

RAND Medicare Advantage (MA) and Part D Contract Star Ratings
Technical Expert Panel October 30th 2018 Meeting

PRESENTATION

Cheryl L. Damberg and Susan M. Paddock



For more information on this publication, visit www.rand.org/t/CF394

Published by the RAND Corporation, Santa Monica, Calif.

© Copyright 2019 RAND Corporation

RAND® is a registered trademark.

Limited Print and Electronic Distribution Rights

This document and trademark(s) contained herein are protected by law. This representation of RAND intellectual property is provided for noncommercial use only. Unauthorized posting of this publication online is prohibited. Permission is given to duplicate this document for personal use only, as long as it is unaltered and complete. Permission is required from RAND to reproduce, or reuse in another form, any of its research documents for commercial use. For information on reprint and linking permissions, please visit www.rand.org/pubs/permissions.

The RAND Corporation is a research organization that develops solutions to public policy challenges to help make communities throughout the world safer and more secure, healthier and more prosperous. RAND is nonprofit, nonpartisan, and committed to the public interest.

RAND's publications do not necessarily reflect the opinions of its research clients and sponsors.

Support RAND
Make a tax-deductible charitable contribution at
www.rand.org/giving/contribute

www.rand.org

RAND Technical Expert Panel Medicare Advantage (MA) and Part D Star Ratings

October 30, 2018



Recap on Threshold Discussion and Update on Analyses

Susan Paddock

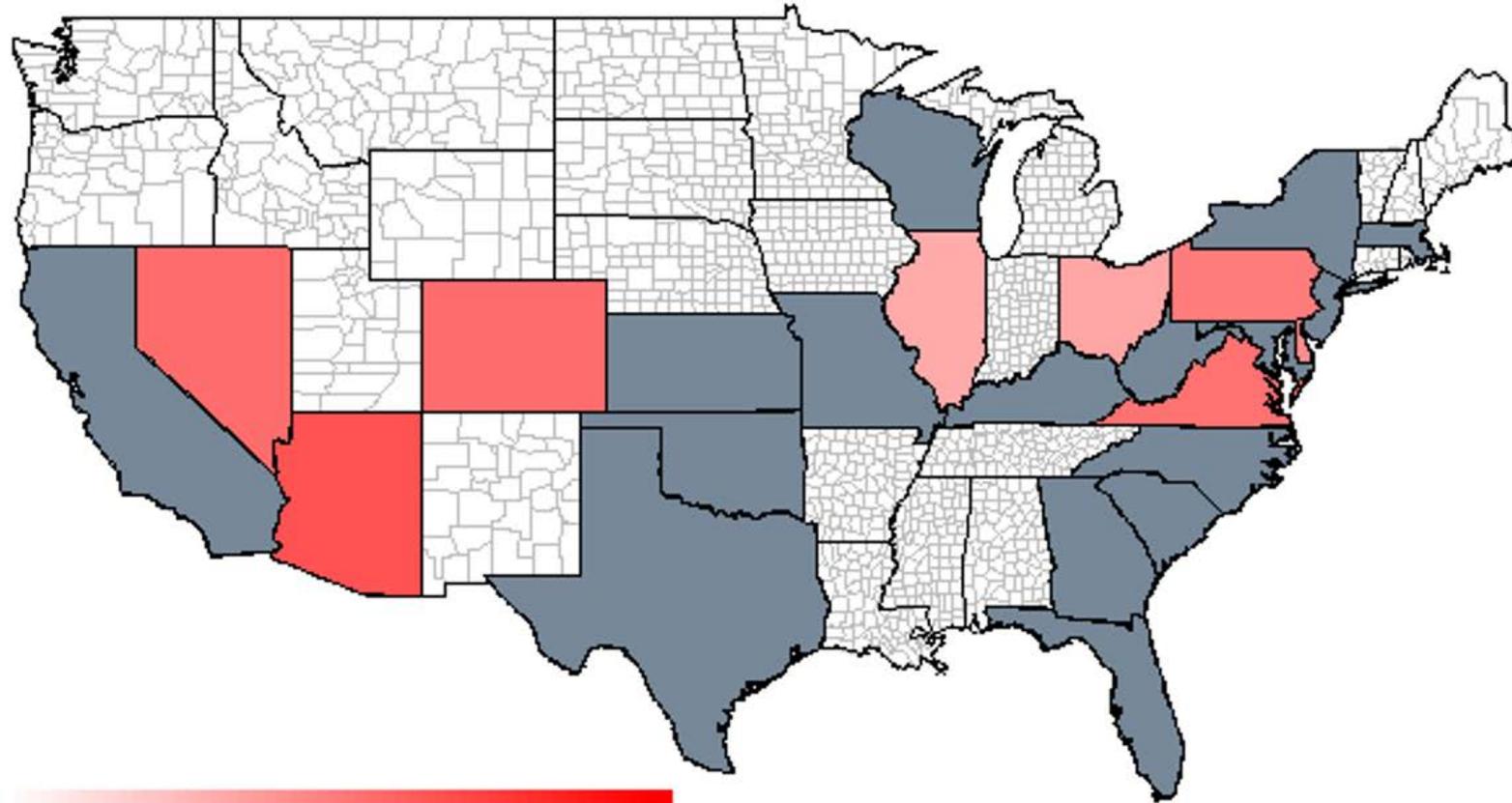


Update on Utility and Feasibility of Constructing and Reporting Star Ratings at Smaller Geographic Units

Justin Timbie, Andy Bogart, Susan Paddock, Cheryl Damberg, and Marc Elliott



Performance can vary substantially across an MA contract's service area



To illustrate, the service area for one HMO contract includes 25 states

Range in performance:
67.1% (Illinois)
to
83.9% (Arizona)

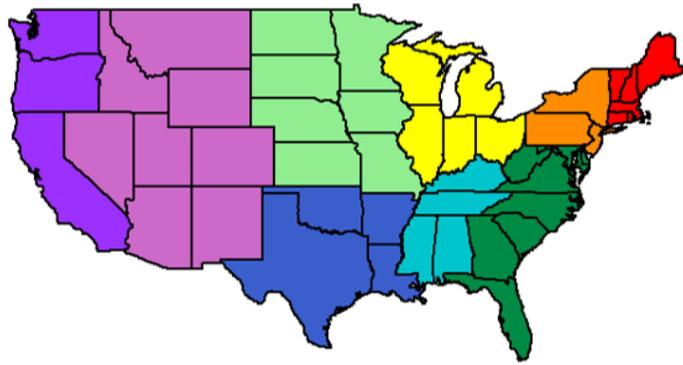
17 states within the service area have small sample sizes

0 20 40 60 80 100
Average Score

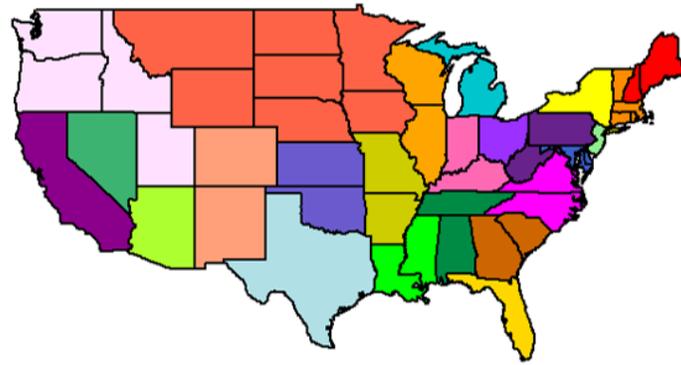
Improving/Maintaining Physical Health (HOS)

