MIS-REPORTING OF PRESCRIPTION DRUG UTILIZATION AND EXPENDITURES IN THE MEDICARE CURRENT BENEFICIARY SURVEY

By John A. Poisal, MBA

Office of Research, Development, and Information

Centers for Medicare and Medicaid Services

Table of Contents

Abstract

One shortcoming of using household surveys to estimate healthcare utilization and expenditures is that respondents tend to misreport their usage of medical services.[1] Using data from both the 1999 Medicare Current Beneficiary Survey (MCBS) and drug utilization data supplied by the survey respondents' pharmacies, we examine the level of mis-reporting of drug use and expenditures in the MCBS.[2] Findings suggest that prescription drug expenditures are under-reported by 17% in aggregate, and that the number of prescriptions used is under-reported by 17.7%. The data show that there are various demographic factors that are predictive of a beneficiary's likelihood to either

[1] Groves RM, Survey Errors and Survey Costs, New York: John Wiley and Sons, 1989

over-report or under-report their medications, as well as the extent to which they misreport their drug use and spending.

Background

Interest in prescription drug expenditures, as they relate to high out-of-pocket costs and possible drug coverage expansion, remains high. Prescription drug spending rose almost 16% in 2001 and is projected to rise an average of 11.1 percent per year between 2002 and 2012.[3] Senior citizens are particularly vulnerable to these rising costs partially due to higher incidence rates of chronic disease, many of which can be effectively treated with prescription medication. Even seniors with employer sponsored insurance (ESI) drug coverage, thought to be the most reliable source of coverage, are becoming increasingly subject to high prescription drug expenses as the number of large employers (500+ employees) offering ESI coverage to Medicare-eligible retirees declined from 57 percent in 1987 to 23 percent in 2001.[4]

Adding a prescription drug benefit to Medicare has been the focus of debate on Capitol Hill and elsewhere for the last several years. Recent findings suggest that Medicare beneficiaries without drug coverage fill fewer prescriptions than their covered counterparts, after controlling for factors like age, supplementary insurance status, and income.[5] Moreover, many beneficiaries skip dosages or avoid filling prescriptions

---

[2] Not all people under-report their drug utilization and spending, however, in the aggregate, MCBS respondents report fewer medication purchases and lower expenses than they actually make and incur. As a result, this analysis often refers to the mis-reporting of drugs as under-reporting.
[3] Heffler, Steve. “Health Spending Projections for 2002-2012.” 2003. Health Affairs 7 Feb 2003 <<http://www.healthaffairs.org/WebExclusives/Heffler_Web_Excl_020703.htm>>
[4] Mercer, *US Mercer/Foster Higgins survey of employer-sponsored health plans - key findings*, . 2002, Mercer Human Resource Consulting LLC and Mercer Investment Consulting Inc.: Washington, DC; Bos, B. *15th Annual Mercer/Foster Higgins National Survey of Employer-Sponsored Health Plans*. 2001. Chicago, IL; DeWitt, D.L., *Emerging facets of retiree benefits*. Business & Health. 1988. 5(10): p. 8-12.
[5] Poisal JA, Murray L, “Growing differences between Medicare beneficiaries with and without drug coverage,” Health Affairs, Jan-Feb 2001, 74-85

entirely due to prohibitively high drug costs.[6] These findings emphasize the importance of prescription drug coverage within the Medicare population.

In response to the legislative proposals to add a drug benefit to Medicare, CMS's Office of the Actuary and the Congressional Budget Office are regularly asked to make cost projections for such proposals, many of which rely on survey prescription drug cost and utilization data. When using survey data for this purpose, several assumptions must be made to accurately project these costs including adjustments for survey mis-reporting, institutional drug usage, and the degree to which demand would increase with the passing of a new benefit.

This paper reports on an attempt to quantify the extent to which prescription drug use and expenditures are mis-reported in one such survey—The Medicare Current Beneficiary Survey (MCBS). The MCBS is an ongoing household panel survey of about 13,000 Medicare beneficiaries, funded by the Centers for Medicare and Medicaid Services (CMS).[7] Annually, CMS produces two file series: The Access to Care series and the Cost and Use series. The Cost and Use series contains data on health care utilization and expenditures for beneficiaries "ever-enrolled" in Medicare, including persons who enrolled in the program or died during the year. This series also includes data on non-Medicare covered services such as prescription drugs, as well as data on Medicare covered services.

Household surveys of health and health expenditures, such as the MCBS, are subject to non-response and mis-reporting of medical events.[8] As a general rule, health events that are farther removed in time and those that are less prominent are less likely to be recalled at the time of interview.[9] Prescription drug purchases are no exception. During each

---

[6] Steinman M, Sands L, Covinsky K, "Self-restriction of medications due to cost in seniors without prescription coverage," Journal of General Internal Medicine, Vol. 16, No. 12, Dec. 2001, 793-799

[7] Adler G, "A Profile of the Medicare Current Beneficiary Survey," Health Care Financing Review, Summer 1994: 153-163

[8] Chulis GS, Eppig FJ, "Matching MCBS and Medicare Data: The Best of Both Worlds," Health Care Financing Review, Vol. 18, No. 3, 211-229

[9] Cohen SB, Burt VL, "Data Collection Frequency Effect in the National Medical Care Expenditure Survey," Journal of Economic and Social Measurement, 1985, Vol. 13: 125-151

interview, respondents are asked about all of their medication use since their last interview. Using The MCBS takes several steps in an attempt to minimize recall error by beneficiaries. For example, respondents are asked to retain and bring to their interview any prescription bottles, packages, or receipts associated with their medication use. They are also encouraged to make notes on calendars provided by the survey to record all of their health care events. Finally, utilizing CAPI (computer-assisted-personal-interview), MCBS interviewers are furnished with a list of all prescription drugs reported in previous interviews so they can ask whether the respondent has taken any of those drugs during most recent reporting period.

However, to date there have been no efforts to assess what mis-reporting occurs in the wake of these efforts.

Our work provides an answer to that question via a multi-step process. First, we collected and compared data from a survey of MCBS beneficiaries and their pharmacies. We then determined the mis-reporting rates for MCBS sample persons for whom we had complete survey and pharmacy data. Finally, we generalized our results to the entire non-institutionalized MCBS population through a series of micro-simulation models. This effort culminated in an estimate of the direction and magnitude of reporting errors as well as the identification of the social, economic, and demographic correlates of those errors.

Data

Collecting pharmacy data

To test the extent of misreported prescription drug use and spending in the MCBS, a pharmacy follow-back study was designed, and conducted in the first four months of 2000, Four types of MCBS respondents were omitted from the study. Respondents who were institutionalized for all of calendar year 1999 were not asked to participate. Similarly, persons who lived in the community during 1999 but were institutionalized at the time of their spring interview were excluded.[10] Respondents for whom a proxy answered and beneficiaries who were not enrolled in Medicare for all twelve months of 1999 (including deaths) were also excluded.[11]

The remaining survey participants (n=9,384) were asked if they would request patient profiles from all the pharmacies where they obtained their drugs in 1999.[12] Sample persons who had not reported any medication use in 1999 were still asked to participate in the study. In these cases, the beneficiaries were asked to identify the pharmacies that they would normally use to fill a prescription.

Sample persons who agreed to participate were asked to supply the names and addresses of every pharmacy they used during 1999. As a means to help beneficiaries recall their pharmacies, interviewers suggested the use of medicine labels, receipts, phone books, and pharmacy directories. Beneficiaries who reported no prescription drug use during 1999 were asked to supply the names of pharmacies they normally used. Each respondent was asked to sign a pre-printed letter requesting a profile of their 1999 drug utilization from each pharmacy on their list. The letters contained return envelopes addressed to Westat, Inc, the contractor that administers the MCBS for the Centers for Medicare and Medicaid Services.

Of those asked to participate in the pharmacy follow-back study (Table 1), more than half were "complete responses" (meaning not only did they participate, but all of their reported pharmacies submitted prescription profiles on their behalf). A small percentage (6%) of those asked refused to participate, and 4% reported no pharmacies. The 8,406

---

[10] Spring (Round 26) interviews are those interviews conducted between the months of January and April of 2000.

[11] There are times when a sample person is unable to participate in the MCBS interview. Where possible, someone familiar with the beneficiary's health care utilization and expenditures serves as a proxy and answers on their behalf.

respondents who supplied pharmacy names reported 11,102 pharmacies. Westat, Inc. received about three-quarters of the requested profiles (8,126), which were entered into machine-readable format using a computer-assisted data entry system (CADE). About one-fourth of the beneficiaries had unusable data (missing or invalid dates) turned in from one or more of their pharmacies. Only respondents for whom all pharmacies returned usable profiles were examined in this analysis. Thus, the effective response rate was 57 percent (Table 1).

Table 1. Follow-back study status of sample persons in 1999 Cost and Use file.

	Sample Persons	% of All	% of Study participants or non-participants
Total Sample, 1999 Cost & Use	13,106	100%	
Excluded from the follow back:	3,722	28%	100%
A. No 1999 event level drug data collected:			
1. In facility for all of 1999	946		25%
2. New enrollee in 1998 or 1999	638		17%
B. Didn't receive Round 26 interview			
1. Proxy interviews	1,162		31%
2. Spring interview was facility interview	227		6%
3. Deaths and refusals	749		20%
Asked to participate in follow-back study	9,384	72%	100%
A. Refused	570		6%
B. No pharmacies reported by beneficiary	408		4%
C. Reported one or more pharmacies			
1. Partial complete (at least one, but not all pharmacies responded)	813		9%
2. Pharmacy non-response or unusable data	2,291		24%
3. All pharmacies reported usable data	5,302		57%

Including the proper medications from the beneficiary and the pharmacy

[12] The total number of respondents in the 1999 MCBS Cost and Use file is 13,106; not all were selected to participate in the study.

