



Outpatient Prescription Drugs: Data Collection and Editing in the 2021 Medical Expenditure Panel Survey

Methodology Report #37 | January 2024 | Salam Abdus, PhD, Steven C. Hill, PhD, and Rebecca Ahrnsbrak, MPS

Abstract

The Medical Expenditure Panel Survey (MEPS), sponsored by the Agency for Healthcare Research and Quality (AHRQ), is a nationally representative survey of the U.S. civilian noninstitutional population's medical care use and expenditures. Household respondents report prescription drugs obtained by members of the household and the number of times each drug was obtained, while a follow-back survey of pharmacies is the primary source of prices, payers, and drug attributes. This report describes the household and pharmacy data collection and editing processes, the editing and imputation techniques used to supply values for missing data in the pharmacy database, the procedure linking the data reported by the pharmacy to each prescription drug reported by the household, and recent improvements in these procedures. Statistics on these methods are presented for the 2021 MEPS.

Table of Contents

- Introduction..... 1
- Household Component 3
 - Data Collection..... 3
 - Editing Household Component Data 4
- Pharmacy Component 5
 - Data Collection..... 5
 - Supplemental Data 6
 - Editing the Pharmacy-Reported Data 6
 - Imputing NDC* 7
 - Imputing quantity dispensed* 8
 - Editing third-party payers* 8
 - Identifying acquisitions needing price imputation* 8
 - Imputing price and payments*..... 10
 - Results of Preparing the Pharmacy-Reported Data 12
- Matching Pharmacy Data to Household Data..... 12
 - Overview of Matching..... 12
 - Details of Matching..... 12
 - The first approach: Within-person matching* 12
 - The second approach: Imputation from a different person* 13
- Editing Matched Data..... 14
 - Imputed fills with too many days supplied*..... 14
 - Free antibiotics, anti-diabetics, and prenatal vitamins* 15
 - Prices paid by CHAMPVA*..... 15
 - Federal pharmacy prices* 15
 - Editing year in crossover rounds*..... 16
 - Medicare Part D and private insurance*..... 16
 - Resolving payer inconsistencies arising from imputation*..... 16
 - Additional variables*..... 17
 - Editing for confidentiality*..... 17

Outpatient Prescription Drugs: Data Collection and Editing in the 2021 Medical Expenditure Panel Survey

Salam Abdus, PhD, Steven C. Hill, PhD, and Rebecca Ahrnsbrak, MPS

Introduction

The Medical Expenditure Panel Survey (MEPS) is a nationally representative survey of medical care use and expenditures sponsored by the Agency for Healthcare Research and Quality (AHRQ) covering the U.S. civilian noninstitutional population. MEPS includes a Household Component (HC) and a Medical Provider Component (MPC).^a MEPS data can produce estimates for individuals, families, and selected population subgroups. The HC provides data on health status, demographic and socioeconomic characteristics, employment, access to care, and experiences with healthcare. The MPC collects expenditures data from hospitals, physicians, home healthcare providers, and pharmacies identified by MEPS-HC respondents. Its purpose is to supplement the HC information.

Each year, a new panel of households is selected and participates in five rounds of HC interviews, covering 2 full calendar years.^b This set of households is a subsample of households that participated in the previous year's National Health Interview Survey (NHIS) conducted by the National Center for Health Statistics (NCHS) of the Centers for Disease Control and Prevention. The NHIS sampling frame provides a nationally representative sample of the U.S. civilian noninstitutionalized population.

MEPS supports longitudinal analysis; each panel typically consists of five rounds of interviews to collect 2 full calendar years of data. It is possible to analyze even more long-term trends by linking MEPS to the previous year's NHIS. Cross-sectional analysis over a long period, going back to 1996 (when MEPS began), is possible as well.

MEPS defines expenditures as payments from all sources (including individuals, private insurance, Medicare, Medicaid, and other sources) for healthcare services during the year. To construct estimates of expenditures, MEPS collects detailed information about medical events including ambulatory visits (outpatient and office based), inpatient stays, emergency care, home healthcare, dental care, other medical care, and prescription medicines. At each HC interview, the household respondent supplies the names of any prescribed medicine that family members obtained and identifies the pharmacies used. In addition, family members are asked for permission for MEPS to contact the pharmacies. With this permission, the Pharmacy Component (PC), which is a subcomponent of the MPC, collects detailed information from the pharmacies about the drugs obtained, including payments (the sum of which equals the drug price), payers, date each prescription was filled, quantity dispensed, the National Drug Code, and precise drug attributes. Matching drugs reported by the pharmacies in the PC to those reported by the household in the HC is accomplished through the use of supplemental data and specialized software.

The MEPS conducts the PC to collect information that pharmacies can more easily and accurately provide than household respondents. Because pharmacies receive payments for the drugs they dispense, they are generally likely to have more accurate payment information than households, which may be more familiar with their out-of-pocket payments than third-party payments. Some HC respondents may lack documentation, such as explanations of benefits, with details about third-party payments. The PC collects information about payments for drugs, but data users should note that rebates between manufacturers and pharmacies, pharmacy benefit managers, and government programs are not collected. A second motivation for conducting the PC is that some households do not have easy access to details about their medications, such as the number or strength of pills. In addition, some households have many fills for multiple medications during the reference period, and asking them for information about each of these fills would be repetitive and overly burdensome.

^a MEPS also has a survey of employers referred to as the Insurance Component (IC).

^b In 2020, 2021, and 2022, data were collected from some older panels beyond their second year because of the COVID-19 pandemic. See below for more details.

As with all surveys, inconsistent and missing data occur in MEPS. Errors and failures to report information happen. Some pharmacies do not respond to the PC, and some sample members deny permission to contact pharmacies, so PC data are not available for every sampled person. Even when pharmacies respond, the data provided are not always complete.

This report describes the methods used to (1) supply values for missing HC and PC data and (2) link the data reported by the pharmacies to each prescription drug name reported by the household. Table 1 summarizes the key variables edited or imputed. Details of edits and imputation are in the body of the report. The objective of these procedures is to maximize the amount and quality of data available for analysis and to reduce the risk of bias associated with reporting error and missing data. Statistics on these methods are presented for the 2021 MEPS.

This report updates an earlier Methodology Report on the 2011 MEPS prescription drugs.¹ Since the 2011 data were released, several enhancements have been made to improve the quality and analytical capability of the MEPS prescription drug data. Some improvements implemented include the following:

- Beginning with the 2013 data, a new drug name variable supplied by the Multum Lexicon Plus database of Oracle Health was added to the public use file. This drug name is the generic name of the drug most commonly used by prescribing physicians. For the earlier years, this variable has been provided with Addendum Files to MEPS Prescribed Medicines Files for 1996–2013.
- Beginning with the 2017 data, higher imputed prices were allowed to account for the rising prices of specialty drugs. In 2017, this change in editing procedures accounted for more than 95 percent of the increase in total expenditures for prescribed medicines relative to 2016.
- MEPS instrument redesign in 2018 aimed to improve data collection for all sections of the MEPS HC, including (1) prescribed medicines that sample members obtained earlier in the reference period but were no longer taking and (2) medications to treat diabetes and asthma.
- Starting with the 2018 data, the pharmacy types are those reportedly used by the person. For prior data years, the pharmacy types are those reportedly used by anyone in the household.
- Beginning with the 2020 data, the rules used to identify outlier prices for prescription medications were improved based on newer price benchmarks and analyses.² New outlier thresholds were established based on the distribution of the ratio of retail unit prices relative to the National Average Drug Acquisition Cost per unit, which better reflects the prices paid for drugs. As a result, the prices paid for generics are lower in the 2020 data, compared with the 2019 data, and fewer generic fills have third-party payments.

The importance of these changes depends on the research focus. For cross-sectional studies focused on a single year of data, these changes may not be of interest to general users. However, studies that pull data across years or examine trends over time should pay attention to methodological changes that may affect the trends. In particular, attention should be paid when comparing 2016 data to 2017 data (allowing higher imputed prices may account for part of the increase in expenditure), 2017 data to 2018 data (instrument redesign might affect trends), and from 2019 data to 2020 data (effects of the COVID-19 pandemic on data collection, as well as changes in price editing for generics).

This report is organized as follows. First, it describes data collection for the HC and the procedures used to edit the HC data. Second, the discussion turns to data collection for the PC and the procedures employed for editing and imputing data in the PC. Finally, an overview and details of matching pharmacy data to household data and the methods of editing the matched data are presented.

Household Component

Data Collection

The MEPS-HC uses a panel design with two overlapping cohorts of the U.S. noninstitutionalized civilian population combined to produce annual estimates.^{3,4,5} A new cohort of households begins each year and is interviewed five times to collect 2 calendar years of data. Additionally, due to the COVID-19 pandemic, the panels that began in 2018 and 2019 (panels 23 and 24, respectively) each were extended to nine rounds of data collection covering 4 calendar years. As a result, the 2021 MEPS includes data from four panels.

