



2023

Urologic Diseases in America

ANNUAL DATA REPORT

Chapter 1: Introduction and Methods

June 21, 2023

SPONSORED BY

National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK)
National Institutes of Health (NIH)

Acknowledgements

This year's *Urologic Diseases in America: Annual Data Report* was prepared in collaboration between Acumen, LLC and the contract sponsor, National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) of the National Institutes of Health (NIH). The NIDDK team composed of core members Kevin C. Abbott, Julia S. Barthold, and Ziya Kirkali; as well as Kevin Friel, Melanie Hardy, Max Kimpson, Ivonne Schulman, and Neha Shah.

Additional clinical urology contributions were coordinated by Acumen, LLC, led by Chad Ellimoottil (University of Michigan) and John P. Lavelle (Stanford University; Veterans Affairs Palo Alto CA). The Acumen LLC team consisted of core members Kyle Buika, Po-Lun Chou, Myrna Cozen, John C. Hornberger, Zhihang Lin, Suraj Pant, and Lei Sandy Ye, with additional contributions from Dee Dee Aubourg, Yvonne Aubourg, Jeffrey Carwile, Tuyen Duong, Paul Fanelli, Thomas Genova, Kristina Gjika, John Hunt, Sushant Joshi, Xiaofei Lai, Gloriana Lopez Montealegre, Patricia Morales, Wagner Peng, Andrew Rabe, Eril Smith, Victoria Ta, Lidya Tadesse, Tatiana Valentine, and Yiren Alan Wang.

Note

This chapter is the first of four chapters in the *2023 Urologic Diseases in America: Annual Data Report (ADR)*. This chapter introduces the 2023 ADR and describes the methodology underlying the results for Chapter 2: Benign Prostatic Hyperplasia and Associated Lower Urinary Tract Symptoms (BPH/LUTS) in Men; Chapter 3: Urinary Stone Disease (USD); and Chapter 4: Urinary Incontinence (UI). These chapters are available under separate links on the UDA website. Additional details on the methodology and data sources are provided in Appendices A and B, respectively, that accompany this chapter.

Suggested citation

Urologic Diseases in America. 2023. *Urologic Diseases in America: Annual Data Report*. National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD, 2023.

1 Introduction and Methods

The Urologic Diseases in America (UDA) Annual Data Report (ADR) – herein in its third release since 2012 – is sponsored by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) of the National Institutes of Health (NIH). The primary audiences for the ADR are researchers, policy makers, patients, and clinicians. In review with NIDDK and its advisory panel, this ADR discusses gaps in existing epidemiological data and provides suggestions for future investigations. The first ADR was published in 2012, with a second release (or addendum) in 2018.¹ The 2012 ADR included cancerous and non-cancerous urologic diseases. The second release excluded cancerous diseases (as does this third release), which are reported by the National Cancer Institute (NCI).²

The current ADR's primary aim is to document epidemiological trends of benign urological diseases in the past decade among Americans, including trends in management and resource use for three common urologic diseases: benign prostatic hyperplasia and associated lower urinary tract symptoms (BPH/LUTS), urinary stone disease (USD), and urinary incontinence (UI).

This ADR employs a retrospective study design using claims data from two separate sources. This study cohort includes persons aged 18 and older residing in the United States. The age 18-64 cohort data are from a private insurer's claims data. The age 65 and older cohort data are from Medicare Fee-for-Service (FFS) claims data. This ADR study spans the years 2012 to 2021.

The reported measures include disease/condition prevalence, incidence, comorbidities, utilization of diagnostic tests, prescription drugs filled, procedures used (surgery and other non-pharmaceutical treatments), and resource utilization for these disorders. The analyses are stratified by age groups, gender, reported race, dual eligibility status (age 65 and older), and geographic regions.

This ADR is organized as follows. The rest of this chapter summarizes the data sources and methods used (1.1), summarizes the measures reported in the study (1.2), provides an overview of the analytical approach (1.3), and highlights the characteristics of the cohorts (1.4). Additional details of the methodology and data sources are provided in Appendices A and B, respectively, that accompany this chapter.

Drawing on the methodology described in this chapter, this ADR further includes three separate chapters on findings for each of the three urologic conditions. Chapters 2, 3, and 4 present results on BPH/LUTS, USD, and UI, respectively. These chapters are available under separate links on the UDA website.

1.1 Data sources

This ADR includes two cohorts captured in two data sources: (1) adults age 65 years and older entered into the Medicare Fee-for-Service (FFS) data from the Centers for Medicare & Medicaid Services (CMS), and (2) adults age 18 to 64 from Optum’s de-identified Clinformatics® Data Mart Database (CDM). Both of the data sources used below identified claims only, and did not contain any information from the medical record.

→ Medicare FFS data

The data consist of 100% of FFS data, which have been augmented from the 5% sample used in previous editions of the ADR. The data include information on enrollment status, demographics, inpatient stays, outpatient services, home health care, skilled nursing facilities care, hospice care, pharmacy (Part D data), and reimbursements.

→ CDM data

The data consist of private insurer paid claims that include information on plan member eligibility status, demographics, inpatient stays, outpatient services, pharmacy, and reimbursements.

→ Years of data

For this ADR, both data span the years 2012-2021. The latest year covered in the previous edition of the ADR was 2013. The interval of analysis is annual.

