ASPIRE Workshop 5: Data Collection Tools

Katie Derington, PharmD
Outcomes Research Fellow in Ambulatory Care
Kaiser Permanente Colorado & University of Colorado Skaggs School of Pharmacy
   and Pharmaceutical Sciences

Outline

- **Data collection tools**

- **Data management techniques using Microsoft Excel®**

Note: Supplemental Slides on Statistical Analysis are posted on the ASPIRE website under Session 5's "Useful Tools" section:
http://aspirekpco.weebly.com/session-materials.html

KAISER PERMANENTE® thrive

DATA COLLECTION TOOLS

OPEN THE "ASPIRE_DATASET FOR PRACTICE_NO PATIENT DATA" EXCEL DOCUMENT

KAISER PERMANENTE® thrive

Examples of Data Collection Tools

- **Centralized database**
  - Ex. Redcap: https://projectredcap.org/
  - HIPAA compliant, with online and offline data capture capabilities
  - Allows for data validation, branching logic, basic calculations, and can export entered data to most statistical packages

- **Spreadsheet-based, e.g. Excel®**

KAISER PERMANENTE® **thrive**

Excel as a Data Collection Tool

- **Keep in mind what you will be doing with the spreadsheet after entering the data**
  - Consider how to enter data to make analysis easier
- **Suggestions to consider:**
  - Better to enter more "raw" data, and perform calculations in Excel
    - For example, if calculating length of stay: create a column for admission date and discharge date, then calculate difference in dates in another column
  - Enter data into Excel cell in a usable form
    - Can use a separate data dictionary to code entries
  - Avoid entering more than one data point into the same cell
    - Cannot perform calculations
  - Employ data validation techniques to avoid errors

KAISER PERMANENTE® **thrive**

Importance of Creating a Good Data Collection Tool to Minimize Possible Errors
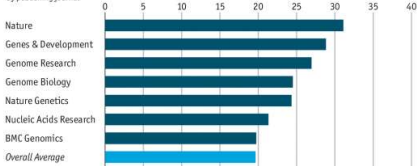
http://www.economist.com/blogs/graphicdetail/2016/09/daily-chart-3.
Accessed September 24, 2016.

KAISER PERMANENTE® **thrive**

DATA MANAGEMENT IN EXCEL®

KAISER PERMANENTE® thrive

Quick Excel® Tips 'N Tricks

- **Excel does not treat blank cells as zero**
- **If there are blank cells and you are performing a function on these cells, use "" to indicate blank**
- **If referring to words in an Excel formula, such as counting the number of females, enter the words in quotes, i.e. "female"**
- **If intend to copy your formula and apply it down an entire column and do not want the cells references in a formula to change, use the $ sign:**
  - **For example: A$1:B$88 vs. A1: B88**

KAISER PERMANENTE® thrive

Basic Spreadsheet Manipulation

- **Using the data fill handle**
  - **Task: enter "4" into HF clinic duration and use the fill handle to drag down 10 cells (then undo)**

	F	G	H	I
	Dischar ge_Dt	HF Clinic Duration	BL_EF_ actual	Beta blocker_ baseline
	37372	4		Metoprolol
	37658		40	Metoprolol

Hold mouse over bottom right corner until a '+' forms. This is the fill handle.

KAISER PERMANENTE® thrive

Basic Spreadsheet Manipulation

- **Using the data fill handle**

Dischar ge_Dt	HF Clinic Duration
37372	4
37658	
37522	
37693	
37671	
37762	
37627	
37476	
37453	

Dischar ge_Dt	HF Clinic Duration
37372	4
37658	4
37522	4
37693	4
37671	4
37762	4
37627	4
37476	4
37453	4

10

KAISER PERMANENTE. thrive

Basic Spreadsheet Manipulation

- **Select the entire column/row/spreadsheet**
  - Task 1: select row 3 of the spreadsheet

	ID	Group	Gender	Age
1				
2	1	COPD	M	80
3	131	NoCOPD	F	82
4	2	COPD	M	61

Select the number 3 on the left. A black arrow should form.

11

KAISER PERMANENTE. thrive

Basic Spreadsheet Manipulation

- **Select the entire column/row/spreadsheet**
  - Task 2: select column D of the spreadsheet

C	D	E

Select column D at the top. A black arrow should form.

