An Episode-Based Framework for Analyzing Health Care Expenditures: An Application of Reward Renewal Models

E. Michael Foster and Fengjuan Xuan

Objective. To illustrate an episode-based framework for analyzing health care expenditures based on reward renewal models, a stochastic process used in engineering for describing processes that cycle on and off with “rewards” (or costs) occurring at the end of each cycle.

Data Sources/Study Setting. Data used in the illustration were collected as part of an evaluation of a national initiative to improve mental health services for children and youth. Participants were enrolled in a longitudinal study at a demonstration site and in a comparison community between 1997 and 1999. The illustration involves analyses of mental health expenditures at the two sites and of the dynamics of service use behind those expenditures.

Data Collection/Extraction Methods. Services data were derived from management information systems as well as patient records at inpatient facilities in the two communities. These data cover services received between 1997 and 2003. The analysis focuses on the year following study entry.

Principal Findings. Between-site differences in expenditures reflect complex between-site differences in the timing of service use. In particular, children at the demonstration stayed in treatment longer but were less likely to return for treatment later. In contrast, children at the comparison site experienced substantially less continuity of care. Costs per day of treatment within an episode were comparable at the two sites.

Conclusions. Reward renewal models offer a promising means for integrating research on service episodes and the dynamics of service use with that on health care expenditures.

Key Words. Episodes of service use, health care expenditures

Many questions in health services research revolve around the issue of timing. This issue arises in research on treatment onset or termination as well as transitions between treatment settings. Measures of treatment timing are

frequently used as indicators of quality or as part of treatment guidelines. For example, several elements of the Health Plan Employer Data and Information Set (HEDIS) involve timing. For example, one indicator tracks whether or not individuals with mental illness discharged from hospitals receive follow-up care within 30 days. HEDIS also tracks whether individuals being treated with antidepressants remain on those medications continuously during the acute phase of their illness (Druss and Rosenheck 1997; Druss et al. 2002; Busch, Leslie, and Rosenheck 2004). Treatment guidelines also refer to timing issues. For example, the American Academy of Pediatrics recommends treatment plans for children with attention problems so that they receive timely medication management and monitoring (Bauchner 2000; Herrerias, Perrin, and Stein 2001; Stein and Perrin 2003; Leslie et al. 2004; Rushton, Fant, and Clark 2004).

The issue of timing is central to analysis of treatment episodes—"a series of temporally contiguous health care services related to treatment of a given spell of illness or provided in response to a specific request by the patient or other relevant entity" (Hornbrook, Hurtado, and Johnson 1985). Researchers have long been interested in the dynamics of treatment episodes. This research involves their length and the factors that influence their beginning and end (e.g., Foster 1998; Goldman et al. 1998).

In contrast, analyses of costs or expenditures¹ often focus on a fixed time period, such as a calendar or fiscal year.² This strategy makes sense for many budgetary or accounting purposes, such as predicting expenditures during a given budget period. However, analyzing the data in this way discards a great deal of information about the dynamics of expenditures. Two individuals with the same total expenditures in a period may differ significantly in the timing of those expenditures. One individual may have large expenditures in a single treatment episode occurring early in the year. For another expenditures may be spread throughout the year across a series of treatment episodes (Bondy et al. 2000). As noted above, these patterns may have a variety of implications for the quality of care and long-term patient outcomes.

This variation in the timing of expenditures—and service use—may have a variety of implications concerning access to and quality of care as well as treatment outcomes. An episodic perspective highlights that expenditures are shaped by several processes. These include the timing with which episodes

Address correspondence to E. Michael Foster, Ph.D., School of Public Health, University of North Carolina, Rosenau Hall, CB# 7400, Chapel Hill, NC 27599-7400. Fengjuan Xuan, Ph.D., is with the Department of Statistics, Pennsylvania State University, University Park, PA.

start and stop as well as the magnitude of expenditures while in treatment. All three may be shaped by different forces. Whether an individual begins treatment may depend on system-level factors, such as the availability of services. How long he or she remains in treatment may depend on other characteristics, such as transportation. Expenditures per day while in treatment may depend on treatment setting and the efficiency of the particular facility or provider delivering services. As discussed below, this multitude of processes has important implications for predicting expenditures as well. In particular, a model that predicts expenditures in a period assuming a single process will often perform poorly. (This possibility is quite striking in light of the difficulties health economists have had in predicting expenditures on mental health services for risk-adjustment purposes [Ettner et al. 1998, 2001; Kapur, Young, and Murata 2000].) Furthermore, the estimated parameters of that model will be difficult to interpret—they will capture a blend of the parameters of the underlying models.

In this paper, we use a class of stochastic processes known as reward renewal models as a means of understanding the dynamics of service use underlying expenditures in a given period. These models are frequently used in engineering to describe any process that cycles on and off. When the process cycles off, a cost occurs, or an expenditure is made. Engineers use these models to understand the cost implications of various strategies for machine maintenance.

