Appendix 1: Deductible Level Imputation

To determine employer deductible levels, we used a benefits type variable that we had for most smaller employers (with approximately 100 or fewer employees). For larger employers, we took advantage of the fact that health insurance claims data are the most accurate source for assessing out-of-pocket obligations among patients who utilize health services. Our claims data contained an in-network/out-of-network individual deductible payment field. For patients who use expensive or frequent services, the sum of their yearly deductible payments adds up to clearly identifiable exact amounts such as $500.00, $1000.00, $2000.00, etc. When even several members have these same amounts, it provides strong evidence that the employer offered such an annual deductible level. It is also possible to detect employers that offer choices of deductible levels when multiple employees have deductibles at two or more levels, such as 20 employees with an exact annual amount of $1000.00 and 12 employees with $500.00. For employer accounts with at least 10 enrollees, we therefore summed each member's in-network (individual-level) deductible payments and number of claims over the enrollment year and assessed other key characteristics such as percentage with Health Savings Accounts. We randomly selected half of the employer account data set that contained both our calculated employer characteristics (independent variables, below) and actual annual deductible levels from the benefits table (dependent variable, after categorization; below). We then used a multinomial logistic model that predicted the 4-level outcome of individual-level deductible ≤$500/$501-$999/$1000-$2499/≥$2500 (again, dependent variable) based on multiple aggregate employer characteristics (independent variables) such as the percentage with Health Savings Accounts and Health Reimbursement Arrangements, the deductible payment per employer in the 75 percentile of payments, the percentage of employees reaching exact deductible levels or with deductible payments but not reaching an exact deductible level, the employer account size, the percentage of enrollees per account with summed whole dollar annual deductible amounts (from claims data) between $0 to <$100, ≥$100 to ≤$500, >$500 to <$1000, ≥$1000 to <$2500, ≥$2500, etc.

The statistical model was as follows:

$$\text{Logit}(\Pr = Y_i) = \beta_0 + \sum \beta_k X_{ki}$$

Where:

Y_i = dependent variable (4-level deductible category)
X_{ki}= k^{th} characteristics for i^{th} employer
$\beta_{0=}$ intercept
$\beta_{k=}$ coefficient for k^{th} characteristic

The SAS code we used to implement this model was:

```
proc logistic data=csn_impute_PLUS_to_be_imputed descending;
class
        d_wusd1perc_0_100_cat d_wusd1perc_100_500_cat d_wusd1perc_500_1000_cat
        d_wusd1perc_1000_2500_cat d_wusd1perc_ge2500_cat
        d_wusd2perc_0_100_cat d_wusd2perc_100_500_cat d_wusd2perc_500_1000_cat
        d_wusd2perc_1000_2500_cat d_wusd2perc_ge2500_cat
        d_wusd3perc_0_100_cat d_wusd3perc_100_500_cat d_wusd3perc_500_1000_cat
        d_wusd3perc_1000_2500_cat d_wusd3perc_ge2500_cat
        d_wusd4perc_0_100_cat d_wusd4perc_100_500_cat d_wusd4perc_500_1000_cat
        d_wusd4perc_1000_2500_cat d_wusd4perc_ge2500_cat;

model real_dduct_cat =
        pyr sampletot hsa_cnt_over_total cdhp_cnt_over_total perc_grp2 perc_grp3 perc_grp4
        perc_grp5 d_wusd1perc_0_100_cat d_wusd1perc_100_500_cat d_wusd1perc_500_1000_cat
        d_wusd1perc_1000_2500_cat d_wusd1perc_ge2500_cat d_wusd2perc_0_100_cat
        d_wusd2perc_100_500_cat d_wusd2perc_500_1000_cat d_wusd2perc_1000_2500_cat
        d_wusd2perc_ge2500_cat d_wusd3perc_0_100_cat d_wusd3perc_100_500_cat
        d_wusd3perc_500_1000_cat d_wusd3perc_1000_2500_cat d_wusd3perc_ge2500_cat
```

```
        d_wusd4perc_0_100_cat d_wusd4perc_100_500_cat d_wusd4perc_500_1000_cat
        d_wusd4perc_1000_2500_cat d_wusd4perc_ge2500_cat
        p75_0_100_dduct p75_100_500_dduct p75_500_1000_dduct p75_1000_2500_dduct
        p75_gt2500_dduct
output out=prob_of_dduct_cat&IOS. p=p_dduct_cat predprobs=i;
run;
```

