Health Informatics Journal

http://jhi.sagepub.com/

Why are we waiting? How poor information systems increase hospital queues

Reeva Lederman

Health Informatics Journal 2002 8: 147
DOI: 10.1177/146045820200800304

Why are we waiting? How poor information systems increase hospital queues

Reeva Lederman

In this paper, we examine the data delivery process in the neurology department of a large public hospital and consider whether information systems failure contributes to queuing in hospitals. This is a particularly important issue given the pressures worldwide on governments to provide funding to reduce hospital queues. What we find in this research is that, where data are lost or delayed in a system, queues for service increase. In systems where data are distributed manually or in a semi-automated fashion, a search for lost data can extend beyond one week. This can cause service queues to increase disproportionately to the time spent searching for the data and have a dramatic impact on all stakeholders in the hospital system.

Keywords
Data loss, image-based systems, queueing theory, simulation

INTRODUCTION

> Timely management of medical imaging information is one of the greatest challenges facing medicine today . . . As medical centres increase in size, illnesses increase in complexity and the demand for rapid transfer of information increases accordingly, the capacity of film-based radiological systems to meet these demands decreases. Films are often unavailable or lost and film storage costs are high. [1]

The push to improve database technology in hospitals has resulted from well-documented problems of data loss and delay in data delivery throughout the health system [1]. In non-automated and semi-automated database systems, where some of the processes are performed manually and some are implemented automatically, timely access to data and the simultaneous receipt and analysis of more than one data item at a time cannot be achieved. Such systems, where there is significant delay in file transfer and a loss or misplacement of files, can lead to delayed decision-making by medical staff as well as lengthened hospital stay and increased administrative workloads.

Recent reports on government expenditure in the healthcare area detail the anger expressed by healthcare professionals that significant funds are being earmarked for the development of hospital databases at a time when staff numbers are being cut, hospital beds closed and queues for treatment increasing. Members of medical and nursing bodies feel that funds would be better spent in ways that will provide direct improvements in medical care [3]. This response mirrors similar reports on patient queuing and treatment delay in Canada [4] and other developed countries, and highlights the difficulties for decision-makers in deciding how to best spend the scarce resources allocated to the public hospital system. My research suggests, however, that judiciously spent funds in the information systems area can, in fact, reduce hospital queues.

In the patient episode under examination in this research, a patient enters a system where they are to undergo a neurological scan (e.g. PET or MRI) which is captured and stored. The stored data are then transferred to a medical consultant on request in time for a scheduled patient consultation. The number of completed consultations that occur is dependent on consultants receiving complete and timely data. Often, however, data are lost or missing and patient consultations are delayed in consequence. The extent to which there is throughput of the service mechanism depends on the number of negative events (instances of lost or missing data) which send patients back into the service queue and may require them to be reprocessed along the path back to the consultant.

The search for available patient data begins in the first data repository, the central file area (CFA), where it is delivered after initial processing in the PET or radiology departments. If the data are found immediately, the data moves straight from the CFA to the doctor in time for a patient consultation and the patient is seen and moved along the workflow process. If data cannot be located in the file store, a sequence of searches take place as in Figure 1, which shows the possible paths that a request for a file might follow. If the data are not in the CFA the search continues through the PET and radiology centres. If the data are still not found, then the PET and radiology archives are searched.

If the data are not available at any earlier point in the information search process shown in Figure 1, the patient may need to be rescanned and end up re-entering the workflow at the first

R. Lederman BA MIS
Department of Information Systems
University of Melbourne
Victoria 3010
Australia
Tel: +61 3 8344 0867
Fax: +61 3 9349 4596
Email: reeval@staff.dis.unimelb.edu.au

