DOI: 10.1111/1475-6773.14278

METHODS BRIEF

HSR Health Services Research

Deductible imputation in administrative medical claims datasets

Betsy Q. Cliff PhD[1] | Julia C. P. Eddelbuettel BS[2] | Mark K. Meiselbach PhD[3] | Matthew D. Eisenberg PhD[3,4]

[1]Department of Public Health Sciences, University of Chicago, Chicago, Illinois, USA

[2]PhD Program in Health Policy, Harvard University, Cambridge, Massachusetts, USA

[3]Health Policy and Management, Johns Hopkins University, Baltimore, Maryland, USA

[4]Johns Hopkins University, Baltimore, Maryland, USA

Correspondence
Betsy Q. Cliff, Department of Public Health Sciences, University of Chicago, 5841 S. Maryland Ave, Chicago, IL 60637, USA.
Email: bqcliff@bsd.uchicago.edu

Funding information
National Institute on Drug Abuse, Grant/Award Number: R01DA044201

Abstract

Objective: To validate imputation methods used to infer plan-level deductibles and determine which enrollees are in high-deductible health plans (HDHPs) in administrative claims datasets.

Data Sources and Study Setting: 2017 medical and pharmaceutical claims from OptumLabs Data Warehouse for US individuals <65 continuously enrolled in an employer-sponsored plan. Data include enrollee and plan characteristics, deductible spending, plan spending, and actual plan-level deductibles.

Study Design: We impute plan deductibles using four methods: (1) parametric prediction using individual-level spending; (2) parametric prediction with imputation and plan characteristics; (3) highest plan-specific mode of individual annual deductible spending; and (4) deductible spending at the 80th percentile among individuals meeting their deductible. We compare deductibles' levels and categories for imputed versus actual deductibles.

Data Collection/Extraction Methods: Not applicable.

Principal Findings: All methods had a positive predictive value (PPV) for determining high- versus low-deductible plans of ≥87%; negative predictive values (NPV) were lower. The method imputing plan-specific deductible spending modes was most accurate and least computationally intensive (PPV: 95%; NPV: 91%). This method also best correlated with actual deductible levels; 69% of imputed deductibles were within $250 of the true deductible.

Conclusions: In the absence of plan structure data, imputing plan-specific modes of individual annual deductible spending best correlates with true deductibles and best predicts enrollees in HDHPs.

KEYWORDS
data analysis, health insurance, high-deductible health plans, research methods

What is known about this topic

- High-deductible health plans are an increasingly common type of benefit structure that may impact health care access, health and consumer finances.
- Research has been hindered by a lack of plan-level information on deductibles in administrative medical claims datasets.

What this study adds

- Using each plan's highest mode of annual individual deductible spending is a reasonably accurate way to identify high- versus low-deductible health plans, and more accurate than more computationally intensive methods.
- When imputing deductibles for a categorical distribution, limiting the sample to plans with ≥50 enrollees increases accuracy.
- All imputation methods are imperfect at predicting deductibles. Claims dataset vendors should include plan structure variables, including deductibles, in data releases so researchers do not have to rely on imputation.

1 | INTRODUCTION

High-deductible health plans (HDHPs) are among the most common type of health plan for the 155 million Americans who receive their health insurance through an employer and the 20 million who purchase insurance on the individual commercial market.[1,2] These plans aim to increase discernment of health care value and judiciousness of medical spending among enrollees through high cost sharing and, often, tax-advantaged savings accounts that incentivize patients to have a financial stake in their health care spending.[3–5] Indeed, HDHPs have been shown to reduce health care spending, largely by reducing utilization.[6–10] Some evidence shows HDHPs decrease valuable medical service utilization and worsen health; other studies show no effects on health and that patients disproportionately reduce potentially wasteful health services.[11–18]

An impediment to research on HDHPs is the lack of data about plan structure. Administrative claims data, ideal for research, often come in two types. The first type includes detailed information about plan structure but often has poor externality validity as it is typically sourced from a single health insurer or small subset of enrollees.[19] The second type has improved external validity by pooling across insurers but does not usually include plan-structure variables necessary to distinguish between HDHPs and plans with lower deductibles, or interpret what binary "HDHP" variables represent.[20] Several research groups have used claim-level deductible spending to impute deductibles, but different methods have been used and none has been validated against a full distribution of true deductible levels.[13,21] This paper aims to fill that gap by comparing several methods of deductible imputation used in previous literature or suggested by experts. We use an administrative claims dataset of the first type (sourced from a single insurer with plan-level deductibles) to validate imputation methods.