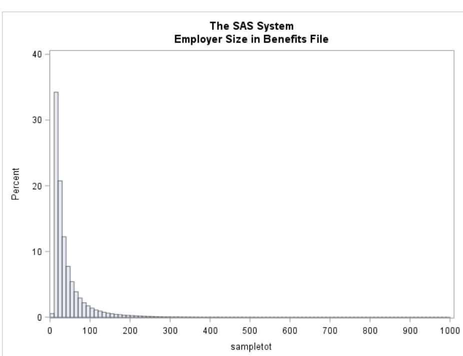
Further explanation of this code is below. Note that all values described are calculated over the benefit year per employer account, and a given employer account could be present for multiple years.

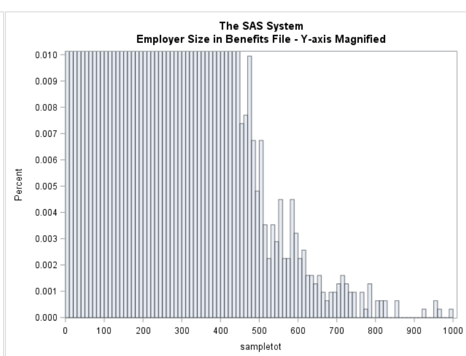
- csn_impute_PLUS_to_be_imputed = name of dataset that contains, at the employer account and benefit year level, accounts with missing deductible levels as well as a random half of the accounts that have actual deductible levels. The other random half is also present in the dataset but with actual deductible levels “hidden” so that they can later be used to validate the predictive algorithm.
- real_dduct_cat = dependent variable; category of actual deductible level from the gold standard source (<=$500, $500-$999, $1000-$2499, ≥$2500)
- pyr = benefit year of account’s information and tied to the calendar year. An employer could have multiple benefit years represented in separate records per account-benefit year.
- sampletot = total enrollees per account during the benefit year
- hsa_cnt_over_total = percent of members per account listed as having a health savings account
- cdhp_cnt_over_total = percent of members per account listed as having a health savings account or health reimbursement arrangement
- perc_grp1. Percentage of enrollees per employer-year who have claims but $0 deductible amounts for all annual claims.
- perc_grp2. Percentage of enrollees per employer-year who have reached their annual deductible, evidenced by the sum of their deductible payments ending in $*0.00. Members must have at least one month after the month of the $*0.00 summation where the deductible field is blank, and all subsequent months must have blank deductible fields, indicating that the member reached his or her annual deductible amount.
- perc_grp3. Percentage of enrollees per employer-year who have an annual deductible amount that does not end in $*0.00.
- perc_grp4. Percentage of enrollees per employer-year who have enrollment during the benefit year where all months show no evidence of utilization (no health insurance claims).
- perc_grp5. Percentage of enrollees per employer-year who might have reached their deductible, as evidenced by having the last month of enrollment of the benefit year with a summed annual deductible amount that ends in $*0.00.
- d_wusd1perc_0_100_cat, d_wusd1perc_100_500_cat, d_wusd1perc_500_1000_cat, d_wusd1perc_1000_2500_cat d_wusd1perc_ge2500_cat. Category of percentage of enrollees with an employer’s most common whole number annual individual deductible payment total (e.g. dollar amount ending in 0.00) per employee that is $0 to <$100, ≥$100 to ≤$500, >$500 to <$1000, ≥$1000 to <$2500, and ≥$2500, respectively.
- d_wusd2perc_0_100_cat, d_wusd2perc_100_500_cat, d_wusd2perc_500_1000_cat, d_wusd2perc_1000_2500_cat d_wusd2perc_ge2500_cat. Category of percentage of enrollees with an employer’s second most common whole number annual individual deductible payment total (e.g. dollar amount ending in 0.00) per employee that is $0 to <$100, ≥$100 to ≤$500, >$500 to <$1000, ≥$1000 to <$2500, and ≥$2500, respectively.
- d_wusd3perc_0_100_cat, d_wusd3perc_100_500_cat, d_wusd3perc_500_1000_cat, d_wusd3perc_1000_2500_cat d_wusd3perc_ge2500_cat. Category of percentage of enrollees with an employer’s third most common whole number annual individual deductible payment total (e.g. dollar amount ending in 0.00) per employee that is $0 to <$100, ≥$100 to ≤$500, >$500 to <$1000, ≥$1000 to <$2500, and ≥$2500, respectively.

