Appendix -
Quality Disclosure and Regulation: Scoring Design in Medicare Advantage

Benjamin Vatter

October 24, 2021

Job Market Paper

Click here for the latest version of the draft and **here for the latest version of the appendix**[*]

Contents

[*]Or visit https://benjaminvatter.com.

1 Additional Figures and Tables

1.1 Figures

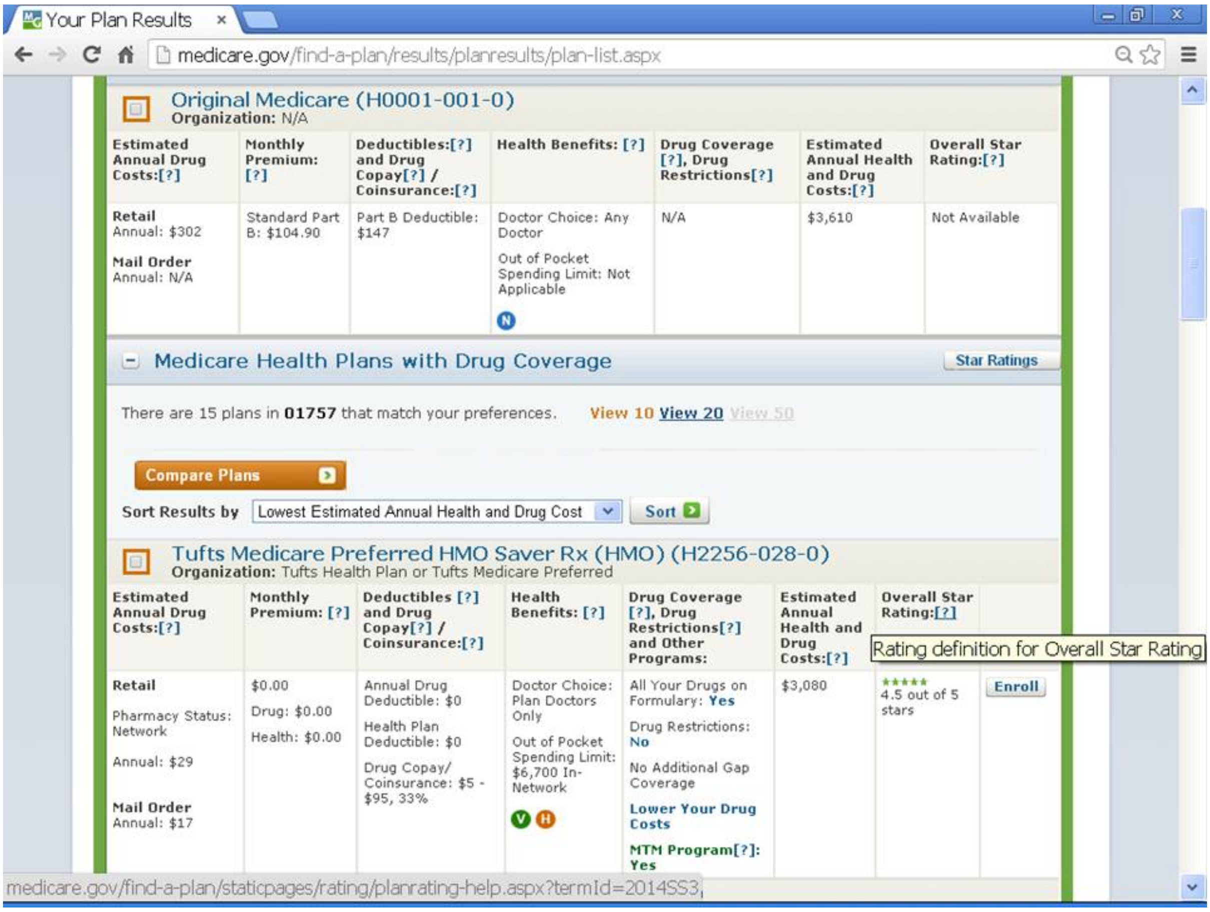
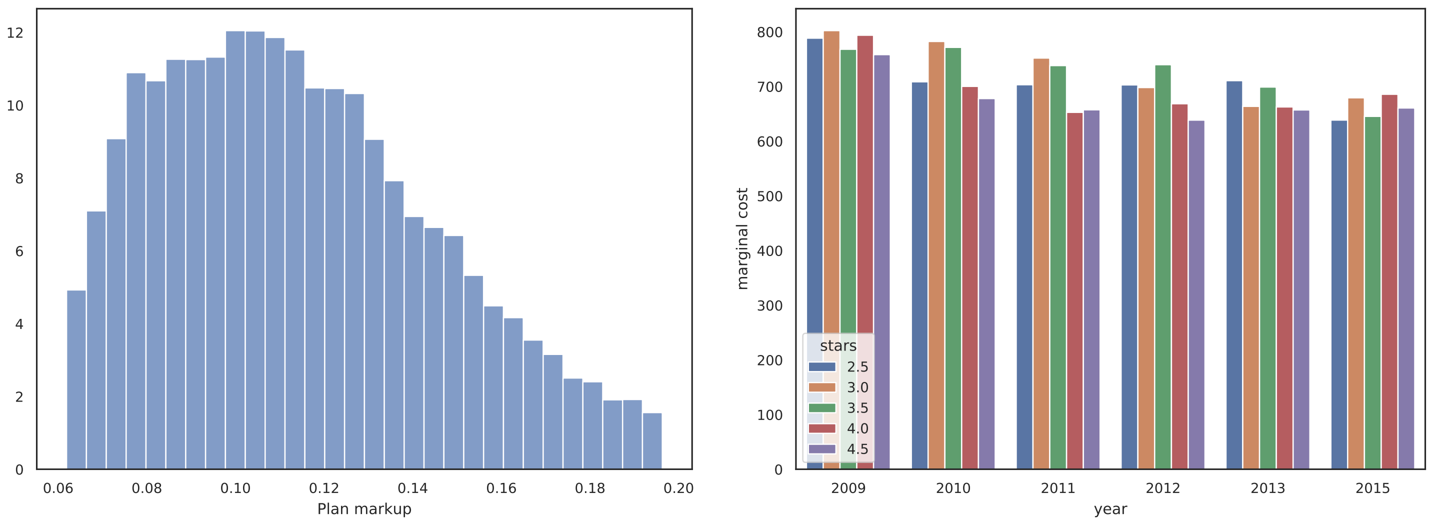


Figure 1: Medicare Plan Finder view

Notes: View of the Medicare Plan Finder platform from 2015. Source: https://byrondennis.typepad.com/theabcsofmedicare/2014/10/starting-to-think-about-it-4.html. For further details see Section 3.



(a) Markups

(b) Marginal cost over years

Figure 2: Estimated marginal costs

Note: These figures display the estimated insurer marginal cost curve of equation (5). Figure (a) shows the implied distribution of markups. Figure (b) shows the average marginal cost by star rating over years. For further details see Section 6.2.1.

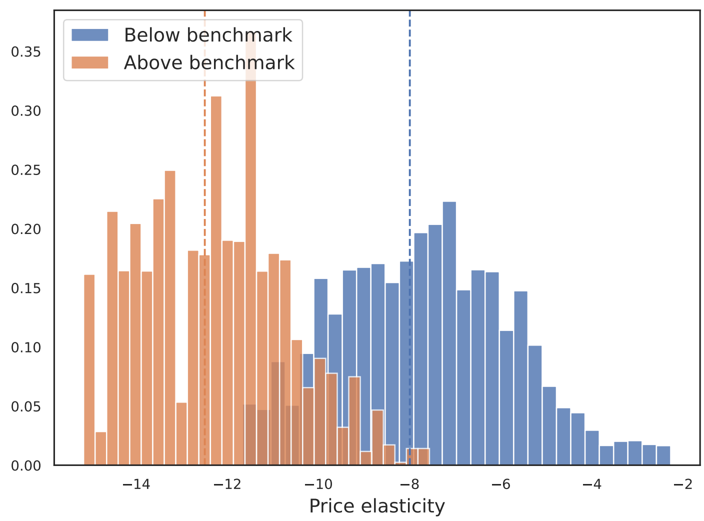


Figure 3: Price elasticity estimates

Note:This figures display the estimated consumer own-bid elasticity for chosen MA products. The figure is split among those plans bidding above and below the benchmark. The dashed lines represent the mean of each group, weighted by beneficiary MCBS weight. For more details see main text Section 5

1.2 Tables

Table 1: Quality Payment Rules

	Year				
Star Rating	2009-2011	2012	2013	2014	2015-2019
			Benchmark Bonus		
≤ 2.5	0.0%	0.0%	0.0%	0.0%	0.0%
3.0	0.0%	3.0%	3.0%	3.0%	0.0%
3.5	0.0%	3.5%	3.5%	3.5%	0.0
4.0	0.0%	4.0%	4.0%	5.0%	5.0%
4.5	0.0%	4.0%	4.0%	5.0%	5.0%
5.0	0.0%	5.0%	5.0%	5.0%	5.0%
New Plans	0.0%	3.0%	3.0%	3.5%	3.5%
Low Enrollment Plans	0.0%	3.0%	3.0%	3.0%	3.5%
			Rebate		
≤ 2.5	75%	66.67%	58.33%	50%	50%
3.0	75%	66.67%	58.33%	50%	50%
3.5	75%	71.67%	68.33%	65%	65%
4.0	75%	71.67%	68.33%	65%	65%
4.5	75%	73.33%	71.67%	70%	70%
5.0	75%	73.33%	71.67%	70%	70%
New Plans	75%	71.67%	68.33%	65%	65%
Low Enrollment Plans	75%	73.33%	58.33%	50%	65%

Note: New plans are those offered by parent organization that has not had any MA contract(s) with CMS in the previous three years. A low enrollment contract is a contract that could not undertake Healthcare Effectiveness Data and Information Set (HEDIS) and Health Outcome Survey (HOS) data collections because of a lack of a sufficient number of enrollees to reliably measure the performance of the health plan (Advance notice of methodological changes, 2012)

Table 2: Correlation between included and excluded measures

	access	intermediate	outcome	patient	process
access	0.102	0.223	0.275	0.264	0.203
intermediate	-0.061	-0.364	0.404	0.193	0.658
outcome	0.375	-0.253	0.095	0.171	0.242
patient	-0.099	-0.267	0.168	0.354	0.238
process	0.142	0.486	-0.031	-0.288	0.033

Notes: On the diagonal: the correlation between included and excluded measures in each category. Above the diagonal: the correlation across those included. Below the diagonal: the correlation across those excluded..

Table 3: Introduced measures

Measure	Category	Year
Adult BMI Assessment	Process	2012
Breast Cancer Screening	Process	2016
Improving Bladder Control	Process	2018
Medication Reconciliation Post-Discharge	Process	2018
Special Needs Plan Care Management	Process	2016
Statin Therapy for Patients with Cardiovascular Disease	Process	2019
Members Choosing To Leave the Plan	Patient	2012
MTM Program Completion Rate for CMR	Process	2016
Statin Use in Persons with Diabetes	Intermediate	2019

Notes: These are the measures introduced to the rating which are used for the analysis of Section 4.

Table 4: Demand Responses to Scoring - Full Table

	I	II	III	IV	V	VI
Weight category		Outcome	Intermediate	Access	Patient	Process
Rounded star rating						
2 stars	-0.000	-0.035***	-0.033***	0.080***	0.174*	0.121***
	(0.003)	(0.006)	(0.008)	(0.019)	(0.068)	(0.020)
3 stars	0.001	-0.036***	-0.027***	0.063***	0.123***	0.132***
	(0.001)	(0.004)	(0.003)	(0.006)	(0.019)	(0.014)
4 stars	0.007***	-0.043***	-0.030***	0.076***	0.170***	0.159***
	(0.002)	(0.004)	(0.003)	(0.007)	(0.021)	(0.015)
5 stars	0.012***	-0.036***	-0.023***	0.081***	0.109***	0.154***
	(0.003)	(0.006)	(0.005)	(0.008)	(0.027)	(0.018)
Rating category weight						
2 stars		0.728***	0.205***	-0.346***	-0.763**	-0.378***
		(0.116)	(0.047)	(0.092)	(0.292)	(0.059)
3 stars		0.770***	0.189***	-0.235***	-0.532***	-0.415***
		(0.077)	(0.019)	(0.022)	(0.084)	(0.043)
4 stars		0.881***	0.219***	-0.287***	-0.719***	-0.511***
		(0.083)	(0.021)	(0.025)	(0.092)	(0.048)
5 stars		0.831***	0.204***	-0.295***	-0.422***	-0.486***
		(0.094)	(0.025)	(0.029)	(0.120)	(0.058)
N	421606	421606	421606	421606	421606	421606
R^2	0.712	0.713	0.713	0.713	0.712	0.713
Mean chosen MA	0.0158	0.0158	0.0158	0.0158	0.0158	0.0158

Notes: This table displays the estimates of the demand response to rating and scoring design. The effect of category weights are relative to others, meaning that a negative number should be interpreted as consumers preferring less information on this category, as it would make the rating more informative of other dimensions. The regression is weighted by the MCBS sampling weights. The omitted category are new plans and plans that don't have star ratings due to insufficient enrollment in previous years. Standard errors in parentheses are heteroskedasticity robusts.

Table 5: Category scores and local ranking effects on demand

	(I)	(II)
Average category score		
Access	0.000266	
	(0.000568)	
Intermediate	0.000574	
	(0.000489)	
Outcome	0.000107	
	(0.000335)	
Patient	0.00197**	
	(0.000656)	
Process	-0.000229	
	(0.000724)	
Ranking		
Local ranking		0.000236*
		(0.000118)
N	421606	421606
R^2	0.715	0.715

Notes: Standard errors are heteroskedasticity robust. For further details see Appendix Section 4.

Table 6: MCBS beneficiary-level data descriptives by enrollment choices

	By Enrollment		**MA only by rating**		
	TM	MA	≤ 2.5	$[3, 4)$	≥ 4
Female	0.529	0.552	0.542	0.550	0.556
	(0.499)	(0.497)	(0.498)	(0.498)	(0.497)
Age	73.836	73.863	72.489	73.457	74.547
	(10.067)	(8.892)	(9.409)	(9.065)	(8.510)
Income ($000)	53.452	41.799	38.556	39.159	46.304
	(70.799)	(39.034)	(29.792)	(33.098)	(47.262)
Part B premium ($)	1264.190	1351.987	1359.038	1349.352	1341.224
	(429.690)	(231.837)	(199.869)	(214.713)	(255.486)
Health status					
Excellent	0.176	0.170	0.145	0.168	0.182
	(0.381)	(0.376)	(0.352)	(0.374)	(0.386)
Very good	0.310	0.324	0.325	0.304	0.345
	(0.463)	(0.468)	(0.468)	(0.460)	(0.475)
Good	0.292	0.290	0.280	0.298	0.282
	(0.454)	(0.454)	(0.449)	(0.457)	(0.450)
Fair	0.141	0.151	0.178	0.162	0.133
	(0.348)	(0.358)	(0.382)	(0.369)	(0.339)
Poor	0.054	0.050	0.060	0.055	0.042
	(0.226)	(0.219)	(0.237)	(0.228)	(0.201)
Knowledge about Medicare					
Read handbook	0.515	0.536	0.538	0.532	0.549
	(0.500)	(0.499)	(0.499)	(0.499)	(0.498)
Satisfied with information provided	0.753	0.764	0.760	0.753	0.776
	(0.431)	(0.424)	(0.427)	(0.431)	(0.417)
Difficulty understanding information	0.316	0.317	0.346	0.322	0.291
	(0.465)	(0.465)	(0.476)	(0.467)	(0.454)
Administrative indicators					
Employer-sponsored insurance	0.467	0.136	0.106	0.089	0.163
	(0.499)	(0.343)	(0.308)	(0.285)	(0.370)
Disability eligibility	0.102	0.090	0.120	0.112	0.060
	(0.303)	(0.286)	(0.324)	(0.316)	(0.237)
ESRD eligibility	0.007	0.004	0.003	0.004	0.004
	(0.086)	(0.063)	(0.051)	(0.067)	(0.064)
Observations	35316	11517	789	4807	4751

Notes: This table presents means and standard deviations (in parentheses) for the MCBS population, weighted according to the MCBS sampling. Income and premium are in 2015 dollars adjusted for the health CPI. The first two columns correspond to the entire MCBS sample split by beneficiaries' enrollment decisions. The last three columns present descriptive for the population that chooses a MA plan that can be linked in the aggregate data (98.9% of all MA enrollees).

Table 7: Quality response to scoring, separeted DD

	(1)	(2)	(3)	(4)
measure	Adult bmi assessment	Breast cancer screening	Enrollment timeliness	Improving bladder control
DD coeff.	0.102***	0.00858	0.0233	-0.000125
	(0.0168)	(0.00608)	(0.0170)	(0.00492)
N	1520	1505	533	1623
R^2	0.899	0.821	0.549	0.716
measure	Medication reconciliation post discharge	Mtm program completion rate for cmr	Special needs plan snp care management	
DD coeff.	0.0352	0.0464**	0.0827	
	(0.0242)	(0.0157)	(0.0492)	
N	400	1179	360	
R^2	0.872	0.864	0.712	

Notes: This table presents the estimates of the below/above median design for each treated measure. The sample is restricted to be within the 25th and 75th percentile of pre-treatement quality to avoid mechanic inflation due to bounded domain. One of the eight treated measures is dropped as there are not enough persistently observed contract between the 25th percentile and the median (treated) for this disagregate analysis.

Table 8: Measure-level scores effect on demand

Measure	Estimate	Measure	Estimate	Measure	Estimate
Rx Timeliness	0.000352	Appeals Auto-Forward	0.000653*	High Risk Medication	-0.00147**
	(0.000414)		(0.000321)		(0.000497)
Access to pcp	0.000906	Appeals Upheld	0.000592	Statin Adherence	-0.00163
	(0.000779)		(0.000311)		(0.00126)
Call timeliness	-0.00119	timely appeals	-0.000610	Diabetes Adherence	0.00267**
	(0.000851)		(0.000356)		(0.00103)
call hold time	-0.000483	Access problems	0.000648*	Hypertension Adherence	0.000274
	(0.000647)		(0.000303)		(0.00107)
pharmacists call center disconnect	-0.000551	Reviewing Appeals Decisions	-0.000300	COPD testing	0.00172*
	(0.00134)		(0.000351)		(0.000743)
Drug Call Center Disconnected	0.00232	Drug Plan Quality Improvement	0.000282	improving mental health	0.00141
	(0.00133)		(0.000602)		(0.000879)
call TTY	0.0000567	plan improvement	0.000615	improving physical health	-0.0000923
	(0.000460)		(0.000525)		(0.000567)
call information	0.000302	Beta-blocker treatment	-0.00143	All-cause readmissions	0.000172
	(0.000483)		(0.00125)		(0.000741)
Call Center - Pharmacy Hold Time	0.000778*	pressure control	0.00129*	coordination	-0.000151
	(0.000365)		(0.000525)		(0.000741)
Part D Beneficiary Access	-0.000303	Diabetes - sugar	-0.000269	Complaints about the Drug Plan	0.00300***
	(0.000294)		(0.000779)		(0.000838)
Part D Beneficiary call center	-0.00124*	Diabetes - LDL	0.000366	complaints	-0.000682
	(0.000571)		(0.000739)		(0.000637)
Part D Call Center - Foreign	0.000359	Diabetes Treatment	0.00145**	Complaints - Benefits	-0.000367
	(0.000396)		(0.000524)		(0.00105)
Part D Call Center - Accuracy	0.000286	Depression followup	0.00102	Complaints - Enrollment	0.00149*
	(0.000461)		(0.00331)		(0.000627)
Doctor communication	0.00129*	mental illness followup	-0.00260*	Complaints - Other	-0.00180**
	(0.000617)		(0.00109)		(0.000646)

Measure	Estimate	Measure	Estimate	Measure	Estimate
Complaints - Pricing	-0.000390	BMI assessment	-0.000974	bladder control	-0.0000290
	(0.000871)		(0.000623)		(0.000596)
Customer Service	-0.0000162	medication review	-0.00233*	PA monitoring	0.00348***
	(0.000391)		(0.00102)		(0.000765)
Part D Members Choosing to Leave the Plan	-0.00122	pain assessment	0.000928	MPF - Composite	0.000532
	(0.00149)		(0.00121)		(0.000336)
quick appointments	0.0000427	LDL Screening	0.00121	LT med monitoring	-0.000803
	(0.000560)		(0.00164)		(0.000759)
Getting Information From Drug Plan	0.000889**	Colorectal Cancer Screening	0.000608	MPF Price Accuracy	0.000361
	(0.000330)		(0.000540)		(0.000382)
Getting Needed Care	0.000822	Flu vac.	0.000123	MPF - Stability	0.000673
	(0.000636)		(0.000586)		(0.000491)
Getting Needed Prescription Drugs	0.000718	Part D Enrollment Timeliness	-0.000449	MPF - Updates	-0.000833
	(0.000450)		(0.000526)		(0.00110)
Member Retention	-0.00351**	Diabetes Care	-0.00166*	osteoporosis management	-0.000933
	(0.00110)		(0.000844)		(0.000625)
member leaving	0.000752	ATD management	-0.00158	osteoporosis testing	-0.000962
	(0.00154)		(0.00113)		(0.000546)
Rating of Drug Plan	0.000141	Diabetes - LDL screen	-0.000646	Pneumonia Vaccine	-0.00223**
	(0.000520)		(0.000809)		(0.000823)
Rating of Health Care Quality	0.000459	Diabetes - eye	0.00157*	Reducing the Risk of Falling	0.000407
	(0.000577)		(0.000730)		(0.000481)
Rating of Health Plan	-0.0000898	Diabetes - kidney	0.000653	Arthritis management	0.0000390
	(0.000619)		(0.000610)		(0.000386)
LDL screening (cardio)	0.000205	Enrollment Timeliness	0.00133	SNP management	0.00183
	(0.000555)		(0.000704)		(0.00125)
functional assessment	0.00238*	Glaucoma Testing	-0.000776	Brst. cancer screening	0.000158
	(0.00108)		(0.000645)		(0.000680)

Notes: These are the estimated coefficients on the measure-level score on choice probability.

Table 9: Mean plan characteristics by star rating groups

	Star Rating		
	≤ 2.5	$[3, 4)$	≥ 4
Premiums (Annual $)			
Part C (MA)	209.789	305.498	426.788
Part D (Prescription)	207.632	264.342	323.412
Total	417.420	569.841	750.200
Bidding (Annual $)			
Benchmark	10480.538	10344.601	9862.400
Bid	8764.563	8830.212	8245.742
Rebate payment	850.536	623.800	388.477
Supplemental coverage indicator			
Dental	0.650	0.622	0.712
Hearing aids	0.445	0.383	0.464
Vision	0.919	0.948	0.947
Plan cost-sharing			
Deductible ($)	21.550	38.768	31.342
Out-of-pocket limit ($)	5188.585	5630.970	5963.785
Inpatient copay	243.636	256.493	271.050
Outpatient copay	192.105	182.724	198.455
Inpatient coinsurance	0.000	0.051	0.053
Fraction HMO	0.589	0.568	0.552
Enrollee-Years (millions)	4.117	35.260	52.928
County-level MA share	0.262	0.284	0.332
Observations	19196	51686	68441

Notes: This table presents unweighted means at the county-year-plan level in the data. Benchmarks presented here are the true plan-level value, computed according to the CMS rules which weight county-benchmarks for a plan by its expected enrollment. All dollar values are in 2015 dollars adjusted for the healthcare CPI.

Table 10: Data descriptive statistics by market, contract and organization

	mean	min	25%	50%	75%	max
By county-year (N=24312)						
Plans	5.731	1.000	2.000	4.000	8.000	41.000
Organizations	2.744	1.000	1.000	2.000	4.000	20.000
HHI	0.010	0.000	0.001	0.004	0.012	0.346
Mean premium	45.225	0.000	15.682	38.420	66.000	312.434
Total MA enrollment	3796.702	11.000	164.070	604.083	2335.917	233710.672
MA county penetration	0.245	0.012	0.145	0.225	0.325	3.137
Share paying any premium	0.658	0.000	0.348	0.773	1.000	1.000
Mean star rating	3.411	-2.000	3.090	3.607	4.000	5.000
By contract-year (N=4238)						
Counties	20.445	1.000	5.000	10.000	21.000	1565.000
Plans	3.989	1.000	1.000	2.000	4.000	155.000
Star rating	3.586	2.000	3.000	3.500	4.000	5.000
Largest plan share of enrollees	0.702	0.065	0.509	0.709	1.000	1.000
Revenue (millions)	14.264	0.000	0.453	3.903	15.364	406.020
Rebate (millions)	1.446	0.000	0.018	0.244	1.235	63.441
By insurer-year (N=1277)						
Contracts	3.319	1.000	1.000	1.000	2.000	62.000
Total market share	0.009	0.000	0.000	0.001	0.004	0.238
Revenue (millions)	47.338	0.000	0.411	5.476	23.942	1749.804
Rebate (millions)	4.799	0.000	0.031	0.484	2.506	208.338

Notes: This table presents descriptive statistics for the main sample used for estimation. At the county-year level the ”mean premium”, ”Share paying any premium” and ”mean star rating” rows represent enrollment-weighted averages. At the contract and organization levels the ”Revenue” columns include part A and B payments from CMS, rebates and premium contributions by enrolles, in millions of dollars. Market share columns do not account for the share of TM. All dollar values are deflated to 2015 prices.

Table 11: Maximum likelihood estimates of individual preferences

	coefficient	std. err
MA x Demographics		
ESRD	-1.284***	(0.277)
Attended college	-0.037	(0.044)
College degree or higher	-0.141**	(0.045)
Disabled	0.040	(0.104)
Employer Sponsored	-1.017***	(0.041)
Female	0.076	(0.067)
Graduated high school	-0.015	(0.052)
Medium Income	0.067	(0.081)
High Income	-0.096	(0.086)
Asian indicator	-0.176	(0.117)
Black indicator	0.090	(0.062)
Hispanic indicator	0.341***	(0.066)
Switch: MA to TM		
Mean preference		
Health - Excellent	-3.727***	(0.113)
Health - Very Good	-3.930***	(0.083)
Health - Good	-3.569***	(0.082)
Health - Fair	-3.717***	(0.111)
Health - Poor	-3.796***	(0.208)
Switch: Across insurers		
Mean preference		
Health - Excellent	-3.104***	(0.153)
Health - Very Good	-3.138***	(0.110)
Health - Good	-3.158***	(0.111)
Health - Fair	-3.168***	(0.151)
Health - Poor	-2.835***	(0.272)
Switch: Within insurer		
Mean preference		
Health - Excellent	-1.948***	(0.223)
Health - Very Good	-2.271***	(0.144)
Health - Good	-1.910***	(0.155)
Health - Fair	-1.848***	(0.221)
Health - Poor	-1.449***	(0.368)
Obs.	36447	
Weighted Log. Likelihood	-5.131	

Notes: This table reports the MLE estimates of the individual preference coefficients of Equation (12) not reported in the main text. Observations are weighted using the MCBS sample weights to obtain nationally representative estimates. For interactions with health status, the omitted category is ”Fair”. For interactions with income, the omitted category is ”Low income”, which corresponds to the first tercile of income distribution in the data. Medium and high income are defined accordingly. For more details see Section 6.2

Table 12: Estimates of mean preference for plan charactersitics

	coefficient	std. err
Second stage: common preferences		
Premium	-1.112**	(0.393)
Benefits	2.915***	(0.383)
Drug deductible	-0.001***	(0.000)
No part D coverage	-1.778***	(0.020)
Enhanced drug benefits	0.093***	(0.012)
Dental cleaning	1.846***	(0.060)
Dental exam	-2.409***	(0.083)
Dental fluoride	-0.438***	(0.024)
Dental xray	0.682***	(0.040)
Hearing aids	-0.229***	(0.031)
Hearing fitting	-0.106***	(0.025)
Vision	-0.032	(0.023)
Third stage: quality preferences		
Access	4.501***	(0.365)
Intermediate	1.839***	(0.042)
Outcome	5.002***	(0.807)
Patient	3.792***	(1.112)
Process	2.315***	(0.161)
Obs.	29004	
Mean Bid elasticity	-8.348	
Mean Premium elasticity ($p^C > 0$)	-0.951	

Notes: This table reports the estimates of common preference coefficients, described in Equation (13) and the posterior mean coefficients described in Equaltion (14). These correspond to the second and third stage estimates of the demand model. Standard errors are computed using the delta method, accounting for the asymptotic distribution of $\hat{\delta}$ and $\hat{\eta}$. For further details see Section 6.2

Table 13: Demand Estimation First Stage

	Premium			Benefits		
	I	II	III	I	II	III
Instruments						
Benchmark	-0.00715***	-0.00274	0.00772***	0.0140***	0.00963***	-0.00818***
	(0.00169)	(0.00168)	(0.00180)	(0.00152)	(0.00143)	(0.00118)
Rebate	0.656***	0.609***	0.398***	-0.317***	-0.275***	-0.162***
	(0.00582)	(0.00567)	(0.00484)	(0.00415)	(0.00400)	(0.00299)
Other endogenous						
Benefits	0.782***	0.836***	1.289***			
	(0.00886)	(0.00878)	(0.00914)			
Premium				0.390***	0.405***	0.463***
				(0.00384)	(0.00369)	(0.00301)
Market FE						
Product controls	No	Yes	Yes	No	Yes	Yes
Contract-year FE	No	No	Yes	No	No	Yes
N	29000	29000	28830	29000	29000	28830
R^2	0.566	0.622	0.887	0.451	0.536	0.896

Notes: This table reports the first stage estimates of the second step of the demand estimation. Version I and II are provided for clarity; version III is the appropriate first stage for the demand estimates in their preferred specification. Standard errors in parentheses are heteroskedasticity robust. For further details, see main text Section 6.1.2.

