Proceedings of the 2002 Winter Simulation Conference
E. Yücesan, C.-H. Chen, J. L. Snowdon, and J. M. Charnes, eds.

SIMULATING WITHOUT DATA

Dean S. Hartley III

Hartley Consulting
106 Windsong Ln
Oak Ridge, TN 37830, U.S.A.

ABSTRACT

As a general rule, simulation requires detailed data to properly represent a problem. Occasionally, a problem may be successfully solved using only reasonable approximations to the data. This paper describes such a situation. The Army's health services conceived the Personal Information Carrier (PIC) as a replacement for the "dog tag." The concept was to support portable personal health records. The problem was to determine the contents of the records to be stored in the PIC.

1 INTRODUCTION

The Personal Information Carrier (PIC) is a dog-tag sized memory card that slides into a PC card adapter made especially for this device. The U.S. Army Medical Research and Materiel Command (MRMC) Telemedicine and Advanced Technology Research Center (TATRC) is developing the PIC to contain the entire medical record of soldiers, both enlisted troops and officers, throughout their military career. The intent is to make these data readily available anywhere the soldiers seek medical treatment.

When soldiers are not deployed, they typically seek medical treatment at a local military hospital or medical facility. The facility maintains the soldiers' records in (paper) files. When the soldiers transfer to another location, they carry the file folders to the new location, where they are again maintained by the local medical facility. When the soldiers are deployed, they don't carry their folders; the units are responsible for the medical files.

This system worked adequately in the slower-paced environment of the past. However, technology now permits new approaches to the problem of making soldiers' medical histories available to the appropriate physicians and of keeping the histories up to date, regardless of where treatments are administered.

Two basic approaches are possible. The first is a centralized system in which all soldiers' records are maintained in a single database and made available electronically at the treatment location. This approach has the disadvantage of relying on communications links. The disadvantage is not severe in peacetime; however, it represents a major vulnerability in wartime. The second approach is decentralized and mimics some of the features of the current paper system. Soldiers would carry their own records with them at all times, synchronizing with a master database when convenient.

The second approach makes the most sense given the military mission. There are disadvantages, such as providing access to appropriate personnel while denying access to the enemy (or to soldiers desiring to modify their own records). These issues will not be addressed here. What will be addressed is the technical issue of the contents of the records to be carried.

2 PIC SYSTEM

Physically, the PIC is a CompactFlash™ memory card, carried in a hardened container, which is about the size of a dogtag (Figure 1). The memory card is removed from the carrier and inserted into a PC Card Adapter for reading by a computer.

Figure 1: PIC and PC Card Adapter

The PIC system hardware is completed with laptop and desktop computers that accept PC Cards. The software consists of database systems to create, hold, and update the medical records.

3 PROBLEM DEFINITION

As mentioned earlier, the larger problem includes the definition of the software system to support the PICs, including such issues as security. The part of the problem addressed in this paper concerns the content of the information to be stored on the PIC. The Army specified the use of CompactFlash™ or some similar storage mechanism. At the time of the problem definition, these devices could hold from 2 to 16 megabytes (MB). It was assumed that the capacities for a given format would grow over time; however, the assumed limit was on the order of 250 MB.

3.1 Item Sizing

The data in current medical files can be divided into two types: textual and image. The textual data consists of the doctors' reports, diagnostic test results, diagnoses, and prescriptions/procedures. The size of textual data for any given encounter is measurable in kilobytes (KB). The image data consists of photographs and x-rays. Image sizes are measurable in megabytes.

3.2 Soldier Records Sizing

The item sizing showed that the PIC could hold most previous-encounter information. This would permit a soldier to carry some information. However, although a system could be designed to read the previous-encounter information from the PIC and permit a doctor to replace the data with the current-encounter information, it was not clear that this information would be relevant for the next encounter. Further, the Army desired to have the complete set of records on the PIC.

Some method was required for determining the storage size implied by this desire. If the size were determined to be unattainable, information to support alternatives was required.

4 SOLUTION DEFINITION

A simple model was devised: find the average number of encounters per soldier per year; find the average PIC size per encounter; multiply the two numbers; and multiply by 35 years. This would yield a point estimate of the total PIC size required. Slightly more information would be required to estimate the variance. Unfortunately, this information was not available.