While the terms “broken,” “repaired,” and “costs of repairs” are regrettable in the context of illness, the reward renewal model has many potential benefits for health services research. Like the two-part model (Duan et al. 1983, 1984), these models can be used to decompose expenditure differences into underlying choices or behaviors. The reward renewal model takes the two-part model one step further: it allows one to decompose the likelihood of being in treatment into the choices to enter and leave treatment. These behaviors may be well worth distinguishing as they may have different determinants and may be sensitive to alternative interventions to influence them.

This paper outlines reward renewal models and illustrates their application using data from a study of children’s mental health services. That study focuses on how improved mental health services (as delivered at a demonstration site) affect mental health expenditures relative to those at a comparison site and explains that difference in terms of the underlying processes shaping episodes of service use. The article also illustrates a multiprocess estimation strategy in which individual unobserved heterogeneity is shared across the three processes (starting treatment, stopping treatment, and costs

per day). This multiprocess model allows us to relax some of the assumptions embedded in the simpler framework.

This paper has three sections. The first briefly reviews prior research on episodes of service use and outlines the basic reward renewal model. The second describes a multiprocess, multilevel generalization of the simple reward renewal model. The third presents an empirical illustration of the method. This paper concludes with a discussion.

PRIOR RESEARCH

Two areas of prior research are relevant—that on services episodes and that on the reward renewal model.

The Use of Episodes in Health Services Research

The notion of a treatment episode can be operationalized in several ways.³ One possibility is to tie episodes to the course of treatment for a specific illness. For example, one might define an episode as beginning with initial treatment for an upper respiratory track infection and ending with the termination of related treatment (Mainous and Hueston 1998). The episode may involve related treatment from different providers. For example, one might include ambulance services in analyses of episodes of treatment in an emergency department (Dean et al. 2001).

Identifying episodes involves two key issues. The first involves identifying the end of an episode. Doing so implicitly involves linking-related services; services that are not linked represent the end of one episode and the beginning of the next. Ascertaining such linkages, however, is difficult, potentially requiring reviews of patient records or interviews with providers. For this reason, researchers also have defined episodes in terms of spells of continuous treatment; the end of the episode is marked by a “clean period” in which services (and/or medications) are not received (Rosen et al. 1998). Such periods might involve interruptions in treatment longer than some threshold, such as 45, 56, or 90 days (Kessler, Steinwachs, and Hankin 1980; Wells et al. 1996; Holmes and Deb 1998; Busch 2002). Services separated by a clean period are identified as separate episodes. This strategy has been used widely in health services research to examine the length of episodes, the time between episodes, medication compliance, and other aspects of service use (Williams et al. 1999; Mundt et al. 2001; Rost et al. 2001; Herings and Erkens 2003; Arria 2003).

A second key issue involves the use of diagnosis in defining episodes. One might limit an episode to services with the same diagnosis, but this strategy poses its own problems, especially in the case where the only diagnostic information is derived from services data (such as insurance claims).⁴ Diagnoses of this sort can be problematic. For example, the diagnosis assigned on a claim may change as the individual remains in treatment. Such a change may reflect an initial “tentative” (i.e., incorrect) diagnosis; the individual’s actual condition may become apparent as he or she moves through treatment (Wingert et al. 1995/6). For example, a patient may present with symptoms suggesting an ulcer; further treatment may reveal that the presence of an anxiety disorder. The change in diagnosis may mark the ending of one episode and the start of another even though the underlying condition did not change.

In the empirical application below, episodes are defined using clean periods of 45 days. As all services provided involve treatment for a mental disorder, we do not use diagnosis in defining episodes. It is worth noting that the methods used below could be applied to treatment episodes regardless of how they are defined.

Reward Renewal Models

Renewal processes represent a class of continuous time Markov models in which an event occurs repeatedly (Gertsbakh 1989, 2000). Key features of this process include the number of times an event has occurred ($N(t)$) by time (t), the amount of time between the i th event and the one that preceded it (X_i), and the time required for n events to occur (S_n). These random variables are all related in that

$$N(t) = \{\max n : S_n = X_1 + X_2 + X_3 + \cdots + X_n \leq t\} \quad (1)$$

Assuming a distribution for the X_i , one can identify key characteristics of the process. For example, if we assume the X_i are independent and exponentially distributed with a parameter of λ (the failure rate), then the expected length of any given interval between events is $1/\lambda$. The expected number of events in a given time period, $m(t)$, equals λt . $m(t)$ is often identified as the renewal function.⁵ S_n equals n/λ . (This renewal process is the familiar Poisson process.)

Renewal processes have been extended in two ways that are relevant for analyzing health expenditures. First, statisticians have developed so-called “alternating renewal processes.” In these processes, a process is “on” for a period of time (Y_i) and then is “off” for a period (X_i). The process cycles on and off over time, with an off period followed by an on period representing a cycle.

$N(t)$ represents the number of cycles that occur by time t . Z_i is the length of a given cycle ($Z_i = X_i + Y_i$).

The second extension of the renewal process is the reward renewal process. In this model, when the process is “off,” costs (R_i) are incurred. Figure 1 describes this process. (This figure was derived from Gertsbakh [1989, 2000].) R_i are assumed to be identically and independently distributed with a mean of $E[R]$. The standard reward renewal model assumes that X , Y , and R are all independent.

