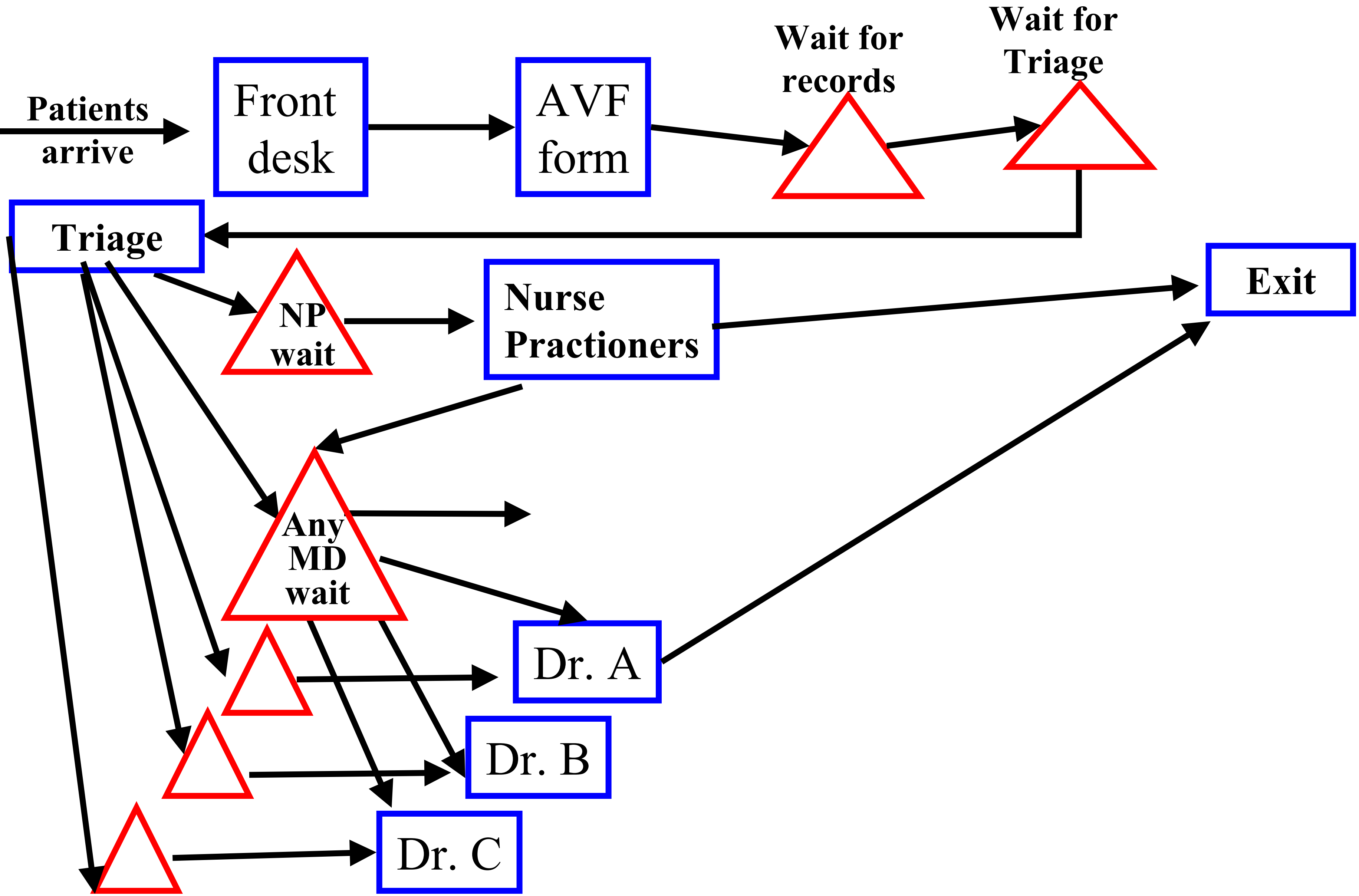
15.760: University Health Service

1. Admin:
   Memory Jogger, The Goal, Mike Hammer
2. What are the sources of variability at UHS?
3. What are the problems UHS is experiencing that should be addressed?
4. Describe the Process Flow Diagram.
5. How to usefully model this system?
6. Utilization and waiting for nurse practitioners.
7. Utilization and waiting for physicians.
8. Recommendations

UHS Process Flow Diagram



sources of variability

- Patients arrival rates by hour, day, week and season.
- Number of MD's on duty - by time of day, day of week.
- Number of NP's on duty - by time of day, day of week.
- MD service rates by Doctor and patient ailment
- NP service rate by nurse and patient ailment.
- Triage service rate.
- Triage coordinator allocation to MD - NPs.

Others

- # of patients wanting to see a special NP or MD,
- by hour, day, week.
- Speed of filling out AVF forms.

Patient Arrival Rates

The clinic is opened to patients from 8:00AM to 5:30PM, a total of 9.5 hrs. Staff are asked to stay until 6:00 PM, So the clinic serves patients 10 hours per day.
For simplification,
we will assume that patients arrive over the 10 hrs.

Patient average arrival rate =
λ =(143 pat/day)/10 hrs/day = 14.3 pat/hr

8 - 9:00 AM on Monday mornings is a peak hour.