Both data sources have been used extensively for clinical, epidemiologic, health services/outcomes, and economics research, with the strengths and limitations of using these data sources discussed in the literature (see Appendix A). Appendix B provides more details on the data files used. In the context of this ADR, Appendix A also describes general limitations relevant to interpreting the findings herein.

1.2 Brief summary of measures

This ADR reports prevalence and incidence of selected diseases, comorbidities, use of diagnostic tests, prescription drugs filled, utilization of procedures (including surgery and other non-pharmaceutical treatments), utilization of healthcare services, and healthcare expenditures. This subsection provides an overview of these outcomes. Diagnoses, testing and procedures were identified by Healthcare Common Procedure Coding System and International Classification of Diseases codes as detailed in section A.1.

The following are definitions of measures:

Period prevalence	Number and percentage of eligible persons in the cohort who had recorded the disease or condition during a given period of time, specifically over each calendar year.
--------------------------	--

Incidence	<p>Number and percentage of persons who had recorded a qualifying diagnosis for a given urological disease each year, and did not have a previous diagnosis for the given disease in claims in a lookback period of 36 months relative to any given incident month in the year.</p> <p>This does not ensure persons have not been diagnosed before the earliest date of the lookback period, but it is considered among the most feasible methods when using claims data.³ Persons who have not had a claims code for a particular diagnosis for a prolonged interval have a reasonable likelihood of having incident (new) disease. The implications of different lookback periods have been discussed in the literature.⁴</p>
------------------	---

➔ **Prevalent versus incident cohorts**

After using the aforementioned methods to identify patients with our diseases of interest, we formed two cohorts: “prevalent cohort” and “incident cohort”. These cohorts were used for the remaining measures. Of note, for purposes of writing clarity, we may refer to the incident cohort as patients who are “newly identified”.

For all metrics based on the “incident cohort”, our study focuses solely on the age 65 and older cohort. This is due to the greater availability of longitudinal data from Medicare FFS, allowing for tracking of patients' follow-up over a period of one year or longer.

Comorbidities	<p>For the prevalent cohort, we report the number and percentage of patients with each disease or condition who were also identified by one of the selected comorbidities (a coexisting condition). We report the proportion of patients with each comorbidity by disease and age group.</p>
Diagnostic tests	<p>For the incident cohort, we report the number and percentage of patients who had a claim for a qualifying diagnostic test within 3 months prior and 12 months after the month of initial diagnosis, to capture diagnostic tests that may be related to the diagnosis.</p>

<p>Prescription drugs</p>	<p>For the prevalent cohort of persons aged 65 and above, we count the number of patients recorded with the disease who were also fully enrolled in Medicare Part D by year. We then report the number and percentage of them who filled at least one prescription in any of the identified relevant pharmacologic classes, and for each individual pharmacologic class. We applied a similar analysis for the prevalent cohort of persons aged 18-64; no Part-D-like restriction is needed as each enrollee is covered for both medical services and prescription drugs.</p> <p>For the incident cohort, we computed the number and percentage of patients with incident disease aged 65 and older who filled any disease-related prescriptions, the percentage shares of first filled prescriptions by type, and the average number of months from initial diagnosis to first filled prescription by type; all within 5 years after initial diagnosis (for incident year 2015).</p>
<p>Procedures</p>	<p>For the prevalent cohort and each disease entity included in the ADR, the number and percentage of patients receiving any disease-related procedures are reported. Further, the number and percentage of patients receiving each major procedure type are reported.</p> <p>For the incident cohort aged 65 and older, we computed the number and percentage of patients with incident disease who underwent any disease-related procedure, the percentage shares of first procedures by type, and the average number of months from initial diagnosis to first procedure by type; all within 5 years after initial diagnosis (for incident year 2015).</p>
<p>Utilization of healthcare services</p>	<p>We report four main service utilization outcomes for the incident cohort aged 65 and older: 1. Evaluation and Management (E&M) visits; 2. Emergency Department (ED) visits, 3. Hospitalization – observation, 4. Hospitalization - inpatient.</p> <p>E&M visits are defined as outpatient visits or other non-institutional health services visits with any diagnosis of the disease. ED visits are defined based on visits with primary diagnosis of the disease. Stays can be distinguished as observation only (either as outpatient or inpatient) or non-observation (all of which are inpatient hospital). Both types of stays are based on those with primary diagnosis of the disease. We report the percentage of incident patients with ED visits within 12 months after initial diagnosis. The same method used to report ED visits was also used to report inpatient-hospital stays and observation stays. We also report the number of E&M visits, ED visits, inpatient-hospital stays, and observation stays on a per-person-per-year basis.</p>

Healthcare expenditures	For the incident cohort, we report Medicare FFS expenditures (in nominal dollars) with primary diagnosis of the disease for persons with incident disease within 12 months after initial diagnosis. Expenditures are also reported on a per person per year basis.
--------------------------------	--

➔ **Stratification variables**

Measures were stratified based on these variables below, with possible modified groupings:

Age groups	For Medicare data, age groups comprise 65-69, 70-74, 75-79, 80-84, and 85 or above. For CDM data, age groups comprise 18-24, 25-34, 35-44, 45-54, and 55-64. For BPH/LUTS, CDM age groups start at 40 (40-44, 45-49, 50-54, 55-59, and 60-64). Comorbidity data are stratified into two age groups for both Medicare (65-74 and 75+) and CDM (18-54 and 55-64; 40-54 and 55-64 for BPH) data.
Gender	Male or female
Geographic regions	The United States is grouped into four regions corresponding to those used by the US Census Bureau: Northeast, Midwest, South, and West.
Race	For Medicare data, race categories include Non-Hispanic White, Black, Hispanic, North American Native, Asian, Other, and Missing. For CDM data, race categories include Non-Hispanic White, Black, Hispanic, Asian, and Missing.
Dual-eligibility status (Medicare only)	Persons with low-income who are eligible for enrollment in both Medicare and Medicaid. This has been used as a proxy for individual level poverty, which otherwise depends on ecologic (i.e., ZIP code or other geographic level) markers of income. ⁵

1.3 Analyses summary

Appendix A provides details on analytical methods, including cohort and claims selection criteria (A1), analytic approach and computations (A2), validation methods (A3), and general limitations (A4). In summary, we queried all claims across Medicare FFS settings or CDM diagnosis files for the years of interest, including a set of data restrictions in the process. For each person, we looked for occurrence of any of the specified medical conditions within a defined time window, and then recorded the person counts where these conditions are met. We then counted the occurrences of the outcomes described in the previous subsection, either within the same time window or a follow-up window, at the annual interval. We either scaled these counts annually by the total count of the cohort, or the total count of persons (prevalent or incident cohort) recorded with the disease of interest. As a result, most of our descriptive statistics are in the form of raw estimates such as total counts/percentages. The analyses were conducted in SAS version 15.2.

1.4 Cohort composition

1.4.1 Age 65 years and older

→ Overall trends

The size of this cohort increased from 24.4 million in 2012 to 25.6 million in 2019, and declined to 24.5 million in 2021. The lower cohort size seen in 2021 has been attributed to the ongoing, well-documented trend toward higher proportion of eligible adults enrolling in Medicare Advantage (Part C) plans and to mortality effects of the COVID-19 pandemic that greatly affected older adults. Among men, the size of the cohort increased from 10.4 million in 2012 to 11.3 million in 2019, and then decreased to 10.8 million in 2021. Among women, the size of the cohort increased from 14.0 million to 14.4 million between 2012 and 2019 and decreased to 13.7 million in 2021. Between 2012 and 2021, women comprised an annual average of 56% of the cohort. About 10% of the cohort were dually-eligible for Medicare and Medicaid in 2021.

→ Comparison with national data

The cohort had more persons identifying as Non-Hispanic White compared to the proportion reported in 2021 US Census data for adults age 65 and older.⁶ Specifically, persons identifying as Non-Hispanic White comprised 85% of the cohort, compared with 75% of the national population in the same age group. 6.3% of the cohort were persons identifying as Black, versus 9.5% of the national population, and persons identifying as Asian, Hispanic, and North American Natives comprised 2.1%, 1.5%, and 0.4% of the cohort; and 4.8%, 9.0%, and 0.6%, of the national population, respectively.

1.4.2 Age 18 to 64 years

→ Overall trends

The size of this cohort increased from 6.1 million in 2012 to 6.3 million in 2017, and then decreased to 5.6 million in 2021. The trends were the same regardless of gender. Among men, the size of the cohort increased from 3.0 million in 2012 to 3.2 million in 2017, and then decreased to 2.9 million in 2021. Among women, the size of the cohort increased from 3.1 million in 2012 to 3.2 million in 2017 and decreased to 2.8 million in 2021. On average, during 2012-2021, the gender composition of this younger cohort was evenly balanced (49.2% women and 50.8% men).

→ Comparison with national data

This younger cohort also had more persons identifying as Non-Hispanic White compared to the proportion reported in 2021 US Census data for adults aged 18-64 (64% versus 59%). Persons identifying as Black, Hispanic, or Asian comprised 8.6%, 12%, and 5.3%, compared to 13%, 19%, and 6.4% of the national population, respectively.

Appendix A: Methodological Details

This methodology appendix describes the analytic approach used in this ADR. The first subsection (A.1) describes the cohort and claims selection criteria for the prevalent and incident cohorts behind all metrics. The second subsection (A.2) provides additional details on the analytic process behind the computations of metrics. The third subsection (A.3) discusses validation methods used in the computation process. The fourth subsection (A.4) summarizes the limitations of the methodology. This appendix also discusses refinements and expansions in methodological approach relative to previous ADRs where relevant.

A.1 Cohort and claims selection criteria

Most summary statistics in this ADR rely on either the prevalent cohort or the incident cohort for a disease of interest. The prevalent cohort denotes persons who have a disease or condition in a given year. The incident cohort denotes persons who are newly identified with a urologic disease during a period of interest, among those without disease diagnosis in a lookback window. This subsection describes the computation process to draw these cohorts.

To identify a patient cohort, the process was the following: (1) query all claims across Medicare Fee-for-Service (FFS) settings or Optum's de-identified Clinformatics® Data Mart Database (CDM) diagnosis files across years of interest, (2) for every person, look for occurrences of any of the specified medical conditions within a defined time window, and (3) record the person count where these conditions are met.

→ Restrictions to data

For all analyses, both the Medicare FFS and CDM data were subject to a number of restrictions. The Medicare FFS data contained the following restrictions: continuous Parts A and B enrollment in each year, age 65+, no Part C enrollment, and residence in the United States (50 states plus the District of Columbia). CDM data were restricted to persons with full annual enrollment in a commercial health plan, aged 18-64 (40-64 for BPH/LUTS analysis), and residence in the United States (50 states plus the District of Columbia).