One can extend the Poisson model by assuming that both X and Y are exponentially distributed with expected lengths μ and ζ , respectively, and that $E[R] = \tau$. Given the assumptions of the model, key characteristics of the process can be described using these parameters. For example, if $A_v(t)$ is the percentage of time the process is “up,” then it can be shown that

$$\lim_{t \rightarrow \infty} A_v(t) = \frac{\zeta}{\zeta + \mu} \quad (2)$$

$1 - A_v(t)$ represents the proportion of time the process is off. Furthermore, the average expenditures per unit of time can be expressed as

$$\lim_{t \rightarrow \infty} \frac{R_v(t)}{t} = \frac{\tau}{\zeta + \mu} \quad (3)$$

(Rigorous proofs of equations (2) and (3) can found in Gertsbakh [2000].)

The basic structure of the model applies well to health care service use and expenditures. Y represents spent time in good health or at least not using services. X represents the length of treatment episodes and R , the associated expenses. Note that in this simple form the model incorporates assumptions that may not fit health services data very well. X and Y may not be exponentially distributed—the risk of leaving or entering treatment may fall (or rise) with

Figure 1: Reward Renewal Process

time in or out of treatment, respectively. X and Y may be correlated—the longer an individual waits to receive treatment, the longer the treatment episode may be. Furthermore, costs of treatment, R_b , are likely grow with episode length. For that reason, we present the results of both a direct application of the simple reward renewal model as well as a more complex extension that relaxes these assumptions. We describe the latter in the next section.

MULTIVARIATE, MULTIPROCESS ANALYSES

As a generalization of the basic framework outlined in the preceding section, one can specify a multilevel, multiprocess (simultaneous equations) model involving three equations: a hazard of starting treatment, a hazard of stopping treatment, and a model for costs per episode day.

$$\begin{aligned} h_{i,e}^X &= h_0^X(t) \exp(\alpha_1 + \beta_1 S_i + \pi_1 \delta_i) \\ h_{i,e}^Y &= h_0^Y(t) \exp(\alpha_2 + \beta_2 S_i + \pi_2 \delta_i) \\ C_{i,e} &= \alpha_3 + \beta_3 S_i + \pi_3 \delta_i + \varepsilon_{i,e} \end{aligned} \quad (4)$$

In this two-level model, episodes of treatment are nested within individuals. i indexes individuals ($i = 1, \dots, N$); e indexes episodes ($e = 1, \dots, E$).⁶ h^X represents the hazard of stopping treatment (leaving state X); and h^Y , the hazard of starting treatment (leaving state Y). c represents the costs per episode day. As an adjustment for the skewed nature of this variable, c is actually log-transformed in our empirical example. ε is an episode-specific error term representing time-varying, unobserved determinants of (log –) costs per day and is assumed to be normally distributed.

For the two hazards, the framework is the standard proportional hazards specification. Under that framework, the two h_0 functions capture the dependency of the hazard on duration, and the covariates (X) (such as site) shift this function up or down proportionately (Lillard and Panis 2000; Singer and Willett 2003). For each outcome, we specify a constant term (α) and a vector of slope coefficients (B) corresponding to the covariates in our empirical example. Given this specification, e^β will represent the proportional effect of a covariate of interest (such as site) on the hazard of interest. e^β is referred to as the hazard ratio, and a ratio greater than 1 indicates that the covariate increased the hazard of interest.

This framework allows us to relax two features of the reward renewal framework. First, we specify the h_0 using splines such that each hazard can rise

or fall as time passes in a state. Second, each equation includes δ , a normally distributed error term representing unobserved heterogeneity or frailty. This term may capture shared, unobserved, or unmeasured determinants of the processes, such as diagnosis or family income. These factors vary across individuals but are constant across episodes. The π terms capture the effect of frailty on the different outcomes. π_1 is constrained to one in order to identify the model. (An alternative would be to constrain the variance of δ to one.)

The parameters of this multilevel model were estimated using the *aML* software (Lillard and Panis 2000).

AN EMPIRICAL APPLICATION

We examined the use of these models in the context of an effort to improve the delivery of mental health services to children and youth involved in multiple child-serving sectors, such as juvenile justice, child welfare, and special education.

Introduction

The so-called “system of care” (SOC) approach relies on interagency coordination and reflects a public health perspective under which responsibility for meeting the mental health needs of children and youth resides at the community level rather than with a single agency. Under such a system, the mental health sector coordinates and delivers services in conjunction with other child-serving agencies, such as juvenile justice and child welfare. The SOC also changes the types of mental health services delivered—it substitutes community-based alternatives for expensive inpatient care.

Such a system could affect expenditures on mental health services in a variety of ways. Such systems typically shift youth from inpatient settings to community-based alternatives, such as partial hospitalization. The latter may be less expensive on a daily basis, but time spent in such settings may be longer (Summerfelt, Foster, and Saunders 1996). Other sectors may shift costs onto the mental health system in a variety of other ways. Youth, for example, may be diverted from juvenile justice and into mental health services (Foster and Connor 2002; Foster, Qaseem, and Connor 2004). Such a diversion may be appropriate and improve mental health outcomes for the youth involved. Nonetheless, expenditures on mental health services are likely to rise.