Table 14: Consumer surplus loss from asymmetric information, across groups

	Current System		**Current System - MA Enrollees**	
	Mean	Std. Dev.	Mean	Std. Dev.
Gender				
Female	-175.814	260.800	-206.181	330.554
Male	-197.447	303.615	-224.354	345.784
Health status				
Excellent	-172.087	268.304	-203.619	321.004
Very good	-181.861	287.578	-215.189	341.945
Good	-186.120	264.959	-207.994	306.529
Fair	-212.210	318.960	-233.411	407.104
Poor	-193.042	276.200	-232.600	310.135
Age group				
< 65	-178.906	215.132	-229.427	271.113
$\in [65, 70)$	-190.430	267.300	-220.884	290.385
$\in [70, 75)$	-178.704	291.771	-197.396	350.788
$\in [75, 85)$	-181.458	287.517	-203.597	320.932
≥ 85	-199.697	357.729	-252.760	509.249
Read Handbook				
No	-187.642	283.545	-223.760	337.021
Yes	-182.762	278.480	-202.148	338.225
Satisfied with information				
No	-187.474	250.677	-240.604	313.646
Yes	-185.016	297.113	-204.691	345.777

Notes: This table compares the consumer surplus loss from assymetric information across demographic groups. Values are per-consumer in yearly dollar. Statistics are computed using the MCBS sampling weight. These values are computed as in Train (2015), with the numbers indicating the difference between the ex-post asymmetric information consumer surplus and the value under full information. For more detail see Section 6.2.2.

2 Quality Disclosure and Regulation

This section provides additional results related to the monopoly regulation example presented in Section 2 in the main text.

2.1 Monopolistic Spencian Distortions

The following proposition provides the condition for the certification described in the main text to be strictly better than full information. I focus on the case in which the firm has separable constant marginal and investment costs such that $C(x,q) = xc + I(q)$ with $I(q)$ being the investment cost. This is the framework of Spence (1975). Throughout, I denote $\pi(q, x^*_F(q))$ the full-information profit of the monopolist under quality q and optimal monopolistic quantities. Subscripts on functions denote partial derivatives.

Proposition 1. *Let $(q^*, x^*_F(q^*))$ be the monopolist's full information optimum. Assume that $\pi(q^*, x^*_F(q^*)) > \pi(0, x^*_F(0))$. Then, total welfare under quality certification is equal to full-information welfare if and only if the following equality holds at $(q^*, x^*_F(q^*))$.*

$$\frac{P_x P_q + x(P_{xx}P_q - P_x P_{xq})}{2P_x + xP_{xx}} = \frac{1}{x}\int_0^x P_q(v,q)dv \tag{1}$$

Proof. First, I restate the problem of the monopolist and the regulator under full information

$$\pi(q,x) = x(P(x,q) - c) - I(q)$$
$$W(q,x) = \int_0^x P(v,q)dv - xc - I(q)$$

The optimal quantity choice of the monopolist is characterized by the first order condition

$$P(x^*_F(q), q) + x^*_F(q) P_x(x^*_F(q), q) = c$$

The optimal choice of quality given $x^*_F(q)$ for both is characterized by the first order conditions

$$\frac{\partial x^*_F(q^m)}{\partial q}(P(x^*_F(q^m), q^m) - c) + x^*_F(q^m)(P_x(x^*_F(q^m), q^m)\frac{\partial x^*_F(q^m)}{\partial q} + P_q(x^*_F(q^m), q^m)) = I'(q^m)$$
$$\frac{\partial x^*_F(q^p)}{\partial q}(P(x^*_F(q^p), q^p) - c) + \int_0^{x^*_F(q^p)} P_q(v, q^p)dv = I'(q^p)$$

where q^m and q^p are the optimal choices of the monopolist and the planner, respectively.

Given that the regulator can restrict the quality by limitting information, she will chose not to do so only if the two solutions agree. This happens only if

$$P_x \frac{\partial x_F^*}{\partial q} + P_q = \frac{1}{x_F^*} \int_0^{x_F^*} P_q(v, q) dv$$

where the function above is evaluate at the q^m. Finally, taking the implicit derivative of $x_F^*(q)$ from its characterizing equation, and substituting in the equation above results in the expression in equation (1). □

2.2 Monopoly Regulation With Private Cost Types

I present an example based on Zapechelnyuk (2020), where the monopolist is privately informed of its investment cost. Let the demand be characterized by $D(p, q) = (q - p)^\alpha$, with $\alpha > 0$. The monopolist is privately informed of his type, $t \in [\underline{t}, \bar{t}]$, such that it's investment cost is $I(q) = c(q)/t$. The type t is distributed according to a distribution G with positive density g. Finally, assume that $c'(q)/q^\alpha$ is strictly increasing and greater then $\bar{t}$ for some $q > 0$.[1] The following proposition, summarizes the result.[2]

Proposition 2. *Denote* $f(t) = (g'(t)/g(t) - \frac{1}{\alpha t + \underline{t}})$ *then*

1. *If* $f(t) > 0$ *for all* $t \in [\underline{t}, \bar{t}]$*, then there exists* $\underline{q} > 0$ *such that the disclosure rule* $\psi(q) = \mathbb{1}\{q \geq \underline{q}\}$ *is optimal.*

2. *If there is* $t^* \in (\underline{t}, \bar{t}]$ *such that* $f(t) \leq 0$ *for all* $t \in [\underline{t}, t^*)$ *and* $f(t) \geq 0$ *for all* $t \in [t^*, \bar{t}]$*, then there exists* $\underline{q} > 0$ *such that* $\psi(q) = q\mathbb{1}\{q \geq \underline{q}\}$ *is optimal.*

Proof. The proof approach is as in Zapechelnyuk (2020), and will be to first recast the problem as a linear delegation problem, then apply the equivalence theorem of Kolotilin and Zapechelnyuk (2018) to convert it to a linear persuasion problem. The final results are based on characterizations of solutions to the linear persuasion problem provided by Kolotilin and Zapechelnyuk (2018) and Kolotilin et al. (2019). I begin by restating the problem. For simplicity I take the constant marginal cost to be zero. As the planner and the monopolist agree on this constant cost, it shouldn't change the type of disclosure qualitatively. Substituting the optimal monopolistic quantity given a quality choice, the firm's profit and the total welfare are given by (see

[1]It is easy to verify that in this example the condition of equation (1) is not satisfied and quality under full information would be underprovided relative to an informed planner's choice.

[2]This the model of Zapechelnyuk (2020), with the distinction that I consider a total-welfare maximizing regulator instead of a consumer-surplus maximizing one. Zapechelnyuk (2020) shows a similar result to proposition 2, under different conditions.

Zapechelnyuk (2020))

$$\pi(t,q) = Kq^{1+\alpha} - \frac{c(t)}{t}$$
$$W(t,q) = \frac{1+2\alpha}{1+\alpha}Kq^{1+\alpha} - \frac{c(t)}{t}$$

Where $K = \frac{\alpha^\alpha}{(1+\alpha)^{1+\alpha}}$. To use the equivalence result we must transform the problem such that the action (q) and the state (t) are both in $[0,1]$. The assumption on $c'(q)/q^\alpha$ guarantees that there is $\bar{q} > 0$ such that the profit of the monopolist of $q > \bar{q}$ are below $\pi(t,0)$ for all t. So to convert the problem, define $y = (\frac{q}{\bar{q}})^{1+\alpha}$ and $\theta = \frac{t-\underline{t}}{\bar{t}-\underline{t}}$, and write

$$\pi_D(\theta,y) = [\theta(\bar{t}-\underline{t}) + \underline{t}]y\bar{q}^{1+\alpha} - K^{-1}c(y^{1/(1+\alpha)}\bar{q})$$
$$W_D(\theta,y) = [\theta(\bar{t}-\underline{t}) + \underline{t}](\frac{1+2\alpha}{1+\alpha})y\bar{q}^{1+\alpha} - K^{-1}c(y^{1/(1+\alpha)}\bar{q})$$

The maximizing action to this problem and the one above are the same. We can interpret this rescaled version as the problem of a principal of objective $W_D(\theta,y)$ that has to determine a set of admissible actions $\boldsymbol{Y} \subset [0,1]$, for an agent of private type θ, who will take an action $y \in \boldsymbol{Y}$. If we constrain the principal to always include $\{0,1\}$ in $\boldsymbol{Y}$, this becomes a linear balanced delegation problem (Kolotilin and Zapechelnyuk (2018) section 5). By theorem 1 in Kolotilin and Zapechelnyuk (2018) there is an equivalent persuasion problem that has the same solution, which is given by the objectives

$$\pi_P(\theta,y) = \int_0^{F^{-1}(y)} (d(\theta)-t)f(t)dt$$
$$W_P(\theta,y) = \int_0^{F^{-1}(y)} (d(\theta)-t\frac{1+2\alpha}{1+\alpha} - r)f(t)dt$$

Where F is the distribution of θ, $r = \frac{\underline{t}}{\bar{t}-\underline{t}}$ and $d(\theta)$ is given by

$$d(y) = \frac{c'(y^{1/(1+\alpha)}\bar{q})y^{-\alpha/(1+\alpha)}}{K(\bar{t}-\underline{t})\bar{q}^\alpha} - r$$

In this persuasion problem, a sender of preferences represented by $W_P(\theta,y)$ commits to a signal structure about a state, θ, provided to an uniformed agent of utility $\pi_P(\theta,y)$, who takes an action y. Given that $f(t)$ is positive, we can see that the receiver would like to take the action $F(d(\theta))$ while the sender would like the strictly lower action $F((d(\theta)-r)\frac{1+\alpha}{1+2\alpha})$. Mapping this back to the original rating problem, we can think about the new state as being the quality that firms have to provide (as $d(y)$ is increasing), and the choice of the agent is the last type to accept the action. So the persuasion problem is stating that the regulator would like lower types (that is, higher cost) to accept the same action.

As the agents profit is linear in the transformed state $d(\theta)$ we can think of the signal as informing the posterior mean belief about it. Hence, for a posterior-mean m, we can see that the optimal choice of the firm would be

$$y^*(m) = \arg\max_{y\in[0,1]} \mathbb{E}_{d(\theta)}[\int_0^{F^{-1}(y)} (d(\theta) - t)f(t)dt|m] = F(m)$$

And the regulator's profit of the posterior-mean m is

$$v(m) = \mathbb{E}_{d(\theta)}[W_P(\theta, y^*(m))|m] = \int_0^m (m - t\frac{1+2\alpha}{1+\alpha} - r)f(t)dt$$

Solutions to the senders problem are often characterized in terms of the concavity/convexity of $v(m)$. The second derivative of $v(m)$ is given by

$$\begin{aligned} v''(m) &= \frac{1}{1+\alpha}f(m) - (\frac{\alpha}{1+\alpha}m + r)f'(m) \\ &= \frac{1}{1+\alpha}(\bar{t} - \underline{t})g(m(\bar{t} - \underline{t}) + \underline{t}) - (\frac{\alpha}{1+\alpha}m + r)(\bar{t} - \underline{t})^2 g'(m(\bar{t} - \underline{t}) + \underline{t}) \end{aligned}$$

Where the second equality follows from substituting in the distribution implied by the definition of θ as a function of t. So, replacing back $t = m(\bar{t} - \underline{t}) + \underline{t}$, we have that $v''(m) > 0$ if

$$g(t) - (\alpha t + \underline{t})g'(t) > 0$$

and $v''(m) < 0$ if the inequality is reversed. It is also important to note that $v'(0+) < 0$ and $v'(1-) < 1$.

Finally, to derive the first results I use proposition 3 part 2 in Kolotilin and Zapechelnyuk (2018). The result implies that the regulator in the delegation case will only allow a specific action and the limiting actions. This implies that the regulator will send a signal if the quality falls equal or above its desired action, and a separate signal if it falls below it. Allowing for the upper limit of actions is irrelevant as the monopolist will never take it. The second case can be seen as two cases. If $t^* = \bar{t}$ then we can apply proposition 3 part 1 of Kolotilin and Zapechelnyuk (2018). Otherwise, it means that $v(m)$ is concave and then convex, which implies that $-v(m)$ is S-shaped, and so theorem 1 from Kolotilin et al. (2019) implies that the optimal signal structure is a lower-censorship. This in turn implies that all actions above a threshold are admissible and so the rating is also a lower censorship rule. □

This result shows that the quality disclosure method can vary substantially depending on the setting and the regulator's information. The first case is one in which the likelihood of high-cost investment is low so that the regulator can establish a strict certification threshold as in the fully-informed regulator example. However, unlike the fully-informed regulator, this certification threshold leads some monopolists to not invest in quality when the cost is high. In the second case of the proposition, the likelihood of high-cost investment is more considerable, and so the planner benefits from yielding some power over quality to the monopolist. This policy allows the producer to scale quality up or down depending on its investment draw. However, the planner still benefits from mandating a minimum level of quality to avoid substantial under-investments. If the regulator was restricted to a handful of signals, it is reasonable that the quality floor would be the first signal's cutoff, using the remaining signal to convey information beyond it.

2.3 Bayesian Consumers

The illustration of Section 2 in the main text assumed that consumers have rational expectations about quality. However, the standard empirical framework for incompletely informed consumers is that they update a Bayesian prior. In this alternative setup, consumers have some prior belief about the quality provided by a monopolist and update it based on the score given by the regulator. For example, consumers might believe that quality is, in general, lognormally distributed. They might hold this belief because they are uninformed about the monopolists' cost structure but have some experience in similar markets that led to this prior. The regulator, however, is informed about the firms' cost and can operate as in the main example.

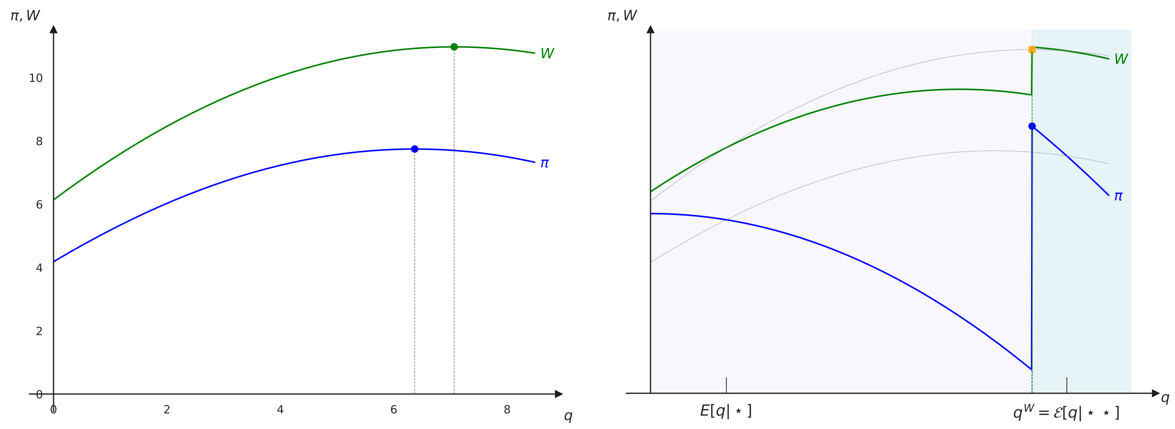
Under mild conditions, this distinction between rational expectation and Bayesian beliefs plays only a small role in the analysis of scoring design. To see this, we can consider the scoring rule illustrated in the main text, in which the designer creates a certification that only allows consumers to distinguish if quality exceeds a threshold value (i.e., q^W in Section 2). As long as consumers' posterior beliefs after seeing the low-quality signal are low enough, the new scores will make the monopolist produce an efficient level of quality. Figure 4 shows this scenario when consumers hold lognormal priors. A minor distinction of this setup is that because consumers belief exceed firm investments, some transfer of welfare occurs. Under regulation, firm's profits are higher and surplus is lower, but total welfare remains the same.

3 Institutional Details and Data

3.1 Additional institutional details

This section presents some additional details about the MA market. For the history of the MA market see McGuire et al. (2011).

The platform: Medicare Advantage organizes plan offering by establishing a unified shopping platform and regulating contract characteristics. Figure 1 displays the view of the platform in 2015. The platform presented consumers with the Original Medicare (TM) option at the top and MA plans at the bottom. Each plan displayed its estimated deductibles, cost-sharing rules and monthly premiums. Importantly, the system showed the "Estimated annual health and



(a) Full information

(b) Regulated

Figure 4: Quality certification under monopolistic provision and Bayesian demand

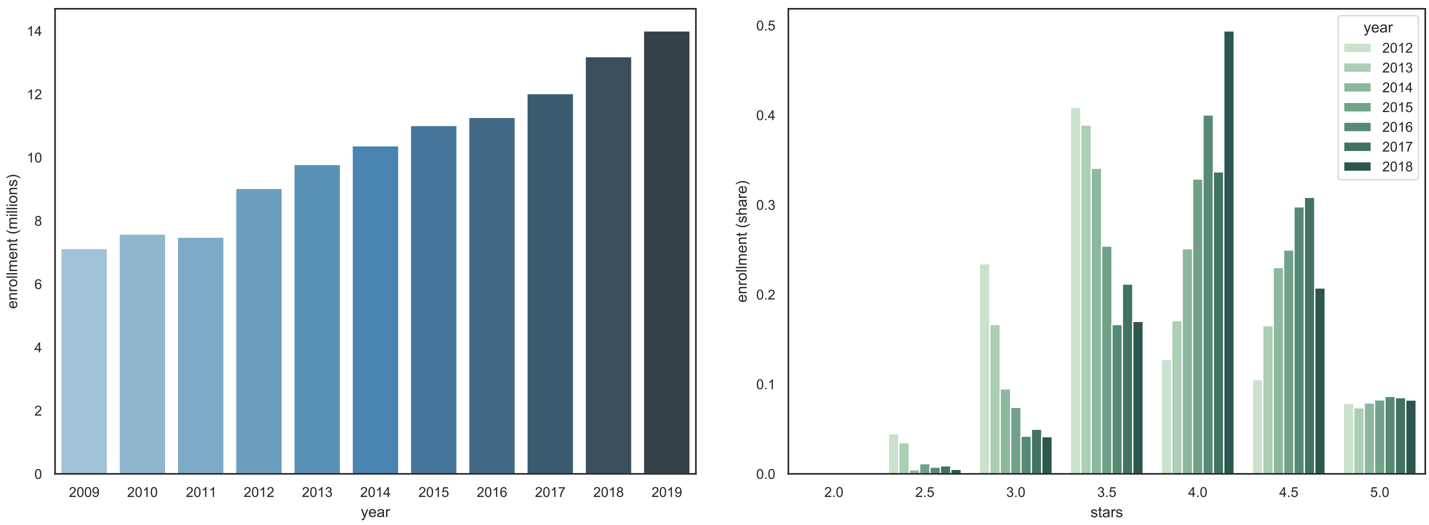
Notes: This figure illustrates how quality certification changes a monopolist's incentive to invest in quality. On the left, the figures illustrate the profit and welfare curves for the market under full information. On the right, these curves are distorted by the presence of a score that reveals to consumers whether quality exceeded the socially optimal level. The short vertical lines on the horizontal axis indicate the location of consumers' posterior beliefs.

drug costs" which are closely related to the benefit levels used in the main text. Also, next to the enrollment button, the system included the MA Star Rating for the plan. Clicking on the question mark in the column name revealed the basic construction details of this rating.

Enrollment and plan offering: Participation in the MA market was low in its early years. Enrollment picked up only after the enactment of the 2003 Medicare Modernization Act. The act modified insurer payments, most notably risk-adjusting them. Since then, enrollment has grown steadily to reach 22 million enrollees (34% of Medicare-eligible) in 2019 (Jacobson et al., 2019). Figure 5.a show the overall enrollment in MA, and 5.b shows the share of each year's enrollment by star rating. The average consumer in 2019 chose among fourteen different contracts, six by Health Maintenance Organizations (HMO), four in Preferred Provider Organizations (PPO), and a remainder of small Private Fee-For-Service (PFFS) and specialty contracts.[3] HMO and PPO plans are the most popular, enrolling 81.8% of the MA population and the majority bundle coverage for prescription drugs (part D).

Pricing regulation: Regulation in the MA market operates through a process known as "bidding" which, despite the name, is not an auction. Every year, insurers submit insurance plan offerings, listing each plan's participating counties and cost-sharing attributes. For each plan, insurers must provide estimates of expected expenses given coverages, benefits, and market characteristics, all of which have to be certified by an actuary. Additionally, insurers submit plan-level estimates of administrative costs and profit margins. The bidding procedure combines these data to form two components: First, the revenue required to cover expenses and margins related to standard and mandatory Medicare coverage, which CMS calls the "bid".

[3]The same firm might offer PPO, HMO, and PFFS contracts. These types primarily represent the network and cost-sharing arrangements of insurance plans.



(a) Total enrollment

(b) Share by star rating

Figure 5: Medicare Advantage enrollment trends

Notes: This figure shows enrollment trends for MA. Figure (a) displays total by years, and (b) the share by star rating.

In the main text, I call this value the plan's *price*. Second, the revenue required to cover supplementary benefits, such as lower copays and maximum out-of-pocket, which I refer to as the plan's additional *benefits*. These additional benefits are not optional and do not include the dental, vision, hearing, or Part D prescription coverage many MA plans offer, which CMS regulates separately.

A plan's bid reflects the revenue requested by an insurer to run a plan equivalent to TM in terms of actuarial value. As such, the bidding process compares a plan's bid against a metric reflective of the cost of running the public alternative, called the plan's "benchmark". To determine these plan-level benchmarks, CMS first computes county-level benchmarks as a percentage over the county's TM cost.[4] A plan's benchmark is then the average of its counties' benchmarks, weighted by the plan's expected enrollment in each. The difference between a plan's bid and its benchmark is fundamental. For plans bidding above the benchmark, CMS will pay the insurer an amount equal to the benchmark per enrollee. Enrollees pay the remaining difference, known as the "basic" MA premium. For plans bidding below the benchmark, CMS pays their bid plus a fraction of the difference, known as the rebate.

Additional benefits are what allows MA insurers to offer more generous coverage than TM. However, to offer these benefits, insurers must fund them through either premiums or rebates. Specifically, insurers must use every dollar of rebates to either fund benefits or buy-down non-MA enrollee premiums. The latter includes both the part B premium charged by CMS to every enrollee regardless of their choice between TM and MA and any part D prescription

[4] Song et al. (2013) provides further details on county-level benchmarks and why they differ from TM costs.

drug premium the plan might charge. Any additional benefits not funded by rebate is paid directly by the consumer under the "supplementary" MA premium.

Overall, the following equations summarize this regulation.

$$\begin{aligned}
\text{Rebate}_j &= \rho_j \max\{B_j - p_j\} \\
\text{Premium}_j &= \underbrace{\max\{p_j - B_j, 0\}}_{\text{basic}} + \underbrace{\max\{b_j - \text{Rebate}_j, 0\}}_{\text{supplementary}} \\
\text{Payment}_j &= \min\{p_j, B_j\} + \text{Rebate}_j + \text{Premium}_j
\end{aligned}$$

Where b_j is the plan's additional cost-sharing benefits, above the minimum requirement, measured in cost-savings for the average unit-risk consumer. The rebate share ρ_j has changed over the years and, since 2012, depends on a plan's previous rating. CMS also requires plans to allocate their rebates to consumer benefits or to reduce other non-MA premiums.

To put the formulas above in perspective, per member per month, the average MA plan in 2015 (by enrollment) submitted a price of \$700, additional benefits equal to \$70, and faced a benchmark of \$782. Among plans with a non-zero premium (43.8%), the average was \$73.5 with 13.4% coming from the basic MA premium. More than half of enrollees chose a zero-premium plan (58%). However, 83.9% of plans had a zero basic premium, resulting in a rebate averaging \$63.8. Every MA plan offered additional benefits to consumers, averaging an actuarial value for medical services of 87.2%.[5]

Comparison to previous descriptions of the market: There is some discrepancy in the literature regarding the bidding process. A recent release of information by CMS, containing the full bidding data, software, and instructions sent to insurers for 2009-2015, informed the description above and solves some of these discrepancies. I discuss the two most noticeable ones: bidding-level and the premium function.

Bidding is often described, erroneously, as taking place at the plan-county level. However, the extent to which previous results based on this slightly misspecified bidding model are biased is likely to be limited. For example, while bidding occurs at the plan-level, insurers can segment their contracts across counties and submit separate bids for each segment. The fact that 99% of all plans offered in more than one county chose not to segment suggests the gains from bidding independently across counties is small. This, in turn, suggests small losses from treating competition as occurring at the county level.

Premiums are often described as being the positive difference between bids and benchmarks. As discussed above, this is the definition of the *basic* MA premium, which constitutes only 12% of the total MA premium paid by consumers. The standard model also describes rebates as being the positive difference between benchmarks and bids. Thus, each plan should

[5]The actuarial values of MA were computed using public CMS software. Due to data limitations, this number is averaged across 2009-2015.

only have either a premium or a rebate. This model is clearly refuted by the data as 45% of plan-counties have both a premium and a rebate. The corrected model, which I present in the main text, can accommodate this observation and matches the data accurately.

Using the wrong bidding model might have important implications for previous work for two reasons. First, before the release of the bidding data, plan bids were not observed in the data. Instead, previous work routinely recovered it by inverting the premium or rebate equations, which were erroneously specified. This might lead to an erroneous accounting of subsidy spending and can overstate how much of the correlation in bids is due to demand and cost. Second, the standard model ignores the supplemental premium which can lead to erroneous accounting of pass-through.

To give a concrete example, consider Kaiser's contract H0425, plan 39, which operates only in Santa Clara county CA. This plan covers 40760 individuals in 2015, making it one of the largest plans in the country. The basic data shows that it has a part C premium of $69, Rebate of $38, adjusted market benchmark of $825, and rebate fraction of 70%. Clearly, the standard model can not reconcile these observations. Under the old model there are two ways we could have recovered the bid. First, we could ignore rebates and invert the premiums to get a bid of $894. Alternatively, we could ignore the premiums and get a bid of $786. The latter is correct. Now suppose we erroneously used the first method to recover the bid, and we are studying the pass-through of benchmarks to bids. Suppose we observe a 1% increase in benchmarks which leads to a 0.5% in the unobserved (given the old data) bid. Inverting again the bid using the first erroneous method would deliver an erred 3.2% decrease in bids.

