

Plan, Geographical, and Temporal Variation of Consumer Assessments of Ambulatory Health Care

Alan M. Zaslavsky, Lawrence B. Zaboriski, and Paul D. Cleary

Objective. To quantify contributions of health plans and geography to variation in consumer assessments of health plan quality.

Data Sources. Responses of beneficiaries of Medicare managed care plans to the Consumer Assessment of Health Plans Study (CAHPS[®]) survey. Our data included more than 700,000 survey responses assessing 381 Medicare managed care (MMC) contracts over a period of five years.

Study Design. The survey was administered to a nationally representative sample of beneficiaries of Medicare managed care plans.

Principal Findings. Member assessments of their health plans, customer service functions, and prescription drug benefits varied most across health plans; these also varied the most over time. Assessments of direct interactions with doctors and their practices were more affected by geographical location, and these assessments were quite stable over time. A health plan's global rating often changed significantly between consecutive years, but only rarely were there such changes in ratings of care or doctor. Nationally, mean assessments tended to decrease over the study period.

Conclusions. Our findings suggest that ratings of plans and reports about customer service and prescription access are affected by plan policies, benefits design, and administrative structures that can be changed relatively quickly. Conversely, assessments of other aspects of care are largely determined by characteristics of provider networks that are relatively stable. A consumer survey is unlikely to detect meaningful changes in quality of care from year to year unless quality improvement measures are developed that have substantially larger effects, possibly through area-wide initiatives, than historical temporal variations in quality.

Key Words. Quality, health care, small-area variation, variance components, Medicare

The Centers for Medicare and Medicaid Services (CMS) have collected Consumer Assessment of Health Plans Study (CAHPS[®]) survey data from beneficiaries of Medicare managed care plans that provide care to almost 5 million patients (Goldstein et al. 2001). A single vendor has collected the

data in consecutive years since early 1998 using a consistent protocol, facilitating comparisons of quality ratings across health plans, geographical regions, and time.

A previous study (Zaslavsky, Landon et al. 2000) found significant variation in Medicare managed care CAHPS scores by region, Metropolitan Statistical Area (MSA), and plan. The most systematic variation occurred for ratings of the plan, and a large fraction of this variation was among plans. For ratings of care, a large fraction of variation was across geographical (MSA and regional) units. This is consistent with our expectation that ratings of plan are largely affected by health plan administrative policies. On the other hand, ratings of care and doctors are determined by general characteristics of the health care delivery system in an area, and networks of health care providers often contract with multiple plans.

Using five consecutive years of available CAHPS data, it is possible to assess how CAHPS scores change over time. The much larger sample size (about eight times that used in previous analyses) also allows more refined estimation of geographical effects, including variation at the state level and substate variations within the same plan. Because CAHPS scores are commonly used to compare plans and evaluate improvement, it is important to know to what extent the scores reflect the quality of the plan at a particular time rather than general characteristics of the area it serves. In this article we address the following questions:

- How much of the variation in consumer assessments is determined by individual plans, and how much by the region, state, and MSA in which the beneficiary resides?
- How much do plan scores change from year to year? How much of the year-to-year variation in plan scores is due to changes affecting entire geographical areas?
- How reliable are estimates of change?
- How correlated are changes in different measures?

This research was supported by a grant from the Commonwealth Fund to Paul D. Cleary and by a contract (#500-95-007) with the Centers for Medicare and Medicaid Services (CMS).

Address correspondence to Alan M. Zaslavsky, Ph.D., Department of Health Care Policy, Harvard Medical School, 180 Longwood Ave., Boston, MA 02115-5899. Paul D. Cleary, Ph.D., and Lawrence B. Zaboriski, Ph.D., are also with the Department of Health Care Policy, Harvard Medical School.

METHODS

Survey Items

The CAHPS survey for Medicare managed care (Goldstein et al. 2001) was based on the CAHPS 2.0 instrument (Agency for Health Care Policy and Research 1999; Hargraves, Hays, and Cleary 2003), except in the first year, in which it was based on CAHPS 1.0; additional items are specific to Medicare (Schnaier et al. 1999). It asks respondents to give four global ratings (with a 0–10 response scale) of their plans, the care they received, their personal doctors, and their specialists. We also analyzed responses to 29 other questions (reporting items) that asked about specific aspects of health care, including access to and interactions with personal doctors, office staff, and specialists; availability of other services; and experiences with plan administrative functions. The survey also asked the respondent's age, educational level, and health status. The questions and information on item response rates and means appear elsewhere (Zaslavsky, Beaulieu et al. 2000; Zaslavsky and Cleary 2002). Two items about making a complaint to the plan and getting a complaint resolved satisfactorily were excluded from the analysis because they were difficult to interpret and had been found in previous analyses to have inconsistent relationships with other items (Zaslavsky, Beaulieu et al. 2000). We excluded 2001 data from the analysis of two items concerning prescription drugs because those items were changed in that year in ways that made them incomparable to the preceding years.

Survey Procedures

The CAHPS Medicare Managed Care (CAHPS-MMC) surveys were conducted in early 1998 (asking about 1997 experiences) and in September–December of 1998, 1999, 2000, and 2001 (asking about experiences in the same year). From each plan, or from several geographic strata within large plans, 600 members (or the entire enrollment, if fewer) were sampled. Sampled beneficiaries were mailed a questionnaire and, if necessary, a replacement questionnaire. They were then contacted by telephone if they had not yet responded and a telephone number could be obtained. Further details appear elsewhere (Goldstein et al. 2001; Zaslavsky et al. 2001; Zaslavsky, Zaborski, and Cleary 2002).

Analyses

For our analyses, we discarded cases whose mailing addresses fell outside the contract service area (CSA, the area in which the plan had agreed to accept enrollment under the given contract) for the corresponding contracts. For the

random effects models we also dropped the lowest-level groupings with very small samples, to simplify the calculations and avoid giving weight to areas or years in which a plan was barely represented. This criterion required a plan by state by MSA sample size of at least 50 cases and a plan by state by year sample size of at least 200 cases.

We use the term “plan” to refer to an MMC contract. A few health plans had multiple contracts in the same state that were consolidated starting with the 2000 survey. For consistency over the study period and to group contracts that were likely to be under a common administrative structure, we recoded the contract identifiers for all years to that of the consolidated contract. Remaining unconsolidated contracts were treated as distinct units because we were unable to determine when they had unified administrations, networks, and management policies. In all analyses, we treated contracts as nested within states. A few contracts (from 1 to 16 in each year) had substantial enrollment in more than one state; in those cases we treated the parts in each state as if they were distinct contracts.

We determined the state and MSA of residence of each respondent, assigning an entire county to an MSA if any part fell within that MSA. The non-MSA (rural) part of the state was treated as a single geographical unit equivalent to an MSA. Estimates omitting the rural respondents were very similar and are not reported.

Analyses of sources of variation were conducted using linear mixed models (Snijders and Bosker 1999). The three variables that were important case-mix adjustors in CAHPS-MMC (age, educational level, and self-reported health status) (Zaslavsky et al. 2001) were entered with coefficients fixed across plans, areas, and years, thus controlling for predictable effects of respondent characteristics (which also are related to survey nonresponse) on scores. Variance components for random effects were estimated by restricted maximum likelihood (REML) estimation using *SAS PROC MIXED* (SAS Institute Inc. 1999).

We could not conduct the full analyses of both spatial and temporal variation within a single model because our software could not estimate simultaneously all the random effects required. We therefore fit two models with different specifications of the random components. The first model included effects for region (10 CMS regions), state, and MSA within state, as well as for plan and the interaction of plan and MSA within state. This model can be expressed mathematically as

$$y_{rsmpk} = \beta^l x_{rsmpk} + \alpha_r + \gamma_{rs} + \delta_{rsm} \\ + \lambda_{rsp} + \kappa_{rsm} + \varepsilon_{rsmpk},$$

where y_{rsmpk} represents the response to an item from subject k enrolled in plan p and residing in MSA m , both within state s in region r , and $\beta'x_{rsmpk}$ is the (vector) product of case-mix coefficients by individual-level covariates. The remaining terms represent random effects, each with expectation 0 and variances σ^2_{α} , σ^2_{γ} , σ^2_{δ} , σ^2_{λ} , σ^2_{κ} , and σ^2_{ε} respectively. The variance components σ^2_{α} , σ^2_{γ} , σ^2_{δ} , σ^2_{λ} , σ^2_{κ} , represent the variation in responses attributable to the various levels of geography and the plan as identified by the indices of the random effects. The final component σ^2_{ε} quantifies the residual variation of individual responses after controlling for all of the other effects. The second mixed model included effects for state, MSA, and plan and interactions of each with the year of the survey, in a similar form. By including interactions with time, this model adjusts for changes in the distribution of enrollment across plans or areas over time.

For each model, we calculated the total explained variation, excluding the individual-level error variance σ^2_{ε} , and the percentage of this total attributable to each of the random effects in the model. We also calculated the percentage of variance explained by the model random effects, as a fraction of total variance including individual-level error σ^2_{ε} . For the second set of models, we summarized the variation over time of each main effect (state, plan, and MSA) using an intraclass correlation coefficient (ICC) defined as $\sigma^2_{\text{EFF}} / (\sigma^2_{\text{EFF}} + \sigma^2_{\text{EFF*YEAR}})$, where EFF represents any of the three main effects (a persistent effect) and EFF*YEAR the interaction of that effect with time (the corresponding varying effect). Thus, if ICC = 1 the corresponding effect (of plan, state, or MSA) is completely stable over time; conversely, if ICC = 0, the corresponding effect is independent each year. A combined ICC was defined by summing the three main effect variance components, summing the three time interaction components, and then applying the definition of ICC to the combined components. We also plotted time trends in selected variables, adjusted for changes in geographical composition and case mix, using main effects of time from the same model.

To summarize the results of these analyses for the 29 reporting items, they were assigned to groups, based on previous plan-level analyses (Zaslavsky, Beaulieu et al. 2000; Zaslavsky and Cleary 2002): access to care and the doctor's office, interactions with the doctor, access to plan-provided services and equipment and to prescription medications, interactions with the plan's customer service functions, vaccinations, and a single item on advice to quit smoking. Percentages of explained variance were averaged across the items in each group.

We evaluated the detectability of temporal changes by calculating t statistics for the change between consecutive years in each plan's ratings and

report composites (the mean of the plan's scores for each group of items). We subtracted the mean for all plans in the year from each year's scores, thus evaluating changes in *relative* ratings of plans. We calculated standard errors that take into account the partially overlapping sets of respondents for different items (Agency for Health Care Policy and Research 1999). This analysis considered each pair of consecutive years and used all plans that had at least 350 survey responses in both years of the pair.

We also investigated whether changes from year to year in the four rating items were correlated. For each pair of consecutive years, we calculated the covariance matrix $S_{y,y+1}$ of the *changes* between years of plan means for the four ratings. We also calculated the mean (across plans) covariance matrix of sampling error of plan means for the two years, V_y and V_{y+1} (Agency for Health Care Policy and Research 1999). We then corrected S for sampling error by subtracting estimated sampling error, $S_{y,y+1} - (V_y + V_{y+1})$, and converted this corrected estimate into a correlation matrix (Zaslavsky 2000).

RESULTS

Sample Description

The total sample size over five years was 705,848 and a total of 381 contracts were represented, with an average of 239 in each year (Table 1). The number of contracts increased over the first three years of the study and then declined, possibly due to the impact of the Balanced Budget Act of 1999 (Achman and Gold 2002). Response rates to the survey ranged from 75.3 percent in year one

Table 1: Response Rates, Number of Respondents, and Number of Contracts in the Medicare Managed Care CAHPS Survey over Five Years

<i>Year</i>	<i>Response Rate</i>	<i>Respondents in CSA</i>	<i>Number of Contracts</i>
1997	75.30	88,369	182
1998	80.98	135,152	255
1999	81.90	162,478	303
2000	82.73	174,118	257
2001	83.93	145,731	198

Notes:

1. "Number of contracts" is the number of distinct contracts with CMS, after combining contracts that were consolidated in 2000.
2. "Respondents in CSA" excludes respondents whose address was outside the contract service area of the surveyed contract.

to 83.9 percent in year five; analyses of response rates appear elsewhere (Zaslavsky, Zaborski, and Cleary 2002). About 2.0 percent of the cases were excluded because their addresses were outside the CSA in each of years one to three, and about 1.0 percent in years four and five (in which most such cases had been removed from the sampling frame before the survey). The criterion for minimum sample sizes excluded 20,053 cases from the variance component analyses (2.8 percent of the sample).

Geographical and Plan Effects

Table 2 shows estimates of variance components for area, plan, and the interaction of plan with MSA within state, as well as the percent of total measure variability that was explained by these effects. Percent explained variation was largest by far for the item “getting prescriptions through the plan” (explained

Table 2: Percent of Explained Variance Attributable to Geographical Factors and Plan

	<i>Plan</i>	<i>MSA</i>	<i>MSA*Plan</i>	<i>State</i>	<i>Region</i>	<i>Explained/Total</i>
Ratings						
Health plan	55	7	14	24	0	5.1
Health care	36	20	11	24	9	1.9
Doctor	31	18	19	29	3	1.7
Specialist	39	20	14	13	14	1.3
Access						
Happy with personal MD	35	16	23	26	0	3.0
Doctor knows important facts	27	18	18	27	9	0.4
MD understands health problems	39	24	5	16	16	0.6
Easy to get referral	41	8	20	9	21	2.3
Get advice by phone from doctor's office	30	21	15	22	13	2.4
Routine care as soon as wanted	43	20	13	11	13	2.9
Care for illness as soon as wanted	38	17	10	16	20	2.0
Get needed care	38	15	7	12	29	1.3
Delays in care waiting for approvals	43	11	11	20	16	3.1
Long wait past appointment time	26	20	18	8	28	4.6
Summary of Access Reports	36	17	14	17	16	2.3
Doctor						
Office staff courteous and respectful	47	21	7	7	17	1.2
Office staff helpful	33	27	12	12	15	1.2
Doctor listens carefully	30	22	13	28	8	1.0
Doctor explains things	28	22	17	24	9	0.7
Doctor shows respect	31	24	12	20	13	1.0
Doctor spends enough time with you	30	25	13	20	11	1.2
Summary of Doctor Reports	33	24	12	19	12	1.1

continued

Table 2. Continued

	Plan	MSA	MSA*Plan	State	Region	Explained/Total
Services						
Got special medical equipment	48	14	8	10	21	2.6
Problem getting therapy	52	15	3	15	16	3.5
Problem getting home healthcare	29	26	9	14	21	3.4
Plan provided all help needed	59	15	2	9	15	1.6
Summary of Services Reports	47	17	5	12	18	2.8
Prescriptions						
Problem getting prescription drugs	68	1	21	9	2	5.2
Get prescriptions through plan	48	11	6	36	0	28.0
Summary of Prescriptions Reports	58	6	13	22	1	16.6
Customer Service						
Problem getting information	69	3	14	13	0	3.3
Problem getting help on telephone	68	1	8	20	3	4.9
Customer service helpful	66	1	8	19	6	4.7
Problem with paperwork	55	3	9	25	8	6.0
Summary of Customer Service Reports	65	2	10	19	4	4.7
Vaccinations						
Flu shot last year	36	26	10	19	9	1.9
Ever had a pneumonia shot	49	26	12	9	4	3.3
Summary of Vaccinations Reports	43	26	11	14	6	2.6
Advised to quit smoking	17	20	10	25	28	1.3
Mean across all items	43	16	12	18	12	3.6

Notes:

1. Limited to Plan/MSA/State units with at least 50 observations and Plan/State/Year units with at least 200 observations.
2. Each entry (excluding the last column) represents the magnitude of a variance component as a percentage of total variance for all effects included in the table.
3. "Explained/Total" represents the percentage of the total variance of individual-level responses that is explained by all effects included in the table.
4. Percentages in some rows do not total to 100 percent due to roundoff error.
5. The 10 CMS regions were defined as follows: New England (Maine, Vermont, Massachusetts, Connecticut, Rhode Island, New Hampshire), New York / New Jersey (New York, New Jersey), Mid-Atlantic (Pennsylvania, Delaware, District of Columbia, Maryland, Virginia, West Virginia), South Atlantic (Alabama, Florida, Georgia, Kentucky, Mississippi, North Carolina, South Carolina, Tennessee), East Midwest (Illinois, Indiana, Michigan, Minnesota, Ohio, Wisconsin), Midwest (Iowa, Kansas, Missouri, Nebraska), Mountain (Colorado, Montana, North Dakota, South Dakota, Utah, Wyoming), Southwest (Arkansas, New Mexico, Oklahoma, Texas, Louisiana), Pacific (Arizona, California, Hawaii, Nevada), Northwest (Alaska, Idaho, Oregon, Washington).

variation = 28 percent), probably because responses to this item largely reflected a plan-determined benefit rather than individual experiences. The percentage of variance explained was also large (about 5 percent) for ratings of plan, as was the mean percentage for the customer service items. In these cases, most of the explained variation was attributable to the plan effect, which

accounted for more than half of the variation in each of the items in these groups.