A more complicated model was needed, one that promised both data accessibility and more complete answers. The following data were required:

- Types of encounters;
- Distribution of PIC size per encounter by type;
- Frequency of encounter type (by group)
  - male enlisted,
  - male officer,
  - female enlisted,
  - female officer; and
- Distribution of length of service (by group).

4.1 Encounter Types and Sizes

A proposed database structure for medical encounters, supplied by the Army, and obtained information on encounter types and sizes was analyzed. The structure defined the text fields and the nature of the input: single character, numeric value, free form text, etc. Eleven types of encounters were needed, as follows:

- Physical examinations;
- Visits to a dentist;
- Visits to an eye doctor;
- Minor illnesses;
- Major illnesses;
- Minor traumas;
- Major traumas;
- Visits to a gynecologist;
- Visits to an obstetrician;
- Mammography visits; and
- Visits to a psychiatrist.

The primary driver in separating encounter types was the size and nature of imagery required by the encounters. The secondary driver was the need for follow-up visits. This latter factor differentiates minor illnesses and traumas from major illnesses and traumas. While follow-up visits might be scheduled, as for other encounters, using a distribution pattern, it makes more sense to schedule them based on the original visit. Thus the primary driver was a size-based criterion and the secondary driver was a modeling-based criterion.

4.2 Encounter Frequencies and Service Length Distributions

Initial encounter frequencies were defined through general knowledge and educated guesses. For instance, there is an **annual** physical for soldiers on flight status and dental visits are supposed to be every six months. Minor illnesses might occur once a year and major illnesses less frequently. Each encounter type requires its own distribution for first occurrence and distribution for intervals between occurrences (for each group).

The length of service distribution was more difficult to guess. Most soldiers serve at least two years and 30 years is the basic maximum career (full retirement). A discrete distribution is required to describe what is known or guessed for length of service for each group.

4.3 Model Logic

Figure 2 illustrates the basic model logic, implemented in Arena (Kelton, Sadowski, and Sadowski 1998). Each replication of the simulation represents the service life of one soldier (of a particular group). A random draw selects the length of service. During that period, encounters are scheduled and processed according to their attributes. Data are collected on the PIC size (text) and image PIC size required for each replication. The animation displays the number of soldiers simulated, the max text and image sizes, and the group being simulated. For test and demonstration purposes, the animation also shows the selected years of service for each replication and a moving object traversing the path from “enter service” to “leave service.” During each replication, moving objects traverse the encounter paths, representing the types of encounters that occur.

5 DATA

Despite the title of this paper, some data are required. The data (including distribution types) consist of guesses, based on experience, rather than validated data. Some of the encounters are simple, as illustrated in the Dentist encounter in Table 1 (for simplicity, “average” is used in the table to include mode and mean). The first encounter occurs any time after the beginning of entering service, averages six months after the beginning, and has an exponential distribution. Subsequent encounters take place uniformly between six and seven months after the previous encounter. The text requires between 1.6 and 4.8 KB, with 2 KB the most likely value of a triangular distribution. The image size depends on the type of X-ray imagery required and varies between 1 and 3 MB with a triangular distribution.

The encounters for the eye doctor (random selection for need) and gynecologist (for women only) are similarly simple. The minor illness and minor trauma encounters are also simple, only adding separate distributions for each soldier group. Because the first physical is a guaranteed encounter, it is used to add the PIC text data overhead.

The data for major trauma (Table 2) and illness encounters illustrate the requirements for more complex encounters. The time before the first encounter depends on