Gender	Age	Enroll_Dt
M	80	1/25/02
F	82	10/15/02

12

KAISER PERMANENTE. thrive

Basic Spreadsheet Manipulation

- **Select the entire column/row/spreadsheet**
  - **Task 3: select the entire spreadsheet**
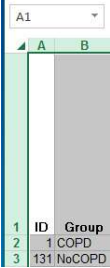
Select the top left corner where there is a triangle. A white plus sign should form.

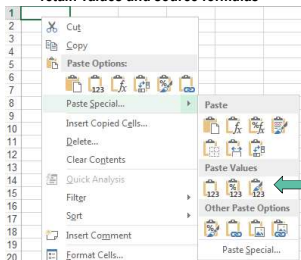## Basic Spreadsheet Manipulation

- **Paste special: can retain formatting or formulas when copy/paste**
  - **Task: copy the first 5 columns and first 10 rows, then use paste special to retain values and source formulas**

Basic Spreadsheet Manipulation

- **Paste special: can retain formatting or formulas when copy/paste**
  - **Task: copy the first 5 columns and first 10 rows, then use paste special to retain values and source formulas**

Select values & source formatting (E)

Freeze Panes(s) in a Dataset

- **Task: freeze top row of dataset**

Freeze Panes(s) in a Dataset

- **Task: freeze both top row and first column of dataset**

	ID	Group	Gender
1			
2	1	COPD	M
3	131	NoCOPD	F

Select first cell next to the column and row of interest (in this case, cell B2)

Freeze Panes(s) in a Dataset

- **Task: freeze both top row and first column of dataset**

While B2 is selected, select freeze panes

Formatting Cells

- **Task: format column F (discharge date) as a date (mm/dd/yy)**

	E	F	G

Select all of column F

nroll_Dt	Dischar ge_Dt	HF Clinic Duration
1/25/02	37372	
10/15/02	37658	
1/11/02	37522	
2/10/03	37693	
11/15/01	37671	

19

KAISER PERMANENTE. thrive

Formatting Cells

- **Task: format column F (discharge date) as a date (mm/dd/yy)**

- Cut
- Copy
- Paste Options:
- Paste Special...
- Insert
- Delete
- Clear Contents
- Format Cells...
- Column Width...
- Hide
- Unhide

Right click on the column and select format cells

20

KAISER PERMANENTE. thrive

Formatting Cells

- **Task: format column F (discharge date) as a date (mm/dd/yy)**

Format Cells

Number | Alignment | Font | Border | Fill | Protection

Category:
- General
- Number
- Currency
- Accounting
- Date
- Time
- Percentage
- Fraction
- Scientific

Sample
Discharge_Dt

Type:
- 14-03-2012
- 14/3/12
- 14/03/12
- 14/mar
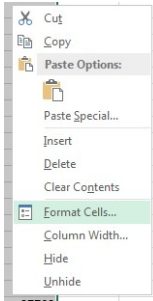
Discharge_Dt
26/04/02
06/02/03
23/09/02
13/03/03
19/02/03
21/05/03
06/01/03
08/08/02
16/07/02
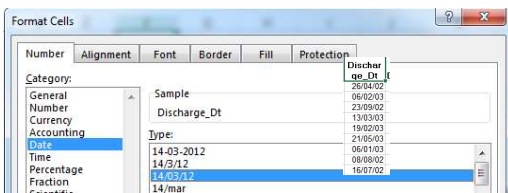
Select date and choose the desired formatting option. Click OK to accept.