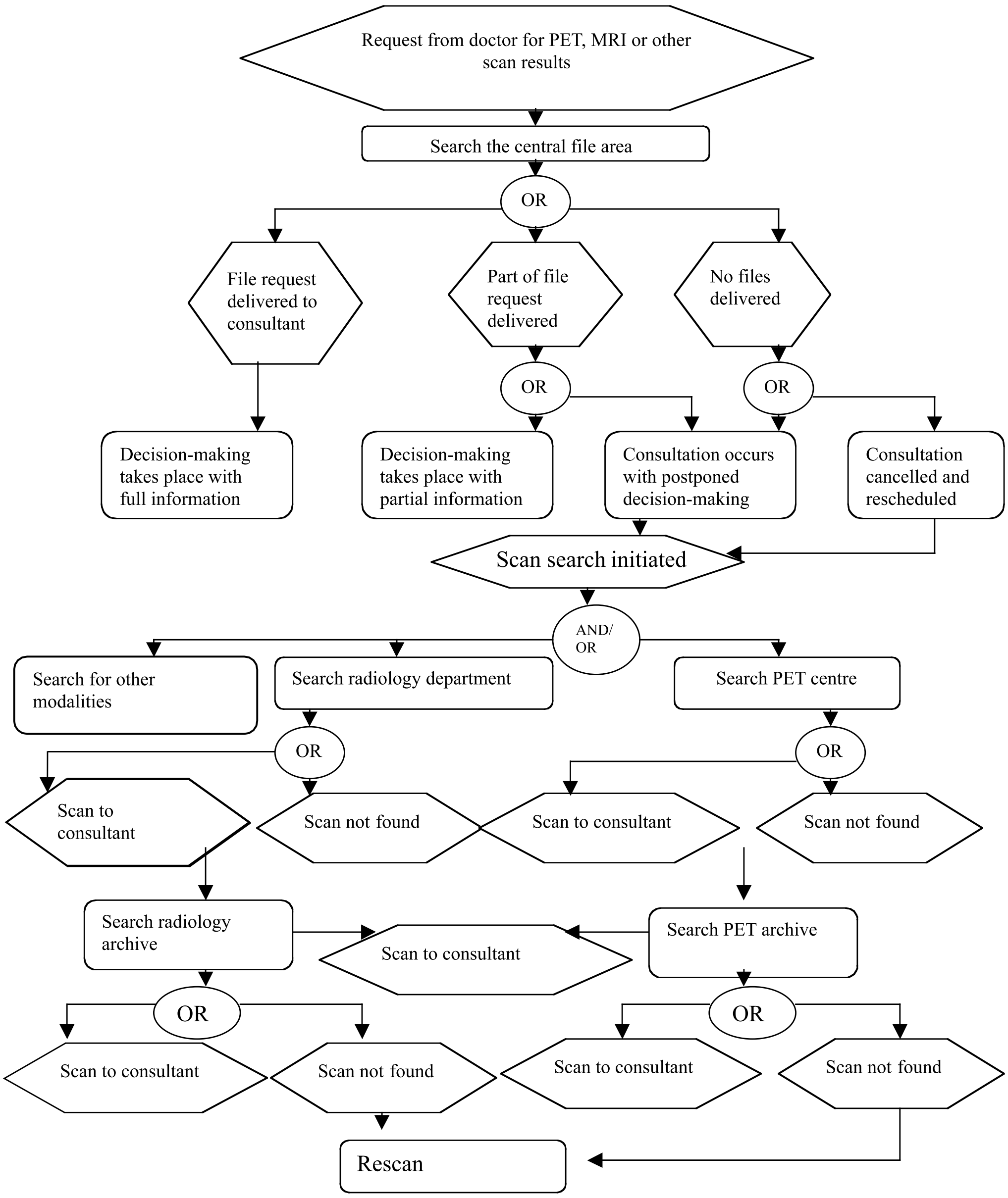

Fig. 1 Information search process

consultation point. This increases queuing at this point. Additionally, the instance of data loss causing the hospital to reprocess these patients increases queuing in the system as a whole and adds to delay in the completion of the individual patient episode.

Queues for service occur at all levels in the hospital process. A patient may need to queue for his or her original appointment with a consultant and for a follow-up consultation. Patients may also queue a number of times to have a number of scans taken. These queues are to a large extent anticipated and the system accounts for them by setting a benchmark waiting time (usually seven days) between consultations, which compensates for the average delay in information flow as well as other internal process delays. These waiting times extend the patient episode, but can mediate against the effects of queuing by giving the system known demand and scheduling parameters to work within. They do not, however, compensate for instances when data goes missing and has not been returned by the benchmark consultation time. In this case, a patient loopback will occur and the patient will be rescheduled or added to the queue for service at some earlier point in the service mechanism.

It will be seen that in a busy hospital system, the queues resulting from patients being returned for reservicing can grow disproportionately to the number of requests for patient data that are not being satisfied at each service point. What concerns us is how the spread of the service time changes and what the maximum length of a patient episode might become as the number of patients requiring reservicing increases, and what the implications of this might be for patient care.

