A SIMULATION MODEL FOR EVALUATION AND DESIGN OF INPATIENT CARE FACILITIES

H. Richard Howland and Richard Irvin Baldock

Westinghouse Electric Corporation

ABSTRACT

The complexity of activities and fluctuations in workload make prospective evaluation and design of inpatient care facilities difficult. This paper describes a simulation which allows the user to select a compromise between service availability and cost by comparing alternatives for different ward sizes, ancillary capacities and admission policies. Reporting includes not only occupancy and workload figures, but also statistics on overloads and excess occupancy. The sequence of care administered to an inpatient is modeled as a Markov chain, whose states are patient level of dependency; ward workloads and ancillary usage depend on the patient's state. Different parameters in the basic model represent the sequences of care corresponding to distinct diagnosis classes; varying patient mixes can be simulated by adjusting arrival rates in the diagnosis categories. This paper focuses on the use of the simulation in comparing admission policies.

INTRODUCTION

The simulation described in this paper allows the administrator of an inpatient care facility to evaluate admitting policies, and the architect or designer to study the effects of different ward sizes and ancillary capacities as well. The administrator of an inpatient care facility wishes to maintain the most nearly even flow of resources within the facility, so as to minimize the unit cost of operations. Choice of admitting policies is one of the decisions available to him. The architect or designer may adjust ward complements and size of ancillary facilities, but is under pressure to plan for greatest utilization of expensive facilities while maintaining adequate service availability. In either case the number and kinds of patients moving through the facility are generated and governed by stochastic phenomena with considerable fluctuations in time. The planning and administrative functions must recognize and cope with these fluctuations. The availability of beds and required services is a function of both the admitting policies and the sizes or capacities, and both must be considered in the design stage.

This work was sponsored by the Westinghouse Health Systems Division.

Many studies have been directed toward the problem of determining the total number of beds. Others have studied admissions systems. Analytic models, such as those of Flagle (3,4) and Young (15,16) provide insight, but do not have sufficient flexibility to study complex cases. Previous simulation models (2,12,13) which dealt with admission scheduling used models of inpatient care which lend themselves to studies of occupancy, but not workloads. On the other hand, the more detailed models of patient care (7,14) have not yet been included in simulations. Milsum et al. survey admissions systems in (11). This simulation is a discrete-time stochastic and dynamic model. Its salient features include division of the hospital into services, a Markov chain model of inpatient care, and a constraint on waiting time before admission. The model of patient care drives the flow of resources while the patient is in the hospital, and allows computations of the workload the patient generates as well as bed occupancy. Patient care is modeled as a static Markov chain, whose states are the patient levels of dependency (full care, intermediate care, and self-care) (1). The patient's level of dependency changes from day to day in accordance with the transition probabilities, which are assumed to be independent of the state of the hospital. The patient uses ancillary facilities (radiology, laboratory, pharmacy) at a rate that is dependent on his state. Surgery is treated the same way formally, but is interpreted as a probability of undergoing surgery on a given day. A separate set of transition probabilities and ancillary usage matrices is estimated for each diagnosis category, as explained in (6).

The concept of maximum waiting time before admission recognizes realistic constraints on decision rules for admission. Young (16) divides patients referred for admission into two classes, urgent and elective, with elective patients deferrable indefinitely. Kolesar (8) studied analytic decision rules based on these two classes. In practice patients must be admitted before a certain number of days have elapsed, though the exact number will vary with the diagnosis category.

The maximum waiting time may not be expressed explicitly, but it can be determined empirically, perhaps by negotiation at the time admission is requested (13). Reducing the maximum waiting times, of course, reduces the room for maneuver in applying admission policies.

The simulation hopefully strikes a balance between desirable detail, availability of meaningful data, and implementation effort. Some factors have been deliberately omitted or treated crudely, and those are described here. It is felt that leveling occupancy or workload from day-to-day can be decoupled, in practice, from load-leveling within a day; since the focus of the simulation is on occupancy and daily workload, the simulation is run in 24-hour time steps. For the same reason, ancillary usage has been treated as a continuous rate, with only surgery treated as a discrete event. Though the hospital is divided into services, beds have not been allocated within a service to private rooms, semi-private, etc.; the simulation work of Goldman et al. (5) showed that over-capacity occupancy and patient waiting time to admission are much more sensitive to the total number of beds than to the allocation scheme.