- d_wusd4perc_0_100_cat, d_wusd4perc_100_500_cat, d_wusd4perc_500_1000_cat, d_wusd4perc_1000_2500_cat d_wusd4perc_ge2500_cat. Category of percentage of enrollees with an employer's fourth most common whole number annual individual deductible payment total (e.g. dollar amount ending in 0.00) per employee that is \$0 to <\$100, ≥\$100 to ≤\$500, >\$500 to <\$1000, ≥\$1000 to <\$2500, and ≥\$2500, respectively.
- p75_0_100_dduct p75_100_500_dduct p75_500_1000_dduct p75_1000_2500_dduct p75_gt2500_dduct. Category of 75th percentile of deductible payments per employer benefit year, categorized as \$0 to <\$100, ≥\$100 to ≤\$500, >\$500 to <\$1000, ≥\$1000 to <\$2500, and ≥\$2500, respectively.

This predictive model outputs the probability that employers had deductibles in the four categories (summing to 1.0) and we assigned the employer to the level that had the highest probability. We overwrote this assignment with the most common whole number deductible amount per year if it was not zero, and with the second most common whole number deductible amount if the most common amount was zero and at least 10 members had the value of the second most common whole number deductible amount. If an employer had members with both enrollment and evidence of utilization, but never had any amounts in the deductible field, we assigned that employer to <\$500 deductible level. If an employer had only members that reached a whole number annual deductible amount such as \$1000.00 or \$2000.00, we assigned the most common deductible amount as the employer's deductible if that amount was greater than or equal to \$1000 and to the 95% percentile value if that number was less than \$1000. If at least 99% of employees had Health Savings Accounts or Health Reimbursement Arrangements, we also overwrote any previous assignment to classify the employer as a high-deductible employer. We assigned employers to have a choice between deductible levels of \$1000 to \$2499 and ≥\$2500 when both were common and one accounted for at least 85% of \$1000-\$2499 or ≥\$2500 deductible levels reached per employer. If we detected employers that had sufficient enrollees with whole number deductible levels both above and below \$1000 (e.g. \$250.00 and \$1500.00), we assigned the employers' category as "choice," applying a similar 85% rule. Finally, for any employer that had gold standard deductible level information in our benefits file, we overwrote any previous imputed deductible level.

Our file that contains actual deductible amounts per employer covers the "small employer" segment of the insurer's business, a segment that generally includes employers with fewer than 100 or so enrollees. However, it does include a modest number of employers with more than 100 enrollees, even up to approximately 1000 enrollees. The histograms below, where the x-axis represents employer size and the y-axis shows the percentage of employers that are that size, demonstrate the distribution of employer sizes. The second plot "magnifies" the y-axis to demonstrate the smaller number of large employers.