We used a three-step approach to simulate geography-based reporting

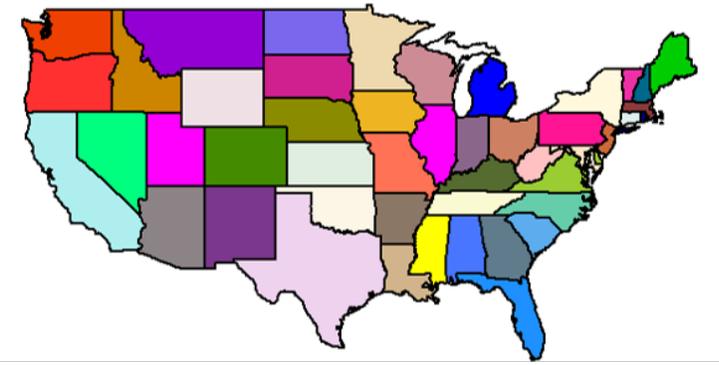
1. Split contracts into smaller units defined by geography
 - We use the term “geographic reporting units” or GRUs



Census Divisions



MA Regions



States/Territories

- Each contract can be split into at most **9** GRUs defined by Census Divisions, **26** GRUs defined by MA Regions, and **54** GRUs defined by States/Territories

We used a three-step approach to simulate geography-based reporting

2. Calculate GRU-level summaries for 7 performance measures
 - Number of GRUs that meet minimum sample size and/or reliability criteria
 - Percentage of MA enrollment in GRUs that meet minimum criteria
3. Compare distributions of performance across GRUs

We simulated geographic-based reporting using 7 quality measures

1. HEDIS: Breast Cancer Screening
2. HEDIS: Osteoporosis Management in Women who had a Fracture
3. HEDIS: Diabetes Care - Blood Sugar Controlled
4. HOS: Improving/Maintaining Physical Health
5. CAHPS: Getting Needed Care
6. CAHPS: Rating of Drug Plan
7. PDE: Diabetes Medication Adherence

We used minimum sample size/reliability requirements listed in the Star Ratings Technical Notes

Measure category	Sample size	Reliability
HEDIS	≥ 500 enrollees & ≥ 30 patients in measure denominator	≥ 0.7 (only if GRU has 500-999 enrollees)
HOS	≥ 30 respondents	--
CAHPS	--	≥ 0.6
PDE	> 30 eligible member- years enrollment	--

A total of 477 contracts were used to conduct geographic simulations

- Contract sample included 421 MA-PD and 56 PDP contracts that received Star Ratings in 2018
 - In the slides that follow we show results for MA-PDs only
- Contract service areas were defined using December 2016 service area files
- Beneficiary-level performance data were submitted by plans
- Beneficiary residence and contract enrollment information were taken from 2016 enrollment file