21

KAISER PERMANENTE. thrive

Formatting Cells

- **Task: format column F (discharge date) as a date (mm/dd/yy)**

Dischar ge_Dt	
26/04/02	
06/02/03	
23/09/02	
13/03/03	
19/02/03	
21/05/03	
06/01/03	
08/08/02	
16/07/02	

Formatted dates

KAISER PERMANENTE. thrive

22

Formatting Cells

- **Task: Custom format ID (column A) to include 3 digits**

Select all of column A

	ID	Group
1		
2	1	COPD
3	131	NoCOPD
4	2	COPD
5	83	NoCOPD
6	129	NoCOPD

KAISER PERMANENTE. thrive

23

Formatting Cells

- **Task: Custom format ID (column A) to include 3 digits**

- Cut
- Copy
- Paste Options:
- Paste Special...
- Insert
- Delete
- Clear Contents
- Format Cells...
- Column Width...
- Hide
- Unhide

Right click on the column and select format cells

KAISER PERMANENTE. thrive
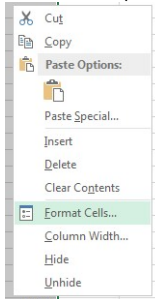
24

Formatting Cells

- **Task: Custom format ID (column A) to include 3 digits**

Select custom format and option to include 3 figures without a decimal point. Replace # signs with zeros.

NENTE. thrive

Formatting Cells

- **Task: Custom format ID (column A) to include 3 digits**

	ID
1	ID
2	001
3	131
4	002
5	083
6	129
7	003
8	019
9	147
10	004
11	124

Formatted IDs to all include 3 digits

KAISER PERMANENTE. thrive

Formatting Cells

- **Task: highlight the row for patient number 147 (row 9) in red**

	ID	Group	Gender	Age	Enroll_Dt	Discharge_Dt	HF Clinic Duration
1	ID	Group	Gender	Age	Enroll_Dt	Discharge_Dt	HF Clinic Duration
2	1	COPD	M	80	1/25/02	37372	
3	131	NoCOPD	F	82	10/15/02	37658	
4	2	COPD	M	61	1/11/02	37522	
5	83	NoCOPD	M	64	2/10/03	37693	
6	129	NoCOPD	M	61	11/15/01	37671	
7	3	COPD	F	76	8/13/02	37762	
8	19	NoCOPD	F	77	10/26/01	37627	
9	147	NoCOPD	M	80	3/21/02	37476	
10	4	COPD	F	74	4/10/02	37453	

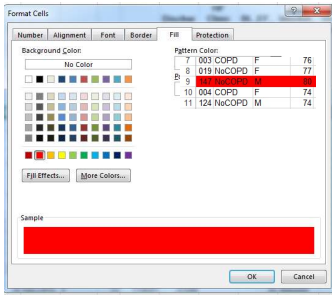KAISER PERMANENTE. thrive

Formatting Cells

- **Task: highlight the row for patient number 147 (row 9) in red**
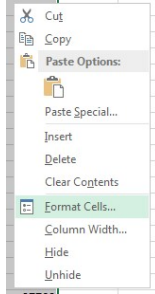
Formatting Cells

- **Task: highlight the row for patient number 147 (row 9) in red**

Formatting Cells

- **Task: highlight the row for patient number 147 (row 9) in red**

7	003	COPD	F		76
8	019	NoCOPD	F		77
9	147	NoCOPD	M		80
10	004	COPD	F		74
11	124	NoCOPD	M		74

Patient 147 highlighted in red

Formatting Cells

- **Task: change columns P, Q, R (SBP, DBP, HR) to be vertical or horizontal**

O	P	Q	R	S
BB D/C'd overall	SBP	DBP	HR	Hypotension

KAISER PERMANENTE. thrive

31

Formatting Cells
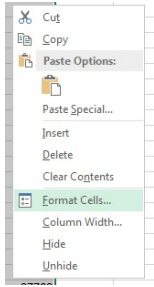
- **Task: change columns P, Q, R (SBP, DBP, HR) to be vertical or horizontal**

✂	Cut
📋	Copy
📋	Paste Options:
	Paste Special...
	Insert
	Delete
	Clear Contents
	Format Cells...
	Column Width...
	Hide
	Unhide

KAISER PERMANENTE. thrive

32

Formatting Cells

- **Task: change columns P, Q, R (SBP, DBP, HR) to be vertical or horizontal**

Format Cells

Number | Alignment | Font | Border | Fill | Protection

Text alignment
Horizontal: Center — Indent: 0
Vertical: Bottom
Justify distributed