To demonstrate the robustness of our imputation algorithm, and its predictive value as employer size increases (given that we do not have benefits information on most large employers), we took advantage of the fact that although this file mostly covers employers with 100 enrollees or fewer, there is some overlap with larger employers (i.e., those with ~100 to 1000 enrollees). A random half of our imputation sample had the actual deductible levels of employers of all sizes "hidden" from the imputation. Thus, this random half included a modest number of employers with 75 to 1000 enrollees. We tested the sensitivity and specificity of the imputation in this overlap zone, categorizing employer sizes as 75-100, 101-400, 401-700, and 701-1000 enrollees (Exhibit 1). At employers with 75-100 enrollees, we found sensitivity of 95.4% and specificity of

98.3% (Exhibit 1a). Sensitivity and specificity increased across employer size to 100%, and Exhibits 1b-1d display these for employers of sizes 101-400, 401-700, and 701-1000.

We used an employer ID and an algorithm that determined linked employer subaccounts to identify an employer's subaccounts per benefit year, and removed benefit years when employers offered both low and high deductible levels.

Rationale for High-Deductible Cutoffs: When Health Savings Account-eligible high-deductible health plans came to market in 2005-2006, the Internal Revenue Service set the minimum deductible level for qualifying high-deductible health plans at $1050 (which could be adjusted upward for inflation annually). The range of this minimum deductible during our study period was $1050-$1250. For these reasons, we defined high-deductible health plans as annual individual deductibles of at least $1000 (otherwise some health savings account plans would be excluded). In addition, choosing this cutoff (as opposed to, e.g., $2000) improves the sensitivity and specificity of the imputation because this is common deductible level and more enrollees per employer meet this threshold. This cutoff is also a "real-world" deductible minimum that allows the most generalizable results. It should also be noted that $1000 was the *minimum* annual deductible level we included and not the mean deductible level. We cannot precisely calculate the mean deductible level of the high-deductible health plan group, but we estimate, using the most common non-zero deductible levels per employer account, an approximate mean deductible of $1900. We defined traditional plans as having deductible levels of ≤$500 after determining that a threshold of ≤$250 would lead to an inadequate sample size for the control group. Again, the mean deductible level of the control group members would be lower than $500.

Appendix Exhibit 1. Validation of Deductible Imputation Algorithm, Stratified by Employer Size

Exhibit 1a. Validation of deductible imputation algorithm, using employer accounts of size 75-100 enrollees.

	Gold Standard[a]=high-deductible (n)	**Gold Standard=low-deductible (n)**
We imputed high-deductible	882,588	24,786
We imputed low-deductible	15,612	511,770

	High-deductible	**Low-deductible**
Sensitivity	98.3%	95.4%
Specificity	95.4%	98.3%

[a]Gold standard was a benefits variable specific to each employer derived from a benefits table and obtained from the health insurer via the data vendor.

Exhibit 1b. Validation of deductible imputation algorithm, using employer accounts of size 101-400 enrollees.

	Gold Standard[a]=high-deductible (n)	**Gold Standard=low-deductible (n)**
We imputed high-deductible	1,998,885	42,655
We imputed low-deductible	20,302	1,748,826

	High-deductible	**Low-deductible**
Sensitivity	99.0%	97.6%
Specificity	97.6%	99.0%

[a]Gold standard was a benefits variable specific to each employer derived from a benefits table and obtained from the health insurer via the data vendor.

Exhibit 1c. Validation of deductible imputation algorithm, using employer accounts of size 401-700 enrollees.

	Gold Standard[a]=high-deductible (n)	**Gold Standard=low-deductible (n)**
We imputed high-deductible	83,393	485
We imputed low-deductible	2,017	122,983

	High-deductible	**Low-deductible**
Sensitivity	97.6%	99.6%
Specificity	99.6%	97.6%

[a]Gold standard was a benefits variable specific to each employer derived from a benefits table and obtained from the health insurer via the data vendor.

Exhibit 1d. Validation of deductible imputation algorithm, using employer accounts of size 701-1000 enrollees.

