine EC, Lang NM. Using a Computerized Fall Risk Assessment Process to Tailor Interventions in Acute Care. In: Henriksen K, Battles JB, Keyes MA, Grady ML, eds. *Advances in Patient Safety: New Directions and Alternative Approaches (Vol. 1: Assessment)*. Advances in Patient Safety. Rockville (MD): Agency for Healthcare Research and Quality; 2008. <http://www.ncbi.nlm.nih.gov/books/NBK43610/>. Accessed June 11, 2015.
- Chodosh J, Edelen MO, Buchanan JL, et al. Nursing home assessment of cognitive impairment: development and testing of a brief instrument of mental status. *J Am Geriatr Soc*. 2008;56(11):2069-2075. doi:10.1111/j.1532-5415.2008.01944.x.
- Katz S, Ford AB, Moskowitz RW, Jackson BA, Jaffe MW. Studies of illness in the aged: the index of ADL: A standardized measure of biological and psychosocial function. *JAMA*. 1963;185:914-919.
- Kroenke K, Spitzer RL, Williams JBW, Löwe B. The Patient Health Questionnaire Somatic, Anxiety, and Depressive Symptom Scales: a systematic review. *Gen Hosp Psychiatry*. 2010;32(4):345-359. doi:10.1016/j.genhosppsych.2010.03.006.



18. Wei LA, Fearing MA, Sternberg EJ, Inouye SK. The Confusion Assessment Method (CAM): A Systematic Review of Current Usage. *J Am Geriatr Soc*. 2008;56(5):823-830. doi:10.1111/j.1532-5415.2008.01674.x.
19. OHDSI. OHDSI/CommonDataModel. <https://github.com/OHDSI/CommonDataModel>. Published 2015. Accessed May 15, 2015.
20. OHDSI. Observational Health Data Sciences and Informatics Vocabulary Resources. <http://www.ohdsi.org/data-standardization/vocabulary-resources/>. Published 2015. Accessed July 9, 2015.
21. Liu S, Ma W, Moore R, Ganesan V, Nelson S. RxNorm: prescription for electronic drug information exchange. *IT Professional*. 2005;7(5):17-23. doi:10.1109/MITP.2005.122.
22. Michelle Dougherty, Jennie Harvell. *Opportunities for Engaging Long-Term and Post-Acute Care Providers in Health Information Exchange Activities: Exchanging Interoperable Patient Assessment Information*. Washington D.C.: Office of Disability, Aging and Long-Term Care Policy, U.S. Department of Health and Human Services; 2011. <http://aspe.hhs.gov/daltcp/reports/2011/StratEng.htm>. Accessed July 31, 2015.
23. Python Software Foundation. Welcome to Python.org. <https://www.python.org/>. Published 2015. Accessed May 15, 2015.
24. "RTI International." *MDS 3.0 Quality Measures User's Manual - v8.0*. Washington D.C.; 2013. <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/NursingHomeQualityInits/Downloads/MDS-30-QM-User%E2%80%99s-Manual-V80.pdf>. Accessed July 31, 2015.
25. OHDSI. ACHILLES for data characterization | OHDSI. 2015. <http://www.ohdsi.org/analytic-tools/achilles-for-data-characterization/>. Accessed May 15, 2015.
26. Centers for Medicare & Medicaid Services. Medicare.gov Nursing Home Compare. <http://www.medicare.gov/nursinghomecompare/search.html>. Published 2015. Accessed May 15, 2015.
27. Rubenstein LZ. Falls in older people: epidemiology, risk factors and strategies for prevention. *Age Ageing*. 2006;35 Suppl 2:ii37-ii41. doi:10.1093/ageing/af1084.
28. Sterke CS, Verhagen AP, van Beeck EF, van der Cammen TJM. The influence of drug use on fall incidents among nursing home residents: a systematic review. *Int Psychogeriatr*. 2008;20(5):890-910. doi:10.1017/S104161020800714X.