→ Prevalent cohort

To identify a cohort of persons with specified medical conditions, a clinician-defined list of qualifying diagnosis codes (International Classification of Diseases [ICD]-9/10 diagnosis [DGN]) or procedure codes (ICD-9/10 procedure [PRC]), Current Procedural Terminology (CPT) codes, and Healthcare Common Procedure Coding System (HCPCS) codes were compiled. For Medicare FFS data, eligible claims to evaluate these codes referred to all claims in the Inpatient Stays (IP), Outpatient Services (OP), Home Health Agencies (HHA), Skilled Nursing Facilities (SNF), Hospice Care Organizations (HS), and Carrier (PB) settings. For CDM data, eligible claims referred to those in the Inpatient Confinement files (for Acute Care Hospitals or SNF) and Medical Claims files.

For Medicare FFS data, institutional OP claims were further restricted to those with presence of HCPCS codes 99201–99205, 99211–99215, 99241–99245, 99271–99275, 99281–99285, 99288 on the claim, indicating office or other outpatient visits for the evaluation, management, and consultation of existing and new patients. Non-institutional claims were further restricted to those where the Service Type field was labelled “Medical care,” “Surgery,” or “Consultation”; and there was no pharmacy, an ambulance, a mass immunization center, an independent laboratory, or other place of service recorded in the Service Place field.

For CDM data, medical claims were primarily further restricted to those where “Professional service: surgery,” “Professional service: emergency room,” “Professional service: consultation,” “Professional service: office visits,” or “Home health/hospice visits” appeared in the Service Type field; and further, no pharmacy, a mass immunization center, an ambulance, or an independent laboratory record was present on the Service Place field.

For DGN/PRC codes, since there was a transition from ICD-9 to ICD-10 in October 2015, codes in both versions were used if the study window crossed this transition point. The ICD 9-10 General Equivalence Mapping served as a guide to obtain comprehensive code mappings between ICD-9 and ICD-10 codes. This was further complemented by clinical expertise to address many-to-many mapping issues and differences in level of details introduced by the ICD transition. Please see associated codebooks accompanying this ADR for all the relevant codes and mapping used.⁷

→ Incident cohort

The derivation of the incident cohort relied on the following process. The restrictions applied for the prevalent cohort were also applied to the incident cohort. The analysis then further restricted the cohort to those who did not indicate a urologic disease during a fixed window prior to the incident year-month. The analysis then further restricted this cohort to those with continuous and full enrollment during a defined follow-up period of varying length, depending on the outcome. Such restrictions allow for studying follow-up of person cohorts and help ensure a consistent sample of persons over time. For all metrics based on the “incident cohort”, our study focuses solely on the age 65 and older cohort. This is due to the greater availability of longitudinal data from Medicare FFS, allowing for tracking of patients' follow-up over a period of one year or longer.

More specifically, for each year-month t , a 36-month lookback window [$t-36, t-1$] was considered. For each year-month t , the Medicare population was restricted to those with continuous and full enrollment in Part A/B claims in the lookback window. Among this group, a person was considered an “incident patient” in year-month t if the person registered a qualifying diagnosis for that time; and did not have a given urologic disease diagnosis in claims in the lookback period. Rolling windows were used to study the incident cohort, in contrast to the use of a fixed incident year in previous editions. The use of rolling windows allows time trends to be shown for incident statistics. Rolling windows also enhance the ability to consider short-, medium-, and long-term outcomes.

A.2 Computations

The statistical analyses behind the 2023 ADR data tables span the following categories of outcomes: prevalence, incidence, comorbidities, diagnostic tests, prescription drugs filled, procedure use, utilization of healthcare services, and healthcare expenditure. These outcomes were compiled annually for the prevalent cohort and on a monthly rolling basis (aggregated annually) for the incident cohort. The discussion below describes the analytical approach for a number of key outcomes in this ADR.

→ Prevalence

Prevalence was presented as the number and percentage of persons in the two age cohorts who were diagnosed with the disease of interest in each year. A patient with disease X was defined as a person who had a qualifying diagnosis code in at least one eligible claim within the calendar year.

For BPH, we also made an exclusion restriction – persons with “symptom” codes should not carry a diagnosis of prostate cancer (ICD9 185 or ICD10 C61). See accompanying BPH/LUTS Excel codebook for symptom codes.

In the case of UI, we classified the subcategories using the following logic in Table A.1.

Table A.1 Classification logic for UI types

UI Type	Classification Criteria
1. FUI	Person has Fistula incontinence (FUI) diagnosis code, regardless of existence of Mixed (MUI), Stress (SUI), Urgency (UUI), Overflow, or Other incontinence diagnosis codes in the same year.
2. MUI	Person has Mixed incontinence diagnosis code, or both SUI AND UUI diagnosis codes. Person classified as MUI, regardless of existence of SUI, UUI, Overflow, or Other incontinence diagnosis codes in the same year.
3. SUI	Person has SUI diagnosis code, regardless of existence of Overflow, UUI, or Other incontinence diagnosis codes in the same year.
4. UUI	Person has UUI diagnosis code, regardless of existence of Overflow or Other incontinence diagnosis codes in the same year.
5. Overflow UI	Person has Overflow incontinence diagnosis code, regardless of existence of Other incontinence diagnosis codes in the same year.
6. Other UI	Person has Other incontinence diagnosis code. An Other incontinence person is not allowed to have any other UI type diagnosis codes in the same year.
7. Otherwise	Person is NOT a UI person.