Evaluation

Since 1994, the Center for Mental Health Services (CMHS) within the U.S. Department of Health and Human Services has funded the development of systems of care through the Comprehensive Community Mental Health Services for Children and Their Families Program. The program provides communities with seed money to establish an administrative structure for the SOC. Communities draw on Medicaid, block grants, and other sources to actually finance services.

CMHS has also funded a national, multisite evaluation, which provides the data for this article. That evaluation includes a quasi-experimental study that matches and compares three system-of-care communities with three similar communities. One pair involves an SOC in Stark County (Canton), Ohio, and a comparison site in Mahoning County (Youngstown), Ohio. The latter represents treatment as usual for children and youth in public systems. Established in the 1970s and administered by the Stark County Family Council, the former links families with education, mental health, child services, health, juvenile justice, and other agencies (Bickman et al. 1997). The target population for the SOC comprises individuals at risk of out-of-home placement who are involved in multiple child-serving sectors, including juvenile justice.

Integration among these agencies occurs at both the administrative and operational levels (Ragan 2003). Regarding the former, representatives of the different agencies participate in a cross-system services planning process. At the operational level, mental health staff are stationed at the other systems, such as juvenile justice. The mental health agency also provides training for personnel in the other systems (e.g., providing juvenile justice personnel with training in the principles of multisystemic therapy).

To evaluate the effect of the SOC on costs and mental health outcomes, a sample of 442 children and adolescents ages 6–17 with serious emotional and behavioral problems who were using mental health services were recruited for a longitudinal study. Study enrollment began in September 1997 and continued through October 1999 with follow-up interviews continuing through December 2000. (In these analyses, however, we only use information from the baseline interview.)

These youth and their care givers participated in a series of interviews. Those interviews provided information on family demographics as well as on the youth's mental health status. Two measures of the latter are the Child Behavior Checklist (Achenbach 1991), a measure of symptomatology, and the Child and Adolescent Functional Assessment Scale (Hodges 1990; Hodges

and Gust 1995; Hodges and Wong 1997), a measure of functional impairment. (Higher scores mean worse functioning.) The analyses given below include baseline values of these measures as covariates.

Services data for these analyses were obtained from the private behavioral health treatment organizations that served as the hub for services in the two communities. These data were extracted from each agency's MIS, which is used for billing purposes. Services recorded in the systems included intake and assessment services, case management, medication monitoring, and individual and group counseling. The SOC also offered day treatment, and the non-SOC offered a short-term crisis residential center. These services were recorded in the MIS for the site involved.⁷ Per-unit costs represented the amounts billed by providers. Services data are available for the period 1997 through 2003.

The MIS data did not include information on inpatient services. As a result, information on the use of such treatment was collected from the records of the two primary inpatient facilities in each county. Three of the four facilities involved general hospitals that delivered short-term inpatient care. The fourth provider (in Mahoning County) was a mental health agency that provided residential, inpatient, and partial hospitalization. Data obtained from three of the providers included services billed to third-party payers, such as private insurance and Medicaid as well as expenses the facilities had to write off because of collection problems. The cost figures for the fourth provider were calculated from standard charges. Average inpatient per diem cost across the two counties was comparable (\$901 and \$835 for comparison and SOC sites, respectively).

Earlier analyses focus on expenditures on mental health services during the year following entry into the study. It was during this period that key mental health outcomes were assessed, and patterns in service use during this time are of particular interest. These analyses reveal that expenditures during the first year were 86 percent higher at the demonstration site (\$3,786 and \$2,036, for the demonstration and the comparison sites, respectively).

RESULTS

We begin with descriptive statistics describing the basic reward renewal model and then examine the results of the multiprocess extension described above.

Basic Model

For this analysis, we delimited treatment episodes using a 45-day "clean period." In the year following study entry, the 442 individuals in the study were

Table 1: Basic Analyses of Episodes

	<i>Demonstration</i>		<i>Comparison</i>		<i>p-Value</i>
	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>	
Description of persons					
Number of episodes	2.56	1.10	3.58	1.88	<.01
Percent of individuals with > 1 episode	91	0.28	99	0.12	<.01
Age	11.13	3.10	11.51	3.27	0.13
Gender (female = 1; male = 0)	0.35	0.48	0.31	0.46	0.31
Race/ethnicity (white = 0; nonwhite = 1)	0.29	0.46	0.53	0.50	<.01
Symptomatology (CBCL)	68.93	9.48	70.17	9.69	0.17
Functioning (CAFAS)	70.75	25.45	78.09	24.64	<.01
Observations	227		215		
Description of episodes					
Mean length of treatment episode (<i>X</i>)	201.60	236.07	101.59	157.97	<.01
Median length of treatment episode (<i>X</i>)	113.00		43.00	NA	<.01
Mean time between treatment episodes (<i>Y</i>)	173.75	179.59	128.36	145.92	<.01
Median time between treatment episodes (<i>Y</i>)	85.00		80.00		0.54
Percent of time receiving treatment during cycle	46		56		<.01
Costs per episode	\$3,982.80	\$6,717.06	\$1,623.30	\$5,063.56	<.01
Costs per episode day	\$44.74	\$113.30	\$55.32	\$106.18	0.13
Observations	591		776		

NA, not applicable.

involved in 1,367 treatment episodes or roughly three per person. The number of episodes per person ranged from 1 to 11. Over 90 percent of the sample had multiple episodes; roughly one-third had four or more.