Conversely, the least variation was explained by model effects for the care, doctor and specialist ratings, and for reports on interactions with doctors (mean = 1 percent), access to care (mean = 2 percent), and advice to quit smoking, indicating that variation in responses to these items largely reflected experiences of individual plan members or the practices of individual doctors. These items had the smallest percentage of variation attributable to the plan effect as well, with relatively strong geographical effects at the MSA, state, or regional level. The vaccinations and services groups were intermediate both in fraction of variance explained and the part of that variance attributable to the plan.

The distribution of explained variance across geographic levels varied across items, but summarizing across all items, about equal shares were explained by state and MSA (18 percent and 16 percent, respectively), with a lesser share explained by region. Regional effects were substantial only for reports on access, doctor communications, services, and advice to quit smoking and for ratings of specialists. Conversely, an average 43 percent of the explained variation was attributable to the plan. The interaction of plan by MSA represents the variation of scores for a given plan across parts of a state. This effect was always smaller than the main effect for plan (mean = 12 percent compared to 43 percent), indicating that each health plan had fairly consistent effects on quality across the areas it served within a state. However, the variance explained by the interaction exceeded half the plan component for ratings of doctors, suggesting that these might have been affected by variations across the local networks providing services for the plans.

When the county was substituted for the MSA as the smallest geographical unit, very similar results were obtained. In models including both county and MSA main effects and plan interactions, the MSA variance components were generally larger than the county components, indicating that assessments for counties within the same MSA tend to be similar (data not shown).

Geographical effects on the ratings and composites by MSA are mapped in an online appendix to this article available at www.blackwell-synergy.com. The maps reveal distinct regional patterns in assessments of health plans and care.

Variation over Time in Geographical and Plan Effects

Table 3 shows estimates of variance components for the interactions of year with plan, MSA, and state. Although the geographic effects were specified

Table 3: Percent of Explained Variance Attributable to Plan, MSA and State, and to Interactions of Each with Year

	<i>Plan</i>	<i>Plan*Year</i>	<i>MSA</i>	<i>MSA*Year</i>	<i>State</i>	<i>State*Year</i>
Ratings						
Health plan	49	10	10	5	24	1
Health care	40	3	20	4	33	0
Doctor	39	6	22	3	29	0
Specialist	47	4	23	0	26	1
Summaries for Report Item Groups						
Access	39	6	19	4	31	1
Doctor	38	4	24	3	31	1
Vaccinations	46	6	28	1	17	2
Customer service	57	14	4	2	22	1
Services	50	3	19	0	28	0
Prescriptions	53	8	15	2	17	5
MD advised to quit smoking	25	0	23	0	53	0
Mean across all items	44	6	18	2	28	1

1. Limited to Plan/MSA/State units with at least 50 observations and Plan/State/Year units with at least 200 observations.
2. Each entry represents the magnitude of a variance component as a percentage of total variance for all effects included in the table.
3. Intraclass correlation can be calculated from these results as in following illustration based on rating of health plan:
 For plan effect, $ICC = (\text{plan effect})/(\text{plan effect} + \text{plan} * \text{year interaction}) \approx 49/(49 + 10) \approx 83\%$.
 Combined $ICC = (\text{sum of main effects})/100\% \approx 49 + 10 + 24 \approx 83\%$.

slightly differently in this model, the variance shares for plan, MSA, and state (combining main effects and time interactions) were comparable to those in the first model (Table 2). (The omitted region component was absorbed into the state component, and the omitted plan by MSA interaction was absorbed into the plan and MSA components.) Conclusions regarding the relative influence of the plan on the various measures are also similar to those presented in Table 2.

Variance components for change over time (interactions with year) were much smaller than the corresponding main effects, indicating that all of the scores were stable from year to year. The overall intertemporal ICC, combining all variance components, ranged from .68 to 1.00 (mean = .90) for individual items and from .83 to .95 for groups of report items. The intertemporal ICC was larger for state than for plan for each rating and group of reports except for prescriptions, signifying that the state effects are more stable over time. In particular, for the access, doctor, and services groups, and care, doctor, and specialist ratings, the ICCs for state effects ranged from .97 to .99.

The contribution of MSA effects to variation over time was always smaller than that of plan effects.

The lowest ICCs were in the same domains for which the effect of the plan is strongest, overall rating of plan (.84), customer service functions (mean = .83), and prescription drugs items (mean = .85). Thus, these items had the greatest year-to-year variation, consistent with the hypothesis that they primarily measure functions that are most directly controlled by the plans and could be modified most readily from year to year. The low ICC for state effects for the prescription items might indicate that when prescription benefits changed, there was a tendency for them to change simultaneously for plans in the same state. The ICCs for the other rating items, and the mean ICCs for the other groups of reports, were greater than .90, indicating that these were relatively stable over time. Scores for services and advice to quit smoking were particularly stable.

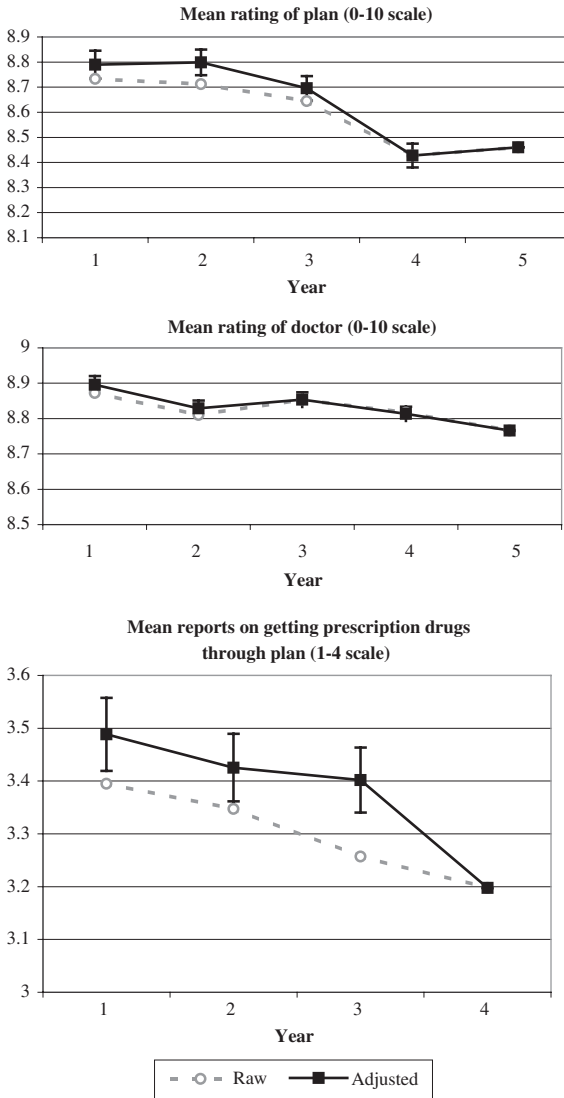
Another way to assess the stability of plan scores is to calculate how much the means typically differ between two plans (within the same state and MSA) in the same year, and for the same plan between years. For rating of plan and the customer service items, the typical (one standard deviation) difference between plans in the same year would be more than twice as large as the change for the same plan in consecutive years. For access and doctor reports and ratings, the typical difference between plans would be about three times the typical change between years. Thus, scores tend to be stable compared to the cross-sectional variation among plans.

To illustrate secular trends, we plotted unadjusted and adjusted means for rating of plan, rating of doctor, and getting prescription drugs through the plan (Figure 1). Adjusted and unadjusted patterns were very similar. Ratings of plan dropped sharply from 1999 to 2001. (Customer service reports followed a very similar pattern.) Ratings of doctor also trended downward over the five years (similar to reports on interactions with doctors). For these variables, the overall decline approximated one standard deviation of the plan effect in the mixed model. Reports on getting prescription drugs dropped sharply between 1999 and 2000. All but two measures declined from the first to last year in which they appeared in the survey.

Statistical Significance of Changes between Years

Table 4 presents changes between consecutive years in the mean ratings and report composites, relative to mean trends. For rating of plan, there were a substantial number of significant changes; 7.6 percent, 22.6 percent, 36.4

Figure 1: Trends in Selected CAHPS Measures, 1997–2001



Adjusted rates are from the mixed models including time effects and are adjusted for plan and area effects and individual case-mix effects. The vertical scale for adjusted rates is shifted to match the unadjusted rates in the last year of each series. Error bars are for the difference of each year’s adjusted mean from the final year’s mean. The final year for “getting prescriptions” is 2000 (Year 4) because 2001 data were incomparable.

Table 4: Statistical Significance of Changes in Relative Ratings of Plans (Plan Mean Score Minus Mean of All Plans) between Pairs of Years

	1997 to 1998			1998 to 1999			1999 to 2000			2000 to 2001		
	No Significant Change		Increase	No Significant Change		Increase	No Significant Change		Increase	No Significant Change		Increase
	Decrease	Change		Decrease	Change		Decrease	Change		Decrease	Change	
Plan	5	145	7	18	144	24	46	136	32	37	99	31
Doctor	4	146	7	10	168	8	12	191	11	10	148	9
Care	2	152	3	9	167	10	6	200	8	9	151	7
Specialist	0	155	2	6	170	10	6	201	7	2	160	5
Access	27	113	17	14	160	12	8	192	14	7	148	12
Doctor	2	150	5	7	175	4	7	198	9	8	150	9
Service	20	123	14	5	179	2	7	198	9	4	159	4
Prescriptions	32	101	24	27	136	23	52	131	31	17	135	15
Customer	37	93	27	12	161	13	23	168	23	15	136	16
Vaccination				14	157	15	11	188	15	3	161	
Advised to quit smoking	4	150	3	3	178	5	6	200	8			3
Total		157			186			214			167	

Note: Changes are classified as significant decrease ($t < -2$), no significant change ($-2 \leq t \leq +2$), or significant increase ($t > +2$).

percent, and 40.7 percent of plans, in the four pairs of years, had significant changes. For ratings of doctor, care, and specialist, on the other hand, there were fewer significant changes (on average 7.4 percent); only slightly more than the 5 percent that would be expected to appear due to random sampling variation. The prescription drugs composite had the largest number of significant changes in each year. These average rates, however, conceal considerable diversity in reliability, because a number of plans with large enrollments and extended service areas were allocated larger than average samples and therefore were measured more reliably. Thus, for some plans relatively small changes would be significant. Reliability is also affected by the varying number of respondents to the different groups of items; for example, over 90 percent of survey respondents provided a rating of plan but only about half rated their care (Zaslavsky and Cleary 2002).

Correlations among Changes

After restricting the analyses to plans that had at least 350 respondents in pairs of years, analyses were based on 157, 186, 214, and 167 plans for pairs of consecutive years. Due to the smaller sample sizes and number of plans in 1997, estimates for the 1997–1998 correlations were insufficiently reliable to be reported, as were those for the specialist ratings (answered by only 51 percent of unit respondents). Correlations for the remaining rating variables in the other three pairs of years (after correction for sampling error) appear in Table 5. Correlations between changes in care ratings and those for either doctor or plan were higher (.64 to .94) than those between changes in doctor and plan ratings (.31 to .51). This result is consistent with previous cross-sectional findings that the plan and doctor ratings characterize distinct dimensions of consumer experiences, while the care rating combines the two (Zaslavsky, Beaulieu et al. 2000).

Table 5: Estimated Correlations of Change for Three Ratings Items, between Consecutive Years

	<i>Rating of Plan by Rating of Doctor</i>	<i>Rating of Plan by Rating of Care</i>	<i>Rating of Doctor by Rating of Care</i>
1998 to 1999	0.51	0.70	0.84
1999 to 2000	0.39	0.64	0.77
2000 to 2001	0.31	0.65	0.94

DISCUSSION

Understanding which aspects of quality vary most by plan, market, and state and over time can help to identify the most appropriate focus of quality interventions. As in our previous study (Zaslavsky, Landon et al. 2000), ratings of plan varied more than other ratings, and most of the variation in ratings of plan and reports on customer service (a strong predictor of plan ratings [Zaslavsky, Beaulieu et al. 2000; Zaslavsky and Cleary 2002]) and obtaining prescription drugs was attributable to the individual plan. The rating of the plan is likely to be driven largely by plan-specific administrative structures and rules (Wagner et al. 2001), while access to prescription drugs is strongly influenced by the plan's benefits design (Zaslavsky, Beaulieu et al. 2000). Conversely, most variation in ratings and reports about health care providers was attributable to geographic factors. These measures are affected by the style of medical practice associated with networks of providers in various areas (which often overlap across plans), and therefore vary mainly by geographic area. The substantial quality variation across MSAs, even within the same plan (manifested as a plan-MSA interaction) might reflect the influence of the local networks of providers who serve a plan's customers in each of the areas in which it operates. In a previous study (Solomon et al. 2002), most of the variation in consumer assessments of primary care within a geographical area was attributable to the medical practice sites, with relatively little effect of health plans.

For most measures the variation among states was larger than the variation among MSAs within states, quantitatively supporting the appropriateness of the state as a unit for evaluation of health care quality. Furthermore, some measures (particularly those related to access and obtaining special services) exhibited substantial regional variation, indicating that contiguous states tended to be similar with respect to these aspects of quality. The state is a natural unit, particularly for Medicare, because (1) it is the unit within which most contracts are defined, (2) quality control and improvement functions of the Medicare Quality Improvement Organizations (QIOs, formerly known as Peer Review Organizations or PROs) operate at the state level, and (3) it is the main unit of geography for comparative reporting of consumer assessments back to consumers and plans. However, there was also substantial substate variation, for the system as a whole (MSA effect) and for particular plans (interaction of plan by MSA). Thus, it might be worthwhile to report assessments for substate parts of large plans, where sample sizes permit.

Most measures trended downward between 1997 and 2001, in many cases substantially, even after removing the effects of plan departures from

the Medicare market. The sharp decline in reports on getting prescription drugs in 2000 might reflect the many plans dropping or restricting those benefits in that year. The decline in ratings could represent disillusionment with managed care among Medicare beneficiaries, and perhaps a reaction to increasing restrictions imposed by plans facing financial pressures that did not allow them to continue to extend the generous benefit designs and care management policies that initially attracted members to managed care. However, because we do not have a comparable series of data from beneficiaries in fee-for-service Medicare, we cannot establish whether these trends were particular to managed care or conversely occurred in traditional Medicare as well.

Differences among plans in most ratings and report items were fairly stable. There was more change from year to year in ratings of plan overall and in reports on customer services and access to prescription drugs, the domains that are most sensitive to the specific policies of the plan. These scores might change more because it is easier for a plan to change its benefits design and administrative structures and policies than to change its network or to modify the practices of health care providers, especially when each plan contributes only a fraction of each provider's caseload.

Differences among states were extremely stable from year to year for all measures except access to prescription drugs. The somewhat larger changes in state effects for prescription drug items may reflect the incentives for plans in a state to change their prescription benefits simultaneously: a plan that retains a more generous benefit than its competitors is likely to be subject to adverse selection.

Plan scores on rating of plan, prescription drugs, and customer service commonly changed significantly between years, but other scores less often changed significantly. The changes between years tend to be smaller than differences among plans within a single year and therefore are more difficult to measure reliably. Future analyses should determine whether the reliability of estimates of change could be improved by combining many items into a single change score, or by using a panel design that samples the same respondents in consecutive years to improve the precision of estimates of change, as in the Current Population Survey. Also, effects of quality improvement efforts might be best measured using items that are specially designed to detect changes in the targeted aspects of quality.

Understanding the levels at which variation in quality occurs, particularly over time, helps to identify the levels at which quality interventions might be most effective. The stability of scores in domains related to direct medical

care suggests that most plans have not yet implemented quality improvement programs that dramatically affect these scores, although perhaps some plans have. The large geographical component of quality is particularly stable over time. Thus, quality differences across areas might be resistant to improvement without system-wide initiatives that are broader than what any single plan is able to implement, while conversely the plan might be the wrong locus for quality improvement except for areas like customer service over which it has the most control. This finding is consistent with the literature documenting large geographical variations in medical practice style and procedure utilization, particularly for the elderly population, with substantial cost implications but no visible clinical rationale or benefit for outcomes (Wennberg and Gittelsohn 1973; Gatsonis et al. 1995; Fisher et al. 2003a, b). Future research should address relationships of consumer assessments to these variations in practice and other measurable characteristics of the local health care system such as physician supply and reimbursement levels.

