MATHEMATICA
CENTER FOR
STUDYING DISABILITY POLICY

DRC Working PAPER

WORKING PAPER NUMBER: 2020-01

A New Approach to Analyzing Opioid Use among SSDI Applicants

January 2020

April Yanyuan Wu*
Peter Mariani
Jia Pu
Andrew Hurwitz

***Corresponding author's contact information:**
Mathematica
111 East Wacker Drive, Suite 920
Chicago, IL 60601
Email: awu@mathematica-mpr.com
Reference Number: 40112.D-MP-15-07

The research reported herein was performed pursuant to a grant from the U.S. Social Security Administration (SSA) that is funded as part of the Disability Research Consortium. The opinions and conclusions expressed are solely those of the authors and do not represent the opinions or policy of SSA or any agency of the federal government. Neither the U.S. government nor any agencies thereof, nor any of their employees, make any warranty, expressed or implied, or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of the contents of this report. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply endorsement, recommendation, or favoring by the U.S. government or any agency thereof.

This page has been left blank for double-sided copying.

ACKNOWLEDGMENTS

The authors wish to thank Jody Schimmel Hyde, Denise Hoffman, and Gina Livermore of Mathematica's Center for Studying Disability Policy for their insightful comments on the draft manuscript. They thank Fei Xing, also of Mathematica, for his comments with the machine-learning method. The authors acknowledge Paul O'Leary of the Social Security Administration (SSA) for his assistance with data from the Structured Data Repository, and Olivier Bodenreider of the National Library of Medicine, National Institutes of Health for his help with the use of RxNorm.

The research reported herein was performed pursuant to an SSA grant that was funded as part of the Disability Research Consortium. The opinions and conclusions expressed are solely those of the authors and do not represent the opinions or policy of SSA or any agency of the federal government. Neither the U.S. government nor any agencies thereof, nor any of their employees, make any warranty, expressed or implied, or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of the contents of this report. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply endorsement, recommendation, or favoring by the U.S. government or any agency thereof.

This page has been left blank for double-sided copying.

CONTENTS

This page has been left blank for double-sided copying.

TABLES

FIGURES

This page has been left blank for double-sided copying.

ABSTRACT

Project Number

MPR 15-07

Title

A New Approach to Analyzing Opioid Use among SSDI Applicants

Author

April Yanyuan Wu, Peter Mariani, Jia Pu, and Andrew Hurwitz

Date

June 2019

Key findings and policy implications

The rising prevalence of opioid use nationwide coupled with the large share of Social Security Disability Insurance (SSDI) applicants with conditions associated with opioid use, such as musculoskeletal conditions, suggests that opioid use may be common and increasing among SSDI applicants. Although applicants cannot qualify for SSDI solely on the basis of drug addiction, in some cases opioid use may exacerbate the effects of other conditions that meet the SSDI medical criteria. Understanding the pattern of opioid use among SSDI applicants has important implications for projecting resources needed to adjudicate SSDI applications and for the size and composition of the SSDI caseload.

To date, little has been documented about opioid use among SSDI applicants. This largely reflects issues surrounding data availability. Since 2007, virtually all SSDI applications have been collected and stored electronically, and SSDI applicants provide information about their medication use. That medication data is recorded as a combination of categorical (pull-down menu) and open-ended text fields, making analysis difficult without using state-of-the-art methods to capture the data in text fields.

This proof of concept study assessed whether machine learning can be used to classify free-form text of medication information. To conduct our analysis, we used the medication information—both categorical and free-form text—from a random sample of 100,000 applicants in 2013, based on data in the Social Security Administration's (SSA) Structured Data Repository (SDR). Specifically, this study capitalized on a supervised machine-learning algorithm using natural language processing to identify opioids recorded in free-form text. We then aligned that information with the categorical opioid data in the SDR for the same applicants. To the extent that this algorithm works based on established criteria in the field, it can potentially be used by SSA to analyze trends of drug use among SSDI applicants.

We applied the machine-learning algorithm to medication records of our sample, documented the reliability of our model, and produced summary statistics of opioid use at the time of the SSDI application. Because our sample only accounts for about five percent of all

applicants in 2013, and includes only those applicants who reported at least one medication, the findings presented here may not be generalizable to the full SSDI applicant population.

We find the following:

- 61 percent of drug records in our sample were recorded in free-text fields and 45 percent of our sample applicants reported medications exclusively via free-text fields.
- All of the 20 most commonly reported drug names from free-text fields are reflected in the pull-down list.
- Among the SSDI applicants who reported using any medications at the time of application, 35 percent reported use of one or more opioids. Among the subset that reported opioid use, 26 percent reported using two or more opioids.
- The most frequently reported opioids include Hydrocodone, Tramadol, Oxycodone, Percocet, and Vicodin.