All individuals contribute one right-censored spell to the analyses. For the 22 percent of the sample that was in services at the end of the observation period, this spell involves a treatment episode (*X*). The censored episodes represent 10 percent of all episodes. For individuals not in treatment at the end of the observation period, the censored spell is one involving the time between treatment episodes (*Y*). The corresponding (censored) gaps in treatment represent 37 percent of all such gaps. The analyses below allow for both types of censoring; in particular, censored spells contribute only the survival function (indicating the spell of interest lasted at least as long as observed).

Table 1 describes the episodes in more detail. It presents the mean number of episodes by site as well as the mean and median length of episodes and time between episodes. (The latter two were calculated based on simple

exponential hazard models.) Table 1 also shows the percentage of time individuals were "down," i.e., in treatment as well as cost per episode and cycle.

One can see that individuals at the demonstration site had one-fourth fewer episodes and that treatment episodes were roughly twice as long (202 versus 102 days). In contrast, the time between episodes was considerably shorter at the comparison site: when those individuals left treatment, they were more likely to return. The net effect is that during a cycle (an episode of treatment followed by a period of no treatment), individuals at the Demonstration were less likely to be in treatment (46 versus 56 percent, $p < .01$).

Table 1 also presents information on expenditures. One can see that costs per episode were higher at the Demonstration (\$3,983 versus \$1,623, $p < .01$), a difference that reflects the difference in episode length. When we examine expenditures per episode day, one can see that the expenditures were lower under the Demonstration (\$45 versus \$55, $p = .13$). This difference was not significant at conventional levels.

As discussed above, all these findings are based on a set of assumptions that may not fit the dynamics of health service use very well. In the next subsection, we relax these assumptions using the multiprocess model specified above.

Multilevel, Multiprocess Model

Table 2 presents key parameter estimates for the multiprocess model. Among the elements of B, we focus on the effect of site. The estimate of the β (-0.53) for the hazard of leaving treatment implies a hazard ratio of 0.59; this estimate means that the risk of exiting services at a point in time was 41 percent lower at the SOC site. The estimate of the β for the hazard of entering treatment (-0.30) implies a hazard ratio of 0.74; this figure suggests that the risk of returning to services at a point in time was 26 percent lower at the SOC site. The results for the third equation suggest that the cost per episode day was equivalent across sites.

Table 2 also provides parameter estimates for the additional covariates. One can see, for example, that higher levels of functional impairment are associated with higher levels of expenditures. Functioning, however, does not appear to influence the timing of episodes.

Note that in terms of the effect of site these estimates have the same substantive implications as those reported in Table 1—individuals at the Demonstration had longer episodes and were less likely to return to treatment after an episode ended. We can, however, reject the more restrictive model

Table 2: Multivariate, Multiprocess Model

	<i>Estimate</i>	<i>SE</i>	<i>T-Stat</i>
Episode length (X)/hazard of leaving treatment			
Constant (α_1)	-2.53	0.60	-4.19
Demonstration	-0.53	0.10	-5.53
Age	0.04	0.02	2.33
Gender (female = 1; male = 0)	-0.07	0.10	-0.69
Race/ethnicity (white = 0; nonwhite = 1)	0.05	0.09	0.52
Symptomatology (CBCL)	0.00	0.01	-0.43
Functioning (CAFAS)	0.00	0.00	0.08
Time between episodes (Y)/hazard of entering treatment			
Constant (α_2)	-11.42	1.23	-9.28
Demonstration	-0.30	0.14	-2.14
Age	-0.04	0.01	-3.45
Gender (female = 1; male = 0)	0.00	0.08	-0.02
Race/ethnicity (white = 0; nonwhite = 1)	0.05	0.07	0.74
Symptomatology (CBCL)	0.01	0.01	0.96
Functioning (CAFAS)	0.00	0.00	0.81
Costs per episode day			
Constant (α_3)	2.89	0.36	8.07
Demonstration	0.01	0.10	0.06
Age	0.01	0.01	1.00
Gender (female = 1; male = 0)	0.13	0.09	1.32
Race/ethnicity (white = 0; nonwhite = 1)	-0.12	0.09	-1.24
Symptomatology (CBCL)	-0.01	0.01	-1.69
Functioning (CAFAS)	0.01	0.00	2.95
Var (ε)	0.95	0.03	28.71
Heterogeneity			
Var (δ)	0.44	0.05	8.77
π_2	0.36	0.10	3.51
π_3	1.37	0.18	7.83
Log Likelihood: observations	-6,843.12		

Note: The models also include nine splines for the two hazard models. Complete parameter estimates are available from the first author.

($p < .0001$).⁸ In particular, we can reject the null hypothesis that the variance of δ is zero, suggesting that episode length was shaped by unobserved, time-invariant determinants. Both π terms were statistically significant, suggesting that both the time between episodes and costs per day were influenced by the same unobserved factors that influence episode length. In particular, because both are positive, the factors that raise the likelihood of leaving treatment increase the likelihood of entering treatment and the costs per day of treatment.