From a queuing perspective we see the

negative impacts of data loss which causes patients to be returned through the system for reservicing. We can extrapolate the following outcomes for the relevant stakeholders and use the queuing theory to consider the extent of their impact:

- from the point of view of the hospital system overall, queuing can significantly increase the length of stay for individual patients and is thus a burden on resources
- queuing costs the hospital department the consultant time taken in repeat consultations
- queuing costs the hospital department the administrative time taken in rescheduling repeat consultations
- queuing wastes the time of the consultant in reservicing individual patients who are returned to the queue
- queuing costs the patient in increased patient episode time.

What we will see from the queuing theory is how, if all of these impacts are magnified disproportionately, the greater the likelihood of data rejection at each point in the service mechanism.

THE SCAN PROCESS

A successful route through the service mechanism for a patient undergoing a scan in the neurology department can be represented as a series of steps:

1. Patient enters the system.
2. Patient undergoes a scan process.
3. Patient makes an appointment to see a consultant.
4. Consultant requests relevant patient scans to be found and delivered to the consultation (from the CFA).
5. Scans are retrieved and delivered.
6. Patient is seen by consultant.
7. Treatment decision is made.
8. Patient continues through the system.

If information is unable to be retrieved in step 5, then it will be unavailable at step 6 and the patient may continue to move back to step 3 or even repeat step 2 until a final meeting where the doctor is satisfied with the amount of information available and is able to proceed to step 8. Where step 7 cannot be completed, queues will begin to develop at the earlier steps as the patient re-enters the service loop.

Returns and loopbacks in the system can make both the information and patient flows irregular, and we will see that they can extend the patient episode beyond the in-built benchmarked delays disproportionate to the additional service time required to see one extra patient twice.

It is the semi-automated nature of the process that prevents information being available to the consultant. While scans are taken in digitized form, the process for transferring them around the hospital is largely manual. Scans are generally stored in hard copy form with one copy being shared across the system by nurses, doctors, radiologists and any other professionals who may have an interest in a particular patient. While users are supposed to return all scans to the central file area, they do not always do this in a timely fashion and, in fact, the unreliable nature of the system encourages people to hold on to patient data rather than returning them, increasing opportunities for data to be lost and for delays in data transfer to occur.

In the information search process we see in Figure 1, there is an opportunity for data to be lost or its transfer delayed at almost every point in the process. In the best system, delay will be linear proportional. That is, as the arrival rate of patients increases, delay will increase proportionately. However, because of loopbacks in the patient flow, the situation can become irregular. That is, if you double the effective arrival rate of patients, by increasing the rate of rejection at each service point, the queue will more than double and the delay will increase non-linearly.

In a system where we have a slotted queue, that is, appointment times are of a fixed length, we can use the theoretical model of Bertsekas and Gallager [5] to illustrate the effect of patients being returned to the loop as a result of data delay or rescan. In their analysis, the time slots were occupied by computer data which may be retransmitted when errored.

We can compare this to our situation where patients are turned back at various points in the service regime due to error (data not being found) and are required, similarly, to re-enter the queue and service mechanism.

Under this model we can show that, as the probability of service rejection, p, increases, both the mean and its variance (second central moment) increase at a greater rate than p. We can derive values for the delay and variance relative to the benchmark of a $p = 0$ service model, where there is no possibility of service rejection, that we see in Table 1.

The queuing theory suggests that with finite p there is both a lengthening of the service time and a disproportionate lengthening of the expected range of service times compared with the standard. That is, as the probability of rejection increases, the average and variance in the

Table 1 Range of values for a given probability (p) of service rejection

Probability of service rejection	Average service time	Variance	Standard deviation
0	1	0	0
.2	1.25	.3125	.56
.4	1.66	1.111	1.05
.8	2	20	4.47

service time increases at a greater rate. The lengthening of the service time implies a lower overall throughput of successful treatments per unit time. In our system, then, we might expect that having to look for data in a second data repository, where the chance of finding it is reduced, might disproportionately increase the delay for patient service.

Under our model, patients are generally recalled by the consultant after one week and then possibly recalled a second time if data are not available at the first consultation. If data are not available at the second appointment then it is likely a rescan will be ordered or the consultant will take steps to locate the data rather than continue to wait indefinitely.

