U.S. FOOD & DRUG ADMINISTRATION

Center for Biologics Evaluation and Research
Office of Biostatistics and Epidemiology

CBER Surveillance Program Biologics Effectiveness and Safety Initiative

A Structured Review of Electronic Coding Algorithms for Identifying Immunocompromised Cohorts Using Administrative Claims and Electronic Health Records

Final Report

May 24, 2021

Prepared by: Patrick Saunders-Hastings[1], Timothy Burrell[2], Graça Dores,[3] Kinnera Chada,[3] Hui-Lee Wong,[3] Azadeh Shoaibi[3]

Author Affiliations:
[1] Gevity Consulting Inc., Ottawa, Ontario, Canada
[2] IBM Watson Health, Cambridge, Massachusetts, United States of America

[3] Center for Biologics Evaluation and Research, Food and Drug Administration, Silver Spring, Maryland, United States of America

This page is intentionally left blank

Table of Contents

Table of Tables

Table of Figures

List of Acronyms

AFHSB	United States Armed Forces Health Surveillance Branch
AHRQ	Agency for Healthcare Research and Quality
AIDS	Acquired Immunodeficiency Syndrome
BEST	Biologics Effectiveness and Safety
CBER	Center for Biologics Evaluation and Research
CDC	Centers for Disease Control and Prevention
CNS	Central Nervous System
CI	Confidence Interval
CPT	Current Procedural Terminology
EHR	Electronic Health Record
FDA	Food and Drug Administration
GEMs	General Equivalence Mappings
HCPCS	Healthcare Common Procedure Coding System
HIV	Human Immunodeficiency Virus
ICD-9-CM	International Classification of Diseases, Ninth Revision, Clinical Modification
ICD-10-CM	International Classification of Diseases, Tenth Revision, Clinical Modification
ITT	Intention to Treat
LOINC	Logical Observation Identifiers Names and Codes
MPR	Medication Possession Ratio
NDC	National Drug Code
NEC	Not Elsewhere Classified
NOS	Not Otherwise Specified
NPV	Negative Predictive Value
PICO	Population, Intervention, Comparator, Outcome
PID	Primary Immunodeficiency Disease
PPV	Positive Predictive Value
RX	Treatment
SME	Subject Matter Expert
U.S.	United States
WBC	White Blood Cell

A Summary

The United States (U.S.) Food and Drug Administration (FDA) Biologics Effectiveness and Safety (BEST) Initiative conducted a structured literature review (through May 17, 2020) to develop validated coding algorithms for identifying immunocompromised individuals in large administrative healthcare databases. The studies selected for this targeted review used billing codes in claims or electronic health record (EHR) databases. A key challenge in developing this algorithm was the lack of a standard, widely accepted definition for 'clinical immunosuppression'. As a result, we used a previously-published, independently-validated algorithm presented by Greenberg and colleagues.[1] In this study, the authors developed six categories of immunosuppression, based on previously published classification schemes.[2-4] The overall and category-specific validity of the Greenberg algorithm is summarized as follows, with sensitivity values consistently lower than other performance measures:

- **Overall:** positive predictive value (PPV) 94.4% (95% confidence interval [CI] 88.8–97.7%); negative predictive value (NPV) 94.3% (95% CI 91.0–96.6%); sensitivity 87.4% (95% CI 80.6–92.5%); specificity 97.6% (95% ci 95.0–99.9%)
- **Human immunodeficiency virus (HIV)/Acquired immunodeficiency syndrome (AIDS):** PPV 100.0% (95% CI 47.8–100.0%); NPV 99.8% (95% CI 98.7–100.0%); sensitivity 83.3% (95% CI 35.9–99.6%); specificity 100.0% (95% CI 99.1–100.0%)
- **Organ transplant:** PPV 95.7% (95% CI 78.1–99.0%); NPV 99.1% (95% CI % 97.3–99.8%); sensitivity 88.0% (95% CI 68.8–97.5%); specificity 97.7% (95% CI 98.3–100.0%)
- **Rheumatologic/inflammatory:** PPV 91.7% (95% CI73.0–99.0%); NPV 97.8% (95% CI % 95.6–99.1%); sensitivity 75.9% (95% CI 56.5–89.7%); specificity 99.4% (95% CI 97.9–99.9%)
- **Hematological malignancy:** PPV 88.9% (95% CI 73.9–96.9%); NPV 97.9% (95% CI 95.9–99.1%); sensitivity 80.0% (95% CI 64.4–90.9%); specificity 99% (95% CI 97.3–99.7%)
- **Solid malignancy:** PPV 75.0% (95% CI 34.9–96.8%); NPV 98.5% (95% CI % 96.6–99.5%); sensitivity 54.5% (95% CI 23.4–83.3%); specificity 99.4% (95% CI 97.9–99.9%)
- **Other intrinsic immune conditions:** PPV 71.4% (95% CI 53.7–85.4%); NPV 99.2% (95% CI 97.2–99.8%); sensitivity 89.3% (95% CI 71.8–97.7%); specificity 97.5% (95% CI 95.4–98.8%)

The findings from this literature review — specifically related to the Greenberg algorithm — were leveraged to develop a comprehensive-code based algorithm for identifying immunocompromised individuals. Codes were mapped from ICD-9-CM to International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM) via forward–backward mapping, using General Equivalence Mappings (GEMs) for reference.[i] The draft algorithm was then reviewed by clinical subject matter experts (SMEs) from FDA Center for Biologics Evaluation and Research (CBER) (GD), IBM (TB), and Acumen. In addition, previous efforts to identify immunocompromised individuals by CBER (unpublished) and the Agency for Healthcare Research and Quality (AHRQ) informed development of the algorithm.

The final algorithm proposes nine categories for users seeking to identify individuals that may be immunocompromised. The algorithm was developed with a focus on optimizing sensitivity, and the application of these categories, and of certain codes within these categories, may vary depending on user priorities.

Most categories (with the exception of transplants and related conditions) required at least two claims on different days for a positive identification. In the literature, performance improved when more than one code was required. Additionally, this was supported by clinical judgement that these chronic conditions would be associated with more than one claim in true cases. An exclusion criterion was proposed for individuals whose most recent claim was over six months before the index date (to be defined based on research question), based on the clinical judgement that immunocompromising conditions can resolve

[i] Additional information about GEMs and the methodology for forward and backward mapping can be found at Centers for Medicaid and Medicare Services. (2017). 2018 ICD-10-CM and GEMs. Available at https://www.cms.gov/Medicare/Coding/ICD10/2018-ICD-10-CM-and-GEMs. Researchers used the following website to map ICD-9-CM codes to ICD-10-CM: https://www.icd10data.com.

and may not result in immunosuppression for life. The algorithm for the nine disease groups is summarized as follows:

1) **HIV/AIDS:** ≥2 claims for ANY HIV/AIDS diagnostic or procedural codes.
    a. This category is viewed as immunocompromised regardless of treatment status (treatment [RX]-independent)
    b. EXCLUDE cases where the most recent claim is MORE THAN 6 months prior to the index date.
2) **Hematologic Malignancy and Related Conditions**: ≥2 claims for ANY relevant diagnostic codes.
    a. This category is viewed as immunocompromised regardless of treatment status (RX-independent).
    b. EXCLUDE cases where the most recent claim is MORE THAN 6 months prior to the index date.
3) **Immune deficiencies (treatment-independent)**[ii]: ≥2 claims for ANY treatment-independent immune deficiency diagnostic codes.
    a. This category is viewed as immunocompromised regardless of treatment status (RX-independent).
    b. EXCLUDE cases where the most recent claim is MORE THAN 6 months prior to the index date.
4) **Immune deficiencies (treatment-dependent):** ≥2 claims for ANY treatment-dependent immune deficiency diagnostic codes AND one of:
    a. ≥2 claims for ANY codes for chemotherapy or radiation
    b. ≥2 claims for ANY codes for immune-modulating or immunosuppressant therapy
    c. EXCLUDE cases where the most recent claim is MORE THAN 6 months prior to the index date.

    NOTE: Codes for brachytherapy (internal radiation via the insertion of implants) may also be considered by users seeking to apply a more inclusive definition of radiation therapy. It is recommended that this category be ignored in regular applications of the algorithm, as brachytherapy is less likely to be immunosuppressive, but it may be of value to those with a specific interest in brachytherapy.
5) **Solid Malignancy:** ≥2 claims for ANY solid malignancy diagnostic codes AND one of:
    a. ≥2 claims for ANY codes for chemotherapy or radiation
    b. ≥2 claims for ANY codes for immune-modulating or immunosuppressant therapy
    c. EXCLUDE cases where the most recent claim is MORE THAN 6 months prior to the index date.

    NOTE: Codes for brachytherapy may also be considered by users seeking to apply a more inclusive definition of radiation therapy. It is recommended that this category be ignored in regular applications of the algorithm, but it may be of value to those with a specific interest in brachytherapy.
6) **Transplant and Related Conditions:** ≥1 claims for ANY transplant or related condition diagnostic codes AND one of:
    a. ≥2 claims for ANY codes for chemotherapy or radiation
    b. ≥2 claims for ANY codes for immune-modulating or immunosuppressant therapy
    c. EXCLUDE cases where the most recent claim is MORE THAN 6 months prior to the index date.

    NOTE: Codes for brachytherapy may also be considered by users seeking to apply a more inclusive definition of radiation therapy. It is recommended that this category be ignored in regular applications of the algorithm, but it may be of value to those with a specific interest in brachytherapy.
7) **Rheumatologic/Inflammatory:** ≥2 claims for ANY rheumatologic or inflammatory condition diagnostic codes AND one of:
    a. ≥2 claims for ANY codes for chemotherapy or radiation

[ii] Treatment-independent categories include conditions that are viewed as immunocompromising. Treatment-dependent categories include conditions associated with treatment viewed as immunocompromising. Additional information on the distinction between treatment-independent and treatment-dependent categories is provided in Section C2.

b. ≥2 claims for ANY codes for immune-modulating or immunosuppressant therapy
c. ≥2 claims for ANY steroid diagnosis codes
a. A 90-day medication possession ratio (MPR) ≥75% for ANY steroid prescription codes
b. EXCLUDE cases where the most recent claim is MORE THAN 6 months prior to the index date.

NOTE: Codes for brachytherapy may also be considered by users seeking to apply a more inclusive definition of radiation therapy. It is recommended that this category be ignored in regular applications of the algorithm, but it may be of value to those with a specific interest in brachytherapy.

NOTE: It should be indicated that steroid MPR is based on an Intention to Treat (ITT) analysis, as the ratio depends on days' supply reported in a claim and does not account for patient compliance. For example, receiving a 90-day steroid prescription would automatically put a patient above the MPR ≥75% threshold, albeit with no guarantee of steroid use.

8) **Dialysis:** ≥2 claims for ANY dialysis diagnostic or procedural codes
   a. EXCLUDE cases where most recent claim is MORE THAN 6 months prior to the index date.
9) **Intermediate conditions** ≥2 claims for ANY intermediate condition diagnostic codes AND one of:
   a. ≥2 claims for ANY codes for chemotherapy or radiation
   b. ≥2 claims for ANY codes for immune-modulating or immunosuppressant therapy
   c. EXCLUDE cases where treatment claim predates diagnostic claim OR where most recent claim is MORE THAN 6 months prior to the index date.

   NOTE: These conditions were viewed as unlikely to be immunocompromising but were included for optional use by users seeking to be as inclusive as possible.

   NOTE: Codes for brachytherapy may also be considered by users seeking to apply a more inclusive definition of radiation therapy. It is recommended that this category be ignored in regular applications of the algorithm, but it may be of value to those with a specific interest in brachytherapy.

As an initial step in assessing the feasibility of using the algorithm to identify immunocompromised individuals, the algorithm was applied in the IBM MarketScan® Research Databases (Commercial and Medicare Supplemental), a large collection of commercially insured individuals in the U.S. Statistics describing the frequency and proportions of immunocompromising disease categories were generated.

B Background

Among other responsibilities, the U.S. FDA is mandated to protect public health by ensuring the safety and efficacy of drugs, biologics, and medical devices.[iii] FDA CBER protects and advances the public health by assuring the safety and effectiveness of biologics and related products, including blood products, vaccines, allergenics, tissues, and cellular and gene therapies. CBER initiated the BEST Initiative to monitor the safety and effectiveness of biologic products using large datasets of clinical and administrative healthcare data.

Some products that CBER regulates, such as vaccines, have higher risk for adverse events among those who are immunocompromised.[2] A person with a weakened immune system is considered immunocompromised which increases the risk and severity of primary infections, reactivation of infections[5], and serious outcomes such as sepsis.[1,3] Although the number of immunocompromised individuals is unknown, the prevalence of immunosuppressed individuals in the U.S. has been estimated to be approximately 3%.[5] A validated claims-based algorithm with known performance metrics affords identification of these conditions in biologic surveillance studies.

[iii] U.S. Food and Drug Administration. What We Do. March 28, 2018. https://www.fda.gov/aboutfda/whatwedo/

The objective of this review was to assess the availability of validated algorithms for identifying immunocompromised individuals from administrative claims data. These code algorithms could draw on a variety of standard code sets, including the International Classification of Diseases (ICD), the Healthcare Common Procedure Coding System (HCPCS), Current Procedural Terminology (CPT®), National Drug Codes (NDCs), and Logical Observation Identifiers Names and Codes (LOINC). Information from this review was used in development of algorithms where validated algorithms were not available.

A structured literature review of code algorithms for identifying immunocompromised cohorts using administrative claims and EHRs was conducted, leveraging findings from U.S. and international studies to inform algorithm development. The focus of the review was on algorithms derived from administrative claims data (i.e., claims-based), while algorithms derived from EHRs that used standard billing code sets (i.e., EHR-based) were also considered. The draft algorithm was then subject to review by clinical SMEs IBM (TB, JB), FDA CBER (JC, DT), and Acumen, and testing in the IBM MarketScan® Research Databases (Commercial and Medicare Supplemental), a large collection of U.S. administrative insurance claims data accessed via the Treatment Pathways online analytic platform. **Section C**summarizes the literature review methodology and findings; **Section D** provides a clinical case definition for immunosuppression, which could be used in further assessing the performance of the proposed algorithm via chart review validation studies; **Sections E** and **F** present the algorithm and its associated assumptions and decisions, respectively; **Section G** presents the approach for and results of the initial application of the algorithm to characterize the immunocompromised population in a claims database; and **Section H** provides discussion and conclusions.

C Literature Review

C1 Methods

The workgroup developed a literature review search strategy based upon a Population, Intervention, Comparator, and Outcome (PICO) framework. The PICO framework for this review can be summarized as follows:

- **Population:** *any population group (human), immunocompromised, immunosuppressed, HIV, AIDS, malignancy, transplant, dialysis, autoimmune*
- **Intervention:** *any intervention or no intervention*
- **Comparator:** *any comparator, placebo*
- **Outcome:** *immunocompromised, immunosuppressed, HIV, AIDS, malignancy, transplant, dialysis, autoimmune*

The setting for eligible studies was any clinically observable environment that led an individual to seek care.

The review process began with conducting comprehensive searches of existing publications available in the CBER[iv] and Center for Drug Evaluation and Research Sentinel[v] databases. CBER previously developed a draft immunocompromised algorithm (unpublished) that served as a starting point for algorithm development efforts. Next, a structured review of the peer-reviewed literature was conducted, using PubMed, Medline, and Google Scholar to identify relevant resources. Only English language publications were selected for review. No restriction was imposed on publication date for the PubMed search (inception to May 17, 2020). The Google Scholar search was limited to January 2000 through May 17, 2020.

[iv] U.S. Food and Drug Administration. Innovation and Regulatory Science. Accessed on July 10, 2020. https://www.fda.gov/vaccines-blood-biologics/science-research-biologics/innovation-and-regulatory-science
[v] Sentinel. Publications and Presentations. https://www.sentinelinitiative.org/communications/publications

Targeted and *ad hoc* searches of the gray literature were conducted, including clinical guidelines and reports from organizations such as the United States Armed Forces Health Surveillance Branch (AFHSB) and the Agency for Healthcare Research and Quality (AHRQ). AHRQ previously developed criteria for identifying individuals at high and intermediate risk of immunosuppression, which was leveraged to further expand the code list developed by CBER clinicians. A snowballing technique was also applied, wherein the bibliographies of relevant studies were scanned for additional publications. As the literature search found few articles that comprehensively covered immunodeficiency it became clear that articles needed to be further categorized into those that provided information to build an algorithm framework, and those that provided information on subcategories of diseases within an algorithm framework.