Text control
☑ Wrap text
Shrink to fit
Merge cells

Right-to-left
Text direction: Context

Orientation — 90 Degrees

OK | Cancel
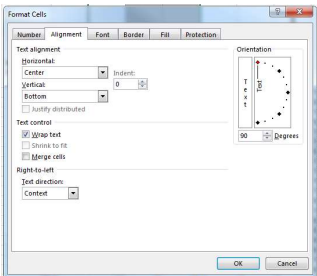
Enter 90 degrees orientation for vertical text or 180 degrees orientation for horizontal
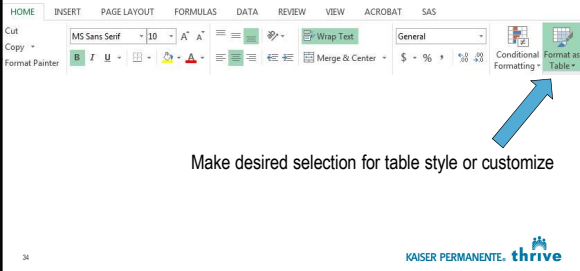
KAISER PERMANENTE. thrive

33

Format as a Table

- **Task: change excel spreadsheet to a table**

Make desired selection for table style or customize

KAISER PERMANENTE. thrive

Format as a Table

- **Task: change excel spreadsheet to a table**

If do not select this, Excel will add headers to your table
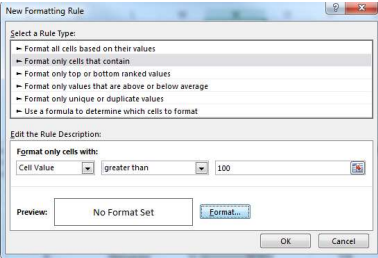
KAISER PERMANENTE. thrive
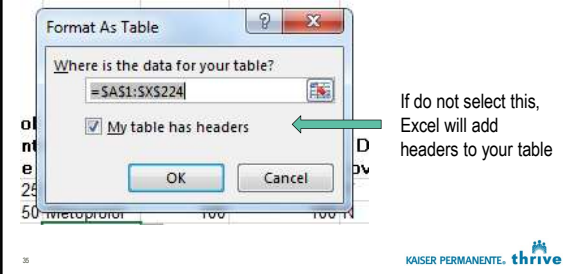
Conditional Formatting

- **Task: use conditional formatting to shade cells in "metoprolol equivalent_discharge" (column N) with value >100 in green**
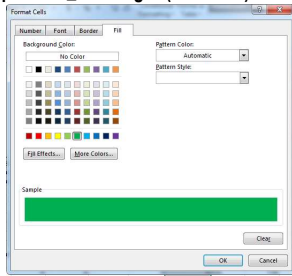

Conditional formatting is under the home tab

KAISER PERMANENTE. thrive
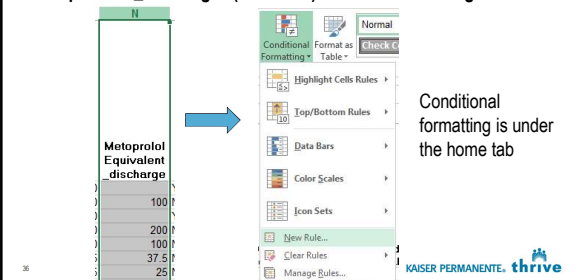
Conditional Formatting

▪ **Task: use conditional formatting to shade cells in "metoprolol equivalent_discharge" (column N) with value >100 in green**

Select format

Conditional Formatting

▪ **Task: use conditional formatting to shade cells in "metoprolol equivalent_discharge" (column N) with value >100 in green**

Select green and OK

Conditional Formatting

▪ **Task: use conditional formatting to shade cells in "metoprolol equivalent_discharge" (column N) with value >100 in green**

100
200
100
12.5
150
200
100
50
50

Cells with values >100 are highlighted in green

Data Validation

- **Task: create a column after "HF clinic duration" for "smoking status" with the following choices: current, past, never**
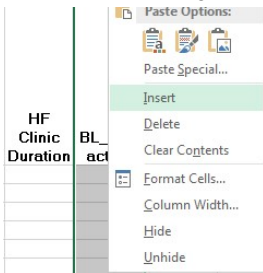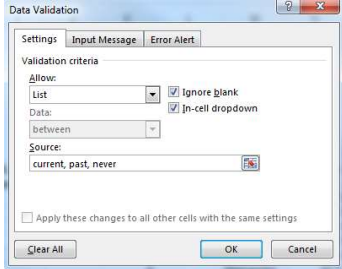
Highlight entire column next to HF clinic duration. Right click and select insert.
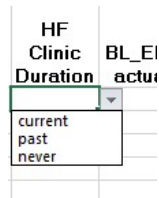
Data Validation