A previous study found significant differences in Medicare CAHPS scores associated with plan characteristics (Landon, Zaslavsky, and Cleary 2001); the impact of organizational structures spanning several contracts within the same state or in different states bears further study. Further research also could examine the experiences reported by beneficiaries who disenrolled from plans (excluded from the main CAHPS managed care survey) and the impact on variation of the declining number of plans remaining in the Medicare managed care market (Achman and Gold 2002). The Medicare managed care plans that we studied had a single payer, operated under identical regulations, enrolled members from the same population, and received reimbursements under a common geographically based structure. We might expect greater variation in the more segmented commercial market, although it is not obvious how this would affect the relative variation within and between geographical areas.

Health plans, purchasers, and consumers will be interested in trends for plans and areas. The results we have presented here should help in the interpretation of longitudinal CAHPS data.

ACKNOWLEDGMENTS

We thank Matthew Cioffi for expert preparation of datasets, Marc Ciriello for assistance with document preparation, and Elizabeth Goldstein and Amy Heller of CMS and the other members of the CAHPS-MMC project team for

their efforts and expertise in conducting the surveys on which these analyses are based.

REFERENCES

- Achman, L., and M. Gold. 2002. *Medicare+Choice 1999–2001: An Analysis of Managed Care Plan Withdrawals and Trends in Benefits and Premiums*. New York: Commonwealth Fund.
- Agency for Health Care Policy and Research. 1999. *CAHPS™ 2.0 Survey and Reporting Kit, Publication No. AHCPR99-0039*. Rockville, MD: Agency for Health Care Policy and Research.
- Fisher, E. S., D. E. Wennberg, T. A. Stukel, D. J. Gottlieb, F. L. Lucas, and E. L. Pinder. 2003a. “The Implications of Regional Variations in Medicare Spending. Part 1: The Content, Quality and Accessibility of Care.” *Annals of Internal Medicine* 138 (4): 273–87.
- . 2003b. “The Implications of Regional Variations in Medicare Spending. Part 2: Health Outcomes and Satisfaction with Care.” *Annals of Internal Medicine* 138 (4): 288–98.
- Gatsonis, C. A., A. M. Epstein, J. P. Newhouse, S. L. Normand, and B. J. McNeil. 1995. “Variations in the Utilization of Coronary Angiography for Elderly Patients with an Acute Myocardial Infarction. An Analysis Using Hierarchical Logistic Regression.” *Medical Care* 33: 625–42.
- Goldstein, E., P. D. Cleary, K. M. Langwell, A. M. Zaslavsky, and A. Heller. 2001. “Medicare Managed Care CAHPS: A Tool for Performance Improvement.” *Health Care Financing Review* 22 (3): 101–7.
- Hargraves, J. L., R. D. Hays, and P. D. Cleary. 2003. “Psychometric Properties of the Consumer Assessment of Health Plans Study (CAHPS) 2.0 Adult Core Survey.” *Health Services Research* 38 (6): 1509–27.
- Landon, B. E., A. M. Zaslavsky, and P. D. Cleary. 2001. “Health Plan Characteristics and Consumer Assessments of Health Plan Quality.” *Health Affairs* 20 (4): 274–86.
- SAS Institute, Inc. 1999. *SAS/STAT User’s Guide, Version 8*. Cary, NC: SAS Institute, Inc.
- Schnaier, J. A., S. F. Sweeny, V. S. L. Williams, B. Kosiak, J. S. Lubalin, R. D. Hays, and L. D. Harris-Kojetin. 1999. “Special Issues Addressed in the CAHPS™ Survey of Medicare Managed Care Beneficiaries.” *Medical Care* 37 (3, supplement): MS69–78.
- Snijders, T., and R. Bosker. 1999. *Multilevel Analysis*. Thousand Oaks, CA: Sage.
- Solomon, L. S., A. M. Zaslavsky, et al. 2002. “Variation in Patient-Reported Quality among Health Care Organizations.” *Health Care Financing Review* 23 (4): 85–100.
- Wagner, E. H., R. E. Glasgow, C. Davis, A. E. Bonomi, L. Provost, D. McCulloch, P. Carver, and C. Sixta. 2001. “Quality Improvement in Chronic Illness Care: A Collaborative Approach.” *Journal of Quality Improvement* 27 (2): 63–80.
- Wennberg, J. E., and A. Gittelsohn. 1973. “Small Area Variations in Health Care Delivery.” *Science* 182 (117): 1102–8.

- Zaslavsky, A. M. 2000. "Using Hierarchical Models to Attribute Sources of Variation in Consumer Ratings of Health Plans." *Proceedings, Sections on Epidemiology and Health Policy Statistics, American Statistical Association* 9–14.
- Zaslavsky, A. M., N. D. Beaulieu, B. E. Landon, and P. D. Cleary. 2000. "Dimensions of Consumer-Assessed Quality of Medicare Managed Care Health Plans." *Medical Care* 38 (2): 162–74.
- Zaslavsky, A. M., and P. D. Cleary. 2002. "Dimensions of Plan Performance for Sick and Healthy Members on the Consumer Assessments of Health Plans Study 2.0 Survey." *Medical Care* 40 (10): 951–64.
- Zaslavsky, A. M., B. E. Landon, N. D. Beaulieu, and P. D. Cleary. 2000. "How Consumer Assessments of Managed Care Vary within and among Markets." *Inquiry* 37 (2): 146–61.
- Zaslavsky, A. M., L. Zaborski, and P. D. Cleary. 2002. "Factors Affecting Response Rates to the Consumer Assessment of Health Plans Study (CAHPS[®]) Survey." *Medical Care* 40 (6): 485–99.
- Zaslavsky, A. M., L. B. Zaborski, Ding, J. A. Shaul, M. J. Cioffi, and P. D. Cleary. 2001. "Adjusting Performance Measures to Ensure Equitable Plan Comparisons." *Health Care Financing Review* 22 (3): 109–26.

CAHPS® Managed Care Maps

An on-line supplement to

Plan, Geographical, and Temporal Variation of Consumer Assessments of Ambulatory Health Care

Alan M. Zaslavsky

Lawrence B. Zaborski

Paul D. Cleary

Harvard Medical School, Cambridge, Massachusetts

Contact: zaslavsk@hcp.med.harvard.edu

The maps in this on-line supplement were prepared using the same data and composite (item grouping) definitions as the main article. Thus, for most measures (excluding only the prescription drugs composite), five years of data are combined. For each item or group of items, we calculated means for all respondents in each Metropolitan Statistical Area (MSA), weighted to be representative of the Medicare managed care enrollment (MMC) of that MSA. An MSA was defined to include all counties falling at least partly within the MSA boundaries. Non-MSA counties that had managed care enrollees at addresses in their plan's service area are treated as a single unit within each state, equivalent to an MSA. Of the 274 distinct areas mapped, 237 represent MSAs and 37 represent non-MSA areas.

We used the CAHPS® analysis macro, version 3.4b, to calculate means and standard errors, taking into account the unequally weighted sample design and the correlations among items in multi-item composites. Unequal weights arise primarily because of the unequal enrollments of managed care plans and the differing distributions of service areas and enrollment for each plan within a state. The analysis macro also casemix-adjusted the MSA measures, using a single national model, for age, education, self-reported health status, and year of response. Of these variables, the first three have typically been used in casemix adjustment of MMC data and the last is included to minimize confounding of geographic variation with general trends in quality scores.

The estimated means were grouped into quartiles for plotting, and mapped using a spectrum from red (lowest mean scores) to purple to blue to green (highest mean scores). The non-MSA areas are plotted using the same color sequence but are distinguished by being cross-hatched.

Eleven national maps were generated, representing four ratings, six composites and one single report item. Eleven additional maps present the same data for the northeastern quarter of the country alone, for better legibility in this area, where there are many geographically compact adjoining MSAs that are hard to distinguish in the national map. For the convenience of the viewer, the maps are presented as separate single-page files for viewing, as a single 22-page color PDF file for printing, and as a single 22-page grayscale PDF file for printing on black-and-white printers.

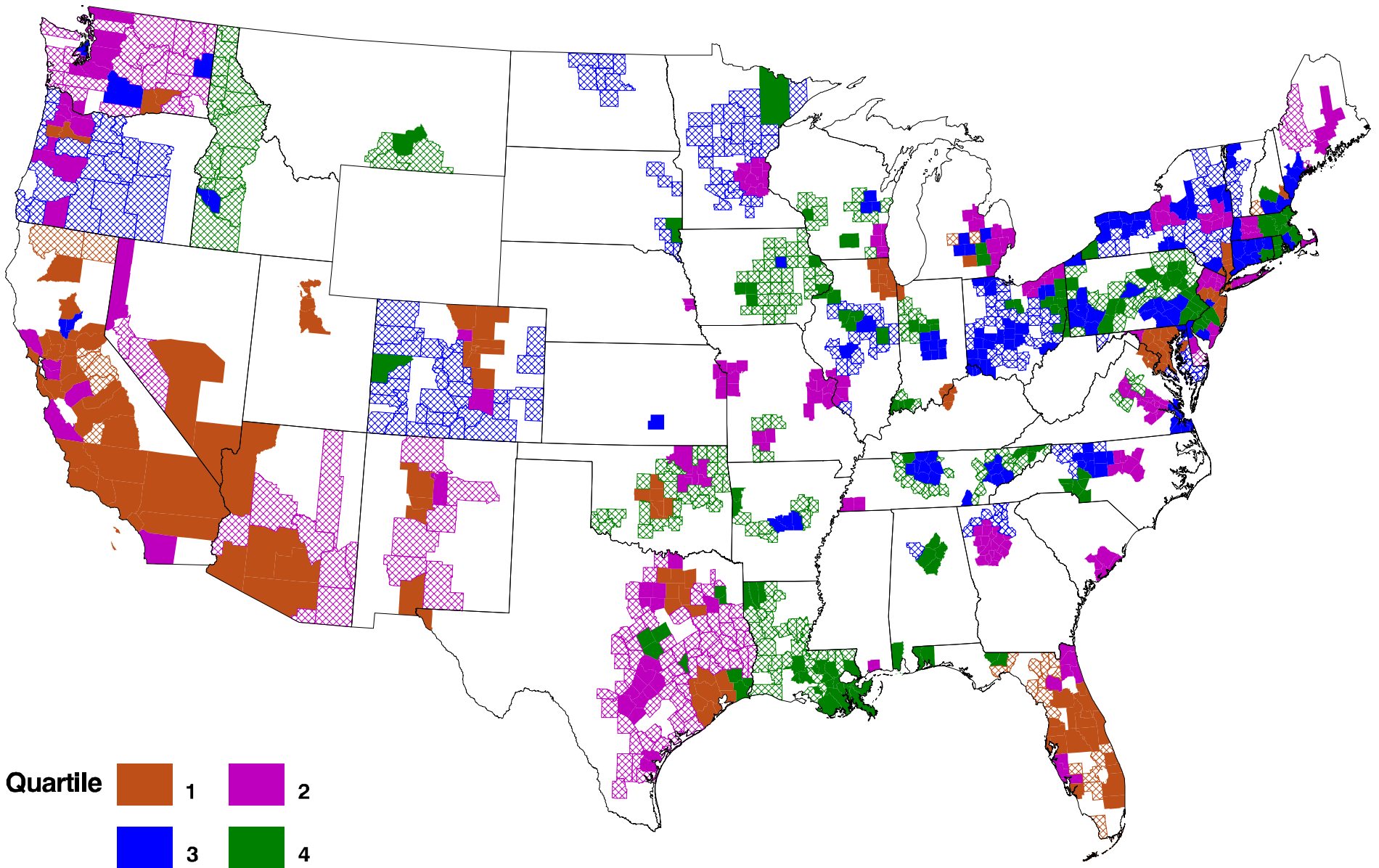
Several limitations should be kept in mind when reading these maps. First, the maps do not display variation within MSAs. Similarly, they do not display variation across the non-MSA area of a state. Our use of cross-hatching is meant to alert the reader that the overall quality quartile of such an area, which can sometimes be extensive, does not necessarily represent the quality of every part of the area. Second, the quartile classifications are only approximate, due to sampling error in the data, although the measures are highly reliable. Most errors of classification would occur because an area that was close to the quartile break was estimated (due to sampling error) to be on the other side of the break. Thus these errors, when they occurred, are only minimally misinforming. (See Gelman A, Price PN (1999), All maps of parameter estimates are misleading, *Statistics in Medicine*, 18:3221-3234, for a discussion of the difficulties of adequately representing both patterns and precision in maps.)

The table summarizes the precision of the estimates and classifications for the various measures. The F -statistic tests the hypothesis that all MSA means are equal, and that null is rejected strongly for every measure. The average intergroup reliability is calculated as $1/F-1$ and indicates how well the measures distinguish among areas. For almost all measures (except perhaps the “advice to quit smoking” item) this reliability is excellent. The remaining columns summarize the confidence with which the comparative statements in the maps can be interpreted. For each measure and MSA, we the macro calculates a t statistic to test the hypothesis that the area’s score equals the national average; values $t < -2$ or $t > +2$ are significant at the .05 level. Thus for most measures the majority of areas could be confidently classified as above or below average. To calculate the final column, we assumed that each area’s population mean had a posterior distribution centered at the sample estimate and with standard deviation equal to the standard error of that estimate, and calculated the mean probability that a population measure was actually in a different quartile than the sample estimate. As explained earlier, this largely represents the number of plans whose scores were close enough to the boundary between quartiles to be misclassified, but most of these errors are correspondingly minor in that the score was in fact close to the boundary.





It is not our objective here to interpret these results. We merely note that the patterns are not consistent across measures. While measures of clinician-patient interactions and health care quality show broadly similar patterns, benefits- and system-driven measures such as vaccinations and prescription drug availability show quite distinct patterns. We hope that these maps will stimulate generation of hypotheses for further research.

