UCLA HLTADM 413 Operations Management in Healthcare

Homework Module #2

Systems and Process in Healthcare

You are the operations management analyst that supports operational decisions in your organization. There are 3 issues that have come up for which the leadership team needs your analytic expertise. The first problem centers around diagnostic technology investment decisions. The second focuses on analyzing the timing by hour of the day of hospital admissions and discharges. The third entails showing the impact of virus spread if one can "Bend the Curve."

The grading of the Excel analysis will be based on your arriving at the correct answers in the cells indicated. Even if you arrive at the correct answers in the cells, full credit will not be given if you hardwired in numbers into your formulas, rather than inserting the appropriate cell references.

You will be submitting an Excel file and a Word document.

Point allocations for the problems are:

1. Problem 1: 15 points
    a. 5 points for the write-up (put the writeup in the text box in the Excel worksheet)
2. Problem 2: 15 points
3. Problem 3: 20 points
    a. 5 points for the write-up (put the write-up in a word document)

Note that in turning in your homework, you are asserting the following:

- ✓ The answers represent your own work. You did not copy answers from any other students, or access any of this class's previous years' materials to complete this work.
- ✓ You can consult with other students in your class, but they should NOT send you their file, or a part of their file, that you then copy and paste into your own file and claim as your own work, or copy their write-up, changing a few words here and there. If a fellow student offers some help to you via zoom or other online collaboration tool, you did NOT take screenshots of their work they may have shown you.
- ✓ Similarly, if you are being asked for help, you should not email your completed file to the student(s) who is asking for help.
- ✓ To do any of these things constitutes academic misconduct, and will result in an F for the course.

To submit your work:

To submit your work:

1. TYPE YOUR NAME in the Excel worksheet called *Academic Conduct Signature* to verify that the work you are turning in represents your own work.
2. Rename the Excel file and Word file with <yourlastname>hw2
3. Submit both the Excel file and the Word file to Canvas. Do NOT email the files to me.

Problem 1: Diagnostic Technology Investment Decisions

Your organization is considering the purchase of 3 dimensional (3D) mammography screening technology to replace its current mammography technology.

The 3D mammography has been found to have a slightly higher True Positive rate (called the sensitivity of the test), and a slightly higher True Negative rate (called the specificity of the test) compared to the 2D technology, particularly in younger women and women who have dense breast tissue.[1] This means that it will be better at detecting breast cancer if it is present. And, it will decrease the need for unnecessary follow-up testing in women whose screening showed potential cancer in their mammogram, but in fact did not have cancer.

In preparation for assessing the financial impact of this decision, you have been asked to prepare an analysis that calculates the positive predictive value and the negative predictive value of the current and new technologies, assuming the information provided below. These comparisons are one step in assessing the cost effectiveness of investing in new technology.

Diagnostic tests are never 100% accurate. The “Father Guido Sarducci” of diagnostic testing models is that knowing the **base rate**, and the **sensitivity** and **specificity** of the diagnostic test, you can calculate the positive predictive value and the negative predictive value of the test. All of these terms are defined below.

You may be surprised to see that the positive and negative predictive values depend greatly on the base rate of the disease in the population. For example, many people think that if the True Positive rate of a diagnostic test is 90%, then if a person is tested and it comes back positive, then the probability that the patient has the disease is about 90%. But this is NOT TRUE!

This analysis will show you why this is the case.

Determining the Diagnostic Characteristics of a Test

Before diagnostic tests are used in actual practice, they have to be tested to determine how well they differentiate those who have disease from those who do not. This is accomplished by using a population of test cases in which it is already known whether each test case in that population actually does or does not have the disease of interest. So, there are two populations: 1) those that actually <u>have</u> the disease; and 2) those who <u>do not have</u> the disease.

Does the diagnostic test sort out the two populations correctly? That is, does the test come back positive for those who we know have the disease, and does it come back negative for those who do not have the disease?

