Multiple Attribution of Episodes for Physician Profiling in Medicare: A Preliminary Investigation

Robert L. Houchens, Ph.D.
Scott McCracken, M.B.A.
William Marder, Ph.D.
Robert Kelley, M.S.
William Anderson, Ph.D.

Thomson Reuters Healthcare
5425 Hollister Avenue
Suite 140
Santa Barbara, CA 93111-2348

•

MedPAC
601 New Jersey Avenue, NW
Suite 9000
Washington, DC 20001
(202) 220-3700
Fax: (202) 220-3759
www.medpac.gov

•

A study conducted by staff from Thomson Reuters Healthcare for the Medicare Payment Advisory Commission

June 2009 • No. 09–4

Multiple Attribution of Episodes For Physician Profiling in Medicare
A Preliminary Investigation

Final Report

Prepared for

Medicare Payment Advisory Commission

601 New Jersey Avenue, NW
Suite 9000
Washington, DC 20001

Prepared by

Robert L. Houchens, Ph.D.
Scott McCracken, M.B.A.
William Marder, Ph.D.
Robert Kelley, M.S.
William Anderson, Ph.D.

Thomson Reuters Healthcare
5425 Hollister Avenue, Suite 140
Santa Barbara, CA 93111–2348

THOMSON REUTERS HEALTHCARE

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

The purpose of this study was to compare multiple episode attribution to single episode attribution for profiling physicians on their treatment of Medicare patients. Multiple attribution occurs when a single episode is attributed to more than one physician for the purpose of physician performance measurement. This study is a follow-on to a previously reported study that assessed the stability of physician efficiency estimates using single attribution (Houchens, et al., 2007).

Researchers have expressed concerns over the ability of episode groupers to appropriately attribute episodes involving multiple physicians. For example, if a patient with congestive heart failure is treated by an internist, a cardiologist, and a cardiac surgeon, how should accountability be spread among the three specialists? This issue is especially germane to the Medicare population because a large proportion of patients tend to be elderly, have multiple conditions, and be treated by multiple physicians.

In the context of physician efficiency this issue is important because standards for resource utilization tend to be disease-specific, and a single physician may be held accountable for each episode of care regardless of the number of physicians involved in treatment. To the extent that some physicians more often treat episodes with multiple physicians, and to the extent that efficiency varies among the physicians, the one physician held accountable could be unfairly penalized or rewarded if the other physicians are more inefficient or more efficient. Also, multiple attribution could increase the number of episodes attributed (at least partially) to physicians, increasing the statistical precision of performance measures for them.

The multiple attribution problem is well-known. For example, in an August 2005 statement to CMS, the Association of American Medical Colleges made the following points:

- *Fair and accurate models for attributing care when multiple physicians treat patients must be implemented.*

- *AAMC requests that CMS give careful consideration to the methodology used to attribute care to individual physicians or groups when more than one physician/group treats a patient. Furthermore, CMS should develop a quality improvement/performance system that correctly attributes responsibility for and delivery of care and thereby minimizes potential unintended consequences.*

However, few solutions have been put forward. Most stakeholders recognize that even with perfect information it would be difficult to equitably divide responsibility for complicated mixtures of resource utilization among multiple physicians treating a single patient in an episode. Therefore, the best we can do is try to arrive at a reasonable approximation to reality in these situations, but which is perhaps better than a “winner-take-all” strategy.

Single attribution is designed to identify the “decision-maker,” perhaps the primary care physician, and hold this individual responsible for all care rendered. Multiple attribution acknowledges that the decision-maker, if there is one, has incomplete control over treatment by specialists and other physicians, even if the decision-maker referred the patient to those other physicians. Depending on the care model, there might be a place for both approaches.

This study explores the extent and the nature of the problem. We calculate the percentage of episodes and total dollars for episodes involving multiple physicians. We also compare physicians performance indices based on multiple attribution strategies to performance indices based on single attribution strategies. We created eight indices of physician performance, all based on a ratio of observed to expected episode payment ratios. Four indices are based on total episode payments and Four are based solely on evaluation and management (E&M) payments. These indices are defined in the Methods section of this report.

Physician performance indices using each multiple attribution strategy are compared with indices using the other multiple attribution strategies as well as with the single attribution indices to test the sensitivity of the results to the attribution method. One advantage of the multiple attribution strategies is likely to be a higher attribution rate for episodes because the single attribution methods makes no assignment when the episode dollars fall short of the 35 percent threshold for all physicians.

A recent study (Pham et al., 2007) that analyzed Medicare claims from 2000 through 2002, along with the Community Tracking Study Physician Survey in 2000 and 2001, determined that the conventional means of assigning one physician to a patient is problematic for physician profiling and pay-for-performance. This study based assignment on an entire year of data, not episodes. It is possible that, even with single attribution, the assignment of physicians to episodes rather than an entire year will mitigate the problems highlighted by that study.

The data for our study were provided by the Medicare Payment Advisory Commission (MedPAC). They comprised all 2002 and 2003 Medicare claims for patients residing in six metropolitan statistical areas (MSAs): Boston, MA; Greenville, SC; Miami, FL; Minneapolis, MN; Orange County, CA; and Phoenix, AZ. The data are described in the Data section of this report. The detailed results of applying MEG to the data are contained in the Appendix.

The Results section describes the levels of agreement among the various performance measures as well as the relative stability—the year-to-year persistence— of the indices. The final section, Conclusions and Recommendations, contains a broad assessment of the results, some important caveats to the study, and some considerations for future studies.

DATA

MedPAC provided the study data, composed of all medical claims during the calendar years 2001 through 2004 for Medicare beneficiaries residing in the six study MSAs: Boston, Greenville, Miami, Minneapolis, Orange County and Phoenix. Table 1 displays the number of claims, by year, for each claim source.

Table 1: Number of Medicare Claims, by Source and Year.

Claim Source	Year				Total
	2001	2002	2003	2004	
HHA	147,523	159,901	178,903	197,674	**684,001**
MEDPAR	575,519	591,412	618,358	633,141	**2,418,430**
Physician	47,342,026	51,054,090	55,980,215	57,916,665	**212,292,996**
Outpatient	14,961,933	16,035,609	16,855,777	18,030,835	**65,884,154**
Total	**63,027,001**	**67,841,012**	**73,633,253**	**76,778,315**	**281,279,581**

Note: An entire facility record, whether inpatient, outpatient, or home health agency, was grouped to a single episode based on the principal diagnosis code present on the claim. Physician claims may contain charges for a number of services provided to a patient during a visit. The diagnosis, date, and payment for each service are detailed on a claim line and each claim line is evaluated separately and assigned to the appropriate episode group.

Key Variables in the Raw Data

The following data elements, which are necessary for episode creation, were extracted from the raw data files and placed in a uniform format:

- Patient ID – a unique and encrypted patient identifier.
- UPIN – a unique physician identification number.
- Diagnosis Codes – the reconfigured claims records contained up to 11 diagnosis codes assigned using the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) diagnosis coding system.
- Procedure Codes – Current Procedural Terminology (CPT) procedure codes, Healthcare Common Procedure Coding System (HCPCS) procedure codes, and ICD-9-CM procedure codes were extracted from the original data. Each claim record contained one procedure code.
- MSA – the patient's metropolitan statistical area.
- Standardized Payment – as described below, claims payment amounts were standardized to remove local market payment differences among episodes.
- Age – patient age, in years.
- Gender – patient gender.
- Date of Service – the date of outpatient service or the date of admission.
- Claim Number – a unique record identification number.
- Length of Stay- inpatient length of stay.

Standardized Payments

MedPAC established methods for standardizing payments for physician profiling applications with episode groupers (Medicare Payment Advisory Commission, 2006). Briefly, for each type of claim, MedPAC standardized payments as follows:

- Hospital inpatient services — A standardized amount was created for each Diagnosis Related Group (DRG) for each year and applied to all records uniformly.
- Skilled Nursing Facility (SNF) services— SNF Medicare Provider Analysis and Review records were merged to the DataPro SNF Stay file. This information was combined with specific standardized amounts of resource utilization groups from CMS.
- Long-term care hospital services —For discharges that occurred on or after October 1, 2002, a standardized amount for each DRG was applied. For discharges prior to this date, local area wage-index adjustments from each hospital's payment were backed-out, assuming local area wage indexes acted as a proxy for underlying costs.
- Rehabilitation/psychiatric hospital services —Total Medicare payments and total length of stay were calculated for each DRG, a DRG-level per diem amount was created and then multiplied by the length of stay for each record.
- Home health — the home health case-mix weight on each claim was multiplied times the base payment rate for the appropriate fiscal year.
- Physician services — The allowed charge was adjusted by the physician geographic adjustment factor (GAF) to create a standardized payment amount.
- Ambulatory Surgical Center (ASC) services — HCPCS codes were used to match records to ASC payment rate files. Consistent with Medicare payment rules the payment rate was reduced for multiple surgical procedures on the same claim.
- Clinical laboratory services — A record was classified as a clinical lab service if the HCPCS for a record on the carrier file matched a HCPCS code on the clinical lab fee schedule. The standardized payment rate for each lab record is the national limitation amount (NLA) for the service.
- Anesthesia services —The base and the time units were summed for each anesthesia record and multiplied by the anesthesia conversion factor for the appropriate year. Payments for services provided with and without medical direction were adjusted consistent with Medicare payment rules.
- Hospital outpatient services — HCPCS codes were used to match outpatient records to an outpatient prospective payment system payment rate file and a standardized payment amount was assigned to each record.

In this study, the total payment for an episode is the total of the standardized payments for the claims contained in that episode. Throughout this report the term "payment" is shorthand for "standardized payment."

Berenson-Eggers Type of Service (BETOS)[1]

The BETOS coding system was developed primarily to analyze the growth in Medicare expenditures. The coding system assigns each and every HCPCS codes to a single BETOS code, which represents a readily understood clinical category. BETOS codes were added to professional and outpatient claims.

BETOS codes are broadly classified under seven major categories:

1. Evaluation and Management
2. Procedures
3. Imaging
4. Tests
5. Durable Medical Equipment
6. Other
7. Exceptions/Unclassified

The category of Evaluation and Management (E&M) played a special role in the assignment of episodes to physicians, as explained in the Methods section.

[1] See www.cms.hhs.gov/HCPCSReleaseCodeSets/20_BETOS.asp (last accessed 9/9/2007) for more information on BETOS categories.

METHODS

The Appendix contains a description the Medical Episode Grouper (MEG^{TM}), which was used to produce episodes for our analyses. Throughout this report, the term "payment" is shorthand for "standardized payment." We used SAS software for all analyses.

Physician Performance Indices

For each physician we calculate four multiple attribution performance indices, two based on E&M payment weights (M_E and M_{ER}), and two based on total payment weights (M_T and M_{TR}). We also calculate four single attribution performance indices, two based on E&M payments (S_E and S_{ER}), and two based on total payments (S_T and S_{TR}). The indices are defined as follows.

Mulitple Attribution Indices

For physician k, the first multiple attribution performance index, M_E, weights each episode payment ratio (observed / expected) by the physician's share of E&M payments. For example, if the E&M dollars in an episode involving three physicians are \$100, \$200, and \$700 then the performance ratio for that episode is weighted 0.1, 0.2, and 0.7 for each physician, respectively. Thus, each physician's average performance ratio is a weighted average, with weights proportional to the physician's share of the E&M dollars for each episode:

$$M_E(\text{k}) = \frac{\sum_{i=1}^{n_k} w_{ik}\left(\frac{o_i}{e_i}\right)}{\sum_{i=1}^{n_k} w_{ik}}$$

$n_k =$ number of episodes involving physician k

$$w_{ik} = \frac{\text{E\&M dollars for physician k in episode i}}{\text{E\&M dollars for all physicians in episode i}}$$

$o_i =$ observed total standardized dollars for episode i

$e_i =$ expected total standardized dollars for episode i

Note that a physician can be involved in an episode without a payment for an E&M visit, in which case the physician's weight for that episode would be zero (e.g., $w_{ik} = 0$).

The second multiple attribution performance index, M_{ER}, is the ratio of the weighted sum of observed episode payments to the weighted sum of expected episode payments, with weights equal to the physician's share of E&M payments:

$$M_{ER}(\text{k}) = \frac{\sum_{i=1}^{n_k} w_{ik} o_i}{\sum_{i=1}^{n_k} w_{ik} e_i}$$

It is easily shown that M_{ER} is an (expected) dollar-weighted version of the index M_E. To see this, substitute (e_i*w_i) for (w_i) in the formula for M_E. The result is the formula for M_{ER}.

The third multiple attribution performance index, M_T, weights each episode payment ratio (observed / expected) by the physician's share of total payments:

$$M_T(\mathrm{k}) = \frac{\sum_{i=1}^{n_k} t_{ik}\left(\frac{o_i}{e_i}\right)}{\sum_{i=1}^{n_k} t_{ik}}$$

$$t_{ik} = \frac{\text{Total dollars for physician k in episode i}}{\text{Total dollars for all physicians in episode i}}$$

The fourth multiple attribution performance index, M_{TR}, is the ratio of the weighted sum of observed episode payments to the weighted sum of expected episode payments, with weights equal to the physician's share of total payments:

$$M_{TR}(\mathrm{k}) = \frac{\sum_{i=1}^{n_k} t_{ik} o_i}{\sum_{i=1}^{n_k} t_{ik} e_i}$$

Again, it is easy to see that M_{TR} is an (expected) dollar-weighted version of the index M_T. To see this, substitute (e_i*t_i) for (t_i) in the formula for M_T. The result is the formula for M_{TR}.