Patient peak arrival rate = λ =
14.3 [average rate] X (163/143)[the Monday factor]
X (18.2/14.3)[the 8-9 AM factor] = 20.7 patients/hr

ASSUMPTIONS OF THE QUEUEING MODELS

- Poisson arrivals/exponential service times
- steady state
- $\rho < 1$, when computing the queue lengths and waiting times
- Constant # of servers
- FIFO service
- Single-line queue (to MD's)
- Infinite queue capacity
- Ignore special priority emergencies
- Ignore special priority requests

Triage and arrival rates

Let's interpret the second column of data in Exhibit 6
as follows: 28% of patients are triaged to the NP queue,
48% are triaged to the queue for first available MD,
and 24% queue up for a specific MD.
(I.e., let's assume the fraction of patients that request
to see a specific NP is zero.)
However, since 5% of patients seen by NP's get sent
on to MD's (bottom of p. 4), these patients wait twice
and are seen twice, so the total load on the system
might be thought of as greater than 100% of initial demand.

Therefore, let's assume that the percentage of patients
who see an MD = 24% + 48% + [(28%) x 0.05] = 73.4% of
initial arrivals and the arrival rate to the NP queue
is 28% of initial arrivals.

Arrival Rates

Patients average arrival rate to an NP

Average over week,
λ = 14.3 pat/hr (28%) = 4.0 pat/hr.

Average over peak hour
λ = 20.7 pat/hr (28%) = 5.8 pat/hr.

Patients average arrival rate to an MD

Average over week,
λ = 14.3 pat/hr (73.4%) = 10.5 pat/hr

Av. over peak hour λ = 20.7 pat/hr (73.4%) = 15.2 pat/hr

Number of Servers

(From exhibit 4)

Weekly average number of NP servers = 3.15

Average number of NP servers during peak Monday hour = 2

Weekly average number of MD servers = 2.9

Average number of MD servers during peak Monday hour = 2 *utilization rate, length of queue and waiting time* for the MD's and NP's.

Sadly, as in real life, data and models are imperfect. As a result, one must often make decisions with admittedly imperfect understanding of the systems under study. The data in the case permit a range of analytical approaches, two of which are illustrated below.

The first method uses as inputs the expected service time, patient arrival rates and number of servers given in the case to compute utilization rates. The second method uses an inputs waiting time and patients arrival rate from the case to compute the queue length using Little's formula. Then, the queue length combined with the number of servers gives the utilization rate using table A2.

METHOD #1

M, λ, S (from case)	r by definition, $\rho = \lambda M/S$
ρ, S	L from Table A2.
L, λ	W(*) with Little's formula, $W = L/\lambda$
	(* compare with case data)

Method #2

W, λ	L with Little's formula, $L = \lambda W$
L, S,	ρ from Table A2
ρ, λ, S	$M = \lambda/\rho S$ by definition

NP Calculations

A. Method #1

Average for the Week.

M = 32.8 min/patient (Ex5) = .547 hrs/pat

S = 3.15 (Ex4)

λ = 4.0 patients/hour (above page)

$\rho = \lambda M/S$ = (4.0) (.547)/3.15 = .695

From table A-2, we have L = 1.15 patients

W = L/λ = 1.15 pat/4.0 pat/hour = .29 hrs = 17 min

Peak Hour

M = 32.8 min/pat (above) = .547 hr/pat

S = 2 (Exhibit 4)

λ = 5.8 patients/hour (above page)

$\rho = \lambda M/S$ = (5.8)(.547)/2 = 1.58

Method 2: Nurse Practitioners

Average for the Week

$W = 6.7$ min $= .112$ hrs

$L = \lambda W = 4.0$ pat/hr$(0.112$ hrs$) = .418$ patients

$S = 3.15$ …from Table A2, $\rho = 0.57$

$M = S\rho/\lambda = (3.15)(0.57)/4.0 = .449$ hrs/pat

Thus the service rate is $1/(.449) =$

2.23 pat/hr and $\rho/M = .57/.449 = 1.27$ pat/hr,

Peak Hour

$S = 2.$

$M = .449$ hrs/pat, since we assume the same service rate as during regular hours, i.e., their service rate is independent of the patients arrival rate.

$\rho = \lambda M/S = (5.8)(.449)/2 = 1.30$, again greater than one, so inventory builds up.

Method #1 : MD Calculations

Average for week

M = 19.4 minutes/patient = .323 hrs/patient
S = 2.9 (exhibit 4)
λ = 10.5 pat/hr
$\rho = \lambda M/S$ = 10.5(.323)/(2.9) = 1.17, implying that queues build all day. This is suspicious since there is no indication that some patients camp overnight or are sent home. Perhaps:

- **triage sends patients to NP's when the MD's are busy.**
- **Staff stays late**

Peak Hour

M = .323 hrs/patient, S = 2
λ = 15.2 pat/hr, ρ = (15.2)(.323)/2 = 2.45 !!

MD CALCULATIONS: Method #2

Average for Week

W = 25.2 min = .42 hrs

L = λW = 10.5 pat/hr (.42 hr) =>

4.4 pat in queue.

Using S = 2.9 and L = 4.4 in Table A-2 gives

ρ = .86

M = Sr/l = (2.9)(.86)/10.5 = .238 hrs/pat

Thus the service rate is 1/.238 = 4.21 pat/hr

and ρ/M = (.86)/.238 = 3.61 PATIENTS/hr,

neither of which is very close to the reported value of 3.1 patients per hour.

LEARNING POINTS

1. Queueing models do not predict the customer waiting times with great accuracy.
This could be because one or more of the assumptions underlying the model was seriously violated, or it could be that inaccurate data was reported in the case.

2. Analysis suggests that the MD's were more heavily loaded than the NP's, and that the entire staff was heavily loaded during peak hours.
3. The models provide a structure for thinking about the operating system. The formulas $L = \lambda W$, $\rho = \lambda M$, and $W = \lambda (M^2 + \sigma^2)/2(1-\rho)$ do stimulate thought about the important parameters of the System, their relationships with each other and the decision variables available to Ms. Angell.