All titles and abstracts were reviewed[vi], and 27 articles were reviewed in full text. Of these, 19 were retained for extraction and informed algorithm development. A Microsoft® Excel spreadsheet was developed to extract relevant data. The data elements collected are provided in **Table 1**. A relevance ranking was assigned based on the judgement of the reviewer and the available information on study location ("Country"), the algorithm specifications ("Algorithm"), and the measures of validity and diagnostic accuracy (e.g., positive predictive value [PPV] and negative predictive value [NPV]). Relevance rankings were assigned based on the following criteria:

- ***Ranking 1:*** U.S. claims- or EHR-based validation study (i.e., reporting measures of validity and diagnostic accuracy)
- ***Ranking 2:*** U.S. study that reported a claims- or EHR-based definition but no independent validation OR a non-U.S. validation study
- ***Ranking 3:*** Non-U.S. study that reported a claims- or EHR-based definition but no independent validation

Table 1. Data elements recorded in the extraction spreadsheet.

Data Element
Author
Publication Year
Article Relevance (Ranking 1-3)
Full Citation
Country of Study
Data Source
Years Included
Population Eligibility Criteria
Validation Method
Disease Definition
Algorithm Incidence Rules
ICD-9/ICD-9-CM Codes
ICD-10/ICD-10-CM Codes
Other Codes
PPV % (95% Confidence Interval [CI])
NPV % (95% CI)
Other Performance Measures
Comments

Abbreviations: ICD-9-CM, International Classification of Diseases, Ninth Revision, Clinical Modification; ICD-10-CM, International Classification of Diseases, Tenth Revision, Clinical Modification; PPV, Positive predictive value; NPV, Negative predictive value; 95% CI, 95% confidence interval

[vi] Since this was not a systematic review, authors did not track the total number of abstracts screened after de-duplication.

C2 Results

Of the 19 articles retained for data extraction (see **Appendix A** for additional details), articles by Greenberg and colleagues[1] and Sunesen and colleagues[6], respectively, provided information on how to construct an overall algorithm framework for immunocompromised conditions. The algorithm framework was deemed to be critical as it helped to define the scope and breadth of diseases and conditions to be considered to be an immunocompromised condition. The remaining 17 articles provided information on specific diseases or disease groups.

2.a Background of Algorithm Framework

CBER developed a draft algorithm for identifying an immunocompromised population in 2018; the current algorithm in this report is an update and expansion of that previous algorithm. The initial CBER draft algorithm and the algorithm proposed herein are modified versions of an algorithm developed and validated by Greenberg and colleagues,[1] hereafter referred to as "the Greenberg algorithm". The Greenberg algorithm used ICD-9-CM codes, which have subsequently been mapped to ICD-10-CM using GEMs. The previous CBER draft algorithm also added codes and new disease categories (e.g., dialysis) based on clinical insights and an approach developed by AHRQ.[7]

For this algorithm development and update effort, CBER assembled a committee of clinical SMEs to guide the update process and edit, expand, and refine the algorithm structure and code lists. As with the previous iteration, the current algorithm utilizes an approach similar to Greenberg and colleagues.[1] Greenberg and colleagues identified 4,438 patients who were diagnosed with sepsis occurring between 2010–2012 from a consortium of academic medical care systems and used a list of ICD-9-CM codes in six categories to identify those who were immunocompromised. These categories were based on work by Tolsma and colleagues,[3] Rubin and colleagues,[2] and Poutsiaka and colleagues,[4] though these studies did not involve validation. Using the Greenberg algorithm, three categories are considered definitely immunosuppressive: human immunodeficiency virus infection (HIV), hematologic malignancies, and other intrinsic immune conditions. Patients with the following conditions in other categories were considered immunosuppressed only if they received an immunosuppressive medication during the studied hospitalization: solid malignancies, solid organ transplantations, and rheumatologic/inflammatory conditions. From the 4,438 patients with sepsis, 421 were randomly selected for manual chart review and were categorized as immunocompromised or not; this was considered the "gold" standard. The Greenberg algorithm had an overall PPV of 94.4% (95% CI 88.8–97.7%), NPV of 94.3% (95% CI 91.0-96.6%), sensitivity of 87.4% (95% CI 80.6–92.5%) and specificity of 97.6% (95% CI 95.0–99.9%). This Greenberg algorithm was subsequently used by Greenberg and colleagues in a study published in 2018.[8]

The previous CBER draft algorithm from 2018 was the starting point from which the current algorithm was built. As mentioned earlier, the 2018 algorithm modified the 2016 Greenberg algorithm.[1] Two noted changes were the inclusion of radiation therapy and dialysis codes. Immune cells are sensitive to radiation and the effect of radiation on the immune system is complex with many physiological aspects still unknown. The effect of ionizing radiation results in dead or damaged cells which release a variety of substances that can result in either immunosuppression or, paradoxically, have immune stimulating properties.[9] Dialysis codes were also added to the previous CBER algorithm as it was noted that AHRQ had classified dialysis-dependent conditions as being related to immunosuppression. The inclusion of dialysis is supported by the literature indicating chronic kidney disease patients and uremic patients have associated immune dysfunction.[10-12]

2.b Current Algorithm Framework

For the current update to the CBER draft algorithm from 2018, the committee, comprised of clinical SMEs from FDA CBER (JC, DT), IBM (TB, JB), and Acumen, edited and expanded the algorithm in several ways. For each category, codes were re-checked and additional codes were added, if necessary.

The current algorithm retains the structure of having two major categories of conditions: conditions deemed as sufficient to identify an immunocompromised individual regardless of treatment status and those that require additional evidence of treatment (for example conditions which require treatment with an immunosuppressive medication). The nomenclature for some of the categories used in the Greenberg algorithm were changed and additional categories were added. In particular, a category was added for "intermediate" immune conditions. Conditions in this category are viewed as milder conditions that may not be associated with significant immunosuppression but that could be included by a researcher or health surveillance project seeking to be more inclusive.

Many of the decisions regarding classification of specific conditions were based on the judgment of clinical SMEs. The rationale for the decisions was based on clinical experience and, when possible, these decisions were supported by published, peer-reviewed literature.

2.c Literature Search Results for Specific Diseases or Groups of Conditions

For most of the specific diseases in the categories above, literature searches were conducted to identify published material on administrative codes (with validation when available) associated with the condition(s). The intent of the literature search was not to conduct a full systematic review, but rather to identify articles to establish, confirm, or support the use of codes and methodologies for major categories and if possible, selected specific conditions. Each publication reported either measures of diagnostic accuracy associated with claims-based algorithms (i.e., codes derived from administrative insurance claims databases) or EHR-based algorithms (i.e., codes derived from admission or discharge medical records for disease management). Additional publications identified approaches for identifying cases of immunocompromised individuals using administrative claims or EHR data but did not validate their approach. We have summarized the literature below by specific diseases or groups of conditions.

i HIV/AIDS

Sunesen and colleagues defined a population of patients with immunosuppressive disorders to investigate the subsequent risk of anal squamous cell carcinoma.[6] Patients were enrolled in the Danish National Patient Registry from 1978–2005 with a first-time hospital contact or procedure for HIV infection, solid organ transplantation or autoimmune disease, or a first-time record of hematologic malignancy. The authors used ICD-8 and ICD-10 codes to define the categories and used ICD-10 codes of B20-24 for HIV/AIDS. A single diagnosis of HIV/AIDS was sufficient to categorize an individual as immunosuppressed as long as the encounter was not an emergency department encounter. This coding algorithm was not validated.

Antoniou and colleagues validated a case-finding algorithms for HIV using administrative data from charts of 2,040 randomly selected patients from two medical practices in Toronto, Ontario, between April 2007 and March 2009.[13] The relevant codes for HIV were ICD-9: 042, 043, 044 and ICD-10: B20-B24. A total of 48 algorithms were constructed and compared against chart audits as the standard. The algorithm of three physician claims over a three-year period had a sensitivity and specificity of 96.2% (95% CI 95.2–97.9%) and 99.6% (95% CI 99.1–99.8%), respectively. The specificity of all algorithms with more than one physician claim had specificity exceeded 99%.

ii Hematologic Malignancies

Sunesen and colleagues used the following ICD-10 codes to identify hematologic malignancies: C80-85, C88, and C90-96. A single diagnosis was sufficient to categorize an individual as immunosuppressed.[6] Validation was not conducted.

Additional articles that covered specific conditions of hematological malignancies were found.

The identification of lymphoma using administrative data was systematically reviewed by Herman and colleagues.[14] Eleven articles with algorithms were identified, of which one by Setoguchi and colleagues presented validation measures.[15] This study identified patients 65 years and older from a Medicare claims database in Pennsylvania using ICD-9-CM codes 200.xx, 201.xx, 202.xx (except 202.5x and 202.6x) and CPT codes.[14,15] Validation was done by comparison with cancer registry reports. The four proposed algorithms had specificity greater than 99%; an algorithm with two codes recorded within 2 months had the best PPV (62.83%) and sensitivity (79.81%). Herman and colleagues suggested that the presence of a single diagnostic code for an individual is insufficient to identify lymphoma.[15]

Brandenburg and colleagues validated an algorithm for multiple myeloma based on at least one ICD-9-CM code 203.xx from two populations (2005-2011 and 2008-2012) from a large health system in Michigan.[16] For the first population (i.e., 2005–2011), a tumor registry was used to validate diagnoses and several algorithms were tested. The PPVs of the algorithms ranged from 54% to 88% and the sensitivities ranged from 30% to 88%. Based on the results of the analysis of this first population, the authors chose and applied an algorithm using a diagnosis code occurring before and after a diagnostic procedure code for the second population. For the second population, medical chart review was conducted for validation. The authors reported a PPV of 86% (95% CI 79-92%).

In a study by Setoguchi and colleagues mentioned above, the authors also identified leukemia cases using four algorithms (specific ICD codes not provided).[15] The specificity of the algorithms was high (above 99%) but the sensitivity and PPV were low (sensitivity 41.8–73.6%, PPV 18.8–43.2%).

The validity of using acute myeloid leukemia codes (ICD-9-CM 204.xx-208.xx) was assessed via chart review using records in a children's hospital in Philadelphia. In a paper by Kavcic and colleagues,[17] an algorithm approach using one or more hospitalizations with a relevant ICD-9-CM diagnosis code yielded a sensitivity of 100% and PPV of 31%. When administrative codes for chemotherapy were added in the algorithm, the sensitivity increased to 95.7% and the PPV increased to 100%.

iii Solid Malignancy

Greenberg and colleagues provided a list of conditions and related ICD-9-CM codes for solid malignancies.[1] To be considered immunocompromised, a patient needed additional codes indicating treatment. Greenberg and colleagues did not specify treatment codes (e.g., immunosuppressive medications) that were included in their analysis, though a lists of codes to consider for treatment-dependent categories is provided in **Appendix B**. Greenberg and colleagues used the following ICD-9-CM codes to identify solid malignancies:

140.x-199.x	Organ/system malignant tumors
209.x	Neuroendocrine tumors
235.x-239.x	Neoplasms of uncertain behavior

The validation measures for a solid malignancy diagnosis with immunosuppressive medication were as follows: PPV 75.0% (95% CI 34.9–96.8%), NPV 98.5% (95% CI 96.6–99.5%), sensitivity 54.5% (95% CI 23.4–83.3%). specificity 99.4% (95% CI 97.9–99.9%).

Setoguchi and colleagues identified patients with malignancies using four algorithms.[15] In addition to lymphoma and leukemia, discussed above, the authors also identified lung, colorectal, stomach and breast cancers. The authors did not provide the specific ICD codes used, though validation was conducted. All four algorithms had very high specificity (greater than 98%), sensitivities ranged from 46.9% to 88.0%, and PPV ranged from 44.8% to 81.7%.

Other studies used administrative data to establish cohorts for some of these conditions but did not conduct validation. Using ICD-9-CM codes 140-208, Shayne and colleagues defined a cohort of those with cancer for the purpose of studying length of hospital stays.[18]

iv Organ Transplant

Greenberg and colleagues (2016) provide a list of conditions and related ICD-9-CM codes for organ transplantation. To be considered immunocompromised, a patient needed additional codes indicating treatment. Greenberg and colleagues did not specify treatment codes (e.g., immunosuppressive medications) that were included in their analysis, though a list of codes to consider for treatment-dependent categories is provided in **Appendix B**. Greenberg and colleagues reported the following ICD codes for identifying organ transplantation:

996.8	Complications of transplanted organ
V42.x	Organ transplant status

Organ transplant with immunosuppressive medications was associated with the following measures of validation: PPV 95.7% (95% CI 78.1–99.9%), NPV 99.1% (95% CI 97.3–99.8%), sensitivity 88.0% (95% CI 68.8–97.5%), specificity 99.7% (95% CI 98.3–100.0%).

v Rheumatologic/Inflammatory

Greenberg and colleagues provide a list of rheumatologic/inflammatory conditions and related ICD-9-CM diagnosis codes.[1] To be considered immunocompromised, a patient needed additional codes indicating treatment for the rheumatologic/inflammatory condition. Greenberg and colleagues used the following ICD-9-CM codes to identify rheumatological/inflammatory conditions:

135.x	Sarcoidosis
277.3	Amyloidosis NOS
277.31	Familial Mediterranean fever
277.39	Amyloidosis NEC
340.x	Multiple sclerosis
341.x	Other Central Nervous System (CNS) demyelination
357.x	Acute infective polyneuritis
422.x	Acute myocarditis
446.x	Polyarteritis nodosa and allied conditions
495.9	Allergic alveolitis/pneumonitis/NOS
516.x	Other alveolar and parietoalveolar pneumonopathy
555.x-558.x	Enteritis and colitis
695.4	Lupus erythematosus
710.x	Diffuse connective tissue disease
711.x	Arthropathy with infection
712.x	Crystal arthropathies
714.x	Rheumatoid arthritis & inflammatory polyarthropathies
720.x	Ankylosing spondylitis and other inflammatory spondylopathies
725.x	Polymyalgia rheumatica

CNS; Central nervous system; NOS: Not otherwise specified; NEC: Not elsewhere classified

Rheumatologic/inflammatory diagnoses with receipt of an immunosuppressive medication was associated with the following measures of validation: PPV 91.7% (95% CI 73.0–99.0%), NPV 97.8% (95% CI 95.6–99.1%), sensitivity 75.9% (95% CI 56.5–89.7%), specificity 99.4% (95% CI 97.7–99.9%).

Additional autoimmune conditions were later added to this category by CBER. There are a large number of autoimmune diseases and patients with an autoimmune disease may be immunocompromised due to "intrinsic alteration in immune system, immunosuppressive therapies or both".[6] There is no consensus

criteria for defining autoimmune diseases.[19-21] However it is generally considered that there are over 80 autoimmune diseases, some of which are very rare. Roberts and colleagues provided ICD-9 and ICD-10 codes for 21 of these diseases.[22] Validation was not performed to assess the performance of these codes.

Among patients hospitalized with a diagnosis for autoimmune disease, Sunesen and colleagues reported a list of immunosuppressive conditions (
Table 2).[6] Validation was not conducted in this study.

Table 2. Autoimmune conditions reported as immunosuppressive by Sunesen and colleagues.