3.1.1 The Star Rating Program: This section provides some additional details on the construction of the MA Star Ratings, their history, additional market behaviors, and some discrepancies with previous literature.

Brief history: Improving the quality of care and beneficiary general health is one of CMS' most important strategic goals (Centers for Medicare and Medicaid Services, 2016). To this end, CMS has undertaken several initiatives to gather and display the quality of MA plans. Following the enactment of the Balanced Budget Act of 1997, CMS began collecting information on a set of standardized quality measures through a combination of surveys and insurer reports. A summary of the gathered data was first presented to consumers in the November 1999 edition of *Medicare & You*, a handbook mailed annually to Medicare-eligible enrollees. The impact of this first implementation was noticeable, as studied by Dafny and Dranove (2008). In 2007, CMS began summarizing the quality information into five quality domains (e.g., "Helping You Stay Healthy"), with values described by one to five stars. In 2009, the star rating program took on its current form, with a single overall rating displayed to consumers next to the name, premium, and cost-sharing attribute of a plan. Figure 1 shows a view of this in the Medicare Plan Finder website. Figure 6.a shows an excerpt of the Medicare & You handbook that

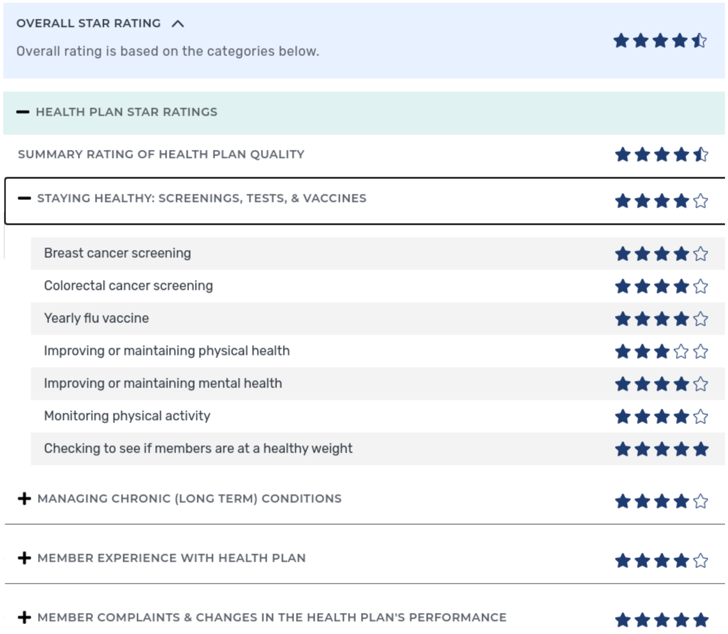
How do I compare the quality of Medicare health and drug plans?

When you visit Medicare.gov/plan-compare to find and compare health and drug plans, you'll see a star rating system for Medicare health and drug plans. The overall star rating gives an overall rating of the plan's quality and performance for the types of services each plan offers.

For plans covering health services, this is an overall rating for the quality of many medical/health care services that fall into 5 categories and includes:

1. **Staying healthy—screening tests and vaccines:** Whether members got various screening tests, vaccines, and other check-ups to help them stay healthy.
2. **Managing chronic (long-term) conditions:** How often members with certain conditions got recommended tests and treatments to help manage their condition.
3. **Member experience with the health plan:** Member surveys of the plan.
4. **Member complaints and changes in the health plan's performance:** How often members had problems with the plan. Includes how much the plan's performance improved (if at all) over time.
5. **Health plan customer service:** How well the plan handles member calls and questions.

(a) Medicare & You excerpt



(b) Plan Compare rating details

Figure 6: Information sources for rating design

Notes: The name of categories shown to consumer is different from the ones used in the main analysis. The figure on the right shows the rating details for a plan, this is visible after clicking on the details of the star rating in the main platform. For further details see Section 3.

describes the ratings for consumers.

Included and excluded measures: CMS first classifies measures as either included in the rating or excluded, and then into one of five categories: outcomes (e.g., improving or maintaining physical health), intermediate outcomes (e.g., controlling blood pressure), patient experience and complaints (e.g., consumer rating of the plan's customer service), access (e.g., the processing time for appeals), and process (e.g., colorectal cancer screenings).[6] The correlation across and among the included and excluded groups, by category, is presented in Table 2. Included quality measures belonging to the same category are considered to represent a similar type of quality and are assigned an identical weight in the overall rating. As shown in Figure 7, the number of measures in each category has varied over the years. As there's a single weight for a category, this has created significant variation in the information represented by the Star Ratings.

Stochastic quality: MA features substantial cross-sectional variation in quality. The distribution of continuous ratings, before their final discretization, is well dispersed as shown in Figure 8.[7] While it might appear contradictory for a firm to invest in a quality in the interior of a rating interval – as consumers do not observe it –, there is a simple explanation for it.

[6]Most quality measures can be thought of as "the fraction of enrollees obtaining service X." For example, CMS computes the outcome measure *improving or maintaining physical health* as the fraction of the population that saw their physical health improved or sustained during the last year.

[7]Star ratings of 1 and 1.5 are rarely observed. Rating between 2 and 3 are less rare but not often provided by top insurers. Otherwise, large firms provide contracts covering the range of stars.

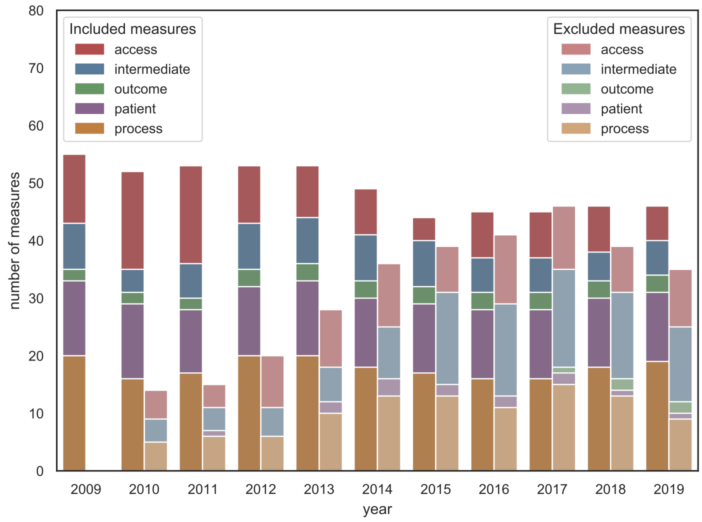


Figure 7: Number of included and excluded quality measures

Note: The weight assigned to each included measure before 2012 was equal to 1. Starting 2012 it was set to 3 for measures categorized as outcomes and intermediate outcomes, 1.5 for patient and access and 1 for process. Excluded measures have a weight of zero. For more details see Section 3.

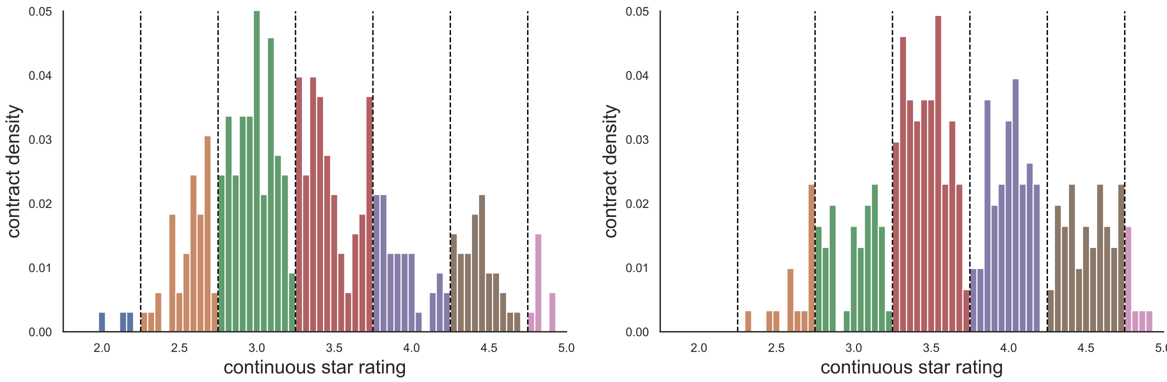
Investments in MA are contractual arrangements with providers and third-party services. Insurers can change quality by restructuring their network and forming incentive contracts, but the final delivery of quality is rarely in their control. For example, to reduce waiting times for primary care appointments, the insurer can expand its physicians network. However, demand shocks might turn these investments into waiting times higher or lower than what the insurer intended. Accounting for the stochastic nature of quality will be important for the identification of investment costs. It also introduces a moral hazard problem to the designer's problem through unobserved investment, as in Holmstrom (1982).

The Quality Bonus Payment (QBP) program: In addition to the star ratings, CMS also provides direct pecuniary incentives for insurers to improve their scores. The Affordable Care Act of 2010 mandated this system of Quality Bonus Payments (QBP), which took effect in enrollment-year 2012. QBP introduced two changes to the bidding system of insurers. First, it modified the percentage paid in rebates for contracts bidding below the benchmark from a uniform 75% to be increasing in the previous year's star rating. Second, it introduced a *reward factor* that increases the benchmark for plans with a higher previous year rating. Overall, after 2012 the rebate paid to a contract j bidding p_{jt} against a benchmark B_{jt} in year t is given by

$$\text{Rebate}_{jt}(p_{jt}) = \max\{\rho(r_{jt-1})(B_{jt}(1+\tau(r_{jt-1})) - p_{jt}), 0\} \tag{2}$$

where $\rho(r_{jt-1})$ and $\tau(r_{jt-1})$ are the *rebate share* and *reward factor* of a plan j that obtained a rating of r_{jt-1} in year $t-1$, respectively. Table 1 shows the formula for $\tau(\cdot)$ and $\rho(\cdot)$.

There is some discrepancy in the literature about the rating year that enters the QBP calculation. In particular, some have modeled it as depending on the current year (i.e., R_{jt}



(a) Year 2012

(b) Year 2015

Figure 8: Distribution of underlying continuous aggregate star rating

Notes: Every contract within a same-color bin obtains the same rating, denoted on the horizontal axis. The continuous star rating corresponds to the un-rounded component in the main text equation (1).

being a function of r_{jt}). However, two pieces of evidence clarify that this is not the case. First, the rating is required at bidding time as the rebate plays an important role in the bidding regulation. This takes place between June and September of the preceding year ($t-1$). Second, the following year's ratings (r_{jt}) are released mid October together with the open enrollment period. Therefore, it can not be that QBP rules use the same ratings as those used by consumers in the same year.

Star ratings formula and reconstruction from public data: CMS provides six data sets required to compute the star ratings (the names correspond roughly to the names of files within the CMS zipped rating files)

1. **Stars**: The per measure star levels computed by CMS
2. **Data**: The quality measurements for included measures
3. **Display**: The quality measurement for excluded measures
4. **Cutpoints**: The cutpoints used to transform measure-level data to measure-level stars
5. **Summary**: The summary stars of each contract
6. **Domain**: Domain-level stars of each contract

Importantly, two steps require intensive manual work. First, the weights associated with each measure are not provided in the CSV files but have to be recovered from a series of PDF files included in the zipped folders. Second, the direction of improvement is not always clear for each measure from the data. This is particularly true for excluded (display) measures. However, these can be recovered manually by examining CMS communication documents each year, announcing changes to the rating program. Some of the most difficult to pin down are (with their direction of improvement)

1. Call center pharmacy hold time: decreasing
2. Appeals auto-forward: decreasing
3. Doctors who communicate well: increasing
4. MPF composite: increasing
5. Call center calls disconnected when costumer calls drug plan: decreasing
6. Beneficiary access and performance problems: decreasing in 2010-2011, increasing in 2012-2014
7. Call center calls disconnected when calling pharmacist: decreasing in 2009
8. Complaints about enrollment: decreasing in 2009
9. Call center beneficiary hold time: decreasing in 2011
10. MPF price accuracy: increasing in 2013-2014

Some additional considerations have to be taken when processing the files. First, special care has to be taken with the names of measures, as they haven't been consistent over the years. Second, some measures change from being presented as percentages to being equivalent fractions. Third, some quality measures are duplicated among the part C and part D files. Contracts that offer both part C and part D coverage must be adjusted to not double count the measures. Some part D measures are only relevant for pure part D contracts and have to be excluded for MAPDP plans.

The adjustment factor discussed in the main text is composed of two components. First, CMS computes an *improvement star* for contracts that have maintained or improved their quality over the preceding years. The data and formula used to compute this are not always available. Hence I take it as given. As the model developed in this paper is static, this is not particularly troublesome. Second, CMS computes a *reward factor* based on how the weighted average and variance of a contracts' measure-level stars compare with the same values of other contracts. Specifically, letting $\bar{s}_j$ and $WV(s_j)$ denote the weighted average and variance of measure-level stars for contract j, the reward factor is defined as

$$\text{reward}(s_j) = \begin{cases} 0.4 & WV(s_j) < Q30(WV(\boldsymbol{s})) \text{ and } \bar{s}_j > Q85(\bar{\boldsymbol{s}}) \\ 0.3 & Q30(WV(\boldsymbol{s})) \leq WV(s_j) < Q70(WV(\boldsymbol{s})) \text{ and } \bar{s}_j > Q85(\bar{\boldsymbol{s}}) \\ 0.2 & WV(s_j) < Q30(WV(\boldsymbol{s})) \text{ and } Q65(\bar{\boldsymbol{s}}) < \bar{s}_j \leq Q85(\bar{\boldsymbol{s}}) \\ 0.1 & Q30(WV(\boldsymbol{s})) \leq WV(s_j) < Q70(WV(\boldsymbol{s})) \text{ and } Q65(\bar{\boldsymbol{s}}) < \bar{s}_j \leq Q85(\bar{\boldsymbol{s}}) \\ 0 & \text{otherwise} \end{cases}$$

Where $Q30(\bar{s})$ is the 30-th percentile of the average rating among all contract for the year, and $Q65, Q70, Q85$ define analogously.

The continuous star rating of a plan is defined based on five values:

1. $\bar{s}_j$: the weighted average star rating of a plan, across its measures, including the improvement star.
2. $\bar{s}^o_j$: the weighted average star rating of a plan, across its measures, without the improvement star.
3. $\text{reward}(s_j)$: the reward factor computed using the improvement star.
4. $\text{reward}^o(s_j)$: the reward factor computed without using the improvement star.
5. improv_j: an indicator of whether the contract j was awarded an improvement star by CMS.

The continuous star rating of contract j is thus

$$cr_j = \begin{cases} \bar{s}^o_j + \text{reward}^o_j(s_j) & \text{if } \bar{s}^o_j + \text{reward}^o_j(s_j) \leq 1.75 \text{ or } \text{improv}_j = 0 \\ \bar{s}_j + \text{reward}_j(s_j) & \text{if } \bar{s}^o_j + \text{reward}^o_j(s_j) \in (1.75, 3.75) \text{ and } \text{improv}_j = 1 \\ \max\{\bar{s}_j + \text{reward}_j(s_j), \bar{s}^o_j + \text{reward}^o_j(s_j)\} & \text{if } \bar{s}^o_j + \text{reward}^o_j(s_j) \geq 3.75 \text{ and } \text{improv}_j = 1 \end{cases}$$

The overall star rating of contract j is cr_j rounded up to the nearest .5 decimal.

We can map the complicated description above to the simple formula given in equation (1). First, for each measure m define $\psi_m(\cdot)$ as the step function that transforms quality measurements to measure-level stars, multiplied by the relative weight that corresponds to the measure given its category. Second, define the adjustment factor as

$$\omega_j = \begin{cases} \text{reward}^o_j(s_j) & \text{if } \bar{s}^o_j + \text{reward}^o_j(s_j) \leq 1.75 \text{ or } \text{improv}_j = 0 \\ w^{\text{improve}} s^{\text{improve}}_j + \text{reward}_j(s_j) & \text{if } \bar{s}^o_j + \text{reward}^o_j(s_j) \in (1.75, 3.75) \text{ and } \text{improv}_j = 1 \\ w^{\text{improve}} s^{\text{improve}}_j + \text{reward}_j(s_j) & \text{if } \bar{s}^o_j + \text{reward}^o_j(s_j) \geq 3.75 \text{ and } \text{improv}_j = 1 \\ & \text{and } \bar{s}_j + \text{reward}_j(s_j) \geq \bar{s}^o_j + \text{reward}^o_j(s_j) \\ \text{reward}^o_j(s_j) & \text{otherwise} \end{cases}$$

where w^{improve} is the relative weight of the improvement star and s^{improve}_j is contract j's improvement star.

Finally, it is worth pointing out that there are three types of contracts for which CMS does not compute a star rating as above. First are new contracts of organizations that did not offer any contracts in the previous three years. For QBP benchmark bonuses, these were treated as three stars between 2012 and 2013 and as 3.5 stars afterward. For rebate purposes, these are treated as 3.5 stars throughout the years. Second are new contracts of existing organizations. These are assigned the average star rating of the firm's other contracts, weighted by enrollment, in the earliest year possible, up to three years back. Finally, the third group is composed of contracts with low enrollment. For QBP benchmarks, they were treated as having three stars before 2015 and as 3.5 afterward. For rebate purposes, they were treated as 4.5 stars in 2012, 3 stars in 2013-2014, and 3.5 starting 2015.

3.2 Data Construction

This work combines two data sources: public data from CMS, and individual-level data from the Medicare Current Beneficiary Survey (MCBS). Here I describe briefly the cleaning steps used when constructing the data used for estimation. The data used for constructing the quality ratings was discussed previously and omitted from this description.

3.2.1 Public Data: I construct a plan-county-year level panel by combining several publicly available data sets. First, I use public *Enrollment* files to recover the average number of enrollees per plan-county-year. Plans that have fewer than 11 enrollees have a missing enrollment number in this data. When needed for specific computations, I replace this missing value with 6. These plans are small by definition and do not affect any major computations or estimations. I use the *Contract* definition files to recover parent organization names. This requires extensive data-cleaning as names are often misspelled or change across years. I cross-reference these with the public *Landscape* files which also contain organization names. Using these files to also construct the core of the panel, as the landscape files list plans by county with their premiums, deductibles, and contract numbers. I drop a small fraction of plans that disappear between September of the previous year and March of the current. These are likely to be plans removed by CMS due to regulation issues, or plans that failed to enroll any consumer. I also remove part-D only plans. I combine the landscape file with the *Plans* files to incorporate additional details, including the star ratings. I use the *Penetration* files to add information regarding the population of each county, including the number of Medicare eligibles and total number enrolling in MA. I also add information from the *Dual Eligibles* files to count the number of dual Medicare/Medicaid eligible population per county-year. I use the *Payment* files to add information about total payments to each plan from CMS in subsidies and rebates. I derive the local benchmarks for each plan using the *Ratebook* files. I use the detailed plan-level *Benefits* files to add information on each plan's cost-sharing attributes. Using the detailed *Bid* data I add each plan's bid, rebate, premiums, benchmark, share of rebates allocated to each premium reduction, and benefits.

Using this detailed panel I exclude Special Needs Plans, dual Medicare/Medicaid plans, and those that are offered exclusively to some employers and not the common market. I also exclude a small fraction of plans that were not approved to operate in a county despite submitting a bid. I adjust all dollar values according to the 2015 Medical CPI. Table 9 presents descriptive statistics of the resulting panel by star rating and Table 10 by aggregation level.

3.2.2 MCBS: The MCBS is provided as a collection of files in different formats for different years. This data poses several challenges. First, column name, placement of variables among files, and categorical variable definitions are not consistent across years. Second, some variables exist in the SAS version of files and not in the flat text files and vice-versa. This involves a lengthy manual work of linking columns across years, which I did by starting from a crosswalk provided by the MCBS for linking 2015 and 2013 data. I automated a process which reads the SAS description files, recovered the format of each column, applying it directly to categorical variables in the flat text files. This also flagged missing columns which I recovered separately.

I merged the MCBS individual-level data with the public plan-level data using the contract and plan names when avaialble. For 2009-2011 and 2015, the MCBS include the part C plan name which provides an accurate match. In 2012-2013 I used the part D contract number together with an "enrolled in MA" indicator. As most MA plans bundle part D coverage, these contract names are almost always MAPDP contracts. 731 out of 12392 beneficiary-years have more than one contract number in the data. For those, I keep the first contract (chronologically) that is also observed in the aggregate data. Overall, 98.9% of single contract consumers have a match in the aggregate data and only 9 beneficiary-years with more than one contract have no match in the aggregate data. This results in 29312 beneficiary-years to match at the plan level. 15088 of those have more than one plan associated to their contract. I first match by premium, matching 7858. 1477 are then matched based on dental benefits, and 9 on eye benefits (both of these columns exist in the MCBS and the aggregate data). The remaining 3207 beneficiary-year are assigned to the most popular plan within their contract-year-county. Table 6 presents descriptive statistics of the resulting rotating matched panel.

4 Demand and Quality Responses to Scoring

This appendix section provides additional analysis, detail, and supportive evidence for the excercises of Section 4 of the main test.

4.1 Demand Responses

Understanding the design. The Star Rating program's design consists of hundreds of cutoffs and weights, collapsing nearly sixty dimensions of quality. It is improbable that consumers understand all of this complexity because the details are not readily available, and the system does not reveal the quality measurements to them.[8] However, aspects of the design make it well approximated by only a few parameters in category space. This point is presented in detail in Appendix Section 6.1. Under this representation, the rating design consists of regions within a space. Main text Figure 5 shows this for two dimensions. In the case illustrated, the rating consists simply of the slope of the two lines and their position.

Consumers have a variety of sources of information to develop an understanding of those slopes and locations. The slope consists of the relative contribution categories to the rating, determined by their overall weight. The simplifying feature is that measure weights are nearly uniform within a category (3 for outcome and intermediate, 1.5 for patient and access, and 1 for process). Therefore, it suffices that consumers know the category-level weight and the number of measures in each category. The latter point is simple, as CMS sends each member a booklet called *Medicare & You* which contains some of this information (shown in Figure 6(a)). Different information materials are available to consumers that can complement the previous and serve to understand the weight. First, consumers can click to see the rating details of a plan and get a category-level score, shown in Figure 6(b). This additional information is not enough to determine quality at that level but is sufficient to form an idea of the contribution of each category. Second, large changes are covered in news articles and discussed in promotional material by insurers. Together consumers can use these sources to understand the weighting system.

The cutoff positions can be derived by combining the previously discussed information with preexisting knowledge of the market. MA consumers are either retirees or disabled, which means they have likely had significant interaction with their local health care markets. Therefore, quality uncertainty stems mainly from network arrangement over a known distribution of quality. Nevertheless, by looking at the relative distribution of ratings in her market and understanding the available quality of providers, a consumer can form an understanding of cutoffs. For example, if only a scanty fraction of plans have five stars but the consumer knows that there are many high-quality providers, it stands to reason that the five-stars cutoff is very high. Suppose the same happens, but there are many low-quality providers and one medium-quality provider. In that case, the good plans likely have the medium quality provider in their network, and the standard is low.

More granular ratings. Clicking on the details of a plan's rating in the Plan Compare platform shows a more granular score. However, I assume that consumers largely use these scores to derive some knowledge of the overall system instead of comparing them across plans. There are several reasons for this assumption. First, historically CMS used to present more granular ratings to consumers before 2009. The old system is described as offering consumers a

[8] Quality measurements are publicly available for research, but some computational effort is required in parsing the data. The rating details are even more obscure and require reading hundreds of pages of documentation. See Appendix Section 3.1.1.

click-through, or additional material, providing them with category-level scores for each plan. However, there appears to have been a consensus that consumers either failed to understand the detailed rating or found it too difficult to make the multidimensional comparison across plans. Because of this, CMS opted for constructing a summarized score for consumers.

The second reason is that the granular rating is still not transparent. To understand this score, consumers would need to understand something about the measure-level design, which is more demanding than the category-level representation. The third and final reason is that industry participants believe that the click-through rate for the additional information section is low.[9] Hence, it is possible that a consumer clicks a couple of details and infers something about the design, but it is unlikely that they compare measure by measure.

This assumption has testable implications. In particular, it implies that, conditional on the rating-year, consumer choices should not vary with the granular score. To validate it, I modify the regression of Section 4 to be

$$y_{ijt} = \boldsymbol{\alpha}' \boldsymbol{sc}_{jt} + \eta_{rt} + \gamma_{c(j)} + \mu_{m(i)} + \xi_t + \boldsymbol{x}_{ijt} \boldsymbol{\lambda} + \epsilon_{ijt}$$

Where $\boldsymbol{sc}_{jt}$ is a vector of measure-level scores for plan j, and η_{rt} is a year-rating fixed-effect. Table 8 present the estimated effect using the full set of measures scores. The results show that very few measure-level scores have a statistically significant effect on demand once the rating-year is included. Among those that appear relevant, complaints and access problems might be correlated with choice due to word-of-mouth information. Diabetes care might be a case where some consumers are willing to pay the cost of comparing at this granular level across conditions. However, as the main goal of the rating is to generate additional quality for consumers, and insurers are risk-adjusted, likely, additional information for diabetics is not of first order importance.

The CMS system also provides consumers with a slightly more aggregate definition of scores. These are often called "domain-level" ratings, falling into categories such as "keeping the population healthy". The definitions of these domain level scores have changed over time, and their data is scarce. However, CMS built these by aggregating scores over categories. Therefore, to test whether consumers used these additional sources of information, I repeat the previous exercise, substituting measure-level scores with category-average scores. The first column of Table 5 shows the resulting estimate. The only column with a statistically significant coefficient is the patient-level category. As the effect is small and only on the patient level, it is likely caused by word-of-mouth information.

A different alternative is that consumers compare ratings in levels and as a relative ranking within their market. For example, a consumer seeing a single 5-star plan and many 3-star plans might infer something more about the 5-star than just the scoring regions. To test this, I

[9]For this point, I interviewed the MA Star Analytic team manager at one of the largest insurers.

compute a local ranking for each product, counting the number of alternatives in each market with a strictly lower score. The average resulting ranking in the data is about five. I run the same regression as before but replacing measure-level scores with this local ranking. The second column of Table 5 shows the estimated coefficient. The result indicates a small and nearly insignificant effect.

I find these estimates supportive of the modeling choice of focusing on the aggregate level scores as the only source of quality information. While there is a potential for other sources such as word-of-mouth, and some role for consumer exploration of the detailed score, it appears that the average enrollee is not significantly affected by these behaviors. Therefore, these considerations do not appear of first-order importance for the scoring design problem at hand.

4.2 Quality Responses

Pre-trends and effects over time. To study the pre-trends and the scoring effects over time, I modify the triple-difference design of Section 4 to have only two treatment groups of pre-introduction quality. I consider two possible definitions: by quartiles of quality, comparing the second to the third quartile; by groups of predicted measure-level scores, comparing those predicted to receive a score of two or three to those predicted to receive four. The latter definition is closest to the main analysis, while the former depends less on the endogenous choices of the regulator. In both cases, the highest and lowest qualities have been truncated to deal with the bounded domain issues raised in the main text. The first two columns of Table 15 show that replacing the treatment group definition in the main text's triple-difference regression with either of these definitions delivers a similar result. The third and fourth columns show the effect over time, as estimated by the regression:

$$\underbrace{y_{kjt}}_{\text{quality}} = \sum_{\tau=-4}^{3} \beta_\tau \underbrace{G_{kj}}_{\substack{\text{treatment}\\ \text{group}}} \underbrace{\mathbb{1}\{\mathcal{T}_{kt} = \tau\}}_{\substack{\text{years to}\\ \text{Introduction}}} + \underbrace{\gamma_{kj} + \mu_{jt} + \xi_{kt}}_{\text{pairwise fixed effects}} + \epsilon_{kjt}$$

The advantage of this simplification is that pre-trends can be easily explored visually in Figure 9. As shown, the difference between the groups is not statistically significant before the introduction and is afterward. Thus we can not reject that the groups were trending similarly before treatment.