This results in four numbers of interest:

True Positives and False Negatives in the Population that actually **HAS** the Disease

True Positive Rate or Sensitivity: Does the test return a POSITIVE test result IF THE TEST CASE ACTUALLY HAS THE DISEASE? This is the **TRUE POSITIVE (TP)** rate. It is the probability of a POSITIVE TEST given THE TEST CASE ACTUALLY HAS THE DISEASE. This is also known as the test's SENSITIVITY.

In statistics, it is written as: P(+| Disease) , read as "probability of testing positive given disease."

False Negative or 1 minus the Sensitivity: The test **error** that can occur with the population that actually HAS the disease is that the test says they do NOT have the disease. This is called a **FALSE NEGATIVE (FN)** rate. It is the probability of a NEGATIVE TEST given that the person ACTUALLY HAS THE DISEASE. It is equal to 1 minus the SENSITIVITY.

In statistics, it is written as: P(-| Disease), read as "probability of testing negative given disease."

You can see that both of these metrics are conditioned on HAVING the disease.

True Negatives and False Positives in the Population that actually **DOES NOT HAVE** the Disease

True Negative Rate or Specificity: Does the test return a NEGATIVE test result IF THE TEST CASE ACTUALLY DOES NOT THE DISEASE? This is the **TRUE NEGATIVE (TN)** rate. It is the probability of a NEGATIVE TEST given THE PERSON ACTUALLY DOES NOT HAVE THE DISEASE. This is also known as the test's SPECIFICITY.

In statistics, it is written as: P(-| no Disease) , read as "probability of testing negative given no disease."

False Positive or 1 minus the Specificity: The test **error** that can occur with the population that actually DOES NOT HAVE the disease is that the test says they DO have the disease. This is called a **FALSE POSITIVE (FP)** rate. It is the probability of a POSITIVE TEST given that the person ACTUALLY DOES NOT HAVE THE DISEASE. It is equal to 1 minus the SPECIFICITY.

In statistics, it is written as: P(+| no Disease), read as "probability of testing positive given no disease."

Test Summary

The table below summarizes the characteristics of the test. The two <u>columns</u> sum to 1 because they represent the two populations, those that have the disease, and those that don't. So, each column represents 100% of these two populations. So, the TP and FN probabilities sum to 1, and the FP and TN probabilities sum to 1.

	Has Disease? YES	Has Disease? NO
Test +	TP	FP
Test -	FN	TN
	1	1

The summary of the four probabilities of the test characteristics are:

TP = P(+| Disease) , "probability of testing positive given disease" (called Sensitivity)

FN = P(-| Disease), "probability of testing negative given disease" (equals 1-Sensitivity)

TN = P(-| no Disease) , "probability of testing negative given no disease" (called Specificity)

FP = P(+| no Disease) , "probability of testing positive given no disease" (equals 1-Specificity)

Mammography Diagnostic Testing for your analysis

For purposes of your analysis, you will assume: 1) the current technology has sensitivity of .76 and specificity of .96; and 2) the new 3D technology has sensitivity of .81 and specificity of .98. [2]

Breast Cancer Base Rates (Prevalence)

According to breast cancer statistics, the percentage of women who have breast cancer (its prevalence) varies based on age as shown below:[3]

Age 30 . . . . . . 0.49% (or 1 in 204)	0.49%
Age 40 . . . . . . 1.55% (or 1 in 65)	1.55%
Age 50 . . . . . . 2.40% (or 1 in 42)	2.40%
Age 60 . . . . . . 3.54% (or 1 in 28)	3.54%
Age 70 . . . . . . 4.09% (or 1 in 24)	4.09%

Commented [1]: Alt text:
Age 30: 0.49% or 1 in 204 = 0.49%
Age 40: 1.55% or 1 in 65 = 1.55%
Age 50: 2.40% or 1 in 42 = 2.40%
Age 60: 3.54% or 1 in 28 = 3.54%
Age 70: 4.09% or 1 in 24 = 4.09%

These numbers are called the "base rate." You can see that the base rate (prevalence) of breast cancer depends on a woman's age.

