Supplementary Online Content

Henderson J, Bouck Z, Holleman R, et al. Comparison of payment changes and Choosing Wisely recommendations for use of low-value laboratory tests in the United States and Canada. *JAMA Intern Med*. Published online February 10, 2020. doi:10.1001/jamainternmed.2019.7143

This supplementary material has been provided by the authors to give readers additional information about their work.

eTable 1. Frequency of Exclusions from Vitamin D Screening Measure

Number of unique beneficiaries with one or more primary care visits excluded for each of the five most common reasons as determined by median rank. Beneficiaries were included at most once within each group of exclusions but may have contributed to more than one exclusion group.

Reason for Exclusion (rank)	CA-Ontario	US-Veterans	US-Commercial
Renal disease (1)	1,268,607 (1)	1,259,203 (1)	3,482,814 (4)
Other (2)	836,761 (2)	493,621 (3)	6,065,485 (1)
Vitamin D Deficiency (2)	3,184 (21)	776,578 (2)	5,605,593 (2)
Bone Disorders (4)	359,462 (6)	416,485 (4)	3,854,325 (3)
Obesity (5)	553,616 (4)	317,687 (5)	1,952,462 (5)

eTable 2. Diagnostic and Procedural Coding Systems

This table lists the procedure and diagnostic medical coding systems used to define these measures within each jurisdiction. Abbreviations: (CCI) Canadian Classification of Health Interventions, (CIHI) Canadian Institute for Health Information, (CPT4) Current Procedural Terminology, Fourth Edition, (LOINC) Logical Observation Identifiers Names and Codes, (ICD-9 and ICD-10) 9th and 10th revisions of the International Statistical Classification of Diseases and Related Health Problems.

	Healthcare Jurisdiction		
	Ontario	US Veterans	US ESI
Vitamin D and T3 lab tests	CCI, CIHI Fee Codes	LOINC	CPT4
Primary Care Visits	CCI, CIHI Fee Codes	CPT4	CPT4
Exclusions for Vitamin D Measure	ICD-10	ICD-9	ICD-9
Hypothyroidism Diagnoses	ICD-10	ICD-9	ICD-9

eTable 3. Demographic Information for Unique Beneficiaries

Number of unique beneficiaries examined for low-value Vitamin D screening or person-months at-risk for low value T3 testing.

	Vitamin D			T3		
	CA-Ontario	US-Veterans	US-Commercial	CA-Ontario	US-Veterans	US-Commercial
Study Population, n	9,571,350	3,683,660	37,569,125	344,948	204,500	2,849,865
Age groups, n (%)						
18-34	3,929,149 (41.1%)	684,384 (18.6%)	12,844,035 (34.2%)	72,936 (21.1%)	11,933 (5.8%)	353,021 (12.4%)
35-44	1,931,266 (20.2%)	492,387 (13.4%)	8,047,287 (21.4%)	81,229 (23.5%)	20,974 (10.3%)	528,842 (18.6%)
45-64	3,710,935 (38.8%)	2,506,889 (68.1%)	16,677,803 (44.4%)	190,783 (55.3%)	171,593 (83.9%)	1,968,002 (69.1%)
Sex, n (%)						
Female	4,949,781 (51.7%)	383,017 (10.4%)	20,541,324 (54.7%)	280,223 (81.2%)	46,119 (22.6%)	2,304,112 (80.8%)
Male	4,621,569 (48.3%)	3,300,643 (89.6%)	17,027,801 (45.3%)	64,725 (18.8%)	158,381 (77.4%)	545,753 (19.2%)
Region, n (%)						
US: Midwest, CA: East	1,245,249 (13.0%)	768,945 (20.9%)	8,570,248 (22.8%)	38,030 (11.0%)	44,767 (21.9%)	634,063 (22.2%)
US: Northeast CA: North	847,407 (8.9%)	431,712 (11.7%)	7,637,238 (20.3%)	23,637 (6.9%)	21,841 (10.7%)	606,372 (21.3%)
US: South CA: Central	4,907,966 (51.3%)	1,680,017 (45.6%)	14,294,261 (38.0%)	204,182 (59.2%)	92,115 (45.0%)	1,079,721 (37.9%)
US: West CA: West	2,570,727 (26.9%)	802,986 (21.8%)	7,067,378 (18.8%)	79,099 (22.9%)	45,777 (22.4%)	529,709 (18.6%)

eTable 4. Estimated Rates and Trends for Low-value Vitamin D and T3 Testing Before and After Choosing Wisely Recommendations

