Need for Risk Adjustment in Adapting Episode Grouping Software to Medicare Data

Thomas MaCurdy, Ph.D., Jason Kerwin, and Nick Theobald, Ph.D.

Episode grouper software offers a potential framework for developing important components of a pay-for-performance system for healthcare providers. If the costs for treating health conditions can be computed, then policymakers can in principle benchmark different providers' cost distributions and reward the most efficient. This article applies two of the most prominent commercial groupers and examines the properties of the cost distributions calculated for their constructed episodes. The analysis reveals that episode cost distributions exhibit substantial variation and skewness, suggesting the need for innovative risk adjustment methods prior to utilizing groupers for the purpose of physician profiling.

INTRODUCTION

Pay for Performance (P4P) can be broadly defined as "any type of performance-based provider payment arrangements including those that target performance on cost measures" (Dudley and Rosenthal, 2006). One technology suggested by many in the policy and health-care communities as offering a framework for measuring physician performance is episode grouping software. This software allots health-care claims (e.g., hospital inpatient, physician, post-acute care,

etc.) into episodes of care in order to identify service patterns and their associated costs. In principle, episodes produced by groupers provide a data source for profiling the resource utilization of individual physicians benchmarked against episode-based standards, such as the mean cost per episode, or against a composite performance measure that integrates rankings across types of episodes within relevant peer groups.

The Medicare Payment Advisory Commission (MedPAC) examined episode grouping software and concluded on a preliminary basis, that these products have face validity from a clinical perspective, seem to be able to identify practice patterns, and possess risk adjustment capabilities that could account for differences in disease severity and the presence of co-morbidities (MedPAC, 2005, 2006, 2007). The Senate Finance Committee has also made optimistic statements about the potential of episode grouper software: "Ultimately, episode groupers could give providers and payers more specific, actionable information that could lead to meaningful reductions in inappropriate care patterns. Medicare should develop its own open-source technology platform that includes information on both episodes of care and per-capita resource use (Senate Finance Committee, 2008)." A U.S. Government Accountability Office (GAO) report focusing on physicians' practice patterns recommends that "CMS develop a system that identifies individual physicians with inefficient

practice patterns and, seeking legislative changes as necessary, uses the results to improve the efficiency of care financed by Medicare." According to the GAO, CMS has tools available to evaluate physicians' practices for efficiency and could implement these tools in ways similar to other purchasers (GAO, 2007) if given the appropriate authority.

Health plans and insurers have been experimenting with episode grouping software products to manage medical costs and monitor physician performance (Lake, 2007). Little independent research exists on using episode groupers for measuring physician performance, and none on using them with Medicare data. The Institute of Medicine (IOM) states that "numerous challenges must be faced in the development, implementation, and ongoing evaluation of performance measures... Multiple methodological considerations—risk adjustment reflecting patient populations of varying acuity, small sample sizes at the individual practitioner level, ... and attribution of responsibility among multiple providers ... have already been identified as high priority areas for further research..." (IOM, 2006). Previous research has questioned both the adequacy of episode measurement relying upon diagnosis codes as the fundamental link between claims and the assumptions necessary to attribute episodes to providers (Beckman, 2007). Physician responses to early grouping efforts have ranged from positive to skeptical, depending upon the degree of familiarity with the software (Lake, 2007).

A fundamental issue determining the usability of grouping software concerns its capacity to generate episodes of care that constitute coherent units of analysis comparable across providers. Ideally, each episode should be constructed in a way that exhibits cost homogeneity

sensitive only to the decisions made by the providers deemed responsible for the care. Measuring performance based upon non-homogeneous episode classifications would be akin to measuring the properties of fruit without distinguishing between apples and oranges. Using Medicare data, this article examines the cost variability obtained within episode categories constructed using the Episode Treatment Groups (ETG) and Medical Episode Grouper (MEG) groupers, two leading commercial episode products. The central research question addressed in this analysis concerns whether the construction of episodes by groupers yields sufficient cost homogeneity to make comparisons feasible across providers; or are additional steps in risk adjustment needed prior to introducing grouper processes into candidate P4P systems under consideration for Medicare?

BACKGROUND ON EPISODE GROUPERS

Most episode groupers are proprietary software designed to assign raw medical claims into sets of clinically coherent episodes. This study presents findings constructed from two major commercial episode groupers: Episode Treatment Groups (ETG), developed by Ingenix; and Medical Episode Grouper (MEG), developed by Thomson/Reuters. Each of these products incorporates over two decades of development and refinements by their respective owners.