Staggered Difference-in-Differences. Recent research in Econometrics has raised concerns about staggered difference-in-differences designs like the one used above (Goodman-Bacon, 2021; Callaway and Sant'Anna, 2020; Baker et al., 2021). However, the structure of the treatments used in this work is somewhat different from the canonical example used in this literature. Unlike standard staggered differences-in-differences, where the same treatment is assigned to different units over time, measure entry can be seen as different treatments assigned to different

Table 15: Quality Response to Scoring, alternative specification

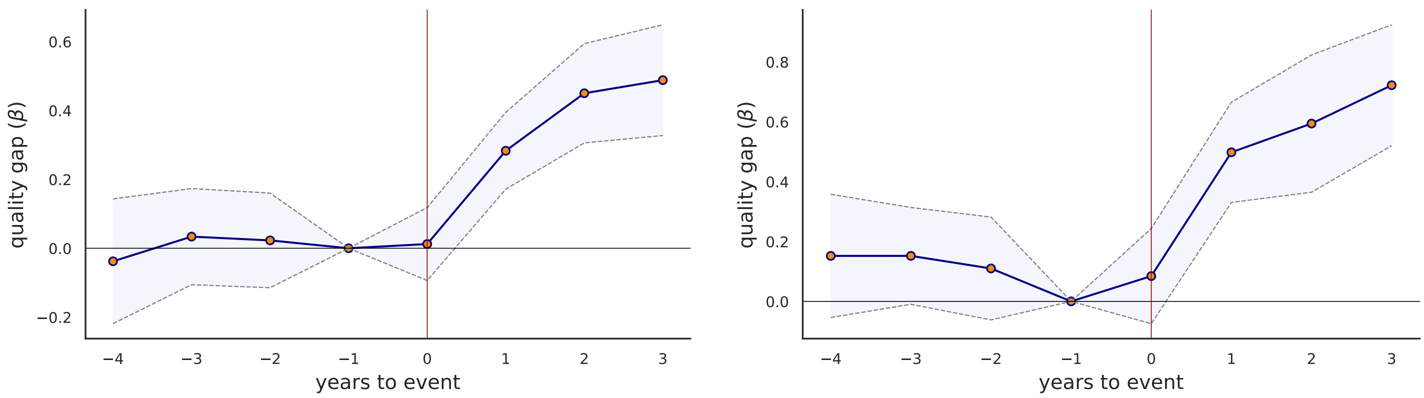
	Aggregate				Over time			
	Medians		Predicted score		Medians		Predicted score	
Treatment group	0.142***	(0.035)	0.143**	(0.0431)				
-4					-0.0379	(0.0924)	0.152	(0.105)
-3					0.0338	(0.0712)	0.152	(0.0825)
-2					0.0228	(0.0701)	0.110	(0.0876)
-1					0	(.)	0	(.)
0					0.0122	(0.0540)	0.0845	(0.0811)
1					0.283***	(0.0569)	0.498***	(0.0854)
2					0.450***	(0.0735)	0.594***	(0.117)
3					0.488***	(0.0820)	0.722***	(0.103)
N	170129		168972		39206		4954	
R^2	0.695		0.693		0.676		0.566	

Notes: The sample sizes differ between columns because the "Over time" design restricts the sample to contract-measures observed at least two years before and after introduction. The "Predicted score" definition is stricter because it also requires a score between 2 and 4 at introduction. Additionally, a 20% random subsample of non-treated contract measures are added to help capture the contract-year fixed effect. The magnitudes of the "Over time" coefficients are larger because the normalized overall "treated" effect is -0.176 (standard error 0.0736). Standard errors are clustered at the contract level.

units. However, regardless of this distinction, the concern remains that by aggregating the effect of different treatments, the pooled regressions used in the previous subsection and in the main analysis might deliver a biased estimate of the treatment effect. To alleviate this concern, I conduct two robustness checks. The first is the event study analysis done in the previous subsection. The data structure behind this event study falls within the *stacked regressions estimators* category (Baker et al., 2021). This sort of analysis has been used, for example by Jung et al. (2021), to show robustness to dynamic treatment effects. Second, Table 7 shows the Differences-in-Differences coefficient for each entering measure, run separately from the rest. As can be seen, all but one of the coefficients are positive, as the main estimate. The one negative coefficient is extremely small and statistically insignificant and is therefore unlikely to bias the main result of this section.

5 Model

This section provides additional details about the model that have been omitted from the main text for ease of exposition.



(a) By pre-introduction quartile

(b) By pre-introduction predicted measure-score

Figure 9: Quality response to scoring

Notes: These figures show the estimated coefficient of the differences-in-differences analysis of supply responses to scores. In both cases plans of low pre-introduction quality are compared to those of high quality. Figure (a) shows the comparison of plans in the second quartile of pre-introduction quality relative to those in the third quartile. Figure (b) shows those predicted to receive a score of two or three in the measure-score relative to those predicted to receive four. In both cases the coefficient on the year before treatment has been normalized to zero. The solid line depicts the estimated coefficient, while the shaded area is the 95% confidence interval. For further details see Section 4.

5.1 Demand

The total premium consumers pay is a sum of several factors, taking the form:

$$p^{\text{total}}_{ijmt} = \underbrace{p^B_i - PR^B_{jmt}}_{\text{part B}} + \underbrace{p^C_{jmt}}_{\text{part C}} + \underbrace{p^D_{jmt} - PR^D_{jmt}}_{\text{part D}} \tag{3}$$

Where p^B_i is the consumer's part B premium, p^C_{jmt} is the MA-specific (or part C) premium, p^D_{jmt} is the premium associated with bundled prescription drug coverage, if they exist. The part B premium is an income-adjusted amount required to enroll in Medicare independent of coverage choice. Plans can opt to allocate rebate dollars to reduce their part B and part D premiums, which are denoted by PR^B_{jmt} and PR^D_{jmt}. The next section indicates how total premium relates to the plan's price.

5.2 Supply

5.2.1 *Insurers' pricing problem:*

Plan j in market m at year t obtains an additional revenue given by

$$R(p_{jmt}, z_j t) = \underbrace{p^D_{jmt}}_{\substack{\text{part D}\\ \text{premium}}} + \underbrace{R_{jmt}(p_{jmt})\kappa^e_{jmt}}_{\text{extra rebate}}$$

CMS requires plans to allocate their rebates among benefits or premium reductions. Each plan determines a fraction of rebates to allocate to part B premium reductions (κ^b_{jmt}), part D reductions (κ^d_{jmt}), extra benefits (κ^e_{jmt}), and increasing consumers' coverage on standard Medicare-covered health care services ($1-\sum_{l\in\{e,d,b\}}\kappa^l_{jmt}$). Because premium reductions are either payments directly to CMS or a transfer within the firm, they do not appear as added revenue. CMS strictly regulates improvement to standard Medicare coverage, requiring insurers to submit cost assessments based on CMS utilization models which are certified by actuaries. Because of this, I assume that plans offer these additional benefits at cost. The only remaining free source of rebate revenue is the κ^e_{jmt}, which appears in the profit. I assume that in the short run, each plan's rebate allocation fraction is an exogenous feature.[10]

Having allocated the rebates, a plan's benefits and total premium are expressed as

$$b_{jmt} = b_0 + \bar{b}_{jmt} + \kappa^e_{jmt} R_{jmt} \tag{4}$$

$$p^{\text{total}}_{ijmt} = \underbrace{(p^B_i - R_{jmt}\kappa^b_{jmt})}_{\text{part B}} + \underbrace{(p^D_{jmt} - R_{jmt}\kappa^d_{jmt})}_{\text{part D}} + \underbrace{(\max\{p_{jmt} - B_{jt}, 0\})}_{\text{MA basic}} + \underbrace{\bar{b}_{jmt} - R_{jmt}(1 - \sum_{l\in\{e,d,b\}} \kappa^l_{jmt})}_{\text{MA supplementary}} \tag{5}$$

Where $\bar{b}_{jmt}$ is the additional Medicare cost-sharing benefits level. TM's standard coverage, b_0, is present in equation (4) as CMS requires MA plans to offer an actuarial value at least equal to TM. The Part C premium (p^C_{jmt}), is decomposed into its two ingredients: the MA basic premium and the MA supplementary premium. As described in Section 3, the latter constitutes the majority of the MA-specific premium. These equations reveal the effect of benchmarks on premiums. Above the benchmark, a dollar increment in prices creates an equivalent increase in premiums, while below it, the same is equivalent to only $\rho_{jt}(1 - \kappa^e_{jmt})$ additional dollars of premium.

This pricing model deviates from the setting presented in the Section 3 in two ways. First, it reduces the complex pricing problem to a single choice variable per plan-market. This simplification is mostly innocuous as whether insurers allocate rebates to one premium or the other is irrelevant from a welfare perspective. Moreover, changes to the scoring regulation are unlikely to alter the relationship between premiums and benefits systematically. The second deviation of the model is in assuming that pricing occurs at the market level. This assumption is standard in the literature and is likely to be of minor consequence because plan segmentation is unrestricted. The alternative of setting up prices at the plan level would imply insurer markets are national, vastly overstating the degree of competition in MA.

[10]For the 13% of plans without rebates in the data, I assume that all counterfactual rebates would go to cost-sharing standard benefits. This assumption's effect is minimal on consumer premiums and firm revenue and does not affect this stage's estimates.

5.3 Discussion

This section shows the model I develop and estimate in the main text is flexible enough to generate key market behaviors. I focus on those highlighted by my results: inefficiency of quality provision, and optimal scoring granularity. To highlight the mechanisms, I simplify the model to ignore subsidies, investment risks, multidimensional quality, and multiproduct firm incentives. I discuss the implications of incorporating these model components at the end of this section.

5.3.1 Simplified model: I simplify the utility model for consumers' choices, presented in the main-text equation (3), to be:

$$u_{ij} = \alpha_i p_j + \gamma \mathcal{E}[q|r_j, \psi] + \xi_j + \epsilon_{ij} \tag{6}$$

$$u_{i0} = \epsilon_{0j} \tag{7}$$

Relative to the original model, this expression omits market-year indices, absorbs all non-price and non-quality attributes in the single term ξ_j, and assumes consumers' preferences for quality, $v(q)$, are linear and equal to γq. I also normalize the utility of the outside option to zero, and eliminate the distinction between consumers' premiums (p^{total}_{jmt} in the main text) and insurers' prices.

Ignoring investment risk and focusing on single product firms, the insurer owning plan j chooses quality and prices to maximize

$$\pi_j(\boldsymbol{q}, \boldsymbol{p}, \psi) = D_j(\boldsymbol{q}, \boldsymbol{p}, \psi)(p_j - \theta' q_j - c_j) - \mu_j q_j^2 \tag{8}$$

In contrast to the original model of main-text equation (6), this expression ignores risk adjustments, multidimensional quality, assumes marginal quality costs are linear, and investment costs are quadratic.

5.3.2 Monopolistic quality provision under full information: I begin by discussing the conditions under which a monopolist will over- or underprovide quality in full information, relative to the social optimum, as in Spence (1975). In this scenario, consumers' beliefs satisfy $\mathcal{E}[q|r_j, \psi] = q_j$ and the monopolist's optimality condition for quality are

$$[p_j]: \quad \frac{\partial D_j(q_j, p_j)}{\partial p_j}(p_j - \theta' q_j - c_j) + D_j(q_j, p_j) \qquad = 0$$

$$[q_j]: \left(\frac{\partial D_j(q_j, p_j)}{\partial p_j}\frac{\partial p_j}{\partial q_j} + \frac{\partial D_j(q_j, p_j)}{\partial q_j}\right)(p_j - \theta q_j - c_j) + D_j(q_j, p_j)\left(\frac{\partial p_j}{\partial q_j} - \theta\right) - 2\mu_j q_j = 0 \qquad = 0$$

Combining the first with the second, the firm's optimal quality decision is characterized by

$$D_j(\frac{\partial D_j(q_j,p_j)/\partial q_j}{|\partial D_j(q_j,p_j)/\partial p_j|} - \theta) - 2\mu q_j = 0$$

To match the analysis of the main text, I define efficient quality production as being total-welfare maximizing given ensuing monopolistic prices. That is, q^w is efficient if it satisfies

$$q^w = \arg\max_q \underbrace{\int \frac{1}{\alpha_i}\ln(1+\sum_{j\in J}\exp(-\alpha_i p_j^*(q)+\gamma' q+\xi_j))dF(\alpha_i)}_{\text{Consumer surplus}} + \underbrace{\pi_j(q,p_j^*(q))}_{\text{Firm profit}} \tag{9}$$

Where $p_j^*(q)$ is the monopolist's optimal prices given quality q.

I say that quality is over-provided by plan j if $q_j > q_j^w$, under-provided if $q_j < q_j^w$, and efficient otherwise. This notion of efficiency is motivated by the scoring design literature, that shows scores primarily control quality (Zapechelnyuk, 2020). Therefore, whether a scoring rule is better than a fully informative one depends mainly on its ability to shift quality production towards levels that increase welfare while allowing firms to price at will.

Combining the monopolist's optimality conditions with the total-welfare objective, we can see that the monopolist's choice of quality would be efficient only if it satisfies

$$\int D_i(\frac{\gamma}{\alpha_i} - \frac{\partial p}{\partial q})dF(\alpha_i) + D(\frac{\partial D/\partial q}{|\partial D/\partial p|} - \theta) - 2\mu q = 0$$

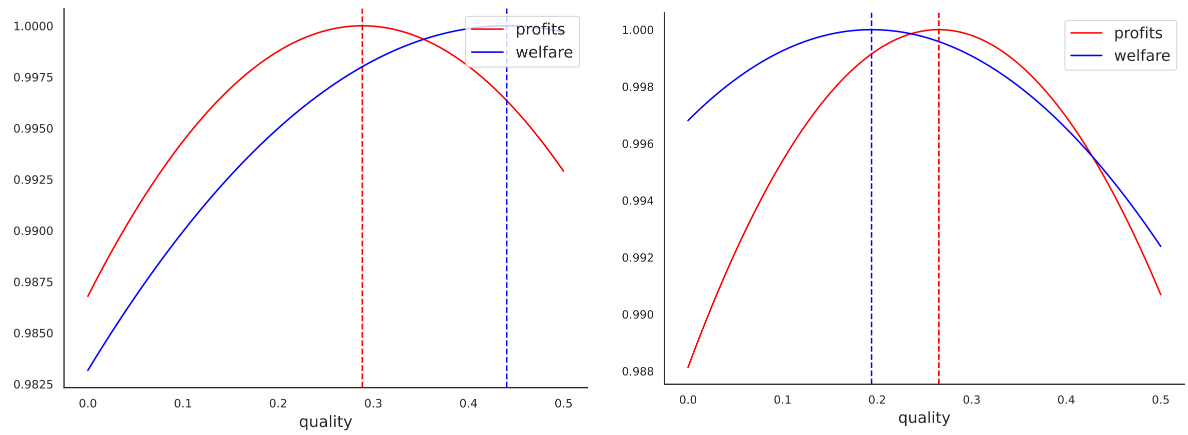
The last two terms equal to zero by the optimality of the firm's actions. Therefore, if $\frac{\partial D/\partial q}{|\partial D/\partial p|} > \theta$, then quality would be overprovided if

$$\frac{1}{D}\int D_i \frac{\gamma}{\alpha_i} dF(\alpha_i) > \frac{\partial p}{\partial q} \tag{10}$$

, efficiently provided if this condition holds with equality, and underprovided if the inequality is reversed. On the left, $D_i\frac{\gamma}{\alpha_i}$ is consumer i's valuation for a marginal increase in quality, measured in units of premiums and weighted by her choice probability. Thus, the left-hand side of this equation is the weighted average valuation for marginal quality increases. On the right, we have the increase in price associated with a marginal increase in quality.

Equation (10) is a direct analog to the condition derived by Spence (1975) in his famous first proposition, although with two minor conceptual difference. First, in Spence's proposition the right-hand side is the derivative of the inverse demand function with respect to price. Here, it is the derivative of the firm's pricing function. Second, Spence's proposition holds constant the quantity sold, which is the monopolist's second control variable in his model after quality (x in his framework). This conditioning means that prices and quality vary in his analysis, keeping the product's demand constant. Here, a variation in quality can change both the price and the demand for products, making this result comparable to Spence's second proposition, which allows quantity to change in response to quality. My result thus shows that Spence's first and second propositions can lead to similar conditions.[11]

To illustrate how this condition can lead to different efficiencies of quality production, I simulate two markets that differ only in their distribution of price preferences but generate opposing quality outcomes. In both markets, consumers' quality preferences are $\gamma = 0.5$ and the firm's cost components are $\theta = 0.2, \mu = 0.1$ and $c = 0$. Consumers' price preferences take on one of three values $\alpha \in \{1.0, 1.5, 3.0\}$. In the first market, shown in Figure 10a, these values are distributed with probability $[0.2, 0.6, 0.2]$, and the monopolist underproviding quality. In Figure 10b, price preferences are distributed with probability $[0.4, 0.2, 0.4]$, respectively, and the monopolist overprovides quality. To further illustrate this effect, Figure 11 shows how condition (10) changes with the distribution of price preference. I take the distribution of price preferences as $[\frac{1-x}{2}, x, \frac{1-x}{2}]$ and let x vary from one to zero.



(a) Underprovision

(b) Overprovision

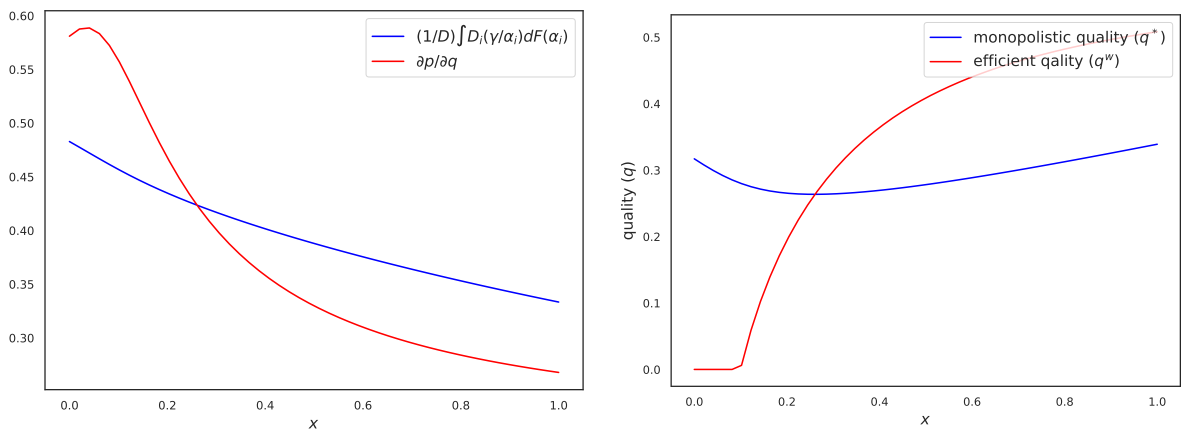
Figure 10: Efficiency with heterogeneous price preferences

Notes: The figures above display scenarios of under and overproduction in markets with heterogeneous price preferences. The common configuration across the two plots is $\gamma = 0.5, \xi = 2.0, \theta = 0.2, \mu = 0.1, c = 0$. In the case of underproduction, the price preferences are distributed with mass points $\alpha \in \{1.0, 1.5, 3.0\}$ and probabilities of $[0.2, 0.6, 0.2]$, respectively. In the case of overproduction the mass points are the same but the probabilities are $[0.4, 0.2, 0.4]$, respectively. Welfare and profit have been normalized to have a maximum of 1 such that they could be displayed in the same scale.

5.3.3 Oligopolistic quality provision under full information:

The analysis above extends to markets with more than one firm. In general, the quality of plan j is efficient if it

[11] Implicitly at play here is the Envelope theorem. As prices are also set to maximize profit, fixing the firm's second choice variable (prices here, quantity in Spence (1975)), or letting it respond to quality does not alter the overall efficiency of quality.



(a) Components of the Spence condition

(b) Quality

Figure 11: Spence condition with heterogeneous price preferences

Notes: The figures above display scenarios of under and overproduction in markets with heterogeneous price preferences. The common configuration across the two plots is $\gamma = 0.5, \xi = 2.0, \theta = 0.2, \mu = 0.1, c = 0$. Price preferences are distributed with mass points $\alpha \in \{1.0, 1.5, 3.0\}$ and probabilities of $[(1-x)/2, x, (1-x)/2]$.

satisfies

$$\left(\sum_{j' \in J} \frac{\partial CS(\boldsymbol{q}, \boldsymbol{p})}{\partial p_{j'}} \frac{\partial p_{j'}}{\partial q_{jk}} + \frac{\partial CS(\boldsymbol{q}, \boldsymbol{p})}{q_{jk}}\right) + \sum_{j' \in J} \frac{d\pi_{j'}(\boldsymbol{q}, \boldsymbol{p})}{dq_j} = 0 \tag{11}$$

Using simulations, it is easy to show that the model can generate different efficiencies. Figure 12 shows simulations for a duopoly market where consumers preferences are $\alpha_i \in \{1, 1.5, 3\}$, distributed with probability $[(1-x)/2, x, (1-x)/2]$, and $\gamma = 0.5$. Consumers have a fixed unobserved preferences for firm 1 given by $\xi_1 = 2$, while their preference for the second firm, ξ_2, varies in the simulation. Firms have equal cost functions ($\theta = 0.2, \mu = 0.1, c = 0$). In the simulations, firm 1 always underprovides quality relative to the social optimum but firm 2, might over or under provide depending on consumers' preferences. In particular, it overprovides quality when consumers' willingness to pay for quality is more heterogeneous and when they dislike the product more.

5.3.4 Quality provision under incomplete information: Generating scenarios of an excessive or insufficient quality provision under incomplete information is trivial. Following the intuition of the monopoly regulation exercise of Section 2, quality efficiency is largely influenced by the optimality of the scores. To illustrate, I replicate the scenario shown in Figure 10b when the regulator uses a simple certification to disclose quality. Consumers are Bayesian and update an exponentially decaying prior. Under complete information, this market features overprovision of quality. Figure 13 shows that depending on the certification threshold, the market can feature excessive, insufficient, or efficient quality.

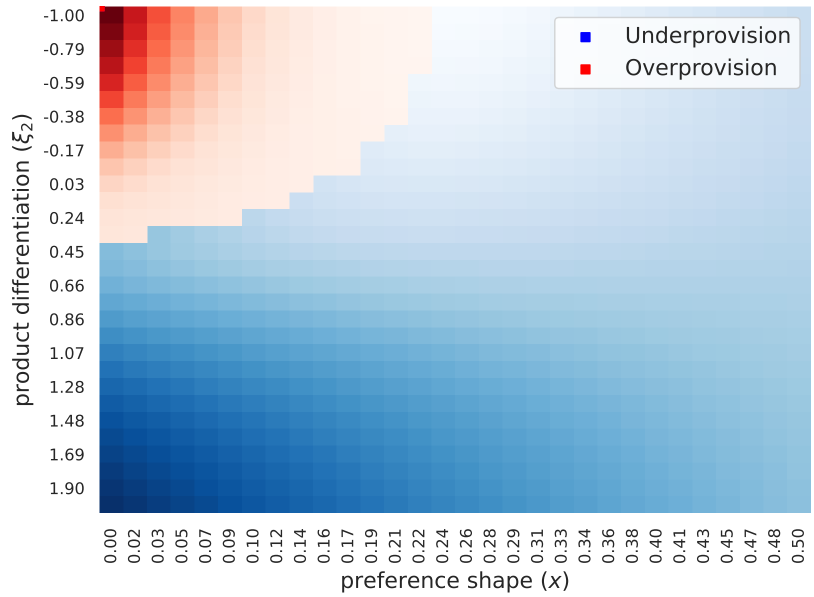


Figure 12: Quality efficiency in duopolistic markets

Notes: The figure above illustrates the efficiency of quality production in a duopolistic market. Red blocks indicate areas where the second firm overprovides quality, while blue blocks indicate underprovision. The intensity of color marks the welfare loss from this inefficiency, with lighter tones indicating less welfare is lost from quality distortions.

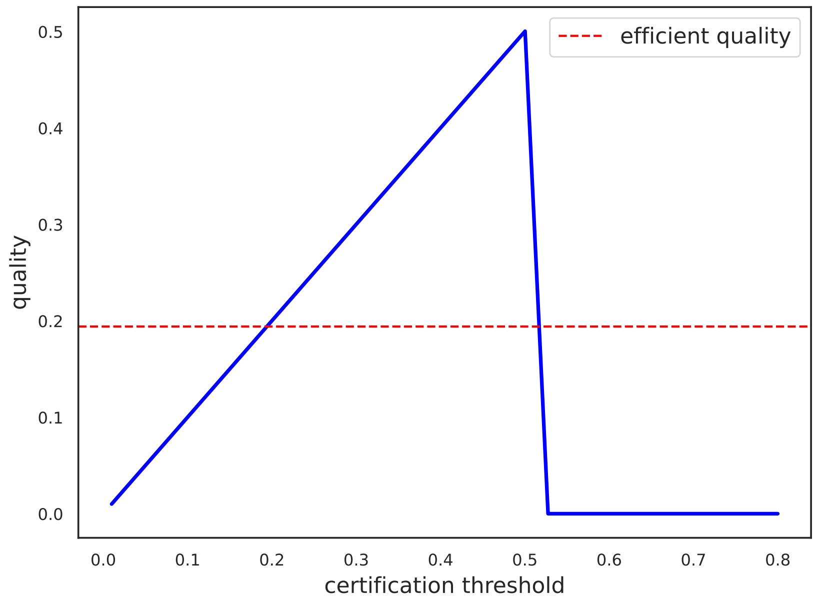


Figure 13: Quality provision in certification market

Notes: The figure presents the monopolistic quality given a certification scoring design. The sharp drop occurs when the monopolist is better not investing at all rather than attaining the certification. Consumers' priors are an exponential distribution with scale parameter 0.3, truncated at 1.

5.3.5 Discussion of simplifications: The preceding analysis simplified some of the model components. First, I assume consumers' quality utility is linear, marginal costs are linear, and investment costs are quadratic. I make the same assumptions when estimating the model, so these do not distort how the analysis compares to the work done in the main text. Second, I ignore risk adjustment and subsidies. These will make firms more likely to overprovide quality as they can offload part of the cost to the government. Third, I ignore investment risk. As firms' profits drop sharply to the left of scores cutoffs, firms with risk will tend to invest at values that are, in expectation, higher than the cutoff itself. This induced risk behavior is verifiable in the main estimates but plays a small role in the overall result. Investment risk is an essential institutional feature and serves to smooth an otherwise non-differentiable problem for firms. It does not appear to affect the scoring designs I find significantly. Finally, I ignore multiproduct firm behavior. This likely plays a role in pricing decisions and potentially affects the scoring design problem. However, it is unlikely that these motives would overturn the model's flexibility in generating over or underprovision or uniquely determine the optimal scoring design.

6 Identification and Estimation

This section presents additional details regarding the implementation of the model, its estimation, and its identification.

6.1 Category-level Rating Rule

The rating rule used by CMS, described in detail in Section 3.1.1, is too complicated for practical use. The primary reason for this is that there are too many dimensions of quality that vary over time. Using it in its full form would require identifying consumers' beliefs and insurers' investment cost for each measure of quality. This task is both untractable and likely not an accurate representation of reality. For example, it seems implausible that an insurer can invest in improving the rate at which physicians review their patients' medication while decreasing the rate at which the same physicians assess the pain of their patients. For similar reasons, CMS groups measures into related categories. Two measures that belong to the same category are given the same weight in constructing the star rating. Therefore, it is useful and natural to approximate the rating system by category-averages. This section describes one approach to this, which I adopt in the analysis of the main text.