The prevalence of opioid use among our sample of SSDI applicants suggests that our approach may lend itself to future work considering trends in opioid use, assessing the effects of efforts to reduce opioid addiction among injured workers, and understanding the relationship between opioid use and SSDI applications and awards. The implications of our findings include the following:

- Our results show the importance of using both medications recorded in the SDR pull-down list and free-form text to study drug use among SSDI applicants. Research that includes only medications recorded in the SDR pull-down list will omit a notable portion of all reported medications.
- While the pull-down list includes a comprehensive list of drug names, applicants are more likely to use the free-form text to record their medications. Applicants may prefer the free-form text because it enables them to be more specific when listing their medications. This suggests a shortcoming in the pull-down list and points to a potential improvement to the application form.
- The relatively high prevalence of opioid use among SSDI applicants with at least one medication suggests that opioid use may indeed be high among all SSDI applicants. The nature of our sample does not permit us to estimate opioid use among the general applicant population but future research using our machine-learning approach could provide that information.
- The SDR medication data, while rich, only offer a point-in-time look at opioid use and cannot distinguish appropriate use from misuse. Our findings should be interpreted with that caveat in mind.

I. INTRODUCTION

Since the 1990s, there has been a marked increase in medical and nonmedical opioid use, opioid addiction, and opioid overdose deaths (Kolodny et al. 2015). In 2016, 19 percent of the United States population filled at least one opioid prescription (Mytelka et al. 2018). Prescription opioid use is even higher among Social Security Disability Insurance (SSDI) beneficiaries who have completed the two-year Medicare waiting period: 44 percent of these beneficiaries under age 65 filled at least one opioid prescription in 2011, and almost one in four filled six or more opioid prescriptions that same year (Morden et al. 2014).

Opioid use may also be common and increasing among SSDI applicants. Between 2007 and 2016, there has been a rise in SSDI awards made to applicants with musculoskeletal conditions (Social Security Administration [SSA] 2017), which are often associated with opioid use (Carnide et al. 2017). Moreover, a recent analysis of state-level data suggests that the wider availability of opioids is linked to increases in SSDI applications (Cutler et al. 2017). However, the extent to which opioid use among SSDI applicants mirror rates among SSDI awardees with Medicare coverage is unclear because SSDI applicants may not necessarily have a reliable source of health insurance coverage.

Understanding the rates and patterns of opioid use among SSDI applicants supports SSA in identifying the health care needs and utilization of this group and can aid in identifying ways to help. For example, this information may help identify opportunities for work-related early intervention programs. In addition, although addiction to opioids does not make a worker eligible for SSDI, per se, the worker may be eligible because of the condition that causes the pain. Hence, a significant portion of applicants are likely to become future SSDI awardees. Yet,

statistics on opioid use among SSDI applicants, who are generally not eligible for Medicare, are not available.

This study lays the groundwork for addressing this knowledge gap by demonstrating the ability to use SSA administrative data to assess opioid use among SSDI applicants. Although SSDI applicants are required to report medication use on their application, they often do so in an open-ended text field, making this information difficult to use in research. In the past, using data from open-ended text fields required the training of coders and manual coding, a process that is labor intensive, time consuming, prone to error, and usually cost prohibitive.

This study is a proof of concept study for the proposition that machine learning can be used to classify free-form text of medication information in SSA's Structured Data Repository (SDR). Specifically, the study capitalizes on a supervised machine-learning algorithm that was developed to classify free-form text about occupations in the SDR into occupation codes (Wu 2018). For this study, we apply a similar technique to identify opioids recorded in free-form text and combine that information to appear on a pull-down menu, which is easier to access for research. If successful, this method could be used more broadly to support research on prescription drug use among applicants.

Using a random sample of 100,000 SSDI applicants in 2013 with at least one reported medication, we successfully applied the machine-learning technique to identify drug and opioid use at the time of the application. We found that on average, each applicant in our sample has 5.18 medication records. Nearly half (45 percent) exclusively used free-form text to report medications, highlighting the importance of analyzing these data when using the SDR as a source of information on medications. We found that 35 percent of applicants in our sample

reported any opioid use at the time of application and about 26 percent of them reported using more than one opioid.

In Chapter II of this report, we discuss the data, and in Chapter III, we describe our methods. Chapter IV summarizes the medication reports and the statistics of opioid use among our sample. Chapter V concludes with a discussion on the limitation of the study and the implications of the findings.

This page has been left blank for double-sided copying.

II. DATA

For over a decade, SSA has stored data collected from SSDI applicants in the SDR. The SDR contains data from virtually all SSDI and Supplemental Security Income (SSI) applicants. The medication information in the SDR comes from Section 7 of application form 3368. In this section, the applicants are asked whether they are taking any medicine at the time of application. If they respond "Yes," they are asked to report medications either by choosing the drug names from a pull-down list of 630 medications or by noting drug names in an open-text field. The form instructions recommend that applicants refer to the labels on their medicine containers to ensure the accuracy of the information. The applicants are also asked to record the reasons for taking the medicine.

This analysis represents a proof-of-concept study and results are not intended to be generalizable to all applicants. For efficiency, we selected a random sample of 100,000 SSDI applicants (including concurrent SSDI-SSI applicants) from 2013 who met certain selection criteria; this sample accounts for 4.5 percent of all applicants in that year. The sample includes initial-level applications with final decisions, and includes only the first such application if someone had more than one application in the year. The sample is restricted to applicants ages 18 to 67 and who reported at least one medication at the time of the application. Because we focus only on applicants who reported at least one medication and those with initial-level final decisions, we cannot estimate opioid use prevalence more generally among the SSDI applicant population.