The grouping algorithms of ETG and MEG share many similarities. Both groupers can build episodes of care using all contacts that a beneficiary has with the health care system over a fixed period of time. Diagnosis codes appearing on claims primarily drive the grouping process, with procedure codes also used in a variety of

circumstances. While the actual grouping process occurs in an opaque component of the software, the output produced by groupers depends on users' decisions regarding the claim types included in the processing, the information on the claims selected for input, and the time periods specified. The payment amounts appearing on claims do not play any role in grouping algorithms, but typically this information is used in post-grouping analyses to assign costs to episodes.

The grouping software classifies episodes as being chronic, acute, or preventive. Whereas acute and preventive episodes invariably have clearly defined start and end dates, chronic episodes typically do not because they often reflect health conditions that began before the study period, became progressively worse, and continued after. Consequently, both groupers invoke administrative rules to define the duration of chronic episodes, with the lengths of chronic care episodes truncated into fixed twelve month intervals with the calendar year being the most common interval. Acute and preventive episodes are generally much shorter in duration, with the majority lasting only a single day (MaCurdy et al., 2008). Many patients experience more than one type of episode of care at the same time.

Despite similarities in their fundamental approaches, ETG and MEG rely on distinct schemes for classifying health conditions and rules for claims assignment which sharply limits opportunities for direct comparison of their results. Version 7.0.1 of the ETG grouper includes a total of 524 base ETG classes (of which 68 do not count as identifiable health conditions since they are categorized as "ungroupable"). ETG further refines the classification of 129 base ETGs into up to 4 severity levels, which yields a total of 679 distinct episode classifications. The ETG vendor recommends using these episode-severity categories as its list of episode types. Version 7.1 of the MEG grouper assigns each episode to one of 560 base MEG disease classifications. In addition, MEG can allot up to 4 "disease stages" to a base MEG episode, with stage 1 representing the lowest level of health complication and stage 4 being death. Distinguishing base MEGs by their disease stages implies thousands of classifications. To avoid a proliferation of episode categories, the MEG vendor recommends using the base MEGs as its list of episode types. Rarely can a MEG episode type be matched directly to an ETG episode type, nor can MEG episode types be grouped in a fashion that makes them directly comparable to groups of ETG episode types. Moreover, an episode depicted as chronic in one grouper may be considered acute in the other grouper. As a result, the episode duration and costs may greatly differ because the groupers define health conditions/treatments differently.

METHODS AND DATA

Our analysis applies the ETG and MEG groupers to the Medicare claims of a 20% random sample of beneficiaries residing in Colorado, with data from all claim types for the years 2002, 2003 and 2004 included in the sample. The beneficiaries had to be continuously enrolled in fee-for-service (FFS) Parts A and B services while alive in the 2002-04 period. Claim types include the following: hospital in-patient (IP), outpatient facility (OP), Part B services (PB), durable medical equipment (DME), skilled nursing facility (SNF), home health (HH), and hospice (HS).

Table 1 shows summary statistics for the sample for both ETG and MEG. Between 2002 and 2004 Medicare paid $585.5 million for 5.05 million claims on behalf of the 20% Colorado beneficiary

sample. The ETG grouper created 672,600 episodes leaving 15% of claims and 5% of costs ungrouped; the MEG grouper produced 661,053 episodes with 23% of claims and 8% of costs left ungrouped. Within the sample, chronic ETG episodes are 23% more costly on average than chronic MEG episodes. Conversely, acute episodes produced by ETG are 27% less costly than acute episodes in MEG.

To produce the findings presented in Table 1, one requires a framework for reporting grouper output using common measures. One encounters certain challenges in applying the ETG and MEG software in a Medicare setting to create episodes of care and in assigning costs to these episodes. Whereas some challenges are specific to the individual software packages, others are common to both groupers. The following discussion initially describes the design features and implementation steps involved in running each grouper with Medicare data, then describes a framework for presenting the results produced by the two groupers using common metrics for measuring episode lengths and costs.

Application of the ETG Grouper to Medicare Data

For each claim, the ETG grouper reviews the diagnosis code(s), procedure code(s), revenue center code(s), provider category, and type of service to determine a classification of the claim. Only claims representing clinical interactions (e.g., office visit, a surgery, or an admission to a hospital or SNF) are allowed to open an episode and these are called "anchor records." An episode is opened by the earliest record that qualifies as an anchor record for an episode. Anchor records are created through the following service or claim types: Evaluation & Management (E&M HCPCS codes), surgery (procedural HCPCS codes), or facility (generally

Table 1
Summary Statistics for Claims, Episodes, and Costs

Statistic	ETG	MEG
Total # Claims	5,049,696	
% Ungrouped	15%	23%
Total # Episodes	672,600	661,053
% Chronic Episodes	50%	40%
% Acute Episodes	50%	60%
Average # per beneficiary	6	6
Total Cost of Claims	$585,447,839	
% Cost of Chronic Episodes	65%	43%
% Cost of Acute Episodes	30%	48%
% Cost of Ungrouped Claims	5%	8%
Chronic Episodes		
Average Cost per Episode	$1,071	$871
Average Length of Episode (days)	113	123
Acute Episodes		
Average Cost per Episode	$498	$690
Average Length of Episode (days)	22	24

