Quality Disclosure and Regulation:

Scoring Design in Medicare Advantage

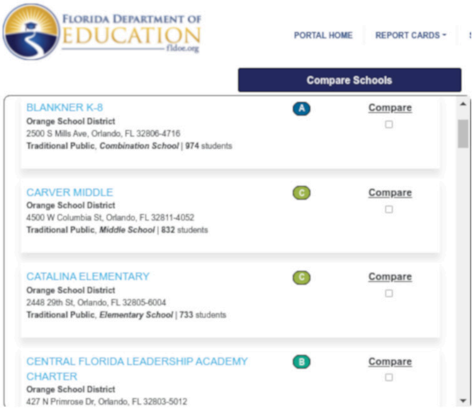
Benjamin Vatter

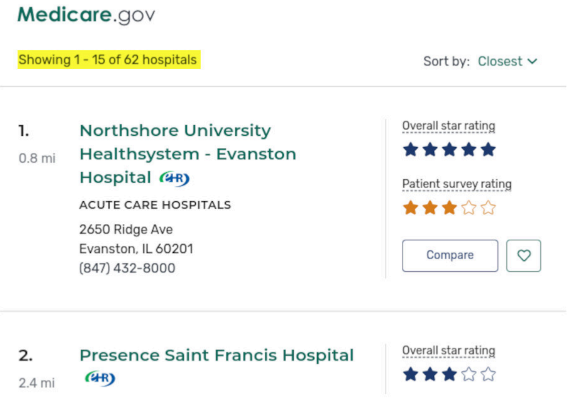
Stanford University

September 19th, 2022

Motivation

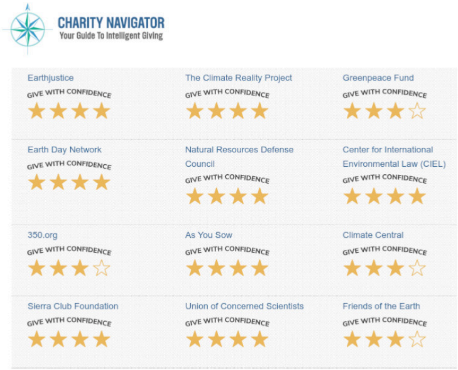
- **Quality scores affect our everyday choices**





Motivation

- **Quality scores affect our everyday choices**



Motivation

- **Quality scores affect our everyday choices**
  - How to design them to maximize welfare?

- Two central mechanisms:
  1. Help consumers choose through added information (Dranove and Jin, 2010)
  2. Affect firms' incentives to invest in quality (Barahona et al., 2020)

Motivation

- **Quality scores affect our everyday choices**
  - How to design them to maximize welfare?

- Two central mechanisms:
  1. Help consumers choose through added information (Dranove and Jin, 2010)
  2. Affect firms' incentives to invest in quality (Barahona et al., 2020)

- Scores can be powerful policy tools, however
  - No systematic guidance on how to design them
  - Poor designs can backfire (gaming) (Feng Lu, 2012)

Overview of the Paper

Q: **How to design welfare-maximizing scores for Medicare Advantage (MA)**?

> Summarize medical and service quality of insurance plans using nine scores (stars)

- Use yearly variation in scoring design between 2009 and 2015 to:
  1. Show that design affects demand and supply of health insurance
  2. Estimate a model of demand, pricing, and quality investments
     - Information asymmetries: consumers' quality information is severely limited
     - Inefficient quality provision: too low on aggregate, distorted by private incentives (Spence, 1975)

- Develop a general empirical scoring design methodology
  > Combine computational methods with insights from information design

$\Rightarrow$ Model + method deliver a welfare-improving design for MA

Preview of Results

- **New design increases surplus by 2.4 monthly premiums per consumer/year**
  - Uses five scores: five stars with discrete increments
  - One-star pools low and medium quality (↓ info) others partition high quality (↑ info)
  - Consumers avoid one-star plans, firms respond by increasing investments (↑ quality)
  - Reward more improvements in quality dimensions consumers' care about (↑ efficiency ↑ info)

⇒ Consumers make more informed choices over higher quality products

- Delivers broad lessons about scoring policies
  - Scores are powerful mechanisms by which to regulate quality
  - Coarse, simple, scores can outperform full-information outcomes at small informational losses

Outline

1 **Institutional Details and Data**
   - Graphical representation of the scoring design problem

2 **Model, Identification, and Estimates**
   - Measurement of the frictions addressed by the scores

3 **Scoring Design**
   - Mechanisms by which optimal scores improve welfare

Three Facts About Medicare Advantage

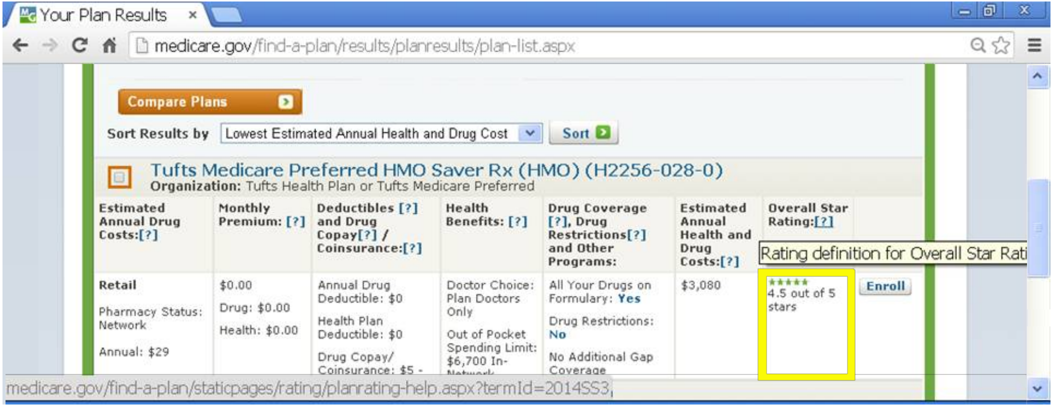
1. National regulated private health insurance market
   - All 65 million Medicare-eligible individuals can opt into MA, about half do
   - Trade-off: greater access vs. better coverage
   - Generous premium subsidies, risk-adjustments for insurers