This table compares utilization and estimated trends in low-value testing before and after the release of CW recommendations across three jurisdictions. Abbreviations: *AAPC = Average Annual Percent Change (relative risk). AME = Average Marginal Effect (absolute risk).*

	CA-Ontario		**US-Veterans**		**US-Commercial**	
	Before	*After*	*Before*	*After*	*Before*	*After*
Vitamin D						
Start End	Dec 2010 Oct 2014	Nov 2014 Jun 2015	Jan 2010 Feb 2013	Mar 2013 Jun 2015	Jan 2010 Feb 2013	Mar 2013 Jun 2015
Tests/100	0.44	0.66	3.13	4.12	2.38	2.47
AAPC, % (95% CI)	29.7[a] (28.8, 30.7)	24.9 (22.7, 27.0)	16.5 (15.4, 17.6)	7.7 (6.8, 8.5)	5.6 (4.9, 6.2)	-2.8 (-3.3, -2.2)
AME Time (95% CI)	0.08 (0.07, 0.08)	0.10 (0.07, 0.13)	0.54 (0.51, 0.58)	0.20 (0.16, 0.25)	0.17 (0.16, 0.19)	-0.12 (-0.14, -0.10)
T3						
Start End	Jan 2012 Oct 2014	Nov 2014 Jun 2015	Jan 2012 Oct 2013	Nov 2013 Jun 2015	Jan 2012 Oct 2013	Nov 2013 Jun 2015
Tests/100	3.00	2.96	0.68	0.73	2.59	2.68
AAPC, % (95% CI)	-0.6 (-1.3, 0.0)	0.3 (-1.0, 1.6)	2.0 (-0.8, 4.9)	5.0 (3.6, 6.5)	-0.4 (-1.1, 0.3)	3.2 (2.8, 3.7)
AME Time (95% CI)	-0.02 (-0.04, -0.00)	0.32 (0.13, 0.51)	0.01 (-0.01, 0.03)	0.08 (0.06, 0.09)	0.00 (-0.02, 0.02)	0.11 (0.09, 0.13)

Notes: [a]*AAPC for Ontario before the release of the CW recommendation was computed using year-over-year ratios beginning in December 2011.*

eMethods.

Regression Models

As described in the main text of the paper, we used quasi-Poisson GEE models (with log link) at the strata-month level to examine trends in rates of low-value utilization of Vitamin D and T3 level tests. The response in each model was the number of low-value blood tests performed with and offset for the number of beneficiaries eligible for the specific test. The covariates in all of these models included: main effects and all interactions for stratification variables (three age groups, two sexes, four geographic regions, and, for US-commercial, a binary employer relation), continuous time terms for the pre- and post-recommendation trends, and seasonal indicators for calendar months. We also included interactions between continuous time terms and stratification variables other than employer relation. In addition, we added trends and associated interactions for 2010 post-hoc to the two US jurisdictions to improve fit and accommodate the more rapid increases in low-value Vitamin D level testing for 2010 relative to subsequent years of the pre-recommendation period observed in scatterplots.

For the Vitamin D model in CA-Ontario, we also included a third continuous time term for January-November 2010 and a December 2010 level-change to accommodate the policy change delisting low-value Vitamin D for reimbursement – both of which were allowed to interact with all demographic stratification variables. For all jurisdictions, in the models for Vitamin D we did not include level-changes for the post-recommendation periods opting instead for linear splines. However, owing to the lack of strong pre-recommendations trend in the T3 models, we did include binary indicators for level-changes and again interacted these with all stratification variables.

We specified clusters in the GEE models to be the (24 or 48) strata determined by demographic variables. We considered both independence and auto-regressive order one working covariance structures within clusters and selected to report the results using the independence working covariance based on a comparison of the uncorrected quasi information criterion (QICu). In light of this, it is important to note that a feature of GEE models is that the resulting estimates are robust to misspecification of the working covariance.

Post-stratification Weighting for Population Estimates

Estimates for the US Commercial population were derived from the Marketscan research database, representing a substantial fraction of insured individuals with employer sponsored insurance in the US during these years. Individuals in this database were eligible for study inclusion during those months for which they had sufficient periods of continuous coverage, defined a $\geq$24 covered days in a month. For the T3 analysis, individuals were eligible for inclusion in any month they had insurance coverage for 24 or more days. Individuals contribute to the denominator in the T3 analysis during months of eligibility during which they also met the criteria for an established hypothyroidism diagnosis. For the Vitamin D analysis, individuals

were eligible to contribute a primary care visit in any month during the study period for which they met this criteria of 24 or more days of coverage in: (1) the month of eligibility, (2) the prior 12 months allowing for sufficient look-back to assess exclusions, and (3) the following month, ensuring sufficient look-forward to assess Vitamin D testing. Eligibility was determined for all beneficiary-months, not just those with primary care visits, to inform the weighting computations described below.