ICD-10 Code*	ICD-10 Description
Hematological system	
D59.0, D59.1	Autoimmune hemolytic anemia
D69.3	Idiopathic thrombocytopenic purpura
Endocrine system	
E05.0	Graves' disease
E06.3	Autoimmune thyroiditis
E27.1	Addison´s disease
Central nervous / neuromuscular system	
G35.x	Multiple sclerosis
G70.0	Myasthenia gravis
Gastrointestinal / hepatobiliary system	
D51.0	Pernicious anemia
K90.0	Celiac disease
K50, M07.4	Crohn´s disease
K51, M07.5	Ulcerative colitis
K74.3	Primary biliary cirrhosis
Skin	
L10.0, L10.2, L10.4, L12.0	Pemphigus / pemphigoid
L13.0	Dermatitis herpetiformis
L40, M07.0-M07.3	Psoriasis
Connective tissue diseases	
M05.x, M06.x, G73.7D, I32.8A, I39.8E, I41.8A, I52.8A	Rheumatoid arthritis
M45.x	Ankylosing spondylitis
M33.x	Polymyositis / dermatomyositis
M32.x, G73.7C, I39.8C, N08.5A, N16.4B	Systemic lupus erythematosus
M35.x, G73.7A, N16.4A	Sjogren's syndrome
D86.x, G53.2, H22.1A, I41.8B, K77.8B, M63.3	Sarcoidosis
Vasculitis syndromes	
M30.0	Polyarteritis nodosa
M31.3	Wegener´s granulomatosis
M31.5, M31.6, M35.3	Temporal arteritis / polymyalgia rheumatica

*ICD-10 codes for autoimmune diseases are reported in the Supporting Information of a manuscript by Sunesun and colleagues.[6]

vi Other Immune Conditions

The ICD-9-CM diagnosis codes for "Other immune conditions" assigned by Greenberg and colleagues formed the basis for Immune Deficiencies categories presented in the proposed algorithm (**Section A** and **Section E**).[1] Codes were further categorized into treatment-independent and treatment-dependent immune deficiencies based on the judgement of clinical SMEs. The ICD-9-CM codes included by Greenberg and colleagues

279.x	Disorders of Immune mechanism
288.0	Neutropenia
288.1	Functional disorders of polymorphnuclear neutrophils
288.2	Genetic anomalies of leukocytes
288.5	Decreased WBC count
288.8	WBC disease NEC
288.9	WBC disease NOS
289.83	Myelofibrosis
289.89	Blood diseases NEC
289.9	Immunologic findings NEC
795.7	Immunological findings NEC
795.79	Nonspecific immune findings NEC and NOS

NOS: Not otherwise specified; NEC: Not elsewhere classified

The reported validation measures and confidence intervals were as follows: PPV of 71.4% (95% CI 53.7-85.4%), NPV of 99.2 (95% CI 97.7–99.8%), sensitivity of 89.3% (95% CI 71.8–97.7%), specificity 97.5% (95% CI 95.4–98.8%).

A validation study of Primary Immunodeficiency Disease (PID) diagnoses was conducted using U.S. Medicaid data by Hernandez-Trujillo and colleagues.[23] Using ICD-9-CM codes 279.06, 279.04, 279.05, and 279.012, with adjudication by medical chart review, the authors found a low PPV of 19.3% (95% CI 11.4–29.4%). Kobrynski and colleagues conducted a study of the prevalence of PID using similar ICD-9 diagnosis codes (279.0, 279.1, 279.2, 279.8, 279.9, 288.1, 288.2), however validation was not performed.[24] Resnick and colleagues also used similar codes for their study covering the epidemiology of PID; validation was also not performed.[25]

Two studies reported poor sensitivity for algorithms to identify neutropenia.[26,27] In 2011, Kim and colleagues published a validation study utilizing administrative claims codes to identify neutropenia diagnosis.[26] The following ICD-9-CM codes were used: 288.00 (unspecified neutropenia), 288.03 (drug-induced neutropenia), 288.09 (other neutropenia), 288.5 (decreased white blood cell count), 288.50 (unspecified leukocytopenia), 288.59 (other decreased white blood cell count, 288.8 (other specified disease of white blood cells), 288.9 (unspecified disease of white blood cell). Outpatient laboratory data were used to determine true cases of neutropenia. The sensitivity of the three proposed algorithms tested were very poor (1-9%) while the specificity was excellent (99-100%). The PPV range was 18-33% and NPV was 98% for all three algorithms.

In 2017, Knerr and colleagues assessed the incidence of neutropenia in patients receiving lung cancer chemotherapy using two different methods: administrative claim codes and laboratory results.[27] Record of any of the following ICD-9-CM codes was considered an occurrence of neutropenia: 288.0 (agranulocytosis), 288.00 (neutropenia), 288.03 (drug induced neutropenia), 288.5 (leukocytopenia), 288.8 (specified disease of WBC), 288.9 (unspecified disease of WBC). Designation of neutropenia status by laboratory results was considered the "gold standard". Administrative claim codes had low sensitivity (26.3%), but high specificity (97.0%) with a PPV of 63.3%, and NPV of 86.9%.

vii Dialysis

Dialysis was not included as an immunosuppressive category in publications by Greenberg and colleagues[1] or Sunesen and colleagues.[6]

Taneja and colleagues examined the predictive accuracy of administrative billing codes to identify dialysis patients who received peritoneal dialysis and hemodialysis.[28] The primary aim of the study was to differentiate between hemodialysis and peritoneal dialysis patients. The authors identified 233 patients with evidence of end-stage renal disease and receipt of dialysis from their list of healthcare diagnosis and procedure claim codes. The types of codes used were ICD-9-CM, ICD-9-CM Procedure, CPT-4 procedure, and HCPCS codes. All patients were verified to have at least one form of dialysis treatment (PPV=100%). The PPV for billing codes at a 30-day, 90-day and 180-day window around an index encounter was high for identifying hemodialysis patients (86.7-93.1%), but low for identifying peritoneal dialysis patients (34.9-67.4%).

D Immunocompromised Clinical Case Definition

Currently there is no universal clinical definition of immunosuppression.[1] This report adds to the diseases and conditions that Greenberg and colleagues developed from previously published classification schemes.[2-4] The additional diseases and conditions are from AHRQ, Sunesen and colleagues,[6] and input from clinicians and SMEs.

Prior validation exercises, including those reported by Greenberg and colleagues,[1] relied on clinician chart review to identify the presence of an immunosuppressive diagnosis or receipt of an immunosuppressive therapy.

E Immunocompromised Coding Algorithm

The aim of this review was to develop an algorithm to identify immunocompromised individuals that could be of potential interest following exposure to a biologic product. As informed by previously published studies, the workgroup has developed an algorithm that includes administrative claims codes for nine disease groups, three of which are viewed as immunosuppressive regardless of treatment status and six of which require accompanying treatment. Refer to **Appendix B** for the complete immunocompromised algorithm as of the date of this report. The algorithm will be revised periodically. To develop a comprehensive list of immunocompromised diagnosis codes for clinical consideration, all ICD-9-CM and ICD-10 codes for immunocompromised conditions were extracted from the 19 articles identified in the literature review (**Appendix A**). To expand the draft code list and sreflect current coding practice, ICD-10-CM diagnosis codes were generated from ICD-9-CM codes using forward-backward mapping via the Centers for Medicare and Medicaid Services (CMS) GEMs files. The expanded draft code list, which included ICD-9-CM and ICD-10-CM codes, was subsequently reviewed by clinical SMEs from IBM (TB), FDA CBER (GD), and Acumen. Specific decisions and assumptions related to construction of the algorithm are summarized in **Section F**. Overall, the clinical SMEs recommended the inclusion of additional codes or exclusion of codes from the expanded draft code list based on clinical relevance and optimizing the balance between specificity and sensitivity.

The proposed algorithm for the nine disease groups presented below are consistent with approaches identified in the peer-reviewed literature. Most categories (with the exception of transplants and related conditions) required at least two claims on different days for a positive identification, as informed by approaches observed in the literature — which found that performance improved when more than one code was required — and the clinical judgement that these chronic conditions would be associated with more than one claim in true cases. An exclusion criterion was proposed for individuals whose most recent claim was over six months before the index date (to be defined based on research question), based on the clinical judgement that immunocompromising conditions can resolve and may not result in immunosuppression for life. The algorithm for the nine disease groups is summarized as follows:

1) **HIV/AIDS:** ≥2 claims for ANY HIV/AIDS diagnostic or procedural codes.
    a. This category is viewed as immunocompromising regardless of treatment status (RX-independent)
    b. EXCLUDE cases where the most recent claim is MORE THAN 6 months prior to the index date.
2) **Hematologic Malignancy and Related Conditions**: ≥2 claims for ANY relevant diagnostic codes.
    a. This category is viewed as immunocompromising regardless of treatment status (RX-independent).
    b. EXCLUDE cases where the most recent claim is MORE THAN 6 months prior to the index date.
3) **Immune deficiencies (treatment-independent)**: ≥2 claims for ANY treatment-independent immune deficiency diagnostic codes.
    a. This category is viewed as immunocompromising regardless of treatment status (RX-independent).
    b. EXCLUDE cases where the most recent claim is MORE THAN 6 months prior to the index date.
4) **Immune deficiencies (treatment-dependent):** ≥2 claims for ANY treatment-dependent immune deficiency diagnostic codes AND one of:
    a. ≥2 claims for ANY codes for chemotherapy or radiation
    b. ≥2 claims for ANY codes for immune-modulating or immunosuppressant therapy
    c. EXCLUDE cases where the most recent claim is MORE THAN 6 months prior to the index date.

    NOTE: Codes for brachytherapy (internal radiation via the insertion of implants) may also be considered by users seeking to apply a more inclusive definition of radiation therapy. It is recommended that this category be ignored in regular applications of the algorithm, as brachytherapy is less likely to be immunosuppressive, but it may be of value to those with a specific interest in brachytherapy.
5) **Solid Malignancy:** ≥2 claims for ANY solid malignancy diagnostic codes AND one of:
    a. ≥2 claims for ANY codes for chemotherapy or radiation
    b. ≥2 claims for ANY codes for immune-modulating or immunosuppressant therapy
    c. EXCLUDE cases where the most recent claim is MORE THAN 6 months prior to the index date.

    NOTE: Codes for brachytherapy may also be considered by users seeking to apply a more inclusive definition of radiation therapy. It is recommended that this category be ignored in regular applications of the algorithm, but it may be of value to those with a specific interest in brachytherapy.
6) **Transplant and Related Conditions:** ≥1 claims for ANY transplant or related condition diagnostic codes AND one of:
    a. ≥2 claims for ANY codes for chemotherapy or radiation
    b. ≥2 claims for ANY codes for immune-modulating or immunosuppressant therapy
    c. EXCLUDE cases where the most recent claim is MORE THAN 6 months prior to the index date.

    NOTE: Codes for brachytherapy may also be considered by users seeking to apply a more inclusive definition of radiation therapy. It is recommended that this category be ignored in regular applications of the algorithm, but it may be of value to those with a specific interest in brachytherapy.
7) **Rheumatologic/Inflammatory:** ≥2 claims for ANY rheumatologic or inflammatory condition diagnostic codes AND one of:
    a. ≥2 claims for ANY codes for chemotherapy or radiation
    b. ≥2 claims for ANY codes for immune-modulating or immunosuppressant therapy
    c. ≥2 claims for ANY steroid diagnosis codes
    c. A 90-day medication possession ratio (MPR) ≥75% for ANY steroid prescription codes
    d. EXCLUDE cases where the most recent claim is MORE THAN 6 months prior to the index date.

NOTE: Codes for brachytherapy may also be considered by users seeking to apply a more inclusive definition of radiation therapy. It is recommended that this category be ignored in regular applications of the algorithm, but it may be of value to those with a specific interest in brachytherapy.
NOTE: It should be indicated that steroid MPR is based on an Intention to Treat (ITT) analysis, as the ratio depends on days' supply reported in a claim and does not account for patient compliance. For example, receiving a 90-day steroid prescription would automatically put a patient above the MPR ≥75% threshold, albeit with no guarantee of steroid use.

8) **Dialysis:** ≥2 claims for ANY dialysis diagnostic or procedural codes
    a. EXCLUDE cases where most recent claim is MORE THAN 6 months prior to the index date.
9) **Intermediate conditions** ≥2 claims for ANY intermediate condition diagnostic codes AND one of:
    a. ≥2 claims for ANY codes for chemotherapy or radiation
    b. ≥2 claims for ANY codes for immune-modulating or immunosuppressant therapy
    c. EXCLUDE cases where treatment claim predates diagnostic claim OR where most recent claim is MORE THAN 6 months prior to the index date.

    NOTE: These conditions were viewed as unlikely to be immunocompromising but were included for optional use by users seeking to be as inclusive as possible.
    NOTE: Codes for brachytherapy may also be considered by users seeking to apply a more inclusive definition of radiation therapy. It is recommended that this category be ignored in regular applications of the algorithm, but it may be of value to those with a specific interest in brachytherapy.

F Assumptions and Decisions

The algorithm presented in **Section E** was reviewed internally as well as with CBER stakeholders and partners. Decisions and assumptions related to algorithm construction are summarized below. Some of these assumptions may be adjusted for future research questions.

F1 General

- Algorithms identified in the literature applied different restrictions on code lists, diagnosis coding position (e.g., principal position codes only), healthcare settings. An inclusive approach to code list development has been taken here, while further tailoring and restriction of settings and risk windows are likely better determined at the study planning stage.
- The proposed algorithm was created for use in administrative claims datasets. If there are data sources that provide information on medication dosage, users can consider adjusting criteria.
- Given the changes made to the Greenberg algorithm specifics, which include mapping ICD-9-CM to ICD-10-CM, adding new codes within categories, and adding entirely new categories, the estimates of category-specific and overall algorithm performance reported in prior studies should not be viewed as directly transferable.
- Chronicity of therapy was only considered for steroid use, but users could consider a similar dimension for other treatment categories. Similarly, only individuals with a relevant claim in the past six months have been considered as immunocompromised based on the clinical judgement that immunosuppression can be transient. Users may wish to tailor this restriction to their research purposes.
- The decision was made to not stratify codes based on the level of immunosuppression risk (high, intermediate, low), as this would entail subjective clinical assessments that could vary broadly across users. Users have the option of stratifying codes if appropriate to their research question.
- The nine categories of immunocompromising conditions are meant to be mutually exclusive. Users may prefer to exclude some categories of conditions as the appropriateness of certain categories may depend on the research question.

- NDC codes were drawn from the FDA NDC Database (https://www.fda.gov/drugs/drug-approvals-and-databases/national-drug-code-directory), last updated July 21, 2020. HCPCS codes were drawn from https://www.hcpcsdata.com/, last updated July 21, 2020. These codes may not include all relevant codes and may be or become outdated. Users should review the most current datasets available prior to using the algorithm.

F2 HIV/AIDS

- Codes for pneumocystosis (ICD-9-CM 136.3; ICD-10-CM B59) were not included, as this is not exclusively associated with HIV (e.g., can also be due to organ transplantation), and the causative process (HIV, organ transplant) is likely to be coded anyway.
- Codes for nonspecific serologic evidence (ICD-9-CM 795.71) and inconclusive laboratory evidence (ICD-10-CM) were excluded as it was decided that these did not provide sufficient evidence of HIV. Both were queried in a large administrative claims database and produced very small counts.
- Recognizing that most HIV-positive individuals on highly active antiretroviral therapy are not immunocompromised, users may want to consider adding a requirement for either non-adherence to treatment or comorbid opportunistic infection, which could serve as a proxy for identifying immunosuppression. This will depend on the research question and information available to the user.
- Individuals with HIV who receive proper treatment may have a lower risk of immunosuppression relative to those who do not. A list of HIV-specific medications was not generated, based on the clinical judgement that this would be of little value in developing a sensitive algorithm to identify broad groups of immunocompromised individuals, which is not stratified on the risk of immunosuppression. Users with a specific interest in HIV/AIDS may consider adding a treatment component to identify individuals at reduced risk of immunosuppression. HIV-positive individuals are commonly on three or more medications concurrently, which might also need to be taken into consideration if a treatment component is to be added.

F3 Hematologic Conditions and Related Conditions

- Clinicians discussed whether codes for polycythemia vera (ICD-10-CM D45) should be included. On the basis of their inclusion in AHRQ criteria for individuals at high risk of immunosuppression and that other myeloproliferative neoplasms are also included, these codes were retained.
- Clinicians discussed whether codes for hematologic conditions that were in remission should be included. As informed by prior AHRQ criteria for immunocompromised individuals, and out of a desire to include this at-risk cohort (which should also not be considered fully immunocompetent), these codes were included.

F4 Immune Deficiencies

- The inclusion of diagnosis codes for chronic hepatitis B was discussed among clinicians. It was decided that individuals receiving treatment for hepatitis B would be included (i.e., only immunocompromised if receiving treatment). It is recognized that many with hepatitis B will not be immunocompromised. An additional limitation is that a list of specific antiretroviral medications relevant to hepatitis B treatment has not been developed for the proposed algorithm, with the condition instead being paired with more general treatments (e.g., interferon). Users with a specific interest in hepatitis may therefore want to specify the algorithm further.
- ICD-9-CM 255.41 (glucocorticoid deficiency) was excluded. Though it maps approximately to the ICD-10-CM code for Addisonian crisis, it was viewed as too general to include. In addition, treatment is unlikely to vary depending on the mechanism (i.e., autoimmune vs non-autoimmune),

meaning assigning a treatment-dependence criterion would not be a useful qualifier for identifying immunocompromised individuals.