A number of editing steps were necessary prior to analysis of the data. From the MCBS, all beneficiary-reported drug names were standardized, correcting any misspelled words as well as reformatting drug names. Over-the-counter medications that were reported by the respondent were dropped.[13]

In preparing the pharmacy profile data, profiles were excluded if they contained prescription drug events with either a missing month or a missing day. Just over 1% of the profiles received were rejected for unusable or missing dates resulting in 84 sample persons (found within the 2,291 persons categorized as, "Pharmacy non-response or unusable data") being dropped from the analysis. As with the beneficiary-reported data, all drug names were standardized and any over-the-counter medications were deleted. All pharmacy-reported data for a given respondent were then concatenated into a single file.

The next step was to ensure that beneficiary-reported drugs and pharmacy-reported drugs were from exactly the same time periods. Unlike the pharmacy profile data, beneficiary-reported data do not have recorded dates of purchase: In the standard MCBS interview, respondents are not asked for this exact date because such a practice would significantly increase the respondent's recall burden, particularly when a medication is refilled several times. Although the survey does not capture dates of drug purchases, it does establish a recall reference period with specific beginning and ending dates. Drug purchases for calendar year 1999 were recorded in four rounds of interviews, numbered 23, 24, 25, and 26. Round 23 took place between the months of January and April, 1999. Because the reference period for any interview is the previous four months, drug purchases recalled in round 23 could have occurred during the end of 1998 or the beginning of 1999. Likewise, the round 26 interview took place between January and April of 2000, meaning recalled drugs could have been purchased in either late 1999 or early 2000. All reported drugs for rounds 24 and 25 (June-December) were purchased in 1999, therefore, the survey data analyzed was limited to those rounds.

[13] MCBS interviewers are instructed to not collect over-the-counter medications.

Including the proper drugs from the pharmacy reports involved a simple process of date comparisons. For each person, all beneficiary-reported drugs collected in rounds 24 and 25 were included and all pharmacy-reported drugs that fell between the beginning date of the round 24 reference period and the ending date of the round 25 interview were included. The results were a total of 101,144 pharmacy-reported drug events and 96,878 survey-reported drug events.

Matching beneficiary-reported medications with pharmacy-reported medications

An initial attempt was made to match beneficiary-reported drugs to pharmacy-reported drugs electronically. For each event in the survey-reported file, a variable, (MATCH_KEY) was created that contained the sample person's personal identification code, the drug name, and a sequence number: there was one record per beneficiary, per drug, per purchase (including refills). For example, for the fictitious beneficiary whose BASEID was 00001234, the records would read in the following way:

BASEID	DRUGNAME	SEQUENCE	MATCH_KEY
00001234	AMOXIL	001	00001234AMOXIL 001
00001234	CIPRO	001	00001234CIPRO 001
00001234	CIPRO	002	00001234CIPRO 002
00001234	FUROSEMIDE	001	00001234FUROSEMIDE001
00001234	FUROSEMIDE	002	00001234FUROSEMIDE002
00001234	FUROSEMIDE	003	00001234FUROSEMIDE003
00001234	FUROSEMIDE	004	00001234FUROSEMIDE004

The same process was carried out on the pharmacy-report file, and the records from the two files were matched on the variable MATCH_KEY. The automated merge produced 64,273 matches, 36,871 events that appeared only in the pharmacy file, and 32,605 events that appeared only in the survey file.

Examination of the 'pharmacy-only' and 'survey-only' records revealed many missed matches. There were many events in which a generic name was reported by one party while the brand name was reported by the other. There were other events in which the drug name was converted into different standardized names. For instance, if a beneficiary reported the name, 'Cardizem', and the pharmacy reported the drug, 'Cardizem SR', then the prescriptions would fail to match during the electronic merge.

This manual review of the electronic match improved the agreement between pharmacy-reported and survey-reported events in the aggregate. The matched figure increased by 9,246 to 73,519. The pharmacy-only figure fell to 27,625 and the survey-only figure dropped to 23,359.

Unmatched survey-reported drugs were further classified into one of two categories: survey-over-reports or omitted-pharmacy-under-reports. A prescription was assigned survey-over-report status if there were any mentions of that drug in the pharmacy file. We assumed that survey-over-reports occurred when the beneficiary "telescoped" a refill, that is, they reported a refill that occurred in an earlier round, or that never occurred at all. A drug was assigned omitted-pharmacy-under-report status if that drug name did not appear in the pharmacy data. We assumed that these events occurred because the beneficiary failed to report all of their pharmacies and that the prescription was filled in one of these 'omitted' pharmacies. It is possible that a fraction of the drugs categorized as survey-over-reports were, in fact, purchases made at omitted pharmacies.

Methods

Models of mis-reporting of prescription drug utilization (Determining the number of prescriptions that should have been reported)

We explored three mis-reporting models. For each scenario, the following definitions apply:

P=Sum of all prescriptions reported by beneficiary's pharmacies[14]

[14] We assumed that all pharmacy-reported drugs were reported accurately.

M=Number of matched prescriptions

O1=Number of non-matched survey-only prescriptions that were deemed a result of survey over-reports

O2=Number of non-matched survey-only prescriptions that were deemed a result of omitted-pharmacy-under-reports

S=All survey reported prescriptions, or M+O1+O2

R=Net adjusted under-reporting rate

The models vary in their assumptions regarding the source and nature of survey-only events. In the first mis-reporting rate model, all survey-reported drugs are divided by all pharmacy-reported drugs. Here, sample persons are assumed to have reported pharmacies completely, and O2 reflects errors of recall on the part of the beneficiary.

MODEL 1: R=1-(S/P)

See under-reporting and over-reporting examples in Appendix A

In the second model, all unmatched survey-reported drugs were assumed to have been the result of an under-reporting of pharmacies. This model was further divided into two possibilities. In model 2A we assumed that all unmatched drugs were perfectly reported, that is, that there was no mis-reporting of drugs obtained from the omitted pharmacies.

MODEL 2A: R=1 – (S/ (P+O1+O2))

See under-reporting and over-reporting examples in Appendix A

In model 2B, we modify the assumption about mis-reporting in omitted pharmacies. We assume that the same reporting percentage observed from the known pharmacy(ies) occurred in the omitted pharmacies.

MODEL 2B: R=1-(S / (P+((O1+O2) * P/M)))[15]

See under-reporting and over-reporting examples in Appendix A

[15] When the beneficiary's pharmacy(ies) reported no drugs, the denominator was set to O2, which in this case is equal to S.

The third model combined aspects of the preceding two. In this model, survey-only events involving drugs encountered in the pharmacy data were considered over-reports and prescriptions not seen in the pharmacy data were classified as omitted-pharmacy-under-reports. As with model 2, two alternative specifications are possible depending upon the assumption of mis-reporting in omitted pharmacies. Model 3A assumed perfect reporting of drugs:

MODEL 3A: R=1-(S / (P+O2))

See under-reporting and over-reporting examples in Appendix A

Model 3B, similar to model 2B, assumed the same rate of over or under-reporting in omitted pharmacies as in reported pharmacies.

MODEL 3B: R=1-(S / (P+(O2 * P / (M+O1))))[16]

See under-reporting and over-reporting examples in Appendix A

We adopted model 3B for both our utilization and expenditure analyses, with one modification. Analysis of the imputations for the constant reporting rate assumption in this model uncovered some cases where the imputed number of estimated additional scripts purchased for under-reporters seemed unrealistic. Consequently, we established a cap on that number of any given beneficiary. First, an unconstrained number of estimated additional scripts was calculated for each beneficiary using the methodology described above. Next, the mean (2.54) and standard deviation (8.01) of these additional prescriptions were tabulated. Finally, for each beneficiary, the number of additional scripts purchased for an under-reporter was truncated at O2 plus two standard deviations above the mean of the estimated additional prescriptions purchased.

For instance, assume a respondent reported 32 prescriptions and their pharmacy reported just 3, of which 1 drug event matched. Further assume that each of the 31 remaining survey-reported prescriptions are classified as omitted-pharmacy-under-reports.

Unconstrained, the model would have resulted in an additional 93 prescriptions from an omitted pharmacy. Under the constraint described, the number of additional scripts is capped at (31+20.5) = 52.

Results -- Utilization

The distribution of net adjusted utilization under-reporting rates using model 3B is shown in Table 2 below:

Table 2: Distribution of Net Adjusted Under-reporting Rates

Percentile of the population	**Net adjusted mis-reporting percentage**
Maximum	100% under-reported
99th Percentile	100% under-reported
95th Percentile	100% under-reported
90th Percentile	67% under-reported
75th Percentile	36% under-reported
50th Percentile (Median)	10% under-reported
25th Percentile	0
10th Percentile	33% over-reported
5th Percentile	67% over-reported
1st Percentile	200% over-reported
Minimum	799% over-reported
MODE	100% under-reported

The net adjusted utilization under-reporting rates of all models are displayed in Table 3 below.

[16] Identical to scenario 2B, when the beneficiary's pharmacy(ies) reported no drugs, the denominator was set to O2, which in this case is equal to S.