I build the approximation based on the following observation. The star rating of a contract is the weighted sum of many step functions, rounded to nearest half-star. Therefore, a smooth function could approximate each step function and have most of its fitting error rounded out. As the weights are category-specific and the distribution of quality within a contract is similar across firms, we could use a single smooth function per category to map average category-level investment to total category score. Thus, the approximation procedure follows this intuition:

1. For each category-year, fit a bounded polynomial that maps the total sum of quality for a category onto the total sum of stars. The bounds guarantee that predictions with the polynomial never escape the logical range of sum of stars for the category. Figure 14.a shows the second-degree polynomial fitted for intermediate category rating in 2015.

2. Next, determine the systematic component of the dispersion around the fitted value. The idea of this step is to capture small systematic differences between firms in the distribution of their quality across measures. This is done by regressing the residual on indicators for plan-type, state, firm, and year. I also include the number of measures in the category, bonus rating, and improvement stars. The predicted value is added as a dispersion adjustment. These values are relatively small and serve to reduce some of the noise introduced by a few outliers. Figure 14.b shows the effect of this adjustment.

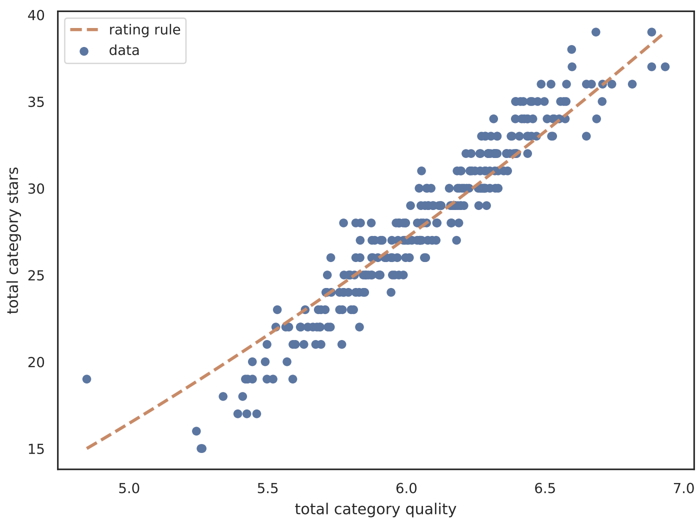
3. Re-calibrate the weight of each category to maximize the fit. This is done by solving a constrained least square, where the constraint is that weights are positive. Figure 14.c shows the resulting distribution of ratings contrasted with the observed ones.

Comparing the predictions of this procedure with the data for each year, the R^2 ranges between 0.91 and 0.946. The maximum absolute error of the model is only half a star, and 78% of plans get the same star rating as in the data. This remaining error is added to the adjustment factor, such that in the baseline, the model predictions are exact.

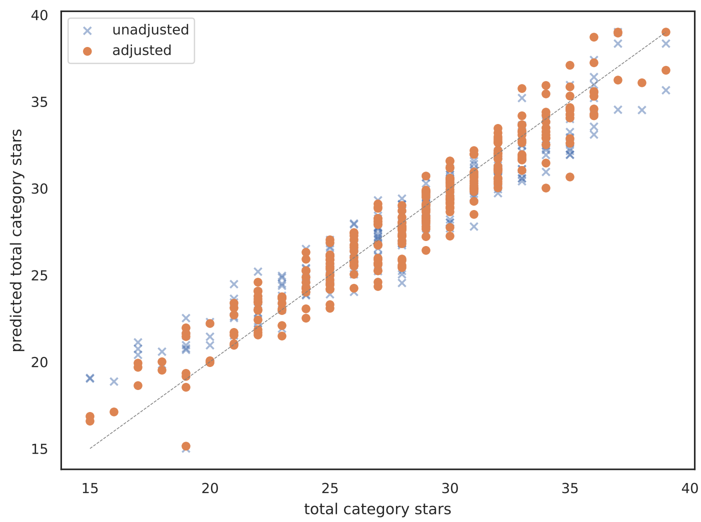
This approximation rule is only used three times in the paper. First, I use it when estimating investment costs. The error introduced by the approximation there is fairly small as the rule is exact in the baseline and only marginal changes are considered. Second, I use it when estimating consumers' beliefs and preferences for quality under the informed choice assumption. There, this approximation is used to compute the domain of quality that can achieve each rating in the baseline. Here, again it is mostly harmless and can be thought of as considering the range of the average category-level quality per rating. Finally, I use it to simulate the baseline in the counterfactual analysis of the scoring design. This is where it is truly leveraged, as the code requires the ability to compute scores under the baseline rule for counterfactual quality outcomes.

6.2 Demand

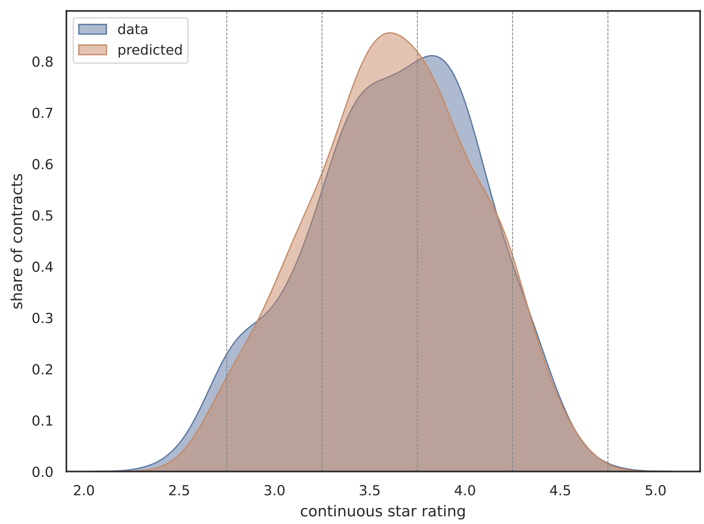
This section provides implementation details for the demand estimator of Section 6.1. The estimator consists of two stages, plus an additional stage for recovering quality preferences and beliefs under the informed choice assumption. I begin by rewriting consumers' utility in a way that facilitates estimation. To do so, I normalize the mean utility of TM to zero and



(a) Step 1: Intermediate 2015



(b) Step 2: Intermediate 2015



(c) Step 3: Overall 2015 rating

Figure 14: Scoring rule approximation

Note: These figures illustrate the first two step of the scoring rule approximation used to reduce the complexity of the problem.

separate mean from individual preferences (e.g., $\alpha_i = \alpha + \tilde{\alpha}_i$), resulting in the modified utilities:

$$\tilde{u}_{ijmt} = \tilde{\alpha}_i(p^{\text{total}}_{ijmt} - p^B_i - p^D_{0mt}) + \tilde{\beta}_i(b_{jmt} - b_{0mt}) + \boldsymbol{\lambda}^d dem_{it} + \boldsymbol{\lambda}^l \boldsymbol{l}_{ijt} + \delta_{jmt} + \varepsilon_{ijmt} \tag{12}$$

$$\delta_{jmt} \equiv \alpha(p^{\text{total}}_{ijmt} - p^B_i - p^D_{0mt}) + \beta(b_{jmt} - b_{0mt}) + \boldsymbol{\lambda}^a \boldsymbol{a}_{jmt} + \eta_{c(j)t} + \mathbb{d}_{mt} + \tilde{\xi}_{jmt} \tag{13}$$

Within the plan-county-year mean preference δ_{jmt}, I have decomposed the sum of the quality-utility and unobserved preferences $(\mathcal{E}[v(\boldsymbol{q})|r_{jt}, \psi_t] + \xi_{jmt})$ into three components $\eta_{c(j)t} + \mathbb{d}_{mt} + \tilde{\xi}_{jmt}$. The first component is the contract-year specific preference $\eta_{c(j)t}$, which captures both quality and any time-invariant contract-specific unobserved taste. The second, $\mathbb{d}_{mt}$, is the market-year preference for MA relative to TM, which captures the variation that exists across counties in the quality of service provided by TM. Finally, $\tilde{\xi}_{jmt}$ captures the residual variation in plan preference that is neither contract nor market specific. Therefore, the key equations associated with each step are

1. MLE estimation of individual preferences

$$\tilde{u}_{ijmt} = \tilde{\alpha}_i p^T_{jmt} + \tilde{\beta}_i b_{jmt} + \boldsymbol{\lambda}^d dem_{it} + \boldsymbol{\lambda}^l \boldsymbol{l}_{ijt} + \delta_{jmt} + \varepsilon_{ijmt}$$
$$\tilde{u}_{i0mt} = \varepsilon_{i0mt}$$

2. Linear regression estimation of common product preferences

$$\delta_{jmt} \equiv \alpha p^T_{jmt} + \beta b_{jmt} + \boldsymbol{\lambda}^a \boldsymbol{a}_{jmt} + \eta_{c(j)t} + \mathbb{d}_{mt} + \tilde{\xi}_{jmt}$$

3. Semi-parametric minimum distance estimator of prior and quality preferences

$$\eta_{c(j)t} \equiv \gamma' \mathbb{E}_{jt}[\boldsymbol{q} | r_{c(j)t}] + \nu_{c(j)}$$

The first step of the estimator is implemented as a weighted-MLE with a nested fixed point problem. Let $\boldsymbol{\theta} = (\{\tilde{\alpha}_i, \tilde{\beta}_i\}_i, \boldsymbol{\lambda}^z, \boldsymbol{\lambda}^l)$, the estimator can be defined as

$$\max_{\boldsymbol{\theta}} \sum_t \sum_i w_{it} \sum_{j \in \mathcal{J}_{mt}} y_{ijmt} \ln(s_{ijmt}(\boldsymbol{\theta}, \boldsymbol{\delta}(\boldsymbol{\theta})))$$

where $\boldsymbol{\delta}(\boldsymbol{\theta})$ is the solution to the Berry (1994) fixed point, given as the fixed point of the equation

$$g(\boldsymbol{\delta}_{mt}) = \delta_{mt} + \log(\boldsymbol{s}^*_{mt}) - \log(\boldsymbol{s}_{mt}(\boldsymbol{\theta}, \boldsymbol{\delta}_{mt})) \quad \forall m, t$$

Where s^*_{mt} is the observed market share vector for a county m and year t.

I estimate this stage by implementing a BFGS minimizer of the negative log-likelihood, solving for each guess of $\boldsymbol{\theta}$ for the vector $\boldsymbol{\delta}$. Each inner fixed-point is solved using fixed point iterations, parallelizing over years, up to an euclidean norm error below 10^{-8}. I provide an analytic gradient for the MLE estimator that accounts for the internal fixed point.

The second stage estimator is the simplest and corresponds to a two-stage least-squares regression. The instruments are discussed in the main text.

The third stage is more involved as it requires solving for the posterior beliefs. The estimator is given by

$$\min_{\gamma,\zeta} \sum_{c(j)} \sum_{t} \sum_{\tau>t} \left(\Delta_t^\tau(\eta_{c(j)t} - \gamma'\mathbb{E}_{jt}[\boldsymbol{q}|r_{c(j)t};\zeta])\right)^2 \tag{14}$$

$$\begin{aligned} s.t \quad & f_c(q;\zeta_c) \geq 0 \quad \forall q, c \\ & \int_{\boldsymbol{Q}_c} f_c(q;\zeta_c)dq = \bar{q}_c^* \quad \forall c \end{aligned}$$

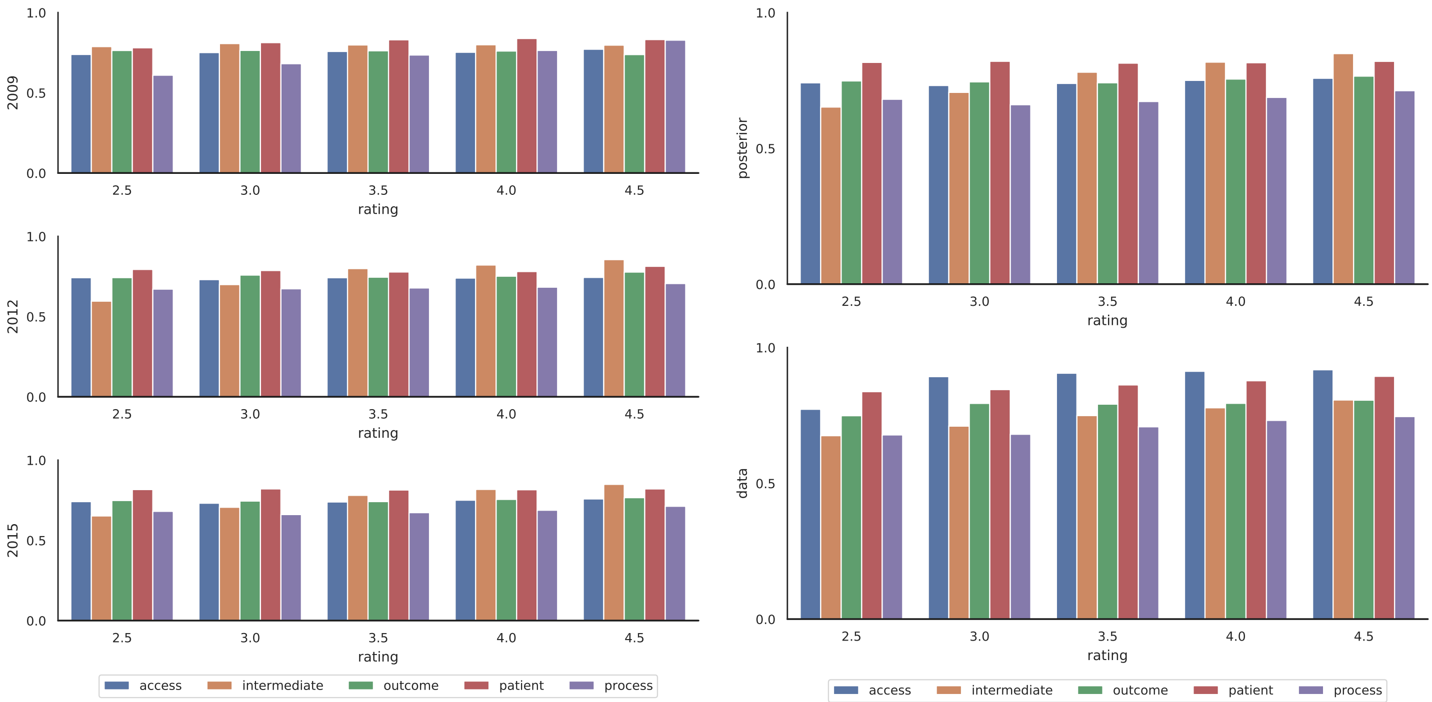
The first constraint imposes that prior of each quality category c has to be positive, and the second imposes that the prior mean of each category equals the sample mean $\bar{q}_c^*$. The domain of integration $\boldsymbol{Q}_c$ of each category is determined as the 20% below the minimum, and the same above the observed maximum, constrained to be within the unit interval. I implement the posterior belief evaluation using techniques from Quasi-Monte Carlo integration (Kuo and Nuyens, 2016), using 250 draws from a Niederreiter grid over the five dimensional space given by $\prod_c \boldsymbol{Q}_c$. For each of the draws, and each contract-year, I evaluate the corresponding scoring rule and determine the appropriate rating. I denote $\boldsymbol{Q}_{jt}(r)$ the set of vectors that would obtains rating r for contract-year jt. For tractability, I assume that priors are independent over categories. Therefore, the numerical form for the posterior quality belief given observed rating r for contract j in year t is

$$\mathbb{E}_{jt}[\boldsymbol{q}|r] = \frac{\sum_{\boldsymbol{q}\in\boldsymbol{Q}_{jt}(r)} \boldsymbol{q} \prod_c \zeta_c^t \boldsymbol{b}_c(q_c)}{\sum_{\boldsymbol{q}\in\boldsymbol{Q}_{jt}(r)} \prod_c \zeta_c^t \boldsymbol{b}_c(q_c)}$$

Where ζ_c are the Fourier coefficients of category c and $\boldsymbol{b}_c(q_c)$ is the vector of Fourier bases evaluated at the point q_c. I re-parameterize the Fourier bases from their usual support of $[0, 2\pi]$ to $\boldsymbol{Q}_c$, for each category. I impose that the first coefficient of each category b_{0c} be equal to one, which guarantees that the prior over each dimension integrates to unity. Conditions under which $\boldsymbol{f}_c$ admits a Fourier series representation are found in many non-parametric Econometrics and Fourier Analysis books, such as Horowitz (2009). Figure 15 shows the estimated posterior beliefs and compares them to the distribution of quality in the data.

6.2.1 *Quality beliefs and preferences:* Consumers' valuation for ratings is identified from their willingness to trade premium increases for scores. As each rating defines a lottery over potential qualities, if the data held variation over the gamut of lotteries, it would be possible to uncover consumers' preferences over them and find a unique subjective utility representation (Anscombe and Aumann, 1963).[12] The problem with this non-parametric identification argument is that discrete scores generate only a fraction of all lotteries. Hence, even under the most generous assumption about the data generating process (DGP) for MA, the variation would be insufficient for identification. Instead, by structuring preferences for quality, I prove

[12]With the right type of variation in scoring rules, one could even identify more complex cases, such as utilities that vary with ratings beyond their effect on information (Lu, 2019).



(a) Across years

(b) 2015, Compared to data

Figure 15: Estimated posterior means

Note: These figures display the estimated consumer posterior beliefs of contract-category quality for different ratings. The figure on the left shows how the average posterior across all contracts evolve over they years, highlighting how the information contained in the rating changes. The figure on the right contrast the posterior belief of consumers for plans obtaining a specific rating and the quality observed in the data in 2015.

a semi-parametric identification result that only uses variation consistent with MA.

The semi-parametric structure comes for the assumption that $v(q)$ is linear such that $\mathcal{E}[v(q)|r, \psi] = \gamma' \mathcal{E}[\boldsymbol{q}|r, \psi]$. This assumption clarifies that the identification challenge consists of separating tastes for quality (γ, finite parameters) from subjective beliefs about them ($\mathcal{E}[\boldsymbol{q}|r, \psi]$, non-parametric function). That is, do consumers prefer plans with five stars over those with four because they believe the quality difference is small yet valuable or large but less valuable? The answer depends on how consumers form their beliefs, for which I consider two alternatives. The first is that consumers understand the category-level design, as supported by the evidence in Section 4. The opposing alternative is that consumers are unaware of the design and instead interpret ratings based on some exogenous experience with ratings.

Assumption 1 (Consumer beliefs)**.** *One of the following holds:*

a) ***Informed choice****: Consumers' posterior beliefs $\mathcal{E}[\boldsymbol{q}|r, \psi]$ are the Bayesian posteriors of a Lipschitz continuous prior density $f : \boldsymbol{Q} \to \mathbb{R}_+$, conditional on the scoring design ψ. Moreover, the support of f is compact, connected, and satisfies $\boldsymbol{Q} \subseteq supp(f) \subseteq [0, 1]^{dim(\boldsymbol{Q})}$.*

b) ***Ignorance****: Consumer's posterior beliefs $\mathcal{E}[\boldsymbol{q}|r, \psi]$ are exogenous and independent of ψ. That is, $\mathcal{E}[\boldsymbol{q}|r, \psi] = \mathcal{E}[\boldsymbol{q}|r]$. Additionally, they are weakly increasing in r and bounded within the space of* *quality $\mathbf{Q}$.*

I will primarily rely on the assumption of informed choice, which is closer to the literature as I discuss in the main text. I connect the observed designs with the DGP through the following definition and assumption.

Definition 1. *(N Linear partitional scores) $\psi : [0,1]^n \to A$ is an N linear partitional score if there exists N partitions $\{[0,r_1),[r_1,r_2),\cdots[r_{n-1},1]\}$ and $\boldsymbol{w} \in \mathbb{R}^n_+$ such that $\psi(\boldsymbol{q}) = a_k \in A$ if and only if $\boldsymbol{w}'\boldsymbol{q}$ is in the $k-$th interval. Where $\{a_1,\ldots,a_N\} \in A$ is some sequence of the messages without repetitions.*

Assumption 2. *(Design variation) The data scoring rule for year t, ψ_t, is drawn from a distribution F_ψ with strictly positive density over the space of N linear partitional scores, with $N \geq 3$.*[13]

The class of linear partitional scores is restricted but natural. For example, quality certification belongs to this class, but a disclosure policy that fully reveals quality above a threshold does not. Yet, both are allowed to be drawn from F_ψ as the assumption permits the scoring rule to take on forms different from linear partitions. Section 6.1 shows that the MA Star Rating design can be accurately represented as N polynomial partitional scores, a class that contains the N linear partitional scores. Finally, I can state the main result of this section.

Theorem 1. *Let assumption 2 hold. If assumption 1.a holds then $(\gamma, f(\cdot))$ are identified. If assumption 1.b holds, then there exists an identified lower bound for γ.*

The proof of Theorem 1 is presented in several steps for clarity and ease of exposition. I focus on the proof of the first part of the theorem under informed choice.[14] The second part, under ignorance, is trivial and shown at the end of the section. I begin by assuming that the rating-year fixed effect $\eta_{rt} = \gamma'\mathcal{E}[q|\psi(q) = r] + \bar{\eta}$ is identified. $\bar{\eta}$ recognizes that any fixed-effect is identified up to a constant. In the context of MA, this corresponds to the mean valuation – over time and products – for MA relative to TM. Identification of this term from individual demand data is standard, and I discuss its non-parametric identification at the end of this section. To develop intuition, I begin by showing the proof for one-dimensional quality. The result relies on the following lemma.

Lemma 1. *Let f, f' be two distinct, Lipschitz continuous, strictly positive densities, supported on $[0,1]$. Then, there exists $\underline{x} < \tilde{x} < \bar{x} \in [0,1]$ such that either*

$$\mathbb{E}_f[x|x \in (\underline{x},\tilde{x})] \geq \mathbb{E}_{f'}[x|x \in (\underline{x},\tilde{x})] \quad \wedge \quad \mathbb{E}_f[x|x \in (\tilde{x},\bar{x})] \leq \mathbb{E}_{f'}[x|x \in (\tilde{x},\bar{x})]$$

with one of the inequalities strict, or the analogous statement hold with the roles of f, f' reversed. Also,

[13] Distributions over the space of linear partitional scores are well defined as the class is parametrized by its slopes and cutoffs.

[14] Throughout, unqualified "identification" of a primitive θ from some infinite data collection $\mathcal{D}$, means that there is a single θ that can generate the observed $\mathcal{D}$.

there exists another $\underline{x} < \bar{x} \in [0,1]$ *such that*

$$\mathbb{E}_f[x|x \in (\underline{x}, \bar{x})] = \mathbb{E}_{f'}[x|x \in (\underline{x}, \bar{x})]$$

Proof. As the functions are distinct, there exists a point in which f and f' differ. By continuity, they also differ in a positive measure neighborhood of that point. Without loss of generality, suppose that $f(x) > f'(x)$ for each point within this neighborhood. As both functions integrate to 1, f can not lie strictly above f'. Therefore, there exist at least one point $\tilde{x}$ such that f crosses f' from below. By continuity there exists $\epsilon > 0$ such that $f(x) > f'(x)$ for all $x \in (\tilde{x}, \tilde{x} + \epsilon)$, and $f(x) \leq f'(x)$ for all $x \in (\tilde{x} - \epsilon, \tilde{x})$. This implies that

$$F(\tilde{x} + \epsilon) - F(\tilde{x}) = \int_{\tilde{x}}^{\tilde{x}+\epsilon} f(x)dx > \int_{\tilde{x}}^{\tilde{x}+\epsilon} f'(x)dx = F'(\tilde{x} + \epsilon) - F'(\tilde{x})$$

and the contrary on $(\tilde{x} - \epsilon, x)$ with weak inequality. Denote $f_{\tilde{x}}^{\epsilon}(x) = \frac{f(x)}{F(\tilde{x}+\epsilon)-F(\tilde{x})}$, and analogously for f'. By definition, $f(\tilde{x}) = f'(\tilde{x})$ implying that $f_{\tilde{x}}^{\epsilon}(\tilde{x}) < f_{\tilde{x}}^{'\epsilon}(\tilde{x})$. As both $f_{\tilde{x}}^{\epsilon}(x)$ and $f_{\tilde{x}}^{'\epsilon}(x)$ integrate to one within $(\tilde{x}, \tilde{x} + \epsilon)$, they must cross. Let $\underline{\epsilon} \in (0, \epsilon)$ be the smallest value such that the two functions cross at $\tilde{x} + \underline{\epsilon}$. Therefore, there must exist also $\bar{\epsilon} \in (\underline{\epsilon}, \epsilon)$ such that $f_{\tilde{x}}^{\epsilon}(x) < f_{\tilde{x}}^{'\epsilon}(x)$ for every $x \in (\tilde{x}, \tilde{x} + \underline{\epsilon})$ and the contrary for every $x \in (\tilde{x} + \underline{\epsilon}, \tilde{x} + \bar{\epsilon})$. Note that $f_{\tilde{x}}^{\bar{\epsilon}}(x)$ and $f_{\tilde{x}}^{'\bar{\epsilon}}(x)$ cross only once within $(\tilde{x}, \tilde{x} + \bar{\epsilon})$ as otherwise it would violate the definition of $\underline{\epsilon}$. Denote this crossing point as $\tilde{x} + \hat{\epsilon}$, and let $\bar{x} = \tilde{x} + \bar{\epsilon}$. Finally, we have that

$$\begin{aligned}
\mathbb{E}_f[x|x \in (\tilde{x}, \bar{x})] - \mathbb{E}_{f'}[x|x \in (\tilde{x}, \bar{x})] &= \int_{\tilde{x}}^{\tilde{x}+\bar{\epsilon}} x f_{\tilde{x}}^{\bar{\epsilon}}(x)dx - \int_{\tilde{x}}^{\tilde{x}+\bar{\epsilon}} x f_{\tilde{x}}^{'\bar{\epsilon}}(x)dx \\
&= \int_{\tilde{x}}^{\tilde{x}+\hat{\epsilon}} x(f_{\tilde{x}}^{\bar{\epsilon}}(x) - f_{\tilde{x}}^{'\bar{\epsilon}}(x))dx + \int_{\tilde{x}+\hat{\epsilon}}^{\tilde{x}+\bar{\epsilon}} x(f_{\tilde{x}}^{\bar{\epsilon}}(x) - f_{\tilde{x}}^{'\bar{\epsilon}}(x))(x)dx \\
&> (\tilde{x} + \hat{\epsilon}) \int_{\tilde{x}}^{\tilde{x}+\hat{\epsilon}} (f_{\tilde{x}}^{\bar{\epsilon}}(x) - f_{\tilde{x}}^{'\bar{\epsilon}}(x))dx + (\tilde{x} + \hat{\epsilon}) \int_{\tilde{x}+\hat{\epsilon}}^{\tilde{x}+\bar{\epsilon}} (f_{\tilde{x}}^{\bar{\epsilon}}(x) - f_{\tilde{x}}^{'\bar{\epsilon}}(x))(x)dx \\
&= 0
\end{aligned}$$

Which delivers the strict inequality. The weak opposite inequality follows by applying the same logic on the interval $(\tilde{x} - \epsilon, \tilde{x})$. This proves the first point of the lemma. For the second part of the lemma, consider the range defined by the previous step: $(\underline{x}, \bar{x})$. On this range, we have one of three scenarios: $\mathbb{E}_f[x|x \in (\underline{x}, \bar{x})] = \mathbb{E}_{f'}[x|x \in (\underline{x}, \bar{x})]$, in which case no more proof is needed, or one of the two inequalities hold. Without loss, as the following argument is symmetric, assume that $\mathbb{E}_f[x|x \in (\underline{x}, \bar{x})] > \mathbb{E}_{f'}[x|x \in (\underline{x}, \bar{x})]$. Define $g(\lambda) = \mathbb{E}_f[x|x \in (\underline{x}, \tilde{x} + \lambda\bar{\epsilon})] - \mathbb{E}_{f'}[x|x \in (\underline{x}, \tilde{x} + \lambda\bar{\epsilon})]$ where $\bar{x} = \tilde{x} + \bar{\epsilon}$. Note that $g(1) > 0$ and $g(0) \leq 0$. Moreover, $g(\lambda)$ is continuous in $\lambda \in [0,1]$, so by the intermediate value theorem there exists $\lambda^* \in [0,1]$ such that $g(\lambda^*) = 0$. The interval for the second part of the lemma is thus $(\underline{x}, \tilde{x} + \lambda^*\bar{\epsilon})$. □

The value of this lemma is that it shows that for any two priors, we can construct two interior regions, defined by three thresholds $(\underline{x}, \tilde{x}, \bar{x})$, such that the posteriors conditional on these regions have opposite orderings. Also, it tells us that there is an interior region where both posteriors are equal. Before stating the theorem for one dimensions, I define the space from which scores are drawn.