Measure	F statistic	Intergroup reliability (percent)	Percent with $ t > 2$	Estimated percent classified in adjacent quartile
Rating of plan	49.6	98.0	73.0	13.9
Rating of doctor	27.4	96.3	68.6	19.1
Rating of health care	28.7	96.5	67.5	21.0
Rating of specialist	16.1	93.8	50.4	29.3
Access composite	44.4	97.7	74.5	16.9
Doctor composite	25.6	96.1	66.1	22.8
Services composite	19.3	94.8	49.6	27.5
Prescriptions composite	6.9	85.5	38.6	30.3
Customer service composite	21.5	95.3	62.8	19.6
Vaccination composite	47.7	97.9	64.6	20.1
Advice to quit smoking	4.5	77.9	30.3	36.1

Health Care Rating. Medicare Managed Care CAHPS, 1997 – 2001



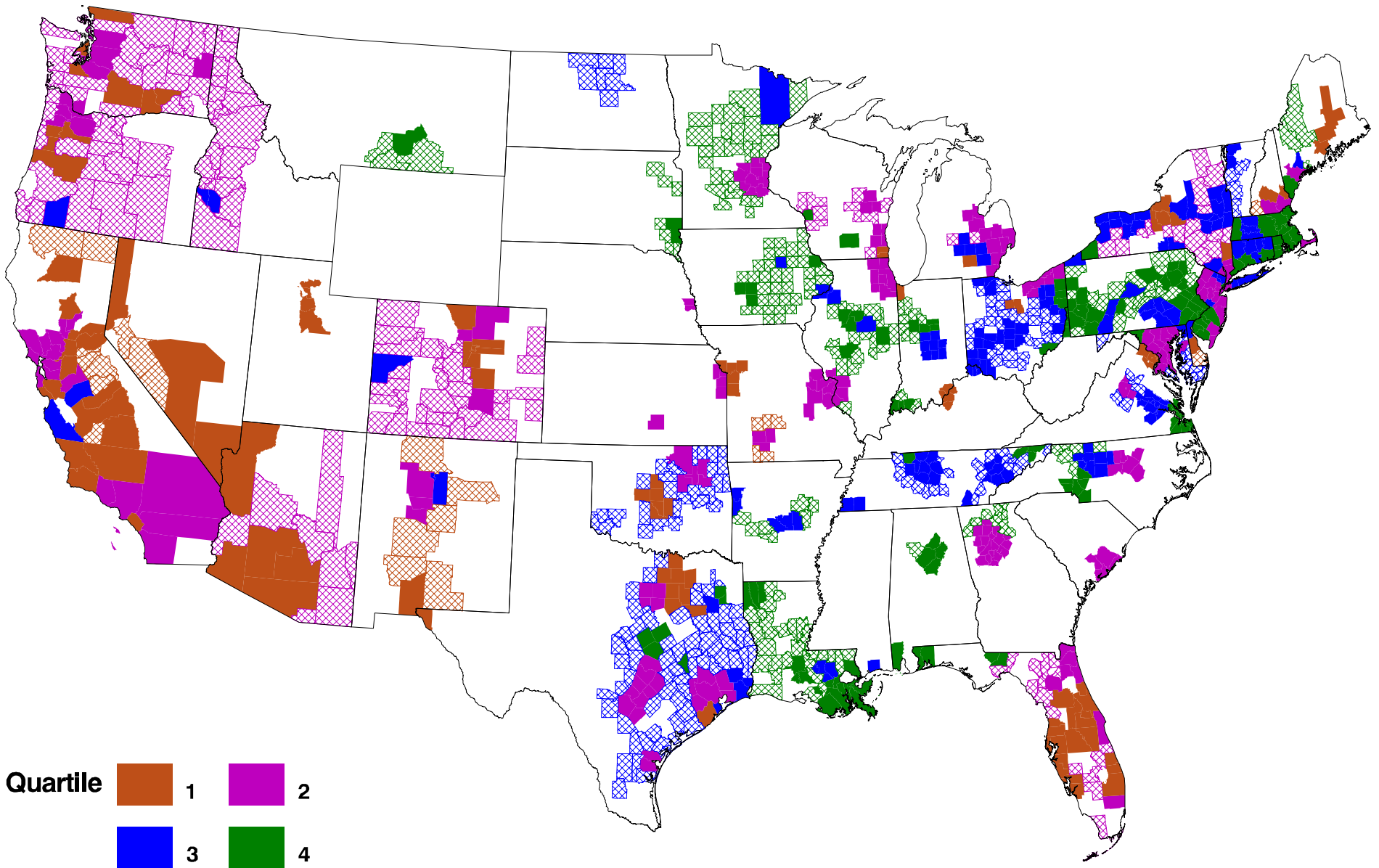
Quartile

	1		2
	3		4

1 = Lowest, 4 = Highest Score

Shaded areas are non-MSA counties

Doctor Rating. Medicare Managed Care CAHPS, 1997 – 2001

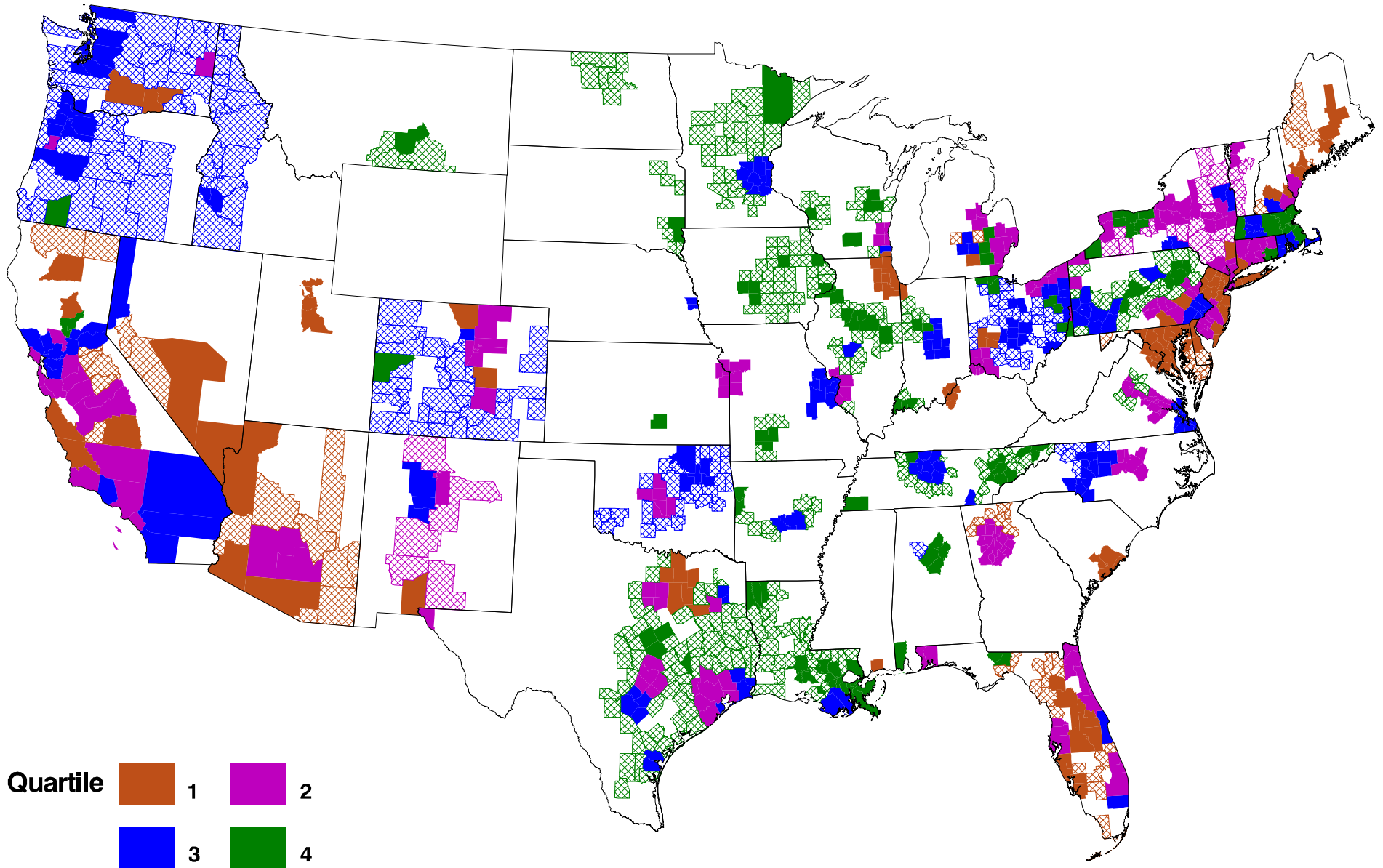


Quartile 1 2
3 4





1 = Lowest, 4 = Highest Score

Shaded areas are non-MSA counties

Plan Rating. Medicare Managed Care CAHPS, 1997 – 2001



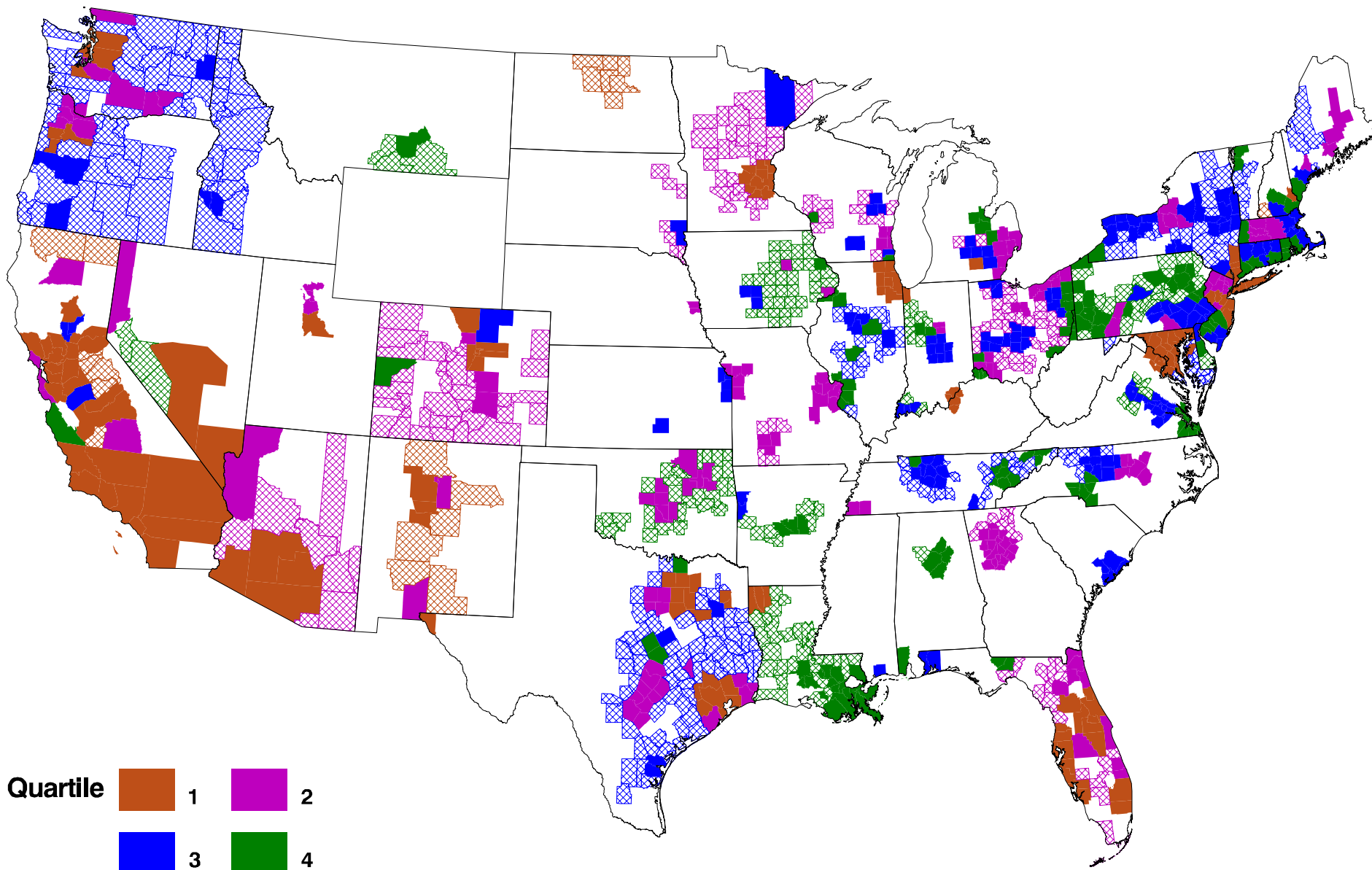
Quartile

	1		2
	3		4





1 = Lowest, 4 = Highest Score

Shaded areas are non-MSA counties

Specialist Rating. Medicare Managed Care CAHPS, 1997 – 2001

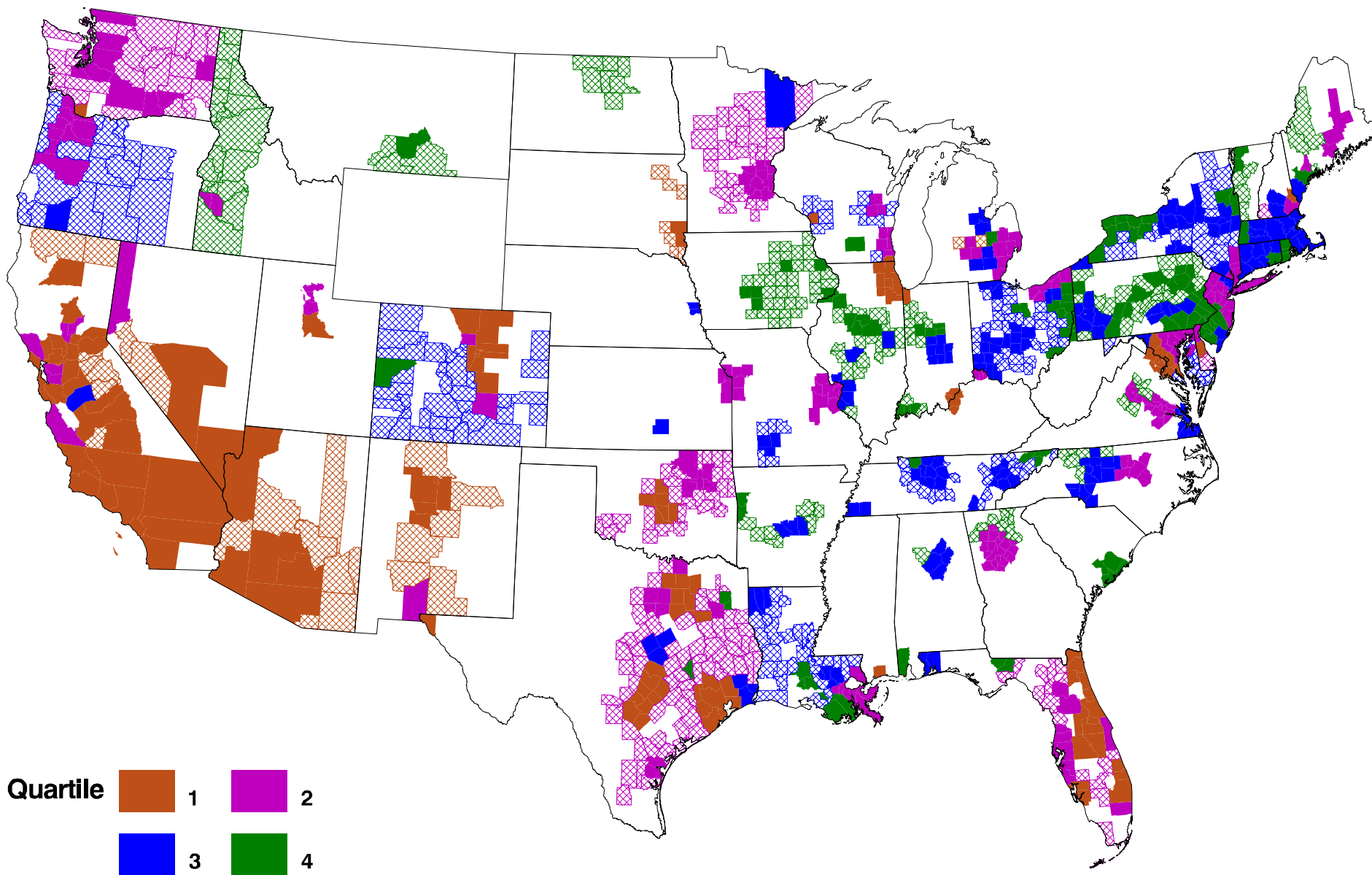


Quartile





	1		2
	3		4

1 = Lowest, 4 = Highest Score
Shaded areas are non-MSA counties

Access Composite. Medicare Managed Care CAHPS, 1997 – 2001



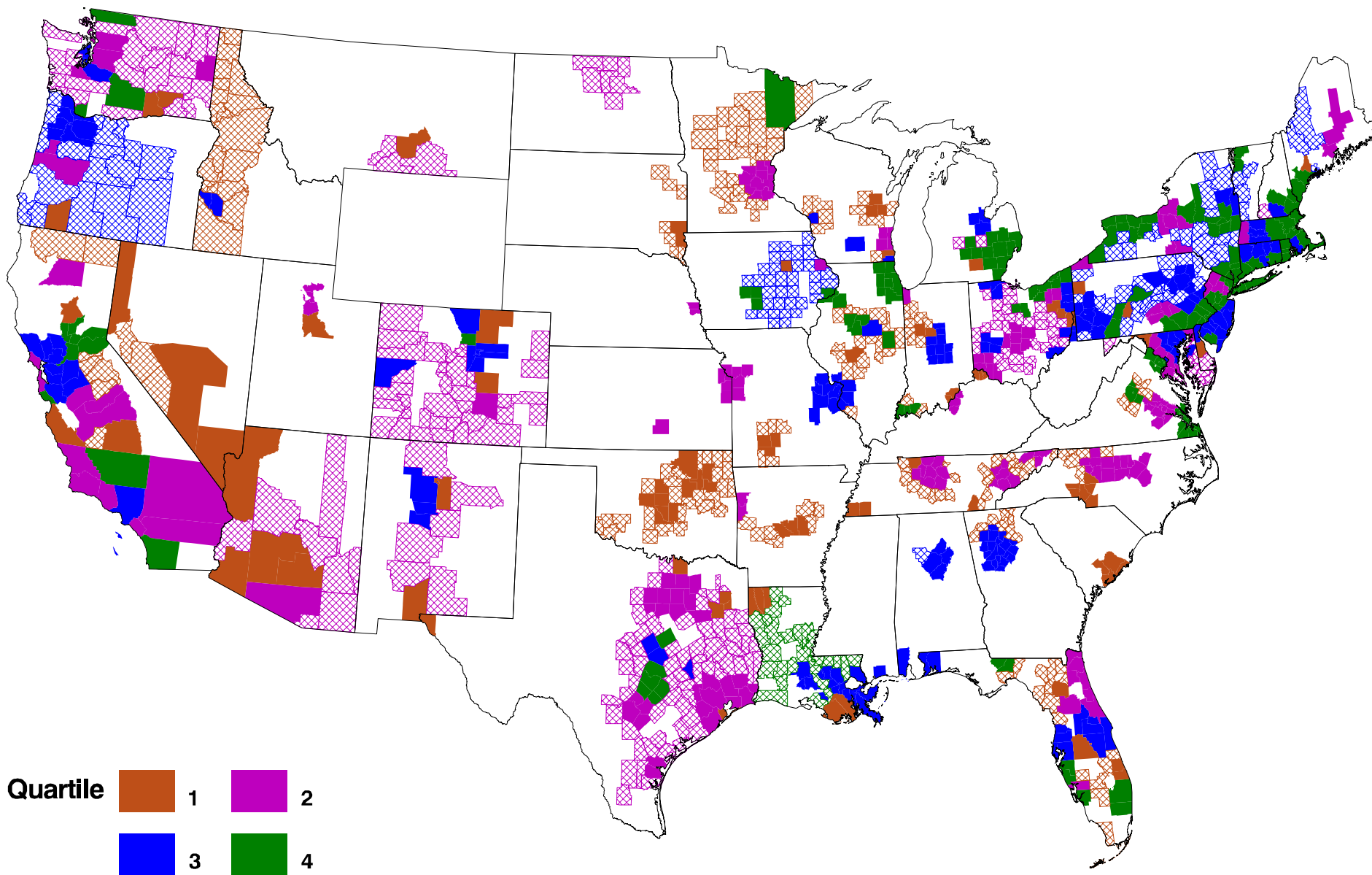
Quartile

	1		2
	3		4

1 = Lowest, 4 = Highest Score

Shaded areas are non-MSA counties

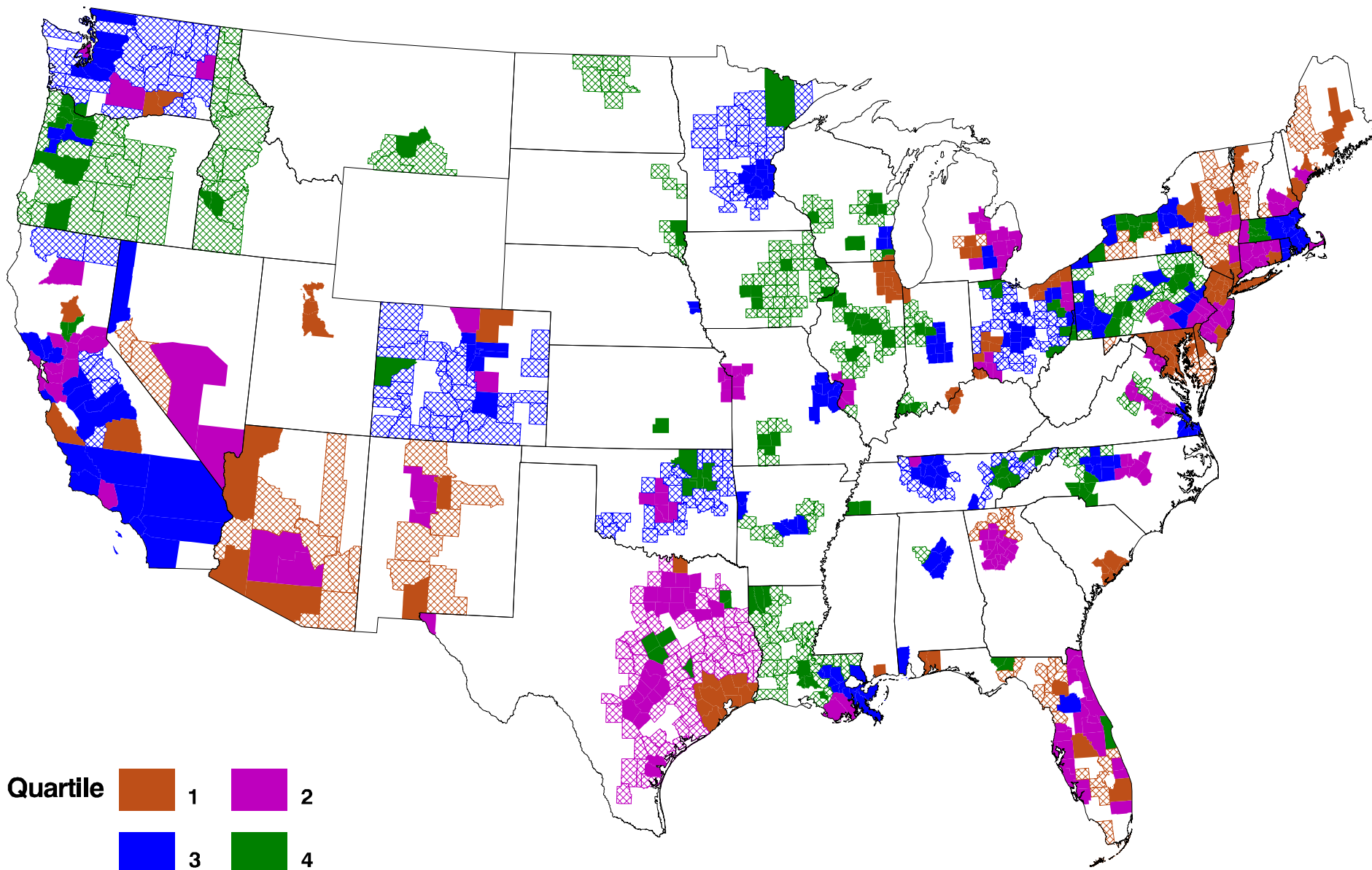
Advise to Quit Smoking. Medicare Managed Care CAHPS, 1998 – 2001