A single respondent in each household typically provides data about each household member during each interview. The MEPS asks that this person be the family member most knowledgeable about health and healthcare use. In an interview, the average recall period (the “round”) is 5 months. The respondent is given a calendar and other materials to aid memory and then asked to retain paperwork such as receipts and insurance benefit statements.

During each interview, the HC gathers information on healthcare services used, including prescribed medicines. The first opportunity to report prescription drugs occurs when the HC asks about nonprescription drug health service events that took place during the round. The respondent supplies the names of any medications prescribed as part of a visit to an inpatient stay, emergency room, hospital outpatient clinic, or doctor’s office and subsequently filled. For hospital stays, the respondent is asked to report only drugs that were prescribed on discharge. The HC does not collect information on drugs administered in healthcare settings.

Another opportunity to mention prescriptions occurs in the dedicated prescribed medicines section of the survey where the respondent can identify prescription medicines not already mentioned. This section of the survey was revised in 2018 to better capture drugs associated with two priority conditions, diabetes and asthma, as well as drugs that sample members are no longer taking. If a respondent reports that a sample member has ever been diagnosed with diabetes, they are asked to report “insulin or any other prescribed medications related to [their] diabetes.” They are then asked whether they obtained “any other diabetic equipment or supplies, typically prescribed by a physician.” (Note that any diabetic supplies purchased without a prescription represent a nonprescription addition to the prescription drug expenditure and utilization data.) If a respondent reports that a sample member has ever been diagnosed with asthma, they are specifically asked whether they “obtained any prescribed medicines related to their asthma.” For each sample member, household respondents are then asked to report (1) “any new prescriptions or refills” obtained at any pharmacy, including mail order or online, and (2) “any other prescriptions, even if [they are] no longer taking the medicine or only take it as-needed.”

In all these sections of the HC interview, the respondent is encouraged to consult records such as medicine bottles or receipts, so the reported medication name is often quite specific. However, the information can also be minimal, for example, “pain pills.” In spring 2022, the HC implemented a searchable lookup tool with commonly reported drugs pre-programmed into the survey instrument. Drugs that are not included in the lookup tool are manually entered as text strings. When drugs are reported, the drug names and any available information on strength and form are entered in a dynamic roster.

In addition to the drug name, the HC collects the following information about each medicine:

- The number of times it was obtained in the round.
- The health condition it was prescribed for (if any) (asked only in the first interview in which it is mentioned).
- The year and month in which the person first used it (asked only in the first interview in which it is mentioned).

The names, addresses, and types of pharmacies that filled each household member’s prescriptions are also requested, along with permission for MEPS to acquire data from these pharmacies. Signed authorization forms allow pharmacies to respond to the PC of MEPS when it is conducted. For the interviews collecting information

about medication obtained in 2021, 58.5 percent of pharmacy permission forms were signed.

In 2021, household respondents reported 144,783 unique drug names obtained by family members in a round (hereafter referred to as “person-round-drugs”), including insulin and diabetic supplies.[°] The end result of HC data collection is a roster of drug names that each sample person obtained in the round, the person’s health condition or conditions associated with each drug, the number of acquisitions, and a roster of the person’s pharmacies.

Because the HC collects no information on expenditures and limited information on drug characteristics, the expenditure information and other details about each prescription medicine—quantities, form, strength, dates obtained, price, payments by source, etc.—come from the PC. A supplementary data source (Multum Lexicon Plus database of Oracle Health) provides information on therapeutic classes and pregnancy category.

Editing Household Component Data

Editing the HC drug data consists mainly of imputing a value for the number of times a drug was obtained in the round when this information is missing or invalid as collected in the HC.

GPI coding. A first step in preparing the HC data for editing is to code each drug name to a Generic Product Identifier (GPI). A proprietary dataset, Wolters Kluwer’s Master Drug Data Base (MDDB®)⁶ Version 2.5 provides the GPI, a 14-digit code that identifies groups of pharmaceutically equivalent drugs that have the same active ingredient(s), dosage form, and strength. The GPI is the common variable linking the PC data to the HC. It allows information to be drawn from other sample members (imputed) in cases of missing information. The first pair of digits in the GPI represents the drug group, the second pair represents the therapeutic class, and the third pair represents the therapeutic subclass. The next four digits of the GPI (digits 7–10) represent the active ingredient(s). The first 8 digits of the GPI are sufficient to identify most active ingredients, but 10 digits are needed for compounds and salts. The remaining digits of the GPI represent dosage form and strength. For the HC, professional coders use the MDDB to identify as many digits of the GPI as possible based on the medication name and any other information provided by the respondent. Typically, 8 to 10 digits can be coded for household-reported drug names, but in 2021, 5,702 (3.9 percent) of person-round-drugs could not be coded.

Imputing the number of acquisitions. The number of times a drug identified by name in the HC was obtained by a person in a round (person-round-drug) needs to be imputed when it is missing or extreme. In the 2021 data, for 4.8 percent of household-reported drug names, the respondent did not know or remember the number of times the drug was obtained during the round. Outlier values for the number of times a household reported obtaining a drug in a round occur as well and are determined by comparing the number of days in the round and the number of acquisitions of the drug reported in the round. Person-round-drugs with more than five fills per month were automatically deemed invalid, with limited exceptions. For each drug, the maximum fills per month across all person-rounds with pharmacy data are also used to label outliers in the household data. Person-round-drugs for which the drug names reported by the household cannot be coded to a GPI, and for which more than one and no more than five fills per month were reported, are reviewed by a pharmacist to assess their plausibility.

For missing and implausible values, a hot-deck procedure imputes a new number of acquisitions, drawing from the donor pool of drugs with valid values. Specifically, days per acquisition are calculated in the donor pool, imputed to recipients with replacement, and then the recipient’s days in the round are divided by the days per acquisition to calculate the number of acquisitions in the round. Four attempts are made to match a person-round-drug to a recipient person-round-drug. The first attempt requires an exact match on drug, dosage form, and strength (14-digit GPI). The second attempt requires an exact match on drug and dosage form (12-digit GPI). The third attempt requires an exact match on drug (8-digit GPI). The fourth attempt requires an exact match on drug group (2-digit GPI), and therapeutic class and subclass are weighting variables in the match.

Weighted match variables used in all four attempts include whether the person used any mail-order pharmacies; insurance status and potential payment sources for the person; private and public health maintenance organization

[°] Each person-round-drug could be acquired one or multiple times during the round.

(HMO) enrollment; length of the round (in months); health conditions treated by any drug; whether the person is enrolled in a high-deductible health plan; and the person's age, state of residence, census division, region, and urbanicity. The weight for each variable is based on its relative contribution to the variation in days per acquisition. Because patients usually have only one fill of an antibiotic, the imputation approach is a little different. Rather than imputing days per fill, the total number of fills is imputed. The donor pool consists of antibiotics with valid values, and imputation does not require an exact match on any variable. Similarly, for person-round-drugs with names that could not be GPI coded, imputation does not require exact match on any variable.

Table 2 summarizes the results of the process of imputing acquisitions in the 2021 HC. Most of the imputations (93.4 percent) were based on exact matches to pharmaceutically equivalent drugs (with the same active ingredients, dosage form, and strength). After this imputation, the 2021 data file contained 333,859 acquisitions.

After these edits and imputation, the HC data are composed of each drug named during every HC interview. These data include a GPI code and the number of acquisitions. The data are now ready to be linked to the payment and other drug details from the PC.

Pharmacy Component

Data Collection

The PC collects data from the pharmacies identified by HC respondents as places where household members obtained prescription medicines during the calendar year. During the second and subsequent HC interviews, sample members are asked to sign permission forms authorizing MEPS to contact their pharmacies and authorizing the pharmacies to release records to MEPS. Only those pharmacies for which an HC sample member signed this permission form and that voluntarily participated are included in the PC.

The data collection protocol for pharmacies is as follows. For small retail pharmacies—that is, pharmacies not associated with a large chain—the data collection staff contacts the pharmacy to explain the study's purpose, verifies receipt of authorization forms, requests specific data elements, and determines whether patient profiles are available. The data collection staff may collect information by telephone or request that the data be sent by fax or mail or submitted electronically so that the necessary data elements can be abstracted. Pharmacy data are received in any format, including hardcopy patient profiles, electronic files with patient profile data, or reports via telephone.

For large retail pharmacy chains, individual pharmacies are grouped by chain. Negotiators follow up with these pharmacies in one of two basic ways:

- If the corporate office of the retail chain prefers that the local stores respond, data collection follows the small retail model.
- If the pharmacy prefers that the data request be handled with a regional or central contact, the negotiator facilitates the most efficient method for data collection.

The Veterans Health Administration (VA) and AHRQ have an interagency agreement for the VA to extract information from centralized administrative data for MEPS HC sample members who sign authorization forms for the VA.

Each pharmacy is asked to provide patient profiles or information about each prescription filled or refilled for each patient during the calendar year. For each acquisition of a drug, the following information is requested:

- The dates the prescription was filled or refilled.
- The National Drug Code (NDC).
- The quantity dispensed.
- The number of days supplied.
- The amount of out-of-pocket payment.