Definition 2. *A scoring rule $\psi : [0,1] \to A$ is an N monotone partitional score if $|A| = N$ and there exists $N-1$ cutoffs $\{r_n\}_{n=1}^{N-1}$ such that*

$$\psi(q) = \begin{cases} A_1 & \textit{if } q \in [0, r_1) \\ \vdots & \\ A_k & \textit{if } q \in [r_{k-1}, r_k) \\ \vdots & \\ A_N & \textit{if } q \in [r_{N-1}, r_N] \end{cases}$$

for some indexing of the message space A, such that $k' \neq k \implies A_k \neq A_{k'}$.

Note that as the message space is irrelevant, any N monotone partitional score is parametrized by its set of partitions. Therefore, distributions over the space of N monotone partitional scores are well defined. I now state and prove that preferences and beliefs about quality are identified if quality is scalar.

Theorem 2. *Suppose $q \in [0,1]$ and $\mathcal{E}[q|\psi(q) = r, \psi]$ is the Bayesian posterior of some prior f given the score and scoring system $\psi(\cdot)$. Assume that f is strictly positive and Lipschitz continuous over $[0,1]$, and the scoring rule ψ is drawn independently from a distribution with full support over the space of N monotone partitional scores, with $N \geq 2$. Then the prior f, and consumers' preferences for quality γ, are identified.*

Proof. Let $\mathcal{I}$ denote the identified set, consisting of tuples of preferences (γ), priors (f), and mean valuations for scored products relative to the outside option ($\bar{\eta}$), that can potentially generate the sequence of identified rating-year fixed effects η_{rt}. By contradiction, suppose that there exist two distinct $(\gamma_0, f_0, \bar{\eta}_0), (\gamma_1, f_1, \bar{\eta}_1) \in \mathcal{I}$. By definition it must be that for any r and any t,

$$\gamma_0 \mathbb{E}_{f_0}[q|r, \psi_t] + \bar{\eta}_0 = \eta_{rt} = \gamma_1 \mathbb{E}_{f_1}[q|r, \psi_t] + \bar{\eta}_1$$

By the second part of Lemma 1, there exists a positive measure partition of $[0,1]$ such that the posteriors of f and f' coincide. Moreover, a scoring rule that generates this partition as a rating interval is drawn with positive probability. Let $\tilde{\psi}$ denote this scoring rule, such that $\mathbb{E}_{f_0}[q|\tilde{r}, \tilde{\psi}] = \mathbb{E}_{f_1}[q|\tilde{r}, \tilde{\psi}]$. Denote $\tilde{\eta}_{\tilde{r}}$ the identified rating-year fixed-effect for that rule and partition.

By Lemma 1 there also exists two partitions of $[0,1]$ such that the posteriors of f_0 and f_1 on these partitions have opposite orderings. Also, a scoring rule that generates such a partition is

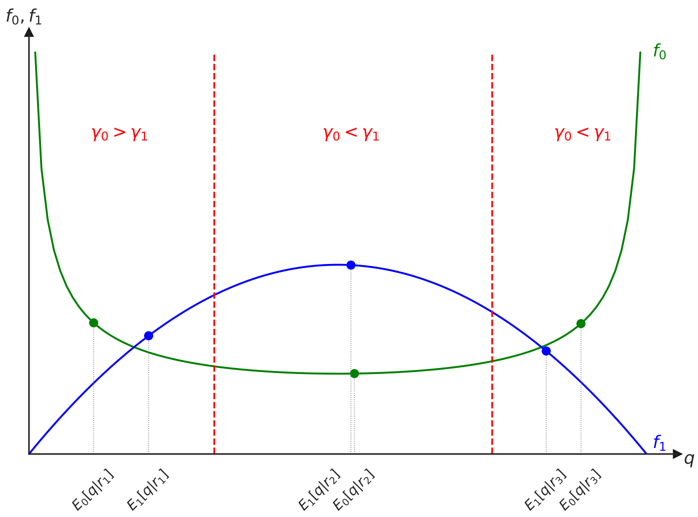


Figure 16: Quality belief and preferences identification

Note: This figure illustrates the intuition behind the identification of consumer priors and preferences for quality. In the example, two alternative priors are compared (f_0, f_1) over three ratings. In each rating segment, the posterior value of both ratings must agree ($\gamma_0 \mathbb{E}_0[q|r_1] = \gamma_1 \mathbb{E}_1[q|r_1]$). The condition on γ implied by the first rating contradicts the other two. This rejects that both priors can represent the same preferences over quality lotteries.

drawn with positive probability. Therefore, there exists ψ and two ratings r, r' such that

$$\mathbb{E}_{f_0}[q|r,\psi] < \mathbb{E}_{f_1}[q|r,\psi]$$
$$\mathbb{E}_{f_0}[q|r',\psi] \geq \mathbb{E}_{f_1}[q|r',\psi]$$

Where the directions of the inequality are assumed without loss, as the argument is symmetric if the opposite inequalities take place. Using this we have that

$$\begin{aligned}
\gamma_0(\mathbb{E}_{f_0}[q|r,\psi] - \mathbb{E}_{f_0}[q|\tilde{r},\tilde{\psi}]) = \eta_{rt} - \tilde{\eta}_{\tilde{r}} = \gamma_1(\mathbb{E}_{f_1}[q|r,\psi] - \mathbb{E}_{f_1}[q|\tilde{r},\tilde{\psi}]) \\
\implies \gamma_0 > \gamma_1 \\
\gamma_0(\mathbb{E}_{f_0}[q|r',\psi] - \mathbb{E}_{f_0}[q|\tilde{r},\tilde{\psi}]) = \eta_{r't} - \tilde{\eta}_{\tilde{r}} = \gamma_1(\mathbb{E}_{f_1}[q|r',\psi] - \mathbb{E}_{f_1}[q|\tilde{r},\tilde{\psi}]) \\
\implies \gamma_0 \leq \gamma_1
\end{aligned}$$

which contradicts that both $(\gamma_0, f_0, \bar{\eta}_0)$ and $(\gamma_1, f_1, \bar{\eta}_1)$ are in the identified set. Therefore $\mathcal{I}$ is singleton. $\square$

The key intuition behind the proof is that one prior can not always be more "optimistic" about the quality in a rating interval than another prior. Figure 16 illustrates Lemma 1, which tells us that under some score one prior acts as if quality is high and the other, as if it is low. Because they both have to explain the same data, their corresponding preferences must be ordered the opposite way. But, as the priors and preferences are fixed, we can find another score such that the first prior acts as if quality is low and the other as if it is high. Which contradicts that with unchanging preferences, both could have generated the same data.

Because the argument of this proof is not constructive, it might be unclear how it translates to the estimation of the parameters. To clarify this, consider the following simple argument for the estimation of γ. In the data, we can estimate the rating-year fixed-effect η_{rt} for each year t and one of the nine distinct ratings r. Suppose the rating intervals in year t are $[k_0 = 0, k_1, \ldots, k_8, k_9 = 1]$, such that $\eta_{rt} = \gamma \mathbb{E}_f[q|q \in [r_{k_{r-1}}, r_{k_r})] + \bar{\eta}$, with the last interval being closed. Now note that we can derive bounds on gamma by doing the following (for $r > r'$)

$$
\begin{aligned}
\eta_{rt} - \eta_{r't} &= \gamma(\mathbb{E}_f[q|r] - \mathbb{E}_f[q|r']) \\
\gamma &= \frac{\eta_{rt} - \eta_{r't}}{\mathbb{E}_f[q|r] - \mathbb{E}_f[q|r']} \\
\implies \gamma &\in (\frac{\eta_{rt} - \eta_{r't}}{k_r - k_{r'-1}}, \frac{\eta_{rt} - \eta_{r't}}{k_{r-1} - k_{r'}}) \\
\implies \gamma &\in \bigcap_{r>r',t} (\frac{\eta_{rt} - \eta_{r't}}{k_r - k_{r'-1}}, \frac{\eta_{rt} - \eta_{r't}}{k_{r-1} - k_{r'}})
\end{aligned}
$$

These bounds are non-trivial. For example, suppose that the true f is uniform, $\gamma = 1$, and the partitions for one year are $\{[0, 1/9), [1/9, 2/9), \ldots [8/9, 1]\}$. This single year of data delivers that $\gamma \in (8/9, 8/7)$. As more years are added and the scoring rule changes, this set narrows delivering the true γ.

With the intuition at hand, I now show that the same arguments extend to vector-valued quality. For the extension, I recognize that the previous result relied on three properties: the continuity of f, partitional ratings in $[0, 1]$ corresponding to intervals, and that the natural ordering on $\mathbb{R}$ is complete. Continuity extends for $f : [0, 1]^n \to \mathbb{R}_+$, but the two other requires some adjustment. Partitions in $[0, 1]^n$ can take on many forms, some of which are not naturally given by the Star Rating. The natural order on vectors, the strong element-wise order, is not complete and so the result of Lemma 1 has to be adjusted to a different, complete, order on $\mathbb{R}^n$.

Lemma 2. *Let f, f' be two distinct, Lipschitz continuous, strictly positive densities, supported on $[0, 1]^n$. For every strictly positive $w \in \mathbb{R}^n_+$, there exists a triplet $a < b < c \in \mathbb{R}_+$ such that either*

$$
\begin{aligned}
&\mathbb{E}_f[w'x|w'x \in (a, b)] \geq \mathbb{E}_{f'}[w'x|w'x \in (a, b)] \\
\wedge \quad &\mathbb{E}_f[w'x|w'x \in (b, c)] \leq \mathbb{E}_{f'}[w'x|w'x \in (b, c)]
\end{aligned}
$$

with one of the inequalities strict, or the analogous statement hold with the roles of f, f' reversed. Also, there exists another $a, b \in \mathbb{R}_+$ such that

$$
\mathbb{E}_f[w'x|w'x \in (a, b)] = \mathbb{E}_{f'}[w'x|w'x \in (a, b)]
$$

Proof. First, note that we can normalize one of $\boldsymbol{w}$ elements to 1 as all comparison can be rescaled relative to it. To distinguish between scalars and vector, I denote vectors in bold letters. Let $w_1 = 1$, and note that f and f' define distributions over the random variable $y = \boldsymbol{w}'\boldsymbol{x}$. To find the implied distribution, define the linear map $\boldsymbol{y} = (\boldsymbol{w}'\boldsymbol{x}, x_2, \ldots, x_n) = W\boldsymbol{x}$, where W is an identity matrix with its first row replace by $\boldsymbol{w}$. Using this and the standard change-of-variables method, the distribution of $\boldsymbol{y}$ induced by the prior f is

$$f_{\boldsymbol{y}}(\boldsymbol{y}) = \frac{1}{|det(W)|} f(W^{-1}\boldsymbol{y}) \quad \boldsymbol{y} \in \{\boldsymbol{y} \in \mathbb{R}_+ \times [0,1]^{n-1} : \sum_{i=2}^{n} w_i y_i \leq y_1 \leq \sum_{i=2}^{n} w_i y_i + 1\}$$

and so integrating over the marginals of $\boldsymbol{y}$

$$f_y(y) = \int_{V(y)} f(y - \sum_{i=2}^{n} w_i v_i, \boldsymbol{v}) d\boldsymbol{v} \quad V(y) = \{\boldsymbol{v} \in [0,1]^{n-1}, \boldsymbol{w}'\boldsymbol{v} \leq y\}$$

Where I used that $W = I + \boldsymbol{u}\boldsymbol{v}'$ where $\boldsymbol{u} = [1, 0, \ldots, 0]$ and $\boldsymbol{v} = [0, w_2, \ldots, w_n]$, which by the matrix determinant lemma implies that $|det(W)| = 1$. The implied distribution by f' is defined analogously. Therefore we have that

$$\mathbb{E}_f[\boldsymbol{w}'\boldsymbol{x} | \boldsymbol{w}'\boldsymbol{x} \in [\underline{y}, \bar{y}]] = \frac{\int_{[0,1]^n} \mathbb{1}\{\boldsymbol{w}'\boldsymbol{x} \in [\underline{y}, \bar{y}]\} \boldsymbol{w}'\boldsymbol{x} f(\boldsymbol{x}) d\boldsymbol{x}}{\int_{[0,1]^n} \mathbb{1}\{\boldsymbol{w}'\boldsymbol{x} \in [\underline{y}, \bar{y}]\} f(\boldsymbol{x}) d\boldsymbol{x}} = \frac{\int_{\underline{y}}^{\bar{y}} y f_y(y) dy}{F_y(\bar{y}) - F_y(\underline{y})}$$

Thus, we only need to prove that $f_y(\cdot)$ and $f'_y(\cdot)$ satisfy the conditions of Lemma 1. First, note that both are positive and supported on $[0, \sum_{i=1}^{n} w_i]$. However, any closed interval of $\mathbb{R}$ is isomorphic to $[0,1]$, so the only property left to verify is continuity.

Let $\{y_n\} \in [0, \sum_i w_i]$ be a convergent sequence to y. Define the extension $\tilde{f} : [-\sum_{i=2}^{n} w_i, 1] \times [0,1]^{n-1} \to \mathbb{R}_+$ such that $\tilde{f}(\boldsymbol{x}) = f(\boldsymbol{x})$ if $x_i \geq 0$, and $\tilde{f}(\boldsymbol{x}) = 0$ otherwise. Note that $\mathbb{1}\{\boldsymbol{v} \in V(y)\} f(y - \sum_{i=2}^{n} w_i v_i, \boldsymbol{v}) \leq \tilde{f}(y - \sum_{i=2}^{n} w_i v_i, \boldsymbol{v})$ for every $y, \boldsymbol{v}$. Also, $\tilde{f}$ is integrable as f is. Hence, by the dominated convergence theorem

$$\begin{aligned}
\lim_{n\to\infty} f_y(y_n) &= \int \lim_{n\to\infty} \mathbb{1}\{\boldsymbol{v} \in V(y_n)\} f(y_n - \sum_{i=2}^{n} w_i v_i, \boldsymbol{v}) d\boldsymbol{v} \\
&= \int \mathbb{1}\{\boldsymbol{v} \in V(y)\} f(y - \sum_{i=2}^{n} w_i v_i, \boldsymbol{v}) d\boldsymbol{v} \\
&= f_y(y)
\end{aligned}$$

corroborating continuity.

□

Having proven this, extending the main theorem is simple. First, I restate the class of scoring rules I assume that data draws from

Definition 3. *A scoring rule $\psi : [0,1]^n \to A$ is an N linear partitional score if there exists $N-1$ cutoffs*

$\{r_n\}_{n=1}^N$ and a vector $w \in \mathbb{R}_+^n$ such that

$$\psi(q) = \begin{cases} A_1 & \text{if } w'q \in [0, r_1) \\ \vdots & \\ A_k & \text{if } w'q \in [r_{k-1}, r_k) \\ \vdots & \\ A_N & \text{if } w'q \in [r_{N-1}, r_N] \end{cases}$$

for some indexing of the message space A, such that $k' \neq k \implies A_k \neq A_{k'}$.

Theorem 3. *Suppose $q \in [0,1]^n$ and $\mathcal{E}[q|\psi(q) = r, \psi]$ is the Bayesian posterior of some prior f given the score and scoring system $\psi(\cdot)$. Assume that f is strictly positive and Lipschitz continuous over $[0,1]^n$, and that the scoring rule ψ is drawn independently from a distribution with full support over the space of N linear partitional scores, with $N \geq 2$. Then the prior f, and consumers' preferences for quality γ, are identified.*

Proof. The proof follows directly by applying the argument in Theorem 2, combined with the observation that if w, x_1, x_2, y_1, y_2 are all strictly positive vectors such that $y_1'x_1 = y_2'x_2$ and $w'x_1 > w'x_2$ then $w'y_1 < w'y_2$. Succinctly, fix a scoring rule slope w, and use lemma 2 to obtain the partitions used for baseline and contradiction. With this, obtain ψ and $\tilde{\psi}$ such that

$$\mathbb{E}_{f_0}[w'q|r, \psi] < \mathbb{E}_{f_1}[w'q|r, \psi] \qquad \mathbb{E}_{f_0}[w'q|r', \psi] \geq \mathbb{E}_{f_1}[w'q|r', \psi]$$

and as before

$$\gamma_0'(\mathbb{E}_{f_0}[q|r, \psi] - \mathbb{E}_{f_0}[q|\tilde{r}, \tilde{\psi}]) = \eta_{rt} - \tilde{\eta}_{\tilde{r}} = \gamma_1'(\mathbb{E}_{f_1}[q|r, \psi] - \mathbb{E}_{f_1}[q|\tilde{r}, \tilde{\psi}]) \implies w'\gamma_0 > w'\gamma_1$$
$$\gamma_0'(\mathbb{E}_{f_0}[q|r', \psi] - \mathbb{E}_{f_0}[q|\tilde{r}, \tilde{\psi}]) = \eta_{r't} - \tilde{\eta}_{\tilde{r}} = \gamma_1'(\mathbb{E}_{f_1}[q|r', \psi] - \mathbb{E}_{f_1}[q|\tilde{r}, \tilde{\psi}]) \implies w'\gamma_0 \leq w'\gamma_1$$

Which delivers the contradiction. □

The intuition behind the proof is as with the one dimensional case. The rating slope gives a complete order to compare beliefs against and allows me to reduce the dimensionality of quality. Although the theorem relies uniquely on the variation in scoring cutoffs, this is not to say that slope variation is not useful. It is quite possible that an alternative proof relies on slope variation keeping cutoffs fixed.

I now turn attention to the second part of Theorem 1. This part is straight forward because, subject to the ignorance assumption, posterior beliefs are fixed and exogenous. Hence, $\mathcal{E}[q|r, \psi] = v_r$ for some fixed v_r. Moreover, for the purposes of the robust scoring design of Section 8, we are only concerned with the lower bounds on preferences. For any pair of fixed effects $\eta_{rt} > \eta_{r't'}$ with $r > r'$ we can obtain a lower bound for γ_k for any category k as

$$\begin{aligned}
\eta_{rt} - \eta_{r't'} &= \sum_i \gamma_i (v_r - v_{r'}) \leq \sum_i \gamma_i (\bar{q}_i - \underline{q}_i) \\
\implies \frac{\eta_{rt} - \eta_{r't'} - \sum_{i \neq k} \gamma_i (\bar{q}_i - \underline{q}_i)}{(\bar{q}_k - \underline{q}_k)} &\leq \gamma_k \\
\implies \max_{r > r', t} \frac{\eta_{rt} - \eta_{r't'} - \sum_{i \neq k} \gamma_i (\bar{q}_i - \underline{q}_i)}{(\bar{q}_k - \underline{q}_k)} &\leq \gamma_k
\end{aligned}$$

Where $\underline{q}_k, \bar{q}_k$ are the minimum and maximum quality in category k, in the market.

Finally, I return to the identification of the fixed effect $\eta_{rt} = \gamma' \mathcal{E}[q|\psi_t(q) = r] + \bar{\eta}$. Intuitively, we can identify this value by looking at how much more money are consumers willing to pay for a product of rating r relative to an unrated equivalent product. Alternatively, we can identify it by examining how much are consumer willing to pay for a fixed product, relative to TM, for different rating levels. In the simple logit choice model of Section 5, this is can be recovered as a rating-year fixed effect, identified up to a location normalization. However, the logit structure is not necessary. η_{rt} is a common product fixed-effect, identified relative to the price coefficient. Lemma 4 in Berry and Haile (2020) provides conditions for the price coefficient and η_{rt} to be identified in a general choice model framework.[15] Notably, beyond the instrumental variable requirements to identify the price coefficient, the key requirement is that η_{rt} enters additively in consumers' unobserved preference for products.

6.2.2 Surplus loss from asymmetric information: Consumers' imperfect knowledge of product quality constitutes the driving force behind the scoring system. It is both the reason for its existence and the mechanism through which it alters quality. Therefore, it is crucial to understand the extent of asymmetric information in the market. Using estimated preferences and belief parameters, I compute consumer surplus loss due to imperfect information following Train (2015). Formally, denote by $u^*_{ijmt} = u_{ijmt} - \gamma'(\mathbb{E}_{jt}[q|r_{jt}] - \boldsymbol{q}_{jt})$ the *true* utility of product j for consumer i in market m and year t, where u_{ijmt} is defined as in Equation (3). The expected loss from asymmetric information to consumer i is given by

$$\Delta CS_{imt} = |\alpha_i|^{-1} \left(\ln \sum_{j \in \mathcal{J}_{mt} \cup \{0\}} e^{u_{ijmt} - \varepsilon_{ijmt}} - \ln \sum_{j \in \mathcal{J}_{mt} \cup \{0\}} e^{u^*_{ijmt} - \varepsilon_{ijmt}} + \sum_{j \in \mathcal{J}_{mt}} s_{ijmt} (u^*_{ijmt} - u_{ijmt}) \right) \tag{15}$$

Table 14 shows the surplus loss for different demographic population, relative to the current system. I estimate that the average Medicare beneficiary loses approximately 185.9 dollars a year of surplus from asymmetric information, which amounts to a yearly average of \$3.74 billion across all Medicare beneficiaries. These losses exhibit wide heterogeneity in the population, particularly across chosen systems, with MA consumers losing an average of \$214.5 a year.

[15] In the Berry and Haile (2020) η_{rt} is an additive component of the identified ξ_{jt}.

Reduced switching across systems drives this disparity, as it limits TM enrollees' benefits from additional information. Interestingly, MA consumers satisfied with the information provided by Medicare about the market, and those that read the handbook, lose less than their counterparts.

There are two things to consider about these estimates. First, as newly entering plans do not have quality measurements or ratings, I assume that consumers' beliefs about them (i.e., their priors) are accurate. This assumption implies that the values presented here underestimate the consumer surplus loss. Second, this analysis ignores the changes a different informational regime would have on the endogenous product quality. The main results of this work will internalize this.

6.3 Supply

6.3.1 Quality shocks: The following results shows that the distribution of quality shocks is identified given the assumptions laid out in the main text.

Theorem 4. *Let $\boldsymbol{z}$ denote the vector of observables used by firms to form beliefs about rival actions. Assume that ϵ^M and ϵ^F are independent of $\boldsymbol{z}$, mean zero and symmetric. Additionally, their distributions have well-defined densities and nowhere-vanishing Fourier transforms. Then, the distribution of the errors are identified and given by*

$$f_{\epsilon_k^M}(\epsilon) = \mathcal{F}^{-1}\left(\mathbb{E}_z\Big[\frac{|\mathcal{F}(f)_{\Delta^F\Phi_k^{-1}(q_k)|\boldsymbol{z}}(t)|^{1/2}}{|\mathcal{F}(f)_{\Delta^M\Delta^F\Phi_k^{-1}(q_k)|\boldsymbol{z}}(t)|^{1/4}}\Big]\right); \quad f_{\epsilon_k^F}(\epsilon) = \mathcal{F}^{-1}\left(\mathbb{E}_z\Big[\frac{|\mathcal{F}(f)_{\Delta^M\Phi_k^{-1}(q_k)|\boldsymbol{z}}(t)|^{1/2}}{|\mathcal{F}(f)_{\Delta^F\Delta^M\Phi_k^{-1}(q_k)|\boldsymbol{z}}(t)|^{1/4}}\Big]\right) \quad \forall k$$

Where $\mathcal{F}(\cdot)$ is the Fourier transform, and Δ^M is the within-market across-firms difference operator, and Δ^F is the within-firm across-market difference operator.

Proof. Denote $x_{jk} \equiv \Phi_k^{-1}(q_{jk})$ the normalized quality. The model implies the following relationship

$$x_{jk} = x^*_{jk} + \epsilon^M_{mk} + \epsilon^F_{fk}$$

Where m is the market in which j is offered, f is the firm of product j, and x^*_{jk} is the unobserved target. Moreover, both shocks are drawn independently across their indices from an unknown, symmetric and differentiable distributions $F_{\epsilon_k^M}(\cdot)$ and $F_{\epsilon_k^F}(\cdot)$. As the results is the same for each category, and there are no connections imposed across them, I omit the category index from this point onward. Given that the mean of x^*_j is the mean of the observed x, we can assume without loss that the we have normalized the inputs such that all variables in the equation have mean zero.

The model further specifies the existence of a vector z such that $x_j|z$ is independent from $x_{j'}|z$ if j and j' are offered by different firms. Also, both shocks are independent of z. Thus, we can write the relationship

$$x_j|z = x_j^*|z + \epsilon_j^M + \epsilon_j^F$$

Consider now two arbitrary products by firm f offered in two distinct markets (m, m'). The difference in normalized qualities satisfies

$$\Delta^F x_f|z = \Delta^F x_f^*|z + \Delta^F \epsilon_{(m,m')}^M$$

As long as there is more than one firm that has products in (m, m'), we have now a non-parametric unobserved measurement error problem with $N > 1$ repeated observations (Schennach, 2016). The standard deconvolution for this class of problem is that

$$\mathcal{F}(f)_{\Delta^F \epsilon_{(m,m')}^M} = \frac{\mathcal{F}(\Delta^F x_f|z)}{|\mathcal{F}(\Delta^M \Delta^F x_f|z)|^{1/2}}$$

Where $\mathcal{F}$ is the Fourier transform, and $\Delta^M \Delta^F x_f|z$ is the result of taking the difference of $\Delta^F x_f|z$ across two of the firms overlapping in (m, m'). This result leverages the fact that $\Delta^M \Delta^F x_f|z$ is symmetric, as $\Delta^F x_f|z$ is independent and of mean zero across firms. Finally, the symmetry and zero-mean of ϵ^M implies that

$$\mathcal{F}(f)_{\epsilon_{(m,m')}^M} = |\frac{\mathcal{F}(\Delta^F x_f|z)}{|\mathcal{F}(\Delta^M \Delta^F x_f|z)|^{1/2}}|^{1/2}$$

Integrating over the distribution of z on both sides and taking the inverse Fourier transform delivers the desired result. The result for the firm-level shock is analogous, but taking first the difference within-market, across firm. □

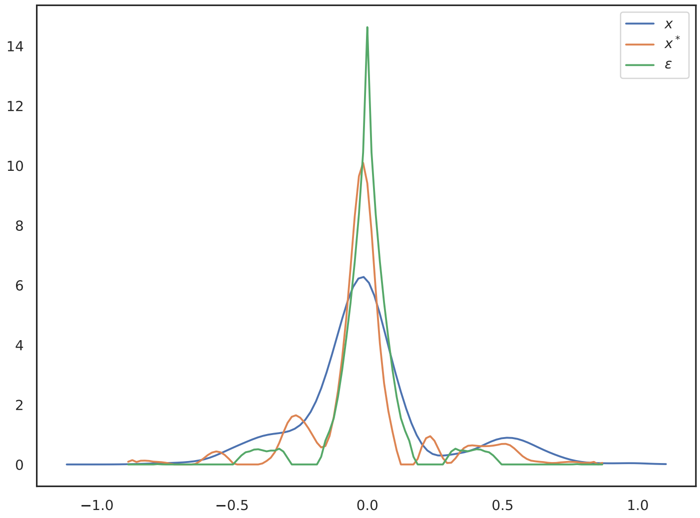
The quality shock distributions are estimated non-parametricaly by solving the conditional deconvolution problem

$$x_{jkt} = x_{jkt}^* + \epsilon_{m(j)kt}^M + \epsilon_{f(j)kt}^F$$

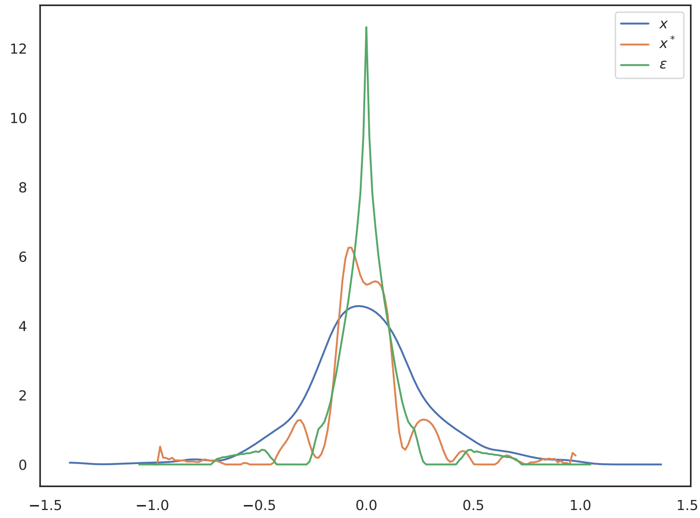
Where $(x^*, \epsilon^M, \epsilon^F)$ are unobserved and x is. In the set of observables used by firms, I include benchmarks, rebate fractions, bundled services, plan types, and market sizes, as well as means, variances, and correlations between the same set of variables for rivals. Additionally, I include indicators for the presence of the top ten firms (by all-time enrollment) in the market. Overall, the vector contains over a hundred attributes observable by firms when investing. However, firms likely use only a few of these to form beliefs as rivals' targets are only relevant insofar they affect demand.[16] This observation suggests a sparse relationship between quality and the conditioning set which the estimator leverages. The estimation proceeds in three steps.