THE COMPUTER SIMULATION MODEL

The simulation is composed of two main programs, a patient generation program, and the simulation itself, which includes statistical reporting, as shown in Illustration 1. The patient generator creates the temporary entities for use in the simulation. The user specifies the average arrival rate for each diagnosis category by age and sex; this data may be empirical or predicted from existing demographic models (9,10). The user also supplies a matrix which specifies the distribution of patients to services by age, sex, and diagnosis category, and the average maximum waiting time by diagnosis category. Patients are created by a Poisson process with average arrival rate as parameter. The maximum waiting time for a patient is picked from a negative exponential distribution with average maximum waiting time as a parameter. Once created, a patient is allocated randomly to one of the services according to the ward distribution by age, sex, and diagnosis category. Each day's batch of referrals for admission is read onto a disk file, which can be re-used to compare directly the effect of different capacities and admissions policies.

The input for the simulation program consists of the Markov chain transition probabilities and ancillary usage matrices for each diagnosis category, nominal ward capacities, ward workload factors, nominal ancillary facility capacities, ancillary workload factors, selection of admission policy, and report generation parameters. Within the simulation model, the permanent entities are the queue of patients desiring admission, the patient storage list, and the services. The major permanent entity is the patient storage list, which contains the attributes of all patients, admitted or waiting. The admission queue is a linked list of pointers into the patient list which designates those patients awaiting admission. The wards are represented by chained lists of pointers into the patient list which designate the patients on a particular service.

The temporary entities within the simulation are the patients occupying the patient list. They have both permanent attributes, which are the ones with which they were created in the patient generator, and temporary attributes, which include the current state, number of days waited to admission, number of days in hospital, and current service, if any.

The relationship between the entities and attributes of entities as time passes constitutes the processing within the simulation. The expanded flow diagram of the simulation, Illustration 2, shows that the processing is driven by the patients. At the end of each simulated day, patients currently in the hospital are assigned new states in accordance with the transition probabilities. If the state is zero, the patient is discharged, and removed from the patient list and all other linked lists containing references to him. New patients are read in daily from the patient file, added to the patient list, and their pointers integrated with those already on the admission queue. It is assumed that once a patient is referred for admission he will not renege, and will arrive when scheduled. The admission queue is ordered by least maximum waiting time, and first-in first-out for equal waiting times. Decision rules can then be applied to determine admissions. Those admitted have their pointers deleted from the admission queue and added to the appropriate service.

The daily workload on each service is computed from the level of dependency of the patients and the workload factors. Daily ancillary workloads are computed from ancillary workload factors and the rate of ancillary usage for each patient. A statistical snapshot records the state of the system at the conclusion of each day. The processing cycles until the patient file is exhausted, at which point the statistics are reported. An arbitrary initial number of days may be skipped in gathering statistics to allow the system come to equilibrium.

SIMULATION OUTPUTS

There are many measures of effectiveness by which a policy or set of capacities can be judged, so a variety of statistics on workload, occupancy, and admissions are collected and printed. Generally, as much attention has been paid to exceptional conditions as to ordinary ones. Statistics of occupancy and over-occupancy are collected and printed, so that utilization of resources can be assessed. Similarly, ancillary workload and overload are collected and printed. Several statistics for the admissions process are printed. In addition to the average patient waiting time, the fraction of patients admitted before their maximum waiting time elapsed is also printed. Though not yet implemented, it is planned to vary admission rates by day of the week, and collect statistics by day, in order to accommodate the need for load-leveling within a week.

As an example, the operation of a hospital is compared for identical patient sets but different admissions policies. The base policy, which represents no attempt to schedule, is to admit patients as soon as they are referred, irrespective of the state of the hospital. For comparison, the second policy is as follows: when in a given ward occupancy is above 90%, no patients other than emergency cases are admitted; otherwise, elective patients, if there are enough, are admitted to bring it up to that level. A patient is admitted on an emergency basis if his maximum waiting time elapses. The maximum waiting times by diagnosis category were estimated by two physicians.

Tables 1 and 2 compare statistics of occupancy and workload on the medical service for the two cases, and Table 3 shows the statistics of the hematology lab workload, which serves all patients. The average utilization is nearly the same in both cases, because the patient set was identical. However, the variance in occupancy and workload shows a marked reduction, and the over-capacity statistics show no days on which ward capacity was exceeded. Table 4 shows the statistics of the admission queue for the second policy. The decision rule used here is only an example; one could use a decision rule based on ward workload instead of occupancy.

The program was written in the NUALGOL (Norwegian University ALGOL) language for the UNIVAC 1106. The simulation occupies 35K words of permanent core storage locations, plus additional space for the variable data structures such as the patient list. The program runs in less than 5 minutes for a year's simulation of a 400 bed facility. The program was written in modular form for flexibility and ease in debugging.