→ Incidence

Given the incident cohort selection process described in the previous subsection, the main outcome for incidence was the following:

$$\% \text{ of incident patients in year X} = \frac{\# \text{ of incident patients in year X}}{\# \text{ of restricted Medicare FFS individuals in year X}} \times 100$$

The restriction in the denominator refers to 36 months of continuous enrollment in the lookback period, age 65+, and residence in the United States (50 states plus District of Columbia); which were applied to all incident cohort analyses.

→ Comorbidities

Comorbidities were identified via the Clinical Classifications Software (CCS), Clinical Classifications Software Refined (CCSR), and Multi-level CCS grouping systems, as well as individual diagnosis codes based on contract team's clinical input. To maximize identification, we looked through claims for each person for the entire year. This multipronged approach allowed us to expand the comorbidity coverage from those covered in previous editions of the ADR. The CCSR grouper is a healthcare database tool, sponsored by the Agency for Healthcare Research and Quality (AHRQ), existing within a broader database suite developed in partnership with industry as well as state and federal government resources.⁸ The CCSR consolidates many ICD-10 codes into clinically informative procedural and diagnostic categories. In turn, users can perform targeted statistical analyses by selecting from these subsets. Diagnosis codes span over 530 clinical categories covering 21 human body systems, while procedure codes encompass over 320 clinical categories within 31 clinical areas.⁹ The full list of comorbidities for each disease is included in the Excel codebooks that accompany this ADR.

→ Diagnostic test

Diagnostic tests were evaluated for rolling incident cohorts. Specifically, an "incident patient" is defined in the same way as the incident cohort, with one additional enrollment restriction: for each year-month t, further impose continuous Part A/B enrollment in a [t-3, t+12] window surrounding the diagnosis month. An incident patient in year-month t is defined to have diagnostic test Y if a person registered a qualifying diagnostic test in claims during this evaluation window. The main outcome of interest is the following:

$$\% \text{ of incident patients in year X who had diagnostic test Y 3 months prior or 12 months after initial diagnosis} = \frac{\# \text{ of incident patients in year X who received diagnostic test Y 3 months prior or 12 months after initial diagnosis}}{\# \text{ of incident patients in year X}} \times 100$$

Note this computation built on but refined the methodology employed in previous editions of the ADR, where a diagnostic test is flagged as long as it occurred in the same calendar year as the incident diagnosis. Using a more refined evaluation window ensured that each incident case was evaluated for diagnostic tests on a [t-3,t+12] months rolling basis and ensured equal treatment for each diagnosis month. The full list of diagnostic tests for each disease is included in the Excel codebooks that accompany this ADR.

→ Prescription drugs

To compile the list of relevant drugs (based on clinical input), we selected drugs based on generic names from Medi-Span, which included all drugs approved by the Food and Drug Administration (FDA), and the CDM drug lookup table.¹⁰ The full list of prescription drugs and their pharmacological classes is included in the Excel codebooks that accompany this ADR.¹¹

We then identified patients who have full and continuous enrollment in Medicare Part D throughout each year. The number and percentage of Part D-enrolled patients among all patients were calculated, and then was used to estimate the number and percentage of Part D-enrolled patients who filled at least one prescription in any of the identified relevant pharmacologic classes, and for each individual pharmacologic class. For the CDM sample, the same prescription outcomes were calculated for all patients with the disease of interest, given that all enrollees have both medical and drug coverage.

Among the incident cohort aged 65 and older, we used the same list of drugs as those analyzed for the prevalent cohort and imposed on the incident cohort an additional restriction: continuous enrollment in Medicare Parts A, B, and D in the 60 months (5 years) after initial diagnosis (for incident year 2015). We then searched for whether any of these prescriptions were filled from the month of initial diagnosis to 60 months afterwards. We then computed the number and percentage of incident patients who met this criterion, the percentage shares of first filled prescriptions by type, and the average number of months from initial diagnosis to first filled prescription.

→ Procedures

In computing procedure use for a given person with BPH/USD/UI, a qualifying primary diagnosis for the disease on that procedure was required. Eligible claims refer to all inpatient institutional, outpatient, and carrier claims. For the prevalent cohort, the number and percentage of patients receiving each major procedure type were calculated. Further, the number and percentage of patients receiving any procedures were compiled. Note that given the more recent time coverage of this edition, some procedures that were introduced in recent years will be covered. The procedures analyzed for each disease were selected based on the contract team's clinical input. The full list of procedures for each disease is included in the Excel codebooks that accompany this ADR.

Among the incident cohort aged 65 and older, we used the same method and list of procedures as those analyzed for the prevalent cohort and imposed on the incident cohort an additional restriction:

continuous enrollment in Medicare Parts A and B in the 60 months (5 years) after initial diagnosis (for incident year 2015). We then searched for whether any of these procedures were performed from the month of initial diagnosis to 60 months afterwards. We then computed the number and percentage of incident patients who met this criterion, the percentage shares of first procedures by type, as well as the average number of months from initial diagnosis to first procedure performed.

→ Utilization of healthcare services

We considered four main service utilization outcomes: 1. Evaluation and Management (E&M) visits, 2. Emergency Department (ED) visits, 3. inpatient hospitalizations, and 4. observation stays.