First, I assign *shock markets* to plans and form the data set used for estimation. To do so,

[16]For example, knowing that one will be competing against Humana, who systematically commands a significant market share, is probably enough to render the attributes related to all other smaller rivals irrelevant.



(a) Outcome quality



(b) Process quality

Figure 17: Quality deconvolution

Note: These figures display two examples of how the deconvolution estimator separated the part of the signal attributed to the shock from the targets. In the figures, x is the observed inverted quality, $\Phi_k^{-1}(q_k)$, x^* is the unobserved target, and ε is the sum of the shocks..

I first select markets with at least two firms that also overlap in another market. Markets that fail to satisfy this condition are not helpful for this estimation step, given the differences in x that appear in the formulas. Most contracts offered in these markets are found only in one market or have the vast majority of their demand in one. However, there is a fraction for which the assignment is less clear. To avoid specifying this manually and arbitrarily, I use the Resident-Matching algorithm with both markets' and contracts' matching preferences specified according to observed enrollment. This way, these contracts are assigned to markets with less loss of valuable data. I then compute the observable market characteristics that enter the conditioning vector $\boldsymbol{z}$, as defined in the main text.

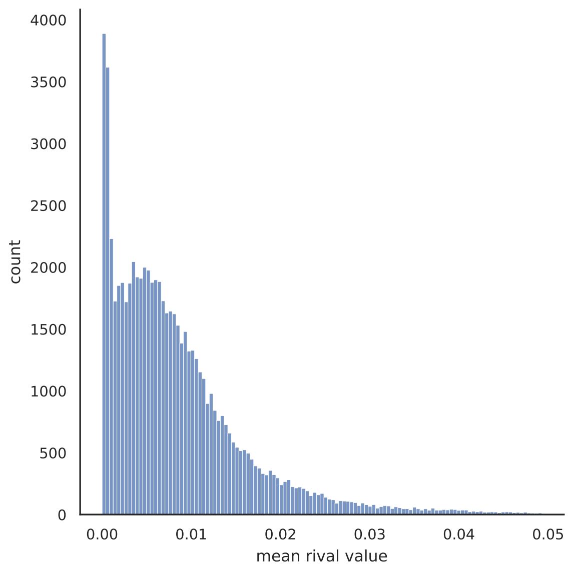
Second, I used the data to estimate the conditional distribution of differences in realized quality (i.e., the distributions of $\Delta^F x_f|z$, $\Delta^M x_f|z$, $\Delta^M \Delta^F x_f|z$). To estimate the non-parametric conditional density, I use the estimator of Izbicki and Lee (2017). Specifically, I use the FlexZBoost implementation of Dalmasso et al. (2020) which combines this estimator with the XGBoost Machine Learning algorithm (Chen and Guestrin, 2016). I use a cosine basis with a maximum of 30 elements, set to minimize the mean squared error of predictions. I use 20% of a random permutation of the data to train the algorithm and the remainder to tune it. This estimator returns an estimate of each density at a collection of points. I ensure the symmetry of the distributions, as assumed in the theorem, by averaging the recovered density and its reverse-order values. Figure 17 illustrates how the deconvolution process separates the observed quality distribution into the investment and shock distributions.

Finally, I use the estimated densities and Fast-Fourier-Transforms (FFT) to evaluate the expression of theorem 4. Specifically, I use an FFT algorithm that exploits the Hermitian symmetry of the estimated distribution.

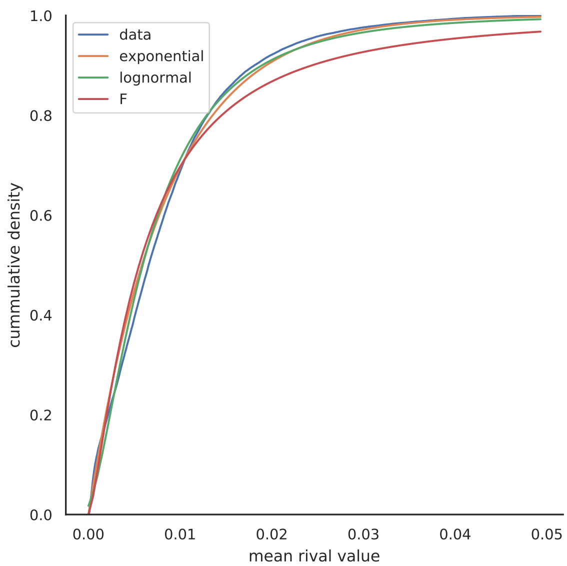
6.3.2 Firm beliefs about rivals: The model outlined in Section 5 states that firms form beliefs about rival target choices based on market observables. The empirical challenge is that these targets are unobserved. However, while the firm might form beliefs about targets, an inspection of the model reveals that firms are only concerned about these targets' effect on consumers' value. Concretely, the firm is affected exclusively by how rivals' choices will alter the joint distribution of the individual demand preferences of consumer i and the average of rival utility values $v_i = \frac{1}{|\mathcal{J}_{-f}|} \sum_{j \in \mathcal{J}_{-f}} \exp(\tilde{u}_{ij})$. Given the demand estimates, I can recover this distribution from the data. Moreover, as the demand model has only a handful of consumer types, the problem is substantially simpler than estimating each rival's joint distribution of qualities in each market. Next, I describe the estimation procedure.

First, I classify the consumer types which govern the dimensionality of the vector of rival values. As the utility of consumer i for MA plan j is given by equation (3), the dimensionality of consumer heterogeneity correlated with v_{if} is limited. In particular, the consumer attributes z_{it} can be ignored from v_{if} as they can be loaded on the outside option's utility. Thus, the joint distribution is over $(\alpha_i, \beta_i, \boldsymbol{l}_i, v_{if})$, where $\boldsymbol{l}_i$ is the consumers lock-in status vector. The challenge is that $\boldsymbol{l}_i$ can take on as many values as MA contracts plus one. However, the insurer only uses v_{if} to predict the effect of changing quality on future profits. This observation implies that dimensionality reductions of the parameters space are harmless as long as they do not systematically alter the likelihood with which consumers will adopt or drop a plan as its quality changes. Therefore, instead of considering the full vector of $\boldsymbol{l}_i$, I consider three lock-in statuses for each consumer: MA, TM, and None. Moreover, I reduce (α_i, β_i) by removing groups that are not statistically different from each other, given the estimates of section 6.1. This procedure reduces consumer types from 4075 to 28 consumer types. I use the demand data to compute the rival value v_{if} for each consumer, firm, and market.

Next, I use the computed rival values to fit a probability distribution for each consumer type across markets. To decide which distribution to use, I first explore the data. Figure 18.a shows the histogram of rival values, and figure 18.b shows some common distribution fitted naively to the data. The figures show that the data looks close to log-normal, except for a mass point near zero. These are markets where the firm is nearly a monopolist, with one or two small value plans competing against it. These markets are plentiful yet minuscule in overall enrollment and account for a small fraction of firms' profits. Accordingly, I fit log-normal distributions to each marginal defined by a consumer type. I allow the mean and variance parameters of the fitted distributions to vary with market characteristics. I use the same variables used when estimating the quality shocks in Section 6.2.2, allowing the mean parameter to vary with the mean of these variables for rivals and analogously for the variance. The average mean-square error (MSE) resulting from predicting the empirical cumulative distribution of rival values



(a) Rival values histogram



(b) Rival values naive fit

Figure 18: Rival value distribution estimation

Note: These figures illustrate the distribution of rival values in the data and the one fitted.

across all marginals was 0.00022.

Finally, I combine the 28 marginals by fitting a Gaussian copula to the logarithm of rival values. I then draw 100 samples from the copula and use the inverse CDF of the fitted marginals to map these back to market-firm-level draws. At the end of the process, I obtain 100 independent draws for each firm-market and consumer type. These draws describe the estimated distribution, which can I contrast with the aggregate rival-values data. The MSE of the fitted empirical CDF is $4.827 * 10^{-5}$.

It is worth noting, however, that the simplification of this approach comes at a cost. The dimensionality of v_i is smaller than that of the set of rival targets only if I do not condition on rivals' vector of non-quality attributes. Doing so would render the method pointless. However, the demand estimates indicate that quality, premiums, and benefits largely dominate v_i. The two last items are unknown to firms when investing in quality and depend on its realization. Hence the extent of firms' uncertainty goes beyond just quality targets, which I capture through the uncertainty over v_i. Nevertheless, the loss is in the exact connection between these beliefs and the equilibrium that ensues. Thus, this approach is an approximation that slightly extends the assumption regarding firms' beliefs to include uncertainty about other product characteristics determined after quality.

6.3.3 Investment costs: The following results shows that marginal investment costs satisfy a useful regression expression.

Proposition 3. *The investment model satisfies the following regressions equation*

$$\dot{\pi}_{jkt} = \frac{\partial I(\Phi^{-1}(\boldsymbol{q}_f) + \boldsymbol{\epsilon}_f^M + \boldsymbol{\epsilon}_f^F, \boldsymbol{\mu}_f)}{\partial x_{jkt}} + \nu_{jkt} \qquad \mathbb{E}[\nu_{jkt}|\boldsymbol{q}_f] = 0 \tag{16}$$

Where $\dot{\pi}_{jkt}$ is the expected marginal profit of the third-stage, given the observation of $\boldsymbol{q}_f$. Moreover, $\dot{\pi}_{jkt}$ can be expressed as a function of only identified distributions.

Proposition 3 contains two simple results. The first is that $\mathbb{E}[\nu_{jkt}|\boldsymbol{q}_f] = 0$ and the second is that $\dot{\pi}_{jkt}$ is a function only of identified distributions. To reduce notation, I omit time indices in this section and denote $\boldsymbol{y}_f = \Phi^{-1}(q_f)$, the normalized observed quality.

Proof. The first results is trivial and follows from appropriately defining $\dot{\pi}_{jk}$ as the derivative of the conditional expectation of the expected third-stage profits for the firm controlling product j. Specifically, define π_f as the expected third stage profit as in equation (6).

$$\pi_f \equiv \sum_m \int \mathbb{E}_{mt}[V_{fm}(\boldsymbol{q}_f, \boldsymbol{q}_{-f})] dF(\boldsymbol{q}_f|\boldsymbol{x}_f)$$

Assuming for now the differentiability of π_f with respect to x_{jk}, we can decompose the derivative as

$$\frac{\partial \pi_f}{\partial x_{jk}} = \mathbb{E}[\frac{\partial \pi_f}{\partial x_{jk}}|\boldsymbol{y}_f] + \nu_{jkt}$$

letting $\dot{\pi}_{jk} \equiv \mathbb{E}[\frac{\partial \pi_f}{\partial x_{jk}}|\boldsymbol{y}_f]$, delivers the first point.

The slightly more complicated step is organizing π_f in a way that is differentiable, tractable, and that allows $\dot{\pi}_{jk}$ to be expressed as as a function of the identified distributions. To do so, I begin by re-expressing π_f as

$$\int\int\int V_f(\Phi(\boldsymbol{x}_f + \boldsymbol{\epsilon}_m + \boldsymbol{\epsilon}_f), \boldsymbol{q}_{-f}) f(\boldsymbol{q}_{-f}|\boldsymbol{\epsilon}_m, \boldsymbol{z}) f_{\epsilon_F}(\boldsymbol{\epsilon}_f) f_{\epsilon_M}(\boldsymbol{\epsilon}_m) d\boldsymbol{q}_{-f} d\boldsymbol{\epsilon}_m d\boldsymbol{\epsilon}_f$$

Where I have made explicit the internal expectation regarding rivals, with $\boldsymbol{z}$ the set of market observables determining these beliefs. This expression also makes explicit the way the market-level shocks affect beliefs about rivals. Finally, I have compressed $\sum_m V_{fm} = V_f$ to avoid confusing the markets that define shocks from those in which the firm participate. Thus, now $\boldsymbol{\epsilon}_m$ is the vector of market-level shocks over all markets and categories.

Define

$$\tilde{V}_{fj}(\cdot|\boldsymbol{\epsilon}_m, r) = \int V_f(\Phi(\cdot), \boldsymbol{q}_{-f}|r_j = r) f(\boldsymbol{q}_{-f}|\boldsymbol{\epsilon}_m, \boldsymbol{z}) d\boldsymbol{q}_{-f}$$

Where $V_f(\boldsymbol{q}_f, \boldsymbol{q}_{-f}|r_j = r)$ is defined as the third-stage profit if qualities are $(\boldsymbol{q}_f, \boldsymbol{q}_{-f})$ but the rating of plan j is fixed at r. This mapping can be used to handle the points of non-differentiability introduced by the rating when evaluating π_f as a function of x_{jk}. Specifically, we can take one of the firm-level shocks, ϵ_{fk} and segment it in intervals that maintain the rating of x_{jk} constant. That is, we can rewrite π_f as

$$\int_{\epsilon_m}\int_{\epsilon_{-k}}\sum_{r\in R_{jk}(\boldsymbol{x}_j,\boldsymbol{\epsilon}_m,\boldsymbol{\epsilon}_{-k})}\int_{e_r(x_{jk})}^{e_{r+0.5}(x_{jk})}\tilde{V}_{fj}(\boldsymbol{x}_f+\boldsymbol{\epsilon}_m+\boldsymbol{\epsilon}_f|\boldsymbol{\epsilon}_m,r)f_{\epsilon_F}(\boldsymbol{\epsilon}_f)f_{\epsilon_M}(\boldsymbol{\epsilon}_m)d\boldsymbol{\epsilon}_m d\boldsymbol{\epsilon}_f$$

Where $R_{jk}(\boldsymbol{x}_j,\boldsymbol{\epsilon}_m,\boldsymbol{\epsilon}_{-k})$ is the set of ratings that plan j can reach through different firm-level shocks in category k, given the other values of integration. Letting $\psi(\boldsymbol{q}_j,\omega_j)$ denote the rating rule as in Equation (1) and R the set of allowed ratings, the feasible set of ratings is defined as

$$R_{jk}(\boldsymbol{x}_j,\boldsymbol{\epsilon}_m,\boldsymbol{\epsilon}_{-k})=\{r\in R|\exists\epsilon_{fk}\in\mathbb{R},\quad r=\psi(\Phi(\boldsymbol{x}_j+\boldsymbol{\epsilon}_m+\boldsymbol{\epsilon}_k),\omega_j)\}$$

and the integration limits are defined as follows: for $r\in R_{jk}(\boldsymbol{x}_j,\boldsymbol{\epsilon}_m,\boldsymbol{\epsilon}_{-k})$ and $r\in(1,5]$ these are

$$e_r(x_{jk})=\Phi_k^{-1}(\psi_k^{-1}(r-0.25-\sum_{k'\neq k}\psi_{k'}(\Phi_{k'}(x_{jk'}+\epsilon_{m(j)k'}+\epsilon_{f(j)k'}))-\omega_j)-x_{jk}-\epsilon_{f(j)k}$$

and equal to $-\infty$ and ∞ for $r=1$ and $r=5.5$, respectively.[17]

Now we can evaluate the derivative safely by using the Leibniz integral rule and the envelope theorem

$$\begin{aligned}\frac{\partial\pi_f}{\partial x_{jk}}=&\int_{\epsilon_m}\int_{\epsilon_{f,-k}}\sum_{r\in R_{jk}(\boldsymbol{x}_j,\boldsymbol{\epsilon}_m,\boldsymbol{\epsilon}_{-k})}\Big(\tilde{V}_{fj}(\boldsymbol{x}_f+\boldsymbol{\epsilon}_m+[\boldsymbol{\epsilon}_{f,-k},e_r(x_{jk})]|\boldsymbol{\epsilon}_m,r)f_{\epsilon_{Fk}}(e_r(x_{jk}))-\\&\tilde{V}_{fj}(\boldsymbol{x}_f+\boldsymbol{\epsilon}_m+[\boldsymbol{\epsilon}_{f,-k},e_{r+1}(x_{jk})]|\boldsymbol{\epsilon}_m,r)f_{\epsilon_{Fk}}(e_{r+1}(x_{jk}))\Big)f_{\epsilon_{F,-k}}(\boldsymbol{\epsilon}_{f,-k})f_{\epsilon_M}(\boldsymbol{\epsilon}_m)d\boldsymbol{\epsilon}_m d\boldsymbol{\epsilon}_{f,-k}\\&-\int_{\epsilon_m,\epsilon_f}\mathbb{E}[\sum_{m'}\gamma_{jm}D_{jm'}(\boldsymbol{p}_m(\boldsymbol{q}_m),r(\boldsymbol{q}_m))|\boldsymbol{q}_f=\Phi(\boldsymbol{x}_f+\boldsymbol{\epsilon}_m+\boldsymbol{\epsilon}_f),\boldsymbol{z}]\theta_k\phi(\boldsymbol{x}_f+\boldsymbol{\epsilon}_m+\boldsymbol{\epsilon}_f)dF(\boldsymbol{\epsilon}_m,\boldsymbol{\epsilon}_f)\end{aligned}$$

This expression, while complicated, has a simple interpretation. The first set of integrals correspond to the change in profit due to increasing the probability of higher ratings, and decreasing that of lower ratings. The second set of integral is the change in profits associated with changing the marginal cost in the third stage. If service marginal costs were independent of quality this term would be zero. Note that all of the densities involved in this equation are identified. The only unknown vector being $\boldsymbol{x}_f$.

The final step is in expressing $\mathbb{E}[\frac{\partial\pi_f}{\partial x_{jk}}|\boldsymbol{y}_f]$ as a function of identified densities. This is simple,

[17] The inverse ψ_k^{-1} might not be unique. In this case, for the lower integration limit we take $\inf\psi_k^{-1}$ and the supremum for upper integration limits.

as by the unobserved error structure

$$\begin{aligned}\dot{\pi}_{jk} &= \int_{-\infty}^{\infty} \frac{\partial \pi_f}{\partial x_{jk}}(\boldsymbol{x}_f) f(\boldsymbol{x}_f|\boldsymbol{y}_f) d\boldsymbol{x}_f \\ &= \int_{-\infty}^{\infty} \frac{\partial \pi_f}{\partial x_{jk}}(\boldsymbol{y}_f - \boldsymbol{e}) f_{\epsilon_F+\epsilon_M}(\boldsymbol{e}) d\boldsymbol{e}\end{aligned}$$

Where the second line follows from a change of variables in the integral. □

7 Scoring Design

7.1 Implementation

Solving the designer's problem can be very demanding. For every guess of ratings, the designer must recompute all firms' profits functions and assess the new equilibria of the bidding and investment stages. However, as in the theory of Bayesian persuasion (Kamenica and Gentzkow, 2011), we can simplify the problem by viewing the planner as choosing posterior values for beliefs and firms' marginal cost, subject to some constraints. To see this, note that a firm's insurance profit, $V_f(\psi, \boldsymbol{q})$ varies with $(\psi, \boldsymbol{q})$ only through their effect on ratings, marginal costs, and the value that their rivals might offer in the market. Therefore, we can reparametrize this profit function as $\tilde{V}_f(\boldsymbol{w}, \boldsymbol{c}, \boldsymbol{v})$ where $\boldsymbol{w} = \gamma' \mathcal{E}[q|\psi(\boldsymbol{q})]$ and $\boldsymbol{c} = C(\boldsymbol{q}, \boldsymbol{z})$ are the vectors of quality values and marginal costs for all products. $\boldsymbol{v}$ is defined as in the previous section, which is an aggregate function of $\boldsymbol{w}$ for rivals and their bids, which are in turn a function of their own costs $\boldsymbol{c}$. Clearly, every value taken by $V_f(\cdot)$ has a direct analog in $\tilde{V}(\cdot)$. Moreover, $\tilde{V}_f$ is a function of finite vectors, while V_f has a function as its arguments. Additionally, because $\boldsymbol{Q}$ is convex and compact, and $C(\boldsymbol{q}, \boldsymbol{z})$ is linear in $\boldsymbol{q}$, $(\boldsymbol{w}, \boldsymbol{c})$ take on values on a convex compact set. The value of rivals, as function of costs and quality, is also bounded within a convex compact set by the same logic.

The reparametrization of firms' insurance profit function, $\tilde{V}_f$ implies that we can evaluate $\tilde{V}_f(\boldsymbol{w}, \boldsymbol{c}, \boldsymbol{v})$ on a finite-dimensional grid. Moreover, note that by the implicit function theorem $\tilde{V}_f$ is differentiable. As $\boldsymbol{C}$ is linear and demand is bounded, the first derivative of $\tilde{V}_f$ is bounded. Thus $\tilde{V}_f$ is Lipschitz continuous, which implies strict bounds on the approximation error from linear interpolation within the grid, and that as the number of grid points increases (uniformly over the domain), the linear interpolation of $\tilde{V}_f$ converges uniformly to the true value. Therefore, I create a grid of $(\boldsymbol{w}, \boldsymbol{c}, \boldsymbol{v})$ and evaluate $\tilde{V}_f$ for every f. This grid, and its interpolation across points, is the key to simplifying the computation of the designer's problem.

Using the grid of equilibrium insurance profits, I can evaluate the total welfare of the market at any scoring rule ψ through the following steps.

1. Solve the optimal choices of $\boldsymbol{x}^*_f(\psi)$ for each f by iteratively finding the choices of target

that maximizes the expectation of $\tilde{V}_f$ minus investment cost for each firm, and then updating their beliefs about rivals to be consistent with their choices.

2. use the equilibrium $\boldsymbol{x}_f^*$ to compute the expected firm profit, use the implied equilibrium bids (a by-product of computing the grid of $(\boldsymbol{w}, \boldsymbol{c}, \boldsymbol{v})$) to evaluate the expectation of consumer surplus.

The most expensive computation step, consisting of solving the equilibria implicit in V_f, is done only once over a large grid, which I recycle for every value of ψ.

I make two simplifying assumptions when solving this procedure. First, I hold the distributional assumption of firms on rival actions fixed, changing only their means to satisfy rational expectations. Second, I hold consumer prior fixed, only changing their posterior belief according to ψ. I interpret these choices as an expression of the short-run effect of changing the scoring rules. Evaluating the long-run effects would involve specifying how a firm's belief distribution is formed and a process for adjusting consumers' priors. Additionally, it would require considering the dynamic effects of quality investment which is beyond this work's scope.

While the simplification above makes the problem computationally tractable, expected total welfare, as a function of ψ is still a complicated function. In particular, it is not everywhere differentiable and suffers from many local flat regions where changes in cutoffs or weights do not affect firms' choices. To find the global constrained optima of this function, I use the algorithm of Malherbe and Vayatis (2017), particularly the implementation found in the DLib library (King, 2009). This algorithm is guaranteed to find the global optima of Lipschitz functions of unknown finite Lipschitz constant over a compact convex space with a non-empty interior.

7.2 Polynomial Reductions

In the main text I refer to a certain class of scores as Polynomial Reductions. The following definition formalizes this class of scores

Definition 4 (polynomial reduction score)**.** *A scoring rule $\psi : \boldsymbol{Q} \to \mathcal{A}$ is a **polynomial reduction score** if there exists a polynomial of finite order m, $P_m : \boldsymbol{Q} \to \mathbb{R}$, and a step function $R : \mathbb{R} \to \mathcal{A}$, such that $\psi(\boldsymbol{q}) = R \circ P_m(\boldsymbol{q})$.*

I will consider this class useful if it can arbitrarily approximate any finite monotone score, define as follows:

Definition 5 (finite monotone score)**.** *A scoring rule $\psi : \boldsymbol{Q} \subset \mathbb{R}^n \to \mathcal{A} \subset \mathbb{R}$ is a **finite monotone score** if $|\mathcal{A}|$ is finite and for any $q, q' \in \boldsymbol{Q}$ we have that $q' \geq q \implies \psi(q') \geq \psi(q)$.*

The MA star ratings are a finite monotone score, and so are many other common scoring systems such as deterministic quality certifications. Note that if the dimension of the quality space is one, then the polynomial reductions and finite monotone scores coincide. The reason is that in one dimension any finite monotone score is a step function, hence the polynomial part of the polynomial reduction score is irrelevant. The following proposition shows that with multiple dimensions the approximation of the polynomial become relevant.

Proposition 4. *Let $\psi : [0,1]^n \to \mathcal{A} \subset [0,1]$ be a finite monotone scoring rule. Then $\forall \epsilon > 0$, $\exists m > 0$ and a polynomial reduction ψ^* of order m, such that $||\psi^* - \psi||_{L_1} < \epsilon$.*

Proof. Fix $\epsilon > 0$. Assume $n > 1$ and that ψ takes on more than one value, otherwise the proof is trivial. Every weakly monotone finite score partitions $[0,1]^n$ into a collection of finitely-many disjoint sets $\mathcal{M}$, such that ψ is constant over set $M \in \mathcal{M}$. Without loss, assume that the Lebesgue measure of every $M \in \mathcal{M}$ is positive, otherwise the scoring set can be ignored under the L1 norm. Note that the boundary between every pair of contiguous sets is a set Δ_M of points, or equivalently, a line segment. Consider a small $\delta > 0$ such that the δ-neighborhood of each boundary $N(\Delta_M, \delta)$ do not overlap. Define the continuous function $f_\delta(x)$ as equal to ψ everywhere but in the boundary of the neighborhoods, and equal to the linear interpolation between the steps of ψ across the boundary. Note that

$$||\psi - f_\delta||_{L1} = \int_{[0,1]^n} |\psi(x) - f_\delta(x)| dx = \sum_{\Delta_M} \int_{N(\Delta_M,\delta)} |\psi(x) - f(x)| dx < \sum_{x' \in M} [\psi(\Delta_M+) - \psi(\Delta_M-)] \frac{\delta}{2}$$

Where $\psi(\Delta_M+)$ and $\psi(\Delta_M-)$ are the values of ψ above and below the boundary. Now as f_δ is continuous, by the Stone-Weierstrass theorem there exists a polynomial P_m such that for all $x \in [0,1]$ we have $|P_m(x) - f_\delta(x)| < \frac{\epsilon}{2}$. Thus, picking $\delta < \frac{\epsilon}{\sum_{x' \in M}[\psi(\Delta_M+) - \psi(\Delta_M-)]}$ we have that

$$||P_m(x) - \psi|| = ||P_m - f_\delta + f_\delta - \psi|| \leq ||P_m - f_\delta|| + ||f_\delta - \psi|| < \epsilon$$

Finally, note that we can pick R in the polynomial reduction ψ^* to further shrink the approximation error. □

7.3 Additional Results for Total Welfare Maximization

This section presents additional results related to the scoring designs shown in the main text Section 7. These consider the designer's objective when $\rho^F = 1$ and $\rho^G = 0$.