Table 3: Model summaries and their associated formulas and net adjusted under-reporting rates

		Under-reporting Rate (unweighted)	Under-reporting Rate (weighted)
1	S / P	4.2%	4.4%
2a	S/ (P+O1+O2)	22.2%	22.2%
2b	S / (P+((O1+O2) * P/M))	27.3%	27.4%
3a	S / (P+O2)	13.3%	13.5%
3b	S / (P+(O2 * P / (M+O1)))	14.7%	14.9%

Note: The net adjusted under-reporting rate for Scenario 3B incorporates the capping procedure described.

Determining what factors are predictive of utilization mis-reporting

A multi-step process was used to determine the overall net adjusted utilization under-reporting rate of prescription drugs in the MCBS. First, we analyzed the demographic data of the follow-back participants to determine those factors that were predictive of a person's reporting status (over-, under-, or perfect-reporter). Second, we determined those factors that were predictive of the degree to which a person under or over reports their prescription use. After these models were developed, they were applied to those beneficiaries not in the follow-back study so that an aggregate estimate of under-reporting could be made.

We analyzed the demographic characteristics of the 5,302 full pharmacy follow-back participants to determine which factors were predictive of utilization under-reporting. The results from the pharmacy follow-back were merged to the 1999 Cost and Use Public Use File. Several variables were analyzed via multinomial logistic regression to test their predictive power of a person's reporting status. Table 4, below, lists the independent variables and our null hypotheses associated with each.

All of these variables are collected in the survey itself, with the exception of Total Average Wholesale Price (AWP) and Drug Coverage. Total AWP was estimated by merging beneficiary-reported drug data with First DataBank's drug pricing compendium, Bluebook. Dependent on each drug's name, form, strength, and prescription size, an average wholesale price is imputed. This imputation scheme is discussed more thoroughly later in the paper.

Drug coverage is also a derived field. For the purposes of this analysis, drug coverage is assigned if a beneficiary has at least one month of drug coverage in 1999. Beneficiaries are categorized as 'covered' if one or more of the following occur:

- Medicare+Choice beneficiaries: They belong to a plan that offers prescription drug coverage as part of its basic benefit package or they purchase such coverage via an added premium
- Medicaid beneficiaries: They are fully entitled, as determined by CMS administrative data, or they self-identify Medicaid drug coverage
- Privately insured beneficiaries: They report a private plan (employer-sponsored or individually purchased) that covers their prescription drugs
- 'Other public' insured beneficiaries: They report drug coverage from state-based pharmaceutical assistance programs, the Veteran's Administration, the Department of Defense, or any other public source
- All beneficiaries: They report any third party reimbursements[17]

Table 4: Variables Included in Original Regression Model

Concept	**Null Hypothesis**
Total Prescriptions	As the number of prescriptions increases, so does the likelihood of mis-reporting
Total AWP	As the level of expenditures increase, so does the likelihood of mis-reporting
Age	The disabled and the oldest old would mis-report drug use to a greater degree relative to the youngest old
Race/Ethnicity (White, African-American, Other)	There would be greater mis-reporting of drugs among non-whites

[17] Very rarely, survey respondents will report that they do not have prescription drug coverage via any private or public plans, yet they report that a private or public plan made a drug payment on their behalf

Health Measures (Self-reported health status, presence of chronic conditions)	Healthier beneficiaries would report their drug use more accurately
Income	Higher income enrollees would report their drug use more accurately
Utilization of Medicare Covered Services	Beneficiaries with increased utilization of Medicare covered services would mis-report more often
Gender	Females would be more likely to mis-report their drug use
Drug Coverage	Covered enrollees would mis-report more often relative to non-covered enrollees
Supplementary Health Insurance (Medicare+Choice, Medicaid, Employer-sponsored, Individually-purchased, Other public, Fee-for-Service Medicare Only	Likelihood is high for differences between certain types of supplemental insurance

Reporting status was one of three types. Under-reporters were those beneficiaries whose reported medications totaled less than the estimated number of prescriptions purchased, where the estimated number of prescriptions purchased equals the sum of pharmacy-reported drugs plus imputed pharmacy drugs. Over-reporters were those enrollees who reported more prescriptions than were estimated purchased. Persons were labeled as perfect-reporters if their reported drug use matched the total estimated purchased drugs. The unweighted frequencies of each category among the pharmacy follow-back participants were 3,037 (57.4%), 1,269 (24%), and 990 (18.7%), respectively.

Study members averaged 365 days in the community during 1999 while non-members spent just 332 days in the community.[18] In order to adjust for this experience, prescriptions per beneficiary were standardized to annual figures (ANN_TOTSCRIP) for modeling development purposes by dividing 365 days by each beneficiary's community days (C_DAYS). The final equation was as follows:

ANN_TOTSCRIP=((365/C_DAYS)*TOTSCRIP
Note: Variable definitions can be found in Appendix B

Relative to perfect-reporters, ten factors were statistically significantly predictive of reporting status as determined by the following multinomial logistic regression equations:[19]

$$LogP(O)/P(U)= \quad Exp(X_i\beta_m) / Exp(X_i\beta_n)$$

$$LogP(U)/P(P)= \quad Exp(X_i\beta_m) / Exp(X_i\beta_n)$$

$$LogP(P)/P(O)= \quad Exp(X_i\beta_m) / Exp(X_i\beta_n)$$

Where P(U)=Probability of being an under-reporter
P(O)=Probability of being an over-reporter
P(P)=Probability of being a perfect reporter
Subscript u=under-reporter
Subscript o=over-reporter
Subscript p=perfect reporter
T=Annualized total prescriptions
C=Number of chronic conditions
V=Number of Dr. visits
Y=(age 65-79=1, All Others=0)
A=African-American=1, All Others=0
O=Other race=1, All Others=0
R=Drug coverage=1, No coverage=0

---

18 This variation is explained by the rules for inclusion in the study: Non-participants included beneficiaries who began receiving Medicare benefits during the year, beneficiaries who died during the year, and those who moved between facilities and the community.

19 The original model tested the following variables: Number of beneficiary-reported prescriptions, total AWP, age category, race and ethnicity, health status, number of chronic conditions, income, number of

M=Medicare risk plan=1, No risk=0

D=Medicaid=1, No Medicaid=0

I=Income

Not all variables were significant for both over-reporters and under-reporters. The relative risk ratios are shown in Appendix C.

Generalizing the Results to the Full Population

To generalize these results to the population as a whole, the multinomial regression model described above was used to impute reporting status. Three new variables were created (Prob_Under, Prob_Over, and Prob_Perfect) dependent on the values of every sample person's demographics. These fields represented probabilities and combined, they summed to 1. Second, a random number between 0 and 1 was assigned to each beneficiary. A person's reporting status was determined based on a comparison of that number to the three probability variables. None of the beneficiaries in the pharmacy follow-back had their reporting status changed. Tables 5 and 6 illustrate the mechanics of the imputation and their corresponding results:[20]

Table 5: Imputation of reporting status variable

BASEID	00000001	00000002
RANDOM	0.75223	0.421134
PROB_UNDER	0.5509	0.5023
PROB_OVER	0.2203	0.2115
PROB_PERFECT	0.2288	0.2862
REPORT_STATUS	OVER	UNDER

inpatient hospitalizations, number of doctor visits, number of home health visits, number of outpatient procedures, gender, prescription drug coverage status, supplementary health insurance status

[20] Application of the model of reporting status to the total population resulted in trivial changes in the relative share of over and under-reporters.

Table 6: Detailed weighted statistics of reporting status variable, pre and post imputation

Status	Follow-back Percent	Imputed Percent	Full Sample Percent
Under-reporters	57.2%	55.7%	56.3%
Over-reporters	23.8%	23.1%	20.3%
Perfect reporters	19.1%	21.2%	23.4%

Following the assignment of reporting status, we estimated the degree to which respondents either under- or over-reported their drug events. As not all variables were predictive of a person's reporting status, separate models were developed to estimate the magnitude of under-reporting and over-reporting within their respective categories. Using only follow-back participants, an inflation factor was computed at the person level to annualize the number of prescriptions that had been identified as either under or over-reported. The factor was computed by dividing the total number of prescriptions in the public use file by the number of survey-reported prescriptions in the follow-back study. Given that the public use file represents one full year, and the follow-back analysis was limited to two rounds, on average, the inflation factor was about 1.5. The result was two new variables, UNDER365 and OVER365, where the former represented the annualized number of under-reported drugs for beneficiaries who were identified as under-reporters and the later represented the annualized number of over-reported prescriptions for beneficiaries identified as over-reporters. The table below demonstrates.

Table 7: Example of calculations to estimate annualized under and over-reported medications

BASEID	00000100	00000200	00000300
REPORT_STATUS	OVER	UNDER	PERFECT
TOTSCRIP	25	30	10
MTOT	20	20	6
FACTOR	1.25	1.50	1.67
U_EVNTS	0	10	0
O_EVNTS	4	0	0
UNDER365	0	15	0
OVER365	5	0	0

Limited to under-reporting follow-back participants, a multi-linear regression model was developed to determine which factors were predictive of the number of annualized under-reported prescriptions. The final equation follows:

$$UNDER365=\alpha_u T_i+\chi_u W_i+\delta_u H_i+\varepsilon_u V_i+\phi_u Y_i+\eta_u A_i+\varphi_u R_i+\kappa_u E_i+\lambda_u P_i+ ß$$

Where Subscript u=under-reporter

T=Annualized total prescriptions

W=Annualized total AWP

H=Health Status

V=Number of Dr. visits

Y=(age 65-79=1, All Others=0)

A= African-American=1, All Others=0

R= Drug coverage=1, No coverage=0

E=Employer Sponsored Insurance=1, No Employer Sponsored Insurance =0

P=Individually Purchased Insurance=1, No Individually Purchased Insurance =0

Note: The output from the model can be found in Appendix D.