SOURCE: ETG and MEG Output, Medicare Claims for Random Sample of 20% of Colorado FFS Parts A and B Beneficiaries, 2002-2004.

room-and-board revenue codes). Claims other than anchor records, such as claims for ancillary services (lab tests, imaging, etc.), are incidental to direct evaluation, management, or treatment of a patient and cannot open an episode. The episode ends when a sufficiently long clean period follows the last claim assigned to the episode, with the clean periods defined by each grouper for each episode type.

The ETG software inputs each claim as a set of service-level records composed of the revenue center and procedure codes on the claim, with each record individually assigned to an episode. For institutional claims, each input record, which we term as a "pseudo-claim," consists of a single revenue center code identifying a form of service, an accompanying procedure code if available, and the diagnoses listed on the parent claim. An institutional claim has as many input records as it has revenue center codes. Medicare institutional claims for IP, SNF, and HH often have more than four diagnosis codes. Since ETG allows only four diagnosis codes per input record, the user must choose which four codes to use. Whereas revenue center codes are universally reported on all institutional Medicare claims, HCPCS/CPT procedure codes—which often reveal more details about the form of service—are rarely available on IP, SNF, and HS claims; in contrast, these procedure codes usually accompany revenue center codes on OP and HH claims. Of the pseudo-claims built from IP claims, 2% have a HCPCS or CPT code; 7% of pseudo-claims from SNF claims have a HCPCS/CPT code; 6% of input records from HS claims have a procedure code; and 72% of pseudo-claims from OP claims and 88% from HH claims have a HCPCS/CPT code.

If a user enters institutional claims as multiple records using all of the service codes available on Medicare claims, the ETG grouper can and often does assign the separate input records from a single parent claim to different episodes. Such assignments result in institutional claims being linked to more than one episode. In the application considered here, over 52% of SNF claims are split across episodes, as are 23% of IP claims, 40% of HH claims, 13% of OP claims, and 15% of HS claims.

For non-institutional services, Medicare's PB and DME claims are readily separated into line items associated with individual HCPCS or CPT codes; these claim types have no revenue center codes. Each input record constructed from a PB and DME claim consists of a single procedure code and its corresponding line-item diagnosis.

Application of the MEG Grouper to Medicare Data

MEG does not refer to the term "anchor records" to open an episode, but rather uses physician services (visits and procedures), hospitalizations (IP), SNF stays, HH, and hospice services (HS) to initiate an episode. As with the ETG software, the MEG grouper ends an episode when a sufficiently long clean period follows the last claim assigned to the episode.

The MEG grouping process inputs each claim as a single record, relying primarily on diagnosis information in its assignments to episodes. Regardless of whether a Medicare claim comes from an institutional or non-institutional source, the MEG grouper accepts one input record per claim. This record distinguishes IP and PB claims from other types of Medicare claims, but it does not differentiate among the other distinct types of Medicare claims as the source of diagnoses. Switching from one of these claims types to another results in no change in constructed episodes. An input record accepts

data on procedure codes appearing on the claim (not revenue center codes). This procedure information is primarily used to determine whether a claim represents an x-ray/lab event—which cannot start an episode—and in some instances to assist the grouper in deciding how to interpret secondary diagnoses on the claim. Medicare institutional claims allow for up to 9 secondary diagnoses, and MEG allows for all these to be included on input records. Typically, however, only the primary diagnosis is used in the grouping process, although the secondary diagnoses can affect the assignment to the stages.

The MEG grouper does not offer the capacity to treat a claim as an aggregate of services potentially linkable to more than one episode. The prospective payment system used by Medicare not only compensates based on diagnoses but also on procedures and the likelihood of various co-morbidities. MEG's inability to associate the services on claims paid under such a system with more than one episode constitutes a potential challenge in applying this software to a Medicare setting.

Common Measures for Comparing Grouper Outputs

To present the results produced by the two groupers on a level playing field, our analysis relies on a framework developed in MaCurdy et al. (2008) for computing episode lengths and costs using comparable rules. This approach exploits the fact that both groupers map claims to episodes, making it possible to see, claim by claim, to which episode the claim was assigned. MaCurdy et al. (2008) also describes the precise settings and options selected in running the ETG and MEG groupers.