2. Highly concentrated: 90% of average county enrollment controlled by 2 firms
   - 4 firms account for 70% of national MA enrollment

3. Quality heterogeneity affects mortality, costs billions in subsidies (Abaluck et al., 2021)
   - Challenging to assess if not for the quality scores

The MA Star Ratings

- Summarize medical and service quality in 1-to-5 stars, in half-star increments



Scoring Design (simplified)

1. Measure plan's performance over five categories of quality
   1. Medical Outcomes
   2. Intermediate Medical Outcomes (chronic conditions)
   3. Access to Care
   4. Patient Experience
   5. Process Measures (preventive, diagnostic care)
2. Give a score of 1-5 to each plan and each category
3. Show consumers the rounded weighted average

Graphical Representation

- **Design**: slope and location of hyper-planes
  - Slope = Weights, Location = Cutoffs
  - In two dimensions design is just lines $\longrightarrow$

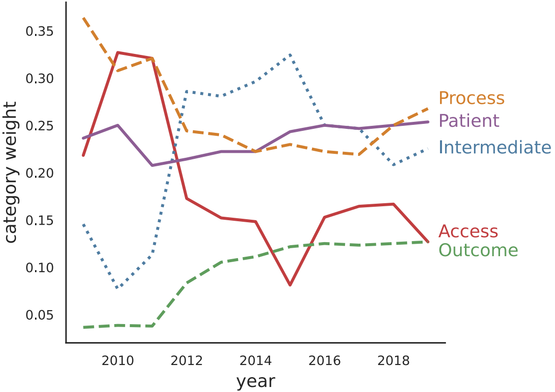
Q: Which lines to draw and how many?

- Scores reveal quality regions, not value

Data and Descriptive Evidence

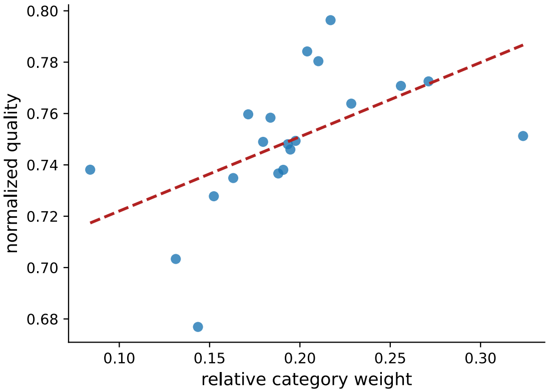
1 Scoring rules

- Hand collected from CMS
- **Substantial variation in design**



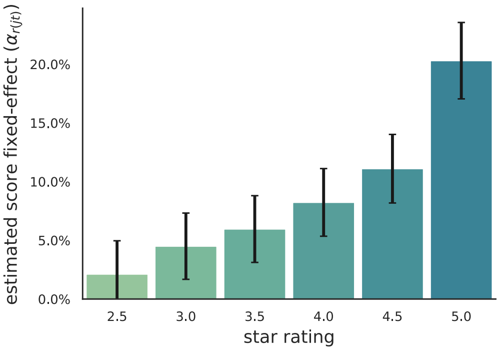
Data and Descriptive Evidence

1. Scoring rules
2. Data on all plans
   - Premiums, coverage, and benefits
   - **Quality: responds to design**



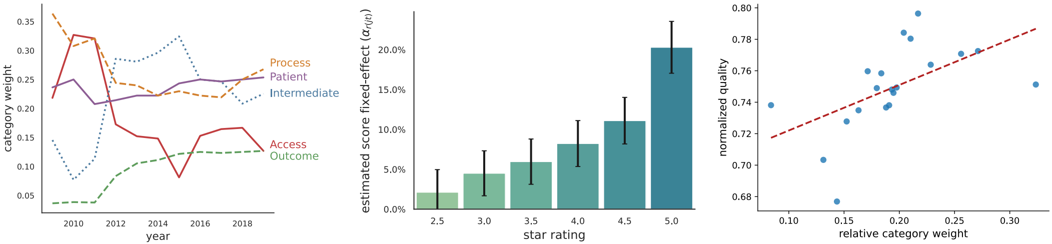
Data and Descriptive Evidence

1. Scoring rules
2. Data on all plans
3. Enrollment data
   - Individual-level representative panel
   - 46,833 enrollment choices
   - **Consumers prefer higher-scoring plans**



Taking Stock: The Designer’s Toolkit

- Plentiful design variation reveals that scores:
  1. Shift demand across products
  2. Affect firms’ quality investments
- To extrapolate to new designs, we must recover the social cost and value of quality
  - Costs: from variation in scoring incentives to invest
  - Value: from variation in WTP for scores



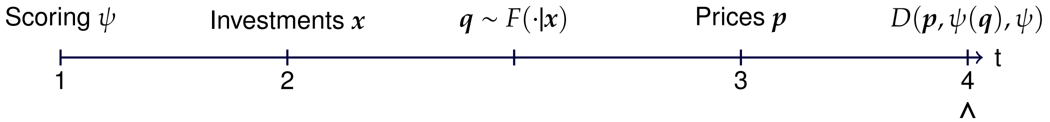
Outline

Model

Designer	**Insurers**	**Nature**	**Insurers**	**Consumers**
Scoring ψ	Investments $\boldsymbol{x}$	Quality $\boldsymbol{q} \sim F(\cdot \vert \boldsymbol{x})$	Prices $\boldsymbol{p}$	$D(\boldsymbol{p}, \psi(\boldsymbol{q}), \psi)$
1	2		3	4
public	private	designer & insurers	public	public

t

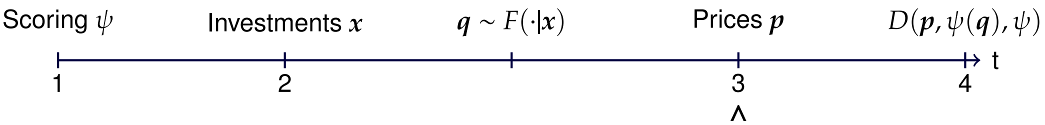
Model



$$u_{ij} = \underbrace{\alpha_i P_j}_{\text{premium}} + \underbrace{\beta_i b_j}_{\text{coverage}} + \underbrace{\mathcal{E}_q[\boldsymbol{\gamma}'\boldsymbol{q}|\psi(q_j), \psi]}_{\text{quality}} + \underbrace{\lambda' z_{ij}}_{\text{Obs. attributes}} + \underbrace{\xi_j}_{\text{unobs. preferences}} + \underbrace{\varepsilon_{ij}}_{\sim\text{T1EV}}$$