TABLE 1

Statistics of Occupancy - Medicine Service

	Occupancy		Over-occupancy	
	Mean	Variance	Mean	Variance
Policy 1	113.0	93.7	2.7	4.7
Policy 2	113.1	13.0	0	-

TABLE 2

Statistics of Workload - Medicine Service

	Workload Units		Overload	
	Mean	Variance	Mean	Variance
Policy 1	104.0	100.3	5.6	21.6
Policy 2	102.0	18.1	0	-

TABLE 3

Statistics of Workload - Hematology

	Workload Units		Overload	
	Mean	Variance	Mean	Variance
Policy 1	5.1	9.6	1.5	0.55
Policy 2	5.0	8.0	0.1	0.08

TABLE 4

Admission Queue Statistics - Policy 2

Fraction of patients admitted without delay	0.74
Fraction admitted within max. waiting time	1.0
Average waiting time for deferred patients	3.9 days

BIBLIOGRAPHY

1. Connor R. J. "An Inpatient Classification System," Ph.D. Dissertation, Johns Hopkins University, 1960.
2. Fetter, R. B., and Thompson, J. D. "The Simulation of Hospital Systems", Operations Research, Vol. 13, No. 5, 1965.
3. Flagle, C. D. "The Problem of Organization for Inpatient Care", Management Science: Models and Techniques, Vol. II, Pergamon Press, N. Y., 1960.
4. Flagle, C. D. "Operations Research in the Health Services", Operations Research, Vol. 10, No. 5, 1962.
5. Goldman, J., Knappenberger, H. A. and Eller, J.C. "Evaluating Bed Allocation Policy with Computer Simulation", Health Services Research, Vol. 3, No. 2, Summer, 1968.
6. Howland, H.R. and Powell, B.A. "A Descriptive Model of Inpatient Care Dynamics". Proc. Fourth Annual Pittsburgh Conference on Modeling and Simulation, Pittsburgh, Pa., April, 1973.
7. Kao, E. P. C. "A Semi-Markovian Population Model with Applications to Hospital Planning", IEEE Trans. Systems, Man and Cybernetics, Vol. SMC-3, No. 4, pp. 327-336, 1973.
8. Kolesar, P. "A Markovian Model for Hospital Admissions Scheduling", Management Science, Vol. 16, No. 6, pp. 384-396, 1970.
9. Love, C. G. and Trebbi, G. "Regional Health Care Planning", IEEE Trans. Systems, Man and Cybernetics, Vol. SMC-3, No. 1, 1973.
10. Mathias, R. A. "Planning Models of Demand for Health Care", Seventh Annual Meeting, Assoc. Advancement Medical Instrumentation, Washington, D.C., April, 1972.
11. Milsum, J. H., Turban, E., and Vertinsky, I. "Hospital Admissions Systems: Their Evaluation and Management", Management Science, Vol. 19, No. 6, pp. 646-665, 1970.
12. Robinson, G. H., et al. "A Simulation Model for the Evaluation of Scheduling Decision Rules for Hospital Elective Admissions", Res. Report HFT-64-5-T, Human Factors in Technology Research Group, University of California, Berkeley, 1964.

13. Robinson, G.H., Wing, P. and Davis, L. E. "Computer Simulation of Hospital Patient Scheduling Systems", Health Services Research, Vol. 3, No. 2, Summer, 1968.
14. Smallwood, R. D., et al. "A Medical Services Requirements Model for Health System Design". Proc. IEEE, Vo. 57, pp. 1880-1887, 1969.
15. Young, J. P. "Stabilization of Bed Occupancy Through Control of Admissions," Hospitals, J.A.H.A., Vol. 39, No. 19, 1965.
16. Young, J. P. "Administrative Control of Multiple Channel Queueing Systems", Operations Research, Vol. 14, No. 1, 1966.