## Appendix

**Table A1. Population Characteristics for Selected Variables from All Five SNF Facilities Included in the Study**

	FACILITY A (N=553)	FACILITY B (N=1111)	FACILITY C (N=1408)	FACILITY D (N=1146)	FACILITY E (N=301)	ALL FACILITIES (N=4519)
Characteristic:	count / %	count / %	count / %	count / %	count / %	count / %
<b>DEMOGRAPHICS</b>						
Age*						
< 65	87 / 16%	238 / 21%	310 / 22%	217 / 19%	48 / 16%	900 / 20%
65-74	76 / 14%	182 / 16%	282 / 20%	190 / 17%	56 / 19%	786 / 17%
75-84	171 / 31%	311 / 28%	413 / 29%	322 / 28%	99 / 33%	1316 / 29%
85+	219 / 40%	380 / 34%	403 / 29%	417 / 36%	98 / 33%	1517 / 34%
Race						
White	440 / 80%	714 / 64%	1311 / 93%	922 / 81%	289 / 96%	3676 / 81%
Black	94 / 17%	357 / 32%	57 / 4%	212 / 19%	5 / 2%	725 / 16%
Other	19 / 3%	40 / 4%	40 / 3%	12 / 1%	7 / 2%	118 / 3%
Gender						
Female	360 / 65%	758 / 68%	846 / 60%	737 / 64%	179 / 60%	2880 / 64%
Male	193 / 35%	353 / 32%	562 / 40%	409 / 36%	122 / 41%	1633 / 36%
Marital Status*						
Married	156 / 28%	205 / 19%	369 / 26%	306 / 27%	90 / 30%	1126 / 25%
Divorced /Separated	48 / 9%	104 / 9%	102 / 7%	102 / 9%	41 / 14%	397 / 9%
Widowed	228 / 41%	257 / 23%	409 / 29%	420 / 37%	119 / 40%	1433 / 32%
Never married	112 / 20%	204 / 18%	156 / 11%	170 / 15%	39 / 13%	681 / 15%
Unknown	9 / 2%	341 / 31%	372 / 26%	148 / 13%	12 / 4%	882 / 20%
<b>MDS OBSERVATION (ANY TIME DURING A RESIDENT'S STAY)</b>						
Cognitive impairment (BIMS summary score $\leq$ 12)	106 / 19%	206 / 19%	258 / 18%	208 / 18%	82 / 27%	860 / 19%
Fall since admission (any stay)	177 / 32%	329 / 30%	335 / 24%	351 / 31%	139 / 46%	1331 / 30%
Impaired functional status (ADL summary score $\geq$ 16)	506 / 92%	958 / 86%	1310 / 93%	1025 / 89%	265 / 88%	4064 / 90%
Impaired transfer (ADL required extensive or total assistance or occurred only once or twice)	497 / 90%	897 / 81%	1283 / 91%	1017 / 89%	268 / 89%	3962 / 88%

Notes: \*Age and Marital status values were from the first admission or quarterly MDS report for a patient and might not reflect changes in status throughout a stay.



**Table A1. Population Characteristics for Selected Variables from All Five SNF Facilities Included in the Study (Cont'd)**