To measure an episode's length, the framework sets its start date as the earliest service date of all the claims grouped into the episode, and its end date as the latest service date of the grouped claims. ETG uses "anchor" claims—claims that represent a clinical interaction—to open episodes. These anchor claims need not even be grouped to the episode. MEG calculates episode end dates as the start date of the last claim in the episode, rather than the end date, with the exception of IP claims for which the end date is used. For the majority of acute episodes, our length assignments for episodes match those of both ETG and MEG. In the case of chronic episodes, however, both groupers set chronic episode lengths to a fixed 12-month period, usually a calendar year. Our calculations of episode lengths cover the period beginning with the date of the first claim and finishing with date of the last claim grouped into the chronic episode, so our measurements of lengths invariably fall short of 12 months.

Neither grouper suggests a method for calculating episode costs; this calculation is left entirely to the user. To measure an episode's cost, this analysis aggregates the cost of claims assigned to the episode, with a claim's expense composed of its Medicare payments, excluding the capital payment portion of IP claims, pass-thru payments, and deductibles and copayments made by beneficiaries. In the case of the MEG, where each claim is assigned to only one episode, this exercise is straightforward. When the ETG algorithm links the services from a single parent institutional claim to different episodes, we allocate the cost of the claim to the episode that was assigned the plurality of the claim's service-level input records.[1] For example, if a claim has ten pseudo-claims, and six are assigned to one episode, while four are linked to another, then the first episode would be allocated the entire

---

[1] This cost allocation rule is arbitrary; the ETG software offers no recommendations.

cost of the original claim and the second episode would receive nothing. In the case of ties, the cost of the parent institutional claim is evenly divided between the two episodes.[2]

FINDINGS

Substantial cost variation exists within episode types, regardless of the grouping software used to build episodes.

Distributions of Costs for Most-Expensive Episode Types

Table 2 presents statistics depicting the distributions of per-episode costs for the 5-highest-cost acute ETGs—highest aggregate costs of episode types—the top-5 chronic ETGs, and for all ETGs combined reported at the bottom of the table. These statistics show the mean and standard deviation for the sample of 2003 complete episodes, along with several percentiles of the distributions to convey the extent of variation in costs across episodes within ETG classifications. The column following the standard deviation lists the coefficient of variation (CV). The final two columns show the share of costs captured by the lowest-cost 50% and the highest-cost 5% of episodes within each ETG classification.

Inspection of the findings in Table 2 reveals that distributions exhibit substantial dispersion in costs across episodes within each ETG, with distributions highly skewed for the highest cost episodes. For each of the top five highest-cost acute and chronic ETGs, the level of costs demarking the most expensive 10% of episodes (i.e., the 90th percentile) always exceeds the level demarking the cheapest 10% (i.e., the 10th percentile) by more than a factor of four, and in many instances by more than

_____
[2] In this application, ties occur for 2.9% of institutional claims and 1.1% of institutional costs.

two orders of magnitude. The table further reveals that the top 5% of episodes in an acute ETG accounts for 15.4% to 41.7% of all costs in that ETG. In comparison, the bottom 50% of episodes account for only 2.9% to 20.3% of costs. For each chronic ETG, the top 5% of episodes account for 26% to 50% of the costs of that ETG, and the bottom 50% only cover 2.6% to 7.7% of costs. Considering the total costs across all ETGs included in the software and Colorado sample (which comprises 642 ETGs), the last row shows that 67.4% of these costs are incurred by the most expensive 5% of episodes, while a mere 2.7% are incurred by the cheapest 50%, displaying an exceptionally high degree of skewness.

Table 3 reports analogous statistics for the top-5-cost acute MEGs, the top-5 chronic MEGs, and for all MEGs combined reported at the bottom of the table. As for ETG results, the cost distributions in this table exhibit very large variation across episodes within individual MEGs with distributions highly skewed toward the highest cost episodes. For each of the top five highest-cost acute and chronic MEGs, the level of costs for the 90th percentile episode always exceeds the level for the 10th percentile episode by at least an order of magnitude, and in most instances it is more than two orders of magnitude larger. Even for pneumonia, which is one of the less extreme cases, the cheapest 10% of episodes cost $40 or less, while the most expensive 10% of episodes cost $7,332 or more per episode. Additionally, the highest-cost episodes account for between 25.2% and 63.8% of costs, whereas the bottom half only capture between 1.0% and 9.3% of total costs. When considering the distribution of costs across all MEGs appearing in our Colorado sample (which includes 463 MEGs), we see nearly an identical high level of skewness, with 67.3% of costs