Single Attribution Indices

We calculated two single-attribution indices based on E&M dollars and two single-attribution indices based on total dollars. The above formulas apply essentially with a weight equal to 1 for the single physician attributed to each episode and equal to 0 for other physicians involved in the episode.

The two indices based on E & M dollars were:

$$S_E(\mathrm{k}) = \frac{\sum_{i=1}^{n_e} \frac{o_i}{e_i}}{n_e}$$

$$S_{ER}(\mathrm{k}) = \frac{\sum_{i=1}^{n_e} o_i}{\sum_{i=1}^{n_e} e_i}$$

where n_e is the number of episodes for which the physician had the highest proportion of E&M dollars (at least 35 %).

The two indices based on total dollars were:

$$S_T(\text{k}) = \frac{\sum_{i=1}^{n_t} \frac{o_i}{e_i}}{n_t}$$

$$S_{TR}(\text{k}) = \frac{\sum_{i=1}^{n_t} o_i}{\sum_{i=1}^{n_t} e_i}$$

where n_t is the number of episodes for which the physician had the highest proportion of total dollars (at least 35 %).

Statistical Methods

Physician rankings using each attribution strategy are compared with rankings using the other attribution strategies to test the sensitivity of the results to the attribution method. For this, we calculate correlations among the physician indices, weighting each physician by his or her number of episodes. One advantage of the multiple attribution strategies is likely to be a higher attribution rate for episodes because the single attribution method makes no assignment when the E&M dollars fall short of the 35 percent threshold for all physicians involved in an episode.

To measure the "stability" of the efficiency measures, we calculate the correlation of each index between 2002 and 2003. Indices that are more stable—have higher correlations between years—may be more desirable, all else equal.

RESULTS

Table 2 shows the total number of physicians along with information on the range of physician sample sizes (episodes "touched" per physician) for 2002 and 2003. These numbers differ from the number of episodes *attributed* to physicians, which varies by attribution method. We give more details on the distribution of samples sizes separately for the subset of physicians with more than 20 episodes. Many of the statistics reported later are restricted to the subset of physicians with at least 20 (attributed) episodes. For the correlations reported later, we weight each physician by his or her number of episodes to account for the lower reliability associated with smaller physician samples. The observed numbers of physicians are lowest in Greenville and highest in Boston. Within each MSA, the number of physicians and the number of episodes per physician is similar between the two years.

Table 2: Counts of Physicians and Episodes Touched per Physician.

MSA	Total Physicians		Physicians with at least 20 Episodes Touched		Episodes per Physician (Among Physicians with at Least 20 Episodes Touched)							
					Mean		10th percentile		Median		90th percentile	
	2002	2003	2002	2003	2002	2003	2002	2003	2002	2003	2002	2003
Total	50,464	53,139	35,533 (70 %)	37,450 (70 %)	343	352	33	34	175	178	828	854
Boston	16,495	17,191	11,111 (67 %)	11,615 (68 %)	314	337	33	34	169	175	771	830
Greenville	2,715	2,948	2,137 (79 %)	2,254 (76 %)	623	613	48	46	410	394	1469	1465
Miami	6,331	6,654	4,787 (76 %)	4,969 (75 %)	409	417	36	38	214	213	963	984
Minneapolis	10,015	10,565	7,098 (70 %)	7,486 (71 %)	271	268	32	31	147	147	657	636
Orange Co.	6,570	6,835	4,450 (68 %)	4,715 (69 %)	343	347	32	33	163	165	772	785
Phoenix	8,338	8,946	5,950 (71 %)	6,411 (72 %)	328	338	32	32	159	166	809	818

The variation across MSAs in episodes per physician might be an artifact of the data. Recall that the data contain only episodes from patients residing in the six MSAs and they exclude episodes from patients residing in other MSAs. For example, Boston physicians might have fewer episodes in these data because Boston physicians might serve a larger proportion of patients outside the Boston MSA than, say, Greenville physicians serve outside the Greenville MSA.

Figure 1 displays the distribution of Boston episodes, by the number of physicians involved in each episode in 2002 and 2003. The horizontal axis reflects the number of physicians, ranging from 1 to 10 or more. The vertical axis shows the percentage of Boston episodes. Recall that the Boston data contain episodes for Boston *patients* and, because our focus is on Boston physicians, a small percentage of episodes were excluded that involved only physicians outside the Boston MSA. However, we retained all episodes that involved at least one Boston physician, even if non-Boston physicians were involved in the treatment.

About 55 percent of the Boston episodes were treated by a single Boston physician. About 20 percent were treated by two physicians. The same pattern was evident for other MSAs, as shown in Figure 2.

However, the pattern is very different with respect to payments because episodes with more physicans tend to have higher payments. Figure 1 shows the pattern for Boston. While Figure 1 showed that approximately 25 percent of *episodes* involved 3 or more physicians, Figure 3 shows that in Boston approximately 75 percent of *payments* involved 3 or more physicians.

While an uncritical inspection of Figure 2 appears to show that the distribution of <u>episodes</u> by number of physicians was quite similar across MSAs, Figure 4 indicates that the distribution of <u>total payments</u> by number of physicians varied considerably. For example, episodes involving 10 or more physicians (yellow bars) accounted for a relatively lower percentage of total dollars in Miami (13 %) and Orange County (12 %) compared with Minneapolis (22 %), Boston (19 %), and Greenville (18 %). Closer inspection of Figure 2 does show that a slightly higher percentage of Minneapolis episodes involve 10 or more physicians.

To further examine the different patterns of payment variation, Figure 5 shows the <u>average payment</u> per episode by the number of physicians involved in an episode for Boston. Not surprisingly, the average payment per episode climbs stealily as a function of the number of physicians involved in the episode, and the pattern is nearly identical for both 2002 and 2003 in Boston. Similar trends can be observed for other MSAs in Figure 6. In the range of 3 to 9 physicians per episode, each additional physician adds nearly another $1,000 to the average episode payment. The pattern for Miami is slightly different from that for other MSAs. Compared with other MSAs, the average cost per episode in Miami tends to be higher for episodes involving 1 to 3 physicians and lower for episodes involving 4 or more physicians.

These findings illustrate the potential importance of multiple attribution versus single attribution strategies for physician profiling. The financial stakes increase as the number of physicians involved in an episode grows.

Figure 1: Distribution of Boston Episodes, by Number of Physicians per Episode, Exluding Episodes without a Boston Physician, 2002.

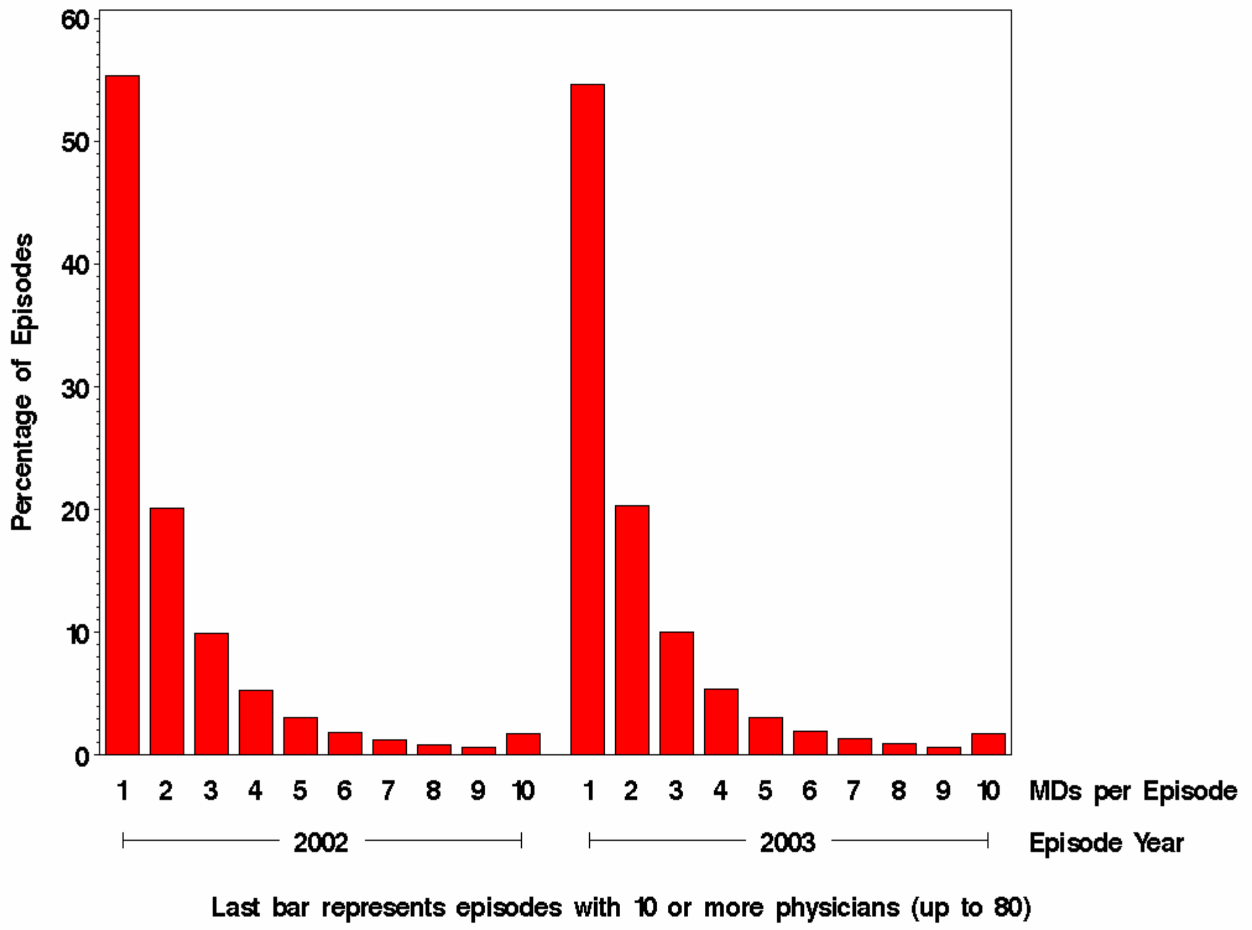


Figure 2: Distribution of Episodes, by Number of Physicians per Episode and MSA of Physician, Exluding Episodes without a Physician in the MSA, 2002.

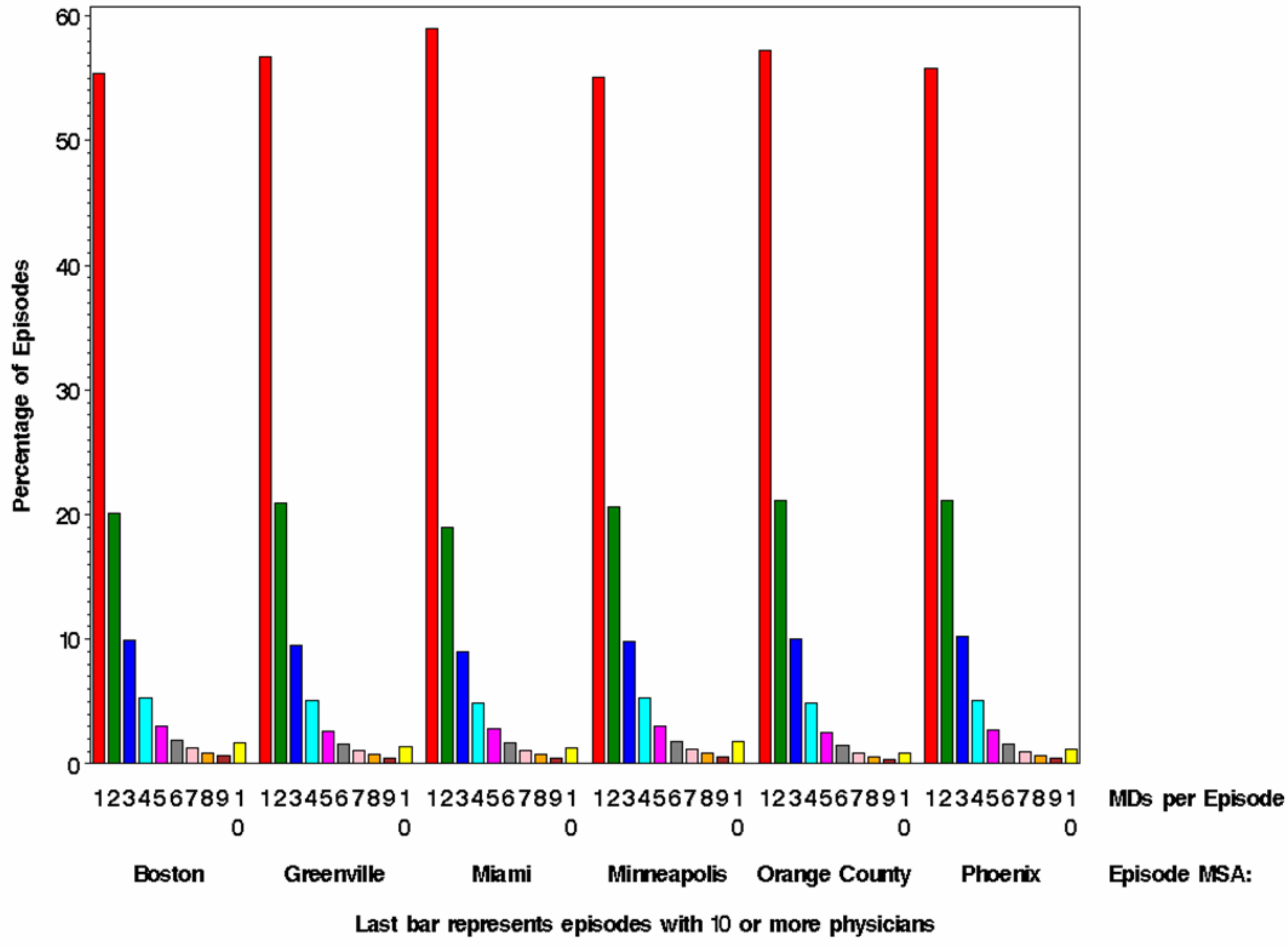


Figure 3: Payment Distribution of Boston Episodes, by Number of Physicians per Episode, Episodes with at Least One Boston Physician, 2002 and 2003.