The lower base rate of breast cancer in younger women is one of the reasons why there is a policy debate of the cost effectiveness of regular mammography screening in younger women. When the base rate is low, and the test is not perfect, a large percent of the tests that come back as positive are in fact False Positive results. But, a small number of them will be True Positives. These False Positive results incur additional worry and testing for women. From society's perspective, the tradeoff is finding a few more actual cases of breast cancer at the expense of a lot more women being falsely identified as having breast cancer.

The logic of diagnostic testing

The purpose for conducting a diagnostic test is to determine whether a patient actually has a disease. We don't know ahead of time if the patient does or does not have a disease—that is why the test is being done.

In terms of conditional probability, we want to know the following:

P(Disease|+), read as "probability of disease given the patient tests positive"

Note that this is NOT the same conditional probability as the test sensitivity, which is P(+|Disease).

This is because the P(Disease|+) also depends on the base rate, or prevalence, of the disease in the population. If the base rate is small, most of the people who test positive will be FALSE POSITIVES if the test is not a perfect test.

Given the base rate of disease in a population and the characteristics of the test, there are two calculations of interest:

Positive Predictive Value

The P(Disease|+) is called the POSITIVE PREDICTIVE VALUE. Given the base rate of disease in the population, it is calculated as:

P(Disease|+) = P(+|Disease)/(P(+|Disease) + (P(+|no Disease)) = TP/(TP+FP) in the population

Negative Predictive Value

The P(no Disease|-) is called the NEGATIVE PREDICTIVE VALUE. Given the base rate of disease in the population, it is calculated as:

P(no Disease|-) = P(-|no Disease)/(P(-|no Disease) + (P(-|Disease)) = TN/(FN+TN) in the population.

These numbers are typically multiplied by 100 to express them as percentages.

An Example

Assuming a population of 10,000 women ages 30-39 (assuming the base rate for Age 30 above applies to this age range), how would you complete the table below? What is the Positive Predictive Value and Negative Predictive Value of the current mammography screening for women aged 30-39?

	Has Disease?				
	YES	NO	row total	P(D|+)	P(no D|-)
Test +	TP	FP	Test +		
Test -	FN	TN	Test -		
column total	N disease	N no disease			

Here are the steps:

1. Base Rate: Assuming 10,000 women, and a breast cancer base rate of .49%, there would be 49 women in the population that actually have breast cancer (.0049 * 10,000). Put this in the cell that says N disease.
2. This means that the size of the population without breast cancer is 10,000 – 49 = 9,951. Put this in the N no disease cell.
3. With a test sensitivity of the current mammography screening of .76, 49 of the women who actually have breast cancer would be detected (.76*49 = 37 women). Put this number in the TP cell.
4. This means that 12 women who actually have breast cancer would be incorrectly labeled as NOT having breast cancer (49-37). Put this in the FN cell.
5. With a test specificity of .96, 9,553 of the women age 30-39 in this population who do not have breast cancer would be correctly identified as not having breast cancer (.96*9,951). Put this number in the TN cell.
6. This leaves 398 women who did not have cancer as incorrectly identified as having cancer (9951-9553). Put this number in the FP cell.
7. Positive Predictive Value = TP/(TP+FP) = TP/Test+ = 37/(37+398) = 37/435 = 8.6%. This means that among those who had a positive test for breast cancer, the probability of actually having breast cancer is actually only 8.56%. Given the low base rate of breast cancer in this population, and the sensitivity of the mammogram testing, this positive predictive value is quite low.
8. Negative Predictive Value = TN/(TN+FN) = TN/Test - =9553/(9553+12) = 9553/9565 = 99.88%. Given the low base rate of breast cancer in this population, and the specificity of the test, this negative predictive value is quite high.

Build your model in Excel

1. You knew an Excel model was coming. Click on the worksheet called "Mammography" in the Module 2 homework file. Set up the input data so that it looks like below. You will need to input the correct information in columns B and C. In column C (# women) multiply the assumed population size by its base rate % in column B.