Figure 2: Baseline Hazard for Exiting Services

The model parameters were insensitive to the number or specification of the nodes of the spline function for the baseline hazard (h_0). While not presented in Table 2, the parameter estimates for the spline function suggest that the hazard is not constant over time (as implied by the exponential model). Figure 2, for example, shows the plotted hazard function for leaving services (which governs episode length). The hazard of leaving services declines sharply after a single day of service and then gradually rises. This shape is consistent with what one sees in the raw data. (Twelve percent of all episodes last only 1 day.)

Figure 2 also provides the baseline hazard for entering services, that function spikes at about 60 days. One can see that the hazard is zero through 45 days. This feature of the graph reflects the definition of the episodes (an individual had to be out of treatment for 45 days for a spell of no treatment to be recorded).

DISCUSSION

This article highlights the application of an episodic model of service use to data on expenditures in a given calendar period. The results suggest that simply finding that costs were higher under the SOC hides a good deal of the story: increased expenditures were driven by substantial shifts in the timing with which clients started and stopped services. In general, individuals at the demonstration site had better continuity of care.

The reward renewal model may have a variety of advantages for policy makers and researchers. For example, expenditures are notoriously difficult to predict, with the proportion of explained variance often quite low (Newhouse, Buntin, and Chapman 1997). One possibility, however, is that conventional models are misspecified. If three processes are at work—governing the starting and stopping of services and expenditures per day—then a one-equation model is misspecified, and explanatory power may be lost. This possibility seems especially likely since different individual, family, or system characteristics drive the different processes.

This article represents only the first step in the application of these models. The work presented here could be extended in a variety of ways. For example, one might analyze condition-specific episodes and the relationship between them (e.g., the relationship between treatment for dementia and that for physical injury). Allowing for multiple and co-occurring episodes would seem essential as one broadened the scope of inquiry beyond a fairly narrow category of services (such as mental health services).

One might also integrate treatment settings into the analyses. One might model the movement between inpatient care, outpatient care, and no treatment over time. This approach would build upon work on multistate models (Aalen, Bjertness, and Soonju 1995; Hougaard 1999; Williams et al. 1999). Such an extension would be analogous to that of the two-part to the four-part model.

Other extensions are more technical in nature. The reward renewal model assumes that individuals cycle in and out of services indefinitely. For a chronic condition, such as many mental disorders, this assumption may be reasonable. However, for other conditions, the model may need to incorporate a final, absorbing state representing (permanent) termination of treatment.

A final extension would address another limitation of the analysis; it would tie the episodes of treatment to episodes of illness. Future work could consider the simultaneous estimation of a growth curve of mental health status and could examine the interplay between an individual's profile of service use and his or her mental health over time. Linking treatment patterns to outcomes would reveal the extent to which the former are really proxies for quality of care.

ACKNOWLEDGMENTS

Data for this paper were collected through the national evaluation of the Comprehensive Community Mental Health Services for Children and Their Families Program (#280-94-0012) funded by the Center for Mental Health

Services. The author would like to thank Elizabeth Gifford and Dennis Shea for helpful comments on an earlier draft of the paper. The author is responsible for any remaining errors or omissions.

NOTES

1. The discussion here uses the terms “costs” and “expenditures” largely interchangeably. Economists, however, differentiate the two—the former refers to the opportunity costs of the resources used in delivering a service. In that case, costs and expenditures (or payments) likely differ (Hargreaves et al. 1998). In the empirical example below, dollar amounts represent expenditures rather than costs per se.
2. There are analyses of costs per episode (e.g., Holmes and Deb [1998]), but these analyses do not use this information to understand costs in a given fiscal or calendar period.
3. This discussion focuses on episodes of care rather than episodes of illness or disease (Wingert et al. 1995/6). The episode of care is most relevant from the perspective of the payor; the latter two, for the patient and provider, respectively. The different types of episodes can overlap substantially, but in many instances they may not. For example, an individual with an illness may never receive treatment; her illness may persist after she leaves treatment; or she may remain in treatment beyond the end of her illness.
4. One also could allow the definition of a “clean period” to vary with diagnosis or even the severity of the condition (Rosen et al. 1998; Wingert et al. 1995/6).
5. An exponential distribution for X_i implies that the process is “memoryless”—the likelihood of an event does not rise or fall as the time since the last event grows. An advantage of the exponential distribution is that the renewal function has an analytical solution. The memoryless assumption may not fit the processes of interest, however. Available alternatives include the γ and the normal distributions. Approximations to other distributions also are available. (For details, see Gertsbakh [1989, 2000].) We relax the memoryless assumption in our empirical illustration below.
6. The period of time between episodes for episode e pertains to the time as episode $e - 1$ ended.
7. Data on child and family outcomes were collected through face-to-face interviews conducted with care givers and their children. Interviews were conducted at study entry and then at 6-month intervals.
8. To test the assumptions embedded in the simpler model, we restricted the model such that the hazard did not vary with duration and there was no unobserved heterogeneity. This involved 12 restrictions (nine parameters for the spline, the variance of δ and the two π terms). We could reject the more restrictive model at $p < .000001$. The parameter estimates of the restricted model are similar to those in Table 2 and are available from the first author.