	Gold Standard[a]=high-deductible (n)	**Gold Standard=low-deductible (n)**
We imputed high-deductible	9950	0
We imputed low-deductible	0	19,664

	High-deductible	**Low-deductible**
Sensitivity	100.0%	100.0%
Specificity	100.0%	100.0%

[a]Gold standard was a benefits variable specific to each employer derived from a benefits table and obtained from the health insurer via the data vendor.

Appendix 2: Definition of Covariates

Comorbidity score: We used version 11.1 of the Johns Hopkins ACG® System[1,2] to calculate members' baseline period morbidity score. The algorithm uses age, gender, and ICD-9-CM codes to calculate a morbidity score and the average of the reference population is 1.0.[2] Researchers have validated the index against premature mortality.[1]

Demographic characteristics: To derive proxy demographic measures, the data vendor linked members' most recent residential street addresses to their 2010 US Census tract.[3] Census-based measures of socioeconomic status have been validated [4,5] and used in multiple studies to examine the impact of policy changes on disadvantaged populations.[6-8] Using 2008-2012 American Community Survey[9] census tract-level data and validated cut-points,[4,5] we created categories that defined residence in neighborhoods with below-poverty levels of <5%, 5%-9.9%, 10%-19.9%, and ≥20%. Similarly, we defined categories of residence in neighborhoods with below-high-school education levels of <15%, 15%-24.9%, 25%-39.9%, ≥40%.[4,5] We classified members as from predominantly white, black, or Hispanic neighborhoods if they lived in a census tract with at least 75% of members of the respective race/ethnicity. We then applied a superseding ethnicity assignment using flags created by the E-Tech system (Ethnic Technologies), which analyzes full names and geographic locations of individuals.[10] We classified remaining members as from mixed race/ethnicity neighborhoods. This validated approach of combining surname analysis and census data has positive and negative predictive values of approximately 80 and 90 percent, respectively.[11]

References:

1. Reid RJ, Roos NP, MacWilliam L, Frohlich N, Black C. Assessing population health care need using a claims-based ACG morbidity measure: a validation analysis in the Province of Manitoba. *Health services research.* 2002;37(5):1345-1364.
2. ACG. The Johns Hopkins ACG System. https://www.hopkinsacg.org/advantage/. Accessed 2 June, 2018.
3. USCensus. Census 2000 Gateway. 2000; https://www.census.gov/main/www/cen2000.html. Accessed 2 June, 2018.
4. Krieger N. Overcoming the absence of socioeconomic data in medical records: validation and application of a census-based methodology. *American Journal of Public Health.* 1992;82(5):703-710.
5. Krieger N, Chen JT, Waterman PD, Rehkopf DH, Subramanian SV. Race/ethnicity, gender, and monitoring socioeconomic gradients in health: a comparison of area-based socioeconomic measures--the public health disparities geocoding project. *American Journal of Public Health.* 2003;93(10):1655-1671.
6. Trivedi AN, Zaslavsky AM, Schneider EC, Ayanian JZ. Relationship between quality of care and racial disparities in Medicare health plans. *Jama.* 2006;296(16):1998-2004.
7. Trivedi AN, Rakowski W, Ayanian JZ. Effect of cost sharing on screening mammography in medicare health plans. *The New England journal of medicine.* 2008;358(4):375-383.
8. Selby JV, Fireman BH, Swain BE. Effect of a copayment on use of the emergency department in a health maintenance organization. *New England Journal of Medicine.* 1996;334(10):635-641.
9. American Community Survey. https://www.census.gov/programs-surveys/acs/. Accessed 28 September, 2017.
10. Ethnic Technologies. Frequently Asked Questions. www.ethnictechnologies.com/faq. Accessed 24 October, 2018.
11. Fiscella K, Fremont AM. Use of geocoding and surname analysis to estimate race and ethnicity. *Health services research.* 2006;41(4 Pt 1):1482-1500.