7.3.1 Added scoring flexibility: : The main analysis presents results for disclosure systems that use either two scores (certification) or nine (best linear substitute). A natural question is what is the optimal number of scores, and what do they accomplish? To explore this, I optimize designs that use at most five scores and at most fifteen scores. The resulting scoring systems are shown in Figure 19. The welfare effect of this added flexibility is shown in Table 16.

The core mechanism and design of scores using two, five, nine, and fifteen distinct signals is very similar. The cutoff placement and the design's reduction weights using five scores are very similar to the one that uses nine. The results indicate that the narrow bands in the nine-scores-based design and the zero-probability lower tail division are virtually meaningless. Therefore, the best linear substitute design shown in the main text is implementable using only five scores. Moving up to fifteen scores results in some intervals that reveal quality in specific points of the distribution. As these intervals are narrow, they act as a full revelation of quality for particular low-quality outcomes. These likely allow some firms with a high investment cost to participate in an otherwise cutthroat market. Notably, the positioning of the last two intervals, which play a crucial role in increasing quality output, remains the same. This final design discards two signals by creating regions of zero length. This confirms that the optimal number of cutoffs is thirteen, although the resulting design is somewhat awkward.

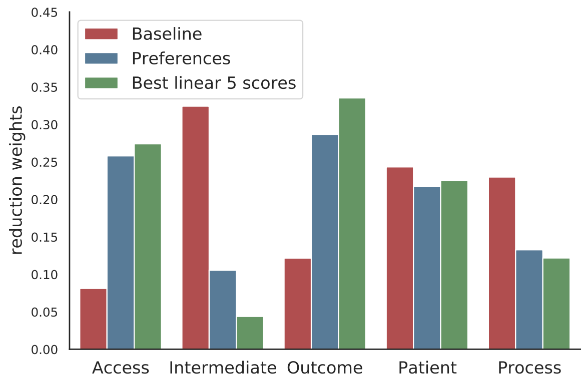
Table 16: Effect of added scoring flexibility on welfare

	Number of scores:			
	2	5	9	15
Δ Consumer Surplus	151.8	146.7	146.5	212.8
Δ Firm Profits	480.5	520.5	522.8	542.5
Δ Gov. Spending	-88.4	-74.9	-76.4	-93.8
Δ Total Welfare	632.4	667.3	669.2	755.3
MA share	56.0%	56.7%	57.1%	59.2%

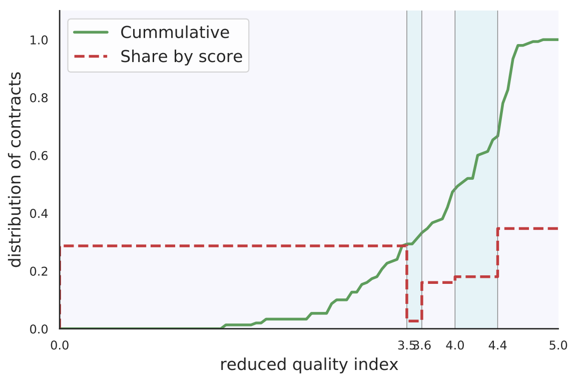
Notes: This table displays the welfare effect of redesigning the scoring system, relative to the MA Star Rating baseline. Columns present the result of adding additional flexibility in the number of scores the designer can used. A separate optimal linear aggregator is computed for each result. All values are in 2015 dollars per Medicare beneficiary.

7.3.2 Higher order reductions: : The scoring designs I present in the main analysis are based on linear reduction functions. However, it might be that the optimal reduction function is not a weighted average but rather some other smooth function, better approximated by a higher-order polynomial. To explore the welfare gains from adding flexibility to the reduction function, I solve the best second and third-order reductions for quality certification and the best second-order reduction for the nine-score design (substitute). Figure 20 displays the resulting cutoff placements for these designs and Table 17, the welfare effects.

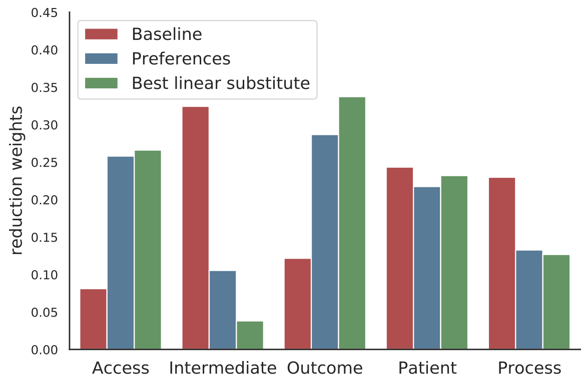
The results indicate that there is virtually no gain from higher-order reduction. The welfare effects are identical, up to an optimization error, and the resulting cutoffs are similar to the ones found in the baseline. Exploring the resulting weights indicates that the solution places only negligible coefficients on the higher-order parameters. This finding is not surprising given that consumers' iso-utility curves in quality space are linear. Therefore, to reduce misclassification regions and create an effective score, it is likely optimal to use linear functions to segment the



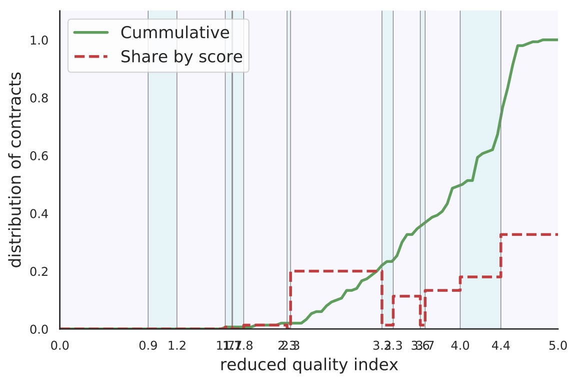
(a) Reduction weights, five scores



(b) Cutoffs, five scores



(c) Reduction weights, fifteen scores



(d) Cutoffs, fifteen scores

Figure 19: Best linear scores using five and fifteen scores

quality space.

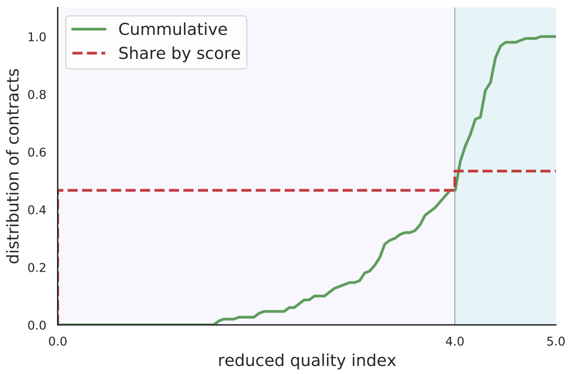
Table 17: Effect of higher order polynomial aggregators on welfare

	Certification			**Substitute**	
Order	1	2	3	1	2
Δ Consumer Surplus	151.8	130.0	130.0	146.5	138.9
Δ Firm Profits	480.5	490.8	490.8	522.8	511.1
Δ Gov. Spending	-88.4	-96.5	-96.5	-76.4	-75.5
Δ Total Welfare	632.4	620.8	620.8	669.2	650.0
MA share	56.0%	55.5%	55.5%	57.1%	57.1%

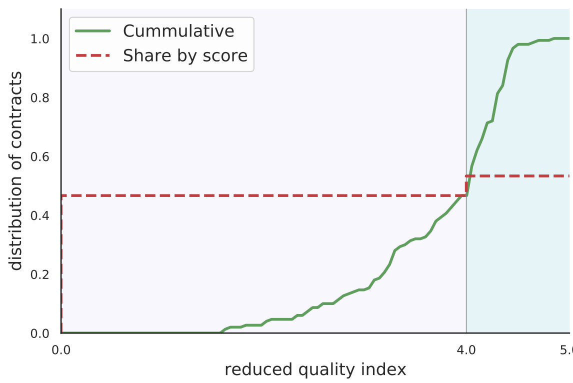
Notes: This table displays the welfare effect of redesigning the scoring system, relative to the MA Star Rating baseline. Columns present the result of adding additional flexibility in the order of the polynomial used to determine the score. All values are in 2015 dollars per Medicare beneficiary.

7.3.3 Preference-based reductions:

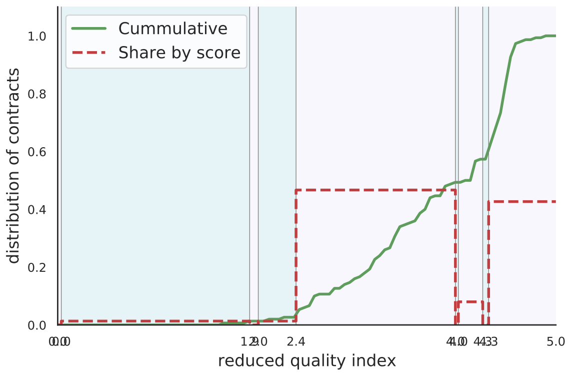
: In the main text, I state that the results of this work suggest that using consumers' preferences as reduction functions is a good rule-of-thumb. Here I present supportive evidence by solving the best cutoff placement using the rule-of-thumb,



(a) Quadratic certification



(b) Third-order certification



(c) Quadratic substitute

Figure 20: Higher order reduction design, cutoff placements

contrasting the resulting welfare with the optimal linear reduction results. Figure 21 shows the resulting cutoff placements and Table 18, the welfare changes.

The results indicate that using consumers' preferences attains nearly the same welfare as the best linear reductions. In certification, the loss from the rule-of-thumb is 3.6%, and in the best linear substitute case, it is but 1.5%. Thus, when consumers' preferences for quality are known, these could be used to create efficient scores at a much lower computational cost.

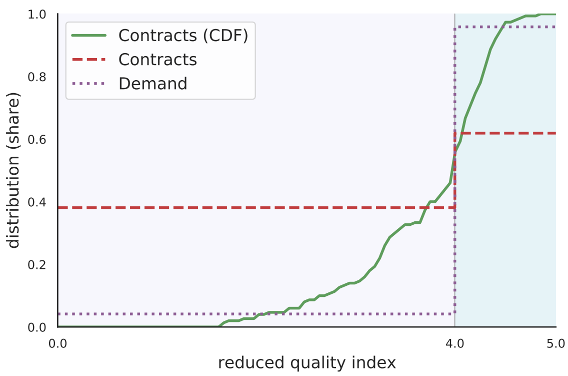
7.4 Scoring results for other social objectives

This section presents the resulting designs when considering different social objectives. In particular, I consider the certification and best linear substitute designs that emerge from maximizing total welfare net of government spending (Full Objective) and only consumer surplus. Table 19 shows the resulting welfare effect of these designs.

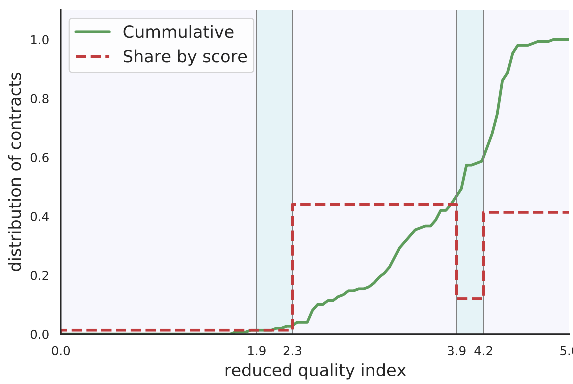
Table 18: Effect of replacing aggregator with preferences on welfare

	Certification		5 Scores		Substitute	
Aggregator	Preference	Linear	Preference	Linear	Preference	Linear
Δ Consumer Surplus	122.8	151.8	149.4	146.7	149.2	146.5
Δ Firm Profits	486.5	480.5	490.7	520.5	491.1	522.8
Δ Gov. Spending	-94.0	-88.4	-75.6	-74.9	-75.5	-76.4
Δ Total Welfare	609.2	632.4	640.1	667.3	640.3	669.2
MA share	54.9%	56.0%	56.1%	56.7%	56.1%	57.1%

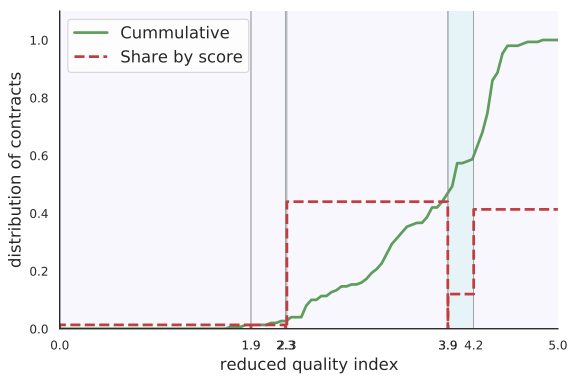
Notes: This table displays the welfare effect of redesigning the scoring system, relative to the MA Star Rating baseline. Columns present the result of replacing the optimal linear aggregator function with consumers' preferences. All values are in 2015 dollars per Medicare beneficiary.



(a) Certification



(b) Five scores



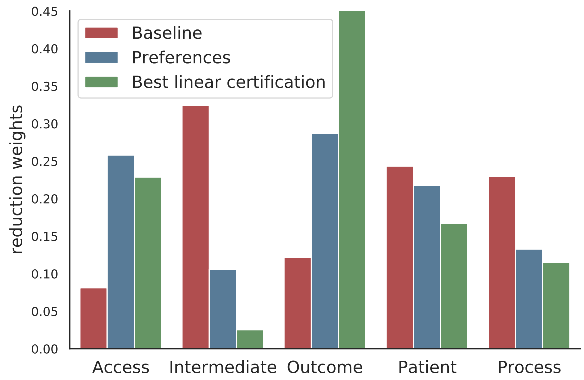
(c) Substitute (9 scores)

Figure 21: Preference-based reduction designs, cutoff placements

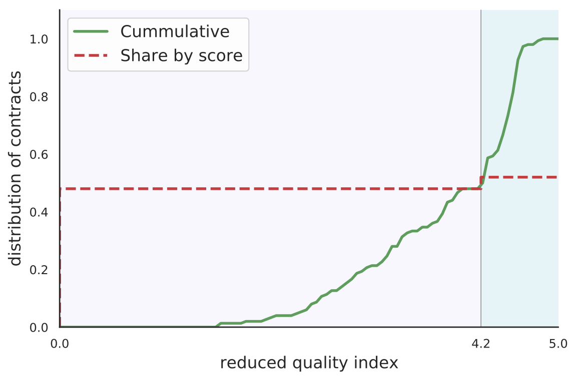
Table 19: Welfare effect of linear scores with different objectives

	Full Objective		**Consumer Surplus**	
	Certification	Substitute	Certification	Substitue
Δ Consumer Surplus	152.6	147.5	211.2	212.8
Δ Firm Profits	476.8	504.0	275.1	274.9
Δ Gov. Spending	-96.3	-81.5	-87.5	-87.4
Δ Total Welfare	629.4	651.4	486.2	487.8
MA share	56.5%	57.1%	48.1%	48.1%

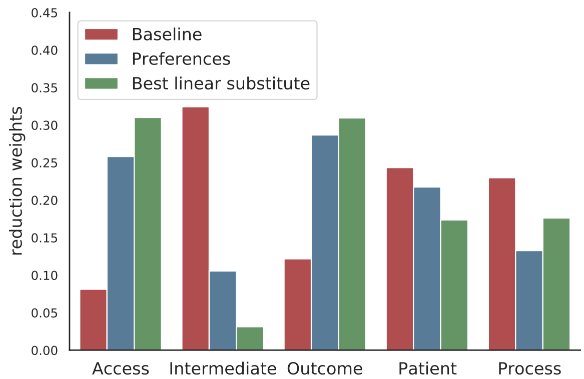
Notes: This table displays the welfare effect of redesigning the scoring system, relative to the MA Star Rating baseline. The first two columns present the welfare effect of the best linear certification and substitute (nine scores) when the objective is to maximize total welfare net of government spending. The third and fourth columns maximize only consumer surplus. All values are in 2015 dollars per Medicare beneficiary.



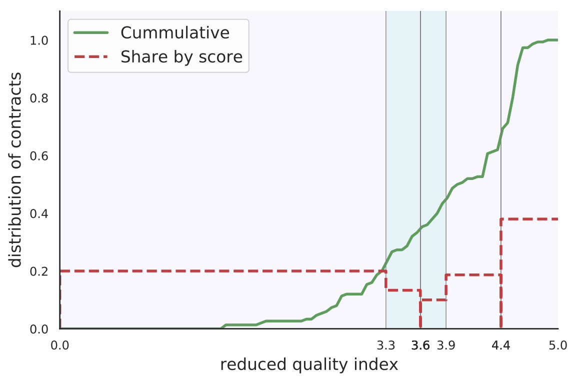
(a) Reduction weights, certification



(b) Cutoffs, certification

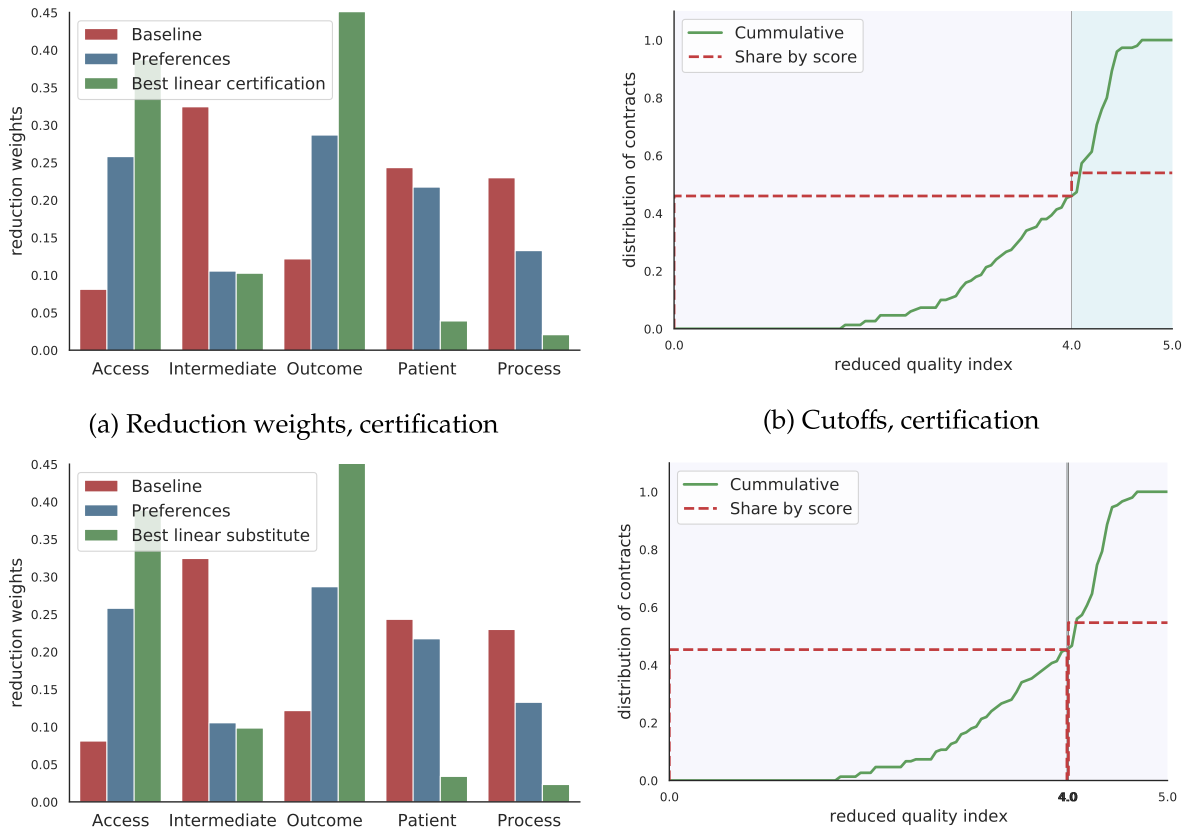


(c) Reduction weights, substitute



(d) Cutoffs, substitute

Figure 22: Full objective certification and best linear substitute



(a) Reduction weights, certification

(b) Cutoffs, certification

(c) Reduction weights, substitute

(d) Cutoffs, substitute

Figure 23: Consumer surplus certification and best linear substitute

8 Robust scoring design

8.1 Alternative formulation of robust scoring design

Proposition 5. *The robust scoring design problem of equation* (13) *can be expressed as*

$$\max_{\psi\in\Psi}\int\left(CS(\psi,\boldsymbol{q},\gamma^*)+\sum_f V_f(\psi,\boldsymbol{q})\right)dF(\boldsymbol{q}|\boldsymbol{x})-\sum_f I(\boldsymbol{x}_f,\mu_f)$$

$$s.t \quad \boldsymbol{x}_f\in\arg\max_{\boldsymbol{x}_f}\int\mathbb{E}_f[V_f(\psi,\boldsymbol{q}_f,\boldsymbol{q}_{-f})]dF(\boldsymbol{q}_f|\boldsymbol{x}_f)-I_f(\boldsymbol{x}_f,\mu_f)\quad\forall f \tag{C1}$$

$$\mathbb{E}_f[\boldsymbol{q}_{-f}]=\int\boldsymbol{q}_{-f}dF(\boldsymbol{q}_{-f}|\boldsymbol{x}_{-f})\quad\forall f \tag{C2}$$

$$\gamma^*\in\arg\min_{\gamma\in\Gamma_\eta}\gamma'\int\sum_{j\in\mathcal{J}}q_j\tilde{D}_j dF(q_j|x_j) \tag{C3}$$

Proof. First, note that consumers' posterior beliefs are irrelevant to the problem. As we have assumed that $\eta_r=\gamma\mathbb{E}[q|r]$ is known, given a draw of $\boldsymbol{q}$ and a scoring rule ψ, the demand for ratings is determined. Second, note that this implies that firm profits are not a function of the γ or the beliefs, conditional on η. Finally, we observe that

$$\begin{aligned}
&\min_{\gamma\in\Gamma_\eta}\int CS(\psi,\boldsymbol{q},\gamma)dF(q|x)\\
\iff&\min_{\gamma\in\Gamma_\eta}\int\int\frac{1}{|\alpha_i|}\left(\ln(\sum_{j\in\mathcal{J}_i}\exp(u_{ij}(\psi(q_j))-\epsilon_{ij}))+\sum_{j\in\mathcal{J}_i}s_{ij}(\psi(\boldsymbol{q}))\gamma'(\boldsymbol{q}_j-\mathbb{E}[\boldsymbol{q}_j|\psi(\boldsymbol{q}_j)])\right)didF(q|x)\\
\iff&\min_{\gamma\in\Gamma_\eta}\int\sum_{j\in\mathcal{J}}\tilde{D}_j(\psi(\boldsymbol{q}))\gamma'(\boldsymbol{q}_j-\mathbb{E}[\boldsymbol{q}_j|\psi(\boldsymbol{q}_j)])dF(q|x)\\
\iff&\min_{\gamma\in\Gamma_\eta}\int\sum_{j\in\mathcal{J}}\tilde{D}_j(\psi(\boldsymbol{q})(\gamma'\boldsymbol{q}_j-\eta_{\psi(\boldsymbol{q}_j)})dF(q|x)\\
\iff&\min_{\gamma\in\Gamma_\eta}\gamma'\int\sum_{j\in\mathcal{J}}q_j\tilde{D}_j dF(q_j|x_j)
\end{aligned}$$

□

References

Anscombe, F. J. and Aumann, R. J. (1963). A Definition of Subjective Probability. *The Annals of Mathematical Statistics*, 34(1):199–205.

Baker, A., Larcker, D. F., and Wang, C. C. Y. (2021). How Much Should We Trust Staggered Difference-In-Differences Estimates? *SSRN Electronic Journal*, (March).

Berry, S. and Haile, P. (2020). Nonparametric Identification of Differentiated Products Demand Using Micro Data. *National Bureau of Economic Research*.

Berry, S. T. (1994). Estimating Discrete-Choice Models of Product Differentiation. *The RAND Journal of Economics*, 25(2):242.

Callaway, B. and Sant'Anna, P. H. (2020). Difference-in-Differences with multiple time periods. *Journal of Econometrics*, pages 1–45.

Centers for Medicare and Medicaid Services (2016). Trends in Part C & D Star Rating Measure Cut Points. 2018.

Chen, T. and Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 13-17-Augu:785–794.

Dafny, L. and Dranove, D. (2008). Do report cards tell consumers anything they don't already know? The case of Medicare HMOs. *RAND Journal of Economics*, 39(3):790–821.

Dalmasso, N., Pospisil, T., Lee, A. B., Izbicki, R., Freeman, P. E., and Malz, A. I. (2020). Conditional density estimation tools in python and R with applications to photometric redshifts and likelihood-free cosmological inference. *Astronomy and Computing*, 30(Ml).

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*.

Holmstrom, B. (1982). Moral Hazard in Teams. Technical Report 2.

Horowitz, J. L. (2009). *Semiparametric and Nonparametric Methods in Econometrics*. Springer, Evanston, 1 edition.

Izbicki, R. and Lee, A. B. (2017). Converting high-dimensional regression to high-dimensional conditional density estimation. *Electronic Journal of Statistics*, 11(2):2800–2831.

Jacobson, G., Freed, M., Damico, A., and Neuman, T. (2019). A Dozen Facts About Medicare Advantage in 2019. (June):2008–2019.

Jung, Y. C., McFarlane, A., and Das, A. (2021). The effect of minimum wages on consumption in Canada. *Economic and Labour Relations Review*, 32(1):65–89.

Kamenica, E. and Gentzkow, M. (2011). Bayesian persuasion. *American Economic Review*, 101(6):2590–2615.

King, D. E. (2009). Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research*, 10:1755–1758.

Kolotilin, A., Mylovanov, T., and Zapechelnyuk, A. (2019). Censorship as Optimal Persuasion. *SSRN Electronic Journal*.

Kolotilin, A. and Zapechelnyuk, A. (2018). Persuasion Meets Delegation. *SSRN Electronic Journal*, pages 1–35.

Kuo, F. Y. and Nuyens, D. (2016). A practical guide to quasi-Monte Carlo methods. (November 2016):1—-50.

Lu, J. (2019). Bayesian identification: A theory for state-dependent utilities. *American Economic Review*, 109(9):3192–3228.

Malherbe, C. and Vayatis, N. (2017). Global optimization of Lipschitz functions. *34th International Conference on Machine Learning, ICML 2017*, 5(1972):3592–3601.

McGuire, T. G., Newhouse, J. P., and Sinaiko, A. D. (2011). An economic history of Medicare Part C. *Milbank Quarterly*, 89(2):289–332.

Schennach, S. M. (2016). Recent Advances in the Measurement Error Literature.

Song, Z., Landrum, M. B., and Chernew, M. E. (2013). Competitive bidding in Medicare Advantage: Effect of benchmark changes on plan bids. *Journal of Health Economics*, 32(6):1301–1312.

Spence, A. M. (1975). Monopoly , Quality , and Regulation. *The Bell Journal Of Economics*, 6(2):417–429.

Train, K. (2015). Welfare calculations in discrete choice models when anticipated and experienced attributes differ: A guide with examples. *Journal of Choice Modelling*, 16:15–22.

Zapechelnyuk, A. (2020). Optimal Quality Certification. *American Economic Review: Insights*, 2(2):161–176.