There are 517,624 reported medication records for the 100,000 applicants in our analysis sample.[1] On average, each applicant in our sample reported using more than five medications, with 25 percent reporting using two or fewer drugs, 50 percent reporting using four or fewer drugs, and 90 percent reporting using 10 or fewer drugs.

Applicants in our sample reported medications through a combination of the pull-down menu and free-form text. About 39 percent of drug records were selected from the pull-down menu, while the rest (61 percent) were recorded in free-form text. At the applicant level, we found that 22 percent reported medications exclusively via the pull-down list, 45 percent of applicants exclusively used free-text fields, and 37 percent used a combination of the pull-down list and free-text fields.

[1] The data file we received includes 538,749 drug records, however there are some individual-level duplicates of drug entries. We consulted with SSA staff and they advised us to delete the duplicates from the data. The de-duplicated file includes 517,624 drug records.

III. METHODS

A. Overview

The project team used a supervised machine-learning algorithm based on natural language processing (NLP) to identify the opioids reported in free-text fields and pull-down list by SSDI applicants. This process had several steps. First, we developed a labeled file where our subject matter expert manually classified a random sample of medication records from our data into opioid versus non-opioid. Then we randomly divided the labeled data into training and test sets. We also identified other resources, including a comprehensive list of opioids, to facilitate the classification process. Next, we used the training set along with other identified resources to develop the machine-learning algorithm. The machine-learning algorithm is based on NLP and implemented in Python. We used the learning algorithms to train the computer to recognize patterns of classification from the matches of text entries to their corresponding categories. Once the computer program was trained, we applied it to the test set for purposes of performance assessment at the end of development. After testing and adjustments, we used the program to code the remaining unlabeled records without further training.

B. Development of training and testing files

A supervised machine-learning approach starts with a subject matter expert to develop a labeled file where observations are labeled to their corresponding categories. For this project, we enlisted the support of a pharmacist on staff at Mathematica to serve as subject matter expert. The subject matter expert labeled a random sample of 2,000 records from both the pull-down list and free-form text responses as opioid versus non-opioid drugs. These labeled data served as the gold standard for our coding effort. We then randomly divided the labeled data into training and test sets. The training set was used to develop the machine-learning algorithm, and the test set was used to assess the algorithm' performance after it had been fully specified.

C. Identifying other resources to classify opioids

According to the literature on drug classification, there is no single, standard drug coding scheme, and code sets can be inconsistent across drug classification systems (DeFalco et al. 2012). For this project, based on the literature and recommendations from our subject matter expert, we used the combination of the Center for Disease Control and Prevention's (CDC) Oral Morphine Milligram Equivalents (MME) file and the National Library of Medicine's (NLM) resource RxNorm,[2] along with the manually coded training file described previously as our gold standard to identify opioids and develop the machine-learning algorithm.

The CDC MME file contains a comprehensive listing of medications at the National Drug Code (NDC) level to help with analyzing prescription data to prevent prescription drug misuse, abuse, and overdose. The file includes a list of opioids with their NDC code (a unique identifier for drugs), product name, generic drug name, drug substance, duration of action (long-acting versus short-acting), drug class code, strength per unit, unit of measurement, and MME conversion factor.

The NLM's RxNorm provides normalized names for clinical drugs and links those names to many of the drug vocabularies commonly used in pharmacy management and drug interaction software. An advantage of RxNorm is that it provides a single system for unambiguously identifying brand-name and generic drugs. This feature is a response to the proliferation of drug identification and classification systems among hospitals, clinics, pharmacies, health systems, manufacturers, and payers—all of which might use an array of different names for the same drug, making it difficult to extract meaningful information and communicate across different systems and databases, as well as to identify the corresponding class of drugs. This feature has a

[2] This study used publicly available data from the NLM; NLM is not responsible for the product and does not endorse or recommend this or any other product.

substantial advantage for our opioid classification. Another advantage of RxNorm is that it specifically supports NLP use and offers a web-based application programming interface (API) for easy access to data.

D. Machine-learning algorithm development

Our machine-learning algorithm is based on NLP, which is a subfield of artificial intelligence concerned with the interactions between computers and human (natural) languages. Identifying opioids from free-text fields can be difficult and the bulk of this section focuses on our approach to classifying those entries. One reason this process is challenging is that when applicants enter medication names in open-text fields, they often also enter other information related to drug use that is not part of the drug name. For example, there are entries such as "vitamin e 400 units 1x daily" and "Dilantin 400 mg at bedtime." Applicants may also enter other invalid free-text entries that include but are not limited to the following: to be provided, unknown, asa, ?, something, have about 20 meds, numerous-unknown kinds, other in the past, not listed on 3441, over the counter meds, see medical records, see paper 3368, will submit info, and a pain killer (don't know). Further, misspelling of the drug names are very common among open-text entries.