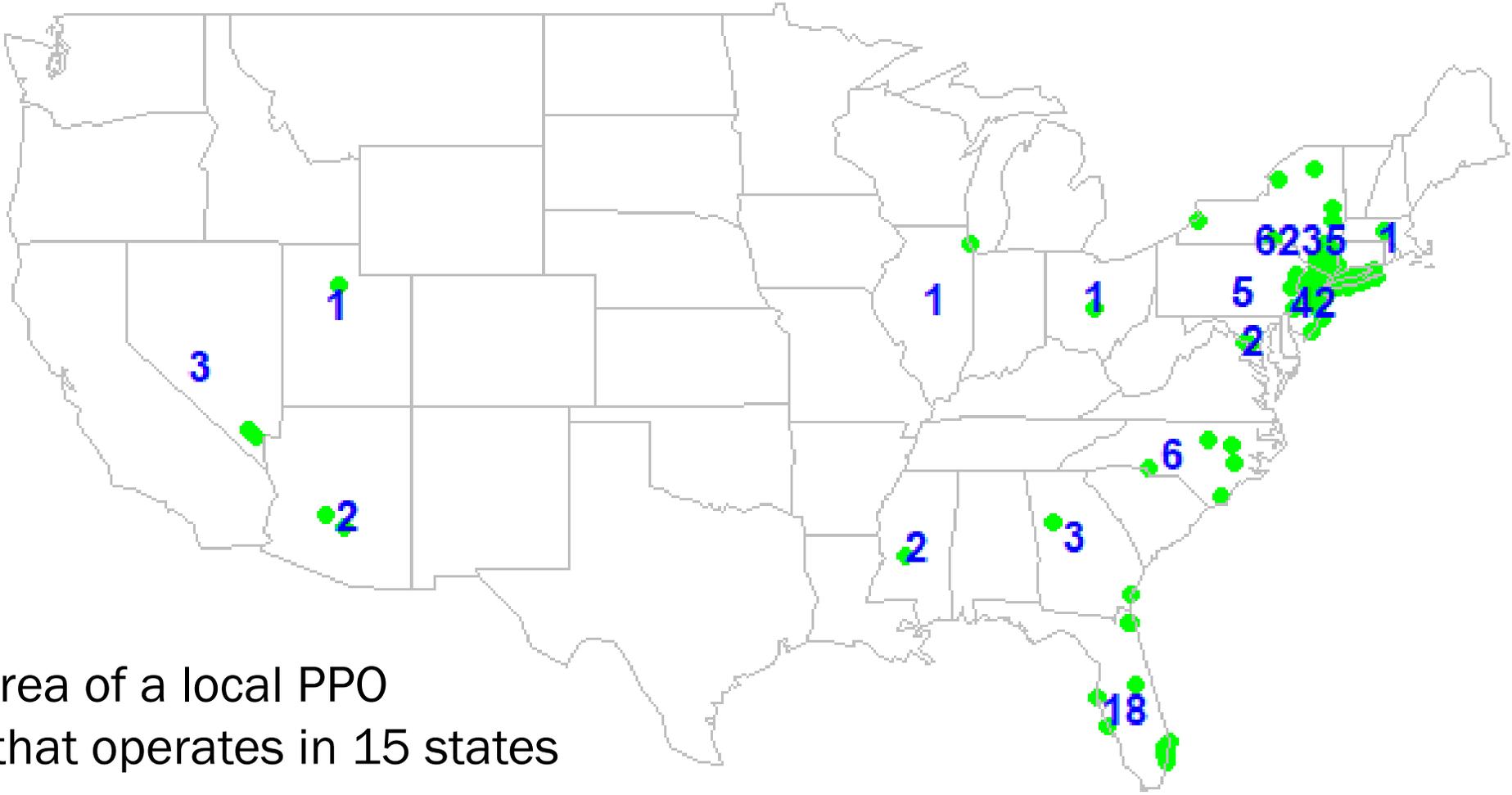
Most MA contracts comprise only a single GRU, although some contracts were split into many units

Number of GRUs within contract	Census Divisions Within Contracts	MA Regions Within Contracts	States Within Contracts	Census Divisions Within Contracts	MA Regions Within Contracts	States Within Contracts
	<i># (%) of contracts</i>			<i>% of MA beneficiaries</i>		
1	316 (75%)	309 (73%)	292 (69%)	53%	51%	47%
2	37 (9%)	31 (7%)	38 (9%)	8%	6%	6%
3	9 (2%)	16 (4%)	20 (5%)	4%	4%	5%
4	9 (2%)	3 (1%)	6 (1%)	3%	1%	2%
5	9 (2%)	7 (2%)	6 (1%)	4%	5%	4%
6-10	41 (10%)	19 (5%)	20 (5%)	29%	7%	8%
11 or more		36 (9%)	39 (9%)		28%	28%

MA enrollment varies widely across GRUs

	Contracts (n=421)	Census Divisions Within Contracts (n=936)	MA Regions Within Contracts (n=1607)	States Within Contracts (n=2204)
Min	206	1	1	1
10th percentile	1,311	2	1	1
25th percentile	3,761	16	4	3
50th percentile	12,218	1,715	58	26
75th percentile	38,780	15,897	5,648	2,904
90th percentile	90,529	51,830	29,068	18,759
Max	1,089,128	1,089,128	1,089,128	1,089,128

The skewed distribution of enrollment may be due to MA contracts that have large service areas with uneven enrollment



Service area of a local PPO contract that operates in 15 states

However, most MA beneficiaries live in GRUs that meet sample size and reliability criteria