1 = Lowest, 4 = Highest Score

Shaded areas are non-MSA counties

Customer Service Composite. Medicare Managed Care CAHPS, 1997 – 2001

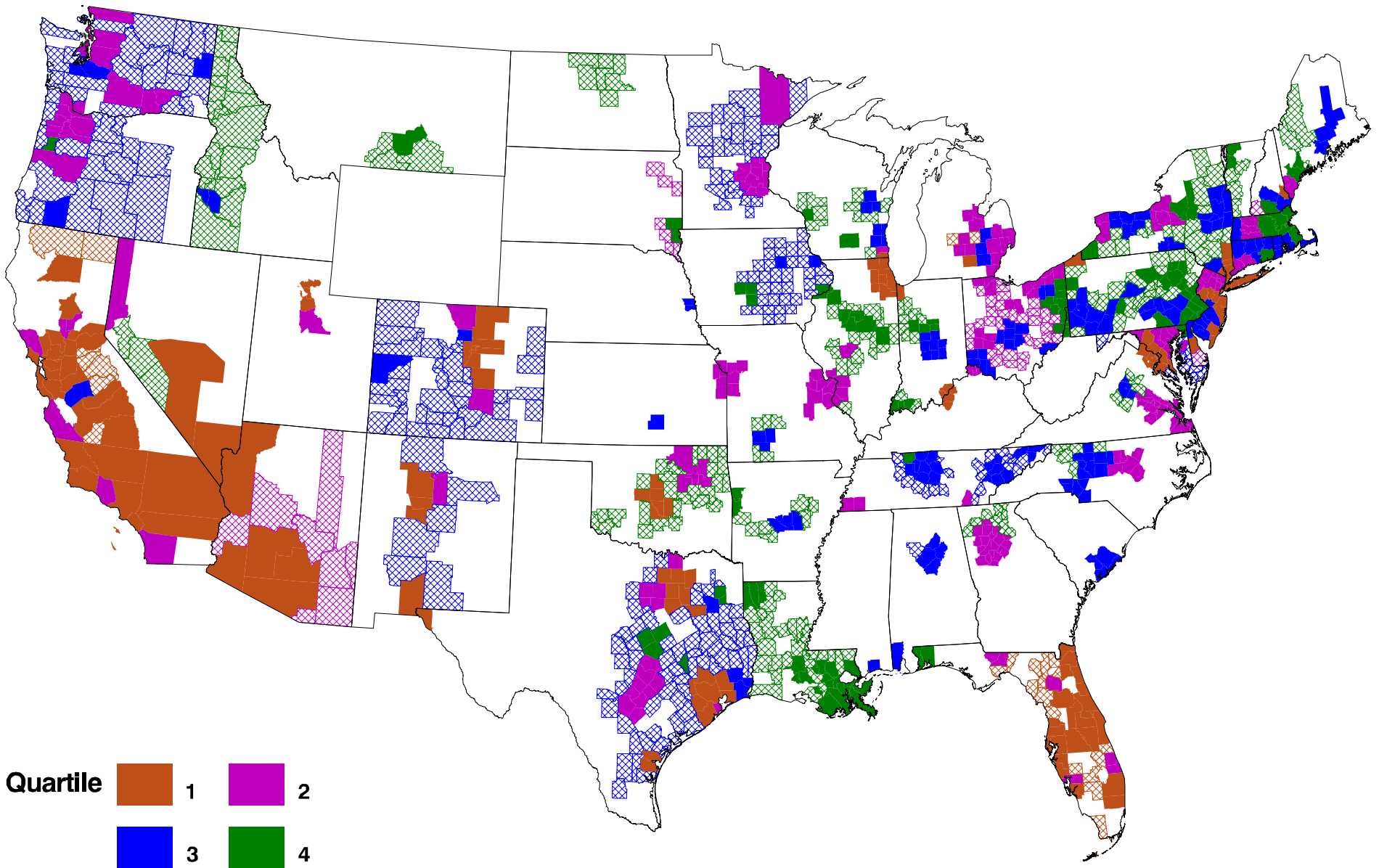


Quartile

	1		2
	3		4

1 = Lowest, 4 = Highest Score
Shaded areas are non-MSA counties

Doctor Composite. Medicare Managed Care CAHPS, 1997 – 2001

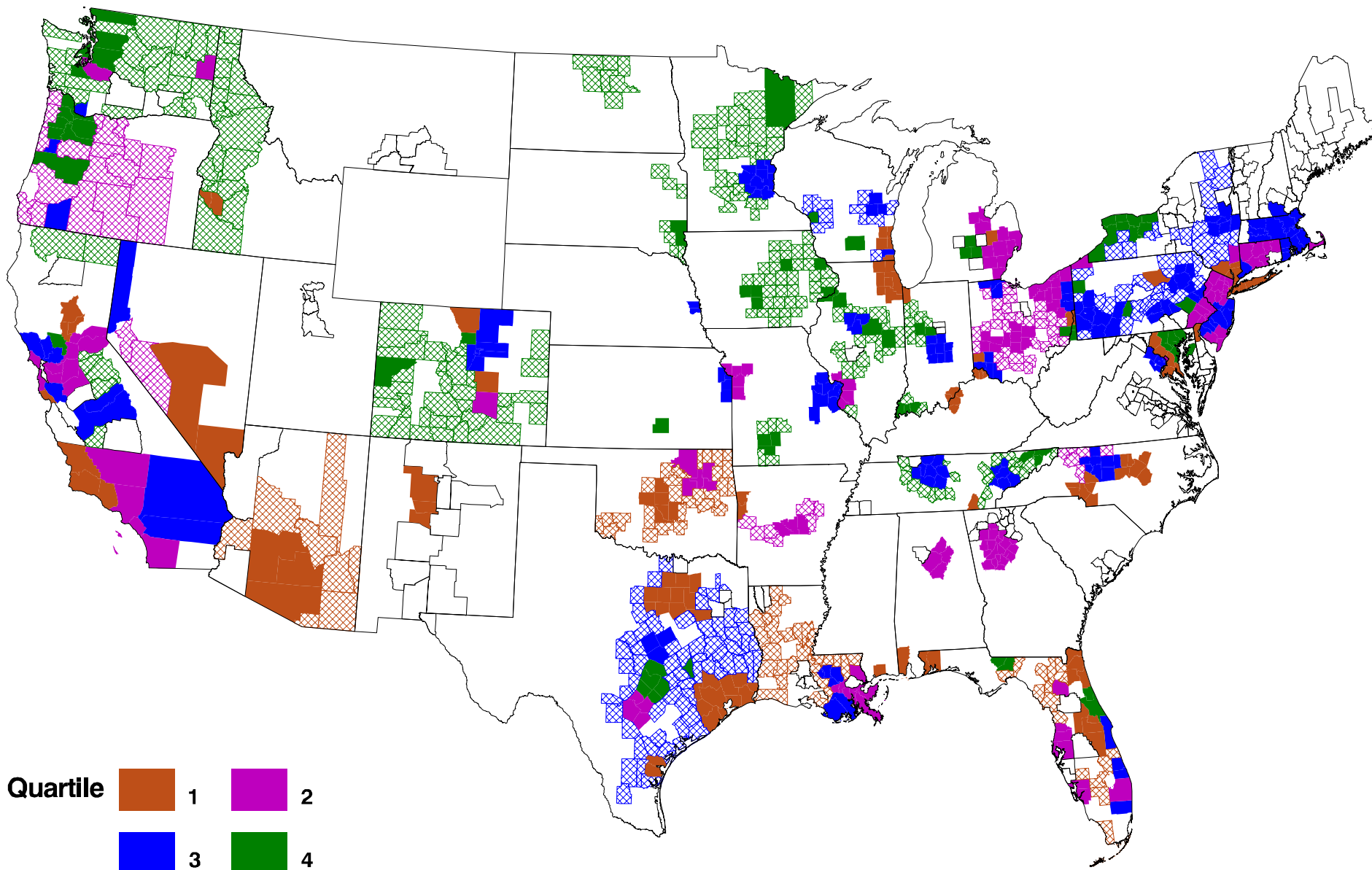


Quartile 1 2
3 4





1 = Lowest, 4 = Highest Score

Shaded areas are non-MSA counties

Prescription Composite. Medicare Managed Care CAHPS, 1998 – 2000

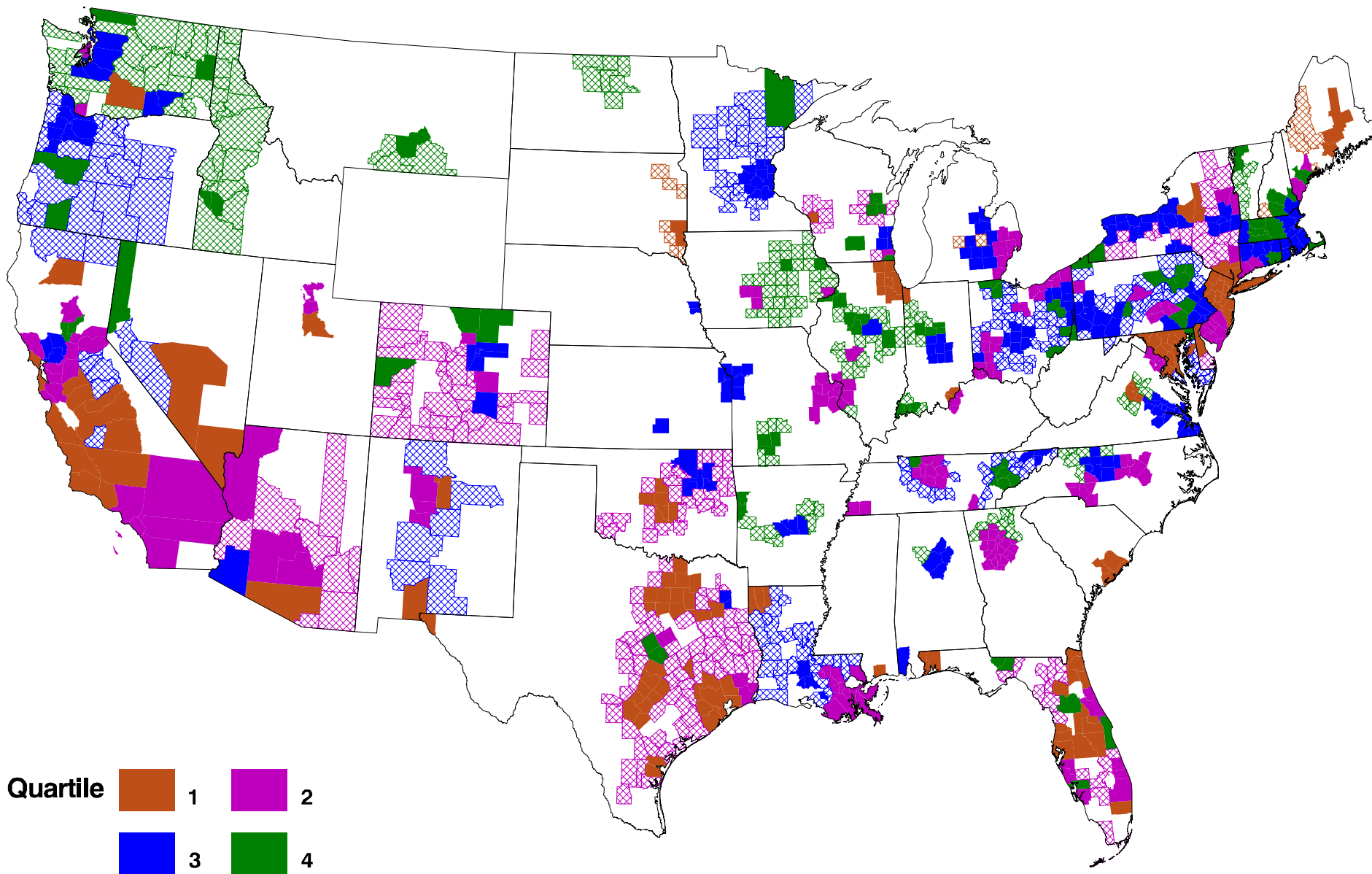


Quartile





	1		2
	3		4

1 = Lowest, 4 = Highest Score
Shaded areas are non-MSA counties

Service Composite. Medicare Managed Care CAHPS, 1997 – 2001



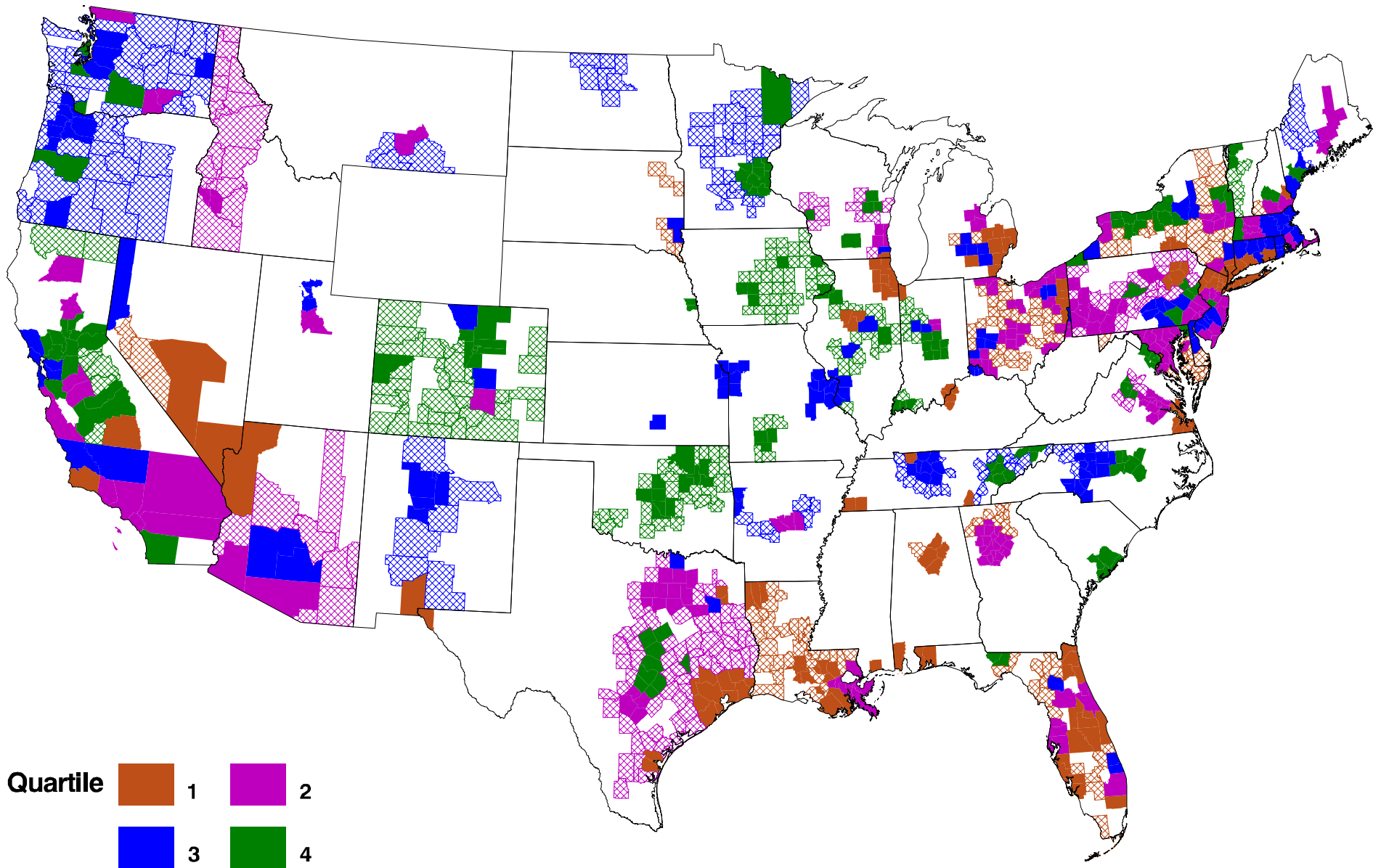
Quartile

	1		2
	3		4

1 = Lowest, 4 = Highest Score