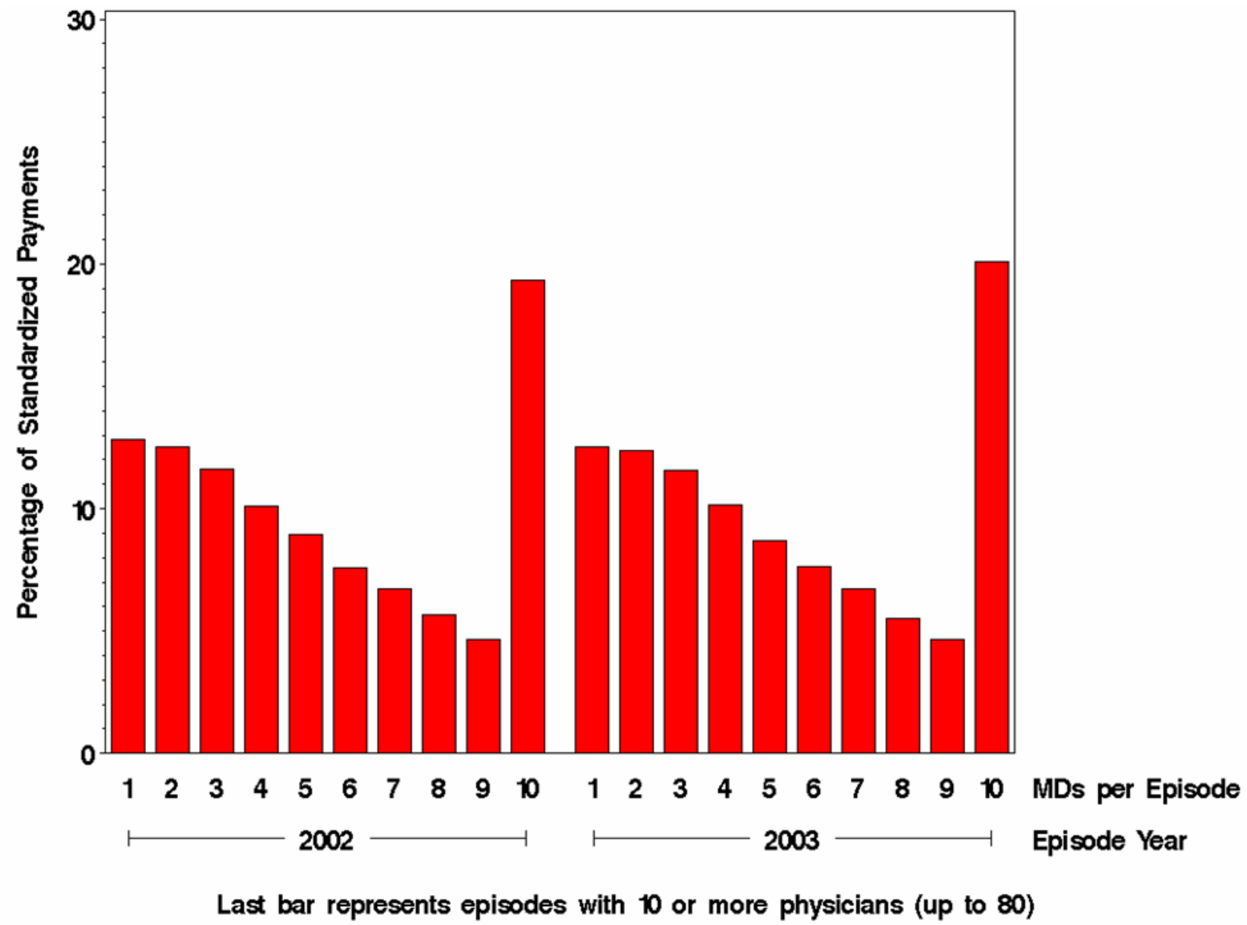


Figure 4: Payment Distribution of Episodes, by MSA and by Number of Physicians per Episode, Episodes with at Least One Physician in the MSA, 2002.

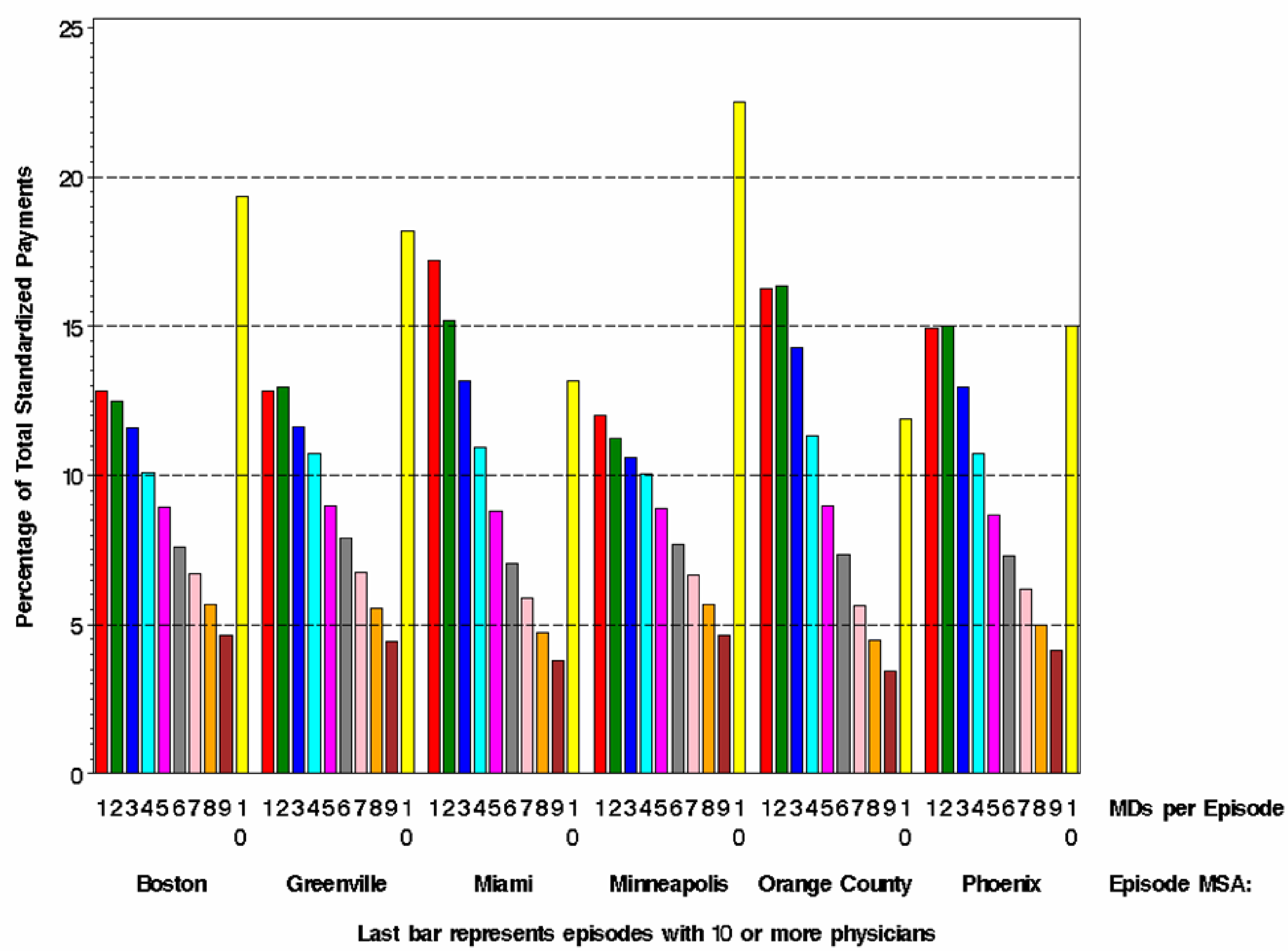


Figure 5: Average Payments of Boston Episodes, by Number of Physicians per Episode, Episodes with at Least One Boston Physician, 2002 and 2003.

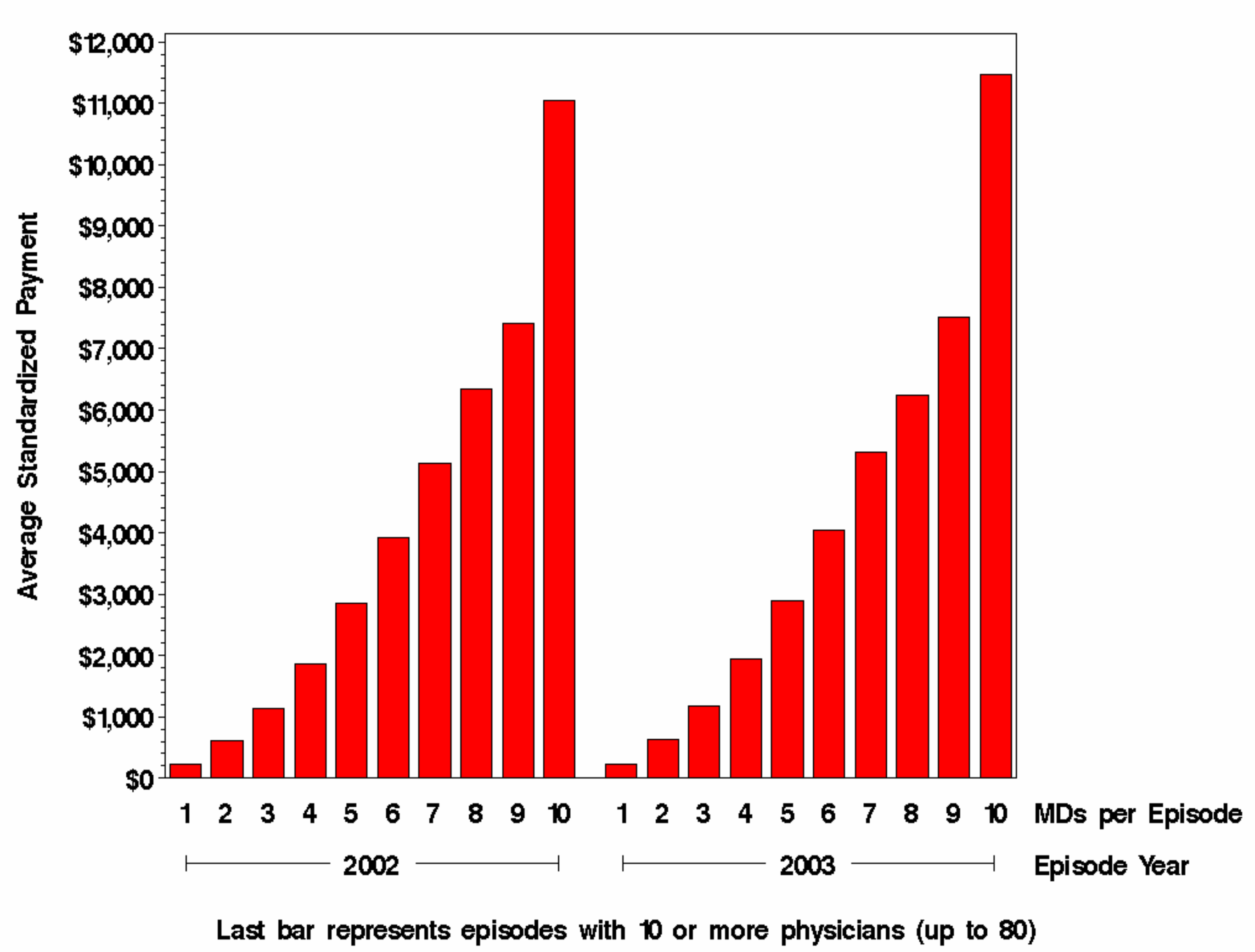
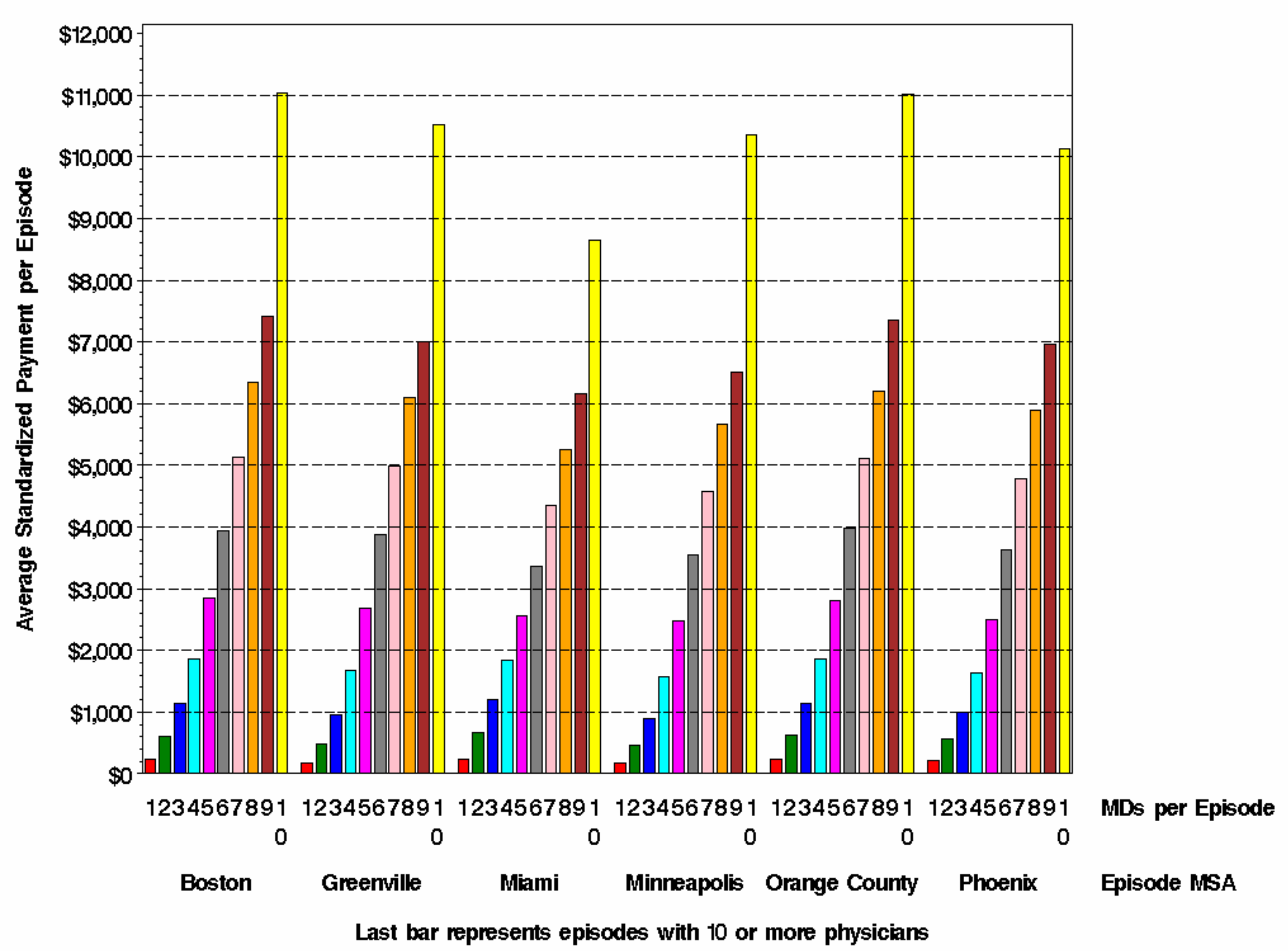