- ICD-9-CM 287.1 and ICD-10-CM D69.1 (qualitative platelet defects) were excluded on the basis of the clinical judgement that these are congenital in nature. Congenital and hereditary thrombocytopenia purpura (ICD-9-CM 287.33/ICD-10-CM D69.42), primary thrombocytopenia, unspecified (ICD-9-CM 287.30), and other primary thrombocytopenia (ICD-9-CM 287.39/ICD-10-CM D69.49) were also excluded as congenital conditions.
- While it was noted that celiac disease (ICD-9-CM 579.0/ICD-10-CM K90.0) was included in a previous paper,[6] this was excluded here on the basis of the clinical judgement that celiac disease was not autoimmune in nature.
- Codes for myositis (ICD-9-CM 728.xx/ICD-10-CM M60.xxx) were excluded on the basis of the clinical opinion that the codes were too non-specific and were most likely to be acquired and not immune-related.
- Codes for vasculitis limited to skin, not elsewhere classified (ICD-9-CM 709.1/ICD-10-CM L95.x) were excluded on the basis of the clinical opinion that only a very small proportion will require immunosuppressive therapy.
- Codes for myopathy from external/toxic causes (ICD-9-CM 359.3; 359.4, 359.81/ICD-10-CM G72.1–.3 and G72.81) were excluded.

F5 Solid Malignancy

- On the basis of clinical consultation, codes for benign carcinoid tumors (ICD-9-CM 209.4x, 209.5x, 209.6x) were excluded.
- On the basis of clinical consultation, codes for hypertensive chronic kidney disease (ICD-9-CM 403.xx), hypertensive heart and chronic kidney disease (ICD-9-CM 404.xx), other and unspecified postsurgical non-absorption (ICD-9-CM 579.3), chronic kidney disease, Stage V (ICD-9-CM 585.5), and end stage renal disease (ICD-9-CM 585.6) were excluded. While some of these individuals may be immunocompromised, the focus is primarily on dialysis-related codes in the context of these conditions, which are captured elsewhere. These codes were therefore excluded due to concerns that the majority of individuals with these codes are unlikely to be immunocompromised.

F6 Transplant and Related Conditions

- Presumed immunocompetent: the clinical opinion was that some forms of organ transplant (specifically skin, bone, corneal and heart valve transplants) were less likely to result in immunosuppression. These have been classified in a separate 'Presumed Immunocompetent' category, to allow testing of their inclusion via sensitivity analysis. Transplantation of other/unspecified organs were kept in the main code list.
- Exclude: Z49 codes were excluded as they were not sufficiently specific to transplant recipients. In addition, transplant patients with these codes were likely to also have a transplant-specific code.
- The inclusion of codes for atherosclerosis (ICD-9-CM 440.xx; ICD-10-CM I25.xxx) and transplant status (ICD-9-CM V42.x; ICD-10-CM Z94.x) was discussed among SMEs. These were ultimately included as, though they do not themselves represent a condition likely to result in immunosuppression, they represent individuals who have received a transplant.

F7 Rheumatologic/Inflammatory Conditions

- Codes for infective arthritis codes (ICD-9-CM 711.0x, 711.9x; ICD-10-CM M00.xx, M01.xx) were excluded, based on the clinical judgement that these are unlikely to represent immunocompromised individuals.

F8 Dialysis

- No additional assumptions or decisions were noted.

F9 Intermediate Conditions

- Two codes were discussed among SMEs, but ultimately excluded:
    - ICD-10-CM B52.0 (Plasmodium malariae malaria with nephropathy: while patients may develop splenomegaly, it is a reactive process and is rare because it is associated with increased mortality
    - R76.0 (raised antibody titer): SMEs determined that the code was too generic to be meaningful

G Algorithm Characterization

G1 Methods

To summarize the epidemiology of immunocompromised individuals among a commercially insured study population in the U.S., the workgroup calculated descriptive statistics of the diagnostic codes included in the immunocompromised algorithm in the IBM MarketScan Research Databases (Commercial and Medicare Supplemental). To gather the broadest range of cases to support a descriptive analysis, the analyses presented herein did not require exposure to a biologic product and did not restrict based on coding position. It is recommended that the proposed algorithm undergo a validation study prior to use, though future analytical studies should also tailor the algorithm specifications according to the study question of interest.

Counts of individual patients who had a diagnosis code related to an immunocompromising disease category within a given query period, rather than counts of particular codes, were presented. As such, counts relate to the first diagnosed immunocompromised condition event for an individual during a given surveillance period (e.g., January 1, 2014, and December 31, 2014), and individuals could only be counted once per surveillance period for each disease category. However, disease categories were queried independently, so individuals could be included in multiple categories. Since we did not estimate the incidence of immunocompromised conditions in the study population, no washout period was applied.

Individuals had to be enrolled for at least one day in the 2014-2018 period, and for 365 days following the first diagnosis of interest to be included in the analysis for a particular year. Age is calculated in Treatment Pathways as if each individual was born on July 1 of their given year of birth. Out of concern that the minimum continuous enrollment requirement could impact the inclusion of infants (i.e., those under one year old), this population group has been left out of the two charts that depict the proportions of immunocompromised individuals. Infants under one year of age were not excluded from queries of the absolute number of patients receiving a diagnosis related to an immunocompromising disease category.

Age- and gender-specific data on MarketScan Research Databases enrollment and counts of individuals receiving a diagnosis code for each immunocompromising disease categories were extracted. Authors executed queries that aggregated all ICD-9-CM and ICD-10-CM for each immunocompromising disease

category (excluding “Intermediate Conditions”). The figures presented below have been drawn from a large patient dataset during the study period of January 1, 2014–December 31, 2018.

G2 Results

Table 3 provides a summary of the eight immunocompromising categories queried.

Table 3. Summary of immunocompromised cohort algorithm categories, MarketScan Research Databases (2014–2018).

Parameter	Enrolled Population	HIV/AIDS	Heme Malignancies	Immune Deficiencies (RX_Indep)	Immune Deficiencies (RX-Dep)	Solid Malignancies	Trans-plants	Rheum & Inflamm	Dialysis
Population Size – N (% of Enrolled)	65,110,796 (100.0)	59,455 (0.1)	138,235 (0.2)	81,086 (0.1)	260,062 (0.4)	246,366 (0.4)	62,883 (0.1)	354,380 (0.5)	32,415 (0.1)
Mean Age at First Diagnosis – Years	NA	44	57	43	52	59	57	51	58
	Count (%)								
Age at First Diagnosis (Years) – N (% of Category)									
0–17	17,478,862 (26.8)	306 (0.5)	4,428 (3.2)	14,776 (18.2)	7,603 (2.9)	2,499 (1.0)	2,124 (3.4)	11,645 (3.3)	326 (1.0)
18–34	18,590,899 (28.6)	12,852 (21.6)	9,292 (6.7)	10,698 (13.2)	25,737 (9.9)	8,111 (3.3)	4,204 (6.7)	44,797 (12.6)	1,868 (5.8)
35–44	9,626,273 (14.8)	13,085 (22.0)	11,615 (8.4)	9,464 (11.7)	37,213 (14.3)	19,537 (7.9)	5,642 (9.0)	49,477 (14.0)	3,079 (9.5)
45–54	9,793,635 (15.0)	20,381 (34.3)	25,675 (18.6)	14,581 (18.0)	63,569 (24.4)	50,280 (20.4)	11,345 (4.4)	80,838 (22.8)	6,334 (19.5)
55–64	7,037,118 (10.8)	11,644 (19.6)	45,017 (32.6)	19,651 (24.2)	81,922 (31.5)	91,325 (37.1)	19,916 (18.0)	103,753 (29.3)	9,932 (30.6)
65+	2,584,009 (4.0)	1,187 (2.0)	42,208 (30.5)	11,916 (14.7)	44,018 (16.9)	74,614 (30.3)	19,652 (31.7)	63,870 (18.0)	10,876 (33.6)
Gender (Female) – N (% of Category)	33,576,048 (51.6)	10,415 (17.5)	62,085 (44.9)	49,642 (61.2)	159,345 (61.3)	151,684 (61.5)	28,614 (31.3)	240227 (67.8)	13,135 (40.5)
Region – N (% of Category)									
Northeast	11,793,510 (18.1)	10,602 (17.8)	30,545 (22.1)	16,182 (20.0)	48,973 (18.8)	50,857 (20.6)	13,298 (21.1)	66,289 (18.7)	5,389 (16.6)
Midwest	13,470,458 (20.7)	6,832 (11.5)	35,200 (25.5)	17,676 (21.8)	60,661 (23.3)	57,835 (23.5)	15,758 (25.1)	85,130 (24.0)	8,674 (26.8)
South	27,017,625 (41.5)	32,772 (55.1)	52,490 (38.0)	35,140 (43.3)	112,611 (43.3)	102,816 (41.7)	25,084 (39.9)	151,797 (42.8)	14,405 (44.4)
West	11,776,544 (18.1)	9,139 (15.4)	19,751 (14.3)	11,922 (14.7)	37,310 (14.3)	34,423 (14.0)	8,640 (13.7)	50,444 (14.2)	3,856 (11.9)
Missing	1,052,659 (1.6)	110 (0.2)	249 (0.2)	166 (0.2)	507 (0.2)	435 (0.2)	103 (0.2)	720 (0.2)	91 (0.3)

Abbreviations: AIDS, Acquired immunodeficiency syndrome; HIV, Human immunodeficiency virus; N, Number; RX-Indep, Treatment-independent; RX-Dep, Treatment-dependent

As shown in **Table 3**, rheumatologic and inflammatory conditions (n=345,380; 0.53%), treatment-dependent immune deficiencies (n=260,062; 0.40%), and solid malignancies (n=246,366; 0.38%) were the most common immunocompromised population categories. As shown in **Figure 1**, absolute numbers of individuals meeting the criteria for each immunocompromised category generally increased with age and were highest in the 55–64 and 65+ age groups.

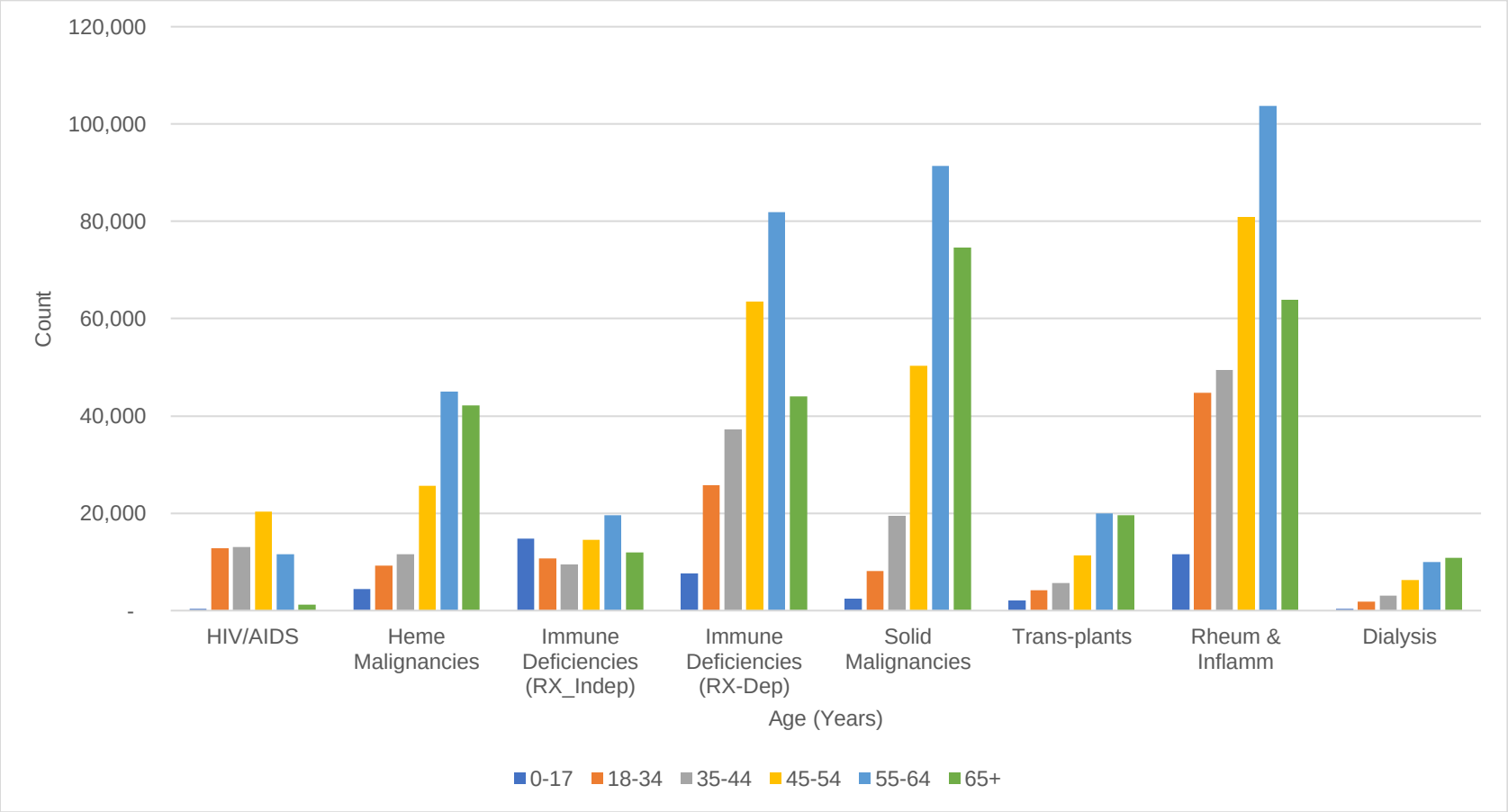

Figure 1. Patients meeting criteria for at least one immunocompromised category in the MarketScan Research Databases, January 1, 2014–December 31, 2017, stratified by age group.

Abbreviations: AIDS, Acquired immunodeficiency syndrome; HIV, Human immunodeficiency virus; RX-Indep, Treatment-independent; RX-Dep, Treatment-dependent

Figures 2–9 present the proportions of patients (aged 1–85+ years) enrolled in the MarketScan Research Databases for at least one day that met the criteria for each of the immunocompromised categories, by age and gender. Patients 85 years or older were grouped to minimize the effect of unstable estimates due to the smaller enrolled population sizes available in this age range in the commercially insured population.

- HIV/AIDS (**Figure 2**) was higher in men than women, with proportions highest among men 40–60 years old.
- Hematologic malignancies and related conditions (**Figure 3**) were distributed fairly evenly between genders until later years, with proportions highest among men over 65 years of age.
- Treatment-independent (**Figure 4**) and treatment-dependent (**Figure 5**) immune deficiencies were more common among women than men between the ages of 25 and 65 years.
- Solid malignancies (**Figure 6**) were distributed fairly evenly between genders though were observed to be higher among women than men in adulthood (35–65 years) and higher in senior men than women (70+ years).
- Transplants and related conditions (**Figure 7**) were higher among men than women starting from about 45 years of age.
- Rheumatologic and inflammatory conditions (**Figure 8**) were higher among women than men across age groups.
- Dialysis (**Figure 9**) was more common among men than women starting from about 45 years of age.

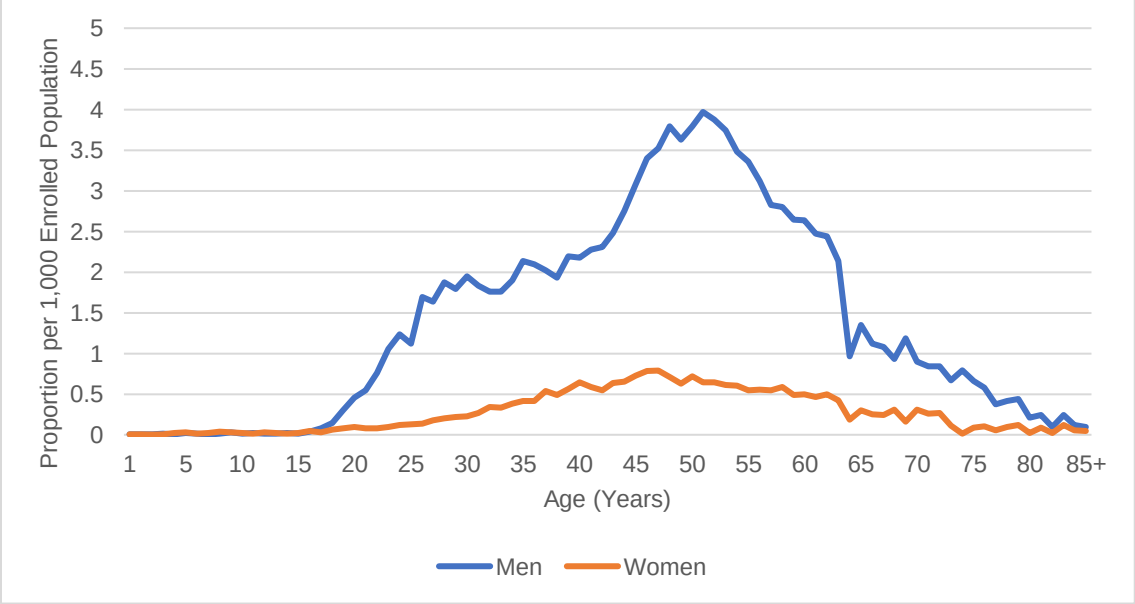

Figure 2. Age-specific proportion of patients (1–85+)* meeting criteria for HIV/AIDS category in the MarketScan Research Databases, per 1,000 enrolled population, by age and gender (January 1, 2014–December 31, 2018).