Shaded areas are non-MSA counties

Vaccination Composite. Medicare Managed Care CAHPS, 1997 – 2001



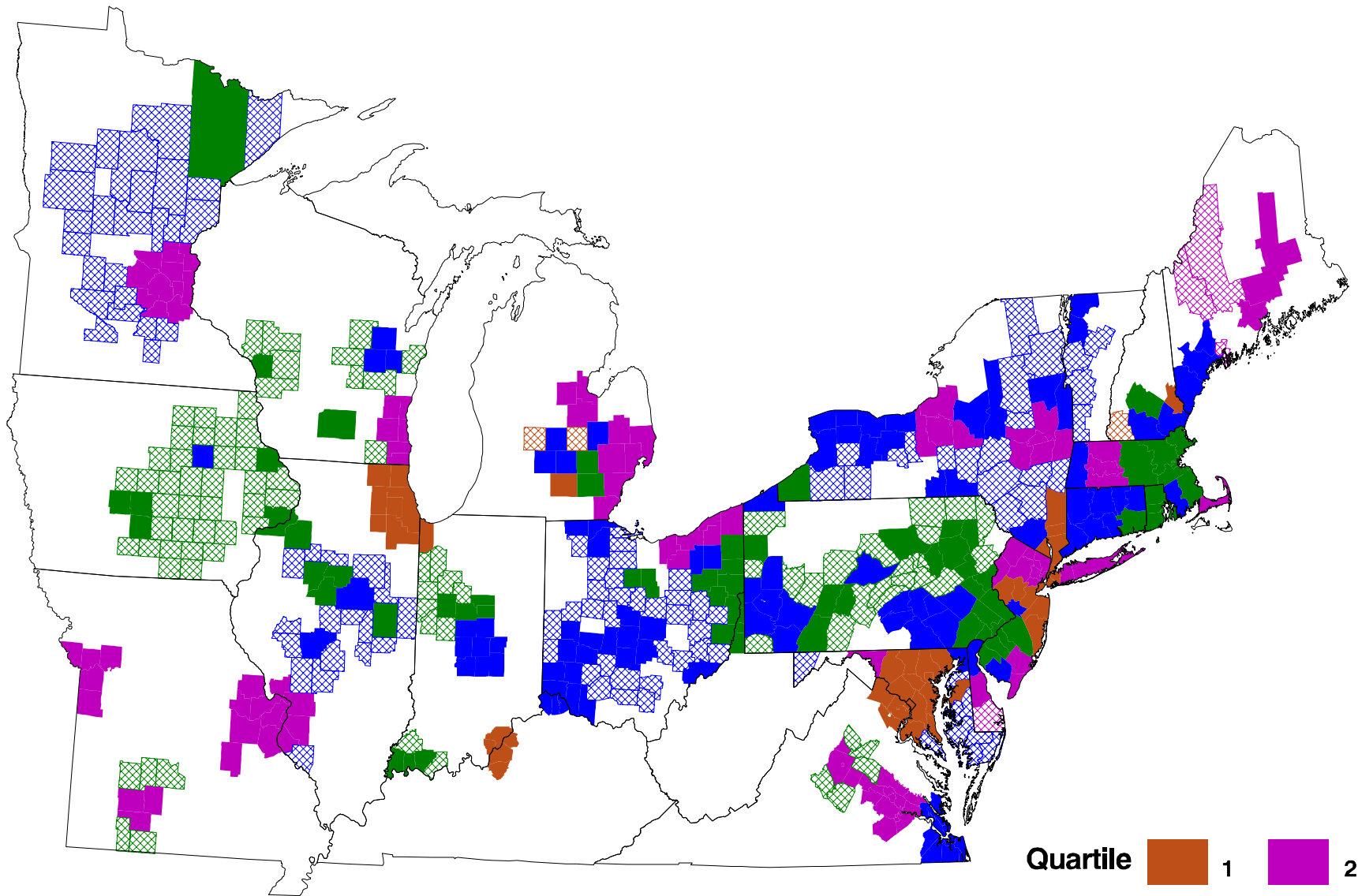
Quartile 1 2
3 4

1 = Lowest, 4 = Highest Score

Shaded areas are non-MSA counties

Health Care Rating. Medicare Managed Care CAHPS, 1997 – 2001

Northeastern States only

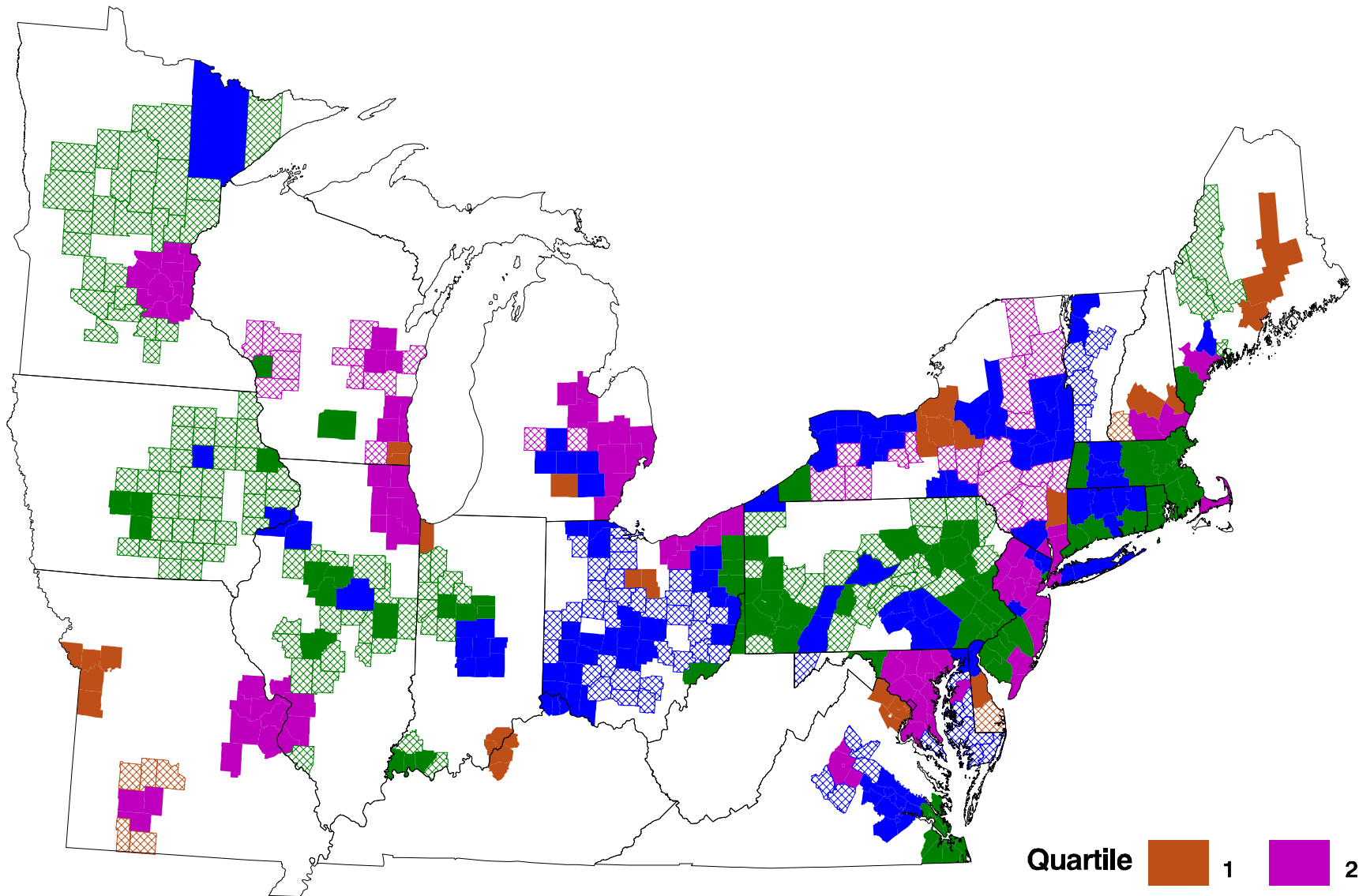


Shaded areas are non – MSA counties

Quartile 1 2
3 4
1 = Lowest, 4 = Highest Score

Doctor Rating. Medicare Managed Care CAHPS, 1997 – 2001

Northeastern States only

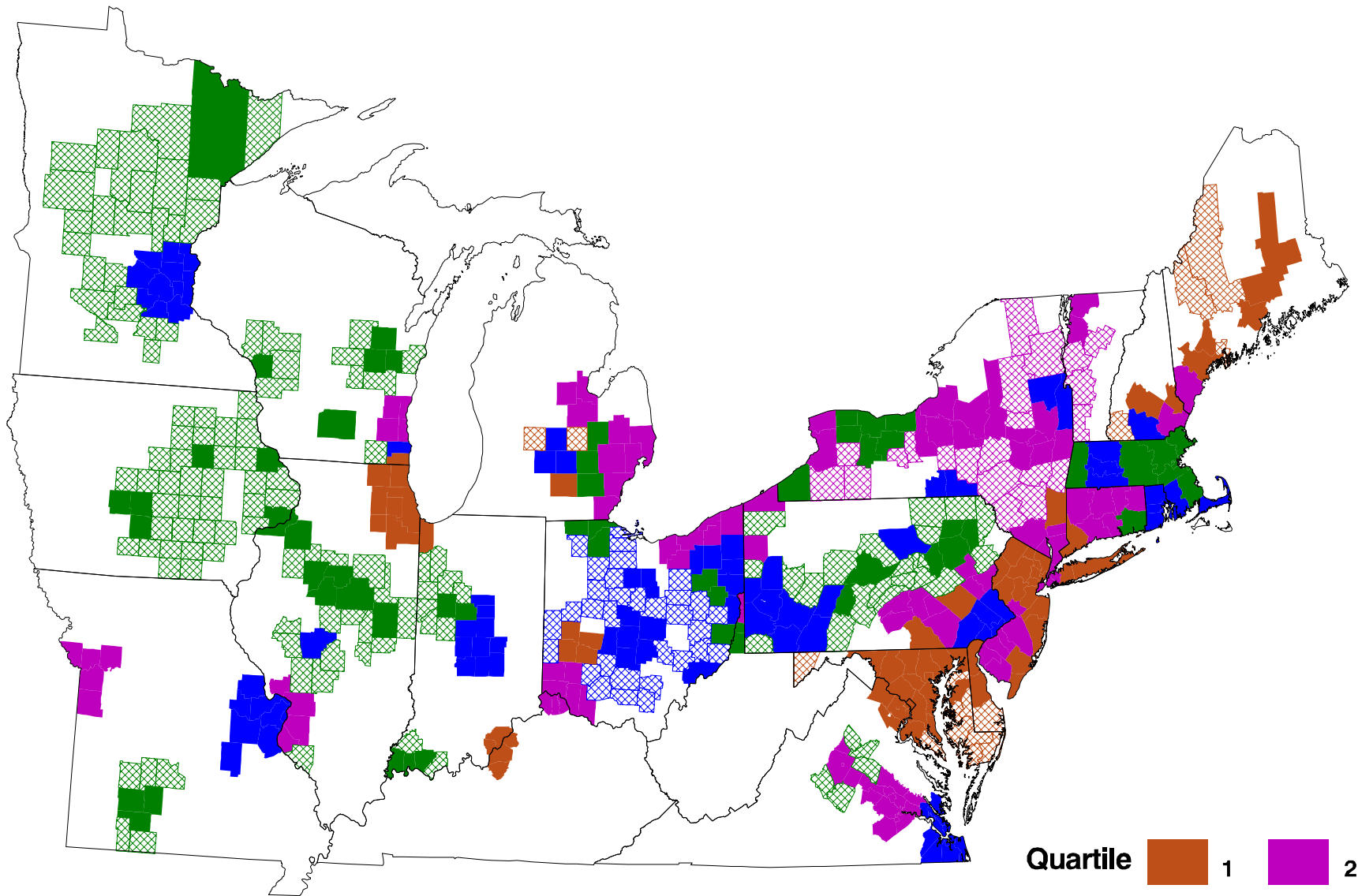


Shaded areas are non – MSA counties

1 = Lowest, 4 = Highest Score

Plan Rating. Medicare Managed Care CAHPS, 1997 – 2001

Northeastern States only

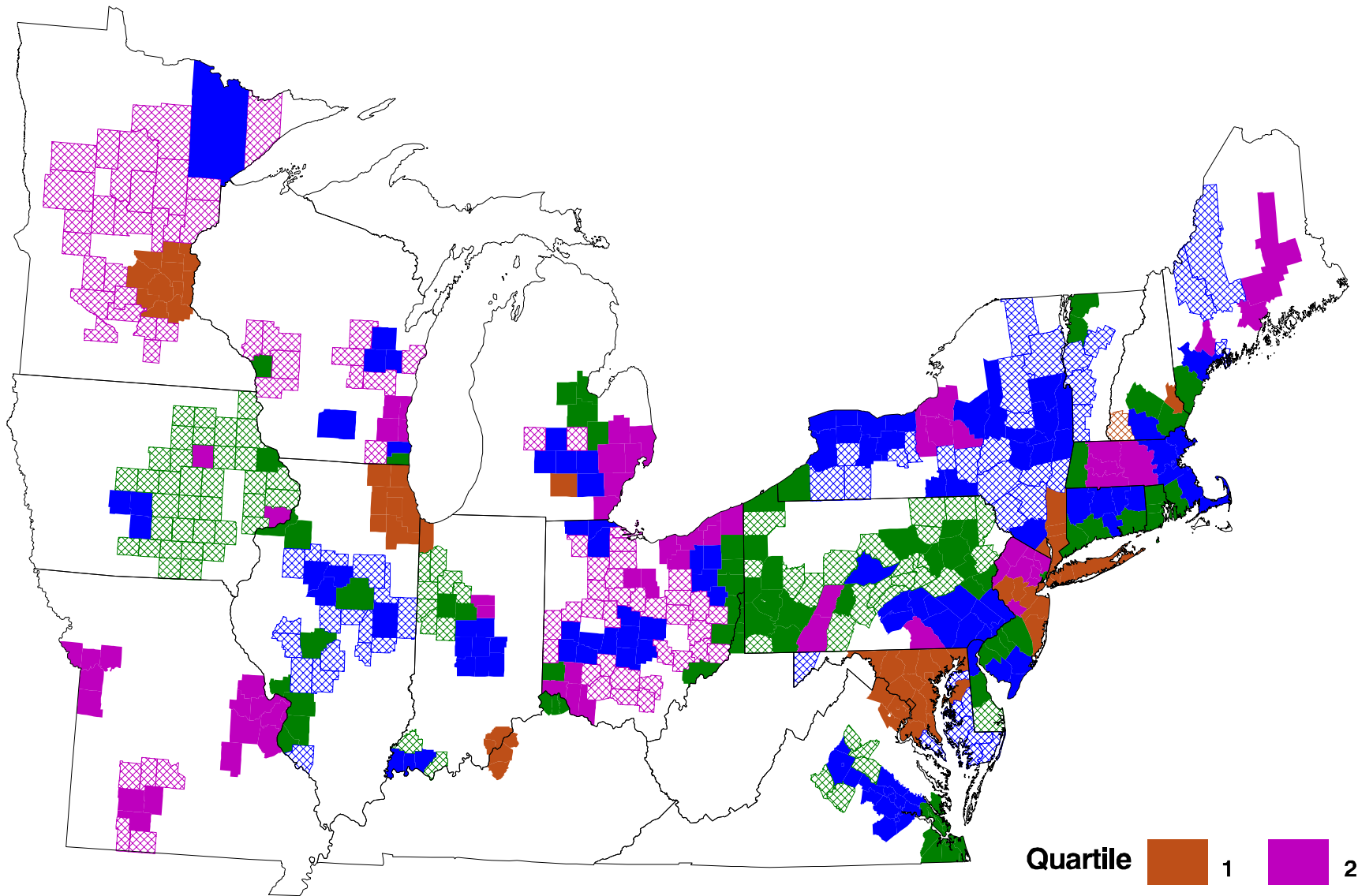


Shaded areas are non – MSA counties

1 = Lowest, 4 = Highest Score