To define E&M visits, we first flagged hospital-based outpatient and carrier claims with CPT codes from 99201 to 99499. Hospital-based outpatient claims are billed with Type of Bill = 13x. We then counted one occurrence of an HCPCS code as one E&M visit (based on revenue unit variable). E&M visits were then counted in each month in which the visits began. For each disease (BPH/USD/UI), only visits with any diagnosis for the corresponding disease were counted.

To define ED visits, we flagged claims in the outpatient and inpatient files identified via Revenue Center Code values of 0450-0459 (Emergency room) or 0981 (Professional fees-Emergency room). We then referred to the following related revenue center codes: 0450 = Emergency room - general classification; 0451 = Emergency room - EMTALA emergency medical screening services; 0452 = Emergency room - ER beyond EMTALA screening; 0456 = Emergency room-urgent care; 0459 = Emergency room-other.

We counted one occurrence of a relevant revenue center code as an ED visit. ED visits were counted in each month in which the visits began. For each disease (BPH/USD/UI), only visits from claims with primary diagnosis for the corresponding disease were counted.

To define inpatient hospitalizations, we first flagged the count of hospitalizations on inpatient hospital FFS claims. Inpatient settings included FFS claims with claim type "60". Inpatient hospital claims with overlapping service dates were considered a single hospitalization. We then considered any claims that began the day after another claim ended and also had the same provider type as a single hospitalization. Provider types included Inpatient Prospective Payment System (IPPS); Critical Access Hospitals (CAH); inpatient rehabilitation facilities; long term care facilities; inpatient psychiatric facilities; and other hospitals. Hospitalizations were counted in each month where the hospitalization stay (admission to discharge) overlapped with the month. For each disease (BPH/USD/UI), only hospitalizations with primary diagnosis for the corresponding disease were counted.

To define observation stays, we considered counts of inpatient and outpatient occurrences of the following related revenue center code: 0762 = Treatment or observation room-observation room. Observation stays were counted in each month in which the observation began. Only observation stays with primary diagnosis for the corresponding disease were counted.

ED visits were computed as percentage of patients with incident disease in each year who also had an ED visit within 12 months after initial diagnosis. Inpatient-hospital stays and observation stays were computed analogously. E&M visits, ED visits, inpatient-hospital stays, and observation stays were also computed on a per-incident-patient per-year basis.

→ Healthcare expenditures

Among the incident cohort aged 65 and older, we flagged FFS claims from the following settings: Inpatient, Skilled Nursing Facilities, Home Health, Hospice, Outpatient, Carrier, and Durable Medical Equipment. For each disease (BPH/USD/UI), we considered only claims with primary diagnosis for the corresponding disease. Expenditures were summed in each month in which the claim falls across patients with incident disease. We then calculated total FFS expenditures within the month of incident diagnosis and the 12 months after, across patients with incident disease in year X. Total expenditures were then scaled on a per-patient per-year basis. Expenditures are reported as raw Medicare paid dollars (nominal).

A.3 Validation

To help ensure quality of the data analysis, the validation process for the 2023 ADR combined several interconnected processes, involving consistency checks, code review, clinical review, literature review, and comparison to the previous ADR. This subsection provides an overview of this multi-pronged approach.

First, we evaluated frequencies within subgroups to ensure that the counts reflected appropriate levels of grouping, all ratios and their numerator and denominator components were consistent, and all sums across subgroups equated to any aggregates reported for the equivalent overall group. Second, programming code reviews were conducted during intermediate and final output stages to ensure that calculations reflect the intended strategy. Third, content was reviewed by clinicians to identify areas of data analysis that may not meet clinical expectations. Fourth, the clinical review was complemented by comparisons to previous findings reporting similar metrics or outcomes, while bearing in mind potential differences in methodology. This involved literature search and review based on relevant key word searches. Large discrepancies were flagged for additional review by the contract team's clinical experts. Fifth, content was compared to the latest previous ADR whenever possible, bearing in mind potential differences in methodology. In some instances, sensitivity checks were conducted to help gauge the precision of the results reported. Lastly, we developed interactive graphical charts on all outcomes to identify unusual patterns or rates and to aid comparisons described above. These graphic tools not only showed aggregate trends, but also those among subgroups. Unusual rates or trends that appeared to deviate significantly from overall trends were flagged for additional review (including the steps above).

A.4 Limitations

Medical coding changed from International Classification of Diseases (ICD)-9 to ICD-10 in 2015, with about 19 times as many procedure codes, and 5 times as many diagnosis codes in ICD-10- Clinical Modification (CM)/Procedure Coding System (PCS) than in ICD-9-CM; trends across years should be interpreted cautiously.¹² Despite the use of the extensive data for 100% of the Medicare FFS population, estimates may vary among populations in databases that do not have data sharing agreements with the sponsor, such as non-dually eligible Medicaid enrollees, persons with Medicare Part C, veterans who may not be enrolled in Medicare or who receive much of their care in the Veterans Affairs (VA) healthcare system, and persons who receive care from tribal entities with access to their own healthcare resources.¹³ Estimates from private insurance databases also may vary due to differences in cohorts' socioeconomic status and geographic differences that are well known to affect access to medical resources and related outcomes (see Section 1.4).¹⁴

Data engineering efforts are actively pursuing ways to better capture and improve the validity of sociodemographic determinants of health, inclusive of data on race and ethnicity.¹⁵ Like changes in ICD coding, future versions of the ADR will need to address how to compare trends across such variables with changing definitions. Further, the ability to assess longitudinal trends across a younger cohort from private insurance database to an older cohort from the Medicare database is affected by the inability to link unique persons over their transition from one system to another. Access to comprehensive data systems that span the age transition, like that of the VA, can - in part, at least - provide additional insights into how estimates vary by age. In addition, accurately capturing prevalence is further complicated by the COVID-19 pandemic, given the substantial avoidance of care observed in the healthcare system. Lastly, our prescription drugs analysis, which is based on prescription filled, is also challenged by the lack of information on actual prescription use.