*Out of concern that the minimum continuous enrollment requirement could impact the inclusion of infants (i.e., those under 1 year old), the proportion of those under 1-year old is excluded from the chart.

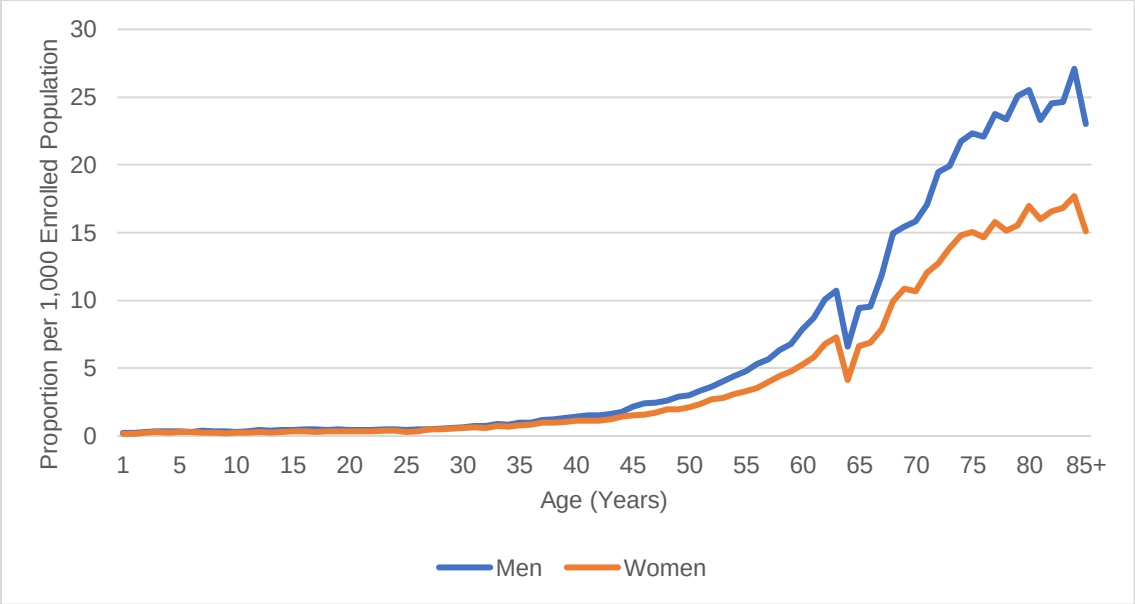

Figure 3. Age-specific proportion of patients (1–85+)* meeting criteria for Hematologic Malignancies & Related Conditions category in the MarketScan Research Databases, per 1,000 enrolled population, by age and gender (January 1, 2014–December 31, 2018).

*Out of concern that the minimum continuous enrollment requirement could impact the inclusion of infants (i.e., those under 1 year old), the proportion of those under 1-year old is excluded from the chart.

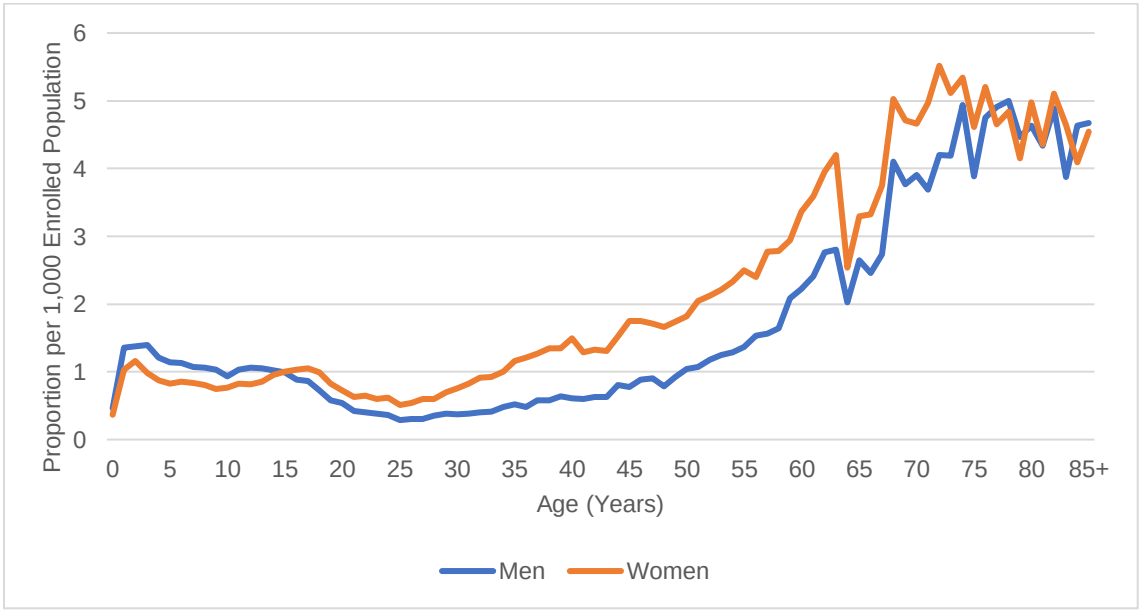

Figure 4. Age-specific proportion of patients (1–85+)* meeting criteria for Immune Deficiencies (Treatment-Independent) category in the MarketScan Research Databases, per 1,000 enrolled population, by age and gender (January 1, 2014–December 31, 2018).

*Out of concern that the minimum continuous enrollment requirement could impact the inclusion of infants (i.e., those under 1 year old), the proportion of those under 1-year old is excluded from the chart.

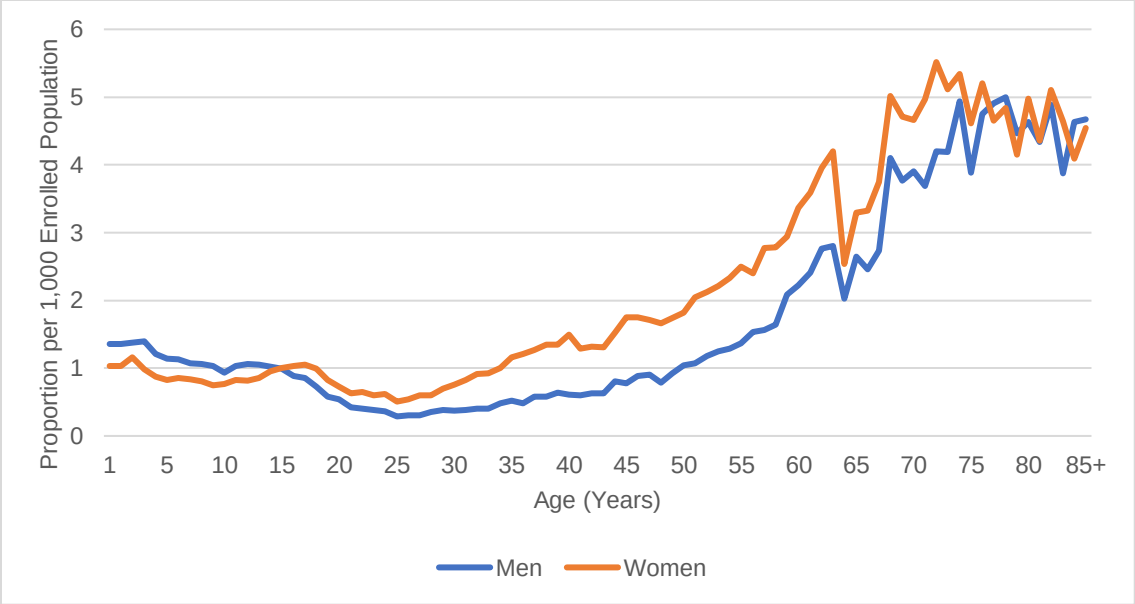

Figure 5. Age-specific proportion of patients (1–85+)* meeting criteria for Immune Deficiencies (Treatment-Dependent) category in the MarketScan Research Databases, per 1,000 enrolled population, by age and gender (January 1, 2014–December 31, 2018).

*Out of concern that the minimum continuous enrollment requirement could impact the inclusion of infants (i.e., those under 1 year old), the proportion of those under 1-year old is excluded from the chart.

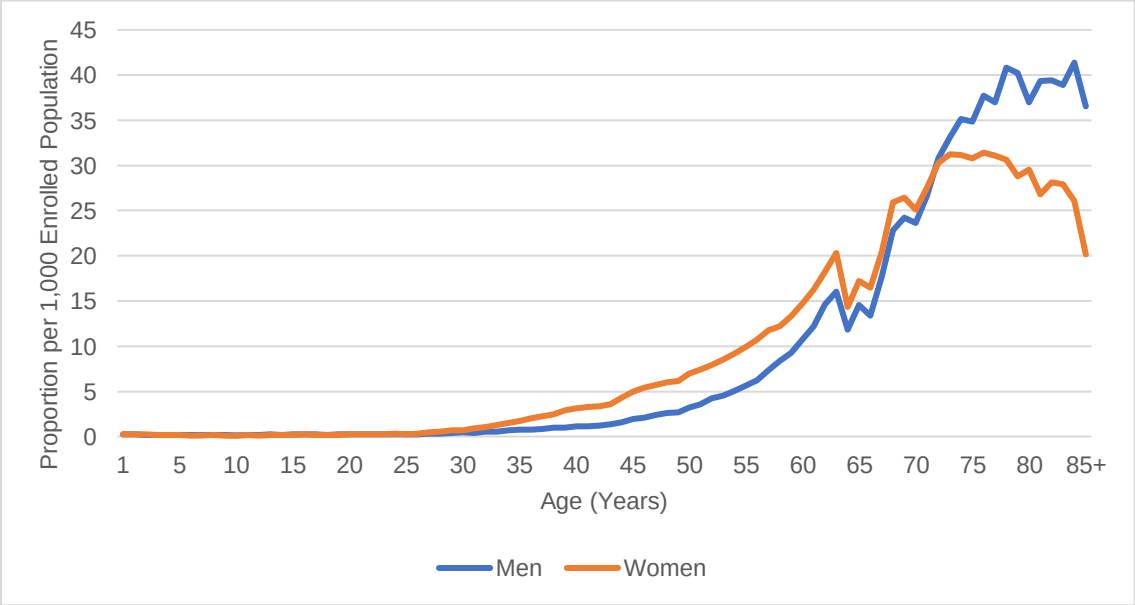

Figure 6. Age-specific proportion of patients (1–85+)* meeting criteria for Solid Malignancies category in the MarketScan Research Databases, per 1,000 enrolled population, by age and gender (January 1, 2014–December 31, 2018).

*Out of concern that the minimum continuous enrollment requirement could impact the inclusion of infants (i.e., those under 1 year old), the proportion of those under 1-year old is excluded from the chart.

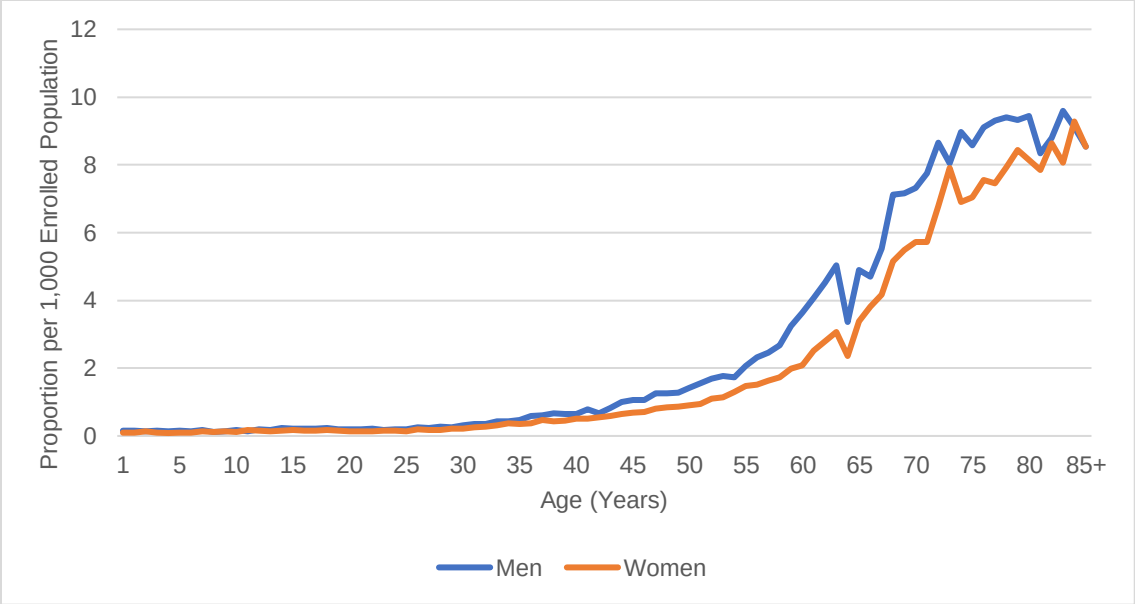

Figure 7. Age-specific proportion of patients (1–85+)* meeting criteria for Transplants and Related Conditions category in the MarketScan Research Databases, per 1,000 enrolled population, by age and gender (January 1, 2014–December 31, 2018).

*Out of concern that the minimum continuous enrollment requirement could impact the inclusion of infants (i.e., those under 1 year old), the proportion of those under 1-year old is excluded from the chart.

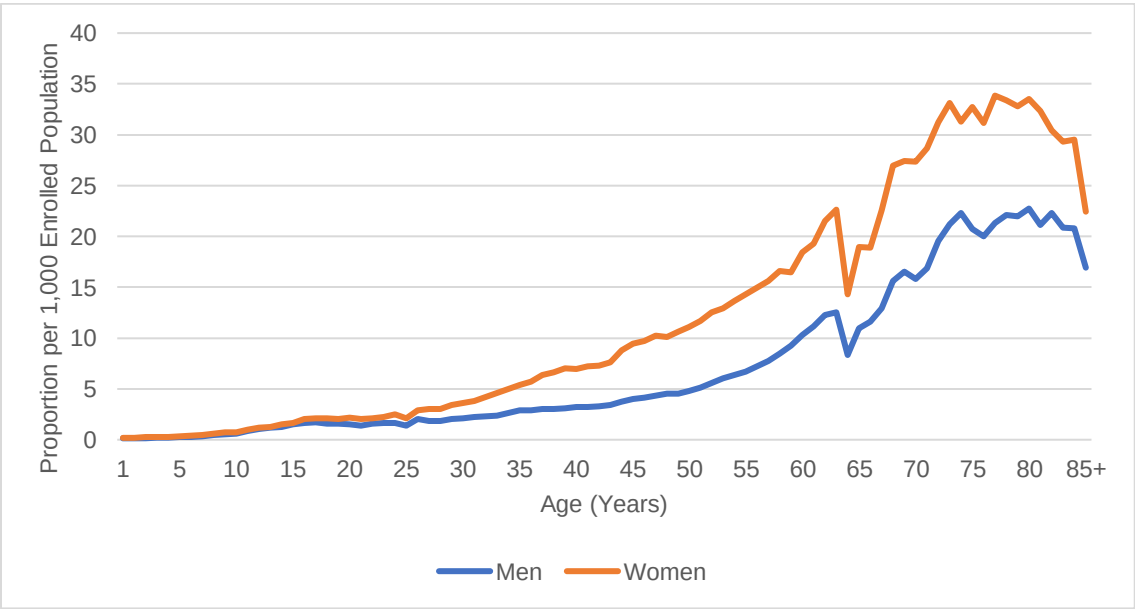

Figure 8. Age-specific proportion of patients (1–85+)* meeting criteria for Rheumatologic & Inflammatory Conditions category in the MarketScan Research Databases, per 1,000 enrolled population, by age and gender (January 1, 2014–December 31, 2018).