	<b>FACILITY A (N=553)</b>	<b>FACILITY B (N=1111)</b>	<b>FACILITY C (N=1408)</b>	<b>FACILITY D (N=1146)</b>	<b>FACILITY E (N=301)</b>	<b>ALL FACILITIES (N=4519)</b>
<b>Characteristic:</b>	<b>count / %</b>	<b>count / %</b>	<b>count / %</b>	<b>count / %</b>	<b>count / %</b>	<b>count / %</b>
<b>HEALTH STATUS</b>						
Alzheimer's	60 / 11%	146 / 13%	104 / 7%	176 / 15%	40 / 13%	526 / 12%
Anemia	222 / 40%	482 / 43%	415 / 29%	598 / 52%	165 / 55%	1882 / 42%
Anxiety	163 / 29%	292 / 26%	302 / 21%	322 / 28%	128 / 43%	1207 / 27%
Aphasia	11 / 2%	37 / 3%	24 / 2%	53 / 5%	22 / 7%	147 / 3%
Arthritis	218 / 39%	453 / 41%	249 / 18%	515 / 45%	84 / 28%	1519 / 34%
Arteriosclerotic heart disease	170 / 31%	290 / 26%	268 / 19%	335 / 29%	102 / 34%	1165 / 26%
Benign prostate hyperplasia	44 / 8%	76 / 7%	59 / 4%	119 / 10%	31 / 10%	329 / 7%
Cancer	73 / 13%	140 / 13%	102 / 7%	189 / 17%	20 / 7%	524 / 12%
Stroke	46 / 8%	158 / 14%	114 / 8%	273 / 24%	62 / 21%	653 / 14%
Constipation	50 / 9%	292 / 26%	160 / 11%	104 / 9%	9 / 3%	615 / 14%
Non-Alzheimer's dementia	165 / 30%	241 / 22%	218 / 15%	253 / 22%	144 / 48%	1021 / 23%
Depression	251 / 45%	461 / 41%	511 / 36%	495 / 43%	165 / 55%	1883 / 42%
Diabetes Mellitus	189 / 34%	395 / 36%	506 / 36%	427 / 37%	101 / 34%	1618 / 36%
Embolisms	10 / 2%	57 / 5%	23 / 2%	42 / 4%	12 / 4%	144 / 3%
COPD	138 / 25%	323 / 29%	336 / 24%	366 / 32%	85 / 28%	1248 / 28%
GERD or GI Ulcer	194 / 35%	322 / 29%	329 / 23%	386 / 34%	157 / 52%	1388 / 31%
Congestive heart failure	138 / 25%	300 / 27%	313 / 22%	340 / 30%	73 / 24%	1164 / 26%
Hemiplegia	39 / 7%	71 / 6%	41 / 3%	97 / 8%	17 / 6%	265 / 6%
Hypertension	421 / 76%	818 / 74%	979 / 70%	862 / 75%	235 / 78%	3315 / 74%
Hypotension	8 / 1%	21 / 2%	22 / 2%	24 / 2%	7 / 2%	82 / 2%
Hypoosmolar hyponatremia	5 / 1%	20 / 2%	15 / 1%	26 / 2%	13 / 4%	79 / 2%
Bipolar disorder	27 / 5%	61 / 5%	61 / 4%	28 / 2%	10 / 3%	187 / 4%
Infection due to resistant organism (MDRO)	21 / 4%	25 / 2%	108 / 8%	30 / 3%	16 / 5%	200 / 5%
Multiple sclerosis	6 / 1%	9 / 1%	12 / 1%	23 / 2%	4 / 1%	54 / 1%
Neurogenic bladder	17 / 3%	17 / 2%	19 / 1%	45 / 4%	8 / 3%	106 / 2%

**Table A1. Population Characteristics for Selected Variables from All Five SNF Facilities Included in the Study (Cont'd)**

	<b>FACILITY A (N=553)</b>	<b>FACILITY B (N=1111)</b>	<b>FACILITY C (N=1408)</b>	<b>FACILITY D (N=1146)</b>	<b>FACILITY E (N=301)</b>	<b>ALL FACILITIES (N=4519)</b>
<b>Characteristic:</b>	<b>count / %</b>	<b>count / %</b>	<b>count / %</b>	<b>count / %</b>	<b>count / %</b>	<b>count / %</b>
<b>HEALTH STATUS</b>						
Osteoporosis	102 / 18%	170 / 15%	93 / 7%	213 / 19%	72 / 24%	650 / 14%
Parkinson's disease	25 / 5%	50 / 5%	49 / 3%	63 / 5%	24 / 8%	211 / 5%
Pneumonia	67 / 12%	118 / 11%	109 / 8%	165 / 14%	28 / 9%	487 / 11%
Schizophrenia	5 / 1%	35 / 3%	14 / 1%	21 / 2%	10 / 3%	85 / 2%
Seizure	36 / 7%	90 / 8%	56 / 4%	107 / 9%	45 / 15%	334 / 7%
Septicemia	16 / 3%	6 / 1%	22 / 2%	22 / 2%	5 / 2%	71 / 2%
Thyroid disorder	118 / 21%	184 / 17%	231 / 16%	262 / 23%	78 / 26%	873 / 19%
UTI	116 / 21%	130 / 12%	179 / 13%	235 / 21%	61 / 20%	721 / 16%
Open wound without complication	25 / 5%	32 / 3%	32 / 2%	56 / 5%	6 / 2%	151 / 3%
Psychosis	60 / 11%	57 / 5%	30 / 2%	82 / 7%	139 / 46%	368 / 8%