- Whether there were any third-party payments, and, if so, the third parties and payment amounts.

If the pharmacy does not provide the NDC, the PC asks instead for the medication name, dosage form, strength, and strength unit.

In addition, pharmacy patient profiles are requested directly from HC family members who reported using one or more of several pharmacy chains that had repeatedly refused to participate in the PC. Families are mailed a request for profiles. To minimize burden on households, the mailing is limited to households that had completed their participation in all rounds of the MEPS, and no follow-up attempts are made for those who do not reply to a mail request. For the 2021 data collection, patient profile requests were mailed to 511 families with 720 patient-pharmacy pairs. Completed profiles were received for 11.1 percent of the pairs. Despite the overall low rate of return, the effort does provide a number of profiles for patients associated with pharmacy chains that otherwise would not be represented in the pharmacy data.^d

Table 3 presents the sample size and participation rates in the 2021 PC. Conditional on having obtained 58.5 percent of signed permission forms sought in the household interview, more than four out of every five eligible pharmacies responded. The responding pharmacies provided data on 257,596 unique acquisitions of drugs during 2021.

To the extent allowed by survey response, the end result of PC data collection is a roster of each HC household member's drug acquisitions during a calendar year. The information that the PC seeks to collect about each of these acquisitions includes the transaction date, the NDC (or analogous identifying drug information), the quantity, the number of days of the drug that the acquisition supplied, and payments to the pharmacy by source.

Supplemental Data

As mentioned above, matching the pharmacy data to the household data requires assigning a common set of codes to the household- and pharmacy-reported drugs. Wolters Kluwer's proprietary Master Drug Data Base (MDDDB[®]) Version 2.5 is the source of such a code, the Generic Product Identifier (GPI), also described above. For the PC, automated and manual coding generally succeed in recording all 14 digits of the GPI based on the NDC (which is more specific than the GPI) or the medication name, dosage form, and strength. A negligible number of acquisitions in the PC cannot be coded to a GPI based on the information the pharmacy provides.

The MDDDB has several other uses besides GPI coding. Information about the patent status and the average wholesale unit price (AWUP) and wholesale acquisition cost unit price (WACUP) used in price editing and imputation come from the MDDDB. The MDDDB is also used to impute the NDC where necessary. (Price editing, price imputation, and NDC imputation are discussed in more detail below.) If the pharmacy reports an NDC found in the MDDDB, the MDDDB provides information on the drug name, dosage form, and strength.

Another important supplemental source of information used for pharmacy data editing is the National Average Drug Acquisition Cost (NADAC) developed by the Centers for Medicare and Medicaid Services. Starting with the 2020 MEPS data, NADAC is used as the primary price benchmark to identify implausibly high or low drug prices (outliers).

Two other datasets are used in editing the pharmacy data. A price list obtained from the VA that includes prices for TRICARE, the VA, and the Indian Health Service is used to determine prices of drugs dispensed by federal pharmacies when total payments are missing. The Multum Lexicon Plus database^e of Oracle Health⁷ provides variables for therapeutic class. These variables are merged into the PC data file by NDC.

Editing the Pharmacy-Reported Data

To maximize the donor pool for imputation, the entire PC data file (257,596 acquisitions in 2021) is edited before it

^d Families were first asked to provide these profiles for the 2004 MEPS data. They were not asked for profiles for 2020, during the COVID-19 pandemic.

^e Cerner Multum, Inc. created the Multum Lexicon Plus database but was recently acquired by Oracle Health.

is matched to the HC. Key aspects of this process include imputing missing NDCs and missing quantity dispensed, editing ambiguous third-party payer information, identifying acquisitions with deficient payment information, and imputing missing payment information.

Most editing and imputation are conducted at the acquisition level, but there are exceptions. For imputing NDCs and matching them to the HC, acquisitions are aggregated to the person-round-drug level. That is, for each person and round, all pharmaceutically equivalent acquisitions (having the same 14-digit GPI, which represents the active ingredient, dosage form, and strength) are grouped together. The GPI can group multiple NDCs, such as brand name drugs and their generic equivalents.

Imputing NDC

Among the 257,596 acquisitions reported by pharmacies in the 2021 data, 7,573 (2.9 percent) had a missing or an invalid NDC. A missing NDC is imputed rather than left missing because it distinguishes between brand name drugs and generics, which is important for identifying price outliers and imputing missing payment data. Furthermore, many analyses of medication use depend on the NDC to identify medications of interest. The NDC can identify the manufacturer, the patent status, and the average wholesale price of a drug. The NDC, therefore, increases the accuracy of any price imputation and the analytic utility of the data.

MEPS uses three approaches to identify the NDCs of acquisitions missing this item in the PC. First, wherever feasible, the NDC is taken from one of the person's other acquisitions in the same round. For each person and round, all acquisitions of drugs with the same 14-digit GPI (representing the active ingredient, dosage form, and strength) are grouped together. Within these groups, if at least one acquisition is missing an NDC and at least one acquisition has an NDC, then the missing NDC is assigned the reported NDC. If more than one NDC in the 14-digit GPI group exists, one is selected at random.

Second, for the residual acquisitions still missing NDC, the MDDDB supplies it. The basic approach is to use matching software to find the best match to a drug product on the MDDDB based on the medication name, quantity units (for example, milliliters), and the GPI. Medication name is used as a match variable because various brand name and generic products, each with a unique NDC, can be grouped together under a single GPI. When a pharmacy reports NDCs, it almost always reports the same NDC for all the person's acquisitions of a drug; so, to preserve homogeneity among the recipient acquisitions, one NDC is imputed to all the acquisitions for the same person-round-drug.

Three attempts are made to impute an NDC by matching to the MDDDB. In the first attempt, one NDC is taken from the MDDDB for all acquisitions with the same person, round, active ingredients, dosage form, and strength. In other words, an exact match on 14-digit GPI is required. For the second attempt, one NDC is imputed from the MDDDB to all acquisitions with the same person, round, and active ingredients (10-digit GPI). Here, the dosage form and strength may differ. An exact match on active ingredients is required, and, again, the medication name is a match variable. The third attempt is similar to the second attempt, except that an exact match on drug name (2-digit GPI) is used.

For the residual acquisitions missing NDCs and without valid GPIs, NDCs are imputed from other pharmacy records. Most of these acquisitions lack the drug name as well as any information about the quantity dispensed and the strength. In these cases, an NDC is imputed from the acquisitions of a person with similar characteristics reported in the HC and for whom the pharmacy reported both drug names and valid NDCs. Match variables include the person's age, sex, geographic division, urbanicity, health conditions, and potential payment sources. In addition to the NDC, the recipient acquisition adopts any missing medication name, GPI, quantity dispensed, or strength.

Table 4 summarizes the results of the NDC imputation process for the 2021 data. Some imputations were from other acquisitions of the same person in the same round. Most of the imputations relied on detailed information provided by the pharmacies.

Imputing quantity dispensed

Pharmacies report quantity dispensed for nearly all acquisitions, but occasionally it must be imputed. If the NDC is imputed and the quantity is missing, then the quantity is taken from the same acquisition that donated the NDC. Otherwise, matching software imputes a quantity from another acquisition. Match variables include the NDC; active ingredients, dosage form, and strength (GPI); and characteristics of the person reported in the HC (age, sex, health conditions, and health status). Exact matching on the drug is required, and heavier weight is placed on the NDC, followed by the dosage form and strength. In the 2021 data, the quantity dispensed was imputed for 0.5 percent of the 257,596 acquisitions (Table 1). Subsequently, quantity was occasionally edited to create greater consistency between price and quantity and greater consistency between number of fills and quantity dispensed.

Editing third-party payers

Sometimes the pharmacy does not clearly identify a third-party payer. For example, a pharmacy might report as a payer a pharmacy benefit management company, which could be under contract with a private insurer, Medicare prescription drug plan, Medicaid, or other public program. For some acquisitions, pharmacies do not report the third-party payer at all. In these cases, information from the HC about insurance coverage and usual third-party payer can indicate the type of payer. For example, if the respondent in the HC reports Medicare Part D as the usual third-party payer, then the unknown payer in the PC is set to “Medicare.”

In the 2021 data, among the 257,596 acquisitions reported by pharmacies, the third-party payer was edited for 16.5 percent of acquisitions (Table 1). Specifically, the category of third-party payer was assigned in three situations using the insurance status reported in the HC:

- The third-party payer was reportedly a pharmacy benefit management company or HMO (39,500 acquisitions). The payer in these cases was coded mainly as private insurance, Medicare, or Medicaid.
- The payer reportedly was a public clinic or some other ambiguous public payer (2,498 acquisitions). The payer in these cases was coded mainly as Medicaid, Medicare, or Other State and Local.
- The third-party payer was unknown (484 acquisitions). The payer in these cases was also coded mainly as private insurance, Medicare, Medicaid, or Other Insurance.