Next, the model was used to predict a new variable, UNDER_RX, for all non-study members for whom an under-reporting figure had to be imputed. For the study's participants, UNDER_RX equaled UNDER365.

An identical process was followed for predicting the annualized number of over-reported drugs among over-reporters in the follow-back. That model's equation follows:

$$OVER365=\alpha_o T_i+\chi_o C+\delta_o I+\varepsilon_o F_i+\phi_o A_i+\lambda_o O_i+ ß$$

Where Subscript o=over-reporter

T=Annualized total prescriptions

C=Number of chronic conditions

I=Income

F=Female=1, Male=0

A=African-American=1, All Others=0

O=Other Ethnicity=1, All Others=0

Note: The output from the model can be found in Appendix E.

The variable OVER_RX, was assigned to all those deemed to be over-reporters. Like the under-reporters, the value for OVER_RX for follow-back participants was set equal to OVER365.

Calculating the final utilization mis-reporting rate

The next step was to transform the annualized number of over- or under-reported drugs to reflect the actual experience of each individual beneficiary. Reduction ratios were developed by dividing the number of prescriptions reported by the beneficiaries (TOTSCRIP) by their annualized number of prescriptions (ANN_TOTSCRIP). Table 8 illustrates how the annualized number of over- and under-reported medications were then multiplied by these ratios resulting in the variables, REAL_UNDER and REAL_OVER.

Table 8: Example of transforming annualized estimates into actual estimates

BASEID	00000100	00000200	00000300
REPORT_STATUS	OVER	UNDER	PERFECT
TOTSCRIP	25	30	10
ANN_TOTSCRIP	30	40	15
RATIO	0.83	0.75	0.67
UNDER365	0	15	0
OVER365	5	0	0
REAL_UNDER	0	11.25	0
REAL_OVER	4.17	0	0

The variable EST_PURCH (Table 9) was created to represent the number of actual prescriptions believed to have been purchased by the beneficiary and was calculated using the following equation:

EST_PURCH=TOTSCRIP+REAL_UNDER-REAL_OVER

Table 9: Example of EST_PURCH variable calculation

BASEID	00000100	00000200	00000300
REPORT_STATUS	OVER	UNDER	PERFECT
TOTSCRIP	25	30	10
ANN_TOTSCRIP	30	40	15
RATIO	0.83	0.75	0.67
UNDER365	0	15	0
OVER365	5	0	0
REAL_UNDER	0	11	0
REAL_OVER	4	0	0
EST_PURCH	21	41	10

To determine the final net adjusted utilization under-reporting ratio, weighted calculations of TOTSCRIP and EST_PURCH were taken resulting in a final under-reporting estimate of 17.7%.

Results – AWP Expenditures

There was a significant amount of overlap between the methods used to determine the utilization mis-reporting estimate for prescription drugs in the MCBS and those used to determine the mis-reporting estimate for medication expenditures in the survey. The methodologies are identical up to and including the selection of model 3B. From there, several other steps were undertaken.

In order to estimate expenditure mis-reporting, all survey-reported and pharmacy-reported events were electronically passed through a published industry source (First DataBank's Bluebook) that assigns a unit average wholesale price (AWP) to each prescription. The imputation algorithm attempts to match on as many characteristics of the drug as possible, including drug name, drug form, drug strength, and prescription size. In cases where form or strength were not collected in the survey, values for those fields are imputed using probabilities that are proportionate to their relative use among the dual-eligible population. CMS has drug utilization data at the National Drug Code (NDC) level for dually-eligible beneficiaries.[21] These data are merged by NDC to the

[21] One of the limitations in this study is the use of drug utilization data by dual-eligible beneficiaries to estimate the relative usage of forms and strengths within drug names for imputation purposes. IMS Health

First DataBank file to provide relative usage counts for the various forms, strengths, and package sizes within each drug. Once a unit AWP is assigned, that value is multiplied by the prescription size to estimate an "event price", or total AWP for that prescription. If we were unable to match on the name of the drug, or if we weren't given a prescription size on which to multiply the assigned unit AWP, we didn't ascribe an event price during this phase of the imputation.

Table 10: Results of Preliminary AWP Imputation by Reporting Status

	Survey-Reported	Pharmacy-Reported
Events	96,882	101,149
Translatable Events (name can be matched to First DataBank)	93,051	98,475
Event Prices Assigned (Unit AWP can be multiplied by prescription size)	84,926	83,405
Average Event Price	$ 50.89	$ 50.00

In order to fill in missing event prices, a two-step imputation process was implemented separately for both the survey-reported and pharmacy-reported data. First, for each of the two groups, average event prices were calculated for all unique drug names. These averages were then merged, by drug name, back onto their respective files. This left 588 drug events with missing event prices in the survey file and 3,042 drug events with missing event prices in the pharmacy file. Those remaining were assigned an event price equal to the average of all other drugs in their respective file (survey- or pharmacy-reported).

Table 11: Results of Final AWP Imputation by Reporting Status

also offers recent drug utilization figures by National Drug Code. We chose to use CMS administrative data for cost reasons.

	Survey-Reported	Pharmacy-Reported
Event Prices Assigned Via First DataBank	84,926	83,405
Event Price Imputed via Individual Drug Name Averages	11,368	14,702
Event Prices Imputed using Overall Event Price Average	588	3,042
Average Event Price Following all Imputation	$ 50.31	$ 49.62

In order to pass the results of our imputation through model 3B, all survey-reported and pharmacy-reported drugs needed to be organized into the same categories as described in the model (P, M, O1, O2). Unlike the utilization analysis, these distinctions needed to be made electronically. This was accomplished in a multi-step process. First, all of the survey-only and pharmacy-only drugs that didn't match electronically (for the reasons described in the section, "Including the proper medications from the beneficiary and the pharmacy") were downloaded to a flat file. For each beneficiary, every manually matched survey-only and pharmacy-only drug was flagged and the data were uploaded again. Next, an unduplicated list of all of the beneficiary's electronically and manually matched drug names was created. Then, all of the beneficiary's non-matched survey-reported drugs were electronically compared to that list. Drug names that matched during that comparison were classified as survey-over reports while those that didn't were flagged as omitted-pharmacy under-reports.

The next step was to tabulate the number of drug mentions that were categorized into the matched (M), survey over-report (O1), and omitted-pharmacy under-report (O2) categories. We then compared these figures to their corresponding estimates that resulted from our manual classifications prepared during our utilization mis-reporting estimation process. All sample persons with matching estimates were considered ready for expenditure analysis and were output to a file. All of the survey-reported drug mentions for the remaining beneficiaries were then downloaded to a flat file. Once more, through a manual process, these events were compared to the electronically and manually matched survey drugs to determine the category (O1, O2) in which the event belonged. All drugs that were deemed to be omitted-pharmacy under-reports were flagged and the data were uploaded again. At this point, we could accurately electronically identify for each sample person all of their drugs that were matched, pharmacy-only, survey over-reports, and omitted-pharmacy under-reports.

Their averages are listed in Table 12 below.

Table 12: Unweighted and Weighted Results of AWP Imputation by Category[22]

SURVEY-REPORTED	Number of Scripts	Unweighted Average AWP	Weighted Average AWP
All Matched	73,519	$ 49.30	$ 49.48
Electronically Matched	64,273	$ 50.30	$ 50.41
Manually Matched	9,246	$ 42.34	$ 42.85
Unmatched	23,359	$ 53.45	$ 53.62
Survey Over-reports	12,779	$ 50.95	$ 51.54
Pharmacy Under-reports	10,580	$ 56.48	$ 56.09

PHARMACY-REPORTED	Number of Scripts	Unweighted Average AWP	Weighted Average AWP
All Matched	73,519	$ 50.21	$ 50.39
Electronically Matched	64,273	$ 50.64	$ 50.79
Manually Matched	9,246	$ 47.20	$ 47.56
Unmatched	27,625	$ 48.04	$ 48.41

After classifying and pricing each drug, the next step was to calculate the total survey-reported AWP expenditures and the total adjusted pharmacy-reported AWP expenditures at the person-level. The formula for determining total adjusted pharmacy-reported AWP was contingent on whether the beneficiary was deemed to have fully reported all of their pharmacies. Where pharmacies were fully reported, the following equation was used:

$$PHARMAWP = M * PharmMat\$ + (P-M) * PharmOnly\$$$

Where:

PharmMat$=Average AWP of beneficiary's matched pharmacy-reported drugs

PharmOnly$= Average AWP of beneficiary's non-matched pharmacy-reported drugs

For beneficiaries for whom pharmacies were deemed under-reported, total adjusted pharmacy-reported AWP expenditures were estimated with the following formula:

[22] For each category of drugs, the Weighted Average AWP was calculated by multiplying the AWP of each drug by that person's MCBS sampling weight, summing those results, and then dividing by the sum of those weights.