Figure 6: Average Episode Payments, by MSA and by Number of Physicians per Episode, Episodes with at Least One Physician in the MSA, 2002.



We now turn to an analysis of the various physician performance measures. First, we look at the stability of the various indices described in the Methods section. We measure stability by the year-to-year correlation for each measure at the physician level, where each physician is weighted by his or her number of episodes per year (averaged over 2002 and 2003).

The year-to-year correlations, shown in Table 3, tend to be fairly high for all of the performance indices. The lowest correlation was 80 percent for index S_{ER} and the highest correlation was 91 percent for indices S_T and M_T. Consequently, the stability of the index might not be a deciding factor for which index to use in physician profiling. However, for each attribution type and weighting method, the mean ratios had consistently higher year-to-year correlations compared with the ratio of means.

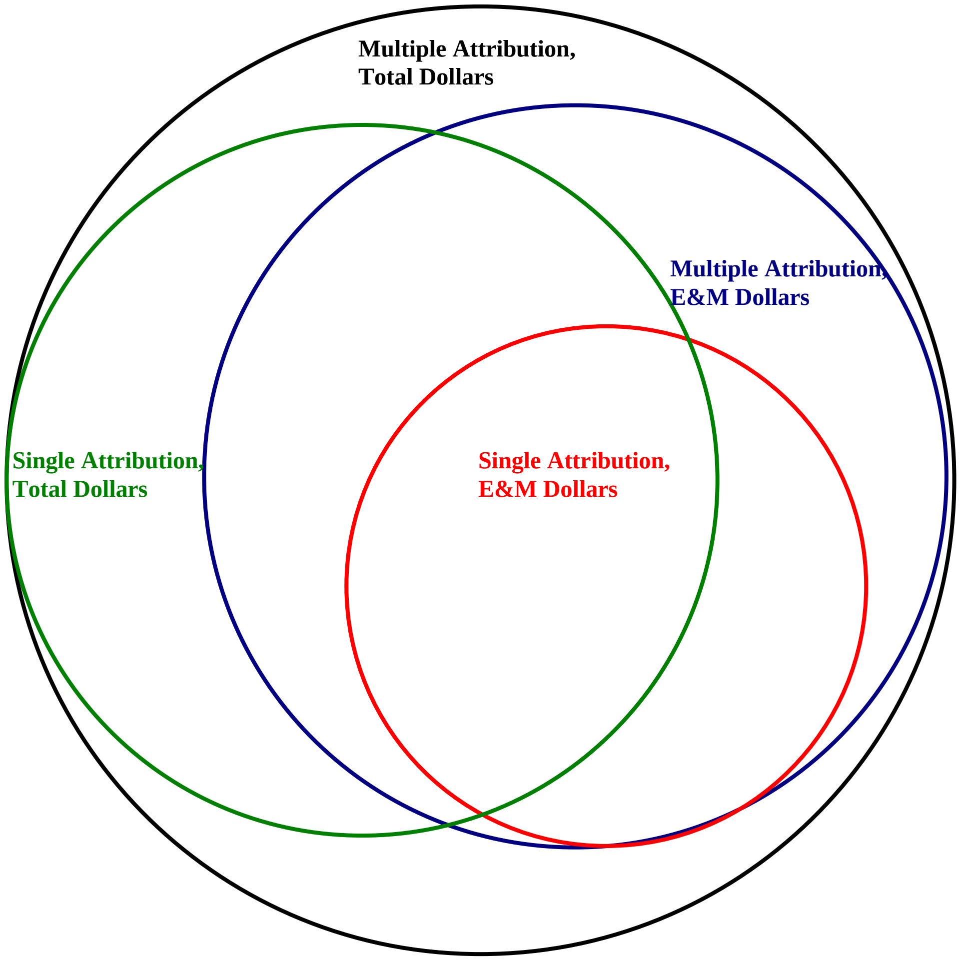
Table 3: Year-to-Year Correlations.

Attribution Type	Weights	Calculation	Index Name	Correlation %
Multiple	**E & M dollars**	**Mean Ratio**	M_E	89.2
		Ratio of Means	M_{ER}	85.0
	Total dollars	**Mean Ratio**	M_T	90.5
		Ratio of Means	M_{TR}	88.6
Single	**E & M dollars***	**Mean Ratio**	S_E	87.4
		Ratio of Means	S_{ER}	80.2
	Total dollars**	**Mean Ratio**	S_T	91.2
		Ratio of Means	S_{TR}	88.3

* For single attribution, the the physician with the most E&M dollars is given a weight of 1 for the episode and all other physicians are given a weight of 0.
** For single attribution, the the physician with the most total dollars is given a weight of 1 for the episode and all other physicians are given a weight of 0.
Note: Correlations were based on physicians with at least 20 attributed episodes in both years and each physician was weighted for his or her average number of episodes per year.

How much of a difference is there among the various performance measures? It helps to recognize the relationship among the groups of episodes attributed under the various attribution methods. First, for any physician, the set of episodes attributed under single attribution based on maximum E&M dollars (S_E and S_{ER}) must be a subset of the episodes attributed under multiple attribution with E&M dollar weights (M_E and M_{ER}). Second, the set of episodes attributed under single attribution based on maximum total dollars (S_T and S_{TR}) must be a subset of the episodes attributed under multiple attribution with total dollar weights (M_T and M_{TR}). Third, E&M dollars are a subset of total dollars in each episode. Therefore, by construction, for any physician the set of episodes contributing to S_E and S_{ER} must be a subset of the episodes contributing to M_E and M_{ER}, which in turn must be a subset of the episodes contributing to M_T and M_{TR}. Similarly, the set of episodes contributing to S_T and S_{TR} must be a subset of the set of the episodes contributing to M_T and M_{TR}. These relationships are displayed in the Venn diagram shown in Figure 7.

Figure 7: Relationship Among Sets of Episodes for Each Attribution Method



Given the overlap in episodes among the attribution methods, we might expect the lowest correlation to occur between single attribution indices based on maximum total dollars (S_T and S_{TR}) versus the indices for attribution methods based on E&M dollars (S_E , S_{ER}, M_E, and M_{ER}), although the difference in episode weights among the other indices could also cause substantial differences.

The correlations among the indices are shown in Table 4. *These correlations are based on physicians who were attributed at least 20 episodes under both methods in the pairing*, and the physicians were weighted by the average number of episodes in the pair. For example, if a physician had 20 episodes attributed using the M_E index and 30 episodes using the S_E index, then that physician would have been included in the calculation of the correlation between M_E and S_E with a weight of 25. The correlations for 2002 and 2003 are both shown (upper and lower entry, respectively), but there is little difference between the years. Since stability was estimated to be higher for the indices based on mean ratios compared with indices based on the ratio of means, we concentrate on the inter-correlations among those measures (highlighed in bold).

First, the correlations are quite high between S_E and M_E (97 %) and between S_T and M_T (95 %). That is, single attribution and multiple attribution indices give very similar results based on E & M dollars, and likewise for indices based on total dollars. The correlations are somewhat lower—

between 85 percent and 90 percent—when comparing indices based on E & M dollars (S_E and M_E) to indices based on total dollars (S_T and M_T), as we thought they might be.

Table 4: Correlations Among Indices (below diagonal) and Number of MDs (above diagonal), 2002 (upper entry in each cell) and 2003 (lower entry in each cell).

Index	M_E	M_{ER}	M_T	M_{TR}	S_E	S_{ER}	S_T	S_{TR}
M_E	******	28,099 29,563	28,099 29,563	28,099 29,563	24,300 25,529	24,300 25,529	24,356 25,690	24,356 25,690
M_{ER}	91 91	******	28,099 29,563	28,099 29,563	24,300 25,529	24,300 25,529	24,356 25,690	24,356 25,690
M_T	**90** **91**	81 82	******	50,131 52,828	24,300 25,529	24,300 25,529	24,356 25,690	24,356 25,690
M_{TR}	81 82	87 89	92 92	******	24,300 25,529	24,300 25,529	24,356 25,690	24,356 25,690
S_E	**97** **97**	88 88	**86** **87**	78 80	******	24,300 25,529	24,356 25,690	24,300 25,529
S_{ER}	84 84	94 94	74 75	82 84	88 88	******	24,300 25,529	24,300 25,529
S_T	**85** **86**	77 78	**95** **95**	88 88	**86** **87**	75 76	******	27,177 28,638
S_{TR}	72 73	79 80	83 83	93 93	74 75	80 81	90 90	******

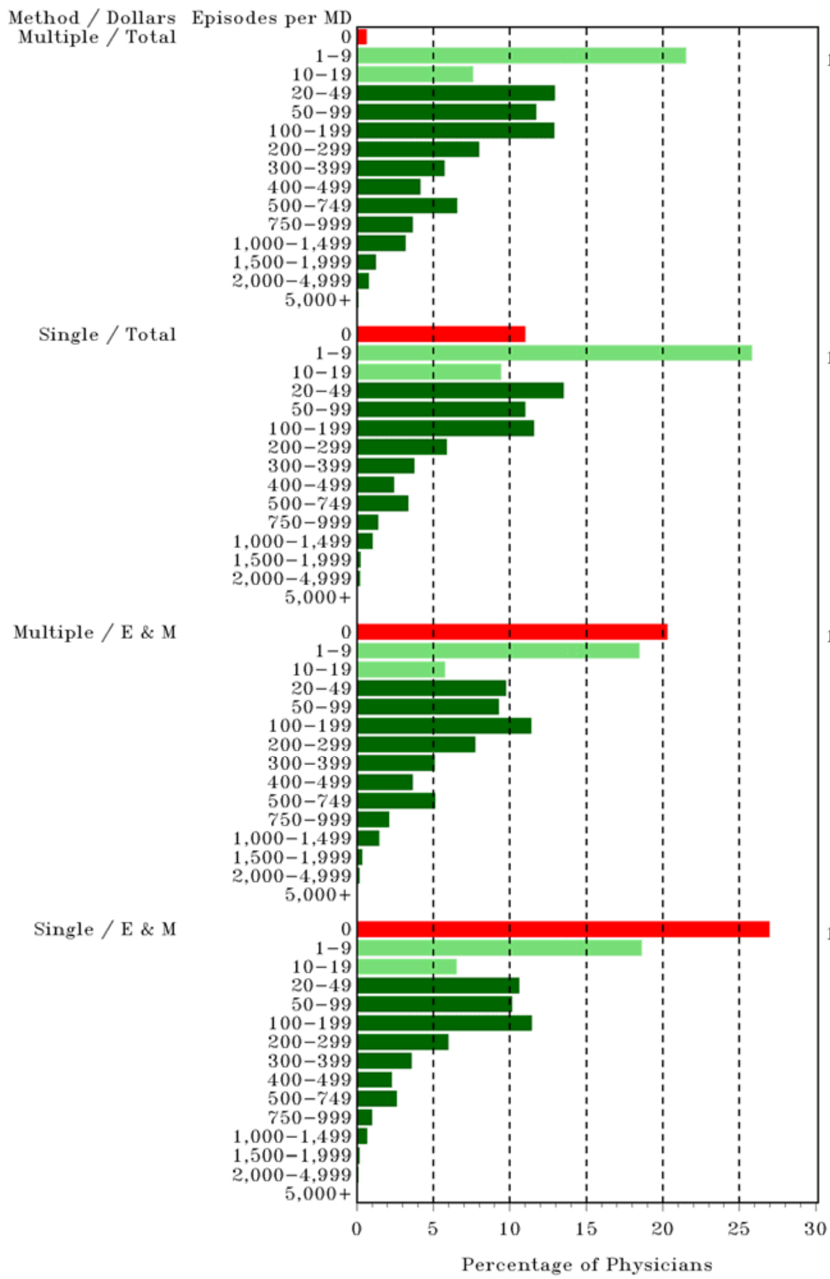
Based on the results from tables 3 and 4, a particular index does not appear to be statistically superior to all other indices. Consequently, the choice of index must rest on other considerations.