*Out of concern that the minimum continuous enrollment requirement could impact the inclusion of infants (i.e., those under 1 year old), the proportion of those under 1-year old is excluded from the chart.

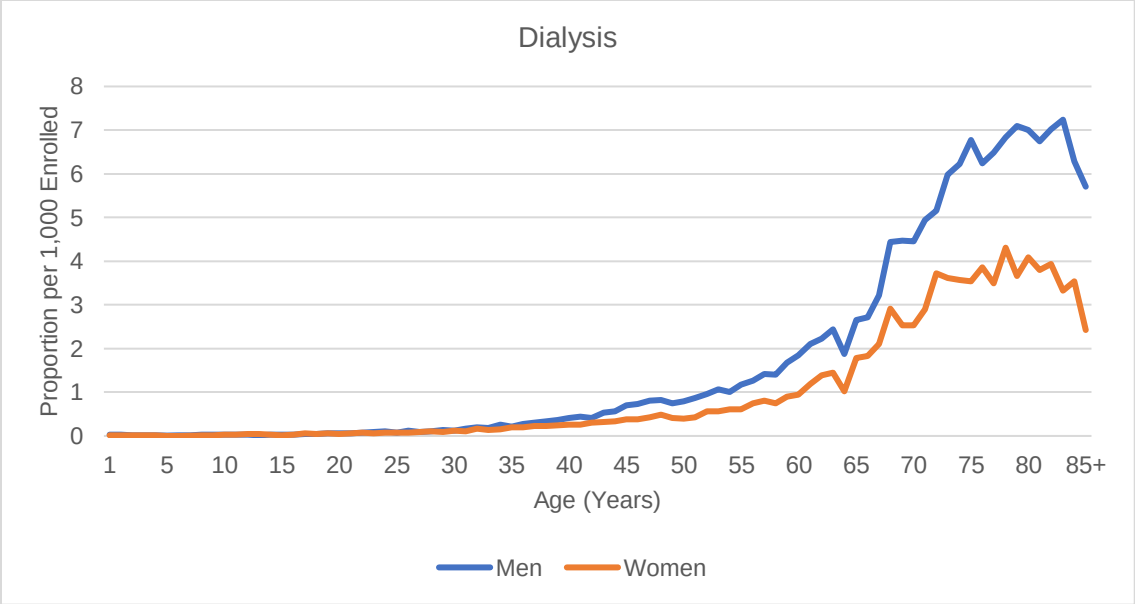

Figure 9. Age-specific proportion of patients (1–85+)* meeting criteria for Dialysis category in the MarketScan Research Databases, per 1,000 enrolled population, by age and gender (January 1, 2014–December 31, 2018).

*Out of concern that the minimum continuous enrollment requirement could impact the inclusion of infants (i.e., those under 1 year old), the proportion of those under 1-year old is excluded from the chart.

Figure 10 presents the proportions of patients (aged 1–85+ years) enrolled in the MarketScan Research Databases for at least one day that met the criteria for at least one of the immunocompromised categories, by age. Patients 85 years or older were grouped to minimize the effect of unstable estimates due to the smaller enrolled population sizes available in this age range in the commercially insured population. Results suggest that proportions of patients meeting criteria for immunosuppression increased with age across categories of interest (except for HIV/AIDS), with peaks in the 65 years and older aged cohorts. The proportions of individuals meeting criteria for the immunocompromised categories decreases at 65 years of age, though this may be due to shifts in enrollment as individuals move off of commercial insurance plans, and proportions begin to increase again after 65 years of age.

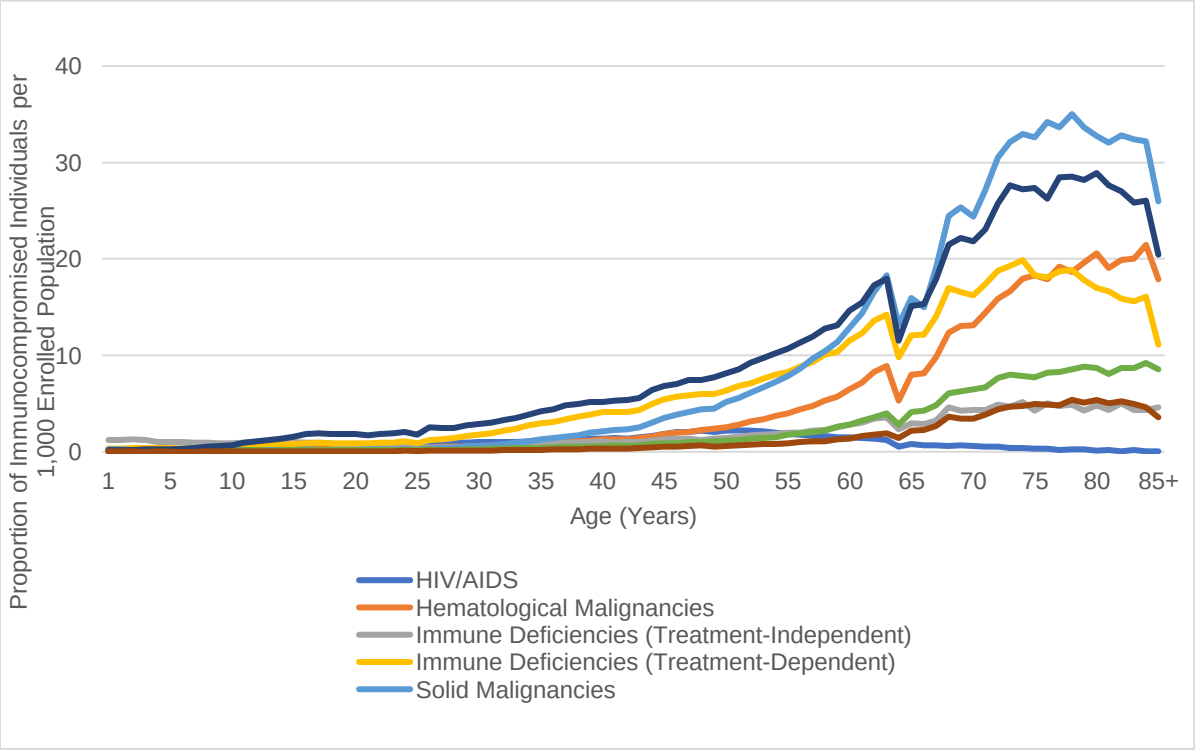

Figure 10. Age-specific proportion of patients (1–85)* meeting criteria for at least one immunocompromised category in the MarketScan Research Databases, per 1,000 enrolled population, by age (January 1, 2014–December 31, 2018).

*Out of concern that the minimum continuous enrollment requirement could impact the inclusion of infants (i.e., those under 1 year old), the proportion of those under 1-year old is excluded from the chart.

H Discussion and Conclusion

The objective of this structured literature review was to understand and assess the validity of algorithms for identifying immunocompromised individuals from administrative claims and EHRs/EMRs using billing codes. It is unclear how diagnostic code-based algorithms would perform differently in EHR compared to claims databases, beyond the differences that already occur within different databases of either EHR or claims. The immunocompromising disease diagnostic code-based algorithms validated in EHR were assessed as supplemental data to support the completeness of the code lists in this report. A structured review found 19 studies judged to be of particular relevance. None reported using ICD-10-CM codes to identify potential immunocompromised individuals. One U.S. EHR-based algorithm validation study reported strong measures of validation performance — PPV of 94.4% (95% CI 88.8–97.7), NPV of 94.3% (95% 91.0-96.6), sensitivity of 87.4% (95% CI 80.6–92.5%) and specificity of 97.6% (95% CI 95.0–99.9%) — associated with six disease categories (three treatment-independent, three treatment-independent) to identify immunocompromised individuals.[1] Based on literature review, past CBER efforts and extensive consultation with clinical SMEs, an algorithm was developed that could be applied to identify immunocompromised individuals using nine mutually exclusive disease categories.

The algorithm was then applied in the MarketScan Research Databases to test the feasibility of algorithm use and conduct initial analyses describing the epidemiology of immunocompromised individuals in a U.S. database of commercially insured patients.

Rheumatological and inflammatory conditions, treatment-dependent immune deficiencies, and solid malignancies were the most commonly reported categories, while dialysis, HIV/AIDS, and transplants and related conditions were the least common. Summing counts of individuals meeting the criteria for each of the eight disease categories led to a total of 1,234,882 individuals (1.9% of the population enrolled for at least one day between January 1, 2014, and December 31, 2018). This is lower than prior estimates, such as a 2013 estimate that 2.7% (95% CI 2.4–2.9%) of the U.S. population was currently immunosuppressed.[5] However, queries reported herein required that individuals be enrolled for 365 days after their first diagnosis, which may have lowered estimates. Also, the “intermediate conditions” category was not applied, which may have decreased counts. Conversely, summing of individuals across disease categories risks potential duplication of individuals across multiple categories, which could have increased estimates.

The average age at first diagnosis for the immunocompromised categories queried ranged from 43 years (treatment-independent immune deficiencies) to 59 years (solid malignancies). The proportion of enrolled individuals diagnosed with an immunosuppression condition increased with age across categories. This is consistent with clinical expectations and past research that found the prevalence of U.S. immunosuppression was highest among those 50–59 years of age.[5]

Women were overrepresented in categories related to treatment-dependent and treatment-independent immune deficiencies, solid malignancies, and rheumatologic and inflammatory conditions, while men were over-represented in categories related to HIV/AIDS, hematologic malignancies and related conditions, and transplants and related conditions, and dialysis. The over-representation of men for HIV/AIDS (82.5% of those meeting criteria) was particularly striking, but is consistent with a Canadian study reporting that 81.2% of HV-infected patients identified between 2007 and 2009 were male.[13] Similarly, the U.S. CDC reported that 81% of new HIV diagnoses in 2017 were in men.[29]

Strengths of this study are the development of an immunocompromised algorithm using the most current diagnostic, procedural and prescription code standards, based on a structured review of code definitions and active engagement with clinical SMEs. To assess the plausibility of the algorithm, it was applied in a large administrative claims database to characterize immunocompromised cohorts in the commercially insured U.S. population and generate descriptive statistics. The study also includes important limitations that should be considered in interpreting findings. Many decisions were based on the subjective judgement of CBER and IBM clinicians, as the scope of the algorithm was too large for every decision to be justified with reference to the existing literature and knowledge base. Further, the analyses conducted

in the MarketScan Research Databases should be viewed as exploratory and generalizable to the U.S. population that is commercially insured, and additional studies among populations with different insurance coverage would be required to validate the results and observations stemming from these queries.

I Acknowledgements

Development of the immunocompromised cohort algorithm and report benefitted from significant engagement with the FDA CBER team members and their partners. We thank them for their contributions and feedback. Additional feedback on the proposed algorithm and draft report was provided by IBM Watson Health, Acumen (Laurie Feinberg, Nirmal Choradia) and Epi Excellence LLC.

J References

1. Greenberg JA, Hohmann SF, Hall JB, Kress JP, David MZ. Validation of a Method to Identify Immunocompromised Patients with Severe Sepsis in Administrative Databases. *Annals of the American Thoracic Society.* 2016;13(2):253-258.
2. Rubin LG, Levin MJ, Ljungman P, et al. 2013 IDSA clinical practice guideline for vaccination of the immunocompromised host. *Clin Infect Dis.* 2014;58(3):309-318.
3. Tolsma V, Schwebel C, Azoulay E, et al. Sepsis severe or septic shock: outcome according to immune status and immunodeficiency profile. *Chest.* 2014;146(5):1205-1213.
4. Poutsiaka DD, Davidson LE, Kahn KL, Bates DW, Snydman DR, Hibberd PL. Risk factors for death after sepsis in patients immunosuppressed before the onset of sepsis. *Scand J Infect Dis.* 2009;41(6-7):469-479.
5. Harpaz R, Dahl RM, Dooling KL. Prevalence of Immunosuppression Among US Adults, 2013. *Jama.* 2016;316(23):2547-2548.
6. Sunesen KG, Nørgaard M, Thorlacius-Ussing O, Laurberg S. Immunosuppressive disorders and risk of anal squamous cell carcinoma: a nationwide cohort study in Denmark, 1978-2005. *Int J Cancer.* 2010;127(3):675-684.
7. AHRQ. Pediatric quality indicators technical specifications. https://www.qualityindicators.ahrq.gov/Modules/PDI_TechSpec_ICD10_v2020.aspx. Published 2020. Accessed April 28, 2021.
8. Greenberg JA, Hohmann SF, James BD, et al. Hospital Volume of Immunosuppressed Patients with Sepsis and Sepsis Mortality. *Annals of the American Thoracic Society.* 2018;15(8):962-969.
9. Carvalho HA, Villar RC. Radiotherapy and immune response: the systemic effects of a local treatment. *Clinics (Sao Paulo).* 2018;73(suppl 1):e557s.
10. Altieri P, Sau G, Cao R, et al. Immunosuppressive treatment in dialysis patients. *Nephrology Dialysis Transplantation.* 2002;17(suppl_8):2-9.
11. Janus N, Vacher L-V, Karie S, Ledneva E, Deray G. Vaccination and chronic kidney disease. *Nephrology Dialysis Transplantation.* 2007;23(3):800-807.
12. Kato S, Chmielewski M, Honda H, et al. Aspects of immune dysfunction in end-stage renal disease. *Clinical journal of the American Society of Nephrology: CJASN.* 2008;3(5):1526-1533.
13. Antoniou T, Zagorski B, Loutfy MR, Strike C, Glazier RH. Validation of case-finding algorithms derived from administrative data for identifying adults living with human immunodeficiency virus infection. *PLoS One.* 2011;6(6):e21748.
14. Herman RA, Gilchrist B, Link BK, Carnahan R. A systematic review of validated methods for identifying lymphoma using administrative data. *Pharmacoepidemiology and drug safety.* 2012;21 Suppl 1:203-212.
15. Setoguchi S, Solomon DH, Glynn RJ, Cook EF, Levin R, Schneeweiss S. Agreement of diagnosis and its date for hematologic malignancies and solid tumors between Medicare claims and cancer registry data. *Cancer Causes Control.* 2007;18(5):561-569.
16. Brandenburg NA, Phillips S, Wells KE, et al. Validating an algorithm for multiple myeloma based on administrative data using a SEER tumor registry and medical record review. *Pharmacoepidemiology and drug safety.* 2019;28(2):256-263.
17. Kavcic M, Fisher BT, Torp K, et al. Assembly of a cohort of children treated for acute myeloid leukemia at free-standing children's hospitals in the United States using an administrative database. *Pediatr Blood Cancer.* 2013;60(3):508-511.
18. Shayne M, Culakova E, Poniewierski MS, et al. Risk factors for in-hospital mortality and prolonged length of stay in older patients with solid tumor malignancies. *J Geriatr Oncol.* 2013;4(4):310-318.
19. Damoiseaux JG, Tervaert JW. The definition of autoimmune disease: are Koch's postulates applicable? *Neth J Med.* 2002;60(7):266-268.
20. Hayter SM, Cook MC. Updated assessment of the prevalence, spectrum and case definition of autoimmune disease. *Autoimmun Rev.* 2012;11(10):754-765.
21. Rose NR, Bona C. Defining criteria for autoimmune diseases (Witebsky's postulates revisited). *Immunol Today.* 1993;14(9):426-430.
22. Roberts MH, Erdei E. Comparative United States autoimmune disease rates for 2010-2016 by sex, geographic region, and race. *Autoimmun Rev.* 2020;19(1):102423.