- Choose among MA plans – or – Medicare + Part D (prescription drug coverage)
- Heterogeneity in WTP for quality $(\gamma/\alpha_i) \Rightarrow$ scoring granularity
- Subjective Bayesian non-parametric priors $\Rightarrow$ scoring cutoffs and weights

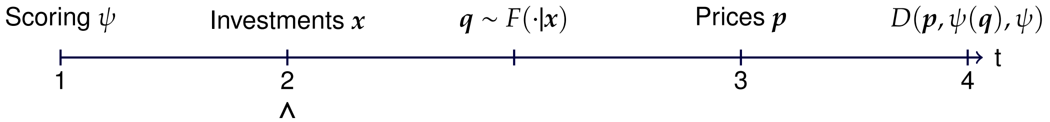
Model



$$\pi_f(\boldsymbol{q}, \psi) = \max_{\{p_j\}_{j \in J_f}} \sum_{j \in J_f} \underbrace{D_j(\boldsymbol{p}, \psi(\boldsymbol{q}))}_{\text{demand}} \big(\underbrace{R_j(p_j)}_{\text{Mg. Revenue}} - \underbrace{C(\boldsymbol{q}_j, \boldsymbol{z}_j, \boldsymbol{\theta}_j)}_{\text{Mg. Cost}} \big)$$

- Multiproduct oligopolistic price competition with risk adjustment
- Quality affects insurance cost:
  - Better hospitals increase claim prices ($\uparrow C$), preventive care reduces hospitalization ($\downarrow C$)

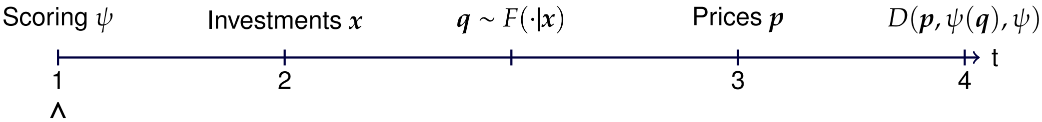
Model



$$\max_{\boldsymbol{x}_f \in \mathbb{R}^{|Q|\times|J_f|}} \underbrace{\int \mathbb{E}[\pi_f(\boldsymbol{q}_f, \boldsymbol{q}_{-f}, \psi)] dF(\boldsymbol{q}_f|\boldsymbol{x}_f)}_{\text{expected insurance profit}} - \underbrace{I(\boldsymbol{x}_f, \boldsymbol{\mu}_f)}_{\text{investment cost}}$$

- Choose investment for each product-category
- Rational expectations about rivals' investments based on market observables (Sweeting, 2009)
- Heterogenous convex investment costs ⇒ equilibrium quality effects

Model



- No optimality imposed on designer's experimentation

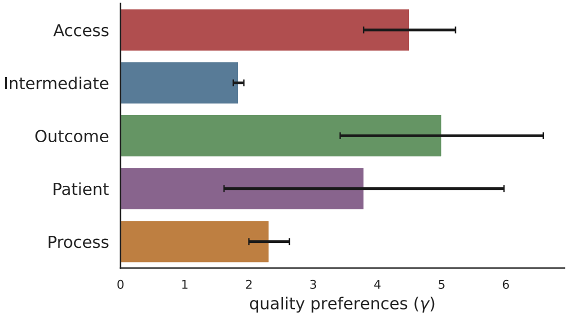
Identification

- Supply model identified from profit optimality conditions
- Revealed preferences identify consumers' WTP for scores
  - **Cannot** tell if WTP comes from beliefs about quality or preferences
  - Example: only readmission risk quality (scalar)
    - Consumers WTP \$100 for plan to have 4 instead of 3 stars, all else equal
    - $\Delta\mathcal{E}(q) = 1\%$ and $\gamma = \$100$ or $\Delta\mathcal{E}(q) = 5\%$ and $\gamma = \$20$?
- Intuition: if consumers understand design, posterior beliefs are bounded
  - Bounds on beliefs + WTP $\implies$ bounds on preferences
    - Consumers knows that $\psi(q) = 3 \iff q \in [0.8\%, 1\%)$ and $\psi(q) = 4 \iff q \in [0, 0.3\%)$
    - Therefore $\Delta\mathcal{E}(q) \in (0.5\%, 1\%) \implies \gamma \in (100, 200)$

$\Rightarrow$ Variation in scoring design generates additional bounds and tightens identification

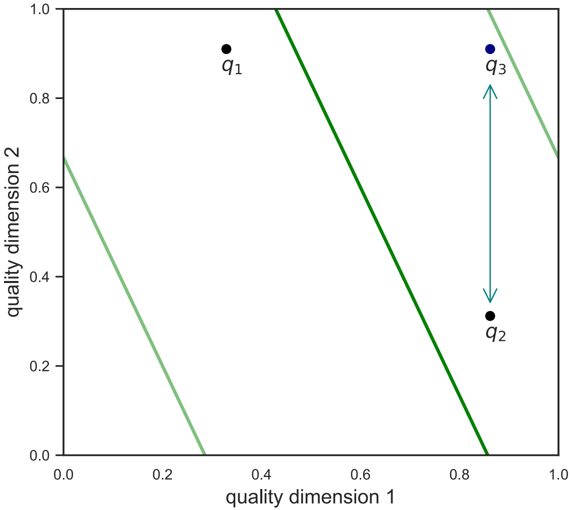
Key Estimates - Information Assymetry

- 1 std. dev. in Outcomes $\approx$ \$1463 in OOP
- Incomplete info lowers surplus by \$185.9 (keeping supply fixed)
- Two sources of information asymmetry:



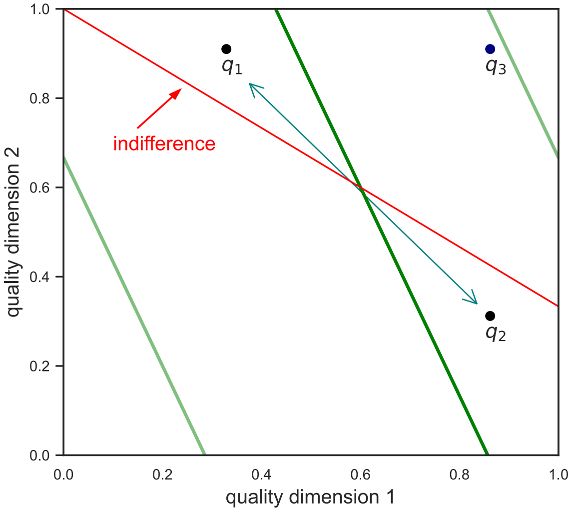
Key Estimates - Information Assymetry

- 1 std. dev. in Outcomes ≈ \$1463 in OOP
- Incomplete info lowers surplus by \$185.9 (keeping supply fixed)
- Two sources of information asymmetry:
  1. Within-scores:
     Best 4-star worth \$257.1 more than worst



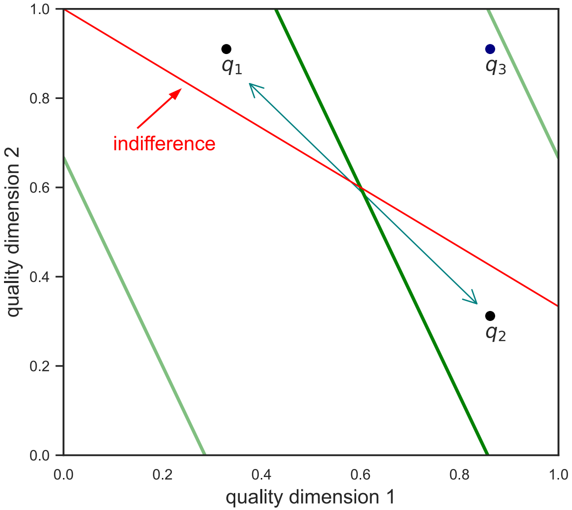
Key Estimates - Information Assymetry

- 1 std. dev. in Outcomes ≈ \$1463 in OOP
- Incomplete info lowers surplus by \$185.9 (keeping supply fixed)
- Two sources of information asymmetry:
  1. Within-scores:
     Best 4-star worth \$257.1 more than worst
  2. Across-scores:
     22.4% of plans ranked opposite to preferences



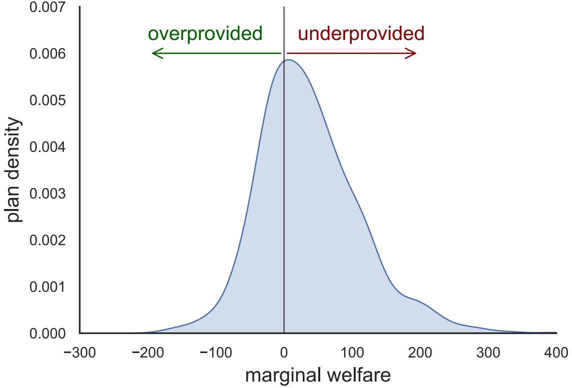
Key Estimates - Information Assymetry

- 1 std. dev. in Outcomes ≈ \$1463 in OOP
- Incomplete info lowers surplus by \$185.9 (keeping supply fixed)
- Two sources of information asymmetry:
  1. Within-scores: **5%**
     Best 4-star worth \$257.1 more than worst
  2. Across-scores: **95%**
     22.4% of plans ranked opposite to preferences



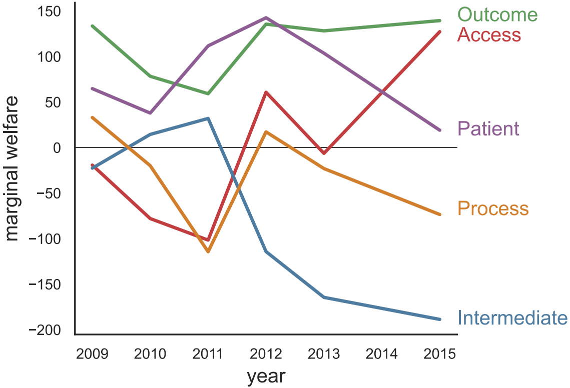
Key Estimates - Quality provision

- Avg insurance markup of 11.2%
  - For top insurers: avg marginal cost is \$771
  - Curto et. al (2019): medical cost is \$680
- Median investment $=$ 24% of insurance profits
- Quality is inefficiently provided:
  1. On aggregate: underprovided $dTW/dq =$\$42.8



Key Estimates - Quality provision

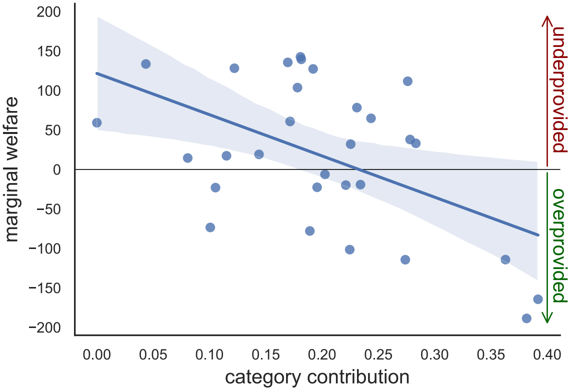
- Avg insurance markup of 11.2%
  - For top insurers: avg marginal cost is \$771
  - Curto et. al (2019): medical cost is \$680
- Median investment $=$ 24% of insurance profits
- Quality is inefficiently provided:
  1. On aggregate: underprovided $dTW/dq =$\$42.8
  2. By category: mixed over/under provision