One difference among the indices is the number of episodes attributed to physicians. Figure 8 shows the distribution of the unweighted number of episodes attributed to physicians under the various attribution and weighting strategies. There are four distributions shown in the figure:

1) Multiple / Total = multiple attribution with total dollar weighting (M_T and M_{TR}),
2) Single / Total = single attribution based on total dollars (S_T and S_{TR}),
3) Multiple / E & M = multiple attribution with E&M dollar weighting (M_E and M_{ER}), and
4) Single / E & M = single attribution based on E&M dollars (S_E and S_{ER}),

In each distribution, the red bar corresponds to the percentage of physicians who were attributed zero episodes under that attribution method. The attribution strategies based on total dollars (top two distributions) had fewer physicians with zero episodes compared with the attribution strategies based on E & M dollars (bottom two distributions) because some episodes did not have any E & M dollars associated with them.

Figure 8: Unweighted Episodes per Physician, by Attribution Method



Method / Dollars	Episodes per MD	MDs	Cumm. MDs	%	Cumm. %
Multiple / Total	0	311	311	0.59	0.59
	1-9	11412	11723	21.48	22.06
	10-19	4019	15742	7.56	29.62
	20-49	6854	22596	12.90	42.52
	50-99	6201	28797	11.67	54.19
	100-199	6831	35628	12.85	67.05
	200-299	4212	39840	7.93	74.97
	300-399	3012	42852	5.67	80.64
	400-499	2187	45039	4.12	84.76
	500-749	3458	48497	6.51	91.26
	750-999	1905	50402	3.58	94.85
	1,000-1,499	1670	52072	3.14	97.99
	1,500-1,999	641	52713	1.21	99.20
	2,000-4,999	395	53108	0.74	99.94
	5,000+	31	53139	0.06	100.00
Single / Total	0	5819	5819	10.95	10.95
	1-9	13707	19526	25.79	36.75
	10-19	4975	24501	9.36	46.11
	20-49	7149	31650	13.45	59.56
	50-99	5820	37470	10.95	70.51
	100-199	6131	43601	11.54	82.05
	200-299	3104	46705	5.84	87.89
	300-399	1969	48674	3.71	91.60
	400-499	1272	49946	2.39	93.99
	500-749	1757	51703	3.31	97.30
	750-999	712	52415	1.34	98.64
	1,000-1,499	521	52936	0.98	99.62
	1,500-1,999	108	53044	0.20	99.82
	2,000-4,999	90	53134	0.17	99.99
	5,000+	5	53139	0.01	100.00
Multiple / E & M	0	10752	10752	20.23	20.23
	1-9	9793	20545	18.43	38.66
	10-19	3031	23576	5.70	44.37
	20-49	5155	28731	9.70	54.07
	50-99	4900	33631	9.22	63.29
	100-199	6038	39669	11.36	74.65
	200-299	4086	43755	7.69	82.34
	300-399	2685	46440	5.05	87.39
	400-499	1921	48361	3.62	91.01
	500-749	2695	51056	5.07	96.08
	750-999	1103	52159	2.08	98.16
	1,000-1,499	747	52906	1.41	99.56
	1,500-1,999	166	53072	0.31	99.87
	2,000-4,999	64	53136	0.12	99.99
	5,000+	3	53139	0.01	100.00
Single / E & M	0	14312	14312	26.93	26.93
	1-9	9857	24169	18.55	45.48
	10-19	3441	27610	6.48	51.96
	20-49	5611	33221	10.56	62.52
	50-99	5370	38591	10.11	72.62
	100-199	6051	44642	11.39	84.01
	200-299	3152	47794	5.93	89.94
	300-399	1876	49670	3.53	93.47
	400-499	1188	50858	2.24	95.71
	500-749	1356	52214	2.55	98.26
	750-999	501	52715	0.94	99.20
	1,000-1,499	339	53054	0.64	99.84
	1,500-1,999	60	53114	0.11	99.95
	2,000-4,999	24	53138	0.05	100.00
	5,000+	1	53139	0.00	100.00

Among strategies based on total dollars, the multiple attribution strategy yielded fewer physicians with zero episodes than the single attribution strategy because under the multiple attribution strategies nearly all physicians who touched the episode had the episode attributed to them, whereas under the single attribution strategy only the physician with the most total dollars had the episode attributed to them.

A similar argument holds for strategies based on E & M dollars: under multiple attribution any physician with an E & M visit would be attributed to the episode whereas under single attribution only the physician with the most E & M visits would be attributed to the episode.

The light green bars correspond to physicians with between 1 and 19 episodes attributed to them, and the dark green bars correspond to physicians with 20 or more episodes attributed to them. The percentage of physicians with at least 20 episodes varies by attribution method, shown in Table 5.

Table 5: Percentage of Physicians Attributed at Least 20 Episodes

Attribution method	Percentage of Physicians Attributed at Least 20 Episodes
Multiple attribution with total dollar weighting (M_T and M_{TR})	70.4 %
Single attribution based on total dollars (S_T and S_{TR})	53.9 %
Multiple attribution with E&M dollar weighting (M_E and M_{ER})	55.6 %
Single attribution based on E&M dollars (S_E and S_{ER})	48.0 %

It is important to bear in mind that the database for this study excludes episodes for patients outside the physician's MSA. Therefore, the percentage of physicians who would be attributed at least 20 episodes is in reality higher than that shown in Table 5. However, if we could assume that the relative percentages among attribution methods are about right in Table 5, we would conclude that the multiple attribution strategies allow a higher percentage of physicians to be "reliably" measured compared with single attribution strategies, assuming a minimum threshold of 20 episodes for physician measurement.

Figure 9 shows the percentage of physicians attributed 20 or more episodes, by specialty[2] and by attribution method. In the columns to the right of the graph, "Total MD-Years" is the total number of MD-Years in the specialty group and "Percent >= 20 Episodes" is the percentage of the Total MD-years that have at least 20 episodes under the given attribution methodology. Many MDs are present in both 2002 and 2003. These MDs are counted once for each year (they can be classified < 20 for both years, >= 20 for both years, or < 20 in one year and >= 20 in the other year). The actual number of MDs in either year is a little over half of the total MD-years. There is little difference in the pattern between 2002 and 2003, so we combined the two years of data for these statistics.

Figure 10 shows the mean number of episodes attributed to physicians for each attribution method by specialty group. The means are based on physicians attributed at least 20 episodes

[2] Some specialties were grouped. For example, thoracic surgery, vascular surgery, and cardiac surgery comprise the specialty group "cardiothoracic surgery." Also, several lower-frequency specialties were omitted entirely from Figures 9 and 10, including anesthesiology, osteopathic therapy, nuclear medicine, infectious disease, peripheral vascular disease, preventive medicine, and unknown specialties, among others. These omitted specialties tended to have fewer episodes per physician, explaining why the specialties in Figure 9 tended to have a higher rate of physicians with at least 20 episodes, compared with the overall rates shown in Table 5.

Figure 9 - Percentage of Physicians with at Least 20 Episodes by Specialty and Attribution Method

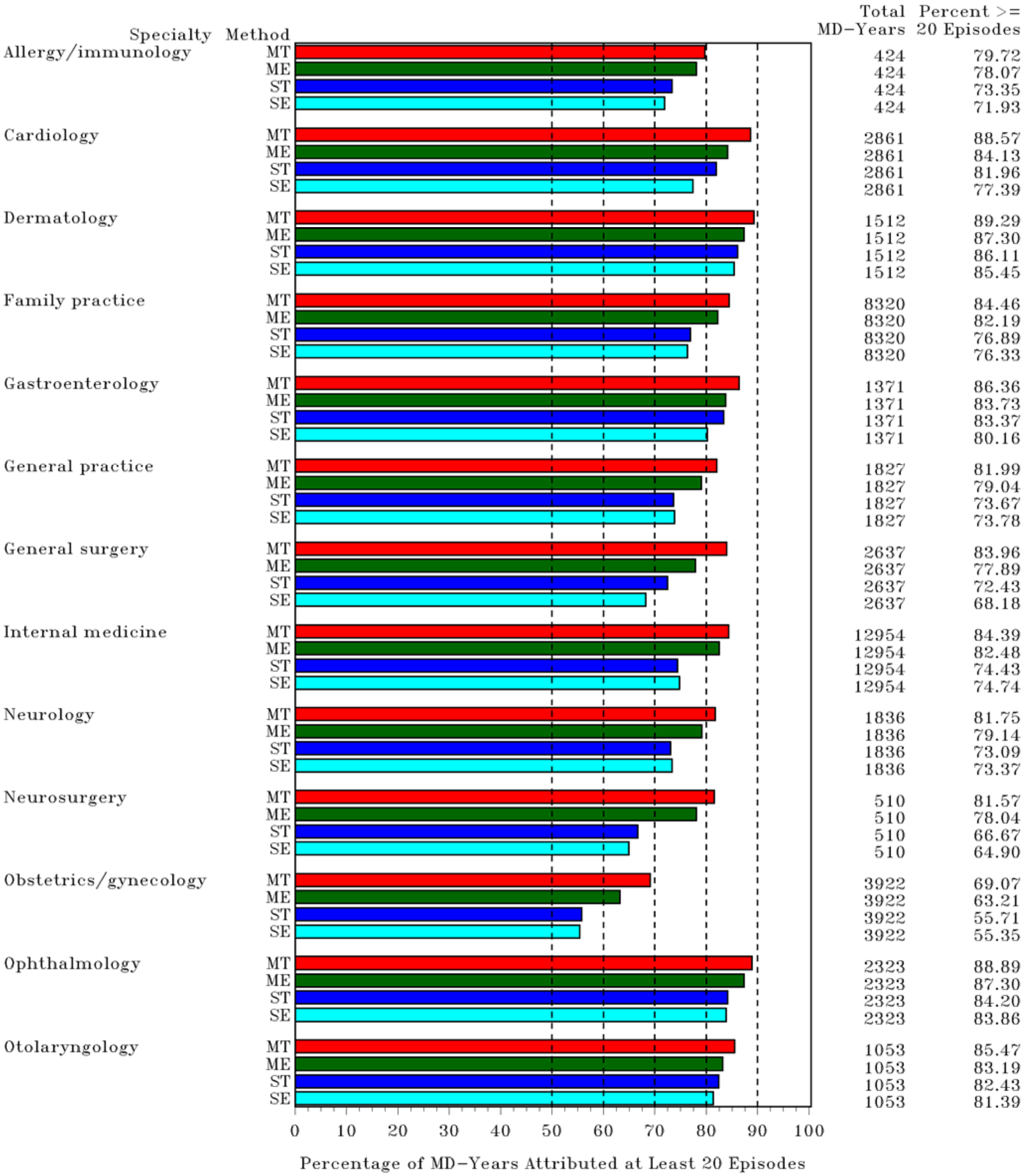
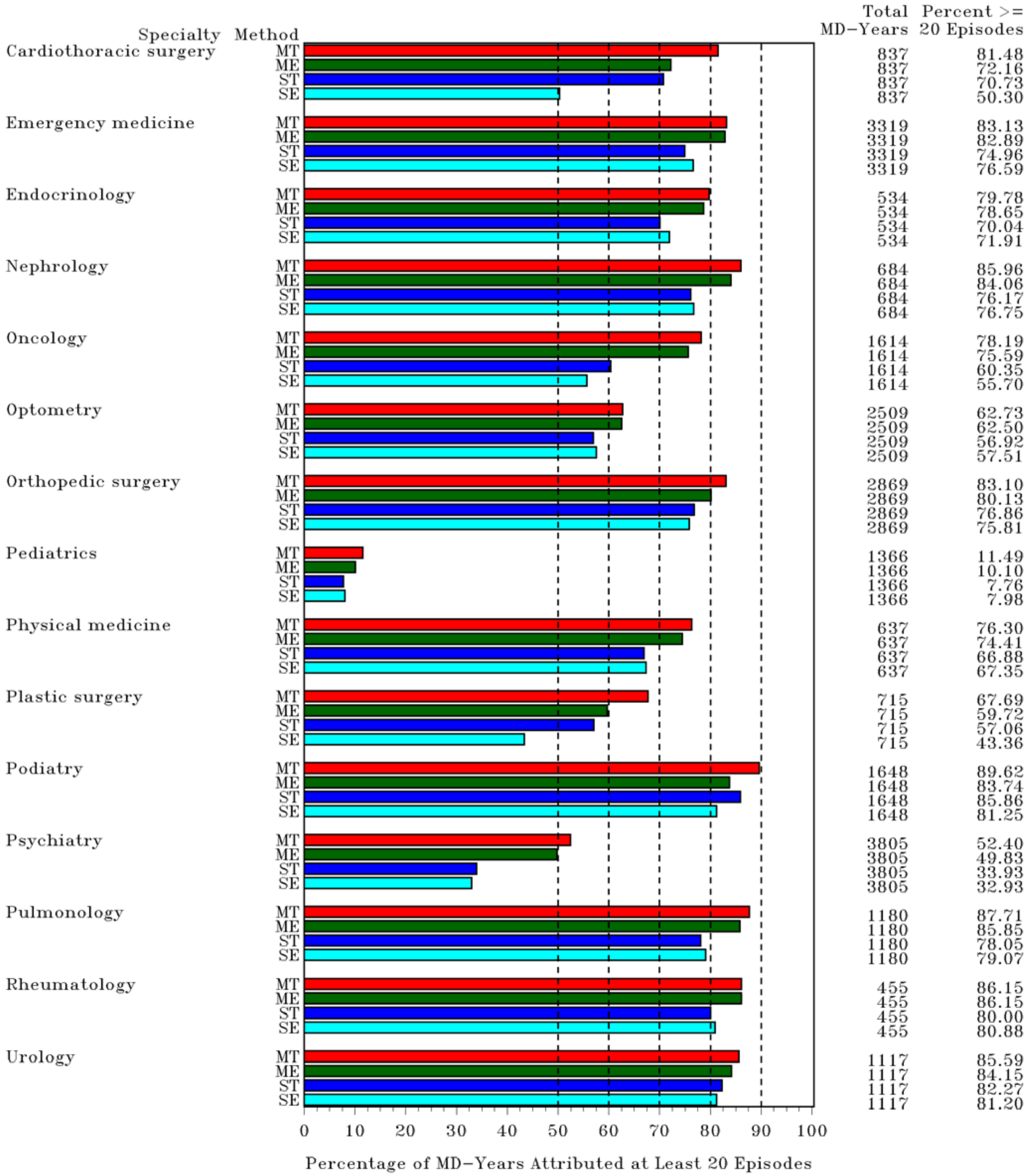