PHARMAWP = M * PharmMat$ + (P-M) * PharmOnly$ + ((O2 * P / (M+O1) * O2$

Where:

PharmMat$=Average AWP of beneficiary's matched pharmacy-reported drugs

PharmOnly$= Average AWP of beneficiary's non-matched pharmacy-reported drugs

O2$=Average AWP of beneficiary's omitted-pharmacy under-reported drugs

Calculating the total survey-reported AWP expenditures for all beneficiaries required just one formula:

SURVAWP = M * SurvMat$ + O1 * O1$ + O2 * O2$

Where:

SurvMat$=Average AWP of beneficiary's matched survey-reported drugs

O1$= Average AWP of beneficiary's survey over-reports

O2$=Average AWP of beneficiary's omitted-pharmacy under-reported drugs

In preparation for the modeling phase of the estimate, total survey-reported AWP expenditures had to be estimated for all of the non-followback participants. For this cohort, event prices were derived by passing their reported prescriptions through First DataBank's Bluebook using the identical 'event price' algorithm described earlier. Drugs with missing AWP values following the Bluebook imputation were imputed by separately calculating average AWP by drug name and merging those values back onto the file.

A SAS-transport dataset was created containing 11,141 records (one per person). Each record contained BASEID, PHARMAWP, and SURVAWP. For non-followback participants, PHARMAWP was missing. This file was downloaded and converted to a Stata file.

Determining what factors are predictive of expenditure mis-reporting

Similar to the utilization mis-reporting method, a multi-step process was used to determine the overall net adjusted under-reporting of prescription drug expenditures in the MCBS. We began by analyzing the demographic data of the follow-back participants to determine those factors that were predictive of a person's AWP expenditure reporting status. Next, we determined those factors that were predictive of the degree to which a person under- or over-reports their prescription expenditures. After these models were developed, they were applied to those beneficiaries not in the follow-back study so that an aggregate estimate of expenditure mis-reporting could be made.

We re-analyzed the demographic characteristics of the pharmacy follow-back participants to establish which factors, if any, were predictive of expenditure mis-reporting. Multiple variables were examined using multinomial logistic regression to test their predictive power of a person's expenditure reporting status. Table 13, below, lists the independent variables and our null hypotheses associated with each.

Table 13: Variables Included in Original Regression Model

Concept	**Null Hypothesis**
Total Prescriptions	As the number of prescriptions increases, so does the likelihood of mis-reporting
Total AWP	As the level of expenditures increase, so does the likelihood of mis-reporting
Age	The disabled and the oldest old would mis-report drug expenditures to a greater degree relative to the youngest old
Race/Ethnicity (White, African-American, Other)	There would be greater mis-reporting of drug expenditures among non-whites
Health Measures (Self-reported Health Status, presence of chronic conditions)	Healthier beneficiaries would report their drug expenditures more accurately
Income	Higher income enrollees would report their drug expenditures more accurately

Utilization of Medicare Covered Services	Beneficiaries with increased utilization of Medicare covered services would mis-report more often
Gender	Females would be more likely to mis-report their drug expenses
Drug Coverage	Covered enrollees would mis-report more often relative to non-covered enrollees
Supplementary Health Insurance	Likelihood is high for differences between certain types of supplemental insurance

Reporting status was one of three types. Under-reporters were defined as those beneficiaries whose reported total AWP expenditures totaled less than the estimated total AWP expenditures. Over-reporters were those enrollees who reported total AWP expenditures greater than was estimated from pharmacies. Persons were labeled as perfect-reporters if their reported AWP drug expenditures matched the total estimated AWP expenses. The unweighted frequencies of each category among the pharmacy follow-back participants were 3,221 (60.8%), 1,564 (29.5%), and 511 (9.7%), respectively.

Not unlike use, there were significant expenditure differences between the follow-back participants and non-participants. The weighted average total AWP for the follow-back participant was $1,364.77, or 16% higher than the $1,173.22 estimated for non-participants. Resembling the method to annualize prescriptions described in the utilization mis-reporting model, total AWP expenses were standardized to annual figures (ANN_TOTAWP) for modeling purposes. The exact equation was as follows:

ANN_TOTAWP=((365/C_DAYS)*TOTAWP)

Note: Variable definitions can be found in Appendix B

Relative to perfect-reporters, there were ten factors that were statistically significantly predictive of reporting status as determined by the following multinomial logistic regression equations:[23]

$$LogP(O)/P(U)= \quad Exp(X_i\beta_m) / Exp(X_i\beta_n)$$

$$LogP(U)/P(P)= \quad Exp(X_i\beta_m) / Exp(X_i\beta_n)$$

$$LogP(P)/P(O)= \quad Exp(X_i\beta_m) / Exp(X_i\beta_n)$$

Where :

P(U)=Probability of being an under-reporter

P(O)=Probability of being an over-reporter

P(P)=Probability of being a perfect reporter

Subscript u=under-reporter

Subscript o=over-reporter

Subscript p=perfect reporter

T=Annualized total prescriptions

W=Annualized total AWP

G=Level of self-reported health status

C=Number of chronic conditions

V=Number of Dr. visits

O=Other race=1, All Others=0

M=Medicare risk plan=1, No risk plan=0

D=Medicaid=1, No Medicaid=0

E=Employer-sponsored=1, No Employer-sponsored=0

F=Female=1, Male=0

[23] The original model tested the following variables: Number of beneficiary-reported prescriptions, total AWP, age category, ethnicity, health status, number of chronic conditions, income, number of inpatient hospitalizations, number of doctor visits, number of home health visits, number of outpatient procedures, gender, prescription drug coverage status, supplementary health insurance status

Not all variables were significant for both over-reporters and under-reporters. The relative risk ratios are shown in Appendix F.

Generalizing the Results to the Full Population

To apply these results to the entire population, the multinomial regression model described above was used to impute reporting status. Three new variables were created (Prob_Under$, Prob_Over$, and Prob_Perfect$) dependent on the values of every sample person's demographics. These fields represented probabilities and combined, they summed to 1. Second, a random number between 0 and 1 was assigned to each beneficiary. A person's reporting status was determined based on a comparison of that number to the three probability variables. None of the beneficiaries in the pharmacy follow-back had their reporting status changed. Tables 14 and 15 illustrate the mechanics of the imputation and their corresponding results:[24]

Table 14: Imputation of AWP expenditure reporting status variable

BASEID	00000001	00000002
RANDOM	0.75223	0.421134
PROB_UNDER$	0.5509	0.5023
PROB_OVER$	0.2203	0.2115
PROB_PERFECT$	0.2288	0.2862
REPORT_STATUS	OVER	UNDER

Table 15: Detailed weighted statistics of AWP expenditure reporting status variable, pre and post imputation

Status	Follow-back Percent	Imputed Percent	Full Sample Percent
Under-reporters	60.9%	59.0%	59.8%
Over-reporters	29.3%	28.0%	28.5%
Perfect reporters	9.8%	13.0%	11.7%

Following the assignment of reporting status, we estimated the degree to which respondents either under- or over-reported their drug expenses using separate models for each category. Using only follow-back participants, an inflation factor was computed at the person level to annualize the level of AWP expenditures that had been identified as either under- or over-reported. The factor was computed by dividing the sum of AWP expenses estimated for all of the beneficiary’s purchases by the sum of their AWP expenditures from the follow-back study. Given that follow-back participants had approximately a full year of community exposure, and the follow-back analysis was limited to two rounds, on average, the inflation factor was about 1.5. The result was two new variables, UNDER365$ and OVER365$, where the first represented the annualized level of AWP expenditures for beneficiaries who were identified as under-reporters and the second represented the annualized level of AWP expenditures for beneficiaries identified as over-reporters. The following table illustrates the mechanic of this operation:

Table 16: Example of calculations to estimate annualized under and over-reported AWP expenditures

BASEID	00000100	00000200	00000300
REPORT_STATUS	OVER	UNDER	PERFECT
Full Year AWP	$ 1,000.00	$ 1,200.00	$ 400.00
SURVAWP	$ 800.00	$ 800.00	$ 240.00
FACTOR	1.25	1.50	1.67
U_AWP	$ -	$ 400.00	$ -
O_AWP	$ 200.00	$ -	$ -
UNDER365$	$ -	$ 600.00	$ -
OVER365$	$ 250.00	$ -	$ -

Restricted to under-reporting follow-back participants, a multi-linear regression model was developed to determine which factors were predictive of the level of annualized under-reported AWP. The final equation follows:

$$\text{UNDER365\$} = \alpha u W_i + \delta u C_i + \chi u V_i + \varepsilon u Y_i + \lambda u L_i + \phi u R_i + \eta u A_i + \varphi u Z_i + ß$$

[24] Application of the model of reporting status to the total population resulted in trivial changes in the relative share of over and under-reporters.

Where:

Subscript u=under-reporter

W=Annualized AWP

C=Number of chronic conditions

V=Number of Dr. visits

Y=Ages 65-79=1, All Others=0

L= Age 80+=1, All Others=0

R=Drug coverage=1, No drug coverage=0

A=African-American=1, All Others=0

Z=Other public coverage=1, No other public coverage=0

Note: The output from the model can be found in Appendix G.

Next, the model was used to predict a new variable, UNDER_RX$, for all non-study members designated to be under-reporters. For the study's participants, UNDER_RX$ equaled UNDER365$.

An identical process was followed for predicting the annualized level of over-reported AWP expenses among over-reporters in the follow-back study. That model's equation follows:

$$OVER365\$=\alpha_o W_i+\chi_o C+\varepsilon_o Y_i+\phi_o L_i+\lambda_o HH_i+ ß$$

Where:

Subscript o=over-reporter

W=Annualized AWP

C=Number of chronic conditions

Y=Ages 65-79=1, All Others=0

L= Age 80+=1, All Others=0

HH=Number of home health visits

Note: The output from the model can be found in Appendix H.