Table 2

Cost Distributions of Top 5 Acute and Chronic ETGs by Total Cost

ETG: Description	2003 Complete Episodes								Fraction of Cost in Bottom 50% of Episodes of this ETG	Fraction of Cost in Top 5% of Episodes of this ETG
	Summary Statistics									
	10%	50%	90%	95%	98%	Mean	Std Dev	CV		
Top 5 Acute ETGs by Cost										
713103L2: Closed fracture or dislocation - thigh, hip & pelvis, SL2	$64	$10,190	$27,704	$31,914	$42,912	$12,088	$12,064	1.00	9.0%	18.1%
437400L4: Bacterial lung infections, SL4	$67	$2,507	$10,747	$14,871	$21,011	$4,338	$6,914	1.59	5.2%	29.4%
713103L3: Closed fracture or dislocation - thigh, hip & pelvis, SL3	$273	$16,089	$30,411	$36,682	$49,557	$15,438	$12,053	0.78	19.1%	15.4%
475600L1: Non-malignant neoplasm of intestines & abdomen, SL1	$223	$589	$1,040	$1,508	$5,413	$927	$2,329	2.51	20.3%	41.7%
316500L3: Spinal trauma, SL3	$55	$1,036	$13,905	$18,737	$26,773	$4,533	$6,657	1.47	2.9%	27.9%
Top 5 Chronic ETGs by Cost										
386500L2: Ischemic heart disease, SL2	$60	$872	$9,883	$17,264	$30,851	$3,483	$7,388	2.12	4.0%	43.6%
386500L1: Ischemic heart disease, SL1	$40	$349	$3,441	$9,917	$15,333	$1,704	$4,202	2.47	3.6%	50.4%
386500L3: Ischemic heart disease, SL3	$268	$6,354	$28,311	$35,689	$51,477	$11,313	$16,080	1.42	7.7%	25.5%
351700L1: Cataract, SL1	$0[1]	$72	$1,599	$2,829	$3,051	$425	$834	1.96	7.0%	37.0%
316000L2: Cerebral vascular accident, SL2	$40	$487	$7,977	$14,786	$28,202	$3,057	$7,268	2.38	2.6%	47.1%
All Chronic and Acute ETGs	$15	$97	$1,272	$3,220	$8,875	$794	$3,296	4.15	2.7%	67.4%

1 The costs for the claims for these episodes do not exceed deductibles, so there are no Medicare payments.

SOURCE: ETG Output, Medicare Claims for Random Sample of 20% of Colorado FFS Parts A and B Beneficiaries, 2002-2004.

Table 3

Cost Distributions of Top 5 Acute and Chronic MEGs by Total Cost

| MEG: Description | 2003 Complete Episodes | | | | | | | | Fraction of Cost in Bottom 50% of Episodes of this MEG | Fraction of Cost in Top 5% of Episodes of this MEG |
| | Summary Statistics | | | | | | | | | |
	10%	50%	90%	95%	98%	Mean	Std Dev	CV		
Top 5 Acute MEGs by Cost										
11: Acute Myocardial Infarction	$63	$7,215	$28,532	$35,641	$48,047	$11,039	$14,979	1.36	8.1%	25.2%
92: Cataract	$0[1]	$72	$1,476	$1,820	$2,898	$404	$739	1.83	5.1%	34.9%
426: Complications of Surgical and Medical Care	$40	$628	$13,650	$19,811	$29,329	$4,426	$7,806	1.76	1.8%	34.7%
510: Pneumonia: Bacterial	$40	$348	$7,332	$11,885	$17,752	$2,870	$5,385	1.88	1.9%	35.5%
397: Cerebrovascular Dis with Stroke	$35	$247	$10,760	$19,074	$33,273	$3,760	$9,592	2.55	1.0%	48.4%
Top 5 Chronic MEGs by Cost										
374: Osteoarthritis	$40	$361	$9,485	$14,405	$20,901	$2,300	$5,314	2.31	2.8%	46.3%
10: Angina Pectoris, Chronic Maintenance	$44	$285	$4,066	$10,493	$19,495	$1,830	$4,830	2.64	3.0%	53.5%
500: Chronic Obstructive Pulmonary Disease	$43	$527	$4,595	$6,209	$9,693	$1,861	$4,008	2.15	4.2%	34.6%
430: Encounter for Preventive Health Services	$15	$43	$226	$454	$1,277	$182	$977	5.37	5.3%	63.8%
13: Essential Hypertension	$33	$121	$441	$809	$2,890	$331	$1,383	4.18	9.3%	55.8%
All Acute and Chronic MEGs	$18	$92	$1,263	$3,286	$8,648	$764	$3,081	4.03	2.7%	67.3%

[1] The costs for the claims for these episodes do not exceed deductibles, so there are no Medicare payments.

SOURCE: MEG Output, All Medicare Claims for Random Sample of 20% of Colorado FFS Parts A and B Beneficiaries, 2002–2004.

captured by the top 5 percent of MEGs and 2.7% captured by the bottom 50 percent.