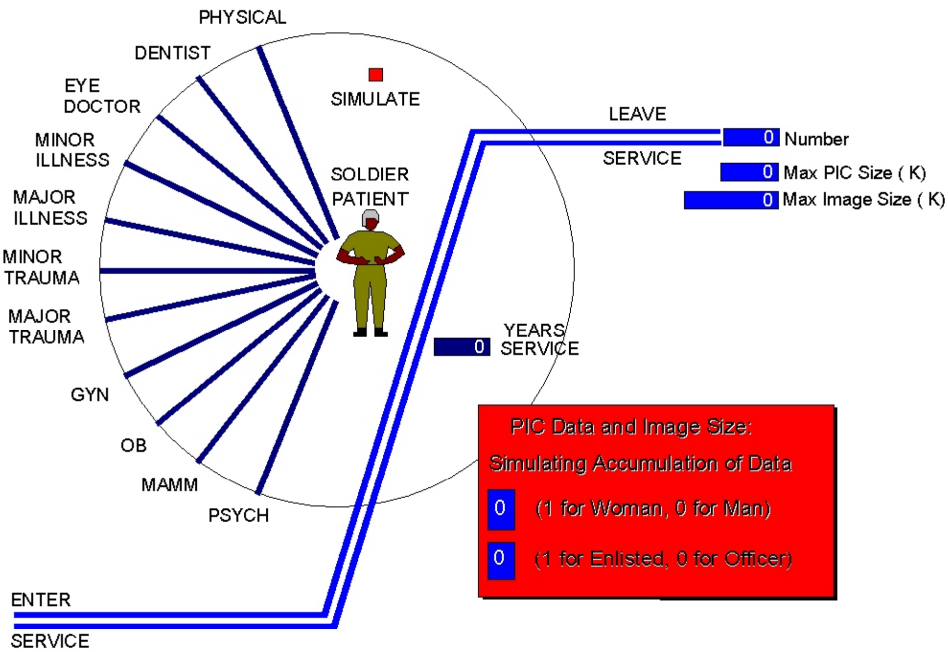

Figure 2: Model Logic

Table 1: Encounters with a Dentist

DENTIST	Distribution	Parameters (as appropriate)			
	Type	Low	Average	High	Std Dev
Time before first encounter	exponential		6		
Interval between encounters	uniform	6		7	
Data size distribution	triangular	1.6	2	4.8	
Image size distribution (if any)	triangular	1024	2048	3072	

Table 2: Major Trauma Encounters

MAJOR TRAUMA (follow-up visits)				Distribution	Parameters (as appropriate)			
				Type	Low	Average	High	Std Dev
Time before first encounter								
Men	Officer			exponential		12		
	Enlisted			exponential		12		
Women	Officer			exponential		18		
	Enlisted			exponential		17		
Interval between encounters								
Men	Officer			exponential		180		
	Enlisted			exponential		120		
Women	Officer			exponential		204		
	Enlisted			exponential		192		
Data size distribution				triangular	1.7	5.1	8.6	
Image size distribution (if any)				triangular	20480	40960	81920	
Follow-up visit data								
	Total # of follow-ups			Percent of encounters with that # of follow-ups				
	1			5.0%				
	2			45.0%				
	3			40.0%				
	4			10.0%				
Delay between follow-ups				triangular	0.2	0.25	0.5	
Data size distribution (follow-up)				triangular	1.7	5.1	8.6	
Image size distribution (if any)				triangular	0	40960	61440	

the soldier's group, as does the interval between encounters. Women may be expected to have major traumas less frequently than men and enlisted personnel to have them more frequently than officers. The text sizes are small, but larger than for dental encounters. The image sizes are quite large, representing the potential requirements for multiple X-rays.

The requirement for follow-up visits (the distinguishing characteristic of **major** traumas) introduces a discrete distribution for the number of follow-ups. Additional data are required to define the intervals between follow-ups and the additional text and imagery required for each.

Some of the encounters presented interesting data problems. For example, mammograms are not recommended until the (female) patient reaches the age of 40. Thus the first encounter must be calculated using an inference on the age of the soldier on entrance to service. Because officers begin service after college (usually) and enlisted personnel begin after high school (usually), the first encounter for officers is set four years later than for enlisted personnel. A normal distribution with a 12 month standard deviation is used to represent the considerable variation in the time of the first encounter. These encounters are significant because four images are made, each requiring 40 MB of storage. The large storage is necessary to support the very high definition needed for digital mammography.

Encounters with an obstetrician are at the other end of the service life: most women bear children before they are 30. To support this, child bearing is represented with a maximum number of batches, with the value drawn from a discrete distribution representing the probability of zero, one, two, or three children for the particular (female) soldier. Three follow-ups are scheduled: the delays are fixed at three months, except for the last interval, which is triangular, between 2.5 and 3.5 months. Sonogram imagery may be required.

These and all other data were selected to put moderate stress on the size requirements, while not exaggerating the requirements excessively. That is, they were chosen to be larger than the guessed at values, but not by much.