**Table A2. Achilles Heel Errors Identified and How Addressed**

ERROR TYPE(S)	HEEL-IDENTIFIED ERROR(S)	HOW ADDRESSED
Issues with how codes from the standard vocabulary were used	Number of persons with at least one procedure occurrence, by procedure_concept_id; 2 concepts in data are not in correct vocabulary (CPT4/HCP/ICD9P)	The only “procedures” in our data set were MDS reports that we consider administrative procedures. These break down into those done at admission and those done for existing patients. Our review of CPT4/HCP/ICD9P did not find any relevant codes for these two types of procedures. Therefore, we ignored this error and continued to use the two SNOMED CT codes we had chosen (108221006: “Evaluation AND/OR management - established patient” and 108220007: “Evaluation AND/OR management - new patient”).
Issues with how codes from the standard vocabulary were used	Number of persons with at least one observation occurrence, by observation_concept_id; 1 concepts in data are not in correct vocabulary (LOINC)	This was an expected issue because we coded fall incidents using the OHDSI concept identifier for the MedDRA preferred term for “Fall” since no appropriate LOINC code could be found.
Date patient data are collected falls outside a valid observation period	Number of procedure occurrence records outside valid observation period; count (n=9,767) should not be > 0	Our original algorithm for generating observation periods from MDS data did not correctly implement all the business rules provided by the MDS 3 Quality Measures user manual. <sup>24</sup> We also identified what appeared to be test MDS records that were timed outside of a patient’s observation periods and provided no data. We corrected these issues and reloaded the procedure occurrence records.
Date patient data are collected falls outside a valid observation period; drug exposure periods with invalid date values	<p>Number of drug exposure records outside valid observation period; count (n=2,195) should not be &gt; 0</p> <p>Number of drug eras outside valid observation period; count (n=2,195) should not be &gt; 0</p> <p>Number of drug exposure records with end date &lt; start date; count (n=179) should not be &gt; 0</p> <p>Distribution of drug era length, by drug_concept_id should not be negative</p>	The reasons for these errors were the same as those that caused the procedure occurrence records outside a valid observation period (see above). Once those issues were corrected, we changed the load procedure to check that a drug order fell within an observation period before adding the records to the Drug Exposure and Drug Era tables.

Table A2. Achilles Heel Errors Identified and How Addressed (Cont'd)

ERROR TYPE(S)	HEEL-IDENTIFIED ERROR(S)	HOW ADDRESSED
Date patient data are collected falls outside a valid observation period	Number of observation records outside valid observation period; count (n=77,898) should not be > 0	The same error that caused the procedure occurrence records to fall outside a valid observation period (see above) was also responsible for most counts of this error. Changing the load procedure to check that an observation fell within an observation period before adding the records to the Observation table corrected all but 26,209 cases. Analysis of these cases found that these occur in two cases where we had intentionally dropped records: (1) when a patient had a single day stay, or (2) a patient's only MDS 3 records occurred before the study start date. We took no action on these remaining cases because all analyses should use correct observation periods to identify valid exposures and observations.
Observation records with invalid values	Number of observation records with no value (numeric, string, or concept); count (n=162,271) should not be > 0	This error exposed a bug in the translation and loading procedure whereby data from validated MDS scales (e.g., the BIMS) were not being loaded properly. Correcting this issue removed the problem.
Drug exposure periods with invalid date values	Number of drug eras with end date < start date; count (n=179) should not be > 0	There were two causes for this error: (1) the source data had a small number of drug records with incorrect start and end dates; and (2) the same issues with the business rules for generating observation periods that affected procedure, drug, and observations above caused the code creating drug eras to create erroneous drug eras for some patients. Both issues were addressed and this error was no longer triggered.