The variable OVER_RX$, was assigned to all those deemed to be over-reporters. Like the under-reporters, the value for OVER_RX$ for follow-back participants was set equal to OVER365$.

Calculating the final AWP expenditure mis-reporting rate

The next step was to convert the annualized level of over or under-reported expenditures to reveal the actual experience of each individual beneficiary. Reduction ratios were developed by dividing the level of AWP expenditures reported by the sample persons (TOTAWP) by their annualized AWP expenditures (ANN_TOTAWP). Table 17 illustrates how the annualized levels of under- and over-reported AWP were then multiplied by these ratios resulting in the variables, REAL_UNDER$ and REAL_OVER$.

Table 17: Example of transforming annualized AWP estimates into actual AWP estimates

BASEID	00000100	00000200	00000300
REPORT_STATUS	OVER	UNDER	PERFECT
TOTAWP	$ 1,000.00	$ 1,200.00	$ 400.00
ANN_TOTAWP	$ 1,200.00	$ 1,600.00	$ 600.00
RATIO	0.83	0.75	0.67
UNDER365$	$ -	$ 600.00	$ -
OVER365$	$ 200.00	$ -	$ -
REAL_UNDER$	$ -	$ 450.00	$ -
REAL_OVER$	$ 166.67	$ -	$ -

The variable EST_PURCH$ (Table 18) was created to represent the level of AWP expenditures believed to have been purchased by the beneficiary and was calculated using the following equation:

EST_PURCH$=TOTAWP+REAL_UNDER$-REAL_OVER$

Table 18: Example of EST_PURCH$ variable calculation

BASEID	00000100	00000200	00000300
REPORT_STATUS	OVER	UNDER	PERFECT
TOTAWP	$ 1,000.00	$ 1,200.00	$ 400.00
ANN_TOTAWP	$ 1,200.00	$ 1,600.00	$ 600.00
RATIO	0.83	0.75	0.67
UNDER365$	$ -	$ 600.00	$ -
OVER365$	$ 200.00	$ -	$ -
REAL_UNDER$	$ -	$ 450.00	$ -
REAL_OVER$	$ 166.67	$ -	$ -
EST_PURCH$	$ 833.33	$ 1,650.00	$ 400.00

To determine the final net adjusted expenditure under-reporting ratio, weighted calculations of TOTAWP and EST_PURCH$ were taken resulting in a final under-reporting estimate of 17%.

Discussion

This analysis addresses the important issue of adjusting survey-reported drug use and expenditure data to account for survey under-reporting. It has demonstrated several critical findings regarding prescription drug data collection among the Medicare elderly and disabled populations, including:

Utilization

- Medicare beneficiaries, on average, under-report 17.7% of their drug utilization, as measured in number of prescriptions filled or refilled (23.3 – Reported, 28.3 – Estimated).
- Many beneficiaries actually over-report their drug utilization.
- Adjusted for under-reporting (Table 19), the data show that approximately 25% of Medicare beneficiaries filled more than 40 prescriptions in 1999.

Table 19: Percentile Distributions of Reported and Estimated Drug Use

Percentile	Estimate Prescriptions /Beneficiary	Reported Prescriptions/Beneficiary
5%	0.0	0.0
10%	0.0	0.0
25%	9.0	6.0
50%	21.5	17.0

75%	40.3	34.0
90%	63.7	54.0
95%	80.4	70.0

- The probability of mis-reporting drug use increases with increased utilization, as well as increases in the number of chronic conditions.
- The most accurate utilization reporters (Table 20) filled between 5 and 10 prescriptions in 1999.
- Although the number of mis-reported drugs increases with utilization, the rate at which they are mis-reported (Table 20) is relatively consistent following the passing of the 15-prescription threshold.

Table 20: Mis-reporting Rates by Estimated Prescriptions per Beneficiary Category

stimated Utilization Category	Estimated Prescriptions/Beneficiary	Reported Prescriptions/Beneficiary	Percent Under- / Over-Reporte
=5	1.1	1.4	27.32% Over-reported
1-10	7.8	6.8	12.08% Under-reported
).1-15	12.6	10.1	20.22% Under-reported
5.1-20	17.7	14.5	18.08% Under-reported
).1-25	22.5	18.3	18.83% Under-reported
5.1-30	27.5	22.9	16.64% Under-reported
).1-35	32.5	26.9	17.27% Under-reported
5.1-40	37.5	31.2	16.68% Under-reported
).1-45	42.5	34.7	18.31% Under-reported
5.1-50	47.6	38.4	19.36% Under-reported
).1-55	52.5	42.7	18.66% Under-reported
5.1-60	57.4	48.5	15.46% Under-reported
60	84.8	68.9	18.71% Under-reported

- Among utilization over-reporters, heavy drug users and minorities tend to over-report to a greater degree.
- Among utilization under-reporters, the number of medication purchases that beneficiaries under-report increases with an increasing number of physician visits, but decreases for those who are privately insured.

Expenditures

- Medicare beneficiaries, on average, under-report 17% of their drug expenses ($1,253.25 – Reported, $1,510.23 – Estimated).
- Many beneficiaries over-report their prescription drug spending.

- The probability of mis-reporting drug spending (as measured by total AWP) increases with increases in expenditures, as well as being enrolled in a Medicare+Choice plan or Medicaid.
- Analysis of the percentile distributions from Table 21 shows that, when adjusted for under-reporting, the median spending level exceeds $1,000, up from an unadjusted figure of $809.

Table 21: Percentile Distributions of Reported and Estimated AWP Expenditures

Percentile	Estimated Total AWP	Reported Total AWP
5%	$ 0	$ 0
10%	$ 0	$ 0
25%	$ 326.85	$ 225.03
50%	$ 1,028.20	$ 809.62
75%	$ 2,108.99	$ 1,720.67
90%	$ 3,468.01	$ 2,891.03
95%	$ 4,595.98	$ 3,936.13

- The most accurate expenditure reporters, as shown in Table 22, tend to be those beneficiaries who purchased between $250 and $500 in drugs, as measured by AWP expenses.

Table 22: Mis-reporting Rates by Estimated AWP Expenditure Level Category

Estimated AWP Category	Average Estimated AWP	Average Reported AWP	Percent Under- / Over-Reported
<=$250	$ 35.09	$ 96.61	175.3% Over-reported
$251-500	$ 374.93	$ 338.91	9.6% Under-reported
$501-750	$ 621.85	$ 535.95	13.8% Under-reported
$751-1000	$ 873.00	$ 715.22	18.1% Under-reported
$1,001-1,250	$ 1,122.71	$ 930.05	17.2% Under-reported
$1,251-1,500	$ 1,369.40	$ 1,103.79	19.4% Under-reported
$1,501-1,750	$ 1,623.46	$ 1,309.45	19.3% Under-reported
$1,751-2,000	$ 1,868.18	$ 1,530.59	18.1% Under-reported
$2,001-2,250	$ 2,125.82	$ 1,673.73	21.3% Under-reported
$2,251-2,500	$ 2,369.07	$ 1,934.03	18.4% Under-reported
>$2,501	$ 4,258.02	$ 3,488.21	18.1% Under-reported

- Among expenditure over-reporters, being an aged beneficiary mitigates the degree to which you over-report.

- The amount of expenses that beneficiaries under-report increases with an increasing number of physician visits, among expenditure under-reporters.
- Beneficiaries frequently report incomplete drug names (eg. Cardizem instead of Cardizem CR) leading to drug cost estimates that are below the actual total expenditure level . This more than offsets the practice of inadvertently reporting more expensive brand name drugs when beneficiaries, in fact, received a less expensive generic drug.
- With respect to average drug prices, Medicare beneficiaries have a propensity to, a) report drug purchases that were, to some extent, higher in cost, and b) not report drug purchases that were somewhat less expensive, marginally offsetting their recall error rate.

Finally, prior to this analysis, the Information and Methods Group (IMG) recommended using the under-reporting estimate for physician visits (30%) as a proxy for the under-reporting level related to prescription drugs.[25] Adjusted for a net-expenditure under-reporting rate of 17%, 1999 MCBS data indicate that outpatient prescription drug spending among the non-institutionalized Medicare population totaled approximately 46.7 billion dollars. Given that level of expenditure, being precise with respect to assumptions made regarding survey mis-reporting takes on added significance as just a one percentage point difference in the under-reporting level estimate can change the total projected annual outlays by nearly 570 million dollars. We believe the analysis described here will help inform policy-makers and contribute to improved accuracy of cost estimates of various prescription drug legislative proposals.

[25] IMG is located within the Office of Research, Development, and Information and is responsible for maintaining and analyzing the MCBS.