6 RESULTS

The simulation was run separately for each group, with totals kept for the text size requirements and for the image size requirements. The number of replications was varied until it became clear that the maximum requirement had been observed sufficiently frequently for a meaningful cumulative distribution to be drawn.

Figures 3 and 4 represent the general results and show the frequency (bars) and cumulative (line) distributions for female officers (1000 replications). The vertical scale varies from zero to one, representing the fraction of patients requiring a given size of PIC. The horizontal scale shows the size.

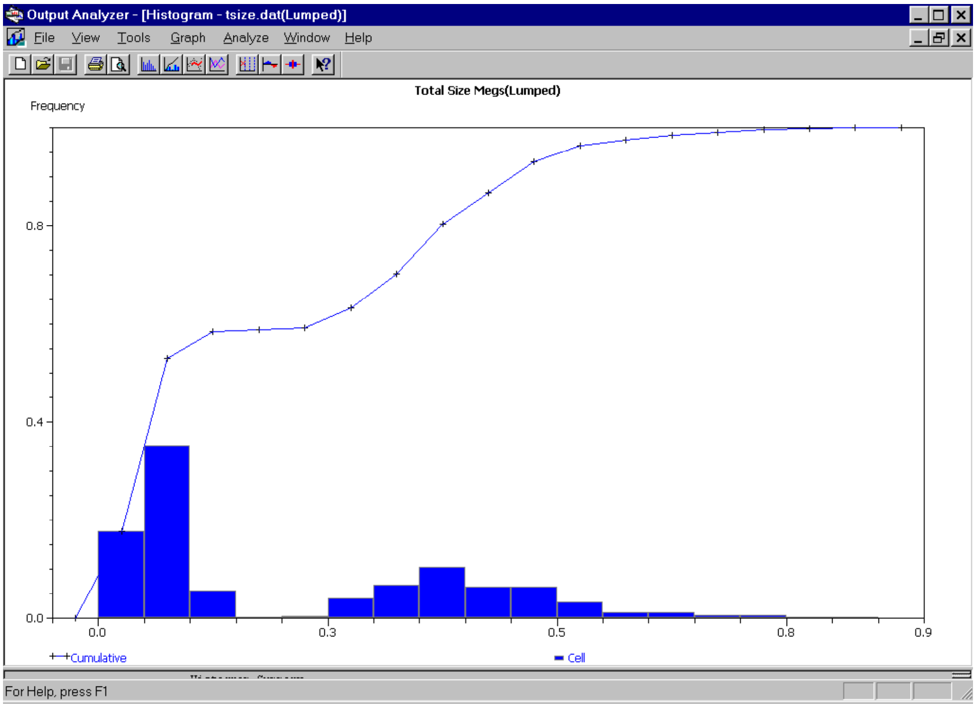

Figure 3: Distribution of PIC Text Size Requirements (Meg=MB)