**Table A3. Comparison of CDM Drug Dispensing Data to Medication Administration Records with Respect to the Prevalence of Exposure to Specific Drugs During the First Week of Each Quarter**

FACILITY	NUM QTRRS	ANTICOAGULANTS					ANTIPSYCHOTICS					ANTIDEPRESSANTS				
		PSP PREV (%)	MIN	MED	MAX	X <sup>2</sup> (df)	PSP PREV (%)	MIN	MED	MAX	X <sup>2</sup> (df)	PSP PREV (%)	MIN	MED	MAX	X <sup>2</sup> (df)
A	11	8.6	1.0	3.3	4.6	4.3(9)	31.9	0.0	9.7	17.0	18.0(11)	50.6	0.4	5.2	8.0	4.1(11)
B	9	15.6	0.3	0.5	2.5	0.7(9)	20.8	1.5	4.2	10.0	9.8(9)	56.6	0.5	3.4	5.0	2.6(9)
C	4	19.8	0.3	0.55	1.8	0.2(4)	17.1	6.3	6.6	8.9	9.1(4)	53.6	0.2	1.0	3.9	0.6(4)
D	8	16.3	0.0	0.7	2.3	0.7(8)	14.6	0.0	0.6	2.8	18(8)*	52.7	0.5	1.5	2.3	0.6(8)
E	6	11.1	0.2	0.7	1.2	0.2(5)	38.8	0.0	2.8	8.9	3.6(5)	54.8	10.3	12.1	13.2	17(5)*

FACILITY	NUM QTRRS	BENZODIAZEPINES					FALL ASSOCIATED DRUGS					SEDATIVES				
		PSP PREV (%)	MIN	MED	MAX	X <sup>2</sup> (df)	PSP PREV (%)	MIN	MED	MAX	X <sup>2</sup> (df)	PSP PREV (%)	MIN	MED	MAX	X <sup>2</sup> (df)
A	11	10.1	1.2	3.8	9.7	5.4(9)	25.0	4.0	9.1	16.8	20.4(11)*	<0.1	1.0	2.0	6.4	-
B	9	6.0	0.8	3.6	4.8	12.1(9)	29.0	1.3	7.2	9.1	13.7(9)	0.1	0.7	1.4	6.8	-
C	4	12.9	2.3	6.7	8.8	9.7(4)*	40.7	2.2	5.4	5.8	2.8(4)	<0.1	4.3	5.9	6.4	-
D	8	11.3	0.6	2.2	3.5	3.6(8)	36.5	0.2	1.7	2.7	0.8(8)	<0.0	2.5	4.8	6.2	1.3(1)
E	6	21.6	0.6	1.9	2.6	0.7(5)	48.6	3.2	4.5	6.1	2.7(5)	0.1	0.0	0.8	1.1	-

FACILITY	NUM QTRRS	STATINS				
		PSP PREV (%)	MIN	MED	MAX	X <sup>2</sup> (df)
A	11	32.1	0.2	1.8	13.1	3.1(11)
B	9	26.2	0.4	1.2	5.5	1.5(9)
C	4	35.1	1.7	2.8	3.7	1.1(4)
D	8	34.6	0.2	1.1	2.4	0.5(8)
E	6	23.2	8.6	9.4	10.7	18.4(5)*

Notes: The data shown are the following: (1) median percent prevalence of administration of the drugs within class according to the pharmacy services provider (PSP) across the measurable quarters; and (2) minimum, median, and maximum absolute difference in percentage prevalence of exposure between CDM and PSP across all quarters for which electronic MAR data were available in the facilities. Also shown are the results of an X<sup>2</sup> goodness of fit test with the number of quarters for which sufficient data were available (df). \*P ≤ 0.05; \*\* P ≤ 0.01.