Distributions of Costs for Common Health Conditions

To compare the extent of cost variation within episode types for the ETG and MEG groupers, one needs to select similar health conditions. One sees in Tables 2 and 3 that bacterial lung infection/pneumonia episodes appear in both, so we select this as our first condition. Table 4 presents the cost distributions for all bacterial lung/pneumonia classifications for each of the groupers. For a second condition, two episode categories of hip fractures show up in Table 2, and a similar clinical condition/treatment would show up in Table 3 had it included the sixth highest episode type by cost for MEG. So, we select all categories of hip fractures as a second health condition for comparisons across groupers, and Table 5 presents the cost distributions for these episode types. Tables 4 and 5 present findings for all episode types linked to these two health conditions. Following the ETG vendor's advice regarding distinguishing ETGs by their severity levels, both tables report the four severity-level classifications for bacterial lung infections and the three levels for hip fractures. To provide a benchmark combining all severity levels, the tables also report findings for the base ETGs associated with these health conditions. Though MEG utilizes disease stages to differentiate episodes by severity, the vendor for this grouper does not recommend presenting results at this level of disaggregation. So, Tables 4 and 5 present cost distributions for only the base MEG associated with pneumonia and hip fractures.

The top rows of Table 4 show the distributions of costs for the ETG episode types linked to bacterial lung infections, with the base episode type listed first followed by cost distributions for the ETGs distinguished by severity levels. The bottom row lists the cost distribution for the MEG episode type associated with pneumonia. Comparing distributions for the base ETG and the MEG reveals similar dispersion; costs from the 10th to 95th percentile range from $58 to $9,307 for ETG compared to $40 to $11,885 for MEG.

Regardless of the episode types considered, one sees heavily skewed cost distributions, with about a third of costs accounted for by the top 5%. Disaggregating costs by ETG severity levels sometimes lowers and sometimes raises cost variation within types, but it does little to reduce the extent of skewness in costs. Although average costs for the ETGs increase with severity level—from $712 for SL1 episodes to $4,338 for SL4 episodes—extensive cost variation remains within the severity-level ETGs, and each distribution shows over a quarter of costs captured by the top 5% of episodes and less than 6% of costs captured by the bottom 50% (for most episode types, more than half have costs fall below $700).

Table 5 presents a parallel structure for presenting costs for hip fracture episodes. Comparing the cost distribution for the base ETG against the MEG distribution reveals similar properties. Both show substantial variation with extensive skewness toward the highest-expense events. Whereas episodes at the 10th percentile cost about $65, episodes at the 95th percentile reach costs around $30,000, nearly 450 times larger. Average costs do increase across ETG categories with the higher severity levels, but the cost distribution for each individual ETG possesses similar properties to that obtained for the base ETG. The degree of skewness in the upper tails of the distributions is smaller than that seen in Table 4, but still implies

Table 4

Distributions of Episode Costs for Pneumonia for ETG and MEG

2003 Complete Episodes

Bacterial Lung/Pneumonia Episode Types		Cost per Episode								Fraction of Cost in Bottom 50% of Episodes	Fraction of Cost in Top 5% of Episodes
		10%	50%	90%	95%	98%	Mean	Std Dev	CV		
ETG Episodes											
437400:	Bacterial lung infections, base	$58	$446	$5,964	$9,307	$13,787	$2,421	$4,547	1.88	3.2%	33.5%
4374001:	Bacterial lung infections, SL1	$40	$160	$2,296	$4,309	$6,105	$712	$1,550	2.18	5.6%	45.1%
4374002:	Bacterial lung infections, SL2	$48	$296	$4,769	$5,607	$9,999	$1,630	$2,922	1.79	3.6%	32.1%
4374003:	Bacterial lung infections, SL3	$76	$671	$5,491	$7,640	$11,155	$2,408	$3,127	1.30	4.9%	25.4%
4374004:	Bacterial lung infections, SL4	$67	$2,507	$10,747	$14,871	$21,011	$4,338	$6,914	1.59	5.2%	29.4%
MEG Episodes											
510:	Pneumonia: Bacterial	$40	$348	$7,332	$11,885	$17,752	$2,870	$5,385	1.88	1.9%	35.5%

SOURCE: ETG and MEG Output, Medicare Claims for Random Sample of 20% of Colorado FFS Parts A and B Beneficiaries, 2002-2004.