Figure 9 (cont.): Percentage of Physicians with at Least 20 Episodes by Specialty and Attribution Method



under the attribution strategy represented by the bar. Generally, multiple attribution with weights based on total dollars (MT) and multiple attribution with weights based on E&M dollars (ME) tend to result in higher average numbers of episodes per physician compared with the single attribution strategies based on maximum total dollars (ST) and maximum E&M dollars (SE). However, Dermatology and Podiatry are notable exceptions. For those two specialties, more episodes tend to be attributed for single attribution based on maximum total dollars (ST) than for multiple attribution based on E&M dollars (ME), indicating that these specialties can have the most total dollars without having any E&M dollars in the episode.

It is also evident from Figure 10 that the average number of episodes per physician varies by specialty group and by attribution method. While Cardiology has the highest average number of episodes attributed under multiple attribution based on total dollars (MT), the average falls off dramatically for the other attribution methods. In contrast, for Opthalmology the average number of episodes remains high for all attribution methods.

The number of physicians attributed at least 20 episodes is not plotted, but it is shown in the right panel of Figure 10. For example, for Allery/Immunology the number of physicians attributed at least 20 episodes is 338 for MT, 331 for ME, 311 for ST, and 305 for SE. This decreasing pattern is repeated for most other specialties. By construction, the number of physicians must be less for SE than for ME and less for ST than for MT because the single attribution methods attribute an episode to the *single physician with the most* E&M dollars or the *most* total dollars, respectively, whereas the multiple attribution methods attribute an episode to *all physicians with any* E&M dollars or *any* total dollars, respectively.

Given this variability in the average number of episodes among physician specialties, we calculated specialty-specific year to year correlations (a measure of stability) for each index. The results are shown in Figure 11. For each specialty, Figure 11 shows the year-to-year correlation for each of the eight indices (M_E, M_{ER}, M_T, M_{TR}, S_E, S_{ER}, S_T, and S_{TR}). The dark blue bars represent measures based on a the average of ratios (M_E, M_T, S_E, and S_T) and the light blue bars represent measures based on a the ratio of averages (M_{ER}, M_{TR}, S_{ER}, and S_{TR}). The year-to-year correlations tend to be higher for indices based on the average of ratios (dark bars) compared with indices based on the ratio of averages (light bars). That is, average episode-level payment ratios tend to be more stable than dollar-weighted episode-level payment ratios. This is to be expected because the dollar-weighted averages emphasize high-payment episodes, which tend to be more variable than low-payment episodes.

Focusing on the dark blue bars (average of ratios), the correlations within specialties are similar across indices for most specialties. Notable exceptions are for Neurosurgery and Cardiovascular Surgery, where the indices based on total dollars have much higher correlations than the indices based on E&M dollars. Also, still focusing on the dark blue bars, the correlations tend to be higher for the multiple attribution index based on total dollars (M_T) compared with that based on E&M dollars (M_E). Likewise, the correlations tend to be higher for the single attribution index based on total dollars (S_T) compared with that based on E&M dollars (S_E).

Finally, the general magnitudes of the year-to-year correlations vary among specialties. For example, the correlations are 90 percent or higher on all indices for Allergy/Immunology, Dermatology, and Opthalmology. In contrast the correlations are 76 percent or lower on all indices for Neurosurgery, Nephrology, and Emergency Medicine. The reason for this is undoubtedly because some specialties tend to treat episodes with lower payment variance while other specialties tend to treat episodes with higher payment variance.

Figure 10: Mean Episodes per Physician for Each Attribution Method, by Specialty Group

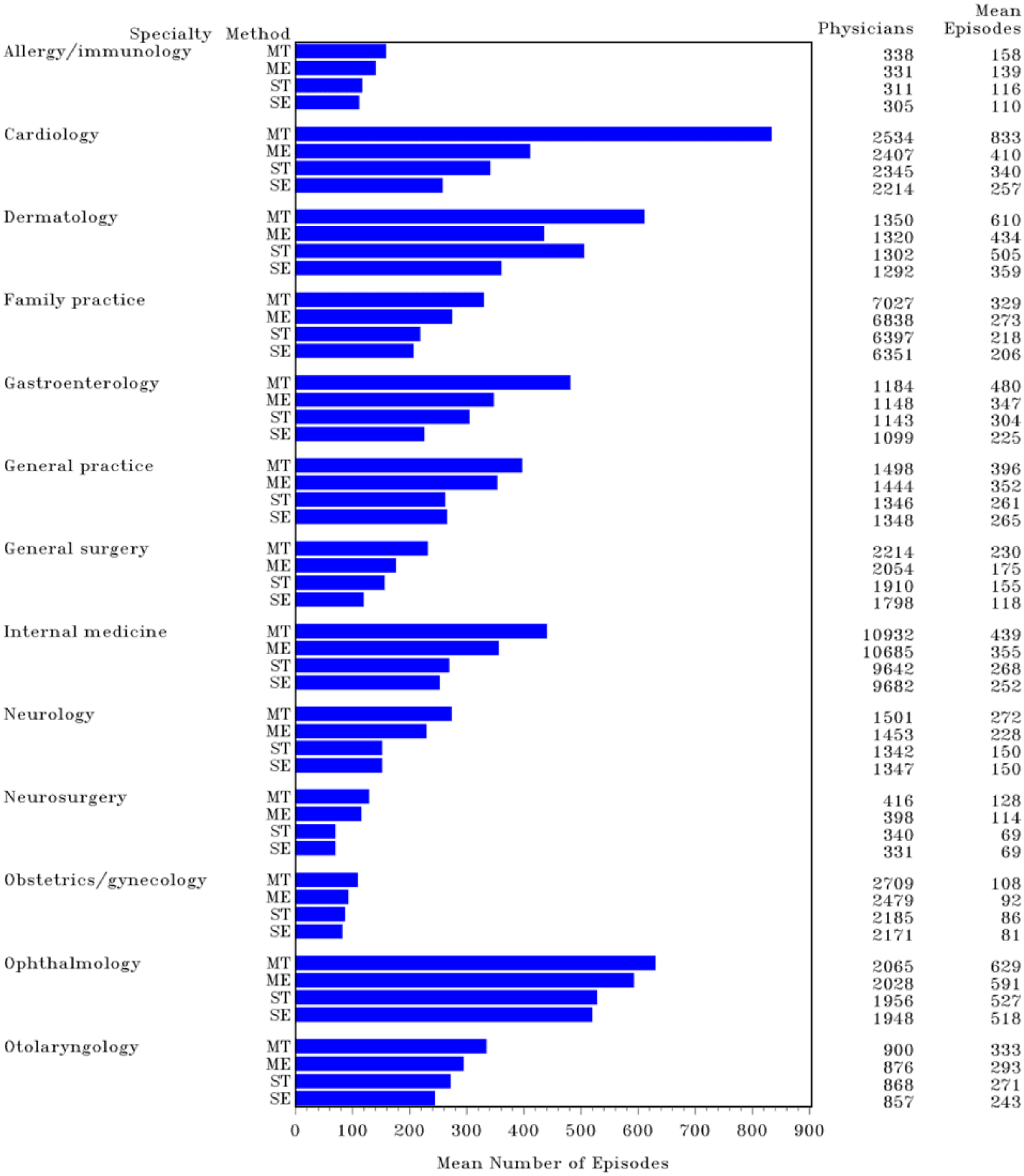


Figure 10 (cont.): Mean Episodes per Physician for Each Attribution Method, by Specialty Group