Measure	Sample Size & Reliability Criteria	Percent of GRUs that meet sample size and reliability criteria				Percent of MA beneficiaries who reside in GRUs that meet sample size and reliability criteria			
		Contracts	Census Divisions Within Contracts	MA Regions Within Contracts	States Within Contracts	Contracts	Census Divisions Within Contracts	MA Regions Within Contracts	States Within Contracts
HEDIS: Breast Cancer Screening	≥1000 enrollees (or 500-999 if reliability ≥ 0.7) & ≥30 patients	91%	56%	40%	35%	100%	99%	99%	99%
HEDIS: Osteoporosis Mgmt.		62%	35%	23%	18%	97%	95%	93%	92%
HEDIS: Diabetes Care A1c Control		97%	52%	31%	23%	100%	96%	93%	89%
HOS: Improving/Maintaining PCS	≥30 respondents	72%	39%	25%	19%	97%	91%	89%	87%
CAHPS: Getting Needed Care	reliability ≥ 0.6	75%	46%	32%	26%	94%	90%	88%	88%
CAHPS: Rating of Drug Plan		84%	46%	30%	25%	94%	90%	87%	87%
PDE: Diabetes Med. Adherence	>30 member-years	100%	68%	51%	46%	100%	100%	100%	100%

Approximately three quarters of MA beneficiaries live in GRUs for which all 7 measures can be reported

Number of Measures with Adequate Sample Size/Reliability	Percent of GRUs				Percent of MA Beneficiaries			
	Contracts	Census Divisions Within Contracts	MA Regions Within Contracts	States Within Contracts	Contracts	Census Divisions Within Contracts	MA Regions Within Contracts	States Within Contracts
0 of 7	0%	31%	48%	53%	0%	0%	0%	<1%
1 of 7	2%	7%	8%	9%	0%	<1%	<1%	<1%
2 of 7	4%	7%	6%	6%	<1%	<1%	1%	2%
3 of 7	7%	5%	6%	6%	<1%	1%	3%	3%
4 of 7	8%	5%	4%	4%	<1%	2%	3%	4%
5 of 7	11%	9%	6%	5%	3%	6%	6%	6%
6 of 7	17%	9%	6%	5%	9%	11%	14%	11%
7 of 7	52%	26%	16%	12%	87%	79%	73%	73%

Comparison of contract-level Star Ratings and state-level GRU Star Ratings

Measure: Breast Cancer Screening

	States within Contract Stars					
Contract Stars	1	2	3	4	5	Total GRUs*
1	100%	--	--	--	--	12
2	4%	85%	12%	--	--	183
3	1%	20%	71%	7%	1%	324
4	--	2%	26%	65%	7%	205
5	--	--	--	3%	97%	33

Note: Row percentages are displayed. 2018 star thresholds were applied.

*Only GRUs that met both sample size and reliability criteria are included. Three GRUs that met criteria are not displayed because the contract was ineligible for the measure due to low reliability.

Summary of preliminary findings

- 69% of contracts operate within a single state (smallest GRU simulated), so simulations reflect status quo for these contracts
- MA enrollment by GRU is highly skewed, driven largely by contracts with large but unevenly populated service areas
- A small percentage of GRUs meet minimum sample size and reliability criteria, but account for most MA beneficiaries
- 3/4 of MA beneficiaries live in GRUs where all 7 measures can be reported
- Contract- and GRU-based measure stars differ somewhat

Questions for TEP

1. Employer group health plans are included in the analysis. Should we reconsider this decision?
2. Should we limit measurement to beneficiaries in GRUs that meet the reliability/sample size criteria?
 - In cases where GRUs do not meet the criteria, would reverting to *contract-level* reporting be appropriate?

Questions for TEP

3. Given the large % of beneficiaries living in state-level GRUs with sufficient sample sizes and reliability, should we consider *smaller* geographic areas?
 - Are there other methods to disaggregate contracts we should consider?
4. Does it make sense to add more measures and/or simulate Star Ratings?
 - What additional analyses should RAND consider?
5. What are the benefits of constructing Star Ratings within geographic areas?