Additional edits to the PC acquisitions (patterned after edits applied to the nonprescription provider data) include the following:

- Correcting the information from pharmacies that mistakenly report a payment from private insurance instead of a Medicare prescription drug plan (910 acquisitions), Medicaid (971 acquisitions), or TRICARE (647 acquisitions).
- Eliminating the inconsistency of pharmacy providers incorrectly reporting a Medicare source of payment instead of a Medicaid payment for those nonelderly who are not Medicare beneficiaries (3,003 acquisitions).
- Eliminating the inconsistency of pharmacy providers incorrectly reporting a Medicaid source of payment instead of a private source of payment for people with private insurance (1,021 acquisitions).

Identifying acquisitions needing price imputation

The primary aim of MEPS is to collect information about healthcare expenditures, including the costs of prescription drugs. Payment information that is altogether or partly missing, incorrectly reported as zero, or otherwise inaccurate will cause distortion in estimated expenditures. Responding pharmacies provided all the information requested for 56.3 percent of acquisitions in 2021 (Table 5). In many cases, the source of incompleteness of payment data is that the third-party payment amounts are missing. Federal pharmacies for a few federal programs are especially likely to omit third-party payment amounts, because these pharmacies report little payment information generally.

Even when payment data appear to be complete, MEPS attempts to detect inaccurate payment data by comparing an

acquisition's price (the sum of payments) to prices provided in the NADAC data and MDDB. If the sum of payments falls outside certain ranges relative to the benchmarks, then MEPS imputes a price and distributes it among sources of payment.

The threshold used for defining what constitutes deficient payment information depends on the pattern of reported payments. Generally, deficiency comes under one of four categories:

- The acquisition lacks any payment information at all.
- The acquisition has missing third-party payments.
- The acquisition has zero third-party payments, and the out-of-pocket amount is unrealistically low as a total price.
- The acquisition's payments sum to an unrealistically high or low price.

The primary method for identifying unrealistically high and low prices ("price outliers") is to assess prices per unit (tablet, capsule, ounces, etc.). The retail unit price (RUP) is the sum of payments for the acquisition divided by the quantity dispensed. Price outliers are identified by comparing the acquisition's RUP reported in the PC to the NDC's national average drug acquisition cost (NADAC) per unit. When NADAC is not available, then average wholesale cost unit price (WACUP) is used. When neither NADAC per unit nor WAC per unit is available, average wholesale unit price (AWUP) is used.

Rules for identifying lower outliers are summarized in Table 6a. When using NADAC per unit or WACUP, the rules depend on patent status, the completeness of the payment data, whether the pharmacy reported discounts or coupons for the fill, and whether the fill was for a person with Medicare Part D who appeared to be in the donut hole. The thresholds are lower for brand name drugs obtained by Medicare Part D beneficiaries who have spent enough to enter the donut hole. The thresholds are lower because, starting in 2019, these acquisitions are discounted 60 percent. Acquisitions in the donut hole are identified using the cumulative, total out-of-pocket payments reported by the pharmacies for acquisitions covered by Part D. The rules for identifying lower outliers also depend on whether the drug was an over-the-counter (OTC) drug and whether there was any third-party payment. The three patent statuses are (1) single source brand name drugs, which are available only from one manufacturer; (2) originator drugs, which are brand name drugs with therapeutically equivalent competitors; and (3) generic drugs, which are available from multiple manufacturers and pharmaceutically equivalent to a brand name drug. The differences between the thresholds for NADAC per unit and WACUP reflect differences in the distributions of RUP relative to NADAC per unit compared with RUP relative to WACUP.² When using AWUP, which is used only as a last resort for identifying outliers, no information regarding OTC or third-party payment is used.

Rules for identifying upper outliers are summarized in Table 6b. The rules depend on patent status and dosage form. Acquisitions are flagged as upper outliers when the RUP exceeds 50 times the NADAC per unit for generics, 8 times the NADAC per unit for single source liquids, and 4 times the NADAC per unit for all other drugs among fills priced at \$16 or more per fill. When using the WAC, cases are flagged as upper outliers (prices too high) when the RUP exceeds 20 times the WACUP for generics, 4 times the WACUP for single source liquids, and 2 times the WACUP for all other drugs. When using AWUP, upper outliers are identified as $RUP \geq 10$ times AWUP, regardless of patent status or dosage form.^f

2021 Results. Table 7 presents the number of acquisitions identified as potentially needing and definitely not needing price imputation in the 2021 data, based on the rules for identifying upper and lower outliers in unit prices. Few of the acquisitions with complete payment data, 1,046 of 144,929 or 0.7 percent, were flagged as lower outliers, and a few more, 2,758 out of 144,929 or 1.9 percent, were flagged as upper outliers. Those flagged as lower outliers were more likely to be generic (772 out of 1,046) than brand name, whereas about two-thirds of those flagged as upper outliers were single source drugs (1,799 out of 2,758). For most of these lower outliers, imputed

^f Until 2019, only AWUP was used to identify price outliers. To see why NADAC and WAC were chosen as preferred price benchmarks, see Ding & Hill (2022).

prices replaced reported prices (865 out of 1,046), mainly because the imputed price was more than the reported price.^g Most of those were cases with zero third-party payment amounts (615 out of 865). About two-thirds of the upper outliers in unit prices were deemed to not need any imputation (1,854 out of 2,758), for example because the total price of the fill was less than \$16.^h Most of those upper outliers flagged for imputation had positive third-party payment amounts (834 out of 904).

More than half of the acquisitions with partial payment data, 34,579 of 63,297, were flagged as lower outliers. A very small number of the acquisitions with partial payment data were flagged as upper outliers (69 out of 63,297). For acquisitions with partial payment data not identified as lower or upper outliers, missing payments were set to zero. For a small proportion of partial payment data flagged as lower outliers that appeared to be partial fills of a small number of pills, missing values were set to zero.ⁱ

Imputing price and payments

Donors for price imputation are the acquisitions with complete price data not identified as outliers and the acquisitions with partial price data whose missing values are set to zero.^j Eight attempts are made to match a donor acquisition to a recipient acquisition. The first two attempts are the highest-quality imputations, where there is a donor with the same NDC, so that the RUP can be directly imputed. The first attempt requires an exact match on NDC and the set of payers, and the second requires an exact match on NDC. The remaining match attempts impute ratios of the RUP to the NADAC per unit, the WACUP, or the AWUP, depending on whether the NADAC per unit (third and fourth attempts) or the WACUP (fifth and sixth attempts) are available, and in cases when they are not available, the AWUP (seventh and eighth attempts). The third, fifth, and seventh attempts require an exact match on GPI and patent status. The fourth, sixth, and eighth attempts requires an exact match on patent status, package unit, and discount flag.

All records flagged as in or after the donut hole are excluded from the donor pool for the second through eighth matches, because if they matched to fills outside the donut hole, then the imputed prices would be too low. Also excluded from the donor pool for the third through eighth matches are all records for generics with high price ratios, because the high ratios are more common among very inexpensive drugs, but these ratios could get matched to expensive drugs. The match variables in all eight attempts are third-party payer, Medicare Part D low-income subsidy participation, enrollment in a high-deductible health plan, the person's number of fills for all drugs in the year up to the date of the recipient fill (to reflect insurance benefit design over the course of the year), the person's age and private and public HMO enrollment, pharmacy name, pharmacy chain name, quantity, and state. The weight for each variable is based on its relative contribution to the variation in both price and the proportion of the price paid out of pocket.

The imputed price is the product of the recipient acquisition's quantity dispensed and an imputed RUP. The calculation of the imputed RUP depends on the attempt at which a match is found for imputation. If the first or second match is successful, then the imputed RUP is the donor acquisition's RUP. If the third or fourth match is successful, then the imputed RUP is based on the donor's ratio of RUP to NADAC per unit and the recipient's NADAC per unit: $\text{Imputed RUP} = (\text{donor's RUP} \div \text{NADAC per unit}) \times (\text{recipient's NADAC per unit})$.

Similarly, if the fifth or sixth match is successful, then the imputed RUP is the product of the ratio of the donor's RUP with the donor's WACUP and the recipient's WACUP. Lastly, if the seventh or eighth match is successful, then the imputed RUP is the product of the ratio of the donor's RUP with the donor's AWUP and the recipient's AWUP. There are no limits on the number of times a donor can match a recipient.

^g The price is not imputed if the imputed price is less than the reported price and if the lower outlier is a discounted single source brand name drug for an uninsured person.

^h Among the additional cases that used the reported, rather than imputed, price are those with total imputed price less than \$2. Also, if only one pill is reported, then we assume it is one bottle of pills, the quantity of pills is increased, and the reported price is used.

ⁱ The rules for determining whether imputed prices were used for upper outliers with partial payment data are the same as the rules used for upper outliers with complete payment data.

^j Free drugs are neither donors nor recipients in price imputation.