Appendix B: Data

This appendix provides further details on the data sources used in the ADR (see Chapter 1 for brief summary). The two data sources are: Centers for Medicare & Medicaid Services (CMS) Medicare Fee-for-Service (FFS) data and Optum's de-identified Clinformatics® Data Mart Database (CDM) data. The specific files underlying these two data sources that were used in our analyses are described below in section B.1 (Medicare FFS) and B.2 (CDM).

B.1 Medicare FFS data

The data used in this ADR to study the age 65+ population is 100% of Centers for Medicare & Medicaid Services's (CMS) Medicare FFS data, supplemented by the Master Beneficiary Summary File (MBSF) and Part D data. The FFS data compose Part A/B claims from the following settings: Inpatient Stays (IP), Outpatient Services (OP), Home Health Agencies (HHA), Skilled Nursing Facilities (SNF), Hospice Care Organizations (HS), Durable Medical Equipment (DME, used for spending analyses only), and Carrier (PB). The Part D Event (PDE) Files contain the National Drug Codes (NDC) to identify prescription medication use. The MBSF was used to generate cohorts of study (i.e., denominator) when calculating rates and demographic stratifications. Claims data were used to calculate the metrics reported in this ADR, and are grouped as institutional and non-institutional claims. PDE files were used to analyze filled prescriptions. More details on these files are provided below:

1. Enrollment and Demographic Data
 - a. The CMS's MBSF contains demographic and enrollment information for each Medicare beneficiary. These datasets are used to create a demographic enrollment data for each Medicare beneficiary for multiple years. This dataset contains information for each Medicare beneficiary such as birth date, death date, sex, race, insurance coverage, and enrollment status.
2. Institutional claims
 - a. Institutional claims files contain records on FFS claims submitted by various health care institutions for reimbursement. A separate dataset exists for each type of institutional claim: IP, HHA, SNF, HS, and OP.
 - b. Each institutional claim, depending on the type of institution, has records of medical diagnostics (ICD diagnosis), procedure (ICD procedure code), FFS reimbursement amount, admission/discharge dates, dates of service, facility provider number, quantity of services, and type of visit (e.g. skilled care, home health aide, physical therapy).

3. Non-institutional claims
 - a. The non-institutional claims files contain records from DME and Carrier (previously Physician [part B] claims file) Files. The Carrier File includes FFS claims submitted by professional providers, including physicians, physician assistants, clinical social workers, and nurse practitioners.
 - b. Each non-institutional claim has records of medical diagnostics (ICD diagnosis), procedure (ICD procedure code), FFS reimbursement amount, dates of service, charges, and allowed amount.
4. Part D Event (PDE) Files
 - a. The PDE files include all transactions covered by Medicare prescription drug plans. These files include information on service date, product service ID (NDC code), quantity dispensed, number of days' supply, cost, payment, and linkages variables to the Characteristics Files.

B.2 CDM data

The data used to study the 18-64 person cohort were provided by Optum's de-identified Clinformatics® Data Mart Database (CDM). It includes paid claims and enrollment information for participants in private insurer plans of a large U.S. managed care health insurance company. The data also include information about plan members who are enrolled in both medical and prescription drug plan.

The CDM data consist of following files: Member Eligibility Files, Inpatient Confinement Files, Medical Claims Files, Pharmacy Claims Files, Lab Test Files, and Provider Files. Member Eligibility Files contain information on member demographics, geographic residence, and eligibility status. Inpatient Confinement Files summarize records for each inpatient serviced in an acute care hospital or a SNF. The Medical Claims Files summarize reimbursements in professional services provided across all places of services (e.g., inpatient and outpatient facilities, labs, physician office). Pharmacy Claims Files summarize prescription drug claims submitted by pharmacies in an outpatient setting. Lab Test Files provide information on lab tests. The Provider Files and Provider Bridge Files provide information on the providers. More details on these files are described below.

1. Member Eligibility Files
 - a. These files contain demographic information, enrollment dates, and insurance information of each member. Generally, it includes information for each member such as identifier, gender, birth information, insurance coverage (plan), race, family identifier, and geographic residence.

2. Inpatient Confinement Files

- a. These files include records of each inpatient episode occurring in an acute care hospital or SNFs of members. These files have records on diagnosis, payments, procedures, and costs. It includes member identifier, admit and discharge dates, admit and discharge diagnosis, length of stay, procedures (CPT, ICD, Revenue codes), diagnosis (ICD, DRG), facility details, and pricing.