- **Task: create a column after "HF clinic duration" for "smoking status" with the following choices: current, past, never**

Data Validation

- **Task: create a column after "HF clinic duration" for "smoking status" with the following choices: current, past, never**
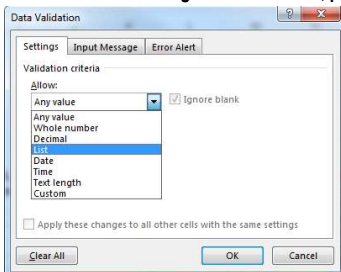
Data Validation

- **Task: create a column after "HF clinic duration" for "smoking status" with the following choices: current, past, never**

Data Validation

- **Task: create a column after "HF clinic duration" for "smoking status" with the following choices: current, past, never**

Dropdown menu for selection with current, past, never
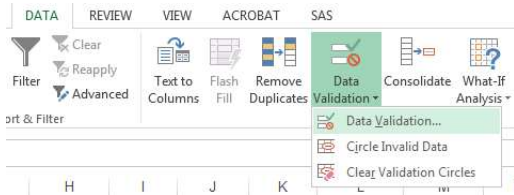

Data Validation

- **Task: change columns "SBP" and "DBP" to only allow numbers**

SBP	DBP
104	64
150	78
86	54
110	80
108	80
100	60
164	80
112	64
124	54

Data Validation

- **Task: change columns "SBP" and "DBP" to only allow numbers**

Data Validation

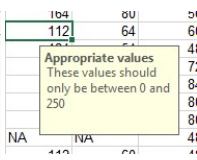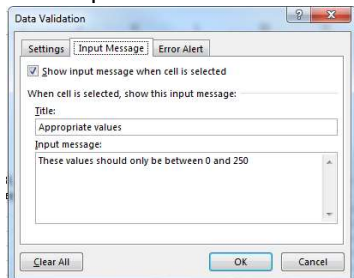- **Task: change columns "SBP" and "DBP" to only allow numbers**
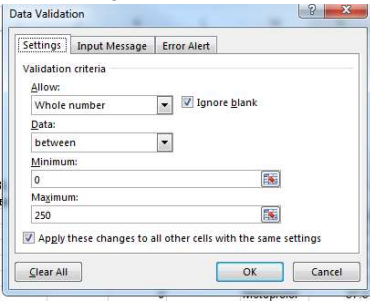
Set criteria specific to expected values in that column

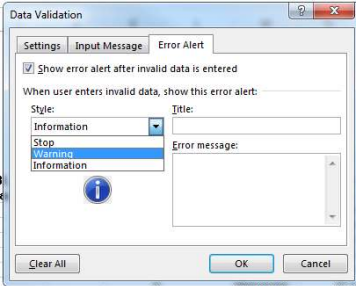# Data Validation

- **Task: Exploration**

Message will appear before user enters any information

Data Validation

- **Task: Exploration**

Stop will prevent user from entering incorrect information. Warning will warn the user (can include a message) and requires that the user okay entering invalidated information.
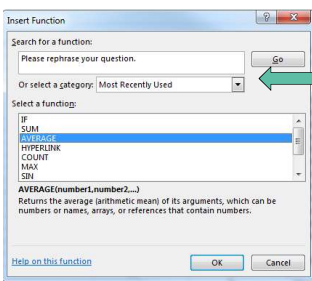
Basic Excel Functions

- **Task: calculate the mean, maximum and minimum age**

Select cell where you would like to insert the function
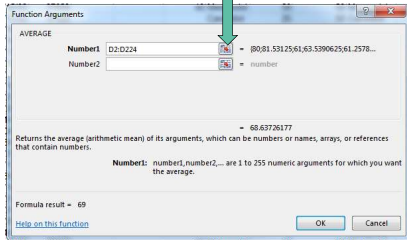
Basic Excel Functions