Some corrections for dispensing fees are made for some very low imputed, undiscounted prices that are not partial fills. To better reflect Medicaid's dispensing fees, if the acquisition is brand name, Medicaid is a source of payment but Medicare is not, and the imputed price is less than \$10, then \$10 is added to the imputed price. For brand name acquisitions with payers other than Medicaid, if the imputed price is less than \$1, then \$1 is added to the imputed price. Lastly, for a generic acquisition, if the imputed price is less than \$0.4, then \$1 is added to the imputed price.

Whether any exception is made to basing the imputed price on the imputed RUP, and how the price is distributed by payer, depends on whether the recipient acquisition had any payment information and whether it was identified as a lower outlier or upper outlier.

Lower outliers with complete payment data: There are two types of cases where the imputed prices are not allocated and the original recipient prices are retained (no change) are: (1) if the imputed price is less than the original recipient price and (2) if it was a discounted single source drug for an uninsured person. Otherwise, the increase in price due to imputation is allocated based on the recipient's household-reported type of insurance or usual third-party payer for drugs. If there is one third-party payer, then the increase is assigned to that payer, and if there is more than one third-party payer, then the increase is allocated according to a hierarchy of the third-party payers. If the household did not report any insurance or usual third-party payer, then the increase is assigned to self-payment.

Upper outliers with complete payment data: There are two types of cases where there are no changes: (1) If only one pill is reported, then quantity is increased, and (2) if the price is less than or equal to \$16 or the imputed price is less than \$2, then there is no change; otherwise, the reduction is taken from the third-party payment. If there is no positive third-party payment, the reduction is taken from the out-of-pocket payment.

Lower outliers with partial payment data: For the infrequent cases where the imputed price is less than the original recipient price, the original price is retained, and missing values are set to zero. More commonly, the imputed price less the reported payment amount is allocated among the reported payers, almost always to fill in a missing third-party payment. That is, if out-of-pocket is the only reported payment, then the increase in price is entirely allocated to the missing third-party payment.

Upper outliers with partial payment data: If only one pill is reported, then quantity is increased. If the price is less than or equal to \$16, then there is no change. Otherwise, the reduction in price is taken entirely out of the positive source(s) of payment reported by the pharmacy provider.

Recipients with no payment data or zero reported payments: The donor acquisition supplies both the unit price and the shares paid out of pocket and by third-party payers. The donor's distribution is sometimes modified using the following hierarchy of rules:

- If the recipient's out-of-pocket payment amount was zero, the recipient's imputed out-of-pocket payment amount is almost always set to zero.
- For the perfect matches on NDC and payers, the donor's proportions by payer are applied to the imputed price.
- For acquisitions from federal pharmacies, the out-of-pocket amounts are set using program rules, and the remaining imputed price is allocated to federal programs.
- If the donor and recipient have different payers, then the share paid out of pocket is imputed and the donor's share covered by the third-party payer is allocated to the recipient's third-party payer.
- If the recipient's pharmacy did not report a third-party payer and the household reported no insurance and no usual third-party payer, then generally the out-of-pocket amount is set to the entire imputed price. If the recipient's acquisition appears to be in the Medicare Part D donut hole, then the price is allocated between out-of-pocket and Medicare based on program rules for that year. In particular, the manufacturer's share does not appear in the MEPS data because the manufacturer pays itself.

- If Medicare Part D catastrophic coverage (beyond the donut hole) appears to apply to the recipient’s acquisition, then 95 percent of the price is allocated to Medicare and the rest to out-of-pocket.

Lastly, unusually high and low prices, and cases with gross inconsistencies across fills purchased by a person, are reviewed by a pharmacist.

Results of Preparing the Pharmacy-Reported Data

The information collected in the PC about an acquisition—the date of the transaction, the NDC (or analogous identifying drug information), the quantity, the number of days of the drug that the acquisition supplied, and payments to the pharmacy by source—now also includes the GPI. Imputation has supplied values for any missing NDC or quantity dispensed. Logical edits have determined third-party payers if necessary, and missing or unrealistic prices have been replaced with imputed values. With these enhancements, the PC data are now ready to be linked to the drug names reported in the HC.

Matching Pharmacy Data to Household Data

Overview of Matching

Two general approaches accomplish matching the data reported by pharmacies to the data reported by households. First, for each of a person’s acquisitions in the HC, an attempt is made to find the same or a similar acquisition obtained by that person in the PC. If this approach fails, the second approach imputes pharmacy data from some other person.

Both matching procedures are conducted at the level of the person-round-drug rather than at the acquisition level. In the HC data, a “drug” is a unique drug name reported for the person in the round. In the PC data, a “drug” is a set of acquisitions of pharmaceutically equivalent drug products identical in the active ingredients, dosage form, and strength (14-digit GPI), whether brand name or generic, by the person in the round. That is, the acquisitions in the PC data are aggregated to the drug level to mirror the structure of the HC data. In 2021, 144,783 person-round-drugs in the HC represented 335,291 acquisitions; in the PC, 257,596 acquisitions aggregated to 140,878 person-round-drugs.

After matching a PC person-round-drug to an HC person-round-drug to create a pair, the PC acquisitions within the aggregated set are unrolled, the HC drug is fanned out into a set of acquisitions, and each HC acquisition is paired with a PC acquisition in the drug set. If the number of acquisitions differs between the HC and PC, then the number of acquisitions is determined by the HC and some randomization is used to allocate the PC acquisitions to the HC acquisitions.

Details of Matching

The first approach: Within-person matching

The first approach to matching—within the person—entails four attempts. First, the procedure seeks PC data exactly matching on person, round, and pharmaceutically equivalent drug (14-digit GPI) for each household-reported drug name. Once a match is made, the PC donor person-round-drug is removed from the donor pool and not matched with any other household-reported drugs for the individual (i.e., the matches are made without replacement). In the second attempt, the residual household- and pharmacy-reported drugs must match exactly on person and round. Weighted match variables, which are not required to match exactly, are the drug group, therapeutic class, active ingredients, and medication name. An HC-PC pair is accepted only if exactly matching on drug group or the reported medication name. In the second attempt, matches are also made without replacement. For the HC residuals of the second attempt, the requirement of matching within the round is removed, and all the person’s pharmacy-reported drugs are candidate donors. This third attempt requires exact matches on person, and active ingredients, dosage form, strength, and medication name are weighted match variables. For the HC residuals of the third attempt, the requirement of exact match on active ingredients is removed (so only exact match on person is required), and all the person’s pharmacy-reported drugs are candidate donors. In this fourth attempt, active ingredients, dosage form, drug group, drug class, therapeutic group, round, and medication name are weighted match variables. A match is

accepted only if it meets quality standards established for the match attempt. In the third and fourth attempts, matches are made with replacement.

One variable in the weighted match is medication name, and matching medication names requires specialized software. We use a method similar to Soundex to match medication name.^k Matching PC and HC drugs within person leaves some HC drugs unlinked to PC data. Some sample persons lack pharmacy data altogether, and some HC drug names remain unmatched for other reasons (for example, the household respondent failed to mention one of multiple pharmacies used). A second approach is needed for this reason.

The second approach: Imputation from a different person

The second approach to matching—imputation from a different person—entails a total of 11 match attempts. It draws on a donor pool of all pharmacy-reported drugs regardless of person (excluding specific free drugs, which are reconciled later). PC drugs with imputed NDCs are part of the donor pool as well. All these attempts match with replacement.

The first attempt requires an exact match on active ingredient, dosage form, and strength (14-digit GPI). Weighted match variables used are the medication name; number of months per acquisition in the round; insurance status and potential payment sources for the person; whether the person is enrolled in any high-deductible health plan; access to Medicare Part D low-income subsidy; name of the pharmacy; whether the HC respondent reported that the person used any mail-order pharmacies; the cumulative number of HC-reported acquisitions of all drugs in the prior and current rounds of the calendar year (a proxy for healthcare utilization); and the person's age, sex, medical conditions, geographic region and division, urbanicity, employment status, and self-reported health status.

The second attempt requires an exact match on active ingredient and dosage form (12-digit GPI), and the third requires an exact match on active ingredient (8-digit GPI). The match variables used in these two attempts are the same as the ones used in the first match attempt, except that in the third attempt, the 10-digit GPI is used as an additional match variable.

The fourth attempt requires an exact match on drug class (4-digit GPI) and whether the drug is single source only (the 10-digit GPI only has single source drugs in the MDDDB). The fifth attempt requires an exact match on drug class (4-digit GPI) only. For the fourth and the fifth attempts, all the weighted match variables for the first attempt are used, in addition to the strength of the drug. The sixth attempt requires an exact match on drug group (2-digit GPI) only, and all match variables from the fourth and fifth attempts are used, in addition to the drug class.

The seventh and the eighth attempts both require exact matches on therapeutic group, with the key difference between the two attempts being that in the eighth attempt, medication name is not used as a match variable (zero weight) because the drug name reported by the household is not specific. The ninth attempt matches nonspecific household-reported steroids (e.g., drug name is “steroid”) to steroids in the pharmacy data, without matching on drug name.