23. Hernandez-Trujillo H, Orange JS, Roy JA, et al. Validity of Primary Immunodeficiency Disease Diagnoses in United States Medicaid Data. *J Clin Immunol.* 2015;35(6):566-572.
24. Kobrynski L, Powell RW, Bowen S. Prevalence and morbidity of primary immunodeficiency diseases, United States 2001-2007. *J Clin Immunol.* 2014;34(8):954-961.
25. Resnick ES, Bhatt P, Sidi P, Cunningham-Rundles C. Examining the use of ICD-9 diagnosis codes for primary immune deficiency diseases in New York State. *J Clin Immunol.* 2013;33(1):40-48.
26. Kim SY, Solomon DH, Liu J, Chang CL, Daniel GW, Schneeweiss S. Accuracy of identifying neutropenia diagnoses in outpatient claims data. *Pharmacoepidemiology and drug safety.* 2011;20(7):709-713.
27. Knerr S, Hu EY, Zeliadt SB. Incidence of Neutropenia in Veterans Receiving Lung Cancer Chemotherapy: A Comparison of Administrative Coding and Electronic Laboratory Data. *EGEMS (Washington, DC).* 2017;5(1):1269.
28. Taneja C, Berger A, Inglese GW, et al. Can dialysis patients be accurately identified using healthcare claims data? *Perit Dial Int.* 2014;34(6):643-651.
29. U.S. CDC. HIV and Men https://www.cdc.gov/hiv/group/gender/men/index.html. Published 2019. Accessed September 20, 2020.
30. Jacobson DL, Gange SJ, Rose NR, Graham NM. Epidemiology and estimated population burden of selected autoimmune diseases in the United States. *Clinical immunology and immunopathology.* 1997;84(3):223-243.
31. Eaton WW, Rose NR, Kalaydjian A, Pedersen MG, Mortensen PB. Epidemiology of autoimmune diseases in Denmark. *J Autoimmun.* 2007;29(1):1-9.

Appendix A. Literature Review Extracted Results

Table A1 below includes a summarized version of the data extraction table used to extract data from papers deemed of interest to immunocompromised cohort algorithms. The 19 papers summarized in the table informed the development of the proposed immunocompromised cohort algorithm.

Table A1. Immunocompromised Conditions Data Extraction Table

Author, Year	Title	Country	Summary	Disease Definition	Algorithm/Criteria	Validity	Claims/ EHR-Based[vii]
Antoniou, 2011	Validation of Case-Finding Algorithms Derived from Administrative Data for Identifying Adults Living with Human Immunodeficiency Virus Infection	Canada	Calculated sensitivity and specificity of algorithms used to identify HIV in patients over the age of 18 covered by the Ontario Health Insurance Plan (OHIP).	3 physician claims within 3 consecutive years, primary care chart review to confirm	ICD-9: 042, 043, 044 and ICD-10: B20 – B24	Sensitivity 96.2% (95% CI 95.2% - 97.9%), specificity 99.6% (95% CI 99.1% - 99.8%)	Claims
Brandenburg, 2018	Validating an algorithm for multiple myeloma based on administrative data using a SEER tumor registry and medical record review	USA	Validated different algorithms to identify multiple myeloma cases using administrative data in patients over the age of 18.	Surveillance Epidemiology and End Results (SEER) Registry case definition used, confirmation using chart review, administrative data, registry data, research database.	Four different algorithms were validated, each with increasing complexity but only the best performing algorithm was chosen and validated a second time. Algorithm 2: 2 or more ICD-9 codes before and 5-90 days after procedures AND 1 or more procedure codes for bone marrow investigation OR 1 or more procedure codes for two different diagnostic tests	PPV:0.81 (0.77-0.85) Sensitivity: 0.73 (0.68-0.77) from tumor registry PPV: 0.86 (0.79-0.92) from Optum Research Database	EHR
Greenberg, 2016	Validation of a Method to Identify Immunocompromised Patients with Severe Sepsis in Administrative Databases	USA	Validated methodology to identify immunocompromised patients with severe sepsis in administrative databases for patients over the age of 18 with severe	Definition using ICD-9 codes, validated using manual chart review	Patient must have HIV/AIDS, hematological malignancies, or other intrinsic immune conditions OR had solid malignancies, organ transplantations, and rheumatologic/inflammatory conditions AND received immunosuppressive medication (6 categories in total)	PPV: 94.4% (95% CI 88.8–97.7), NPV: 94.3% (95% 91.0-96.6), Sensitivity: 87.4% (95% CI 80.6–92.5%) Specificity: 97.6% (95% CI 95.0–99.9%) Performance range	EHR

[vii] Each publication reported on a either a claims-based (i.e., immunocompromised populations identified in codes derived from insurance reimbursement claims) or an EHR/EMR-based (i.e., immunocompromised populations identified in codes derived from administrative medical records) algorithm.

Author, Year	Title	Country	Summary	Disease Definition	Algorithm/Criteria	Validity	Claims/ EHR-Based[vii]
			sepsis identified in a database.		ICD 9 Codes: HIV/AIDS 042 Hematological malignancy 200-208 Other immune condition 279, 288.0, 288.1, 288.5, 288.8. 289.9, 289.83, 289.89, 289.9, 795.7, 795.79 Solid Malignancy 140-199, 209, 235-239. Organ transplant 966.8 ,V42. Rheumatological/inflammatory 135, 277.3, 277.31, 277.39, 340, 341, 357, 422, 446, 495.9, 516, 555-558, 695.4, 710, 711, 712, 714, 720 ,725	for six categories separately (more detail in paper): PPV: 71.4%-100%, NPV: 97.9%-99.8%, Sensitivity: 54.5%-89.3%, Specificity: 97.5%-100%	
Greenberg, 2018	Hospital Volume of Immunosuppressed Patients with Sepsis and Sepsis Mortality	USA	Used previously validated algorithm (Greenberg, 2016) to compare odds of death between immunosuppressed and non-immunosuppressed patients with sepsis in patients over the age of 18 part of the Vizient database.	NR	As reported in Greenberg, 2016.	Sensitivity: 87.4% (95% CI 80.6–92.5%) Specificity: 97.6% (95% CI 95.0–99.9%)	EHR
Herman, 2012	A systematic review of validated methods for identifying lymphoma using administrative data	USA, Canada	Systematic review looking at papers that used and validated algorithms to identify lymphoma using administrative or claims databases in USA or Canada.	Broad definition used for lymphoma including lymphoproliferative diseases, lymphoid leukemia, and plasma cell disorders.	Only one study (Setoguchi et al) with a validated algorithm was identified. Four algorithms were validated but algorithm 2 had the best performance. Other algorithms included the use of CPT codes or only required one or more diagnosis codes without a timeframe. Algorithm 2: 2 or more ICD-9 codes within 2 months ICD-9 codes: 200.XX, 201.XX, 202.XX (except 202.5X and 202.6X)	Algorithm 2: Sensitivity: 79.81%, Specificity: 99.81%, PPV: 62.83% Performance for other 3 algorithms varied with sensitivity: 55.17%-88.71%, specificity: 99.33%-99.86%, PPV: 34.72%-61.52%	Claims
Hernandez-Trujillo, 2015	Validity of Primary Immunodeficiency Disease Diagnoses in United States Medicaid Data	USA	Validated ICD-9 diagnoses codes for Primary Immunodeficiency Disease in patients with at least 6 months of Medicaid enrollment.	Definition based on US Immunodeficiency Network (USIDNET) and European Society for Immunodeficiencies (ESID) criteria.	2 or more repeated primary immunodeficiency disease ICD-9 diagnoses codes AND CPT codes. ICD-9 codes: 279.06, 279.04, 279.05, 279.12 CPT codes: 82784, 96365, 82787	PPV: 19.3 % (95 % CI, 11.4–29.4 %) Performance of individual codes in paper and ranged from 16.7%-33.3%. Performance of individual codes and	Claims

Author, Year	Title	Country	Summary	Disease Definition	Algorithm/Criteria	Validity	Claims/ EHR-Based[vii]
						individual CPT codes in paper and ranged from 11.1%-41.7%.	
Kato, 2008	Aspects of Immune Dysfunction in End-stage Renal Disease	NA	Article did not use a code-based definition in their study	NA	NA	NA	NA
Kavcic, 2013	Assembly of a Cohort of Children Treated for Acute Myeloid Leukemia at Free-Standing Children's Hospitals in the United States Using an Administrative Database	USA	Validated ICD-9 codes to identify Acute Myeloid Leukemia in pediatric patients under the age of 19.	Chart review	At least one hospitalization with ICD-9 codes: 205.xx–208.xx	Using ICD-9 codes only sensitivity: 100%, PPV: 31% Including manual chemotherapy review PPV improved to 100%.	EHR
Kim, 2011	Accuracy of identifying neutropenia diagnoses in outpatient claims data	USA	Validated claims-based algorithms to identify neutropenia from outpatient visits in patients that use Blue Cross/Blue Shield health plans.	Laboratory data using absolute neutrophil counts.	ICD-9 Neutropenia codes: 288.00, 288.03, 288.09, 288.5, 288.50, 288.59, 288.8, 288.9 3 algorithms: 1) Outpatient claim with 288.00, 2) Outpatient code with any listed ICD-9 code, 3) Outpatient code with any listed ICD-9 code AND one or more drug prescriptions for granulocyte colony-stimulating factors	For mild neutropenia Algorithm 1 Sensitivity:4 (3-5) Specificity:100 (100-100) PPV:33 (27-40) NPV:98 (98-98) Algorithm 2 Sensitivity:9 (78-10) Specificity:99 (99-99) PPV:18 (15-20) NPV:98 (98-98) Algorithm 3 Sensitivity:1 (0-1) Specificity:100 (100-100) PPV:56 (33-79) NPV:98 (98-98) For severe neutropenia Algorithm 1 Sensitivity:16 (7-25) Specificity:100 (100-100) PPV:5 (2-8) NPV:100 (100-100) Algorithm 2 Sensitivity:35 (23-46) Specificity:99 (99-99) PPV:3 (2-4) NPV:100 (100-100) Algorithm 3 Sensitivity:6 (0-12) Specificity:100 (100-100) PPV:22 (3-41) NPV:100 (100-100)	Claims

Author, Year	Title	Country	Summary	Disease Definition	Algorithm/Criteria	Validity	Claims/ EHR-Based[vii]
Knerr, 2017	Incidence of Neutropenia in Veterans Receiving Lung Cancer Chemotherapy: A Comparison of Administrative Coding and Electronic Laboratory Data	USA	Validated ICD-9 codes for neutropenia within a cohort of veterans who are receiving a chemotherapy agent for lung cancer and are included in the Department of Veterans Affairs.	Laboratory data using absolute neutrophil counts.	Any one of ICD-9 Neutropenia codes: 288.0, 288.00, 288.03, 288.5, 288.8, 288.9	Sensitivity: 26.3%-46.2% Specificity: 93.9%-97.0% PPV: 14.0%-63.3% NPV: 86.9%-98.9% Range depending on ANC Laboratory results	EHR
Kobrynski 2014	Prevalence and Morbidity of Primary Immunodeficiency Diseases, United States 2001–2007	U.S.	Identified cases of PID in CCE and MC databases: CCE included 5,816,905 enrollees in 2001 increasing up to 28,761,500 enrollees in 2007, aged 64 years and younger; MC included 1,428,884 enrollees in 2001 increasing to 2,964,706 enrollees in 2005, aged 0-21 years.	NR	ICD-9-CM 279.0, 279.1, 279.2, 279.8, 279.9, 288.1, and 288.2	NA	Claims
Poutsiaka 2009	Risk factors for death after sepsis in patients immunosuppressed before the onset of sepsis	U.S.	Studied the epidemiology and risk factors of sepsis in immunosuppressed and immunocompetent individuals	Immunosuppressed individuals were defined as those with human immunodeficiency virus, hematological or solid cancer, solid organ or hematopoietic stem cell transplantation, or neutropenia, or those receiving immunosuppressive medications	Authors did not use a code-based definition in their study	NA	EMR
Resnick, 2013	Examining the Use of ICD-9 Diagnosis Codes for Primary Immune Deficiency Diseases in New York State	U.S.	Identified 2,361 potential cases of PID between the ages 0 and 85+ years.	NR	ICD-9-CM 279.0, 279.10, 279.11, 279.12, 279.13, 279.2, 279.3, 279.4, 279.8, 279.9	NA	Claims
Roberts, 2020	Comparative United States autoimmune disease rates for 2010–2016 by sex,	U.S	Identified potential cases of autoimmune disease in a nationwide cohort	NR	Cutaneous/mucous membranes **Alopecia areata (Other)** ICD-9-CM 704.01	NA	EMR

Author, Year	Title	Country	Summary	Disease Definition	Algorithm/Criteria	Validity	Claims/ EHR-Based[vii]
	geographic region, and race		study between the ages of 0 and 75+ years.		ICD-10-CM L63.0, L63.1, L63.2, L63.8, L63.9 **Dermatitis herpetiformis (TSA)** ICD-9-CM 694.0, 694.2, 694.5 ICD-10-CM L13.0 <u>Endocrine system</u> **Addison's disease (TSA)** ICD-9-CM 255.41 ICD-10-CM E27.1, E27.2, E27.4, E27.40, E27.49 **Autoimmune hypoparathyroidism (TSA)** ICD-9-CM 252.1 ICD-10-CM E20.9 **Diabetes mellitus, type 1 (TSA)** ICD-9-CM 250.01, 250.03, 250.11, 250.13, 250.21, 250.23, 250.31, 250.33, 250.41, 250.43, 250.51, 250.53, 250.61, 250.63, 250.71, 250.73, 250.81, 250.83, 250.91, 250.93 ICD-10-CM E10, E10.1, E10.10, E10.11, E10.2, E10.21, E10.22, E10.29, E10.3, E10.31, E10.311, E10.319, E10.32, E10.321, E10.3211, E10.3212, E10.3213, E10.3219, E10.329, E10.3291, E10.3292, E10.3293, E10.3299, E10.33, E10.331, E10.3311, E10.3312, E10.3313, E10.3319, E10.339, E10.3391, E10.3392, E10.3393, E10.3399, E10.34, E10.341, E10.3411, E10.3412, E10.3413, E10.3419, E10.349, E10.3491, E10.3492, E10.3493, E10.3499, E10.35, E10.351, E10.3511, E10.3512, E10.3513, E10.3519, E10.3521, E10.3522, E10.3523, E10.3529, E10.3531, E10.3532, E10.3533, E10.3539, E10.3541, E10.3542, E10.3543, E10.3549, E10.3551, E10.3552, E10.3553, E10.3559, E10.359, E10.3591, E10.3592, E10.3593, E10.3599, E10.36, E10.37,		

Author, Year	Title	Country	Summary	Disease Definition	Algorithm/Criteria	Validity	Claims/ EHR-Based[vii]
					E10.37X1, E10.37X2, E10.37X3, E10.37X9, E10.39, E10.4, E10.40, E10.41, E10.42, E10.43, E10.44, E10.49, E10.5, E10.51, E10.52, E10.59, E10.6, E10.610, E10.618, E10.620, E10.621, E10.622, E10.628, E10.630, E10.638, E10.641, E10.649, E10.65, E10.69, E10.8, E10.9 Gastrointestinal **Celiac disease (TSA)** ICD-9-CM 579.0 ICD-10-CM K90.0 **Crohn's disease (Other)** ICD-9-CM 555.0, 555.1, 555.2, 555.9 ICD-10-CM K50, K50.0, K50.00, K50.01, K50.011, K50.012, K50.013, K50.014, K50.018, K50.019, K50.10, K50.111, K50.112, K50.113, K50.114, K50.118, K50.119, K50.8, K50.80, K50.811, K50.812, K50.813, K50.814, K50.818, K50.819, K50.9, K50.90, K50.911, K50.912, K50.913, K50.914, K50.918, K50.919 **Primary biliary cirrhosis** (UEA) ICD-9-CM 571.6 ICD-10-CM K74.3 **Ulcerative colitis (Other)** ICD-9-CM 556.0, 556.1, 556.2, 556.3, 556.4, 556.5, 556.6, 556.8, 556.9 ICD-10-CM K51, K51.0, K51.00, K51.011, K51.012, K51.013, K51.014, K51.018, K51.019, K51.2, K51.20, K51.211, K51.212, K51.213, K51.214, K51.218, K51.219, K51.3, K51.30, K51.311, K51.312, K51.313, K51.314, K51.318, K51.319, K51.4, K51.40, K51.411, K51.412, K51.413, K51.414, K51.418, K51.419, K51.5, K51.50, K51.511,		