Based on the features of these free-text entries, we developed the following three-step procedure to identify opioids from free-text fields: 1) break the free-text entry into words; 2) remove words unlikely to be drugs; 3) compare the remaining words to lists of known opioids and non-opioids to classify the entry as an opioid or non-opioid.[3]

The first step was to break the free-text entry into words. This process is called word-tokenization in the NLP literature and it has some subtleties, particularly in our application. For

[3] The detailed description of our method is available from the authors upon request.

example, it is difficult to distinguish between symbols that are a part of a medication name (for example, PC-CAP) and those that applicants use to separate lists of drugs or features (for example, Morphine-30mg/2xday). However, failing to recognize that Morphine is a unique word in "Morphine-30mg/2xday" can make it hard to identify Morphine from free-form text, while breaking "PC-CAP" into "PC" and "CAP" will lead to misclassification of the text "PC-CAP" as a non-opioid, even non-drug. In this step, we converted all text to lowercase, removed extra whitespace, removed the "w/" abbreviation for "with" and removed all other characters that aren't numbers, letters, or periods. We then applied multi-word expression tokenization, which allows consecutive tokens to be grouped if they are a member of a specific input set based CDC MME file.

The second step to identify opioids from free-text fields was to remove words unlikely to be drugs from the free-text fields. We first used simple regular expression parsing to delete tokens that start with any non-letter value because all opioid drugs begin with a letter. We then removed conversational English words such as: doctor, daily, day, know, because, remember, ill, and find out. Finally, we removed words that are common in our application, but are neither drugs nor common English words, such as: mg(s), med(s), tabs, and per. We manually looked through the top 150 most frequent tokens in our training set, and removed these non-drug words and their lexical variants.

The third step of our procedure involved matching the remaining words from free-text entries (which are suspected drugs) to lists of known opioids and non-opioids based on the features of the remaining words, to classify the entry as an opioid or non-opioid. This step was based on comparisons to both the CDC MME file and the RxNorm file. To use the CDC MME file, we computed the Damerau-Levenshtein (DL) distance between post-cleaning free-text

entries and the "opioid words" from the CDC MME file to determine the match. The Levenshtein distance represents the number of insertions, deletions, and substitutions required to change one word to another. A modification of Levenshtein distance, DL distance counts transpositions as a single edit. For example, the DL distance between identical entries such as "fish" and "fish" is 0, but "fish" and "ifsh" equals 1. The intuition behind this feature is that if the DL distance is 0, then the free-text entry exactly matches a specific opioid in the CDC MME file, and if the DL distance is small, then the free-text entry is likely be a typographical error of an opioid of the CDC MME file.

We conducted testing to determine the most appropriate DL distance threshold to classify opioids. This involved testing different thresholds of DL distance, including values 1, 2, and 3. We found that while increasing the DL distance leads to an increase in the likelihood of being able to correctly identify an opioid entry, it also increases the likelihood of misclassifying a non-opioid as an opioid. As a result, the overall accuracy rate decreases with the increase of the DL distance. Based on the testing results, we set the DL distance less than or equal to one as the threshold to determine a match.

We also used RxNorm to identify opioids. If a free-text entry was identical to an opioid classified in RxNorm's drug classification systems, we classified the entry as an opioid. If an entry was identical to a non-opioid in RxNorm's drug classification systems, we classified the entry as a non-opioid. A unique feature of RxNorm is that it suggests a spelling substitution when an entry does not match any normalized drug names in its system. When RxNorm suggests an approximate match to an opioid or other drug class for a free-text entry of our data, we rely on the approximate match algorithm of RxNorm to identify the opioid.[4]

[4] The approximate match algorithm developed for the RxNorm API was doing critical work in our algorithm. The token splitting and drug name expansion done by the RxNorm's approximate match algorithm helped us identify

We classified free-text entries as opioids and non-opioids based on a series of rules depicted in Figure 1. These rules are programmed as an algorithm with four steps: 1) if an entry has an exact matched opioid name in RxNorm's classification systems or the CDC MME file (DL distance=0), then we define the record as an opioid; 2) if an entry has no exact matched opioid name found in either resource, but has an exact matched non-opioid name found via the RxNorm API approach, then we define the record as a non-opioid; 3) if an entry has no exact matched opioid name found in either resource, but the entry appears exactly as a dictionary word, then we define the record as a non-opioid; [5] and 4) otherwise, we check the DL distance of the records against opioids in the CDC MME file and define the record as an opioid if the DL distance is less than or equal to one. We also rely on the approximate match algorithm of RxNorm to match a free-text entry to drugs in RxNorm's classification systems, and identify a drug as an opioid if there is a match.

The same approach applies for 630 pre-filled drug names from the pull-down list. The primary difference is that there is no need to break pre-filled drug names into words or remove words unlikely to be drugs when applying the algorithm to drug names on the pull-down list.

Our algorithm was implemented in Python. We used the Python package NLTK (the Natural Language Tool Kit) to implement the three-step procedure that identifies opioids. We also used the Python package Jellyfish for approximate string matching, which is a technique of finding strings that match a pattern approximately. The Requests package is used for API requests. The

drugs otherwise impossible to identify on our training set. Even the RxNorm API's spelling correction algorithm was used in cases where the RxNorm API split our input token into multi tokens, though this was rare enough to only occur when we ran our algorithm on the full data set. See https://rxnav.nlm.nih.gov/RxNormApproxMatch.html for more details on the RxNorm approximate match algorithm.

[5] The resource used in this check is the "Words Corpus" that is available through the python NLP library of the NLTK. The "Words Corpus" has over 230,000 words and is essentially a dictionary of English words available on the UNIX operating system.