The last two attempts do not require any exact matches. The key difference between the tenth and eleventh match attempts is in the weight on medication name.

After matching person-round-drugs, the HC and PC drug records are expanded into acquisitions, as mentioned above. Each drug name reported in the HC interview is fanned out to the number of acquisitions that the household reported; each PC drug, a set of aggregated PC acquisitions, is unrolled back into distinct acquisitions as originally reported in the PC. Then each HC acquisition is paired with a PC acquisition within the drug-to-drug matched set. Because the household reports only the number of acquisitions, they have no natural order. Therefore, the acquisitions take the date order of the PC acquisitions; the order is preserved only in the unique record identifiers. When the household and the pharmacy both report the same number of acquisitions, then the pairing is one to one

^k This matching algorithm is also used in matching the PC to the MDDDB to impute NDC (when missing).

by the date order in the PC. Otherwise, if the number of acquisitions differs between the HC and PC, then the PC acquisitions are paired with the HC acquisitions in a randomization process that follows the date order of the PC and the unique identifiers of the acquisitions in the HC. Note that when the drug match is imputed, the order of acquisitions represented by the record identifier may not be analytically useful because the date order is from a different person or round.

Matching diabetic supplies and equipment is similar. Pharmacy-reported acquisitions of diabetic supplies and equipment are mostly aggregated into one “drug” record for each person-round, so that the PC data structure parallels the HC. Aggregating in this way before matching generates a more accurate representation of the variety of diabetic supplies purchased. Unfolding PC diabetic supplies and equipment acquired in the same day added 1,432 records to the file, so the total number of records increased from 333,859 to 335,291.

Table 8 summarizes the results of matching pharmacy data to household data in 2021. Almost three-fourths (72.9 percent) of drugs were acquired by those sample members who had any pharmacy data. Matching was highly successful for these drugs: 83.6 percent (line 5) of their acquisitions were from their own pharmacies. The imputed pharmacy matches—acquisitions not matched within the person—made good use of the information given by household respondents: 89.0 percent (line 11) were matches to the active ingredient of the drug reported by the household.

In the 2021 PC donor database, 39.1 percent of acquisitions remained unmatched to the HC. A validation study found that household respondents reported as many acquisitions as were found in claims data, but households reported fewer drugs and more acquisitions per drug. The drugs acquired but not reported by the household tended to be for short-term use. They entailed fewer acquisitions and included many anti-infectives, topical agents, and pain medications.⁸

Editing Matched Data

This section describes edits that apply to the matched HC-PC data, including supplemental data merged into the file to enhance analytic utility. These four edits apply to fills with drug details and payment information imputed from another person:

- **Imputed fills with too many days supplied.** This edit corrects for imputing 90-day fills to some people with more frequent refills.
- **Free antibiotics, anti-diabetics, anti-hypertensives, anti-asthmatics, and prenatal vitamins.** This edit allows a price of zero for certain free acquisitions, such as antibiotics, from chain pharmacies offering free drug programs, but in a manner that preserves the anonymity of the chain.
- **Prices paid by the Civilian Health and Medical Program of the Department of Veterans Affairs (CHAMPVA).** This edit corrects prices and payment amounts by source for some individuals with only CHAMPVA.
- **Resolving payer inconsistencies arising from imputation.** This edit reconciles differences in payers related to having matched across persons.

Additional editing applies to some fills:

- **Federal pharmacy prices.** This edit improves the accuracy of prices of acquisitions from federal pharmacies.
- **Editing year in crossover rounds.** This edit allocates to each year purchases reported in the interviews that cover parts of 2 calendar years.
- **Medicare Part D and private insurance.** These edits change the payer from private insurance to Medicare Part D and vice versa for certain people’s acquisitions.

Imputed fills with too many days supplied

When imputing PC data from a different person or different round, mismatch in days supplied per fill can occur. For

example, for some recipients, the donors may have 60 or more days supplied, but the recipients may have more frequent fills, suggesting fewer days supplied. This edit corrects fill details for some of these cases. In particular, if the days supplied are twice or thrice the days in the round and there are more than 90 extra days supplied, then the quantities, days supplied, prices, and third-party payer amounts are reduced for consistency with the higher frequencies of fills.

Free antibiotics, anti-diabetics, and prenatal vitamins

Under certain conditions, payment data are edited for the free antibiotics, anti-diabetics, and prenatal vitamins that pharmacies report. Among the household-reported acquisitions matched to a different person's pharmacy data, all the price and payment variables are set to zero for selected free antibiotics, anti-diabetics, anti-hypertensives, anti-asthmatics, and vitamins. This editing rule applies only if the household respondent reported that the person used only one pharmacy; that pharmacy chain had a free antibiotic, anti-diabetic, anti-hypertensive, anti-asthmatic, or prenatal vitamin program; and the antibiotic, anti-diabetic, anti-hypertensive, anti-asthmatic or prenatal vitamin was included in the chain's program. In addition to maintaining the pharmacy chain's confidentiality for prices of other purchases, the person must have resided in a state with two or more pharmacy chains with free programs. In the 2021 data, 26 acquisitions were edited to a price of zero.

Prices paid by CHAMPVA

If a person-round had CHAMPVA insurance; did not have Medicare, Medicaid, or private insurance; did not obtain medications only at a Department of Defense or an Indian Health Service pharmacy; and did not have a positive payment amount from CHAMPVA, then the price and third-party payments are set according to the VA pricing rules for the year.

Federal pharmacy prices

When the household reported that the sample member used federal pharmacies, but the payment data were imputed from nonfederal pharmacies, a federal price list is used to determine prices for acquisitions. For each NDC, the lowest price on the list is used. These prices reflect the government's cost of acquiring medicines, so an estimate of dispensing costs is added.

Federal prices are assigned to acquisitions in three situations. These situations are defined by information from both the HC and the PC:

- The PC data are from a federal pharmacy that did not report payment information and the HC and PC acquisitions match on person and round.
- The household respondent reported the person used only federal pharmacies in that round, and the PC data are from a different person.
- The household respondent reported the person used at least one federal pharmacy in that round and the PC data, which are from a different person, are from a federal pharmacy that did not report payment information.

In these three situations, the federal price is merged into the matched HC-PC acquisition by the pharmacy-reported NDC. Sometimes, however, the NDC from the PC does not exactly match a drug on the federal price list—because the NDC is imputed, or because there is no negotiated price with the manufacturer of the drug. In these cases, instead of the NDC, the match is by active ingredient, dosage form, and strength (14-digit GPI) with drug name as a weighted match variable. The pharmacy-reported NDC, whether reported or imputed, is replaced with the one selected from the federal price list. The remaining cases retain the price from the PC (whether reported or imputed) because the federal price schedule does not include every drug.

For the acquisitions with prices taken from the federal price list, the out-of-pocket and federal payments are set using federal program rules.

In the 2021 data, 4,109 acquisitions (1.2 percent) were assigned federal prices.

Editing year in crossover rounds

In the crossover round interviews (the third interview for all panels; the fifth and seventh rounds for panels that collected 4 years of data), in which the reference period spans the later part of one year and the early part of the next, for each drug name, the household respondent is asked to report both the number of times the drug was obtained since the last interview and the number of those times the drug was obtained in the previous year. When this information is missing for a drug, the number of acquisitions is allocated to each year. The allocation is based on the date the person started taking the medication and, for drugs with PC data matched exactly to the HC on person and round, the number of acquisitions reported by pharmacies in each year. Otherwise, acquisitions are distributed in proportion to the duration of the crossover round in each year. After this allocation, there were 303,394 acquisitions in 2021; this is the number of acquisitions (records) in the public use file.

Medicare Part D and private insurance

In two types of rare cases, payments are switched between Medicare and private insurance. For Medicare beneficiaries, some private insurance payments are assumed to be Medicare Part D. This edit applies only to acquisitions for which the pharmacy data are matched from the beneficiary's pharmacies, not imputed from another person. For a beneficiary who has an acquisition with a private payment and, according to the HC, has Part D coverage, the private payment is assumed to be from a Medicare Part D plan. In addition, for an elderly person who has an acquisition with a private payment and, according to the HC, lacks private drug coverage, the private payment is assumed to be from a Medicare Part D plan. This edit applied to 773 acquisitions in 2021.

Conversely, some Medicare payments are assumed to be private payments. These edits apply only to acquisitions for which the pharmacy data were imputed from another person. For acquisitions that have an imputed Medicare payment but no private insurance payment, and the person is not elderly, not disabled, and has private coverage but no Medicare Part D, then the Medicare payment is reassigned to private payment. This edit applied to 603 acquisitions in 2021.