Specialist Rating. Medicare Managed Care CAHPS, 1997 – 2001

Northeastern States only

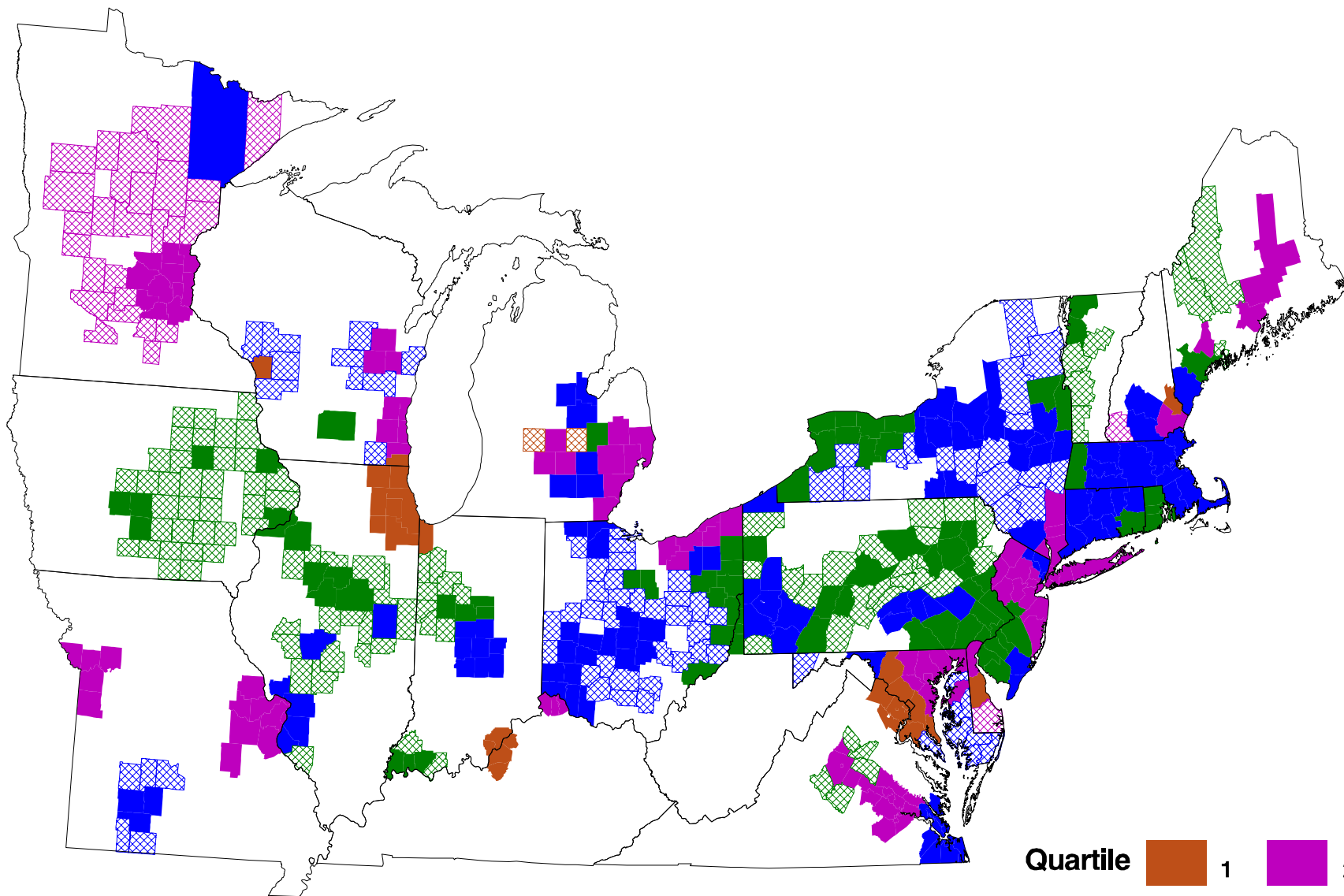


Shaded areas are non – MSA counties

Quartile 1 2
3 4
1 = Lowest, 4 = Highest Score

Access Composite. Medicare Managed Care CAHPS, 1997 – 2001

Northeastern States only



Shaded areas are non – MSA counties

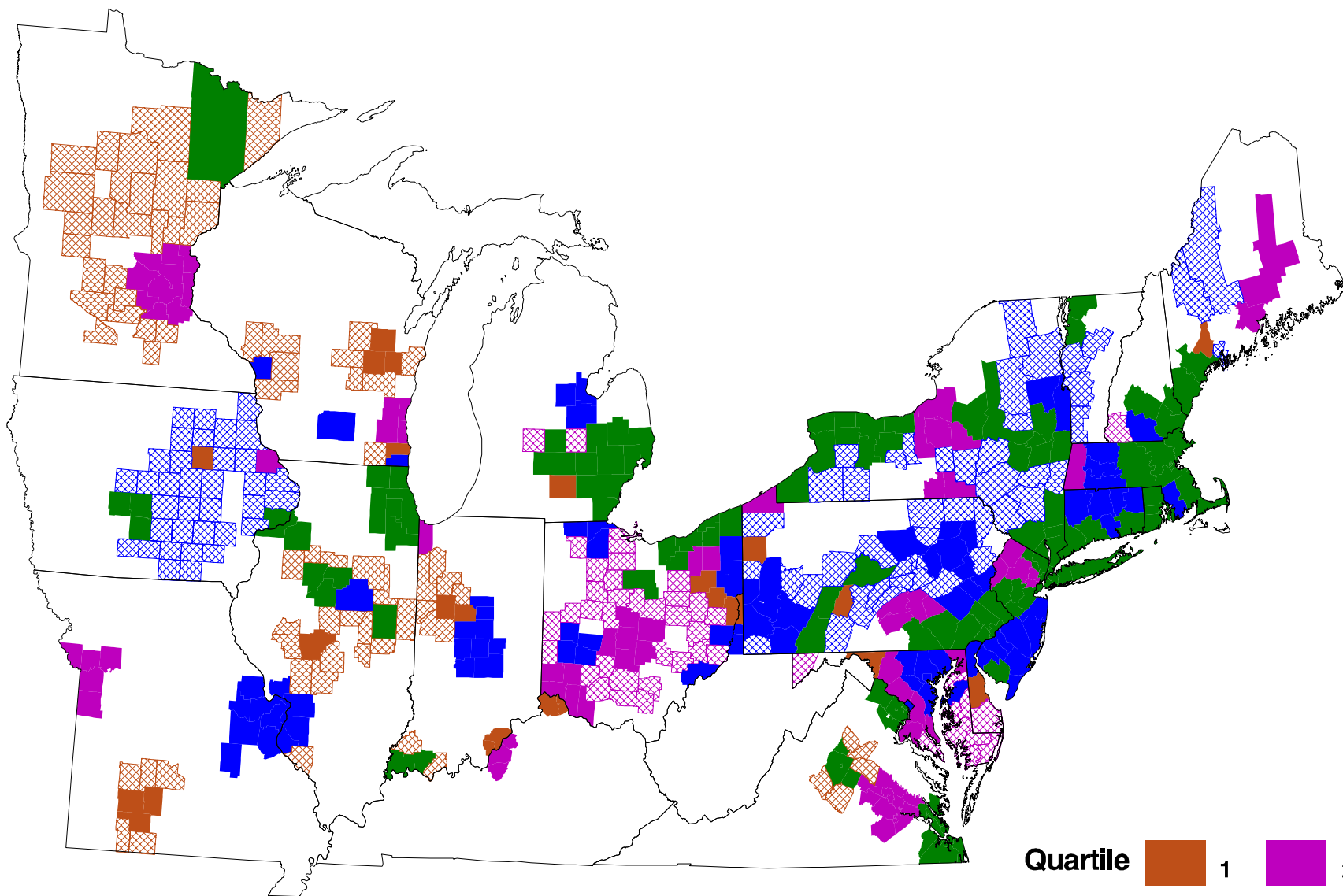
Quartile

	1		2
	3		4

1 = Lowest, 4 = Highest Score

Advise to Quit Smoking. Medicare Managed Care CAHPS, 1998 – 2001

Northeastern States only

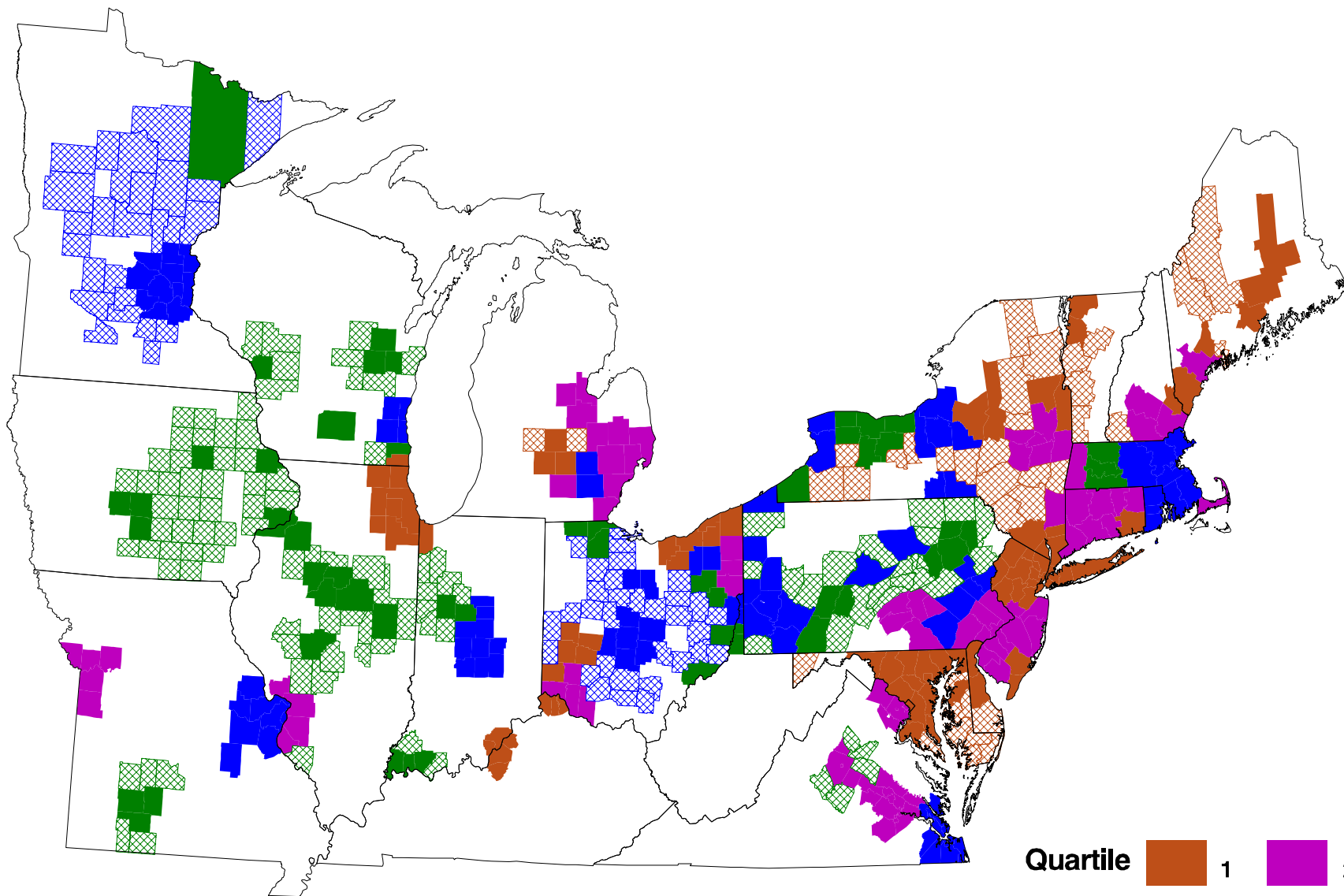


Shaded areas are non – MSA counties

Quartile 1 2
3 4
1 = Lowest, 4 = Highest Score

Customer Service Composite. Medicare Managed Care CAHPS, 1997 – 2001

Northeastern States only

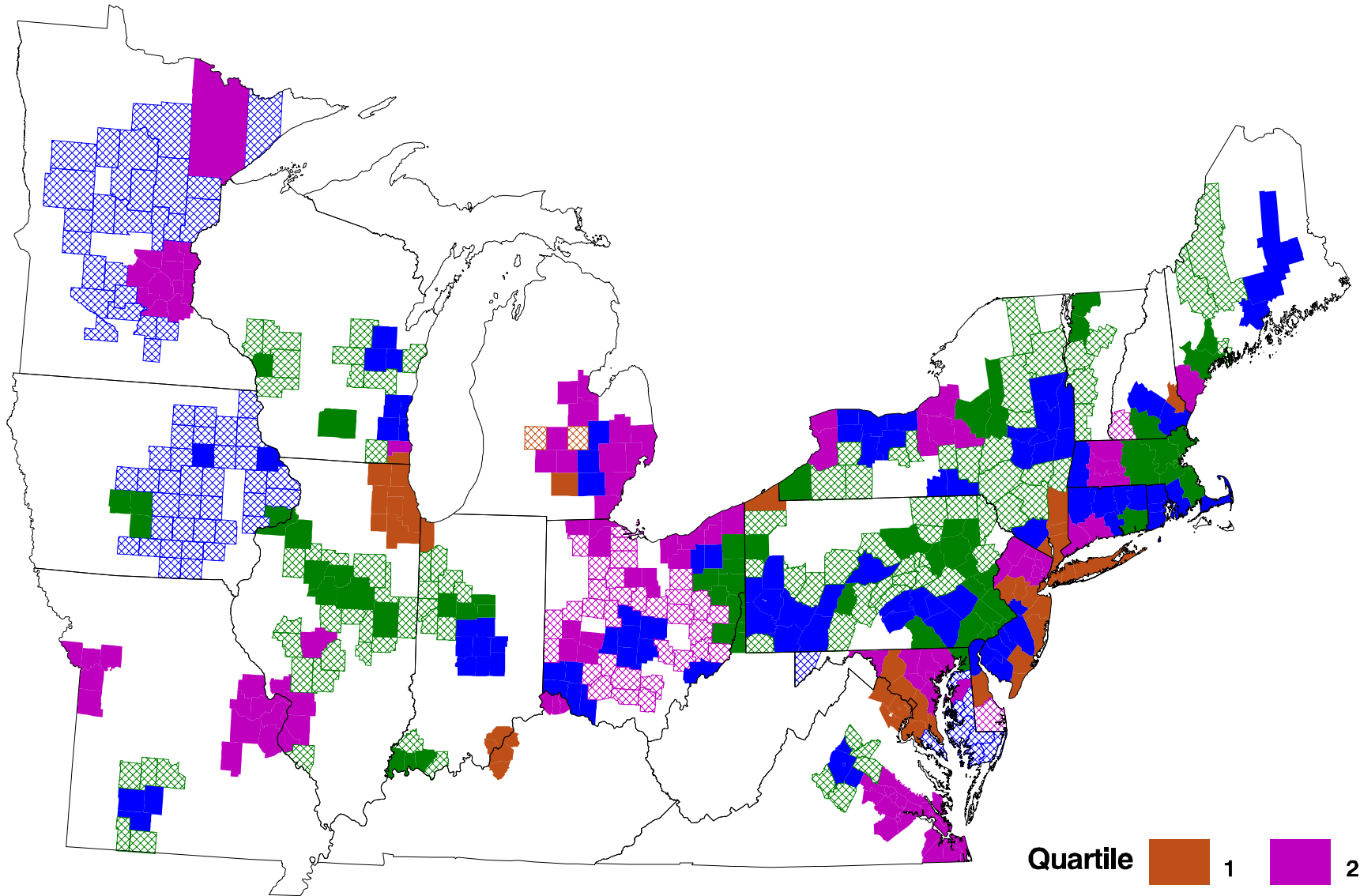


Shaded areas are non – MSA counties

Quartile 1 2
3 4