Researchers studying health insurance design or who want to adjust for deductible spending in their analyses can use the results of this paper to operationalize imputation of plan deductibles. We have included replication code for this purpose. By validating a method for imputing deductibles with common data elements, we hope to both improve the quality of research on health insurance and expand the scope of data that can be used for understanding the effects of HDHPs.

2 | METHODS

Our goal is to impute plan-level deductibles when data do not include them. Then, using these imputed values, assign plans to categories of deductibles and binary high−/low-deductible status for one plan-year of claims.

2.1 | Data

We use 2017 de-identified administrative claims data from the OptumLabs Data Warehouse.[22] We pull medical and pharmaceutical claims for enrollees under age 65 in plans with at least 10 enrollees in employer-based insurance in the USA. We limit the sample to those who are continuously enrolled for 12 months in a single plan for which out-of-pocket spending resets on January 1 to ensure we capture spending for a full plan year.

Similar to most administrative claims databases, the data contain claim-level spending variables, including out-of-pocket spending (deductible, coinsurance, and copayment) and the amount paid by the health insurance plan. We use in-network claims and sum across these fields to find the total amount paid per claim. The data also contain other variables often included in claims datasets: enrollee demographic characteristics (self-reported gender, birthdate), plan characteristics (e.g., network structure), coverage level (individual/family), specific plan identifiers, and an anniversary date (cost-sharing reset date). It is necessary to be able to identify claim-level spending, coverage level, specific plans, and anniversary date for our imputation methods. Unlike many multipayer claims datasets, the OptumLabs data contain variables denoting administratively set annual deductibles that are consistent within plans. We leverage our ability to see both administratively set deductibles and enrollee spending in the same dataset to validate imputation methods.

2.2 | Deductible construction

We derive a claim-level dataset linking individual enrollee medical encounters to their insurance plans. We exclude out-of-network claims, which may not be subject to the general deductible. For plans with separate medical and pharmaceutical deductibles, we use the medical deductible in place of a general deductible and measure only medical spending. For plans with a general (medical + pharmaceutical) deductible, we combine medical and pharmaceutical claims. Within the full employer market, 85% of plans use a general deductible; for plans with a separate pharmaceutical deductible, the average pharmaceutical deductible is $150.[23] We top-coded data at the 99th percentile of deductible spending and bottom-coded at $0 to remove unreasonable values. We define HDHPs as plans with a deductible of ≥$1350, reflecting the Internal Revenue Service minimum deductible limit for HDHPs with a health savings account in 2017.

A general challenge in estimating deductibles is that most plans have separate deductible amounts for individual and family coverage and the way in which individual medical spending contributes to the family deductible varies. Claims datasets may not link family members or include information about the structure of family deductibles though often include variables that denote whether a person is enrolled as a part of a family. Because of these additional complications with estimating family deductibles, we impute deductibles for enrollees with individual-level coverage only. While we believe this is the most straightforward approach, we include in Supporting Information Appendix 1 additional considerations for researchers who wish to estimate family-level deductibles.

2.3 | Imputation

We test four methods for deductible imputation: three are based on methods previously used in peer-reviewed literature or in-progress work and one has been recommended by researchers familiar with claims datasets (Table 1). For all methods, we impute $0 as the deductible for plans with positive total spending but no deductible spending (2% of plans).

2.3.1 | Parametric prediction with spending (regress on spending method)

This method predicts deductibles using plan-specific variations in the observed relationship between individual deductible and total spending amounts.[24] To implement it, we regress each enrollee's annual deductible spending on their total annual spending (plan plus out-of-pocket), common demographic covariates (gender and age), and fixed effects for each plan (details are in Supporting Information, Appendix 1). Using the coefficients from the best-fit regression model, we predict deductibles for each plan at a fixed amount of total spending, which we set at $10,000 to exceed most deductibles. The coding and construction of variables for this method is simple, though processing time to generate predictions can be extensive in datasets with many plans.