Resolving payer inconsistencies arising from imputation

As addressed above, matching pharmacy data to drug names reported in the HC is based primarily on the drug group, therapeutic class, active ingredients, dosage form, and strength. Additional criteria include insurance status and potential third-party payers (although, due to small cell sizes, these variables are not required to match exactly in the second approach to matching, imputation from another person). Despite high-quality drug-to-drug matching, inconsistency in payment patterns can result. The donor acquisition and recipient acquisition may have different payers. For example, the pharmacy (donor)-reported third-party payments can be from an insurer that, according to the HC, did not provide coverage to the sample person (recipient). This section describes the methods used to identify and resolve these inconsistencies.

Acquisitions with inconsistencies are defined as those for which the HC-reported insurance coverage differs between the donor drug record and the recipient drug record. Both the donor and recipient records are classified using a hierarchy of combinations of insurance (see Table 9). The classification reflects common situations that relate to the types of third-party payers and amount paid out of pocket. The most common situation is having only one type of insurance for the entire round, but the classification accounts for other situations as well. When the donor and recipient have different insurance classifications, they are considered inconsistent.

Most situations of inconsistency between a donor's sources of payment and the recipient's insurance require editing or additional imputation for resolution. However, resolving inconsistency between insurance and payers is unnecessary in several situations. First, acquisitions where pharmacy data have been matched within person are always seen as consistent. Second, acquisitions with prices taken from the federal price list are defined as consistent, because third-party payments have already been edited. Third, free acquisitions are consistent, because there are no payments. Fourth, acquisitions are consistent for the small number of drugs covered by Medicare Part B if both the

recipient and the donor are Medicare beneficiaries and one of the following is true: Both have Medicaid as a source of payment; both have private coverage or any prescription drug coverage; or both do not have Medicaid, private, or any prescription drug coverage.

Among the 25,279 fills with inconsistencies requiring editing or imputation, two rare situations rely on editing. First, payments for Medicare Part B drugs for Medicare beneficiaries are allocated to Medicare (78 acquisitions). Second, for acquisitions that appear to occur after the person has exited the Medicare Part D donut hole and are therefore in the catastrophic benefit portion, 95 percent of the price is allocated to Medicare and 5 percent to out of pocket (7 acquisitions).

The method for correcting the remaining inconsistencies (25,194 acquisitions in 2021) is to impute the distribution of payments from an acquisition with consistent payers using a hot-deck procedure. The donor pool is composed of acquisitions with consistent payers: those where the PC data were matched within the person to the HC as well as those from the federal price list. Free acquisitions are excluded, because they lack payment data, and fills for persons who obtained all their fills from military or TRICARE pharmacies are excluded, because no individuals with these characteristics will be in the recipient group, as fills obtained by these individuals already have been edited. Fills with payments that have been edited for consistency are in neither the donor group nor the recipient group. A donor is drawn from the donor pool with replacement. Class variables are, in order of importance: insurance classification, patent status, price categories, whether any mail-order pharmacies were used by the person in the round, more detailed price categories, drug group, drug class, active ingredients, dosage form, and strength. Except for insurance status, if there are fewer donors than recipients in a cell, then cells are collapsed until the ratio of donors to recipients is at least 1:1. The imputed distributions of payments are further adjusted for especially rare combinations of insurance for better alignment between the donor's and the recipient's sources of payment.

Payment information from the new donor acquisition is then used to impute third-party amounts and either copayments or out-of-pocket coinsurance payments. Copayments, including values of zero, are always imputed for Medicaid, the VA, and TRICARE. For example, if the donor has a Medicaid payment, then the recipient acquisition's out-of-pocket amount is set equal to that of the donor acquisition's, and the recipient Medicaid payment is set equal to the recipient price less the donor out-of-pocket amount, preserving the total. For the remaining third-party payers, either a copayment or coinsurance is imputed, depending on whether the donor's out-of-pocket amount is a whole number. If the donor acquisition is generic and the out-of-pocket amount is a whole number, or the out-of-pocket amount is a round multiple of \$5, then the copayment is imputed, and the third-party amount is set as the difference between the price and copayment, preserving the total. Otherwise, the donor acquisition is treated as a coinsurance case. The proportion of the donor acquisition's price paid out of pocket and by each third-party payer is calculated. The recipient acquisition's price is distributed in the same proportions to calculate the recipient out-of-pocket amount and payments by each third-party payer.

Additional variables

Two sets of variables are merged into the dataset for release to the public: Multum drug name and therapeutic classes. The source is the Multum Lexicon Plus database of Oracle Health¹ These variables are merged into the file by NDC.

Editing for confidentiality

Before the data are released to the public, automated masking procedures are reviewed by a pharmacist consultant to ensure the confidentiality of the sample members. Drugs are censored if they are associated with very rare conditions, particularly orphan drugs, or estimated to be used by fewer than 400,000 individuals, unless use of the drug does not reveal specific information about the condition treated (for example, cold remedies). In these cases, the drug name is replaced with a more general therapeutic class name and the NDC is set to "missing." Additional masking ensure

¹ The drug name variable supplied by the Multum Lexicon Plus database was added to the public use file since 2013. This drug name is the generic name of the drug most commonly used by prescribing physicians. For the earlier years, this variable has been provided with Addendum Files to MEPS Prescribed Medicines Files for 1996-2013.

pharmacies are not identifiable. Confidentiality protection affected 10 percent of acquisitions in 2021.

Table 1. Summary of editing and imputation rates for key variables, 2021

Dataset and unit (number)	Variable to be imputed	Percentage with editing or imputation
Household Component (HC) data		
Unique drug names reported for a person in a round (144,783)	Number of acquisitions	5.1%
Pharmacy Component (PC) data		
Unique acquisitions of a drug in a year (257,596)	National Drug Code	2.9%
	Quantity	0.5%
	Third-party payer	16.5%
	Price	31.4%
Matched HC-PC data		
Unique drug names reported for a person in a round (144,783)	Drug details	40.3%

Table 2. Method used to impute the number of acquisitions in the Medical Expenditure Panel Survey Household Component, 2021

Match type	Number	Percentage
Exact match to Household Component donor pool on:		
Active ingredients, dosage form, and strength	7,256	93.4%
Active ingredients and dosage form	19	0.2%
Active ingredients	24	0.3%
Drug group	15	0.2%
Nonexact match		
Antibiotics	139	1.8%
Missing Generic Product Identifier	312	4.0%
Total person-round-drugs requiring imputation	7,765	100.0%

Note: A person-round-drug is a unique drug name within a person and round. Matching used Wolters Kluwer's Generic Product Identifier to identify drug names with the same active ingredients.

Table 3. Sample size and participation in the Medical Expenditure Panel Survey Pharmacy Component, 2021

Participation type	Pharmacies		Person-pharmacy pairs	
	Number	Percentage	Number	Percentage
Eligible sample	9,079	100.0%	17,698	100.0%
Response	7,501	82.6%	14,362	81.2%
Refusal	183	2.0%	1,986	11.2%
Other nonresponse	1,395	15.4%	1,350	7.6%

Notes: The sample for the Pharmacy Component is derived from all persons in the Household Component who signed permission forms to contact the pharmacies from which they reported obtaining drugs during the rounds that include 2021. Person-pharmacy pairs uniquely count the number of HC sample members reported in the HC as using each pharmacy. Other nonresponse includes unlocatable pharmacies and patients who had no services on record. Veterans Health Administration (VA) pairs: In 2021, there were 492 Pharmacy VA pairs. Data were collected for 470 pharmacy pairs, resulting in a completion rate of 95.5%.

Table 4. Method used to impute NDC in the Medical Expenditure Panel Survey Pharmacy Component, 2021

Imputation method	Number	Percentage
Acquisitions	257,596	100.0%
Acquisitions with valid National Drug Codes (NDCs)	250,023	97.1%
Acquisitions missing or invalid NDCs	7,573	2.9%
Among acquisitions missing NDCs	7,573	100.0%
Imputed from same person and round in Pharmacy Component	403	5.3%
Imputed from the Master Drug Data Base	6,753	89.2%
Exact match on active ingredients, dosage form, and strength	6,390	84.4%
Exact match on active ingredients	308	4.1%
Exact match on drug group	55	0.7%
Imputed from an acquisition of a person with similar characteristics	417	5.5%

Note: Matching used Wolters Kluwer's Generic Product Identifier to identify drugs with the same active ingredients.

Table 5. Completeness of payment data in the Medical Expenditure Panel Survey Pharmacy Component, 2021

Payment data completeness	Number	Percentage
Total Acquisitions	257,596	100.0%
Complete payment data	144,929	56.3%
Partial payment data	63,297	24.6%
All payment data missing or zero	49,370	19.2%

Table 6a. Lower thresholds for identifying prices needing imputation using ratio of retail unit price to price benchmarks, by payment data completeness, potential for being in Medicare Part D donut hole, flag for being discounted or being over-the-counter (OTC) medication, any third-party payment, and by patent status, 2021

Pharmacy payment data	Any third-party payment, discounted, in donut hole, or OTC medication	Donut hole	Patent status		
			Single source	Originator	Generic
NADAC per unit					
Complete	Yes		0.01	0.01	0.01
Complete	No		0.85	0.01	0.01
Partial		No	0.95	0.95	0.42
Partial		Yes	0.45	0.45	0.42
WACUP (when NADAC is not available)					
Complete	Yes		0.01	0.01	0.01
Complete	No		0.85	0.01	0.01
Partial		No	0.85	0.85	0.12
Partial		Yes	0.4	0.40	0.12
AWUP (when neither NADAC nor WACUP is available)					
Complete	No	No	0.65	0.20	0.03
Complete	Yes	No	0.4	0.0	0.0
Partial	No or Yes	No	0.75	0.7	0.15
Complete	No or Yes	Yes	0.2	0.2	0.03
Partial	No or Yes	Yes	0.3	0.3	0.15

Note: NADAC = national average drug acquisition cost, WACUP = wholesale acquisition cost unit price. AWUP = average wholesale unit price. When using AWUP as a benchmark, information about third-party payers or OTC medication are not used.