Sklean package (Sci-Kit Learn) is also used for some machine-learning utilities. The use of the Amazon Web Services cloud was critical for rapidly processing the data for this project.[6]

Figure 1. Flow chart depicting the final decision rule to identify opioids

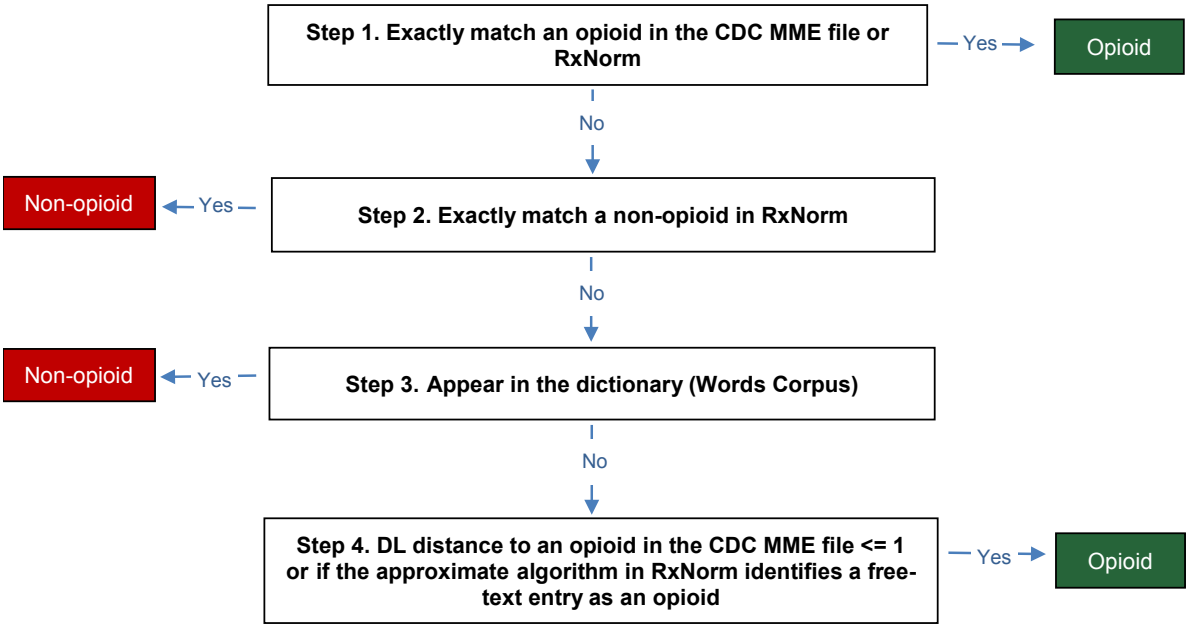

Figure 1 outlines the rules used to identify opioids. These rules are programmed as an algorithm with four steps: 1) if an entry has an exact matched opioid name in RxNorm's classification systems or the CDC MME file, then it is recorded as an opioid; 2) if an entry has no exact matched opioid name found in either resource, but has an exact matched non-opioid name found via the RxNorm API approach, then it is defined and recorded as a non-opioid; 3) if an entry has no exact matched opioid name found in either resource, but the entry appears exactly as a dictionary word, then it is recorded as a non-opioid; and 4) DL distance of the

[6] AWS gave us access to enough computational power to efficiently programming using NLP. Loading the data and all the general and specialized language dictionaries required by our NLP into memory is difficult to do on a standard laptop. Further, AWS gave us the flexibility in our IT infrastructure to evaluate several different NLP software packages used in industry that had different operating system and database requirements. It would have been impossible to support all these different requirements with our on premise infrastructure.

records is checked against opioids in the CDC MME file and define the record as an opioid if the DL distance is less than or equal to one.

E. Performance of the machine-learning algorithm and limitations

The machine-learning approach worked well in identifying opioid medications. We completed a few rounds of tests of and refinements to the algorithm, using standard metrics reported in the computer science literature for precision, including overall accuracy rate, sensitivity, and specificity.[7] Ultimately, we achieved an accuracy rate of over 99 percent for our out-of-sample prediction using the test file of free-text cases, with a sensitivity of 0.984 and perfect 1.0 specificity. This means that our algorithm made very few mistakes on the free-text cases, and the kinds of mistakes it made were undercounting a very small proportion of true opioid cases. Our performance on the pull-down list cases was perfect on our labeled test data, which is not surprising considering the classification problem is much simpler.

Although our machine-learning algorithm performed well, there is one notable limitation. Some applicants recorded what appear to be abbreviations for medications that are not identified in RxNorm's drug name expansion algorithm. For example, applicants filled in the abbreviations such as hydroco/acetam or hydrolapap (an odd abbreviation of hydrocodone + APAP), rather than the full drug name. We employed a conservative approach to classifying drugs, but this approach necessarily omits some abbreviations.

Another limitation of our approach is that we did not differentiate opioid antagonists from agonist opioids. Opioid antagonists are used to treat opioid addiction and block the effects of agonist opioids. For example, Naltrexone can be taken together with opioids to relieve pain and

[7] Sensitivity measures the proportion of true opioids we are able to identify correctly. It is calculated as true positives/(true positives + false negatives). Specificity measures the proportion of the non-opioid we were able to identify correctly. It is measured as true negatives/(true negatives + false positives).

to prevent abuse (for example, Morphine-Naltrexone). For future studies that use a machine-learning method to identify opioids, researchers should consider refining the algorithms to distinguish opioid antagonists from agonist opioids.[8]

[8] We did not differentiate opioid antagonists from agonist opioids because we do not have a comprehensive list of opioid antagonists and developing such a list is beyond the scope and resources of the current project. The lists developed for other Mathematica projects define drugs at the NDC level. Our methodology allows us to identify opioids but not at the NDC code level.