Key Estimates - Quality provision

- Avg insurance markup of 11.2%
  - For top insurers: avg marginal cost is \$771
  - Curto et. al (2019): medical cost is \$680
- Median investment $=$ 24% of insurance profits
- Quality is inefficiently provided:
  1. On aggregate: underprovided $dTW/dq =$\$42.8
  2. By category: mixed over/under provision

     $\Rightarrow$ affected by scoring design



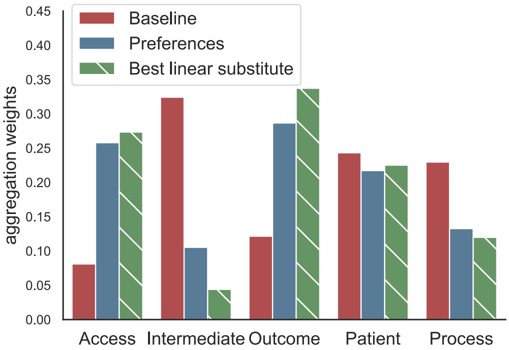
Outline

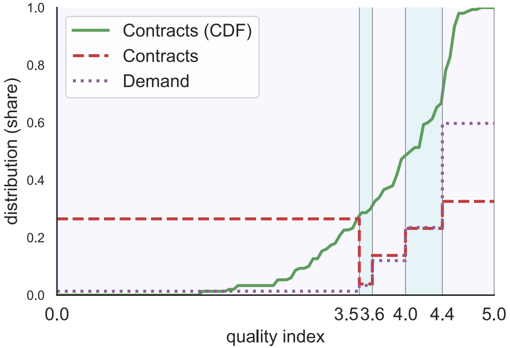
The Designer's Problem

$$\max_{\psi \in \Psi} \quad \mathbb{E}_{\boldsymbol{q}} \Big[\underbrace{CS(\psi, \boldsymbol{q})}_{\text{Consumer surplus}} + \underbrace{\sum_f V_f(\psi, \boldsymbol{q}) - I(\boldsymbol{x}_f^*(\psi), \mu_f)}_{\text{Insurer profit}} \,|\boldsymbol{x}^*(\psi) \Big]$$

- Subject to equilibrium behavior:
  - Firms update investments, prices, beliefs about rivals
  - Consumers update beliefs given design and realized scores
- Empirical scoring design methodology:
  1. Represent scores as composition of aggregator and cutoffs
  2. Use equivalence of scores to distribution over posterior beliefs (Aumann and Maschler, 1995)

Solution: Best Linear Design





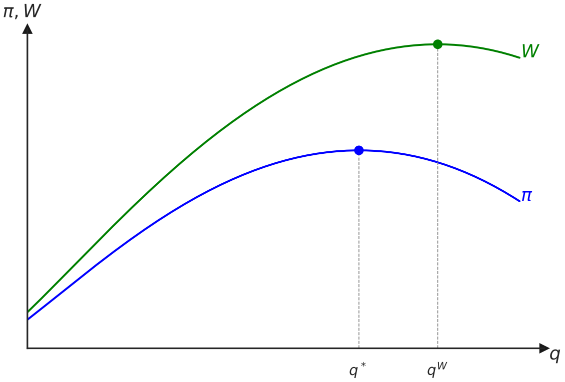
1 Pooling at the bottom: first score pools all low qualities

2 Aggregator: optimal weighting scheme, increase reward on dimensions consumers value

3 Limited granularity: use only five scores; four partition higher quality

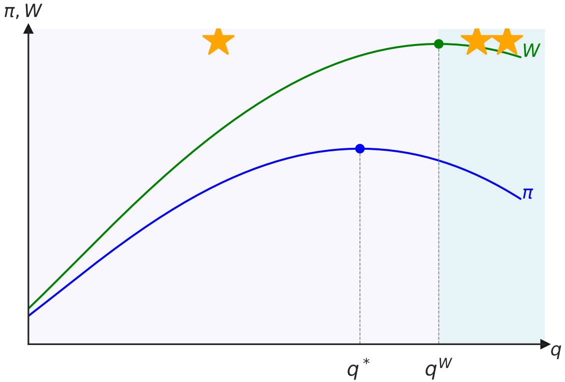
Decomposing the Design: Pooling at the Bottom

- Market power over quality (Spence, 1975; Crawford et al., 2019) : firms under-invest even under full info



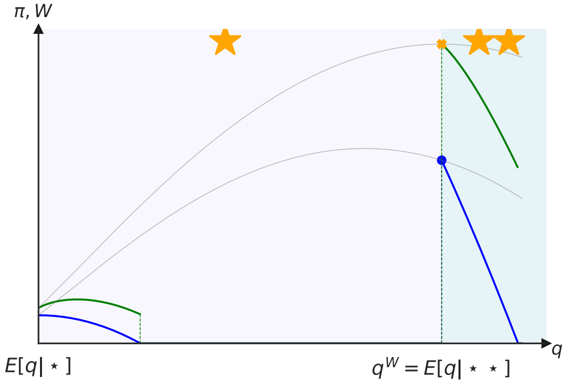
Decomposing the Design: Pooling at the Bottom

- Market power over quality (Spence, 1975; Crawford et al., 2019) : firms under-invest even under full info



Decomposing the Design: Pooling at the Bottom

- Market power over quality (Spence, 1975; Crawford et al., 2019) : firms under-invest even under full info
- Delegation equivalence (Zapechelnyuk, 2020) : certification $\iff$ q^w or 0