3. Medical Claims Files

- a. Medical claims data for inpatient and outpatient professional services such as, outpatient surgery, laboratory, and radiology are included in these files. It includes person information as well as additional information on payment, admission, diagnosis, and procedures. These files contain member identifier, procedures (CPT, ICD, Revenue codes), diagnosis (ICD, DRG), admission and discharge dates, admit types, date and place of service, pricing, and denied claims.

4. Pharmacy Claims Files

- a. These files include claims submitted by pharmacies for prescriptions filled on an outpatient basis. The files have information on member identifier, drug dispensed (NDC), brand name, generic name, quantity and date dispensed, drug strength, days' supply, dollar amounts, and pricing.

5. Lab Test Files

- a. These files contain laboratory test results for all available lab tests, within certain networks (not comprehensive). Files include person identifier, lab claim ID, lab test name, lab sample dates, information on abnormalities, Logical Observation Identifiers Names and Codes (LOINC), and the lab result.

6. Provider Files and Provider Bridge Files

- a. Provider files categorizes providers by unique physicians or facilities. These files contain provider identifier, credentials, affiliations, and geographic location (state).
- b. Provider Bridge files includes information on providers not included in main provider files. These files include information such as Drug Enforcement Agency number, National Provider Identifier, and multiple provider codes.

-
- ¹ <https://www.niddk.nih.gov/about-niddk/strategic-plans-reports/urologic-diseases-in-america>
- ² <https://seer.cancer.gov/statfacts/>
- ³ Epping, Jelena, Siegfried Geyer, and Juliane Tetzlaff. 2020. "The Effects of Different Lookback Periods on the Sociodemographic Structure of the Study Population and on the Estimation of Incidence Rates: Analyses with German Claims Data." *BMC Medical Research Methodology* 20 (1): 1-15. <https://doi.org/10.1186/s12874-020-01108-6>.
- ⁴ Epping, Jelena, Siegfried Geyer, and Juliane Tetzlaff. 2020. "The Effects of Different Lookback Periods on the Sociodemographic Structure of the Study Population and on the Estimation of Incidence Rates: Analyses with German Claims Data." *BMC Medical Research Methodology* 20 (1): 1-15. <https://doi.org/10.1186/s12874-020-01108-6>.
- ⁵ Kimmel, Paul L., Chyng-Wen Fwu, Kevin C. Abbott, Jonathan Ratner, and Paul W. Eggers. 2016. "Racial Disparities in Poverty Account for Mortality Differences in US Medicare Beneficiaries." *SSM-Population Health* (2): 123-129. <https://doi.org/10.1016/j.ssmph.2016.02.003>
- ⁶ United States Census Bureau, Population Division, "Annual Estimates of the Resident Population by Sex, Race, and Hispanic Origin for the United States: April 1, 2020 to July 1, 2021 (NC-EST2021-ASR6H)" (June 2022), <https://www2.census.gov/programs-surveys/popest/tables/2020-2021/national/asrh/nc-est2021-asr6h.xlsx>
- ⁷ These codebooks are entitled, for each respective disease, "ADR_BPH_LUTS_Codebook_2023.xlsx", "ADR_USD_Codebook_2023.xlsx", and "ADR_UI_Codebook_2023.xlsx".
- ⁸ Agency for Healthcare Research and Quality. "Healthcare Cost & Utilization Project User Support." February 10, 2022. <https://www.hcup-us.ahrq.gov/overview.jsp>.
- ⁹ Agency for Healthcare Research and Quality. "Clinical Classifications Software Refined (CCSR)." December 29, 2022. https://www.hcup-us.ahrq.gov/toolssoftware/ccsr/ccs_refined.jsp.
- ¹⁰ Medi-Span is an online drug price reference and analytical tool. Medi-Span's drug database provides prescription and over-the-counter drug product data including drug name, strength, therapeutic class, drug pricing, NDC, and a proprietary Generic Product Identifier (GPI).
- ¹¹ See "ADR_BPH_LUTS_Codebook_2023.xlsx", "ADR_USD_Codebook_2023.xlsx", "ADR_UI_Codebook_2023.xlsx".
- ¹² Centers for Disease Control and Prevention, National Center for Health Statistics. "International Classification of Diseases, (ICD-10-CM/PCS) Transition – Background." https://www.cdc.gov/nchs/icd/icd10cm_pcs_background.htm.
- ¹³ For example, Original Medicare and Medicare Advantage participants are different in-patient characteristics. The former population is generally sicker. There is also considerable geographic variation between Original and Part C selection rates – as with over 90% of Puerto Ricans opting for Advantage to around only 1% of Alaskans. For more discussion see Freed, Meredith, Jeannie Fuglesten Biniek, Anthony Damico, and Tricia Neuman. "Medicare Advantage in 2022: Enrollment Update and Key Trends," Kaiser Family Foundation. <https://www.kff.org/medicare/issue-brief/medicare-advantage-in-2022-enrollment-update-and-key-trends/>.
- ¹⁴ The limitations of race variable in the CDM data have also been documented in the literature. See Nead, Kevin T., Candice L. Hinkston, and Mackenzie R. Wehner. 2022. "Cautions When Using Race and Ethnicity in Administrative Claims Data Sets." *JAMA Health Forum* 3 (7): e221812. <https://doi.org/10.1001/jamahealthforum.2022.1812>.
- ¹⁵ CMS Office of Minority Health. 2022. *The Path Forward: Improving Data to Advance Health Equity Solutions*. Baltimore, MD: Centers for Medicare & Medicaid Services.