This page has been left blank for double-sided copying.

IV. RESULTS

A. All reported medications

Table IV.1 lists the 20 most common drug names chosen from the pull-down list in our sample. The most chosen drug name is Lisinopril (4.3 percent among entries using the pull-down list), which is a non-opioid used to lower blood pressure, followed by Metformin (2.6 percent), a diabetes medication. In addition, medications for heart and metabolic health were commonly reported by SSDI applicants via the pull-down list and five other medications in this category are among the top 20 medications reported in this manner (Metoprolol, Hydrochlorothiazide, Simvastatin, Omeprazole, and Amlodipine). Medications for mental illness are also common among this population, including Trazodone and three others (Xanax, Zoloft, and Cymbalta) in the top 20 pull-down selections. The remaining 9 medications among the top 20 most common drug names chosen from the pull-down menu are all pain medicines, the most common being Aspirin.

While there are 630 pre-filled drug names on the pull-down list, the use of medications among SSDI applicants is highly concentrated on a minority of drugs. The 40 most common medications selected from the pull-down list represent about 50 percent of all pull-down selections. Over 27 percent (173 out of 630) of drug names on the pull-down list have five or fewer entries and 52 drug names (8 percent) were not selected by any applicants in our sample.

Many (32 percent) of free-text entries are unique, and even the most common entries appeared less than one percent of the time. Many SSDI applicants not only reported a drug name in the free-text field, but they also included dosage information (for example, “Lyrica 150mg 2x Daily”) or short sentences describing the medications they were taking (for example, “a lot of

medications"). This highlights the usefulness of applying the NLP process to identify medications in a comprehensive manner.

Table IV.1. The 20 most common drug names chosen/reported from the pull-down list and free-text fields

Ranking on the pull-down list	Ranking on the free-text form	Drug Name	Mean (%, pull-down list)	Mean (%, free-form text)
1	1	Lisinopril	4.32	1.54
2	3	Metformin	2.64	1.07
3	6	Aspirin	2.53	0.92
4	7	Tramadol	2.08	0.86
5	9	Ibuprofen	2.04	0.68
6	4	Gabapentin	1.99	0.94
7		Hydrocodone with APAP	1.67	
8	10	Metoprolol	1.55	0.67
9		Trazodone	1.51	0.42
10	11	Albuterol	1.48	0.62
11		Hydrochlorothiazide	1.47	0.41
12	8	Oxycodone	1.42	0.8
13	17	Simvastatin	1.39	0.53
14	15	Omeprazole	1.34	0.55
15	18	Amlodipine	1.27	0.53
16	14	Xanax	1.14	0.56
17		Flexeril	1.06	0.45
18	13	Zoloft	1.02	0.58
19	16	Percocet	0.98	0.54
20	12	Cymbalta	0.93	0.6
	20	Naproxen	0.9	0.5
	19	Prednisone	0.62	0.5
	2	Hydrocodone		1.16
	5	Vitamin		0.93

Source: Authors' calculations using the SDR sample.

Table IV.1 also lists the 20 most common reported drugs in free-form text. Several interesting patterns emerged. First, all top 20 drug names from the free-text fields are available on the pull-down list. Second, the most frequently reported medications from free-text fields are generally consistent with those from the pull-down list: 16 of the top 20 free-text drugs overlap with the top 20 pull-down list drugs. Similar to the pattern we observed from the pull-down list, medications for pain management, heart and metabolic health, and mental illness are the most commonly reported drugs from free-text fields.

B. Opioids

Among the 630 pre-filled drug names on the pull-down list, we identified 27 drugs as opioids. Table IV.2 lists all 27 opioids we identified along with the number of times each was selected from the list by our sample. Among all medications recorded from the pull-down list, about 9.4 percent are opioids. The most commonly reported opioids from the pull-down lists are Tramadol (22 percent), Hydrocodone with APAP (18 percent), Oxycodone (15 percent), Percocet (10 percent), and Vicodin (8 percent). These top five reported opioids account for about 74 percent of all opioids selected from the pull-down list.

Table IV.2. Prevalence of opioids selected from the pull-down list

Description	Frequency	Percentage of all opioids selected from pull-down list
Tramadol	4200	22.11
Hydrocodone with APAP	3380	17.79
Oxycodone	2872	15.12
Percocet	1974	10.39
Vicodin	1588	8.36
Norco	1196	6.3
Lortab	1040	5.47
Morphine	869	4.57
Methadone	487	2.56
Oxycontin	388	2.04
Ultram	336	1.77
Tylenol with codeine	169	0.89
Dilaudid	162	0.85
Lorcet	109	0.57
Ms contin	75	0.39
Codeine	75	0.39
Roxicet	29	0.15
Duragesic patch	13	0.07
Meperidine	9	0.05
Tylox	7	0.04
Percodan	5	0.03
Promethazine with Codeine	5	0.03
Propoxyphene	5	0.03
Pentazocine	3	0.02
Synalgos-DC	1	0.01
Darvon	0	0.00
Propoxyphene N/APAP	0	0.00

Source: Authors' calculations using the SDR sample.