2.3.2 | Parametric prediction with imputation and plan characteristics (regress on imputed deductibles method)

This method is inspired by an imputation method used in multiple papers published by the same research group.[11,25-27] Those papers

TABLE 1 Imputation methods and implementation intensity.

Imputation method	Brief description	Coding intensity	Approximate processing time
Parametric prediction with spending (regress on spending method)	Use regression to predict deductibles conditional for set spending level with plan fixed effects	Low	12–24 h
Parametric prediction with imputation and plan characteristics (Regress on imputed deductibles method)	Impute readily identifiable deductibles, then use regression to predict nonreadily identifiable	High	24–72 h
Modal deductible spending (mode method)	Impute highest nonzero plan-specific mode.	Low	<5 min
80th percentile of deductible spending (80th percentile method)	Impute based on the 80th percentile of the mode with a set of parameters	Moderate	1 h

Note: Coding intensity based on creating code in Stata 17.0 MP. Exact code for replication is included as a supplement to this publication. Approximate processing time based on time to run full code associated with each method and generate an imputed deductible for each plan in our sample (~2 million individuals in ~17,400 plans). Code was run on a server that operated using a 64-bit system with 8 gigabytes of installed random-access memory (RAM); times may vary depending on server specifications hardware.

were done with a different dataset that included actual plan deductibles; our paper does not comment on the validity of those methods for their specific data and usage. We implement this method in two stages: (1) impute deductibles for a subset of plans where they are easily identified and (2) use regression to predict deductibles for remaining plans. For the first stage, we sum deductible spending to the individual-year level and impute deductibles based on modal spending values.[11] In this stage, we are able to impute deductibles for 69% of plans. For the second stage, we create a set of covariates describing observed deductible spending and plan characteristics and collapse data from the individual to the plan level. Using the subset of plans with an imputed deductible, we regress the imputed deductible amounts on the set of covariates and use generated coefficients to predict deductibles for plans unable to be imputed in the first stage. The method, including detailed imputation rules for the first stage, covariates used in the second stage, and the regression specification, is more fully described in Supporting Information, Appendix 2.

2.3.3 | Modal deductible spending (mode method)

The logic of this method and the following one is that enrollees who meet their deductible will have observable deductible spending clustered at the administratively set deductible level and these clusters can be seen as modal lumps in each plan's overall deductible spending distribution. To implement this method—the simplest of the four tested—we identify the highest modal nonzero deductible spending amount among enrollees in a plan and apply it to all enrollees in that plan.

2.3.4 | 80th percentile of deductible spending (80th percentile method)

This method is based on the method used in Rabideau et al.[13] We begin with enrollee-month level data, where deductible spending is summed across the month. First, by enrollee, we track month-over-month deductible spending and identify enrollees whose total spending is increasing for multiple months without commensurate increases in their deductible spending; we consider these instances of an enrollee meeting their annual deductible limit and flag them. Then, keeping only these enrollees assumed to have met their deductibles, we collapse the data to the plan level and, for each plan, impute the 80th percentile of annual deductible spending as the plan deductible.

2.4 | Analysis

Our analytic dataset includes, at the plan level, actual deductibles for each plan and imputed deductibles from each method described above. We descriptively compare distributions of actual versus imputed deductibles for each method with both scatterplots and statistics. We compute the sensitivity, specificity, and positive/negative predictive value (PPV/NPV) of each method for classifying enrollees into high- versus low-deductible plans. Sensitivity and specificity measure the proportion of high- and low-deductible plans that will be identified through each imputation method, respectively. PPV and NPV measure the proportion of imputed deductibles correctly classified; PPV can be interpreted as the probability that a plan classified as an HDHP through imputation is actually an HDHP. This study was approved by the Johns Hopkins University Institutional Review Board. All analyses were done in Stata version 17.0 MP; exact code is in supplemental content.

3 | RESULTS

Our analytic dataset includes 2,055,822 individuals in 17,425 unique plans. The actual deductible in our data ranges from $0 to $5500, with a mean of $1847 and a median of $1500 (Supporting Information, Appendix Table 1). The most common category of deductible is $500–999, though 59% of plans have a deductible higher than $1500 (Table 2). Compared with a large national sample of employer plans, our data have a similar number of plans with deductibles above $3500 and at the modal deductible level ($500 and $999), but fewer plans with an individual of deductible <$500 (Supporting Information, Appendix Table 2).