		# womer
Age 30 . . . . . . 0.49% (or 1 in 204)	0.49%	49
Age 40 . . . . . . 1.55% (or 1 in 65)	1.55%	155
Age 50 . . . . . . 2.40% (or 1 in 42)	2.40%	240
Age 60 . . . . . . 3.54% (or 1 in 28)	3.54%	354
Age 70 . . . . . . 4.09% (or 1 in 24)	4.09%	409
https://www.cancer.gov/types/breast/risk-fact-sheet#what-is-the-average-american-womans-risk-of-being-diagnosed-with-breast-cancer-		
Assume Population Size is	10,000	

Current Mammography Test Sensitivity (TP)	76%
3D Mammography Test Sensitivity (TP)	81%
Current Mammography Test Specificity (TN)	96%
3D Mammography Test Specificity (TN)	98%

2. Set up your initial table to calculate the Positive and Negative Predictive Values for the 30-39 year old age group for the current technology. Your numbers should look like below, BUT THEY SHOULD HAVE BEEN CREATED BY USING GOOD EXCEL MODEL BUILDING PRINCIPLES by clicking on the appropriate input cells or building formulas in the cells. Correct numbers that are simply hardwired in the cells will receive no credit.

CURRENT TECHNOLOGY

30-39 YEAR OLDS

	Has Disease? YES	NO	row total	PPV	NPV
Test +	37	398	435	8.56%	
Test -	12	9553	9565		99.88%
column total	49	9951			

3. Complete the remaining tables for the Current Technology for 40-49 year olds, and the 3D Technology for 30-39 year olds and 40-49 year olds.
4. SCROLL DOWN TO THE TEXT BOX IN ROW 24 and write a summary of your model results.

[1] https://www.health.harvard.edu/newsletter_article/digital_mammography_better_for_some_women

[2] https://www.komen.org/breast-cancer/facts-statistics/research-studies/topics/3d-mammography-breast-tomosynthesis/

[3] https://www.cancer.gov/types/breast/risk-fact-sheet#what-is-the-average-american-womans-risk-of-being-diagnosed-with-breast-cancer-at-different-ages

Problem 2: Hospital Admission and Discharge Processes

The daily admission and discharge processes in hospitals can cause congestion because a bed is not available for a patient waiting to be admitted. One reason this can happen is the discharge process of hospital physicians. On their daily morning rounds, many physicians typically round on sickest patients first, not on those who are likely to be discharged that day. If instead the physicians first round on the patients most likely to be discharged, then all of the activities that need to occur in the discharge process could get started. The tradeoff, of course, is that it can also take time for physicians' orders to get started on those patients who are staying.

Similarly, the morning admissions process can be driven by the process of scheduling operating rooms. Operating rooms typically open at 7 a.m. So, patients coming out of surgery who require an inpatient bed will require one by sometime in the mid-morning.

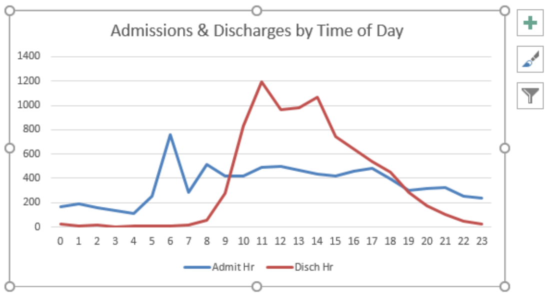
You have been asked to conduct an analysis of the timing of admissions and discharges by hour of day in your hospital for any patient who was in the hospital this past year. You pulled just the time, and not the date, to run this quick analysis.

Here is the first caution in using timestamps. The timestamp is when the data was entered in the computer. This means that it may not be the time the event actually occurred. If a clinician "batches" their data entry tasks, for example, you will see a series of timestamps in which the times are bunched together. For example, if there is a series of discharge timestamps that are all bunched around 6:50 a.m., then the clinician likely entered the data just before change of shift. So, in real life, always look for this potential issue when using timestamps.