To project from values estimated on this sample to the population of individuals with employer-sponsored insurance, we used post-stratification weights calculated from population totals distributed with the database and derived from the Medical Expenditure and Panel Survey (MEPS). Briefly, after applying our inclusion criteria to all individuals represented in the Marketscan database during the months of interest, we computed totals *n(s, t)* of included individuals in stratum *s* during month *t*. The target population *T(s, t)* for stratum *s* in month *t* is taken to be 1/12 of that stratums annual total from MEPS for the year corresponding to *t*. The post-stratification weight *w(s, t) = T(s, t) / n(s, t)* is computed from these totals. These weights are utilized in forming all estimates (and standard errors) of jurisdiction-level aggregates, including the trends shown in Figure 1 and the statistical summaries in Table 2.

For Ontario, Canada and US Veterans we utilized population-level data which did not need to be weighted. Consequently, where weights appear in the definitions below, they should be assumed have constant value one across all strata and time points for these two jurisdictions.

Statistical Summaries

The regression models above estimate rates of low-value utilization of each blood test and corresponding trends for each individual stratum. However, for clarity of presentation we summarize these models at the jurisdiction level by aggregating results across strata. To do so, we make use of the following statistical summaries: average annual percent change (AAPC), absolute and relative total marginal effects, and the average marginal effect of time (AME Time). Each of these is defined below for the sake of completeness.

Average Annual Percent Change

The quasi-Poisson GEE regression models we use in this paper are log-linear, meaning the *relative* rather than absolute rate of change is proportional to time. Consequently, the purpose of AAPC is to summarize the *relative rate of change* implied by our models at the jurisdiction level.

Let *X(s, t)* be the row of the design matrix for strata *s* during month *t*, *N(s, t)* be the number of eligible beneficiaries, *w(s, t)* the post-stratification weight, and *b* be the estimated model parameters. Then, the expected low-value utilization rate per 100 for strata *s* during month *t* is r(s, t) = 100*exp(X(s, t)*b). The expected rate at the jurisdiction level, say R(t), is a weighted sum of these strata-level rates:

$$R(t) = \frac{\sum_{s=1}^{S} r(s,t) \times N(s,t) \times w(s,t)}{\sum_{s=1}^{S} N(s,t) \times w(s,t)}.$$

If *t* has units of months, then for a given time period from t_0 to t_1 the AAPC is then defined to be,

$$AAPC = 100 \times \frac{1}{t_1 - t_0 - 11} \times \sum_{t=t_0+12}^{t_1} \frac{R(t) - R(t-12)}{R(t)}.$$

Total Marginal Effects

The absolute total marginal effect for a recommendation variable is the difference between the expected low-value utilization rate predicted by the model and the counterfactual rate predicted assuming a continuation of pre-recommendation trends. Let $X_{cf}(s, t)$ be the design matrix for predicting this counterfactual and note that $X_{cf}(s, t)$ differs from $X(s, t)$ only if *t* is larger than the recommendation time. Similarly, let $r_{cf}(s, t) = exp(X_{cf}(s, t) * b)$ be the counterfactual rate for strata *s* during month *t* and $R_{cf}(t)$ the aggregate counterfactual rate for the jurisdiction. In this case, if t_0 and t_1 are the first and last months of the post-recommendation period the absolute marginal effect (aME) per 100 beneficiaries is:

$$aME = 100 \times \frac{\sum_{t=t_0}^{t_1} N(t) \times w(s,t) \times (R(t) - R_{cf}(t))}{\sum_{t=t_0}^{t_1} N(t) \times w(s,t)}.$$

Similarly, the *relative total marginal effect* the compares the above difference in tests to the number of expected tests under the counterfactual:

$$rME = 100 \times \frac{\sum_{t=t_0}^{t_1} N(t) \times w(s,t) \times (R(t) - R_{cf}(t))}{\sum_{t=t_0}^{t_1} N(t) \times w(s,t) \times R_{cf}(t)}.$$

Average Marginal Effect of Time

Finally, the average marginal effect of time (AME time) is the average derivative or $R(t)$ with respect to *t* over the course of either the pre- or post-recommendation period:

$$AME\ time = \frac{1}{t_1 - t_0 + 1} \times \sum_{t=t_0}^{t_1} \frac{\partial R(t)}{\partial t}.$$