Table 5

Distributions of Episode Costs for Hip Fractures for ETG and MEG

2003 Complete Episodes

Hip Fracture Episode Types		Cost per Episode								Fraction of Cost in Bottom 50% of Episodes	Fraction of Cost in Top 5% of Episodes
		10%	50%	90%	95%	98%	Mean	Std Dev	CV		
ETG Episodes											
713103:	Closed fracture or dislocation – thigh, hip & pelvis, base	$65	$9,872	$27,704	$31,033	$42,866	$11,742	$11,633	0.99	9.7%	17.8%
7131031:	Closed fracture or dislocation – thigh, hip & pelvis, SL1	$63	$3,871	$20,049	$24,446	$28,874	$7,384	$8,423	1.14	4.8%	22.3%
7131032:	Closed fracture or dislocation – thigh, hip & pelvis, SL2	$64	$10,190	$27,704	$31,914	$42,912	$12,088	$12,064	1.00	9.0%	18.1%
7131033:	Closed fracture or dislocation – thigh, hip & pelvis, SL3	$273	$16,089	$30,411	$36,682	$49,557	$15,438	$12,053	0.78	19.1%	15.4%
MEG Episodes											
348:	Fracture: Femur, Head or Neck	$66	$9,378	$26,015	$29,796	$36,261	$11,351	$11,259	0.99	10.0%	16.8%

SOURCE: ETG and MEG Output, Medicare Claims for Random Sample of 20% of Colorado FFS Parts A and B Beneficiaries, 2002-2004.

significant allocation of total costs to the upper part of the distribution. Also noteworthy, Table 5 reveals skewness at the low-expense end of the distribution. The fraction of costs represented by the bottom 50% of the distribution ranges from 4.8% to 19.1% and the fraction in the top 5% ranges from 15.4% to 22.3%. For all but the highest severity-level ETGs, ten percent of episodes cost far less than $100. Not seen in this table, over twenty-five percent of episodes cost far less than $400.

DISCUSSION

The very high degree of skewness and variation exhibited in the distributions of costs within episode types means that devising a reliable method of risk (or severity) adjustment for episode costs will be an essential ingredient in any integration of "off-the-shelf" episode grouping software into a profiling system for physicians in Medicare. Given the sheer size of the skewness, one questions the homogeneity of cost measures within episode groups. A provider assigned even one of the high-cost episodes could be assessed as inefficient, regardless of this individual's cost ranking on other attributed episodes. The profiling issue becomes whether the assessed providers control the primary circumstances responsible for the highest-cost episodes, a situation unlikely to be true without considerable adjustments being made to reflect costs associated with various preexisting health risk factors. Using one of the prominent commercial grouping products without more refined risk or severity adjustments could readily yield a performance system that improperly rewards providers for factors or behavior beyond their control.

Both the ETG and MEG groupers incorporate features intended to compensate for differences in risk circumstances across episodes, but the availability of these features do not alter the main findings of this article. Principally, these features split base episode types into "severity" gradations, which expands the number of categories for classifying episodes. A category is formulated to include episodes with homogenous characteristics and costs measures. The results presented in this article already account for these severity distinctions in the case of ETGs, and it is in this context we see that substantial variation and skewness in costs remains within episode types. The MEG grouper can also distinguish disease stages in its episode constructions. The above results do not report cost distributions for MEG episodes broken down by stages. Such a construction creates thousands of MEG-staging episode types which leads to smaller cell sizes, especially when entertaining attributions on a per-provider basis. Moreover, we have discovered in our other work that a high variation and skewness in costs remains even within these MEG-staging episode types. The cost distributions of episodes obtained for Medicare populations strongly suggest the need for additional risk adjustment within ETG-severity and MEG-stage categories to create homogeneous cost measures.

The ETG and MEG groupers also have supplementary risk modules that can be applied in conjunction with the groupers to evaluate the actuarial cost of beneficiaries of the sort used to calculate health insurance premiums. The Episode Risk Group (ERG) module for ETG relies on episodes of care as markers of risk to summarize a patient's underlying condition. A total risk score is calculated for a beneficiary by summing the risk weights assigned to the ERGs and demographic brackets in the person's profile. MEG instead recommends adding a DCG (Diagnostic Cost

Grouper) module to estimate a beneficiary's expected costs over fixed horizons, which relies on a person's health conditions and demographic characteristics as predictors of costs. Both the ERG and DCG modules function very much like the HCC risk model used by CMS to risk adjust premiums for Medicare Advantage plans, although neither ERG nor DCG is benchmarked to Medicare beneficiaries relevant for CMS as the payer. In any case, these modules are not primarily intended to risk adjust individual episodes within or across episode types.

A substantial challenge in risk adjusting episodes in Medicare populations concerns the complexity of the medical circumstances experienced by many beneficiaries. Numerous beneficiaries have multiple co-morbidities that jointly determine a patient's health status and resulting receipt of care, and beneficiaries who look quite similar from the perspective of their services can have different underlying causal conditions. These complexities make the task of allocating individual medical services or claims to a single category of care a significant problem. Such a task requires distinguishing which particular health condition constitutes the ultimate source of the provision of each service represented by a Medicare claim. Attributing services to distinct illnesses under these circumstances often becomes a bewildering quandary.