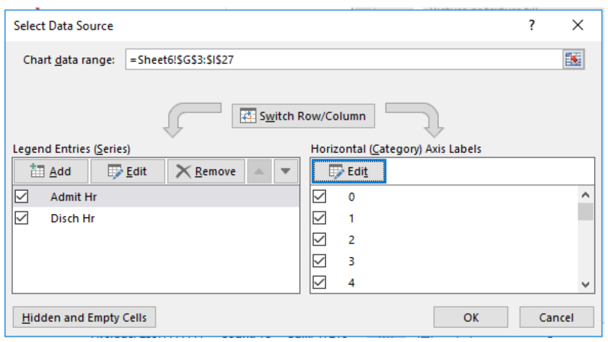
Here are the steps for doing the analysis:

1. The worksheet called ADT Analysis has the data. Each row is the admit time and discharge time of each patient. You need to convert the time into the hour. For example, if the discharge occurred at 2:31, you want the number 2. You will do this in the columns called AdmitHr and DischHr. For cell C2, enter the following formula: =HOUR(B2). Copy and paste this all the way down column C. Repeat for column E, using the discharge time instead of the admit time.
2. Now create a pivot table. For the first pivot table, you want a count of the admissions by admit hour.
3. Copy and paste special values of the numbers from this pivot table somewhere to the right of the pivot table.
4. Now change the pivot table to be the count of discharges by discharge hour.
5. Copy and paste special values of the number of discharges by hour so that it is directly besides the admission counts by hour. See the last page of this document to see what your numbers should be.
6. Now create a line chart that shows the admissions and discharges by hour on the same graph. See below for an example. You should just be selecting the two columns of data for the graph, not the hour numbers of 0 to 23. You can type a title in the Chart Title box.

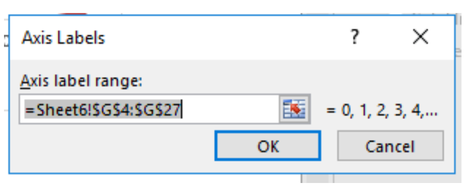
7. To create an x axis that goes from 0 to 23 (NOT 1 to 24), right click on the chart you just made so that it has little circles only around the edges of it as shown below.



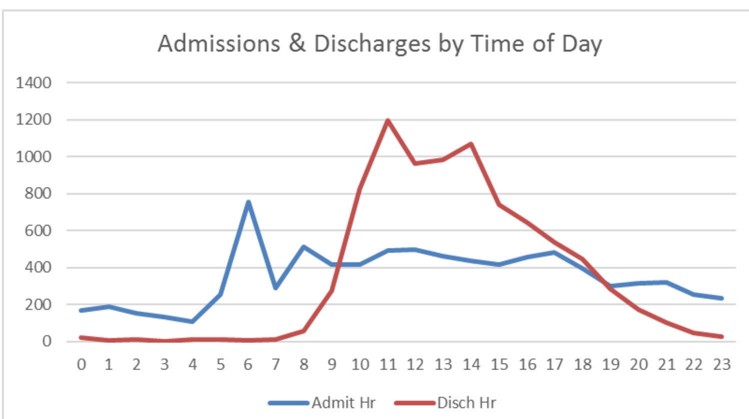
8. In the menu options window that pop up, Choose Select Data.
9. This will bring up a window that looks like below. Click on the Edit button under the Horizontal (Category) Axis Labels.



10. This will bring up a window in which you will enter the range where the 0 to 23 are in your worksheet. Enter this range and click OK. Then click OK on the Select Data Source range window that comes up. This will change your x axis to correctly shows hours 0-23.