Table 6b: Upper thresholds for identifying prices needing imputation using ratio of retail unit price to price benchmarks, by patent status and dosage form, 2021

Measure	Patent status and dosage form		
	Single source liquids	All other brand name drugs	Generics
NADAC per unit	8	4	50
WACUP	4	2	20
AWUP	10	10	10

Note: NADAC = national average drug acquisition cost, WACUP = wholesale acquisition cost unit price. AWUP = average wholesale unit price.

Table 7. Editing categories by patent status in the Medical Expenditure Panel Survey Pharmacy Component, 2011

Type of payment data	Total	Patent status		
		Brand name		Generic
		Single source	Originator	
Total acquisitions	257,596	35,728	4,219	217,649
Complete payment data	144,929	19,446	2,710	122,773
<i>No editing</i>	141,125	17,403	2,630	121,092
<i>Lower outlier</i>	1,046	244	30	772
No change	181	39	2	140
No third-party payment (impute RUP)	615	154	25	436
Positive third-party payment (impute RUP)	250	51	3	196
<i>Upper outlier</i>	2,758	1,799	50	909
No change	1,854	1,588	30	236
No third-party payment (impute RUP)	70	33	3	34
Positive third-party payment (impute RUP)	834	178	17	639
Partial payment data	63,297	8,973	731	53,593
<i>Missing values set to zero</i>	28,649	302	122	28,225
<i>Lower outlier</i>	34,579	8,654	607	25,318
Missing values set to zero	3,891	1,562	44	2,285
Impute RUP	30,688	7,092	563	23,033
<i>Upper outlier</i>	69	17	2	50
Missing values set to zero	26	11	1	14
Impute RUP	43	6	1	36
Free antibiotics, prenatal vitamins, antidiabetics, and glucometers from some pharmacies (no change)	961	312	10	639
Missing all payment data (impute RUP)	48,409	6,997	768	40,644

Note: RUP = retail unit price.

Table 8. Matching and imputation of pharmacy-reported drugs and acquisitions to household-reported drug names in the Medical Expenditure Panel Survey, 2021

#	Description	Drugs		Acquisitions	
		Number	Percentage	Number	Percentage
1	Household-reported totals	144,783	100.0%	335,291	100.0%
2	Person had any pharmacy data	105,618	72.9%	244,429	72.9%
3	Person had no pharmacy data	39,165	27.1%	90,862	27.1%
Pharmacy data					
4	Person had any pharmacy data (line 2)	105,618	100.0%	244,429	100.0%
5	Total matched within person (sum of lines 6, 7)	86,403	81.8%	204,334	83.6%
6	Matched within person-round	74,987	71.0%	183,078	74.9%
7	Matched within person	11,416	10.8%	21,256	8.7%
8	Not matched within person	19,215	18.2%	40,095	16.4%
Unmatched within person					
9	Total not matched within person (sum of lines 3, 8)	58,380	100.0%	130,957	100.0%
10	Imputation with exact match on (rows 11-21)	57,119	97.8%	128,347	98.0%
11	Active ingredient (sum of lines 12, 13, 14)	51,359	88.0%	116,587	89.0%
12	Active ingredient, dosage form, strength	34,661	59.4%	77,559	59.2%
13	Active ingredient, dosage form	6,170	10.6%	14,746	11.3%
14	Active ingredient	10,528	18.0%	24,282	18.5%
15	Drug class and only single source	338	0.6%	841	0.6%
16	Drug class	1,461	2.5%	3,382	2.6%
17	Drug group	2,275	3.9%	4,569	3.5%
18	Therapeutic group (sum of rows 19, 20)	1,605	2.7%	2,857	2.2%
19	Therapeutic group/medicine name	818	1.4%	1,656	1.3%
20	Therapeutic group	787	1.3%	1,201	0.9%
21	Steroid	81	0.1%	111	0.1%
22	Imputation without exact GPI/TG (sum of lines 23, 24)	1,261	2.2%	2,610	2.0%
23	Medicine name	632	1.1%	1,283	1.0%
24	Weighted match variables	629	1.1%	1,327	1.0%

Notes: In the Household Component, a “drug” is a unique drug name reported for the person in the round. In the Pharmacy Component, a

“drug” is a set of acquisitions of pharmaceutically equivalent drug products identical in the active ingredients, dosage form and strength, whether brand name or generic, by the person in the round. Matching used Wolters Kluwer’s Generic Product Identifier to identify drugs with the same active ingredients. Matches are weighted and take into account drug name, types of insurance, health status and chronic conditions, census division and urbanicity, and demographics.

Table 9. Major categories of hierarchical insurance classification for payer consistency editing

Medicare Part D
And Medicaid
Other low-income subsidy (no Part D premium)
And Veterans Health Administration (VA)
Likely in the donut hole based on total expenditures on drugs covered by Part D in prior rounds in the calendar year
And private drug coverage (no Medicaid)
Other
Medicaid (no Medicare Part D)
The whole round, no private insurance
Part of the round, no private insurance
And private insurance
Private drug coverage (no Medicare Part D or Medicaid)
Part of the round, no TRICARE and no VA
The whole round, no TRICARE and no VA
Private insurance (no private drug coverage, Medicare Part D, or Medicaid)
And VA (No TRICARE)
And TRICARE
TRICARE (no private insurance, Medicare Part D, or Medicaid)
No VA
And VA
VA (no private insurance, Medicare Part D, or Medicaid)
No Medicare
And Medicare
Other public
Private insurance only and not elderly
Private insurance only and elderly
Medicare only
Uninsured

References

- ¹ Hill S, Stagnitti M, Roemer M. Outpatient Prescription Drugs: Data Collection and Editing in the 2011 Medical Expenditure Panel Survey. MEPS Methodology Report #29. Rockville, MD: Agency for Healthcare Research and Quality; March 2014. https://meps.ahrq.gov/data_files/publications/mr29/mr29.shtml Accessed January 19, 2024 .
- ² Ding Y, Hill SC. Evaluating Alternative Benchmarks to Improve Identification of Outlier Drug Prices for Medical Expenditure Panel Survey Prescribed Medicines Data Editing. Working Paper #22001. Rockville, MD: Agency for Healthcare Research and Quality; September 2022. https://meps.ahrq.gov/data_files/publications/workingpapers/wp_22001.pdf. Accessed November 10, 2023.
- ³ Cohen J. Design and Methods of the Medical Expenditure Panel Survey Household Component. MEPS Methodology Report No. 1. AHCPR Pub. No. 97-0026. Rockville, MD: Agency for Health Care Policy and Research; 1997. http://www.meps.ahrq.gov/mepsweb/data_files/publications/mr1/mr1.shtml. Accessed November 10, 2023.
- ⁴ Cohen S. Sample Design of the 1996 Medical Expenditure Panel Survey Household Component. MEPS Methodology Report No. 2. AHCPR Pub. No. 97-0027. Rockville, MD: Agency for Health Care Policy and Research; 1997. http://www.meps.ahrq.gov/mepsweb/data_files/publications/mr2/mr2.shtml. Accessed November 10, 2023.
- ⁵ Cohen S. Design strategies and innovations in the Medical Expenditure Panel Survey. Medical Care, 2003 July;41(7):Supplement III-5–III-12.
- ⁶ Master Drug Data Base (MDDDB[®]), Version 2.5. Documentation Manual. Indianapolis, IN: Wolters Kluwer Health, Inc., 2023.
- ⁷ Web Lexicon Plus[™]. [Documentation.] Denver, CO: Cerner Multum, Inc., 2016.
- ⁸ Hill, S.C., Zuvekas, S.H., Zodet, M.W. "Implications of the Accuracy of MEPS prescription drug data for health services research." Inquiry. 2011; 48(3):242-259.

Suggested Citation

Abdus S, Hill SC, Ahrnsbrak R. Outpatient Prescription Drugs: Data Collection and Editing in the 2021 Medical Expenditure Panel Survey. Methodology Report #37. Rockville, MD: Agency for Healthcare Research and Quality, Rockville, MD; January 2024. https://www.meps.ahrq.gov/mepsweb/data_files/publications/mr37/mr37.pdf