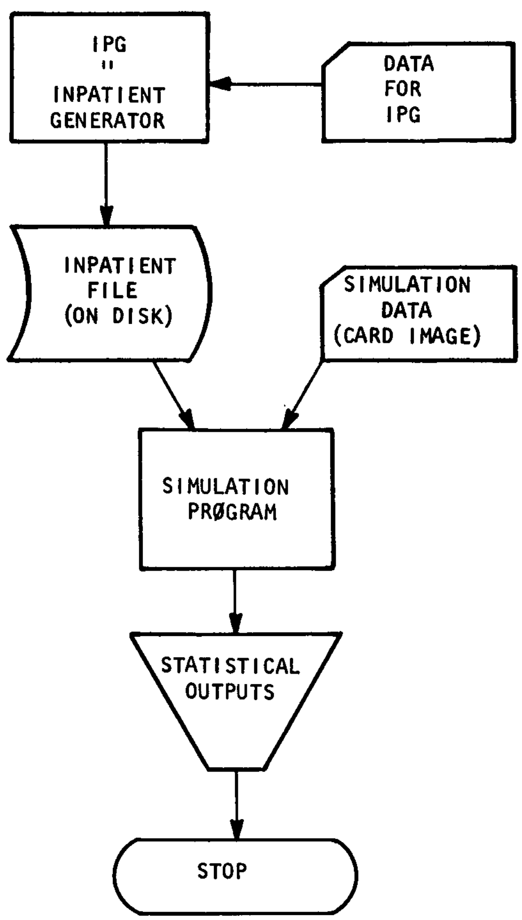

Illustration 1—Patient generation program and simulation program

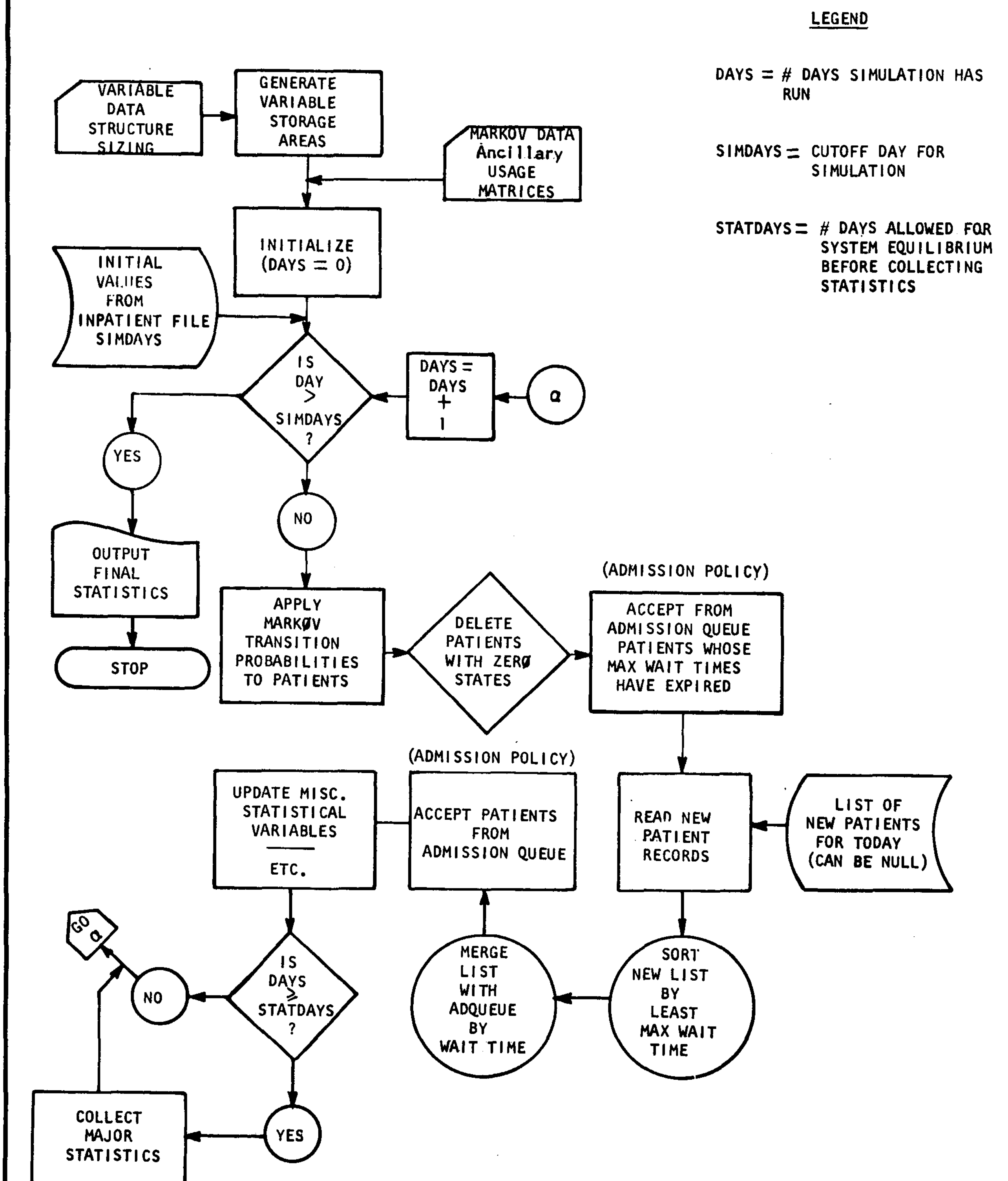

Illustration 2—Expanded flow diagram of simulation

H. RICHARD HOWLAND, Systems Technology Department, Westinghouse Electric Corporation, Research Laboratories, Pittsburgh, Pa. Received B.S.E.E. from M.I.T.; M.S.E.E. and Ph.D. from the University of Pennsylvania.

RICHARD IRVIN BALDOCK, Computer Sciences Department, Westinghouse Electric Corporation, Research Laboratories, Pittsburgh, Pa. Received B.S. in Mathematics from University of Southwestern Louisiana; M.S. in Mathematics from Florida State University.