Decomposing the Design: Pooling at the Bottom

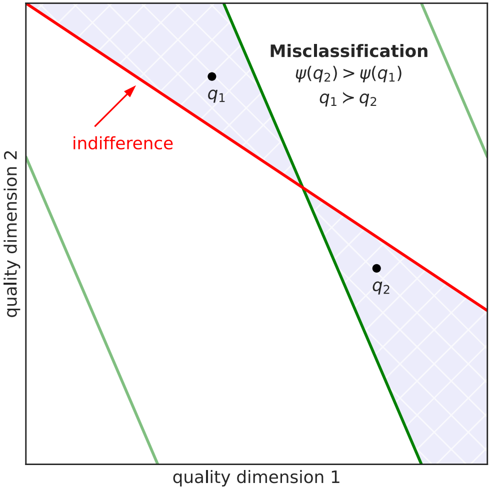
- Market power over quality (Spence, 1975; Crawford et al., 2019) : firms under-invest even under full info
- Delegation equivalence (Zapechelnyuk, 2020) : certification $\iff$ q^w or 0

- **Penalizes underprovision with ↓ demand: 35% of welfare gain** (certification)
  - 62.6% of contract would receive 1 star in baseline, only 26.5% in equilibrium
  - Serve only 1.3% of consumers
  - Quality is 4.3% higher in equilibrium

Decomposing the Design: Aggregator

Aggregation produces two problems:

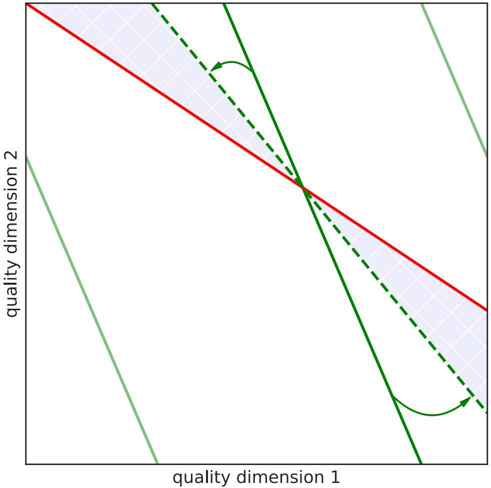
1 Across-scores information asymmetry:



Decomposing the Design: Aggregator

Aggregation produces two problems:

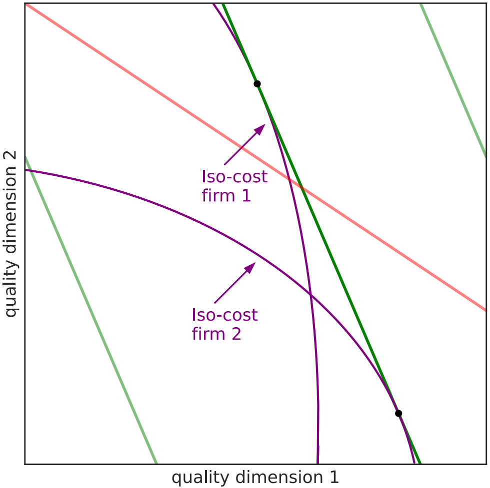
1. Across-scores information asymmetry:
   - Reduced by 97.1% under new weights



Decomposing the Design: Aggregator

Aggregation produces two problems:

1. Across-scores information asymmetry:
   - Reduced by 97.1% under new weights
2. Multitasking moral hazard (Holmstrom and Milgrom, 1991)
   - Firms' allocations ignore preferences

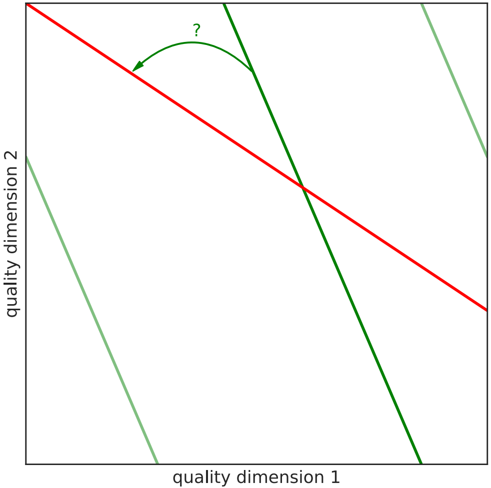


Decomposing the Design: Aggregator

Aggregation produces two problems:

1. Across-scores information asymmetry:
   - Reduced by 97.1% under new weights

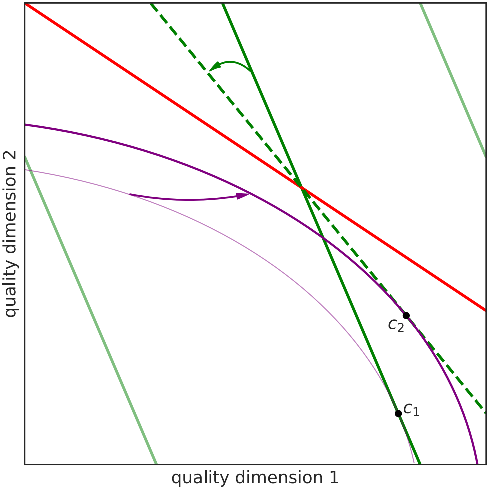
2. Multitasking moral hazard (Holmstrom and Milgrom, 1991)
   - Firms' allocations ignore preferences



Decomposing the Design: Aggregator

Aggregation produces two problems:

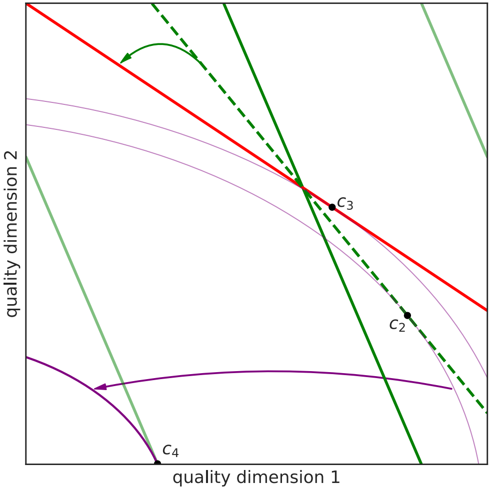
1. Across-scores information asymmetry:
   - Reduced by 97.1% under new weights
2. Multitasking moral hazard (Holmstrom and Milgrom, 1991)
   - Firms' allocations ignore preferences