Means of predicted deductibles range from $1452 to $2286 (Supporting Information, Appendix Table 1). All methods had a predicted minimum of $0, which was bottom coded in the regression-based methods. Predicted plan maximums varied from $2830 to $12,682. Histograms show the distribution of deductibles varies by imputation method (Supporting Information, Appendix Figures 1–5). The regression-based methods show a smoother distribution than the true deductible, reflecting that imputation using these methods is less likely round numbers.

The sensitivity of each method for correctly classifying HDHP versus non-HDHP ranged from 0.79 to 0.93, with the Regress on Spending performing worst and the mode method performing best (Table 2). Specificity across all methods was moderate to high (0.83–0.92). PPV was high for all methods (0.87–0.95). However, NPV is higher for the mode method (0.91) and 80th percentile method (0.86) compared with the regression-based methods, implying that HDHPs can be misclassified as low deductible plans using these methods. Stratifying by the number of enrollees in a plan shows that limiting to larger plans improves precision along most measures (Supporting Information, Appendix Table 3). Varying the percentile threshold in the 80th percentile method does not substantially change results (Supporting Information, Appendix Table 4).

Scatterplots of true against predicted deductible show that all methods have a high concordance with actual deductibles at low deductible levels but that methods using regression underpredict at higher levels of the true deductible (Figure 1). Graphically, the Mode and 80th percentile methods appear to adhere most closely to actual deductibles, though overpredict for most deductible levels. Figure 1

TABLE 2 Categorical deductible distributions and predictive statistics.

Category	True deductible	Regress on spending method	Regress on imputed deductibles method	Mode method	80th percentile method
<$500	12.06	2.04	7.51	12.92	14.10
$500–999	17.01	15.67	22.24	17.72	18.46
$1000–1499	12.08	39.74	24.60	11.83	12.48
$1500–1999	14.09	26.32	14.55	10.99	10.09
$2000–2499	10.83	13.31	6.49	8.14	8.53
$2500–2999	9.64	2.91	11.79	8.44	9.73
$3000–3499	14.12	0	7.44	10.67	10.11
$3500–3999	1.37	0	2.72	3.02	4.16
$4000–4499	1.97	0	2.68	3.11	3.27
$4500–4999	0.24	0	0	1.81	1.79
≥$5000	6.59	0	0	11.34	7.27
Sensitivity		0.79	0.80	0.93	0.90
Specificity		0.89	0.83	0.92	0.89
Positive predictive value (PPV)		0.91	0.87	0.95	0.92
Negative predictive value (NPV)		0.74	0.74	0.91	0.86
Number of plans	17,425	17,425	17,373	17,373	15,884

Note: Categorical distributions describe the percent of plans in each deductible category for each method. For predictive statistics, HDHP is defined as a deductible ≥ $1350, per the 2017 Internal Revenue Service minimum health savings account eligible deductible level. Sensitivity is the ratio of plans correctly classified as HDHPs with each imputation method compared with the total number of HDHPs, defined with actual deductible levels. Specificity is the ratio of plans correctly classified as low deductible (<$1350) with each imputation method over the total number of low deductible plans. PPV is defined as the ratio of imputed deductibles for each method that are correctly classified as HDHPs to the total number that are predicted to be HDHPs. NPV defined as the ratio of imputed deductibles that are correctly classified as low deductible plans per to the total number of those predicted to be low deductible plans. The regress on imputed deductibles method and mode method identified deductibles in plans only when there was positive total spending; 52 plans had no spending and thus were excluded from these imputation methods. The 80th percentile method used only plans in which we could determine that at least 1 enrollee had met their deductible.

shows that predicted deductibles are positively correlated with actual deductibles for all methods, which is also evident in the formal correlations (0.66–0.72; Supporting Information, Appendix Table 1). A positive correlation implies that all imputation methods properly order deductible size in plans relative to each other, even when actual levels of the deductibles are incorrectly imputed.

The overall sensitivity across the categorical distribution of deductible levels is low to moderate for all methods (Supporting Information, Appendix Table 5) and, consistent with Figure 1, imputations perform better at lower deductible levels. The regress on spending method has the lowest sensitivity both for predicting the correct deductible category and the continuous level of a deductible within $250 (0.25 and 0.23, respectively, Supporting Information, Appendix Table 5). The mode method performs best; 72% of plans are correctly classified into each category and 69% of plans have an imputed deductible within $250 of the actual deductible. For this method, limiting imputation to groups with more than 50 enrollees improved sensitivity to 85% of plans correctly classified by category and reduced the average difference between the imputed and actual deductible from $700 to $496 (Supporting Information, Appendix Table 6).