- **Task: calculate the mean, maximum and minimum age**

You can also search for functions

Basic Excel Functions

- **Task: calculate the mean, maximum and minimum age**

Click on this box to select your data range (instead of typing the range)

Enter or select your desired data range. Could also enter values individually in the number boxes

Basic Excel Functions

- **Task: calculate the mean, maximum and minimum age**
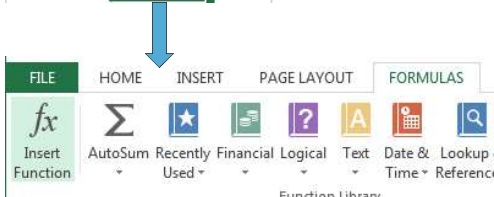
mean	69
max	88
minimum	42

Calculated mean, maximum, and minimum ages in dataset.

CountIF and CountIFS

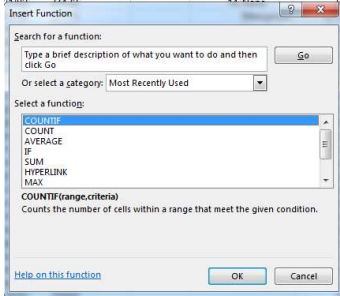- **Task: count the number of females in the sample using Countif**

Select cell where you would like to insert the function
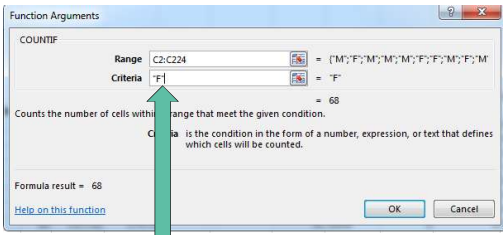
CountIF and CountIFS

- **Task: count the number of females in the sample using Countif**

CountIF and CountIFS

- **Task: count the number of females in the sample using Countif**

If counting something that is a word or letter, put in double quotes

CountIF and CountIFS

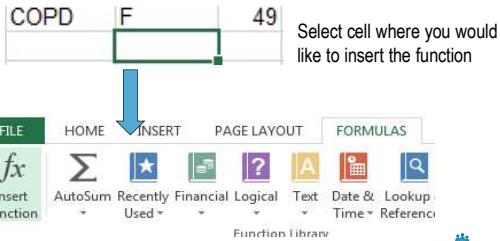- **Task: count the number of females in the sample using Countif**

count F	68

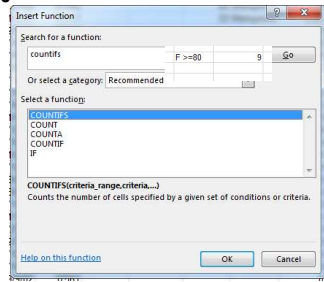Total number of female patients in the dataset.

19

CountIF and CountIFS

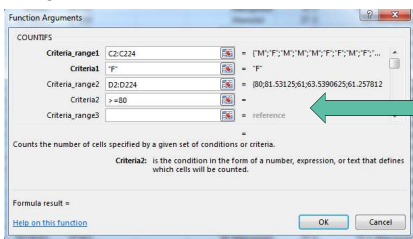- **Task: count the number of females $\geq$ 80 years old in the sample using Countifs**

COPD | F | 49

Select cell where you would like to insert the function

FILE | HOME | INSERT | PAGE LAYOUT | FORMULAS

Insert Function | AutoSum | Recently Used | Financial | Logical | Text | Date & Time | Lookup & Reference

Function Library

CountIF and CountIFS

- **Task: count the number of females $\geq$ 80 years old in the sample using Countifs**

Insert Function

Search for a function:

countifs | F >=80 | 9 | Go

Or select a category: Recommended

Select a function:

COUNTIFS
COUNT
COUNTA
COUNTIF
IF

COUNTIFS(criteria_range,criteria,...)
Counts the number of cells specified by a given set of conditions or criteria.