1 = Lowest, 4 = Highest Score

Doctor Composite. Medicare Managed Care CAHPS, 1997 – 2001

Northeastern States only

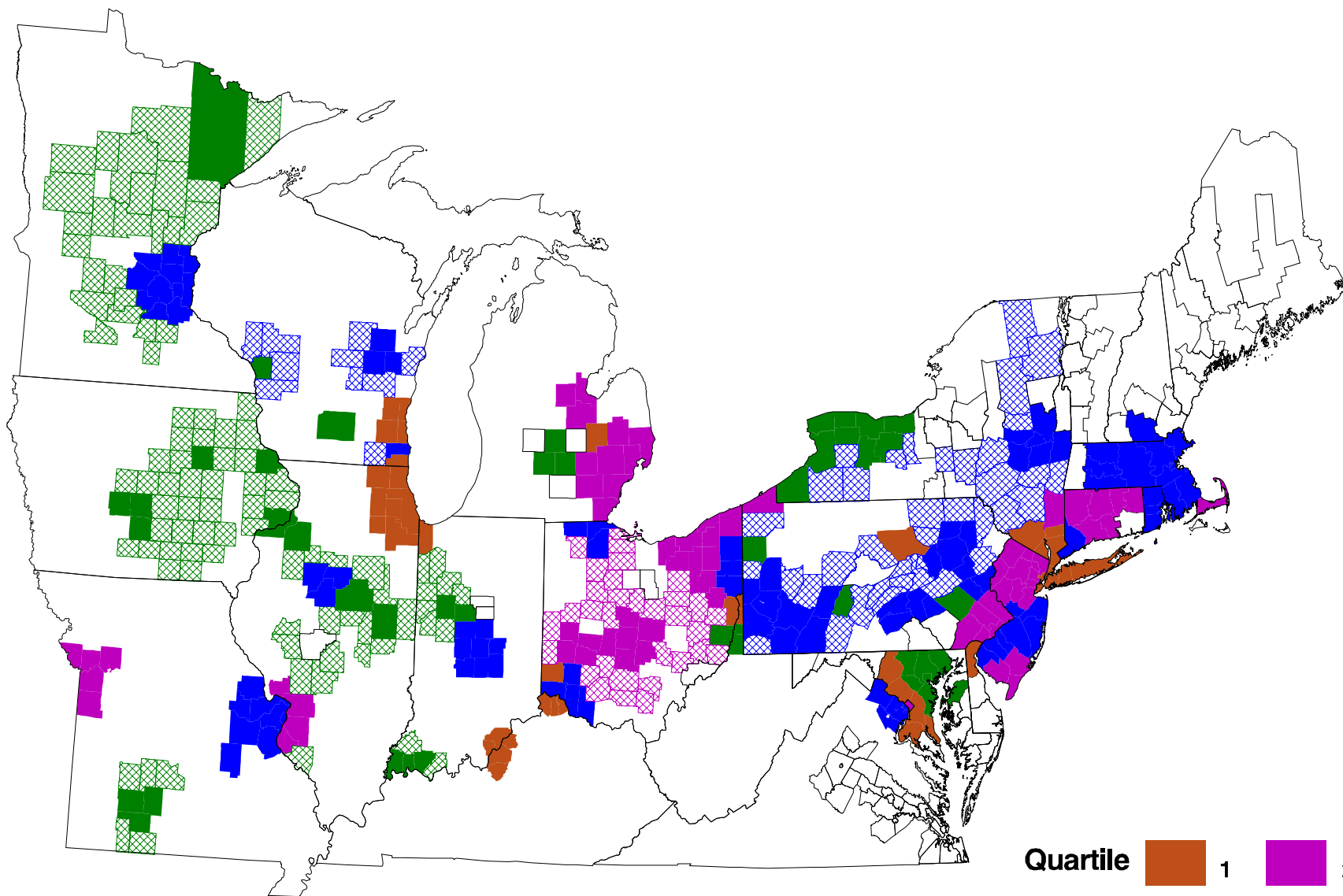


Shaded areas are non – MSA counties

Quartile 1 2
3 4
1 = Lowest, 4 = Highest Score

Prescription Composite. Medicare Managed Care CAHPS, 1998 – 2000

Northeastern States only

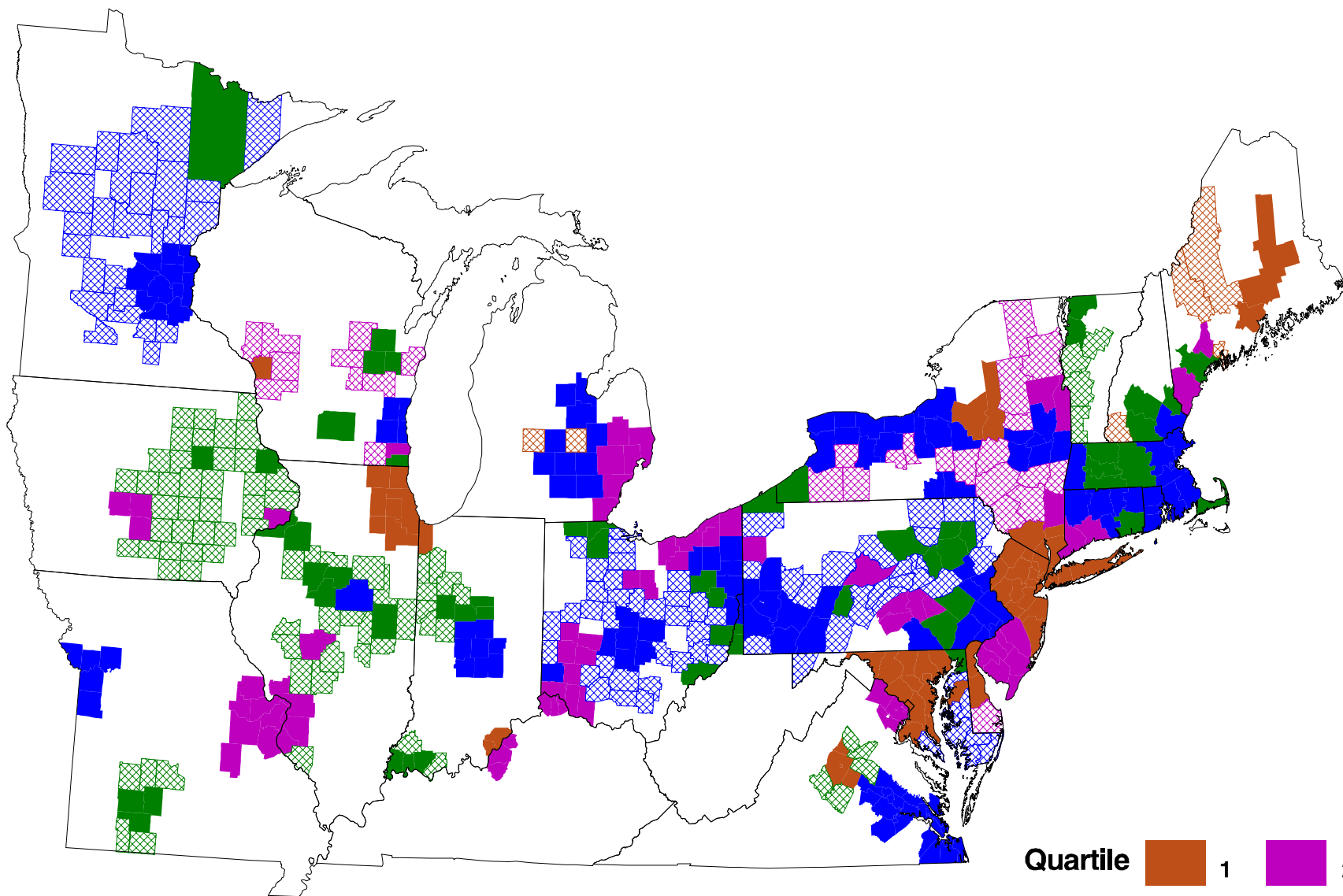


Shaded areas are non – MSA counties

Quartile 1 2
3 4
1 = Lowest, 4 = Highest Score

Service Composite. Medicare Managed Care CAHPS, 1997 – 2001

Northeastern States only

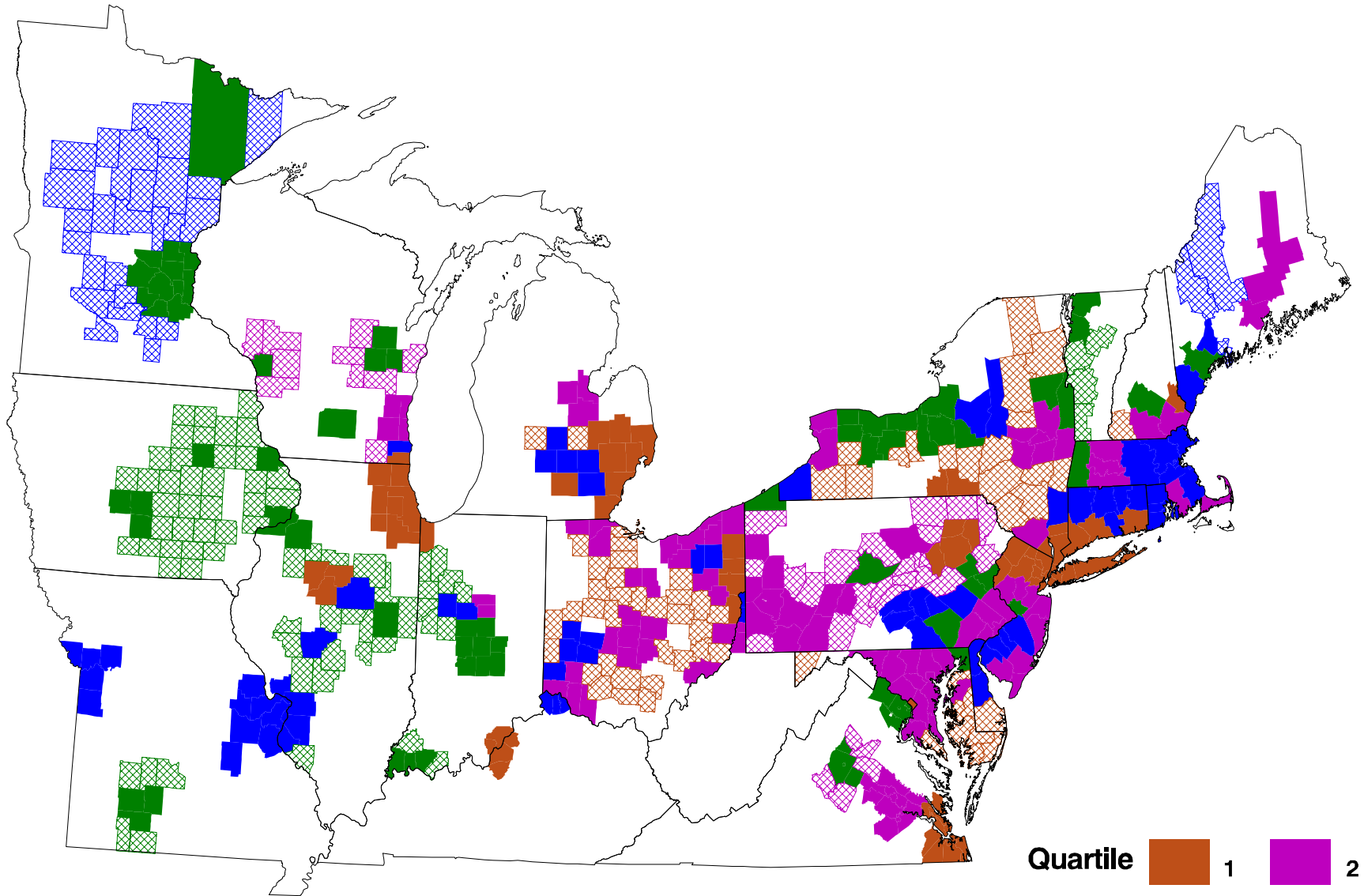


Shaded areas are non – MSA counties

Quartile 1 2
3 4
1 = Lowest, 4 = Highest Score

Vaccination Composite. Medicare Managed Care CAHPS, 1997 – 2001

Northeastern States only



Shaded areas are non – MSA counties

1 = Lowest, 4 = Highest Score