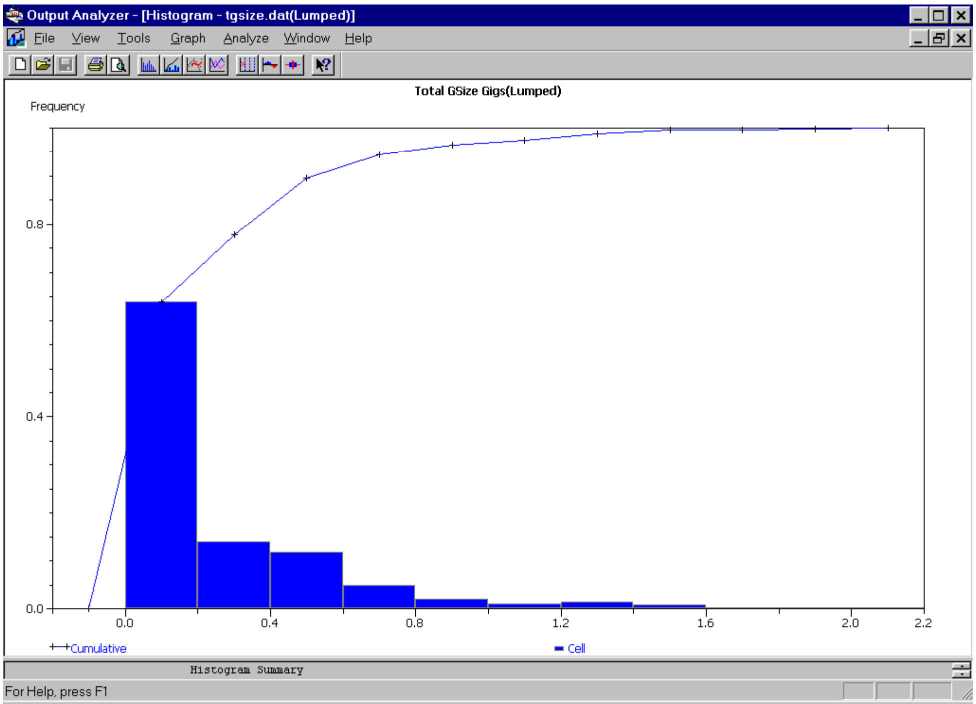

Figure 4: Distribution of PIC Image Size Requirements (Gig=GB)

Figure 3 shows that a one MB PIC is sufficient to hold all text requirements (scale zero to 0.9 MB). Figure 4 shows that two or more gigabytes (GB) are required to hold all of the imagery (scale zero to 2.2 GB). Clearly, a different technology is required for both complete text and imagery. Simple sensitivity analysis showe that the differences in magnitude between text requirements and imagery requirements are such that validated data could not be sufficiently different to change the results: Doubling the text requirements would still leave the text size within the PIC capacity and halving the imagery requirements would still leave the size beyond current PIC capacities.

7 CONCLUSIONS

The conclusion for the PIC investigation is that the current technology will not support soldiers carrying their complete medical records on their persons at all times. The data storage requirements are in the gigabyte range, with the technology only supporting the megabyte range. However, a modified requirement is possible. Table 3 shows that a 15 MB PIC could contain all text records, plus sufficient imagery for identification of a soldier.

Table 3: Proposed PIC Contents

Item	**Size**
Complete medical text records	1 MB
Identification picture	1 MB
Most recent dental X-rays	3 MB
Digital fingerprints	10 MB
Total	**15 MB**

The problem did not require a discrete event simulation to solve, for example spreadsheet tools would have sufficed. However, the original solution envisioned the availability of validated data and the results were not known to be as clear-cut as the analysis showed. In addition, the animation provided by Arena (Figure 2) provided visual impact in discussions with the sponsor, an important factor in real applications. Further, more detailed investigations would require the power of a discrete event simulation. For example, a data requirement that restricts other imagery to the "most recent" category might be supported by a PIC size of 60-100 MB. Validated data would be required to support this possibility.

This example clearly supports the conclusion that "simulating without data" may be sufficient to solve a particular problem. At the very least, simulating with pseudo-data provides an order of magnitude result. It drives the search for an understanding of the problem, such as defining the relevant data factors and determining the appropriate forms for input data distributions. And understanding the problem provides a clear starting point for creating a solution with validated data.

ACKNOWLEDGMENTS

This document was prepared based on the prior work of the author while employed at the Department of Energy facilities in Oak Ridge, TN. Thanks are due to Ms. Barbara Beckerman, who led the project for Oak Ridge support to TATRC.

REFERENCE

Kelton, W. D., R. P. Sadowski, and D. A. Sadowski. 1998. *Simulation with Arena*. Boston, MA: WCB McGraw-Hill.

AUTHOR BIOGRAPHY

DEAN S. HARTLEY III is the Principal of Hartley Consulting. Previously he was a Senior Member of the Research Staff at the Department of Energy Oak Ridge Facilities (Oak Ridge National Laboratory, Y12 Site and East Tennessee Technology Park). He received his Ph.D. from the University of Georgia in 1973. Hartley is a Director of the Institute for Operations Research and Management Science (INFORMS), a past Director of the Military Operations Research Society (MORS), past President of the Military Applications Society (MAS), and a member of the College on Simulation of INFORMS. Hartley's interests include modeling of combat and operations other than war (OOTW), verification and validation of models, psychopharmacology modeling, and simulation. His email address is `<dshartley3@comcast.net>` and his website is `<mywebpages.comcast.net/dshartley3>`.