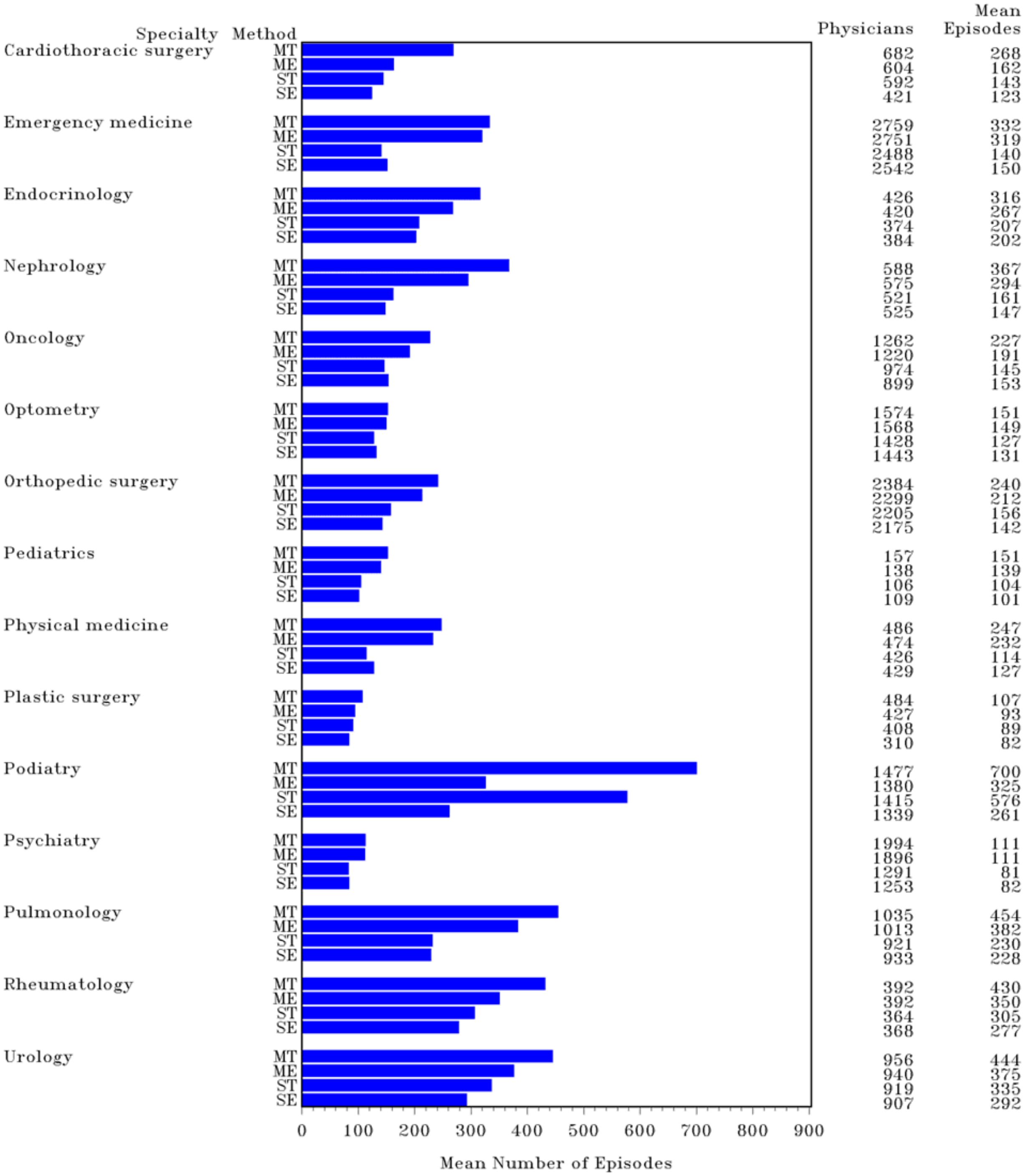


Figure 11: Year-to-Year Correlations for Each Index, by Specialty

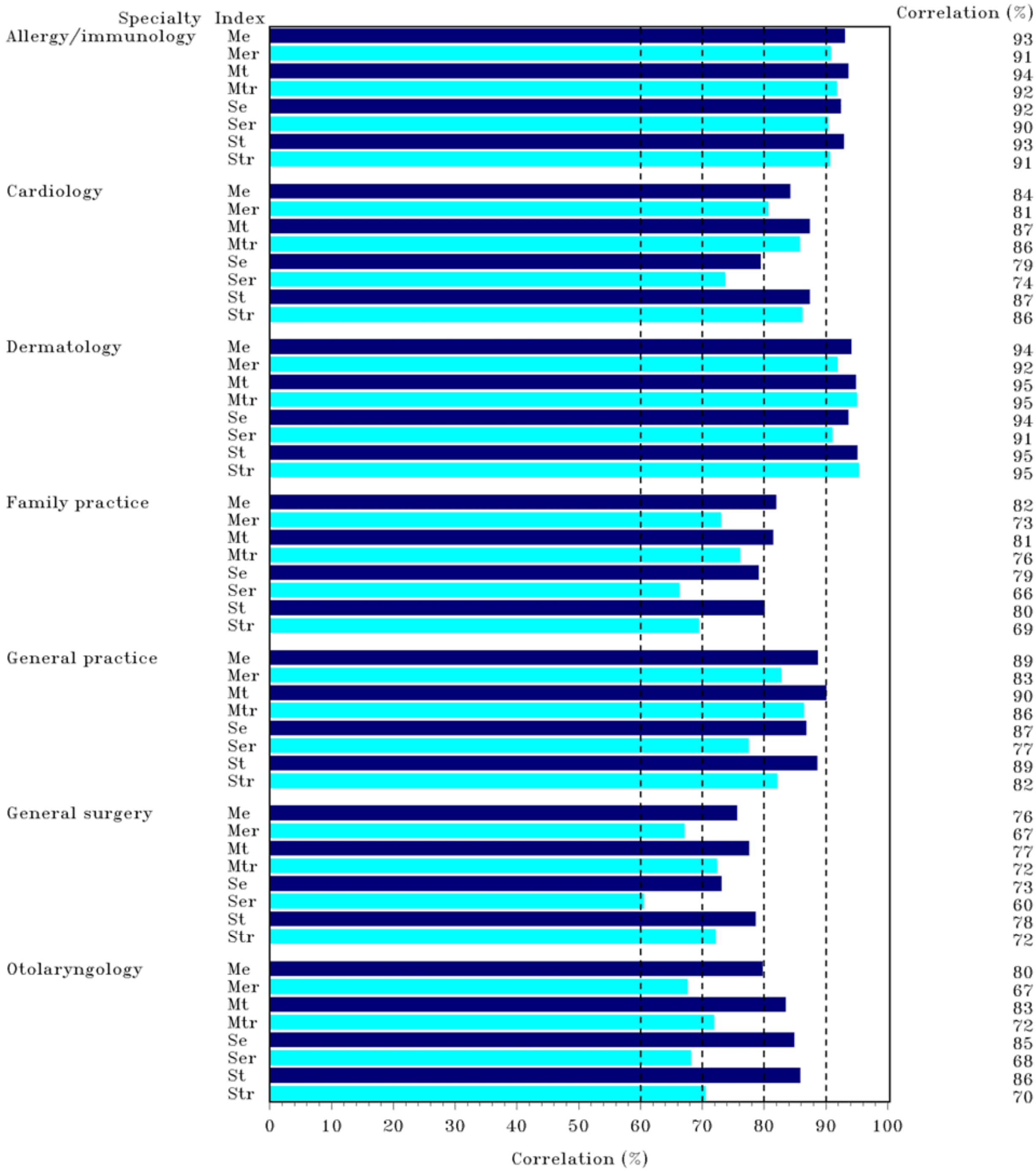


Figure 11 (cont.): Year-to-Year Correlations for Each Index, by Specialty

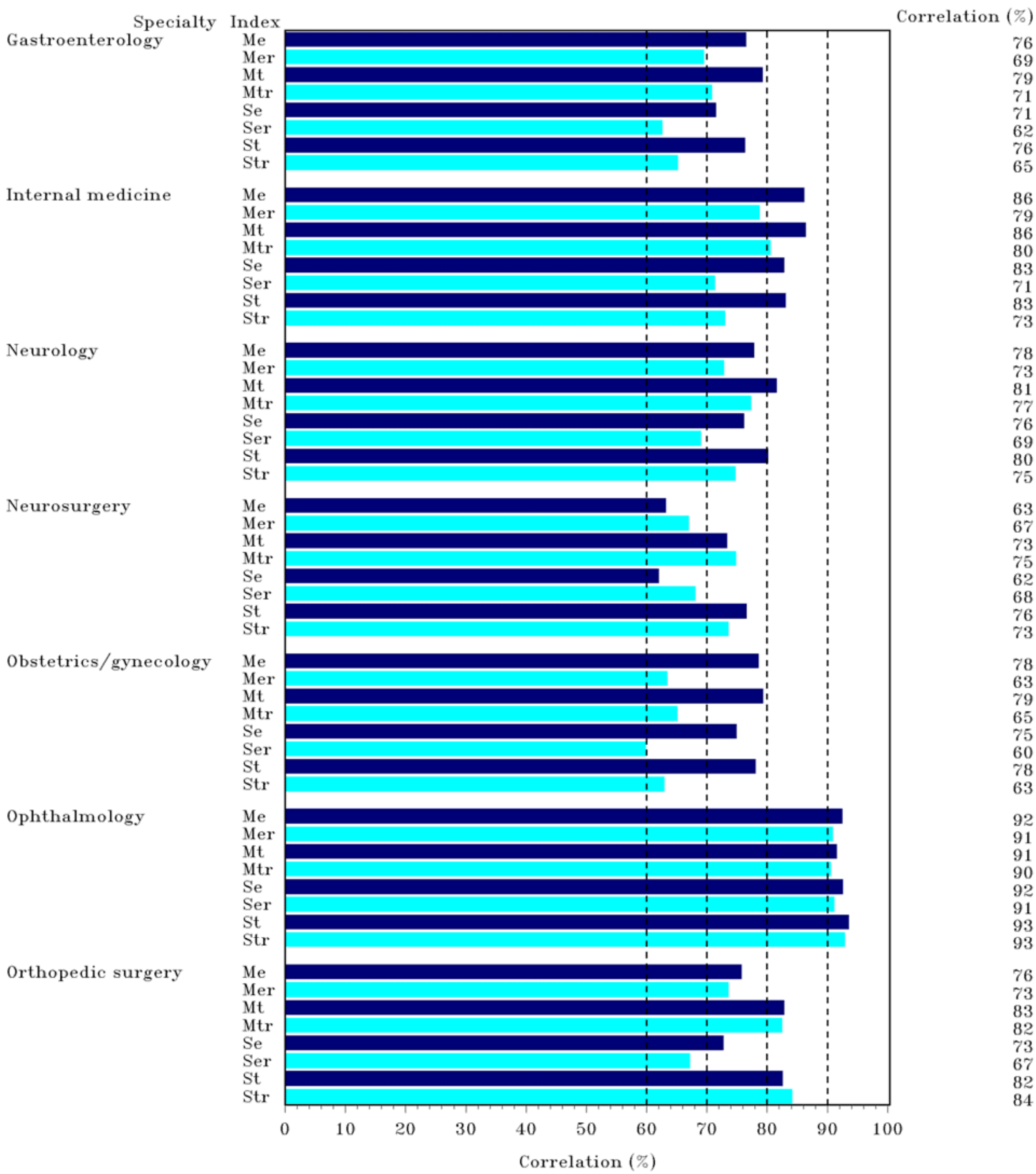


Figure 11 (cont.): Year-to-Year Correlations for Each Index, by Specialty

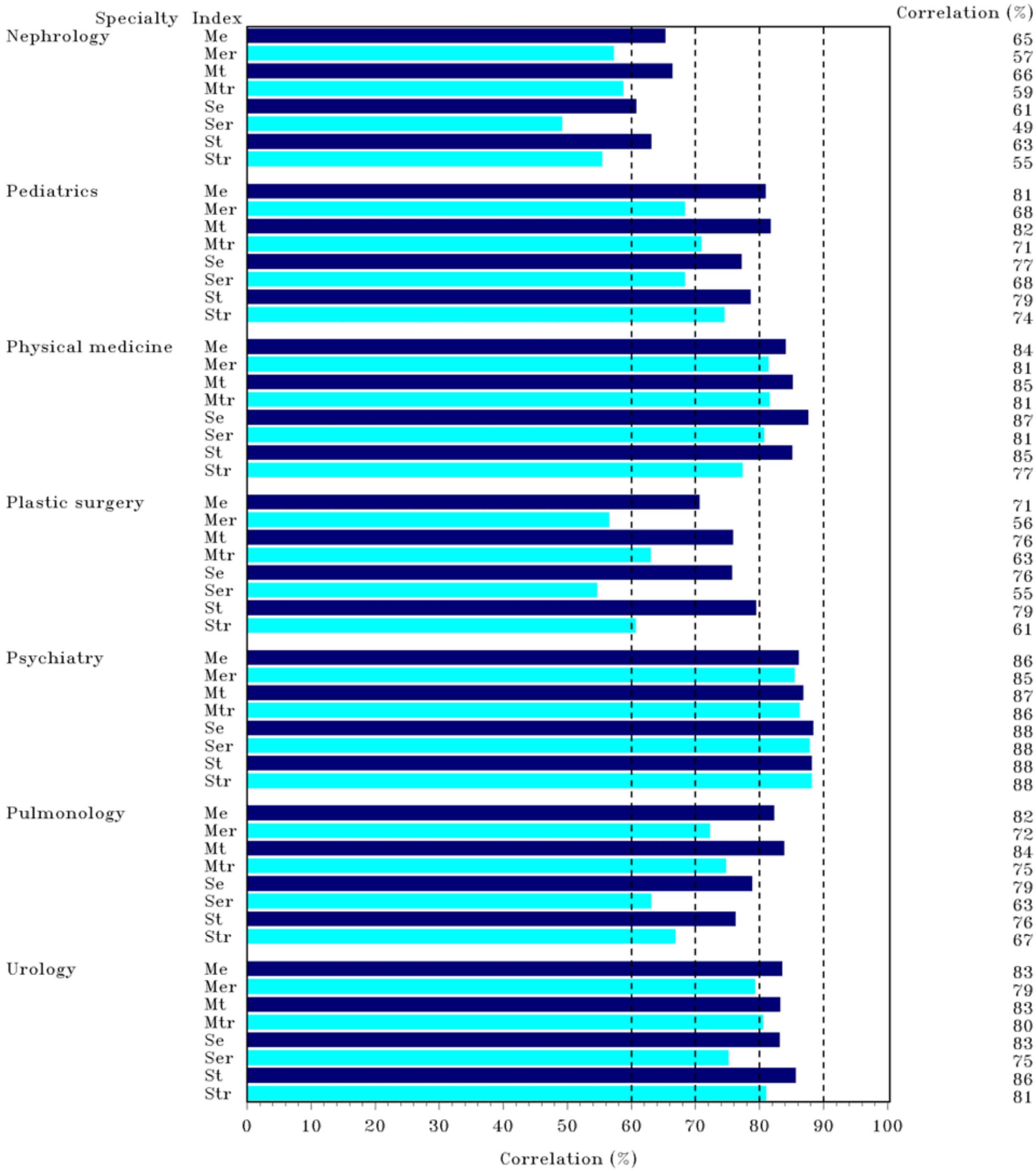
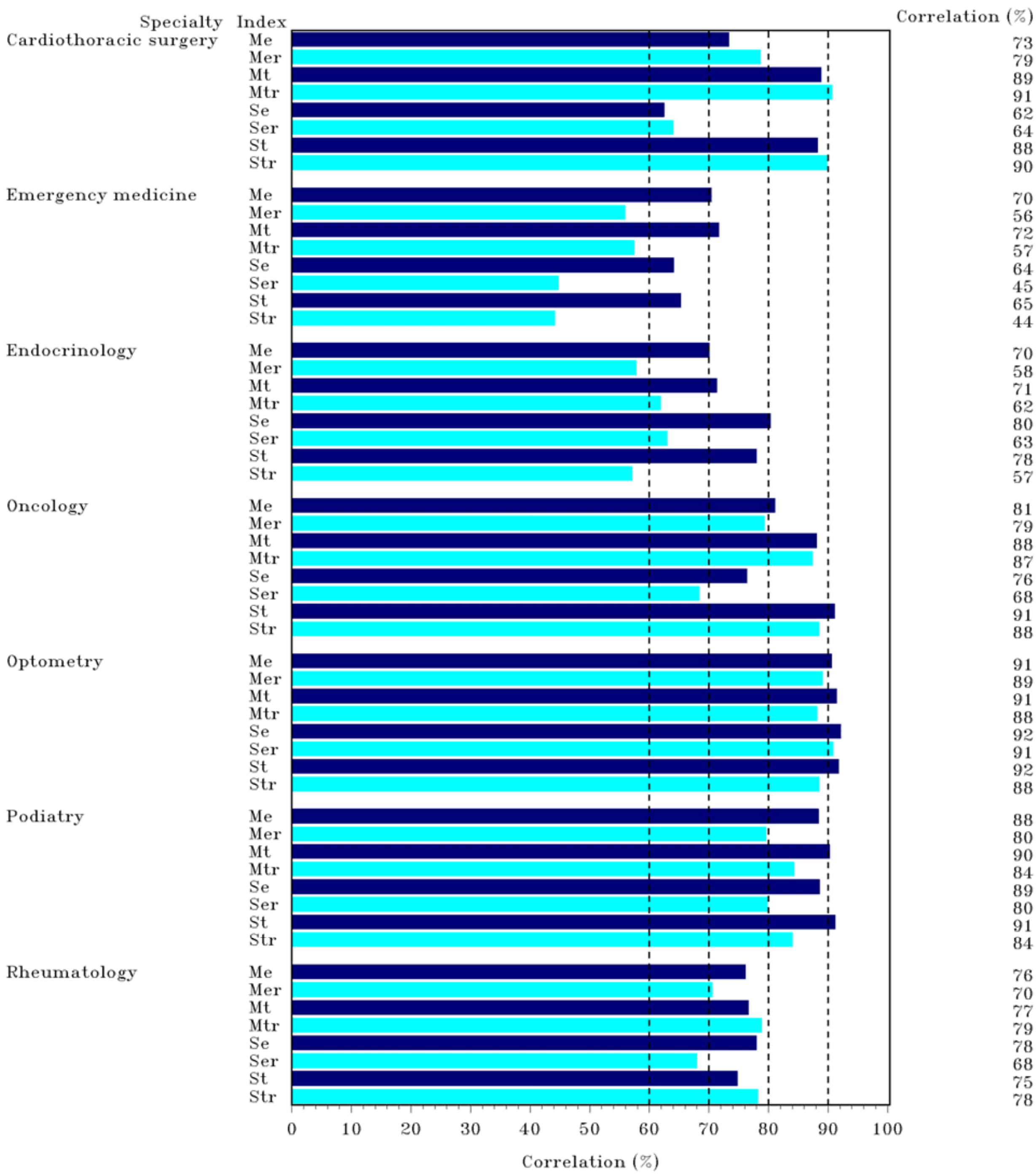