If we demand that patients are serviced within one week we can apply the Bertsekas and Gallagher [5] model (where the chance of data return after one week is $(1-p) + (1-p)p$), and find, given specified rates of rejection, the percentage of patients who will be serviced in the specified time (Table 2).

What we are seeing in Table 2 is that, even with a reasonably low probability of a request for patient data being rejected (.3), nearly one out of every 10 patients will not have their data present if they return to see the doctor after the usual one week benchmarked return time. With higher rates of data delivery failure this increases to more than six out of 10. The implications of this for patient throughput and the completion of the patient episode are enormous.

However, in a busy public hospital, a 100 per cent increase in the service time might be considered tolerable, and in attempting to improve the service time we may only be concerned with the percentage of patients that complete service outside that benchmark. If it is decided that a service time of up to two weeks in length (a 100 per cent blowout) is tolerable, we can examine what percentage of patients will be serviced in this time. We can derive the following table of values for a service time of 2 (2×1 week) or a 100 per cent increase in service time ($(1-p) + (1+p)p + p^3$) (Table 3). The values in Table 3 illustrate the impact of patients returning to the loop and the difficulties which result for the system as data loss increases and patient returns become more prevalent.

While we can see that the likelihood of data being available increases significantly after two weeks, if we were to schedule patient appointments at longer intervals (that is, two weeks instead of one week), we dramatically increase the patient episode having a possibly consequential effect on the patient outcome. Conversely, if we were to decrease the standard time between consultations to less than a week, other considerations would arise. After three or four days patient data may have returned for many patients and, if they were seen earlier than one week, then the length of their patient episode may be decreased, treatments may begin sooner and positive effects may result. However, at a shorter time interval there would be a greater rate of data unavailable, more consultations needing to be repeated, and more use of administrative and consultant time trying to track unavailable data.

What we see in effect is the need for a trade-off between increasing the likelihood of data being available and decreasing the length of the average patient episode. With an online database where data was always available as soon as the scan was completed, this type of trade-off would be unnecessary.

THE SIMULATION APPLICATION

Having considered the possible impact of data loss or significant delay, we need to look at

Table 2 Percentage of patient data files returning within one week given specified probabilities of service rejection

Probability of service rejection	Percentage of patients serviced within 1 week
0	100
.2	96
.3	91
.4	84
.8	36

Table 3 Percentage of patient data files returning within two weeks given specified probabilities of rejection of a file request, leading to reservicing

Probability of data rejection	Percentage of patients serviced within 2 weeks
0	100
.2	96.8
.3	93.7
.4	90.4
.8	87.2

how often such losses might occur under the current semi-automated data delivery system. To examine the impact of queuing in the hospital model with the scan search process described in Figure 1, we created a simulation of the system under examination. From this we are able to compare the likely effects of delay derived from the queuing model with the actual rates of delay observed in the simulation.

Using simulation of the information flows derived from observation and data logging in a neurology department, we examined the impact of data loss on 1,000 patients by tracking the associated computer simulated flows. We estimated the likelihood of data being available at the various nodes in the service chain.

The simulation was built following extensive consultation with the service providers, the consultants. Both the process flow and associated probabilities and delays were extracted from interviews. The simulation results were tested for statistical validity and for consistency against the experience of the consultants. To create the simulation application, we worked out what tasks were required when a consultant sought two separate pieces of data and where the significant points of activity in the process occurred.

Requesting different scan modalities, such as a PET scan and an MRI scan, compounds the possibility of delay, although it is common hospital practice for a request for more than one data mode relating to a single patient to be made at the one time.

Evidence gathered during meetings with hospital staff suggested the possible pathways for a data request through the tasks that we see in Figure 1. At each data repository in the search there is a given probability of data being found and returned to the consultant. If data are not found the search continues, with an average time being allocated to every stage in the search process (see Table 4).

A request for a single data mode (such as PET alone) carries with it a given probability of being delayed at all the points along the service chain (Table 4). When a request for multi-modal data (for example both PET and MRI) is made, all scans will need to travel along their independent paths back to the consultant and may suffer their own separate degrees of delay. Thus if a consultant is waiting for two scans and needs both scans to make a treatment decision, the wait will be equal to the time it takes for the longest delayed scan to be returned.