Another formidable challenge concerns the capability of grouping algorithms to emulate familiar practice patterns observed in treating Medicare beneficiaries. For risk adjustment to be credible and for a grouper to work well within a Medicare setting, it would be advantageous for constructed episodes to capture existing practice protocols and payment regimens. The ETG and MEG grouper algorithms are not designed to follow all the service flows expected under Medicare's program rules, and findings presented in MaCurdy et al. (2008) reveal that these algorithms can indeed perform poorly in mirroring some of the practice patterns seen in Medicare data. With this disconnect, practitioners whose costs may be profiled by a grouper would not have a logical framework for interpreting results.

Adding workable risk adjustment features to episode groupers constitutes only one of a number of major steps required to adapt these software products to profile providers in Medicare. Adaptations must also address the thorny problem of attributing episode costs to individual providers, an especially formidable task for those Medicare beneficiaries experiencing multiple co-morbidities and for those episodes made up of an extensive mixture of medical services. One must further resolve the problem of how to rate providers' resource utilization within episode types, and how to combine such ratings across different types of health conditions to obtain overall scores for individual providers. Neither grouper product offers approaches for carrying out any of these tasks; the needed procedures are left entirely to development by the user. To date, CMS has just begun the work of exploring the viability of using outputs from off-the-shelf grouper software to characterize episodes of care applicable for Medicare providers. Regardless of whether CMS decides to use such outputs or develops measures of its own, the next steps of incorporating additional risk-adjustment methodologies, attribution rules, and scoring protocols applicable for Medicare will undoubtedly require innovative approaches not yet available in either commercial software packages or the existing literature.

REFERENCES

Beckman, H., Mahoney, T., and Greene, R.: *Current Approaches to Improving the Value of Care: A Physician's Perspective.* The Commonwealth Fund. December 2007. Internet address: http://www.commonwealthfund.org/Content/Publications/Fund-Reports/2007/Dec/Current-Approaches-to-Improving-the-Value-of-Care--A-Physicians-Perspective.aspx#citation (Accessed 2009.)

Dudley, R.A. and Rosenthal, B.: *Pay for Performance: A Decision Guide for Purchasers.* Agency for Healthcare Research and Quality. April 2006. Internet address: http://www.ahrq.gov/qual/p4pguide.pdf (Accessed 2009.)

Ingenix: *Symmetry 7.0 Episode Treatment Group Concept Guide.* Ingenix. Eden Prairie, Minnesota. 2007.

Institute of Medicine of the National Academies: *Rewarding Provider Performance: Aligning Incentives in Medicare.* The National Academies. September 2006. Internet address: http://www.iom.edu/CMS/3809/19805/37232.aspx (Accessed 2009.)

Lake, T., Colby, M., and Peterson, S.: *Health Plans' Use of Physician Resource Use and Quality Measures (Submitted to the Medicare Payment Advisory Commission).* Mathematica Policy Research, Inc. October 2007. Internet address: http://www.medpac.gov/documents/6355%20MedPAC%20Final%20Report%20with%20Appendices%201-24-08.pdf (Accessed 2009.)

MaCurdy, T., Kerwin, J., Gibbs, J., et al.: *Evaluating the Functionality of the Symmetry ETG and Medstat MEG Software in Forming Episodes of Care Using Medicare Data (Submitted to the Centers for Medicare and Medicaid Services).* Acumen, LLC. August 2008. Internet address: http://www.cms.hhs.gov/Reports/downloads/MaCurdy.pdf (Accessed 2009.)

Medicare Payment Advisory Commission: *Report to the Congress: Increasing the Value of Medicare Chapter 1: Using Episode Groupers to Assess Physician Resource Use.* The Medicare Payment Advisory Commission. June 2006. Internet address: http://www.medpac.gov/publications/congressional_reports/Jun06_Ch01.pdf (Accessed 2009.)

Medicare Payment Advisory Commission: *Report to the Congress: Medicare Payment Policy.* The Medicare Payment Advisory Commission. March 2005. Internet address: http://www.medpac.gov/publications/congressional_reports/Mar05_EntireReport.pdf (Accessed 2009.)

Medicare Payment Advisory Commission: *Report to the Congress: Medicare Payment Policy.* The Medicare Payment Advisory Commission. March 2007. Internet address: http://www.medpac.gov/documents/Mar07_EntireReport.pdf (Accessed 2009.)

Senate Finance Committee: *Call to Action Health Reform 2009.* Senate Finance Committee. November 2008. Internet address: http://finance.senate.gov/healthreform2009/finalwhitepaper.pdf (Accessed 2009.)

U.S. Government Accountability Office: *Medicare: Focus on Physician Practice Patterns Can Lead to Greater Program Efficiency.* The U.S. Government Accountability Office. April 2007. Internet address: http://www.gao.gov/new.items/d07307.pdf (Accessed 2009.)