4 | DISCUSSION

We found that imputing the nonzero mode for each plan, which requires little coding or computation time, performed as well or better than more complex methods of deductible imputation. This method performed well, even in small groups, for classifying low- and high-deductible plans; it performed well at classifying more granular deductible levels for groups with more than 50 enrollees. Researchers who previously used more computationally intense methods may want to switch, for both simplicity and accuracy.

Additionally, all imputation methods demonstrated a high correlation with true deductibles, indicating that they can correctly order plans in terms of deductible levels. This is particularly useful for researchers who are interested in understanding the relative differences in cost-sharing structures across various health plans.

It is important to note that no imputation method perfectly matched the exact deductibles, nor is able to capture nuances of cost sharing in each plan. The method of imputing nonzero modes, while performing better than other methods, still only predicted a deductible within $250 of the actual deductible 69% of the time. Our findings suggest that, while imputation methods can provide a reasonably

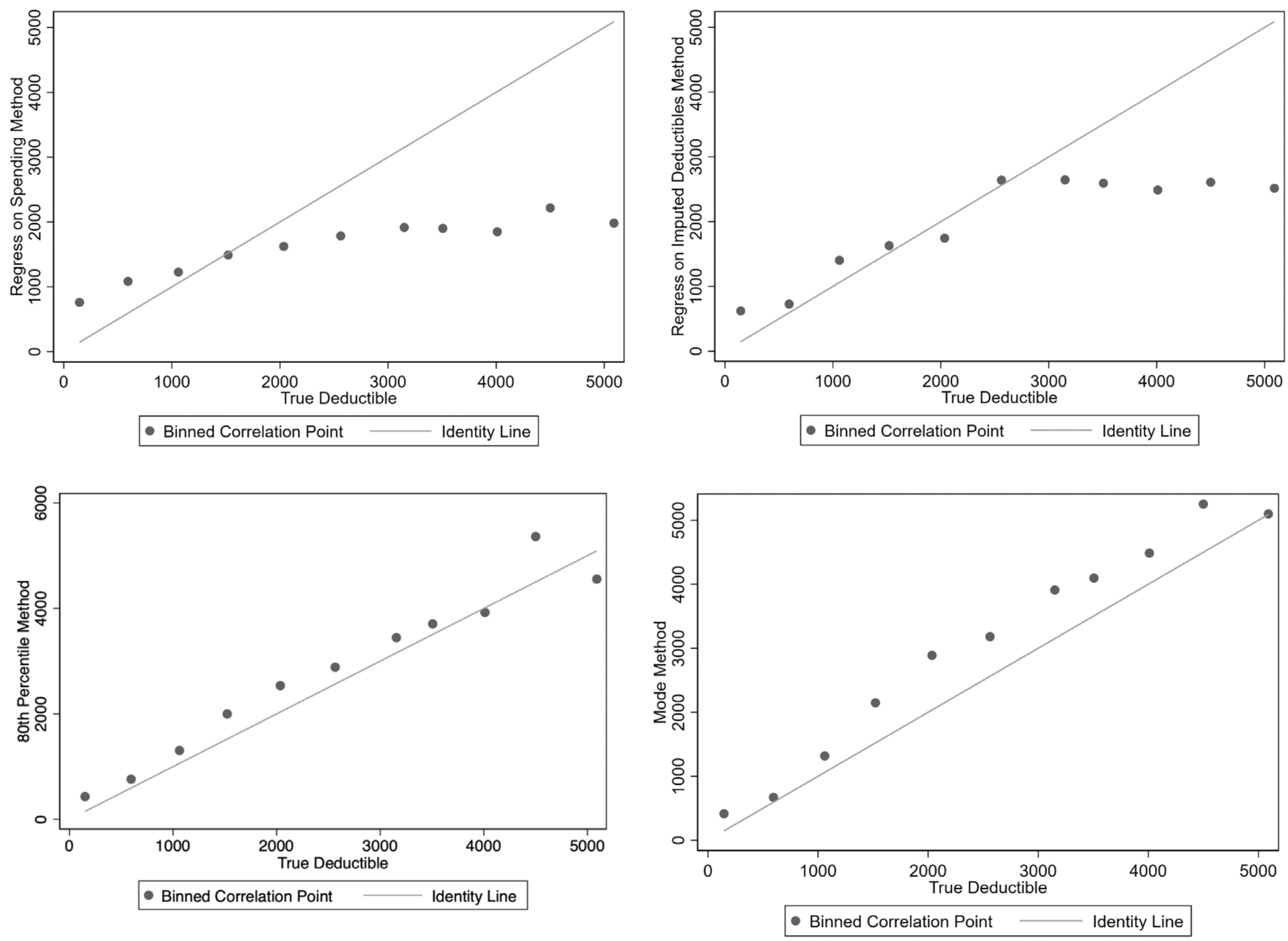

FIGURE 1 Scatterplots of imputed and true deductibles. Each panel in this figure shows imputed deductibles, binned in deciles, for each imputation method plotted against the true deductible. Points on the identity line signify 1:1 match between predicted and actual deductibles. Points under each identity line signify under-prediction and points above the line signify overprediction relative to the true deductible.

good approximation of general deductibles, there is still room for improvement, particularly when it comes to predicting deductibles at higher levels. A solution to these limitations is to include plan structure variables in data releases, which would allow researchers to directly observe the effects of cost-sharing structures, including deductibles, on health and spending outcomes.