Note: Of the 201,923 medication records reported via the pull-down list, 18,997 were opioids.

Among free-text entries, we identified 8.8 percent as opioids. The top ten opioids reported via free-text are Hydrocodone, Tramadol, Oxycodone, Percocet, Vicodin, Norco, Morphine, Lortab, Oxycontin, and Fentanyl (Table IV.3).[9] Most of these frequently reported opioids from free-text fields are available on the pull-down list. The pattern of common opioids we observed based on the free-text fields are generally consistent with that from the pull-down list. The top five reported opioids account for about 62 percent of all opioids reported via free-text fields. While we identified 27 opioids from the pull-down list, we identified more opioids from the free-text fields, indicating that the free-text fields capture a wider variety of opioids.

Table IV.3. Top 20 most commonly reported opioids from the free-text fields

Opioid substances	Frequency	Percentage of all opioids identified from free-text fields
Hydrocodone	5757	20.34
Tramadol	3737	13.19
Oxycodone	3457	12.21
Percocet	2368	8.34
Vicodin	2127	7.5
Lortab	1609	5.66
Norco	1554	5.49
Morphine	1432	5.05
Oxycontin	667	2.34
Fentanyl	646	2.26
Codeine	529	1.84
Methadone	494	1.75
Ultram	472	1.66
Dilaudid	433	1.52
Hydromorphone	415	1.46
Suboxone	268	0.91
Nucynta	217	0.71
Ultracet	140	0.48
Endocet	123	0.42
Opana	117	0.4

Source: Authors' calculations using the SDR sample.

Note: Of the 315,668 medication records reported via free-text entries, 27,674 were opioids. We did not differential brand-name and generic drugs

[9] We proposed to classify opioids by their pharmaceutical classes: full opioid agonists versus partial opioid agonists versus opioid antagonists, and to classify opioids by onset and duration of action, including long-acting, short-acting, and rapid-onset. However, the information from free-text fields does support this analysis because in many cases the applicants only reported the substances of the opioid drugs but not the full drug name. For example, applicants only reported oxycodone, which is a substance for both long-acting opioid oxycodone hcl and long-acting opioid oxycodone hcl-acetaminophen.

When we combine opioids reported via the pull-down list and those reported via free-text form, we see a similar pattern observed among each separate reporting mode (Table IV.4). Considering both sources, the most prevalent opioids reported among our sample are Hydrocodone, Tramadol, and Oxycodone. We also see that opioids of different potencies have been used by this population, from Tramadol, an opioid generally considered to be safer by providers, to high potency opioids such as Fentanyl and Morphine.

Table IV.4. Top 15 most commonly reported opioids in the SDR (pull-down and free-form entries)

Opioid substances	Frequency	Percentage of all opioids identified from pull-down list and free-text entries
Hydrocodone *	9137	19.58
Tramadol	7937	17.01
Oxycodone	6329	13.56
Percocet	4342	9.30
Vicodin	3715	7.96
Norco	2750	5.89
Lortab	2649	5.68
Morphine	2301	4.93
Oxycontin	1055	2.26
Methadone	981	2.10
Ultram	808	1.73
Codeine	773	1.66
Fentanyl	646	1.38
Dilaudid	595	1.27
Hydromorphone	415	0.89

Source: Authors' calculations using the SDR sample.

*We combined Hydrocodone with Hydrocodone with APAP. We did not differential brand-name and generic drugs.

At the individual level, using the combined information from the pull-down list and the free-text fields, we found that among the 100,000 SSDI applicants in our sample, 34,697 (35 percent) reported opioid use at the time of application. About 74 percent of all applicants who reported any opioid use reported using one opioid, 24 percent reported using two or three opioids, and the remaining 2 percent reported using four or more opioids. The majority of opioid users (89 percent) specifically mentioned the word "pain" in their reason for taking the drug. The majority of the rest opioid users are likely using opioids for pain management, but mention something

more specific, including “back,’” “arthritis,” “migraines,” “joint,” and “headache.” A very small fraction of respondents listed “addiction” as the reason for taking opioids.

V. CONCLUSIONS AND IMPLICATIONS FOR POLICY

The rise in the number of SSDI applicants and awardees with musculoskeletal conditions combined with national trends of increasing prescription opioid use, suggests the potential for substantial opioid use in the SSDI population and raises concern for the overall health and safety of this population. Understanding the patterns of and reasons for opioid use among SSDI applicants assists SSA in understanding the health care needs and drug uses of this population and potentially supports SSA in identifying ways to help this group.

Our analysis illustrates an innovative machine-learning approach that can be used to mine the rich but rarely used medication information of SSDI applicants in the SDR. This study demonstrated the ability to use the SDR to provide statistics on patterns of medication use, the prevalence of opioids, and the health conditions prompting medication use among SSDI applicants. Among our non-representative study sample of SSDI applicants with any reported medication, we found that applicants used an average of five drugs at the time of application. Medications for pain management, heart and metabolic health, and mental illness were the most commonly reported drugs.