Results are similar when we impose more conservative criteria: 0.7 reliability for all measures and sample size greater than or equal to 100 for CAHPS

Measure	Sample Size, Reliability Threshold	Percent of GRUs that meet sample size and reliability criteria				Percent of MA beneficiaries who reside in GRUs that meet sample size and reliability criteria			
		Contracts	Census Divisions Within Contracts	MA Regions Within Contracts	States Within Contracts	Contracts	Census Divisions Within Contracts	MA Regions Within Contracts	States Within Contracts
HEDIS: Breast Cancer Screening	500, 0.7	89%	56%	40%	35%	99%	99%	99%	99%
HEDIS: Osteoporosis Mgmt	500, 0.7	62%	35%	23%	18%	97%	95%	93%	92%
HEDIS: Diabetes Care A1c Control	500, 0.7	97%	52%	31%	23%	100%	96%	93%	89%
HOS: Improving/Maintaining PCS	30, 0.7	61%	35%	24%	18%	94%	88%	88%	86%
CAHPS: Getting Needed Care	100, 0.7	60%	33%	20%	15%	86%	74%	72%	70%
CAHPS: Rating of Drug Plan	100, 0.7	77%	36%	21%	16%	89%	75%	70%	69%
PDE: Diabetes Med. Adherence	30, 0.7	99%	68%	51%	46%	100%	100%	100%	100%

Estimating Contract Performance by Geographic Region

**RAND Technical Expert Panel
Star Ratings for Medicare Advantage and Prescription Drug Plans**

**Susan Paddock, PhD; Marc Elliott, PhD; Justin Timbie, PhD; Andy Bogart MS,
Cheryl Damberg, PhD
October 30, 2018**



Topics addressed

- Challenges of reporting contract performance by geographic region
- Potential approaches to estimate contract performance by geographic region

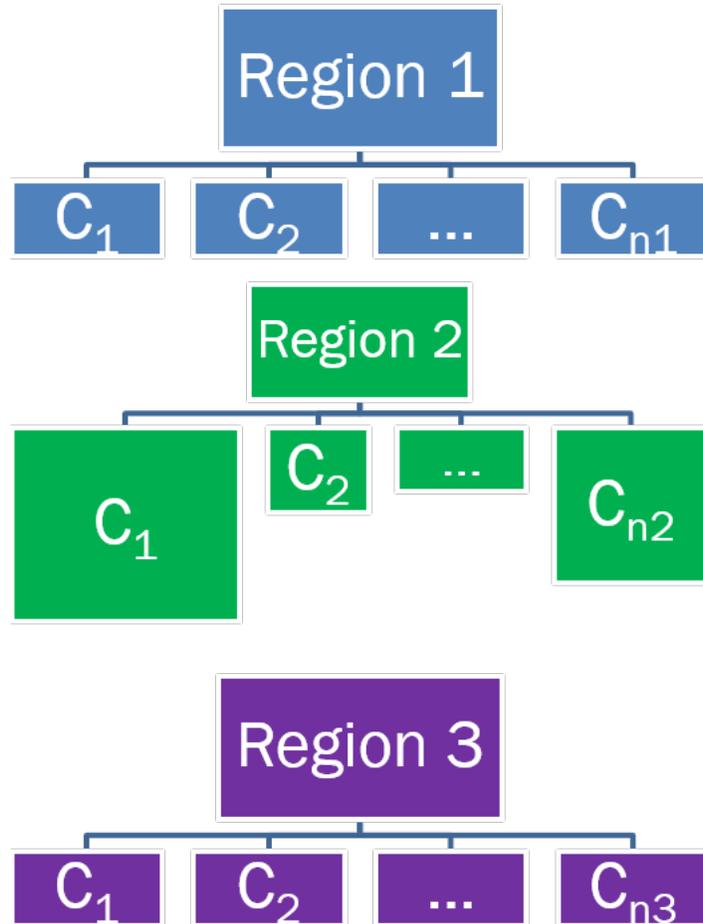
Challenges of geographic reporting

- Small sample sizes per contract in a geographic region might lead to unstable performance estimates
- The severity of the challenge will depend on several factors:
 - Size of geographic region
 - Amount of data available by measure
 - Survey or medical record review versus administrative data-based measures
 - Whether the measure applies to a small number of beneficiaries
- What strategies could be used to address these challenges?