Decomposing the Design: Aggregator

Aggregation produces two problems:

1. Across-scores information asymmetry:
   - Reduced by 97.1% under new weights
2. Multitasking moral hazard (Holmstrom and Milgrom, 1991)
   - Firms' allocations ignore preferences
3. Solution accounts for cost heterogeneity
   - Convex costs vs. (mostly) concave demand gains



Decomposing the Design: Aggregator

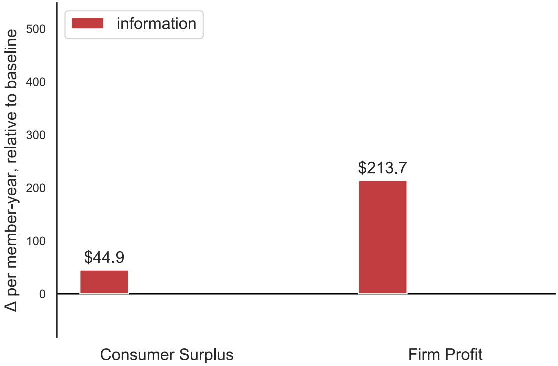
- **Pooling at the bottom + optimal aggregator account for 94% of welfare gains**
  - Pooling increases overall investment
  - Optimal aggregation improves informativeness and allocative efficiency of investments

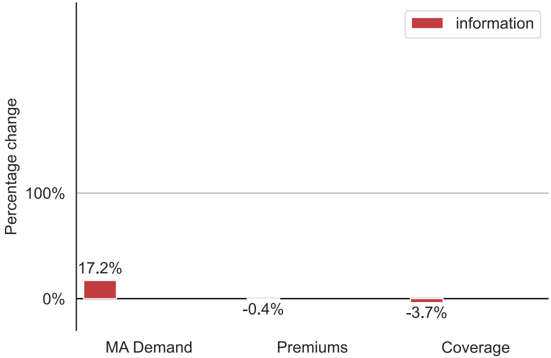
  $\Rightarrow$ High welfare value from optimal certification

Decomposing the Design: Granularity

- Why only five scores at the top?
- Trade-off: efficiency vs. product variety
  - More scores allow more investment actions for firms (delegation equivalence)
  - More actions allow for more heterogeneity: lower quality at lower prices
  - But also more deviations away from efficient production and towards profit maximization
- Limiting factor: ability to generate separating choices for heterogenous firms

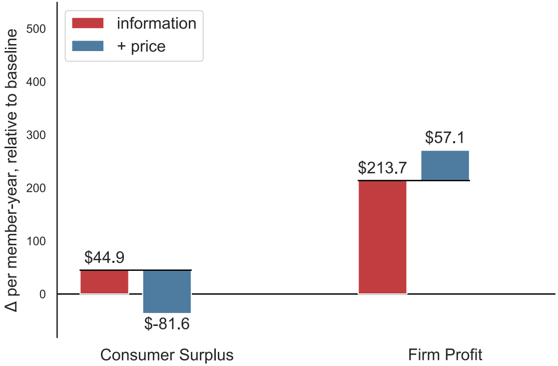
Welfare

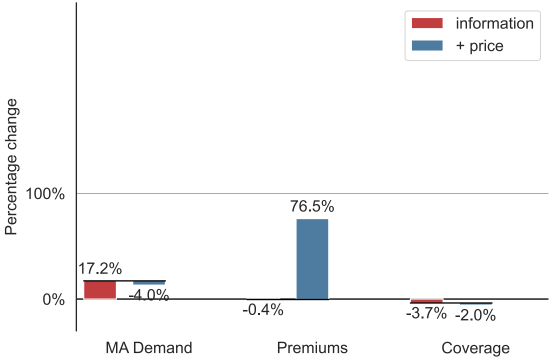




- Holding prices and quality changes information:
  - Products are easier to choose, fewer mistakes
  - Large MA expansion: Consumers select quality that offsets switching costs

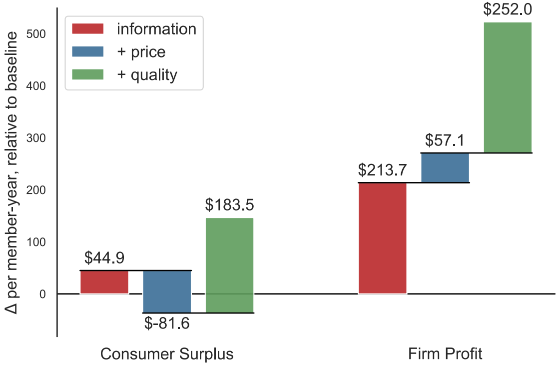
Welfare

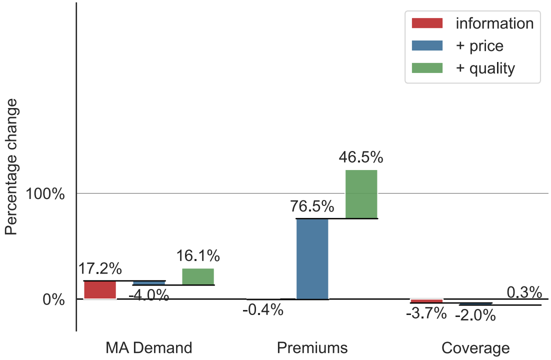




- Holding quality, change information and prices:
  - New information reveals vertical differentiation across products
  - Firms exert market power over prices capturing surplus