Hour	Admit Hr	Disch Hr
0	167	22
1	187	8
2	156	14
3	136	4
4	108	10
5	253	10
6	758	9
7	289	14
8	511	57
9	417	275
10	417	829
11	491	1194
12	497	963
13	463	982
14	437	1071
15	418	742
16	456	643
17	482	540
18	397	448
19	301	287
20	315	172
21	323	101
22	257	49
23	234	26



Problem 3: Exponential Growth

Your organization is concerned that the state leadership is not doing enough to stop the spread of COVID. Your leadership team will be meeting with members of the governor's office because so far their response is that not that many people have the virus, so there is no need to take action. Your goal for the meeting is to get them to understand what exponential growth looks like, because this is the growth model of pandemics. It can appear that things are going fine, but then all of a sudden – bam (as Emeril Legasse says).

Your plan is to set up a model that shows exponential growth over a 90 day time window. You want to be able to show different scenarios of how many people will be infected if nothing is done given: 1) how many people are infected now; 2) how quickly the virus spreads. How fast the virus spreads requires two parts: 1) the rate – e.g. doubling would be a 2, up by 50% would be a .5; and 2) in how many days. So for example 2 in 3 days means the number of people infected doubles every 3 days. So, there are 3 model inputs.

1. Open the worksheet called 3.exp growth.
2. In cell E1, name the cell Inf_Now
3. In cell E2, name the cell Growth_Rate
4. In cell E3, name the cell Days
5. You want column A numbers to start at 1 in cell A2. Then, the number in cell A3 should depend on how many days it takes for the spread rate. For example, if it doubles every 4 days, then you want the number in cell A3 to be a 5. If it is every 3 days, you want the number in cell A3 to be 4. Column A needs to use the information from your Days cell. Type a 1 in cell A2. Now what should you type in cell A3? There should be a formula that adds the number from the Days cell to the number directly above it in column A. Copy and paste this logic down a ways in column A, say out to 90 days.
6. Cell B2 needs to have the starting number of people infected. Do NOT directly type that number in cell B2. It should be typed in cell E1, and the formula in cell B2 should reference your named E1 cell. This makes it clearer to the end users how this model is working.
7. Cell B3 now needs to use both your initial value in Cell B2 and the growth rate input in cell E2 to determine how many people are infected. If the spread rate doubles, then a 2 would be in cell E2. If it goes up by 50%, the number in cell E2 would be 1.5. So, what should the formula be in cell B3? From B3 on down, it should multiply the number right above it by the spread rate in cell E2. Copy and paste this formula all the way down.
8. Now you need to chart the numbers, with the Day Number on the x axis, and the # Infected on the y axis. You have directions on how to chart from problem 2 of this homework.
9. Now, you want to set the x and y axes at fixed numbers, because you are going to create the "what if" model shortly for comparison. <u>Whenever you are comparing data across charts or graphs, the x axis range needs to be the same across the 2 charts, as does the y axis range. The eyes can fool you, because people will look at the picture, and not the axis labels. If anyone ever shows you comparisons, and they have hidden the axis values, watch out.</u> To see how to set the axis range, go to the following website, and scroll down to the heading that says: "How to Change the Axis Range": https://www.got-it.ai/solutions/excel-chat/excel-tutorial/change-axis-values-in-excel/how-to-change-axis-values-in-excel For the x axis, fix it at 90 days. For the y axis,

set it at a high enough number to capture the biggest number you might expect from the initial conditions of your inputs.

10. Create a replica of this worksheet. NO, don't retype everything in a new worksheet. Right click on the worksheet name, choose "Move or Copy." When the dialogue box comes up, click on the bottom box that says "Create a Copy" and hit OK. This will create an exact replica worksheet of the model you just created. Rename this worksheet Scenarios.
11. Change the Growth Rate to a lower number. For example, if it is a 2 in your first scenario, change it to a 1.5 in this one. This shows the impact of the # Infected over time if you can "Bend the Curve" for doubling every x days to going up by only 50% every x days.
12. To view the two worksheets side by side, Choose the View Tab and then choose New Window. Now you can make both of the worksheets fit side by side on your screen. You can see the original ("base case") graph on one side of the screen, and your revised scenario right next to it on the other side of your screen. Use these two graphs in your writeup.
13. Show your model to your friends and family and ask them if they think this model qualifies you to be an Excel god or goddess.