Our study has several limitations. We use data from a single health insurer that, while large, is not representative of all health insurance plans and variables may not translate. We made decisions about which variables to use based on external generalizability and tractability for each imputation method, though acknowledged our methods would be difficult with datasets that do not include basic variables denoting groups or coverage levels. Our methods are not validated for family deductibles, which may be structured differently than individual deductibles. Finally, we used published literature as well as our own experiences to choose methods in our imputations. It is possible we left out a valid method or valid variation on an above method. We hope our results will help to standardize methods used in this type of research so that studies can be better compared and evidence more easily synthesized.

ACKNOWLEDGMENTS

This project was partially supported by grant number R01DA044201 from the National Institute on Drug Abuse (NIDA). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIDA.

ORCID

Betsy Q. Cliff https://orcid.org/0000-0002-0120-7439
Mark K. Meiselbach https://orcid.org/0000-0003-3583-5588

REFERENCES

1. Claxton G, Rae M, Wager E, et al. *Employer Health Benefits: 2022 Annual Survey*. Kaiser Family Funcation; 2022.
2. Health Insurance Coverage of the Total Population. Kaiser Family Foundation. 2021. Accessed May 15, 2023. https://www.kff.org/other/state-indicator/total-population
3. Bundorf MK. Consumer-directed health plans: a review of the evidence. *J Risk Insur*. 2016;83(1):9-41.
4. Wharam JF, Ross-Degnan D, Rosenthal MB. The ACA and high-deductible insurance—strategies for sharpening a blunt instrument. *N Engl J Med*. 2013;369(16):1481-1484. doi:10.1056/NEJMp1309490