Table A4. Comparison of Selected Nursing Home Compare (NHC) Quality Measures Between the CDM Data Set and Data that Were Publicly Reported by the Centers for Medicare and Medicaid Services

FAC.	PERCENT OF [LONG STAY] RESIDENTS WHO RECEIVED AN ANTIPSYCHOTIC MEDICATION (419; N031.02)					THE PERCENTAGE OF [SHORT STAY] RESIDENTS WHO NEWLY RECEIVED AN ANTIPSYCHOTIC MEDICATION (434; N011.01)				
	NHC PREV. (%)	MIN	MED	MAX	X <sup>2</sup> (df)	NHC PREV. (%)	MIN	MED	MAX	X <sup>2</sup> (df)
A	31.3	0.3	3.4	7.8	4(7)	2.9	0.3	1.8	7.8	-
B	18.8	0.5	2.1	5.1	3.7(7)	3.1	0.3	4.1	7.5	80.4(5)**
C	16.3	0.0	2.7	4.5	4.5(7)	3.0	0.5	1.2	5.4	23.4(6)**
D	13.4	0.0	0.6	2.8	1.1(7)	1.0	2.0	4.0	4.7	52.1(3)**
E	34.6	0.6	3.2	7.5	4.0(7)	2.7	0.0	2.0	8.0	-

FAC.	PERCENT OF [LONG STAY] RESIDENTS WHO HAVE DEPRESSIVE SYMPTOMS (408; N030.01)					PERCENT OF [LONG STAY] RESIDENTS EXPERIENCING ONE OR MORE FALLS WITH MAJOR INJURY (410; N013.01)				
	NHC PREV. (%)	MIN	MED	MAX	X <sup>2</sup> (df)	NHC PREV. (%)	MIN	MED	MAX	X <sup>2</sup> (df)
A	2.1	0.0	1.8	4.3	-	2.8	0.0	1.3	3.9	-
B	1.0	0.0	0.8	1.6	0.3(2)	1.7	0.0	0.7	3.1	8.1(2)*
C	0	0.0	0.9	1.8	-	2.1	0.7	0.9	3.6	27.5(7)**
D	2.5	0.0	0.6	4.9	12.2(3)**	2.2	0.8	2.0	5.1	32.8(10)**
E	11.7	0.1	1.35	6.2	5.1(7)	5.2	0.5	1.7	5.3	2.7(3)

FAC.	PERCENT OF [LONG STAY] RESIDENTS WHO SELF-REPORT MODERATE TO SEVERE PAIN (402; N014.01)					PERCENT OF [SHORT STAY] RESIDENTS WHO SELF-REPORT MODERATE TO SEVERE PAIN (424; N001.01)				
	NHC PREV. (%)	MIN	MED	MAX	X <sup>2</sup> (df)	NHC PREV. (%)	MIN	MED	MAX	X <sup>2</sup> (df)
A	20.5	0.6	3.3	5.8	5.3(11)	31.7	0.4	4.7	19.5	6.2(10)
B	22.1	0.3	1.6	6.2	6.2(11)	39.4	0.6	4.8	8.9	15.3(11)
C	14.1	0.8	2.9	7.1	14.9(11)	20.5	2.2	4.9	17.2	60.7(11)**
D	20.9	0.3	2.6	8.9	12.5(11)	32.9	0.0	2.2	6.4	7.3(11)
E	19.7	0.3	2.7	14.5	3.3(8)	29.9	0.0	6.5	21.5	7.5(7)

FAC.	PERCENT OF [LONG STAY] RESIDENTS WITH A URINARY TRACT INFECTION (407; N011.01)				
	NHC PREV. (%)	MIN	MED	MAX	X <sup>2</sup> (df)
A	4.3	0.3	2.5	5.1	7.4(3)
B	1.8	0.0	0.7	1.9	2.6(2)
C	4.0	0.1	1.2	5.1	3.9(5)
D	3.6	0.3	0.7	1.8	1.6(6)
E	4.8	0.5	1.9	3.3	1.5(2)

Notes: The comparison covers 11 quarters, from the second quarter of 2011 through the fourth quarter of 2013—except for the two antipsychotic medication measures, for which data were available for only 7 quarters across all facilities, and all other measures for Facility E for which MDS 3 data were available for 8 quarters. The data shown are the following: (1) the median percent prevalence of the measure according to NHC across the measurable quarters, (2) the minimum, median, and maximum absolute difference in percent prevalence across all quarters for which electronic MAR data were available in the facilities; and (3) the results of a X<sup>2</sup> goodness of fit test with the number of quarters for which sufficient data were available (df). \*P ≤ 0.05; \*\* P ≤ 0.01.