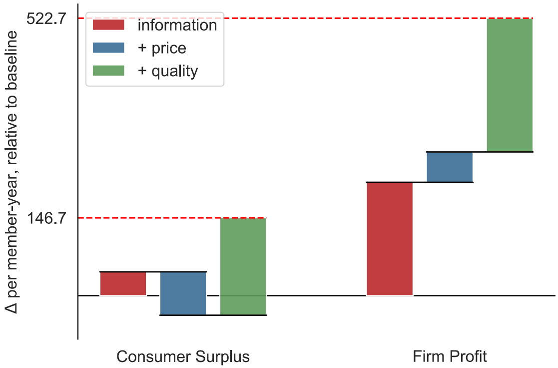
Welfare

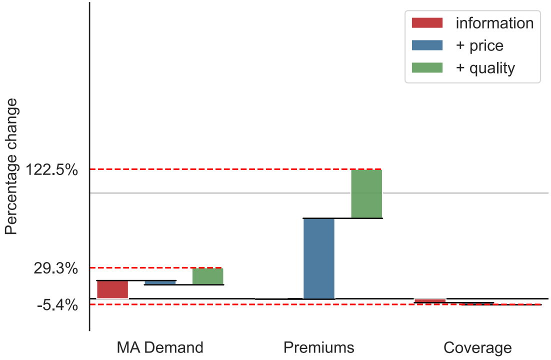




- Full equilibrium changes:
  - Total welfare increases by 285%, firms' benefit from additional expansion
  - Welfare gains primarily driven by quality regulation effect

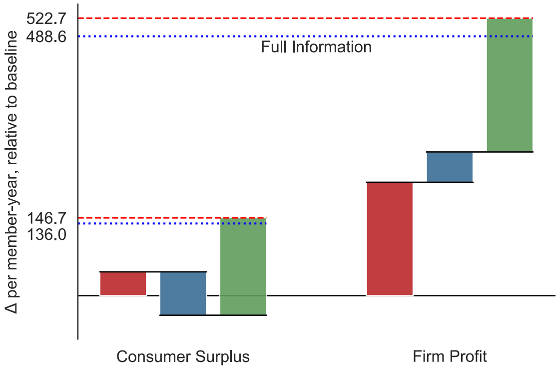
Welfare

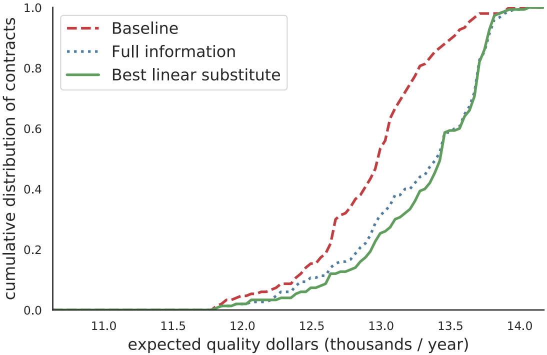




- Total welfare is \$669.3 per member per year
- Surplus gain ≈ 2.4 total monthly premiums

Welfare





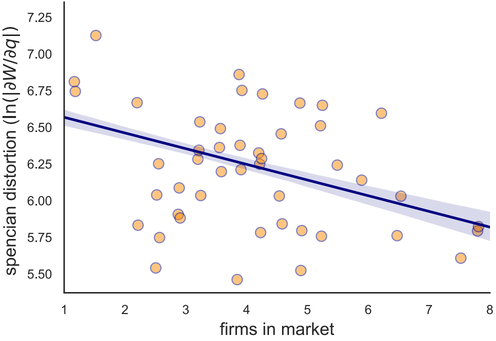
- Full information allows exercise of market power over quality, reduces welfare
- New scores dominate only because of equilibrium quality effects

Competition and Regulation

- Markups increase by 37.2% under new design
  - $\uparrow$ vertical differentiation

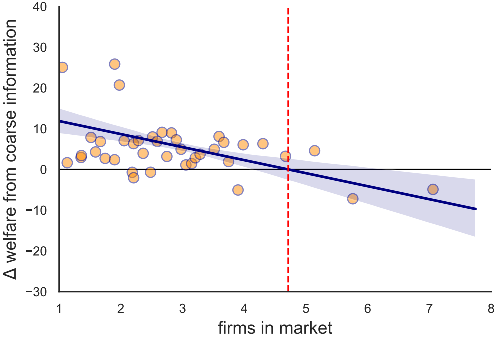
    $\Rightarrow \downarrow$ 7.3% semi-elasticity of substitution across

- Additional competing firm associated with:
  - $\downarrow$ 0.3pp markup increase
  - $\uparrow$ 1.8pp quality increase
  - $\downarrow$ 5.4% spencian distortion in full information



Competition and Regulation

- Markups increase by 37.2% under new design
  - ↑ vertical differentiation
    ⇒ ↓ 7.3% semi-elasticity of substitution across
- Additional competing firm associated with:
  - ↓ 0.3pp markup increase
  - ↑ 1.8pp quality increase
  - ↓ 5.4% spencian distortion in full information
- Gains from coarse information vanish at 5 firms
  - 9.9% of consumers better under full info



Explaining the Differences in Designs

Why is CMS's design systematically different than the optimal?

1. Strong preferences for quality chronic care (Intermediate) and lower-cost hospitals (Outcome)
   - Paternalism or dynamic considerations for future subsidized care
   - Nudging the market with scores is enormously costly:

     $\Rightarrow$ would have to value 10% reallocation of quality by \$14 billion, orders of magnitude above cost

2. CMS might be risk averse to misrepresenting consumers' preferences
   - Medicare plays a delicate political and social role, objective might be $\max_{\psi \in \Psi} \min_{\gamma \in \Gamma} TW(\psi, \gamma)$
   - CMS's weight nearly optimal for robust design

     $\Rightarrow$ optimal robust design improves upon CMS by using the same economic forces as before

Conclusions

- Scores are powerful quality regulation policies:
  - Adapting MA's design to equilibrium effects increases welfare by $43 billion
- Suggests potential for redesigning scores using theory and empirical work
  - Challenges policy focus on granularity, (ex-ante) informativeness, cognitive bias considerations
  - $\Rightarrow$ A simple well-designed sticker can outperform full information outcomes
- **Empirical Scoring Design** methodology for disclosure policies
  - Data-driven solution for an extensive policy problem

Thank You!

bvatter@stanford.edu