Figure 11 (cont.): Year-to-Year Correlations for Each Index, by Specialty



CONCLUSIONS AND RECOMMENDATIONS

We defined and analyzed eight physician performance indices. Four based on multiple attribution and four based on single attribution strategies. Half of the indices were based on E&M dollars and half were based on total episode dollars. One set of indices was based on the average of episode-level ratios (episode ratio = observed total payment / severity-adjusted expected total payment). Another set was based on the ratio of averages (ratio = mean observed episode total payment / mean severity-adjusted expected total payment).

The indices based on the average of episode-level ratios tended to be more stable—had higher year-to-year correlations—than indices based on the ratio of averages. Moreover, the indices based on total dollars tended to be slightly more stable than indices based on E&M dollars. However, the correlation was high (95 %) between the single attribution index based on maximum total dollars (S_T) and the multiple attribution index based total dollar weights (M_T). Therefore, analysts might have a slight preference for one of these indices (S_T or M_T) over the others, but the choice between them should be based on other considerations.

Therefore, the decision to use single attribution or multiple attribution cannot be made based on the results in this study. Instead, analysts must decide what makes the most sense for the purpose of their study and for the outcomes in their application.

This study has the following limitations:

1. Only eight indices were defined and analyzed in this study. Other indices could have been defined, such as those based on residuals in Houchens et al. (2007).
2. Standardized payments were the basis for measuring episode resource intensity and physician "efficiency." For example, hospital payments were the same for every patient hospitalized with a given diagnosis related group. This standardization no doubt masked some true episode cost variation.
3. The episodes in this analysis were based on the MSA of the patient, not on the MSA of the physician. For example, all episodes for Boston physicians were based solely on patients residing in the Boston MSA. This excluded episodes for patients outside the Boston MSA that were treated by Boston physicians. Even so, this would only affect our results to the extent that episodes for patients outside the MSA were different from episodes inside the MSA for a given physician.

We recommend that MedPAC should expand the analyses in the present study to address the third limitation. If some physicians treated a large number of patients outside their own MSA, then their estimated mean episode payment ratio could have been biased to the extent that those patients had different treatment patterns compared with patients in the physician's own MSA. At the least, the larger sample of episodes could produce a more reliable estimate of their mean episode payment ratios.

APPENDIX

This appendix contains more details concerning the Medical Episode Grouper (MEG), the physician attribution rules, and the application of MEG to the claims data.

MEG – Medical Episode Group™

An episode of care describes a series of related health care services for the treatment of a given spell of illness. Episodes can be comprised of inpatient admissions, outpatient services, and prescription drugs. The Medical Episode Grouper (MEG) was commercially released in 1998.

All episode grouping methods are built on two central concepts; a disease classification system and an episode grouping logic.

Disease Staging is the disease classification system that forms the foundation of MEG episode groups. Disease Staging defines levels of biological severity for specific diseases – episodes of care – where illness severity is defined as the risk of organ failure or death. The severity levels include:

Stage 0: History of or exposure to a disease.

Stage 1: The disease involves no complications

Stage 2: The disease involves local complications

Stage 3: The disease involves multiple sites, or has systemic complications

Stage 4: Death

In the definition of the Disease Staging criteria, most diseases begin at Stage 1 and continue through Stage 4. There are several exceptions to this rule. Some self-limiting diseases, such as cataracts, do not include a Stage 3 or 4. Other criteria begin at either Stage 2 or 3 since they are often complications of other diseases (e.g., bacterial meningitis, which can be a complication of sinusitis, otitis media, or bacterial pneumonia). Stage 0 has also been included in the classification of diseases for patients with a history of a significant predisposing risk factor for the disease, but for whom there is currently no pathology (e.g., history of carcinoma).

The MEG episode grouping logic dictates the accumulation of claims into episode groupings, and allocates claims into discrete episodes of care. The logic employed by MEG includes:

Starting Points - An episode of care is initiated by a contact with the health delivery system and is generally the first claim received for a given disease. The MEG methodology allows physician office visits and hospitalizations to initiate patient episodes. The coding of claims for imaging services and laboratory tests are not always reliable. Therefore, such claims can join existing episodes but they cannot create new episodes.

Clean Periods - Episode Duration – The MEG episode logic is designed to capture all relevant treatments related to a given episode. The end of an episode cannot be directly determined from medical claims. Therefore, episodes are deemed complete when a specified “clean period” has passed without claims related to an episode that has been initiated.

Episode Severity – Episode severity is defined as the highest Disease Staging severity stage observed during the episode.

Multiple Diagnosis Codes – It is often the case that a professional claim (or claim lines) will have two or more diagnosis codes associated with a single procedure code. The episode grouper determines which diagnosis code is most related to the procedure, ensuring the accurate allocation of claims to episode groups.

Lookback – Frequently, tests are ordered and performed before a patient has been diagnosed. Since lab tests and imaging studies cannot initiate an episode, a "lookback" logic links an established episode to these claims if they are clinically related to the episode's disease. If such an episode is found within 15 days of a lab or imaging claim, it is added to the episode.

Inclusion of Non-specific Coding - Non-specific, initial diagnoses are relatively common in the billing of treatments for patients. The inclusion logic is a process that examines each episode after the initial grouping to determine whether a non-specific episode (e.g., Episode Group 179, "Other Gastrointestinal or Abdominal Symptoms") can be included with a clinically related specific episode (e.g., Episode Group 138, "Appendicitis").

Drug Claims - MEG has been designed to 'group' drug claims into episodes. National Drug Code (NDC) information is reviewed by clinical and coding experts and mapped to each episode group. This mapping is then used to assign drug claims to episodes.

Complete Episodes –Episodes of care are created from claims datasets that span a given period of time and are used to profile and evaluate the economic efficiency of physicians. Since it cannot be known whether an episode that was initiated by a claim near the beginning of the dataset is the true beginning of the episode or would join an existing episode created earlier if the data had been available, the true payment for the episode could be understated. Analogously, it cannot be known whether episodes near the end of the dataset would extend beyond this date if more data were available. Episodes which may understate the true payments for treatment are removed from the dataset prior to analysis. The remaining are considered to be complete.

A complete episode is defined to be an episode that begins later than the beginning date of the claims data set plus the number of days of the episode clean period. For example, if a given dataset is comprised of claims occurring on or after January 1 and an episode with a clean period of 30 days begins on January 15, it cannot be known whether this is the true beginning of the episode or if it would have been created in December. In this case, the episode would be considered incomplete and would not used when profiling the physician responsible for the episode.

Based on the episodes created from the four years of medical claims, two study periods – 2002 and 2003 – were established. Episodes with a beginning date in 2002 were assigned to that year. Similarly, the 2003 episode data set was determined. The episodes falling into 2002 and 2003 are complete episodes because a full year of claims data preceded the 2002 data and followed the 2003 data. This ensured that episodes bridging two years would not be eliminated from the study data. Consequently, some 2002 episodes extended into 2003 and some 2003 episodes extended into 2004.

Risk Adjusted Expected Episode Payments

Risk adjusted expected episode payments were calculated for the 2002 and 2003 episodes. *Episode severity* was measured by the stage of disease for that episode. *Patient complexity* was measured by the DCG relative risk scores (RRS) for the patient treated in the episode. The RRS is an estimate of the expected medical payments for a patient based on the patient's age, gender, and the medical conditions for which the patient was treated over a specified period of time, usually one year (Ellis and Ash, 1995; Ash, et al., 2000).

For each MEG, a table was constructed with rows for each integer stage of disease, and with up to 5 columns corresponding to five RRS categories corresponding to consecutive ranges of relative risk score values. For each MEG, the ranges for the five RRS categories were determined by maximizing the variance explained over the entire episode file.

Each cell in the MEG table was then populated with an "expected payment" calculated as the average payment taken over all episodes within the table cell (excluding outlier payments). Finally, these expected episode payments were assigned to each episode based on the episode's MEG disease, the stage of disease, and the patient's RRS. These expected payments were used in the multilevel models.

Summary of Episode Grouping Results

Since lab and imaging claims can only join an existing episode and not create one, there will be claims that cannot be grouped to episodes. Table A1 summarizes the percent of claims and payments that comprised disease-specific episodes – 'grouped' - and those that could not be associated with an episode – 'ungrouped'. In 2002, 96.5 percent of the payments and 84.5 percent of the claims were assigned to episodes. In 2003, 96.6 percent of payments and 85.6 percent of claims were grouped. These results are consistent with earlier episode studies conducted by MedPAC.

Table A 1: Summary of Grouped and Ungrouped Claims, 2002 and 2003.

2002	**Payments**	**Percent**	**Claims**	**Percent**
Ungrouped	$276,264,425	3.5	11,544,787	15.5
Grouped	$7,662,107,877	96.5	62,709,123	84.5
Total 2002	$7,938,372,302	100.0	74,253,910	
2003				
Ungrouped	$301,534,386	3.4	11,391,248	14.4
Grouped	$8,649,500,603	96.6	67,587,729	85.6
Total 2003	$8,951,034,989	100.0	78,978,977	100.0

Outliers were removed from the data set to mitigate the likelihood that a single extreme episode would unduly influence the analysis of a physician's performance. Outliers were defined as episodes with total payments falling below the 1^{st} percentile (low outliers) and above the 99^{th} percentile (high outliers) of episode payments within each integer stage of a MEG. Standardized payments were missing on about 0.7 percent of the claims in the data, and episodes created from these claims were also excluded from this study.

Table A2 and Table A3 display the effects of removing outliers from the 2002 and 2003 episode datasets. In 2002, 7.8 percent of the episodes were eliminated representing 6.1 percent of the claims and 16.0 percent of the payments. In 2003, 8.2 percent of episodes representing 6.1 percent of claims and 16.6 percent of payments were outliers.

Table A 2: Episode Exclusions, 2002.

	Episodes		**Claim Records**		**Payments**	
	Number	**Percent**	**Number**	**Percent**	**Total**	**Percent**
Episodes 1-560	8,455,433	100.0	62,709,123	100.0	$7,662,107,877	100.0
Exclusions	663,010	7.8	3,817,305	6.1	1,228,408,980	16.0
Low Outliers	570,505	6.7	1,112,733	1.8	$9,444,155	0.1
High Outliers	83,956	1.0	2,687,917	4.3	$1,218,964,825	15.9
Missing Pmt.	8,549	0.1	16,655	0.0	Unknown	0.0
Study Episodes	7,792,423	92.2	58,891,818	93.9	$6,433,698,897	84.0

Table A 3: Episode Exclusions, 2003.

	Episodes		Claim Records		Payments	
	Number	**Percent**	**Number**	**Percent**	**Total**	**Percent**
Episodes 1-560	9,011,921	100.0	67,587,729	100.0	$8,649,500,603	100.0
Exclusions	735,523	8.2	4,118,416	6.1	1,438,949,909	16.6
Low Outliers	630,502	7.0	1,181,317	1.7	$10,796,370	0.1
High Outliers	89,441	1.0	2,907,054	4.3	$1,428,153,539	16.5
Missing Pmt.	15,580	0.2	30,045	0.0	Unknown	0.0
Study Episodes	8,276,398	91.8	63,469,313	93.9	$7,210,550,694	83.4

REFERENCES

Houchens, RL, McCracken, S, Marder, W, Kelley, R. The Use of an Episode Grouper For Physician Profiling in Medicare. A Preliminary Investigation. Washington, DC: MedPAC, 2007.

Medicare Payment Advisory Commission. Report to the Congress: Increasing the value of Medicare. Washington, DC: MedPAC, 2006.

Pham, HH, Schrag, D, O'Malley, AS, Wu, B, Bach, PB. Care Patterns in Medicare and Their Implications for Pay for Performance. *New England Journal of Medicine*, 2007, 356:1130-9.