APPENDIX A
Description of Various Under- and Over-reporting Models

MODEL 1: UNDER-REPORTING EXAMPLE

Survey-reported drugs: 12

Pharmacy-reported drugs: 16

Matched drugs: 8

Net Adjusted Under-reporting rate = 1-(12 / 16) = .25 = 25%

MODEL 1: OVER-REPORTING EXAMPLE

Survey-reported drugs: 22

Pharmacy-reported drugs: 12

Matched drugs: 10

Net Adjusted Under-reporting rate = 1-(22 / 12) = -.833 = -83%

MODEL 2A: UNDER-REPORTING EXAMPLE

Survey-reported drugs: 12

Pharmacy-reported drugs: 16

Matched drugs: 8

Survey-over-reports: 1

Omitted-pharmacy-under-reports: 3

Omitted pharmacy's expected number of reported drugs (estimated additional scripts purchased): = 1 + 3 = 4

Net Adjusted Under-reporting rate = 1-(12 / (16+ 4)) = .40 = 40%

MODEL 2A: OVER-REPORTING EXAMPLE

Survey-reported drugs: 22

Pharmacy-reported drugs: 12

Matched drugs: 10

Survey-over-reports: 10

Omitted-pharmacy-under-reports: 2

Omitted pharmacy's expected number of reported drugs (estimated additional scripts purchased): = 10+2 = 12

Net Adjusted Under-reporting rate = 1-(22 / (12+12)) = .083 = 8.3%

MODEL 2B: UNDER-REPORTING EXAMPLE

Survey-reported drugs: 12

Pharmacy-reported drugs: 16

Matched drugs: 8

Survey-over-reports: 1

Omitted-pharmacy-under-reports: 3

Reporting ratio for known pharmacy: 16 / 8 =2

Omitted pharmacy's expected number of reported drugs (estimated additional scripts purchased): = 4 * 2 = 8

Net Adjusted Under-reporting rate = 1-(12 / (16+ 8)) = .50 = 50%

MODEL 2B: OVER-REPORTING EXAMPLE

Survey-reported drugs: 22

Pharmacy-reported drugs: 12

Matched drugs: 10

Survey-over-reports: 10

Omitted-pharmacy-under-reports: 2

Reporting ratio for known pharmacy: 12 / 10 =1.20

Omitted pharmacy's expected number of reported drugs (estimated additional scripts purchased): = 12 * 1.2 = 14.4 = 14

Net Adjusted Under-reporting rate = 1-(22 / (12+14)) = .154 = 15.4%

MODEL 3A: UNDER-REPORTING EXAMPLE

Survey-reported drugs: 12

Pharmacy-reported drugs: 16

Matched drugs: 8

Survey-over-reports: 1

Omitted-pharmacy-under-reports: 3

Omitted pharmacy's expected number of reported drugs (estimated additional scripts purchased): = 3

Net Adjusted Under-reporting rate = 1-(12 / (16+3)) = .368 = 36.8%

MODEL 3A: OVER-REPORTING EXAMPLE

Survey-reported drugs: 22

Pharmacy-reported drugs: 12

Matched drugs: 10

Survey-over-reports: 10

Omitted-pharmacy-under-reports: 2

Omitted pharmacy's expected number of reported drugs (estimated additional scripts purchased): = 2

Net Adjusted Under-reporting rate = 1-(22 / (12+2)) = -.571 = -57.1%

MODEL 3B: UNDER-REPORTING EXAMPLE

Survey-reported drugs: 12

Pharmacy-reported drugs: 16

Matched drugs: 8

Survey-over-reports: 1

Omitted-pharmacy-under-reports: 3

Reporting ratio for known pharmacy: 16 / (8+1) =1.78

Omitted pharmacy's expected number of reported drugs (estimated additional scripts purchased): = 3 * 1.78 = 5.34 = 5

Net Adjusted Under-reporting rate = 1-(12 / (16+5)) = .428 = 42.8%

MODEL 3B: OVER-REPORTING EXAMPLE

Survey-reported drugs: 22

Pharmacy-reported drugs: 12

Matched drugs: 10

Survey-over-reports: 10

Omitted-pharmacy-under-reports: 2

Reporting ratio for known pharmacy: 12 / (10+10) =.60

Omitted pharmacy's expected number of reported drugs (estimated additional scripts purchased): = 2 * .60 = 1.20 = 1

Net Adjusted Under-reporting rate = 1-(22 / (12+1)) = -.692 = -69.2%

APPENDIX B
Description of Variables names and their sources

Variable Name	Description	Source
ANN_TOTAWP	Annualized Average Wholesale Price (AWP)	Generated
ANN_TOTSCRIP	Annualized TOTSCRIP	Generated
BASEID	Unique beneficiary identifier	Cost and Use, 1999 and Pharmacy Follow-back Study
BI_AF_AM	Binary, African-American	Cost and Use, 1999
BI_CAID	Binary, Medicaid insurance	Cost and Use, 1999
BI_EMP	Binary, Employer-sponsored insurance	Cost and Use, 1999
BI_OLD	Binary, ages 80 and up	Cost and Use, 1999
BI-OP	Binary, Other Public insurance	Cost and Use, 1999
BI_OTHER	Binary, Other race (not White or African-American)	Cost and Use, 1999
BI_PHI	Binary, Individually-purchased insurance	Cost and Use, 1999
BI_RISK	Binary, Medicare+Choice insurance	Cost and Use, 1999
BI_RXCOV	Binary, prescription drug coverage	Cost and Use, 1999
BI_YOUNG	Binary, ages 65-79	Cost and Use, 1999
C_DAYS	Number of days in community, 1999	Cost and Use, 1999
CHRONIC	Number of chronic conditions	Cost and Use, 1999
DRVISITS	Number of doctor visits	Cost and Use, 1999
EST_PURCH	Estimated number of prescriptions believed to have been purchased by the beneficiary in 1999	Generated
EST_PURCH$	Estimated level of AWP expenditures believed to have been purchased by the beneficiary in 1999	Generated
FACTOR	Inflation factor, defined as	Generated

	(TOTSCRIP/MTOT)	
HEALTH	Self-reported health of beneficiary (ordinal scale, 1-5)	Cost and Use, 1999
MTOT	Total number of beneficiary reported prescriptions in Rounds 24 and 25	Pharmacy Follow-back Study
O_EVENTS	Number of over-reported drugs among over-reporters: MTOT-TOTALCAP	Pharmacy Follow-back Study
O_AWP	Level of over-reported AWP expenses among over-reporters: SURVAWP-PHARMAWP	Pharmacy follow-back, generated
OVER_RX	Annualized imputed number of over-reported prescriptions among over-reporters	Generated
OVER_RX$	Annualized imputed level of over-reported AWP expenses among over-reporters	Generated
OVER365	Annualized number of over-reported drugs among over-reporters	Pharmacy Follow-back Study
OVER365$	Annualized level of over-reported AWP expenses among over-reporters	Pharmacy follow-back, generated
PHARMAWP	Capped level of AWP expenses estimated to have been purchased by beneficiary in 1999	Pharmacy follow-back, generated
PROB_OVER	Probability of being designated as an over-reporter	Generated
PROB_OVER$	Probability of being designated as an AWP expenditure over-reporter	Generated
PROB_PERFECT	Probability of being designated as a perfect-reporter	Generated
PROB_PERFECT$	Probability of being designated as an AWP expenditure perfect-reporter	Generated

PROB_UNDER	Probability of being designated as an under-reporter	Generated
PROB_UNDER$	Probability of being designated as an AWP expenditure under-reporter	Generated
RANDOM	Randomly generated number between 0 and 1	Generated
RATIO	Reporting rate (TOTSCRIP / EST_PURCH)	Generated
REAL_OVER	Transformed number of over-reported drugs among over-reporters	Generated
REAL_OVER$	Transformed level of over-reported AWP expenses among over-reporters	Generated
REAL_UNDER	Transformed number of under-reported drugs among under-reporters	Generated
REAL_UNDER$	Transformed level of under-reported AWP expenses among under-reporters	Generated
REPORT_STATUS	Beneficiary reporting status ('Under-reporter','Over-reporter','Perfect-reporter')	Generated
SURVAWP	Total level of beneficiary reported AWP expenses in Rounds 24 and 25	Pharmacy follow-back, generated
TOTALCAP	Capped total number of prescriptions estimated to have been purchased by beneficiary in 1999	Pharmacy follow-back, generated
TOTAWP	Total AWP expenses for 1999	Cost & Use, 1999
TOTSCRIP	Total prescriptions for 1999	Cost & Use, 1999
U_EVENTS	Number of under-reported drugs among under-reporters: TOTALCAP-MTOT	Pharmacy Follow-back Study
U_AWP	Level of under-reported AWP	Pharmacy follow-back, generated

	expenses among under-reporters: PHARMAWP-SURVAWP	
UNDER_RX	Annualized imputed number of under-reported prescriptions among under-reporters	Generated
UNDER_RX$	Annualized imputed level of under-reported AWP expenses among under-reporters	Generated
UNDER365	Annualized number of under-reported drugs among under-reporters	Pharmacy Follow-back Study
UNDER365$	Annualized level of under-reported AWP expenses among under-reporters	Pharmacy follow-back, generated