Author, Year	Title	Country	Summary	Disease Definition	Algorithm/Criteria	Validity	Claims/ EHR-Based[vii]
					K51.512, K51.513, K51.514, K51.518, K51.519, K51.8, K51.80, K51.811, K51.812, K51.813, K51.814, K51.818, K51.819, K51.9, K51.90, K51.911, K51.912, K51.913, K51.914, K51.918, K51.919 Hematopoetic **Acquired hemophilia A (TSA)** ICD-9-CM 286.52 ICD-10-CM D68.311 **Autoimmune hemolytic anemia (TSA)** ICD-9-CM 283.0 ICD-10-CM D59.0, D59.1 **Autoimmune neutropenia (TSA)** ICD-9-CM 288.09 ICD-10-CM D70.8 **Immune thrombocytopenic purpura (TSA)** ICD-9-CM 287.31 ICD-10-CM D69.3 Musculoskeletal **Polymyositis/dermatomyositis (UEA)** ICD-9-CM 710.3, 710.4 ICD-10-CM M33.0, M33.00, M33.01, M33.02, M33.1, M33.10, M33.11, M33.12, M33.2, M33.20, M33.21, M33.22, M33.9, M33.90, M33.91, M33.92 **Rheumatoid arthritis (UEA)** ICD-9-CM 714, 714.0, 714.2, 714.3, 714.31, 714.32, 714.33, 714.4, 714.8, 714.81, 714.89 ICD-10-CM M05.60, M05.611, M05.612, M05.619, M05.621, M05.622, M05.629, M05.631, M05.632, M05.639, M05.641, M05.642, M05.649, M05.651, M05.652, M05.659, M05.661, M05.662, M05.669, M05.671, M05.672, M05.679, M05.69, M05.70, M05.711, M05.712, M05.719, M05.721, M05.722,		

Author, Year	Title	Country	Summary	Disease Definition	Algorithm/Criteria	Validity	Claims/ EHR-Based[vii]
					M05.729, M05.731, M05.732, M05.739, M05.741, M05.742, M05.749, M05.751, M05.752, M05.759, M05.761, M05.762, M05.769, M05.771, M05.772, M05.779, M05.79, M05.80, M05.811, M05.812, M05.819, M05.821, M05.822, M05.829, M05.831, M05.832, M05.839, M05.841, M05.842, M05.849, M05.851, M05.852, M05.859, M05.861, M05.862, M05.869, M05.871, M05.872, M05.879, M05.89, M05.9, M06.00, M06.011, M06.012, M06.019, M06.021, M06.022, M06.029, M06.031, M06.032, M06.039, M06.041, M06.042, M06.049, M06.051, M06.052, M06.059, M06.061, M06.062, M06.069, M06.071, M06.072, M06.079, M06.08, M06.09, M06.1, M06.20, M06.211, M06.212, M06.219, M06.221, M06.222, M06.229, M06.231, M06.232, M06.239, M06.241, M06.242, M06.249, M06.251, M06.252, M06.259, M06.261, M06.262, M06.269, M06.271, M06.272, M06.279, M06.28, M06.29, M06.30, M06.311, M06.312, M06.319, M06.321, M06.322, M06.329, M06.331, M06.332, M06.339, M06.341, M06.342, M06.349, M06.351, M06.352, M06.359, M06.361, M06.362, M06.369, M06.371, M06.372, M06.379, M06.38, M06.39, M06.80, M06.811, M06.812, M06.819, M06.821, M06.822, M06.829, M06.831, M06.832, M06.839, M06.841, M06.842, M06.849, M06.851, M06.852, M06.859, M06.861, M06.862, M06.869, M06.871, M06.872, M06.879, M06.88, M06.89, M06.9, M08.00, M08.011, M08.012, M08.019, M08.021, M08.022, M08.029, M08.031,		

Author, Year	Title	Country	Summary	Disease Definition	Algorithm/Criteria	Validity	Claims/ EHR-Based[vii]
					M08.032, M08.039, M08.041, M08.042, M08.049, M08.051, M08.052, M08.059, M08.061, M08.062, M08.069, M08.071, M08.072, M08.079, M08.08, M08.09, M08.20, M08.211, M08.212, M08.219, M08.221, M08.222, M08.229, M08.231, M08.232, M08.239, M08.241, M08.242, M08.249, M08.251, M08.252, M08.259, M08.261, M08.262, M08.269, M08.271, M08.272, M08.279, M08.28, M08.29 Neurological system **Chronic inflammatory demyelinating polyneuropathy (TSA)** ICD-9-CM 357.81 ICD-10-CM G61.81 **Guillain–Barré syndrome (TSA**) ICD-9-CM 357.0 ICD-10-CM G61.0 **Multiple sclerosis (UEA)** ICD-9-CM 340 ICD-10-CM G35 **Myasthenia gravis (TSA)** ICD-9-CM 358.0, 358.00, 358.01 ICD-10-CM G70.0, G70.00, G70.01 Systemic disorders **Scleroderma (UEA)** ICD-9-CM 710.1 ICD-10-CM M34.0, M34.2, M34.81, M34.82, M34.83, M34.89, M34.9 **Sjögren's syndrome (UEA)** ICD-9-CM 710.2 ICD-10-CM M35.0, M35.00, M35.01, M35.02, M35.03, M35.04, M35.09 **Systemic lupus erythematosus (UEA)** ICD-9-CM 710.0		

Author, Year	Title	Country	Summary	Disease Definition	Algorithm/Criteria	Validity	Claims/ EHR-Based[vii]
					ICD-10-CM M32.0, M32.1, M32.10, M32.11, M32.12, M32.13, M32.14, M32.15, M32.19, M32.8, M32.9		
Rubin, 2014	2013 IDSA clinical practice guideline for vaccination of the immunocompromised host.	U.S.	Evidence-based guidelines for vaccination of immunocompromised adults and children.	Minimally immunocompromised HIV patients described include (a) asymptomatic HIV-infected adults with CD4 T-cell lymphocyte count ≥200 cells/mm3 (weak, low) (b) asymptomatic HIV-infected children aged 9 months–5 years with CD4 T-cell lymphocyte percentages of ≥15 (weak, very low) Guidelines provided for the following patients: those receiving intensive chemotherapy or those who have received anti–B-cell antibodies within 6 months; those with primary (congenital) complement deficiencies; those with Phagocytic Cell Deficiencies (e.g., CGD, Leukocyte Adhesion Deficiency, Chediak–Higashi Syndrome; those with Innate Immune Defects that Result in Defects of Cytokine Generation/Response or Cellular Activation (e.g., Defects of the Interferon-gamma/Interleukin-12 Axis); those with	Article did not use a code-based definition	NA	NA

Author, Year	Title	Country	Summary	Disease Definition	Algorithm/Criteria	Validity	Claims/ EHR-Based[vii]
				minor Antibody Deficiencies; those with Major Antibody Deficiencies Who are Receiving Immunoglobulin Therapy; those with combined Immunodeficiencies; HIV-infected; cancer patients; hematopoietic stem cell transplant patients; sold organ transplant recipients; those with chronic inflammatory disease on immunosuppressive medications; those with asplenia or sickle cell diseases			
Shayne, 2013	Risk factors for in-hospital mortality and prolonged length of stay in older patients with solid tumor malignancies	U.S.	Identified a cohort of 386,377 hospitalized cancer patients with known mortality aged ≥ 65 years to study risk of prolonged hospital stay and mortality	NR	Diagnosis of malignant disease ICD-9-CM 140-208 By cancer type (ICD-9 CM) lung (162–163), breast (174–175), esophageal (150), gastric (151), pancreatic (157), colon (153), rectal (154), other intra-abdominal (152, 155–156,158–159), ovarian (183), endometrial or cervical (179–182), bladder (188), renal (189), prostate (185), testicular (186), brain (191–192), head and neck (140–149, 160–161), sarcoma (170–171) and melanoma (172), metastatic disease (ICD-9-CM 196–199)	NA	claims
Setoguchi, 2007	Agreement of diagnosis and its date for hematologic malignancies and solid tumors between Medicare claims and cancer registry data	U.S.	Calculated PPV to validate 4 algorithms to identify potential cases of hematologic and solid malignancies in 6,996 patients aged ≥ 65 years .	Confirmed cases were recorded in the Pennsylvania State cancer registry during the study period.	4 Algorithms tested for each cancer type **Definition 1**: Any of the following ≥1 cancer diagnosis + any diagnosis or procedure codes related to complications of cancer or palliative care in two weeks followed by another	Lung **Definition 1** PPV: 75.89% Sensitivity: 56.35% Specificity: 99.79% **Definition 2** PPV: 66.04% Sensitivity: 76.19% Specificity: 99.54%	claims

Author, Year	Title	Country	Summary	Disease Definition	Algorithm/Criteria	Validity	Claims/ EHR-Based[vii]
					diagnosis of cancer within 12 months. ≥1 diagnostic procedure with biopsy followed by ≥ 2 cancer diagnoses at two different occasions within 12 months (recorded on different dates from the procedures). ≥1 cancer diagnosis + any surgery related to cancer during the same hospitalization and/or visit. ≥1 cancer diagnosis + any cancer chemotherapy during the same hospitalization and/or visit ≥1 cancer diagnosis + any radiation therapy during the same hospitalization and/or visit ≥1 cancer diagnosis + hematopoietic cell transplantation during the same hospitalization and/or visit (for leukemia only) ≥1 cancer diagnosis + oral chemotherapy dispensing within 2 weeks after the diagnosis **Definition 2:** ≥2 diagnoses of cancer within 2 months **Definition 3:** Cases defined by using Definition 1 or 2 **Definition 4:** ≥1 diagnosis of cancer Hematologic malignancies (lymphoma) ICD-9-CM codes 200.xx, 201.xx, 202.xx (except 202.5x and 202.6x [**Reported in Herman 2012**] Algorithms for Solid Malignancies (leukemia), lung, colorectal, stomach and breast cancers: ICD codes not reported	**Definition 3** PPV: 64.83% Sensitivity: 80.06% Specificity: 99.49% **Definition 4** PPV: 45.19% Sensitivity: 86.69% Specificity: 98.78% <u>Colorectal</u> **Definition 1** PPV: 70.95% Sensitivity: 67.25% Specificity: 99.62% **Definition 2** PPV: 69.40% Sensitivity: 80.36% Specificity: 99.51% **Definition 3** PPV: 63.84% Sensitivity: 83.98% Specificity: 99.35% **Definition 4** PPV: 44.82% Sensitivity: 88.02% Specificity: 98.51% <u>Stomach</u> **Definition 1** PPV: 59.78% Sensitivity: 69.92% Specificity: 99.93% **Definition 2** PPV: 55.65% Sensitivity: 81.36% Specificity: 99.90% **Definition 3** PPV: 51.96% Sensitivity: 84.32% Specificity: 99.88% **Definition 4** PPV: 35.05% Sensitivity: 89.41% Specificity: 99.75% <u>Breast</u> **Definition 1** PPV: 81.74% Sensitivity: 46.91%	

Author, Year	Title	Country	Summary	Disease Definition	Algorithm/Criteria	Validity	Claims/ EHR-Based[vii]
						Specificity: 99.84% **Definition 2** PPV: 76.56% Sensitivity: 78.89% Specificity: 99.62% **Definition 3** PPV: 74.55% Sensitivity: 83.03% Specificity: 99.56% **Definition 4** PPV: 50.19% Sensitivity: 87.23% Specificity: 98.65% <u>Lymphoma</u> **Definition 1** PPV: 61.52% Sensitivity: 55.17% Specificity: 99.86% **Definition 2** PPV: 62.83% Sensitivity: 79.81% Specificity: 99.81% **Definition 3** PPV: 56.59% Sensitivity: 83.31% Specificity: 99.74% **Definition 4** PPV: 34.72% Sensitivity: 88.71% Specificity: 99.33% <u>Leukemia</u> **Definition 1** PPV: 41.08% Sensitivity: 41.76% Specificity: 99.93% **Definition 2** PPV: 43.18% Sensitivity: 52.20% Specificity: 99.92% **Definition 3** PPV: 37.71% Sensitivity: 61.54% Specificity: 99.88% **Definition 4** PPV: 18.82% Sensitivity: 73.63% Specificity: 99.63%	

Author, Year	Title	Country	Summary	Disease Definition	Algorithm/Criteria	Validity	Claims/ EHR-Based[vii]
Sunesen, 2010	Immunosuppressive disorders and risk of anal squamous cell carcinoma: a nationwide cohort study in Denmark, 1978–2005	Denmark	Identified potential cases with chronic conditions associated with immunosuppression in a nationwide cohort study between the ages of 0 and 96 years.	NR	HIV/AIDS ICD-10 B20-24 Hematologic malignancies ICD-10 C80-85, C88, C90-96 Hematological system ICD-10 D59.0, D59.1, D69.3 Endocrine system ICD-10 E05.0, E06.3, E27.1 Central nervous / neuromuscular system ICD-10 G35, G70.0 Gastrointestinal / hepatobiliary system ICD-10 D51.0, K90.0, K50, M07.4, K51, M07.5, K74.3 Skin ICD-10 L10.0, L10.2, L10.4, L12.0, L13.0, L40, M07.0-M07.3 Connective tissue diseases ICD-10 M05, M06, G73.7D, I32.8A, I39.8E, I41.8A, I52.8A, M45, M33, M32, G73.7C, I39.8C, N08.5A, N16.4B, M35, G73.7A, N16.4A, D86, G53.2, H22.1A, I41.8B, K77.8B, M63.3, Vasculitis syndromes ICD-10 M30.0, M31.3, M31.5, M31.6, M35.3	NA	EHR
Taneja, 2014	Can Dialysis Patients be Accurately Identified Using Healthcare Claims Data?	U.S.	Calculated PPV to validate algorithm to identify potential dialysis (hemodialysis and peritoneal) cases in 233 patients with evidence of end-stage renal disease between the ages of 18 and 63 years.	Confirmed cases require medical records include description of dialysis modality received (hemodialysis or peritoneal).	Hemodialysis-related ICD-9-CM diagnosis 458.21, V56.0, V56.1, V56.31 ICD-9-CM procedure code 38.95, 39.27, 39.42, 39.43, 39.95. CPT-4 procedure 36145, 36800, 36810, 36815, 36831, 36832, 36833, 36838, 90935, 90937, 90939, 90940, 93990, 99512 HCPCS procedure A4674, A4680, A4690, A4700, A4705, A4706, A4707, A4708, A4709, A4712, A4714, A4730, A4740, A4750, A4755, A4770, A4774, A4780, A4790, A4800, A4801, A4802, A4820, A4850, A4870, A4890, A4918, A4919, A4929, E1510, E1520, E1530, E1540, E1550, E1560, E1575, E1580,	PPV At least one form of dialysis treatment 100% Peritoneal-related billing codes: 34.9% (95% CI: 20.6%- 49.1%) ±30 days 67.4% (95% CI: 53.4-81.4%) ±90 days 67.4% (95% CI: 53.4-81.4%) ±180 days ICD-9-CM (±30/±90/1±80 days) 54.98	Claims

Author, Year	Title	Country	Summary	Disease Definition	Algorithm/Criteria	Validity	Claims/ EHR-Based[vii]
					E1590, E1600, E1610, E1615, E1620, E1625, E1635, E1636 Peritoneal dialysis-related ICD-9-CM diagnosis 996.56, 996.68, V56.2, V56.32, V56.8 CPT-4 procedure 49420, 49421, 90945, 90947 HCPCS procedure A4653, A4671, A4672, A4673, A4719, A4720, A4721, A4722, A4723, A4724, A4725, A4726, A4728, A4760, A4765, A4766, A4860, A4880, A4900, A4901, A4905, E1632, E1592, E1594, E1630, E1634, E1638, E1640	50%/50%/50% V56.2 0%/50/50% 996.68 50%/50%/50% Hemodialysis-related billing codes: 86.7% (95% CI: 81.6-91.8%) ±30 days 90.8% (95% CI: 86.4-95.1%) ±90 days 93.1% (95% CI: 89.3-96.8%) ±180 days ICD-9-CM (±30/±90/1±80 days) 38.95 86.8%/92.1%/93.4% 39.27 100%/100%/100% 39.95 91.8%/95.1%/96.7% V56.0 93.9%/93.9%/93.9% V56.1 100%/100%/100%	

Abbreviations: CCE: Commercial Claims and Encounters; CDC: Centers for Disease Control and Disease Prevention; CI: Confidence Interval; CPT: current procedural terminology; EHR, electronic health record, EMR: electronic medical record; HCPS: Healthcare Common Procedure Coding System; HIV: human immunodeficiency virus; ICD-9-CM, International Classification of Diseases, Ninth Revision, Clinical Modification; ICD-10-CM, International Classification of Diseases, Tenth Revision, Clinical Modification; MC: Multi-State Medicaid; NA, not applicable; NR, not reported; PID: Primary Immunodeficiency Disease; PPV, positive predictive value; TSA: disease characterized by tissue-specific autoantigen; UEA: disease characterized by ubiquitously expressed autoantigen

Appendix B. Immunocompromised Coding Algorithm

Refer to the enclosed Excel file titled “FDA_BEST_Immunocompromised_Population_Definition_ 2021_Q2” for the complete list of codes for identification of immunocompromised patients.