5. Cliff BQ, Krenz C, West BT, et al. Attitudes about consumer strategies among Americans in high-deductible health plans. *Med Care*. 2019; 57(3):187-193. doi:10.1097/MLR.0000000000001056
6. Haviland AM, Eisenberg MD, Mehrotra A, Huckfeldt PJ, Sood N. Do "consumer-directed" health plans bend the cost curve over time? *J Health Econ*. 2016;46:33-51.
7. Brot-Goldberg ZC, Chandra A, Handel BR, Kolstad JT. What does a deductible do? The impact of cost-sharing on health care prices, quantities, and spending dynamics*. *Q J Econ*. 2017;132(3):1261-1318. doi:10.1093/qje/qjx013
8. Agarwal R, Mazurenko O, Menachemi N. High-deductible health plans reduce health care cost and utilization, including use of needed preventive services. *Health Aff*. 2017;36(10):1762-1768. doi:10.1377/hlthaff.2017.0610
9. Buntin MB, Haviland AM, McDevitt R, Sood N. Healthcare spending and preventive care in high-deductible and consumer-directed health plans. *Am J Manag Care*. 2011;17(3):9.
10. Fronstin P, Roebuck MC. Health care spending after adopting a full-replacement, high-deductible health plan with a health savings account: a five-year study. *EBRI Issue Brief*. 2013;388:1–15.
11. Wharam JF, Zhang F, Wallace J, et al. Vulnerable and less vulnerable women in high-deductible health plans experienced delayed breast cancer care. *Health Aff*. 2019;38(3):408-415. doi:10.1377/hlthaff.2018.05026
12. Galbraith AA, Soumerai SB, Ross-Degnan D, Rosenthal MB, Gay C, Lieu TA. Delayed and forgone care for families with chronic conditions in high-deductible health plans. *J Gen Intern Med*. 2012;27(9): 1105-1111. doi:10.1007/s11606-011-1970-8
13. Rabideau B, Eisenberg MD, Reid R, Sood N. Effects of employer-offered high-deductible plans on low-value spending in the privately insured population. *J Health Econ*. 2021;76:102424.
14. Eisenberg MD, Haviland AM, Mehrotra A, Huckfeldt PJ, Sood N. The long term effects of "consumer-directed" health plans on preventive care use. *J Health Econ*. 2017;55:61-75.
15. Eisenberg MD, Kennedy-Hendricks A, Schilling C, et al. The impact of HDHPs on service use and spending for substance use disorders. *Am J Manag Care*. 2022;28(10):530-536.
16. Meiselbach MK, Kennedy-Hendricks A, Schilling C, et al. High deductible health plans and spending among families with a substance use disorder. *Drug Alcohol Depend*. 2022;241:109681.
17. Cliff BQ. Do high-deductible health plans affect price paid for childbirth? *Health Serv Res*. 2022;57(1):27-36.
18. Mazurenko O, Buntin MJ, Menachemi N. High-deductible health plans and prevention. *Annu Rev Public Health*. 2019;40:411-421.
19. Wallace PJ, Shah ND, Dennen T, Bleicher PA, Crown WH. Optum labs: building a novel node in the learning health care system. *Health Aff*. 2014;33(7):1187-1194.
20. Blewett LA, Mac Arthur NS, Campbell J. The future of state all-payer claims databases. *J Health Polit Policy Law*. 2023;48(1):93-115.
21. Wharam JF, Zhang F, Eggleston EM, Lu CY, Soumerai SB, Ross-Degnan D. Effect of high-deductible insurance on high-acuity outcomes in diabetes: a natural experiment for translation in diabetes (NEXT-D) study. *Diabetes Care*. 2018;41(5):940-948. doi:10.2337/dc17-1183
22. OptumLabs and OptumLabs Data Warehouse (OLDW). Eden Prairie, Minn. 2017.
23. Claxton G, Rae M, Long M, Damico A, Foster G, Whitmore H. *Employer Health Benefits: 2017 Annual Survey*. Kaiser Family Foundation/Health Research & Educational Trust; 2017. https://www.kff.org/health-costs/report/2017-employer-health-benefits-survey/
24. Cliff BQ, Li M. High-deductible health plans: effects across income. Presented at: Health Policy Research Seminar; March 11, 2022; Chicago, IL. https://publichealth.uic.edu/health-policy-administration/health-policy-research-seminar/
25. Garabedian LF, Zhang F, LeCates R, Wallace J, Ross-Degnan D, Wharam JF. Trends in high deductible health plan enrolment and spending among commercially insured members with and without chronic conditions: a natural experiment for translation in diabetes (NEXT-D2) study. *BMJ Open*. 2021;11(9):e044198.
26. Wharam JF, Zhang F, Eggleston EM, Lu CY, Soumerai S, Ross-Degnan D. Diabetes outpatient care and acute complications before and after high-deductible insurance enrollment: a natural experiment for translation in diabetes (NEXT-D) study. *JAMA Intern Med*. 2017; 177(3):358-368.
27. Wharam JF, Wallace J, Argetsinger S, et al. Diabetes microvascular disease diagnosis and treatment after high-deductible health plan enrollment. *Diabetes Care*. 2022;45(8):1754-1761. doi:10.2337/dc21-0407

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Cliff BQ, Eddelbuettel JCP, Meiselbach MK, Eisenberg MD. Deductible imputation in administrative medical claims datasets. *Health Serv Res*. 2024; 1-7. doi:10.1111/1475-6773.14278