APPENDIX C
Relative risk ratios for reporting status

REP	RRR	STD. ERR.	T	P>|T|	[95% CONF.	INTERVAL]
UNDER-REPORTER						
ANN_TOTSCRIP	1.02012	.0033735	6.02	0.000	1.013515	1.026769
CHRONIC	1.117897	.0433334	2.88	0.004	1.03593	1.206349
INCOME_C	1.002726	.0013524	2.02	0.044	1.000073	1.005386
MPAEVNTS	1.013994	.0029756	4.74	0.000	1.008166	1.019856
BI_YNG	.741002	.0654738	-3.39	0.001	.6229249	.8814608
BI_RXCOV	1.203728	.0857882	2.60	0.010	1.046465	1.384624
BI_AF_AM	1.044207	.1790865	0.25	0.801	.7455314	1.462538
BI_OTHER	.6271836	.1355481	-2.16	0.031	.4102124	.9589162
BI_RISK	1.56772	.1891474	3.73	0.000	1.236898	1.987023
BI_CAID	1.519315	.2183676	2.91	0.004	1.14558	2.014977
OVER-REPORTER						
ANN_TOTSCRIP	1.038283	.0035828	10.89	0.000	1.031269	1.045346
CHRONIC	1.077724	.0563766	1.43	0.153	.9724742	1.194364
INCOME_C	1.000153	.0017065	0.09	0.929	.9968062	1.003511
MPAEVNTS	1.006567	.0034203	1.93	0.055	.9998699	1.013308
BI_YOUNG	.8007253	.0858158	-2.07	0.039	.6487068	.9883679
BI_RXCOV	.8356108	.0734555	-2.04	0.042	.7030821	.9931206
BI_AF_AM	1.601988	.2821842	2.68	0.008	1.133389	2.26433
BI_OTHER	1.207341	.3083749	0.74	0.461	.7310036	1.99407
BI_RISK	1.55242	.2134558	3.20	0.001	1.184953	2.033842
BI_CAID	1.275708	.2367956	1.31	0.190	.8859092	1.837018

(PERFECT REPORTERS ARE THE COMPARISON GROUP)

APPENDIX D
Results of multiple linear regression, R_UNDER365

SURVEY LINEAR REGRESSION

PWEIGHT:	C99WGT	NUMBER OF OBS =	2904
STRATA:	SUDSTRAT	NUMBER OF STRATA =	64
PSU:	SUDUNIT	NUMBER OF PSUS =	491
		POPULATION SIZE =	8769604.1
		F(12, 416) =	15.44
		PROB > F =	0.0000
		R-SQUARED =	0.1271

R_UNDER365	COEF.	STD. ERR.	T	P>|T|	[95% CONF.	INTERVAL]
ANN_TOTSCRIP	.0827889	.0265151	3.12	0.002	.0306724	.1349053
ANN_TOTAWP	.0015831	.000506	3.13	0.002	.0005886	.0025776
_IGENHELTH_2	-.8999554	.9018077	-1.00	0.319	-2.67249	.8725793
_IGENHELTH_3	-.4527387	.9828935	-0.46	0.645	-2.38465	1.479173
_IGENHELTH_4	1.199128	1.155594	1.04	0.300	-1.072232	3.470489
_IGENHELTH_5	4.416433	2.080764	2.12	0.034	.3266185	8.506248
MPAEVNTS	.0985806	.0192787	5.11	0.000	.0606877	.1364735
BI_YOUNG	-1.874666	.7879259	-2.38	0.018	-3.423362	-.3259697
BI_RXCOV	2.342071	.7941136	2.95	0.003	.7812123	3.902929
BI_AF_AM	3.981971	1.839239	2.17	0.031	.3668815	7.59706
BI_EMP	-3.015366	.9407723	-3.21	0.001	-4.864487	-1.166245
BI_PHI	-2.804111	.8997573	-3.12	0.002	-4.572616	-1.035606
_CONS	7.074387	1.29127	5.48	0.000	4.53635	9.612424

APPENDIX E
Results of multiple linear regression, R_OVER365

SURVEY LINEAR REGRESSION

PWEIGHT:	C99WGT	NUMBER OF OBS	=	1269
STRATA:	SUDSTRAT	NUMBER OF STRATA	=	65
PSU:	SUDUNIT	NUMBER OF PSUS	=	326
		POPULATION SIZE	=	3815147.3
		F(6, 256)	=	29.66
		PROB > F	=	0.0000
		R-SQUARED	=	0.2691

R_OVER365	COEF.	STD. ERR.	T	P>|T|	[95% CONF.	INTERVAL]
ANN_TOTSCRIP	.1732639	.015702	11.03	0.000	.1423451	.2041826
CHRONIC	.4518025	.1929737	2.34	0.020	.0718191	.8317859
INCOME_C	-.0204454	.0063774	-3.21	0.002	-.0330032	-.0078877
BI_FEMALE	-1.095882	.4450839	-2.46	0.014	-1.972295	-.2194699
BI_AF_AM	3.079474	.8844495	3.48	0.001	1.337909	4.821039
BI_OTHER	2.825821	.8334886	3.39	0.001	1.184603	4.467039
_CONS	1.437244	.6971909	2.06	0.040	.0644093	2.810079

APPENDIX F
Relative risk ratios for expenditure reporting status

REPORT	RRR	STD. ERR.	T	P>|T|	[95% CONF.	INTERVAL]
UNDER-REPORTER						
ANN_TOTSCRIP	1.04274	.0088527	4.93	0.000	1.025494	1.060277
ANN_TOTAWP	1.000202	.0001152	1.76	0.080	.9999761	1.000429
_IGENHELTH_2	.878652	.13271	-0.86	0.392	.6530668	1.18216
_IGENHELTH_3	1.315806	.1888364	1.91	0.056	.9925492	1.744343
_IGENHELTH_4	1.085319	.2053783	0.43	0.665	.748362	1.573993
_IGENHELTH_5	1.183464	.3459059	0.58	0.565	.6664874	2.101445
CHRONIC	1.140309	.0613496	2.44	0.015	1.02594	1.267428
MPAEVNTS	1.018277	.0053191	3.47	0.001	1.007881	1.02878
BI_FEMALE	1.2477	.1393492	1.98	0.048	1.001902	1.553799
BI_OTHER	.4610349	.1238251	-2.88	0.004	.2720146	.7814035
BI_RISK	1.946958	.2913845	4.45	0.000	1.451015	2.612409
BI_CAID	1.810141	.3830344	2.80	0.005	1.194484	2.743118
BI_EMP	1.309765	.1468208	2.41	0.016	1.05089	1.632412
OVER-REPORTER						
ANN_TOTSCRIP	1.049681	.0090032	5.65	0.000	1.032143	1.067518
ANN_TOTAWP	1.000351	.0001138	3.09	0.002	1.000128	1.000575
_IGENHELTH_2	1.02427	.1547527	0.16	0.874	.7612271	1.378206
_IGENHELTH_3	1.446507	.2433406	2.19	0.029	1.039435	2.013002
_IGENHELTH_4	1.307621	.283259	1.24	0.216	.8544167	2.001218
_IGENHELTH_5	.9599647	.3102434	-0.13	0.899	.5087802	1.811258
CHRONIC	1.12764	.0660925	2.05	0.041	1.005	1.265247
MPAEVNTS	1.012854	.0054205	2.39	0.017	1.002261	1.023558
BI_FEMALE	1.202219	.1348332	1.64	0.101	.9644924	1.49854
BI_OTHER	.7379603	.2141953	-1.05	0.296	.4172549	1.305162
BI_RISK	1.727666	.2956669	3.19	0.001	1.234392	2.418055
BI_CAID	1.525787	.3182734	2.03	0.043	1.012812	2.298578
BI_EMP	1.152943	.1506402	1.09	0.277	.8919428	1.490318

(PERFECT REPORTERS are the comparison group)

APPENDIX G
Results of multiple linear regression, R_UNDER365$

SURVEY LINEAR REGRESSION

PWEIGHT:	C99WGT	NUMBER OF OBS =	3091
STRATA:	SUDSTRAT	NUMBER OF STRATA =	64
PSU:	SUDUNIT	NUMBER OF PSUS =	496
		POPULATION SIZE =	9371125.4
		F(8, 425) =	23.37
		PROB > F =	0.0000
		R-SQUARED =	0.1383

R_UNDER365	COEF.	STD. ERR.	T	P>|T|	[95% CONF.	INTERVAL]
ANN_TOTAWP	.2266565	.0423352	5.35	0.000	.1434478	.3098651
CHRONIC	41.34224	15.97401	2.59	0.010	9.945794	72.73869
MPAEVNTS	4.474903	1.079446	4.15	0.000	2.353283	6.596523
BI_OLD	-230.3471	102.4492	-2.25	0.025	-431.7079	-28.98622
BI_YOUNG	-298.833	93.76712	-3.19	0.002	-483.1295	-114.5365
BI_RXCOV	192.5127	33.95269	5.67	0.000	125.7797	259.2457
BI_AF_AM	200.2031	91.72257	2.18	0.030	19.92513	380.4811
BI_OP	302.4975	147.848	2.05	0.041	11.9067	593.0884
_CONS	251.7298	89.43593	2.81	0.005	75.94617	427.5135

APPENDIX H
Results of multiple linear regression, R_OVER365$

SURVEY LINEAR REGRESSION

PWEIGHT:	C99WGT	NUMBER OF OBS =	1564
STRATA:	SUDSTRAT	NUMBER OF STRATA =	65
PSU:	SUDUNIT	NUMBER OF PSUS =	366
		POPULATION SIZE =	4705472.8
		F(5, 297) =	25.76
		PROB > F =	0.0000
		R-SQUARED =	0.1294

R_OVER365	COEF.	STD. ERR.	T	P>|T|	[95% CONF.	INTERVAL]
ANN_TOTAWP	.0933267	.0323955	2.88	0.004	.0295764	.157077
CHRONIC	55.63052	17.13858	3.25	0.001	21.90392	89.35712
HHAEVNTS	-.2236337	.0909783	-2.46	0.015	-.4026678	-.0445996
BI_OLD	-200.2182	52.65341	-3.80	0.000	-303.8336	-96.60277
BI_YOUNG	-160.098	58.29061	-2.75	0.006	-274.8067	-45.38923
_CONS	279.7384	66.18351	4.23	0.000	149.4975	409.9794