Table 4 Time expended in moving through the process at each date repository

	Time to process (days)	Standard deviation	Probability of data being found at this point
Go to central file	1.5	1	
Search central file	2	1	.8
Search PET centre	3	1	.7
Search radiology	3	1	.7
Search PET archive	2	1	.95
Search radiology archive	2	1	.95
Rescan	20	5	

RESULTS OF THE SIMULATION

What was modelled in this simulation was a request covering two data modalities, an MRI and a PET. We sought to examine the spread or dispersion of the times of return of the information sought. In many cases the two scans would have followed different paths and arrived at different times.

In Table 5 we see the percentage of files likely to have been returned to a consultant after seven days. In order to check the statistical significance of the results in Table 5 we calculated, from multiple simulations, the standard error of the difference for the two sample means from Table 5. The observed difference is highly significant. Therefore we have a high level of confidence in both means and the percentages achieved.

DISCUSSION OF THE RESULTS OF THE SIMULATION

As presented above, using the figures from Table 5 we could estimate that if a consultant saw a patient approximately one week after their scans were done, then in 93.78 per cent of cases the earlier scan would have arrived and in 68.26 per cent of cases the later one would also have arrived.

In this we have presumed a request for two data modes, so if a doctor wanted to use this data to determine a treatment plan after one week, it is likely that in more than 6 per cent of cases he would have no data on which to do so, and in over 31 per cent of cases he would have only one piece of data. We note that while the

Table 5 Percentage of files returned to consultant after 7 days

7 days elapsed	Percentage of files returned to consultant
Percentage of cases where one data item was available	93.78
Percentage of cases where both data items were available	68.26

benchmark 7 days waiting time is sufficient to satisfy all but 6.22 per cent of cases in practice, there is significant staff time spent locating data close to the benchmark time to try to achieve or lower this figure. It is presumed that these 6.22 per cent of patients for which there is no data available will re-enter the queue for servicing as will many of the 31.74 per cent for whom one data item is available.

What we see here is some manifestation of the queuing phenomena that we anticipated from the earlier stated theory. We can see in Table 4 the time taken for individual search requests to be processed. We can see that it is likely that more than 70 per cent (where p = .7) of files will be found by the time that the search moves to the PET or radiology centres and by this time, for an individual file, only 6.5 days should have been expended (one day to go to the CFA, two days to search the CFA, two days to search the PET centre). We see in the simulation, however, that after seven days nearly 32 per cent of requests for a second data item are unfulfilled when the search is simulated for 1,000 data requests. What is happening here is that, as data requests rejoin the queue for processing from the beginning, the time for fulfillment of data requests increases disproportionately.

While only a small number of data files are significantly delayed, the service queues that result have an extreme effect on the delivery of data, and, in a large proportion of cases, one data item, in a request for two, is delivered after the benchmark consultation time. This makes decision-making difficult for consultants and can cause treatment to be either further delayed or implemented without the consultant having full information.

In reality, doctors often order many more than two data modalities, and the more requests made at one time the higher the likelihood that one or more data items will not be returned by the benchmarked time, and the higher the feedback delays across the system as a whole.

CONCLUSION

In conclusion, the delays that can be anticipated from queuing theory are in fact apparent in the service mechanism in semi-automated systems such as the one under examination. Where consultants rely on timely data for decision-making [2], this delay is likely to have a negative effect on patient care and on patient queues for service, and gives rise to calls for further investment by hospitals in the type of database technology that might help remedy such problems. While it is already well known that fully automated systems provide a range of benefits not dealt with in this article, including more efficient use of ancillary staff time and the freeing up of massive storage areas from the demands of hard-copy data, further research into the extent to which they can reduce queuing leading to enhanced patient care would be most worthwhile.

REFERENCES

[1] Choplin R H, Boehme J M, Maynard C D. Picture archiving and communication systems: an overview. *Radiographics* 1992; **12**: 127.

[2] Chabriais J, Aubry F, Balleyguier C, Todd-Pokropek A, Menu Y. Towards the integration of the digital medical image folder within the computerized patient folder: PACS and image networks. *Revue de Pneumologie Clinique* 2000; **56**(2): 140-55.

[3] *The Age*, 16 May 2001: 1.

[4] McDonald P, Shortt S, Sanmartin C, Barer M, Lewis S, Sheps S. Waiting lists and waiting times for health care in Canada – more management!! More money?? *NHRDP*, July 1998.

[5] Bertsekas D P and Gallager R G. Data networks. Englewood Cliffs, NJ: Prentice-Hall, 1987.