REFERENCES

- Aalen, O. O., E. Bjerntness, and T. Soonju. 1995. "Analysis of Dependent Survival Data Applied to Lifetimes of Amalgam Fillings." *Statistics in Medicine* 14 (16): 1819–29.
- Achenbach, T. M. 1991. *Manual for the Child Behavior Checklist and 1991 Profile*. Burlington, VT: University of Vermont.
- Arria, A. M. 2003. "Drug Treatment Completion and Post-Discharge Employment in the TOPPS-II Interstate Cooperative Study." *Journal of Substance Abuse Treatment* 25 (1): 9–18.
- Bauchner, H. 2000. "ADHD: A New Practice Guideline from the American Academy of Pediatrics. Attention Deficit Hyperactive Disorder." *Archives of Disease in Childhood* 83 (1): 63.
- Bickman, L., T. W. Summerfelt, J. M. Firth, and S. M. Douglas. 1997. "The Stark County Evaluation Project." In *Valuating Mental Health Services: How Do Programs for Children "Work" in the Real World*, edited by D. A. Northrup and C. Nixon, pp. 231–58. Newbury Park, CA: Sage Publications.
- Bondy, J., S. Berman, J. Glazner, and D. Lezotte. 2000. "Direct Expenditures Related to Otitis Media Diagnoses: Extrapolations from a Pediatric Medicaid Cohort." *Pediatrics* 105 (6): NIL-1–7.
- Busch, S. H. 2002. "Specialty Health Care, Treatment Patterns, and Quality: The Impact of a Mental Health Carve-Out on Care for Depression." *Health Services Research* 37 (6): 1583–601.
- Busch, S. H., D. Leslie, and R. Rosenheck. 2004. "Measuring Quality of Pharmacotherapy for Depression in a National Health Care System." *Medical Care* 42 (6): 532–42.
- Dean, J. M., D. D. Vernon, L. Cook, P. Nechodom, J. Reading, and A. Suruda. 2001. "Probabilistic Linkage of Computerized Ambulance and Inpatient Hospital Discharge Records: A Potential Tool for Evaluation of Emergency Medical Services." *Annals of Emergency Medicine* 37 (6): 616–26.
- Druss, B., and R. Rosenheck. 1997. "Evaluation of the HEDIS Measure of Behavioral Health Care Quality." *Psychiatric Services* 48 (1): 71–5.
- Druss, B. G., C. L. Miller, R. A. Rosenheck, S. C. Shih, and J. E. Bost. 2002. "Mental Health Care Quality under Managed Care in the United States: A View from the Health Employer Data and Information Set (HEDIS)." *American Journal of Psychiatry* 159 (5): 860–2.
- Duan, N., J. P. Newhouse, C. N. Morris, and W. G. Manning. 1983. "A Comparison of Alternative Models for the Demand for Medical Care." *Journal of Business and Economic Statistics* 2: 115–26.
- . 1984. "Choosing between the Sample-Selection Model and the Multi-Part Model." *Journal of Business and Economic Statistics* 2 (3): 283–9.
- Ettner, S. L., R. G. Frank, T. G. McGuire, and R. C. Hermann. 2001. "Risk Adjustment Alternatives in Paying for Behavioral Health Care under Medicaid." *Health Services Research* 36 (4): 793–811.
- Ettner, S. L., R. G. Frank, T. G. McGuire, J. P. Newhouse, and E. H. Notman. 1998. "Risk Adjustment of Mental Health and Substance Abuse Payments." *Inquiry* 35 (2): 223–39.