We produced detailed statistics about opioid use among our analysis sample. Among applicants with at least one reported medication, 35 percent reported using at least one opioid at the time of the application, and 26 percent of opioid users reported using more than one opioid.[10] The most commonly reported opioid substances are Hydrocodone, Tramadol, Oxycodone, Percocet, and Vicodin. Most (89 percent) of opioid users stated pain is the reason they use the

[10] Roughly 17 percent of all SSDI applicants in 2013 do not report any medication use at the time of application. Assuming the same proportion applies to our sample, the prevalence of opioid use among all SSDI applicants will be 29 percent. Because our sample only accounts for about five percent of all applicants in 2013, our findings should be interpreted with caution.

medicine. The high prevalence of opioids among our sample suggests that opioid use may indeed be high among SSDI applicants and suggests the need for research to produce representative statistics.

This study also highlights the importance of including the free-text entries when analyzing SDR data on medications. Among applicants in our sample, 45 percent reported medications exclusively via free-text fields. Hence, research that includes only medications recorded in the SDR pull-down list may omit roughly half of all reported medications. In the case of opioids, estimates based on data from the pull-down list alone would conclude that 16 percent [11] of our sample used opioids, compared with an estimated rate of 35 percent based on data from the pull-down list and free-text entries.

Another interesting finding that emerged from this study is that applicants were more likely to use the free-text fields even though the pull-down menu includes a comprehensive list of drug names. All of the 20 most common drug names reported in the free-text fields are included on the pull-down list or find the use of a pull-down list cumbersome. This may suggest that SSDI applicants prefer the control and specificity afforded by free-text entry. Greater use of the free-text fields might also signal the inefficiency of the pull-down list, as it may be overwhelming for applicants to have to scroll through 630 drug names. Further, the pull-down list has not been updated since 2007. This suggests that potential improvement to the application form may be needed both to reflect current, commonly-reported medications and to allow users to more easily navigate the extensive list of drug names.

[11] Among the 100,000 randomly sampled applicants, 16 percent selected an opioid from the pull-down list. If we use the total number of people who selected at least one entry from the pull-down list as denominator, the proportion is 29 percent.

The findings reported here represent the tip of the iceberg of knowledge that can be obtained by using machine learning to mine the rich medication data in the SDR. The success of the machine-learning method used in this study implies that this method could be used in the future to code all of the SDR prescription drug data and support numerous research efforts. For example, related research might include efforts to better understand the extent to which opioids are affecting the number of SSDI applicants and awardees and eventually consider the impact of policies designed to replace inappropriate use of opioids with other treatments. Such analyses could use the SDR alone or could link SDR data to data from other SSA research files such as the Disability Analysis File (all beneficiaries since 1996) and the Disability Research File (all applicants in the last 10 years).

This page has been left blank for double-sided copying.

REFERENCES

Carnide N, Hogg-Johnson S, Côté P, Irvin E, Van Eerd D, Koehoorn M, and Furlan AD. "Early Prescription Opioid Use for Musculoskeletal Disorders and Work Outcomes: A Systematic Review of the Literature. *The Clinical Journal of Pain*, July 2017, vol. 33, issue 7, pp. 647–658

Cutler, David, Ellen Meara, and Susan Stewart. "Has Wider Availability of Prescription Drugs for Pain Relief Affected SSDI and SSI Enrollment?" Presented at the Disability Research Consortium Annual Meeting, Washington, DC, August 2017.

Defalco F.J., P.B. Ryan, and M. Soledad Cepeda. "Applying Standardized Drug Terminologies to Observational Healthcare Databases: A Case Study on Opioid Exposure." *Health Services Outcomes Research Methodology*, vol. 13, no. 1, 2012, pp. 58–67.

Kolodny A., D.T. Courtwright, C.S. Hwang, P. Kreiner, J.L. Eadie, T.W. Clark, G.C. Alexander. "The Prescription Opioid and Heroin Crisis: A Public Health Approach to An Epidemic of Addiction." *Annual Review of Public Health*, vol. 36, 2015, pp. 559–574.

Mytelka, Christine M., Melanie K. Kuester, Colin R. Gray, and Lucas Everheart. "How Changing Opioid Prescribing Patterns Can Impact Risk Scores." *Issue in Brief: Medicaid Expansion CDPS + Rx Analysis*. Seattle, WA: Milliman, 2018.

Morden, Nancy E., Jeffrey C. Munson, Carrie H. Colla, Jonathan S. Skinner, Julie P.W. Bynum, Weiping Zhou, and Ellen R. Meara. "Prescription Opioid Use among Disabled Medicare Beneficiaries: Intensity, Trends, and Regional Variation." *Medical Care*, vol. 52, 2014, pp. 852–859.

U.S. Social Security Administration. Annual Statistical Supplement to the Social Security Bulletin, 2017.

Wu, April Yanayun. "New Occupational Statistics for SSDI Applicants and Their Policy Implications." Working paper. Chicago, IL: Mathematica Policy Research, 2018.

This page has been left blank for double-sided copying.

www.mathematica.org

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ TUCSON, AZ ■ SEATTLE, WA ■ WASHINGTON, DC ■ WOODLAWN, MD

MATHEMATICA
CENTER FOR
STUDYING DISABILITY POLICY