1) Use data only from the contract, geographic region, and time period of interest

- Advantages
 - Easy to understand
 - Precise estimates might be obtained for some measures and some combinations of contracts and geographic regions
 - Aligned with current Star Ratings methodology
- Disadvantages
 - Some contracts and some measures will have too small of denominators to be measured by region
 - Ignores potentially informative additional data that could be used to obtain more precise estimates

2a) Borrowing information across contracts



Assume: Overall performance within region is informative about performance for contracts in the region

Borrow information from all contracts within a region to increase stability of estimates for each contract within region*

“C” denotes contract. n_1 , n_2 , n_3 are the total numbers of contracts in region 1-3, respectively.

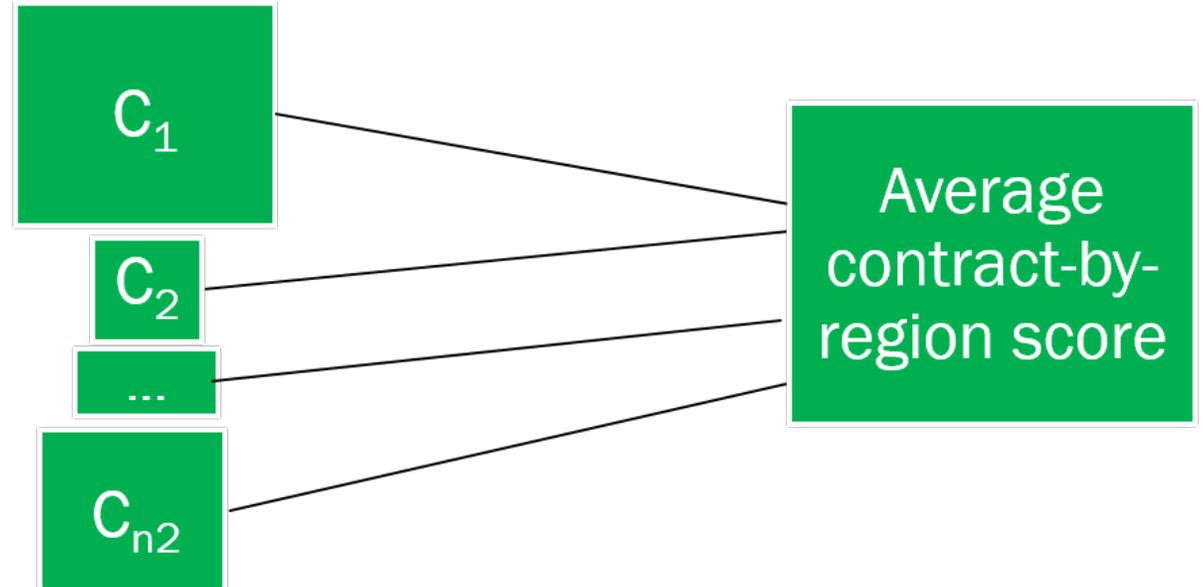
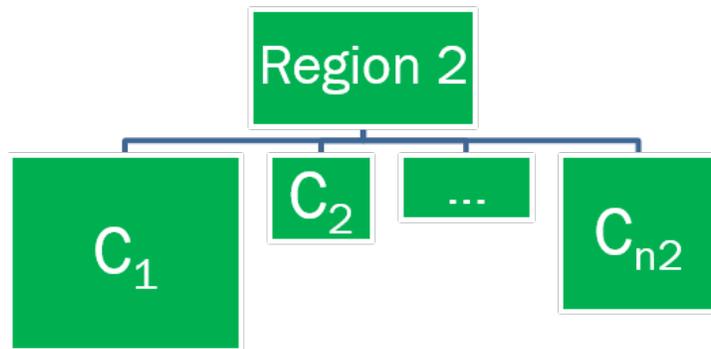
* Ghosh M, Rao JNK (1994). Small Area Estimation: An Appraisal. *Statistical Science*, 9(1), 55-93

2a) Borrowing information across contracts

Use average of contract scores within region to stabilize estimates for each contract within the region

A contract-by-region score would be estimated as a *weighted average* of:

- a) Raw contract-by-region score
- b) Average of all contract-by-region scores

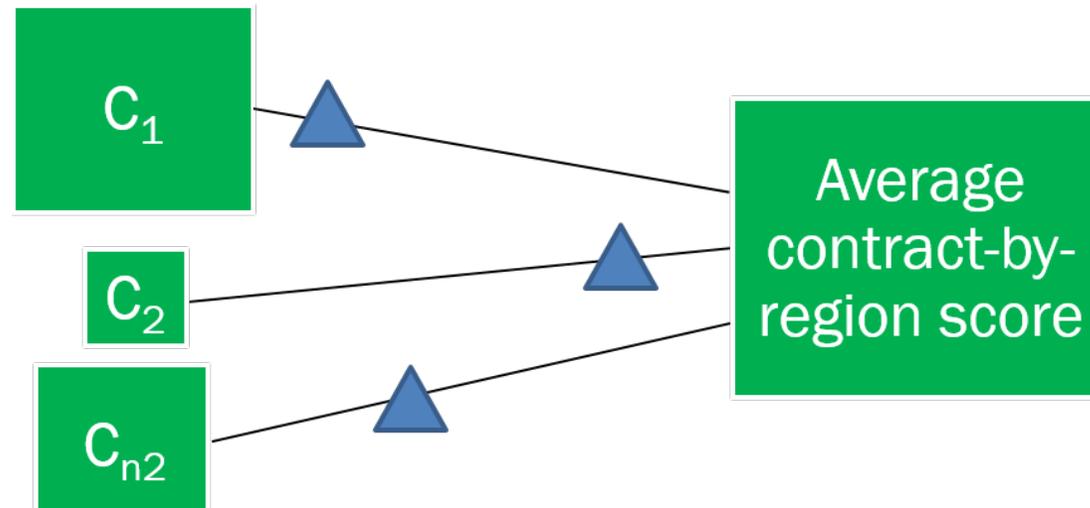


2a) Borrowing information across contracts

The weighting of the average contract-by-region score is greater when:

- A contract has little data in the region
- There is little variation in the contract-by-region scores – the average is more representative of overall performance when this variation is low

The weighted average (▲) is called a *shrunk estimate*



2a) Borrowing information across contracts

- Advantages
 - Reduces the mean squared error of the set of contract-by-region estimates
 - Increases the precision of contract-by-region estimates
 - Some precedent in CMS VBP and reporting programs (e.g., Hospital Compare outcomes measures)
- Disadvantages
 - Requires agreement that shrinkage is appropriate and acceptable
 - Could lead to biased estimates, particularly for small contracts in regions
 - Small contracts are likely to have estimates pulled toward the average contract score
 - Shrunk estimates could not be compared to cut points derived under current Star Rating methodology
 - Shrunk estimates are compressed relative to non-shrunk estimates

2b) Borrowing information across contracts: Incorporating contract characteristics

- The shrinkage ‘target’ of the average contract-by-region score might not be ideal
- The target could be modified to include contract characteristics, such as contract size, SNP status, beneficiary summary statistics (e.g., %LIS/DE), and sponsor
- Advantages
 - Similar advantages as for borrowing strength without contract characteristics
 - Improve the precision and accuracy of contract-by-region estimates
 - Improve the precision of contract comparisons within region
 - Adjust for bias from applying shrinkage for small contracts within region
- Disadvantages
 - Requires agreement that shrinkage is appropriate and acceptable
 - Shrunken estimates could not be compared to cut points derived under current Star Rating methodology

3) Pool contract+region data over time

- Advantages
 - Easy to understand and implement
 - Precedent in other CMS value-based purchasing and reporting programs
 - Estimates on a comparable (non-shrunk) scale relative to Star Ratings thresholds
- Disadvantages
 - Uses current and older data

Discussion questions

- What are the practical strengths and weaknesses of the proposed approaches of shrinkage and pooling information to increase precision of geographic reporting for contracts?
- Which contract characteristics are important to consider for inclusion in a shrinkage approach?
- What criteria might be of interest to stakeholder groups in evaluating these options for increasing the precision of contract performance estimates within region?