- Foster, E. M. 1998. "Does the Continuum of Care Influence Time in Treatment? Evidence from the Fort Bragg Evaluation." *Evaluation Review* 22 (4): 447-69.
- Foster, E. M., and T. Connor. 2005. "The Public Costs of Better Mental Health Services for Children and Adolescents." *Psychiatric Services* 56: 50-5.
- Foster, E. M., A. Qaseem, and T. Connor. 2004. "Can Better Mental Health Services Reduce the Risk of Juvenile Justice System Involvement?" *American Journal of Public Health* 94 (5): 859-65.
- Gertsbakh, I. B. 1989. *Statistical Reliability Theory*. New York: Marcel Dekker Inc.
- . 2000. *Reliability Theory with Applications to Preventive Maintenance*. New York: Springer.
- Goldman, W., J. McCulloch, B. Cuffel, D. A. Zarin, A. Suarez, and B. J. Burns. 1998. "Outpatient Utilization Patterns of Integrated and Split Psychotherapy and Pharmacotherapy for Depression." *Psychiatric Services* 49 (4): 477-82.
- Hargreaves, W. A., M. Shumway, T. W. Hu, and B. Cuffel. 1998. *Cost-Outcome Methods for Mental Health*. New York: Academic Press.
- Herings, R. M., and J. A. Erkens. 2003. "Increased Suicide Attempt Rate among Patients Interrupting Use of Atypical Antipsychotics." *Pharmacoepidemiology and Drug Safety* 12 (5): 423-4.
- Herrerias, C. T., J. M. Perrin, and M. T. Stein. 2001. "The Child with ADHD: Using the AAP Clinical Practice Guideline. American Academy of Pediatrics." *American Family Physician* 63 (9): 1803-10.
- Hodges, K. 1990. *Child and Adolescent Functional Assessment Scale (CAFAS)*. Ypsilanti, MI: Department of Psychology, Eastern Michigan University.
- Hodges, K., and J. Gust. 1995. "Measures of Impairment for Children and Adolescents." *Journal of Mental Health Administration* 22 (4): 403-13.
- Hodges, K., and M. M. Wong. 1997. "Use of the Child and Adolescent Functional Assessment Scale to Predict Service Utilization and Cost." *The Journal of Mental Health Administration* 24 (3): 278-90.
- Holmes, A. M., and P. Deb. 1998. "Provider Choice and Use of Mental Health Care: Implications for Gatekeeper Models." *Health Services Research* 33 (5): 1263-84.
- Hornbrook, M. C., A. Hurtado, and R. Johnson. 1985. "Health Care Episodes: Definition, Measurement, and Use." *Medical Care Review* 42: 163-217.
- Hougaard, P. 1999. "Multi-State Models: A Review." *Lifetime Data Analysis* 5 (3): 239-64.
- Kapur, K., A. S. Young, and D. Murata. 2000. "Risk Adjustment for High Utilizers of Public Mental Health Care." *Journal of Mental Health and Policy Economics* 3 (3): 129-37.
- Kessler, L. G., D. M. Steinwachs, and J. R. Hankin. 1980. "Episodes of Psychiatric Utilization." *Medical Care* 8 (12): 1219-27.
- Leslie, L. K., J. Weckerly, D. Plemmons, J. Landsverk, and S. Eastman. 2004. "Implementing the American Academy of Pediatrics Attention-Deficit/Hyperactivity Disorder Diagnostic Guidelines in Primary Care Settings." *Pediatrics* 114 (1): 129-40.

- Lillard, L. A., and C. W. A. Panis. 2000. *aML Multilevel Multiprocess Statistical Software*. Los Angeles: EconWare.
- Mainous, A. G., and W. J. Hueston. 1998. "The Cost of Antibiotics in Treating Upper Respiratory Tract Infections in a Medicaid Population." *Archives of Family Medicine* 7 (1): 45-9.
- Mundt, J. C., G. N. Clarke, D. Burroughs, D. O. Brenneman, and J. H. Griest. 2001. "Effectiveness of Antidepressant Pharmacotherapy: The Impact of Medication Compliance and Patient Education." *Depression and Anxiety* 13 (1): 1-10.
- Newhouse, J. P., M. B. Buntin, and J. D. Chapman. 1997. "Risk Adjustment and Medicare: Taking a Closer Look." *Health Affairs* 16 (5): 26-43.
- Ragan, M. 2003. "Building Comprehensive Human Service Systems." *Focus* 22 (3): 58-62.
- Rosen, A. K., R. L. Houchens, T. B. Gibson, and A. Mayer-Oakes. 1998. "Developing Episodes of Care for Adult Asthma Patients: A Cautionary Tale." *American Journal of Medical Quality* 13 (1): 25-35.
- Rost, K., P. Nutting, J. Smith, J. Werner, and N. Duan. 2001. "Improving Depression Outcomes in Community Primary Care Practice: A Randomized Trial of the QuEST Intervention. Quality Enhancement by Strategic Teaming." *Journal of General Internal Medicine* 16 (3): 143-9.
- Rushton, J. L., K. E. Fant, and S. J. Clark. 2004. "Use of Practice Guidelines in the Primary Care of Children with Attention-Deficit/Hyperactivity Disorder." *Pediatrics* 114 (1): e23-8.
- Singer, J. D., and J. B. Willett. 2003. *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. Oxford: Oxford University Press.
- Stein, M. T., and J. M. Perrin. 2003. "Diagnosis and Treatment of ADHD in School-Age Children in Primary Care Settings: A Synopsis of the AAP Practice Guidelines. American Academy of Pediatrics." *Pediatric Review* 24 (3): 92-8.
- Summerfelt, W. T., E. M. Foster, and R. Saunders. 1996. "Mental Health Services Utilization in a Children's Mental Health Managed Care Demonstration." *Journal of Mental Health Administration* 23 (1): 80-91.
- Wells, K. B., R. Sturm, C. D. Sherbourne, and L. S. Merideth. 1996. *Caring for Depression: A RAND Study*. Cambridge, MA: Harvard University Press.
- Williams, C. L., B. M. Johnstone, J. G. Kesterson, K. A. Javor, and A. D. Schmetzer. 1999. "Evaluation of Antipsychotic and Concomitant Medication Use Patterns in Patients with Schizophrenia." *Medical Care* 37 (4 Suppl Lilly): AS81-6.
- Wingert, T. D., J. E. Kralewski, T. J. Lindquist, and D. J. Knutson. 1995/6. "Constructing Episodes of Care from Encounter and Claims Data: Some Methodological Issues." *Inquiry* 32: 430-43.