Help on this function | OK | Cancel

CountIF and CountIFS

- **Task: count the number of females $\geq$ 80 years old in the sample using Countifs**

Function Arguments

COUNTIFS

Criteria_range1 | C2:C224 | = {"M";"F";"M";"M";"M";"F";"F";"M";"F";...
Criteria1 | "F" | = "F"
Criteria_range2 | D2:D224 | = {80;81.53125;61;63.5390625;61.257812
Criteria2 | >=80 | =
Criteria_range3 | | = reference

=
Counts the number of cells specified by a given set of conditions or criteria.

Criteria2: is the condition in the form of a number, expression, or text that defines which cells will be counted.

Formula result =

Help on this function | OK | Cancel

Use >= to denote "greater than or equal to"

CountIF and CountIFS

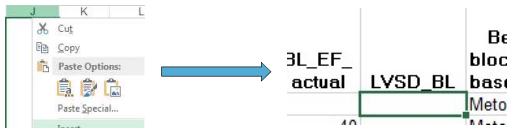- Task: count the number of females $\geq$ 80 years old in the sample using Countifs

F >=80	9

Number of female patients greater than or equal to 80 years old.

IF Statements

- Task: insert a column after "BL_EF_actual" called "LVSD_BL" that displays a value of 1 if the patient's EF is <40% and 0 if EF is >40%, and blank if the EF is not available
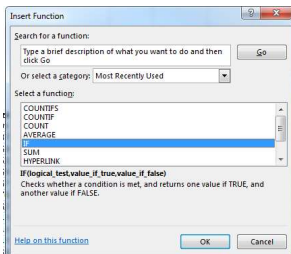

I typically enter the formula into one cell first, then drag down to the rest of the cells (if applicable) later

IF Statements

- Task: insert a column after "BL_EF_actual" called "LVSD_BL" that displays a value of 1 if the patient's EF is <40% and 0 if EF is >40%, and blank if the EF is not available

IF Statements

- **Task: insert a column after "BL_EF_actual" called "LVSD_BL" that displays a value of 1 if the patient's EF is <40% and 0 if EF is >40%, and blank if the EF is not available**

Since there are missing EF values in column I, need to embed an IF statement using "" to denote blank

KAISER PERMANENTE. thrive

IF Statements

- **Task: insert a column after "BL_EF_actual" called "LVSD_BL" that displays a value of 1 if the patient's EF is <40% and 0 if EF is >40%, and blank if the EF is not available**

Second IF statement indicating to show a 1 if I2<40, or otherwise to show a 0

KAISER PERMANENTE. thrive

IF Statements

- **Task: insert a column after "BL_EF_actual" called "LVSD_BL" that displays a value of 1 if the patient's EF is <40% and 0 if EF is >40%, and blank if the EF is not available**
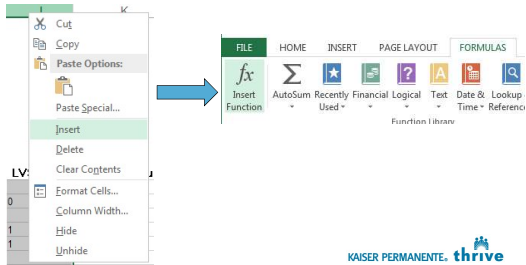
Final IF formula: =IF(I2="","",IF(I2<40, 1, 0))

This could be read as: IF I2 is blank, then insert blank. Otherwise, if I2 is less than 40 insert 1. Otherwise (if I2 is NOT less than 40), insert 0.

When this formula is dragged down the column, I2 will change to refer to I3, then I4, etc.
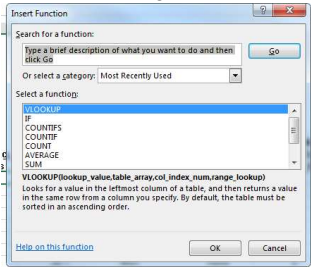
KAISER PERMANENTE. thrive

Vlookup

- **Task: insert a column after the one you just created (LVSD_BL) that includes the discharge EF data available in Sheet 1**

Vlookup

- **Task: insert a column after the one you just created (LVSD_BL) that includes the discharge EF data available in Sheet 1**

Vlookup

- **Task: insert a column after the one you just created (LVSD_BL) that includes the discharge EF data available in Sheet 1**

References the first "linking" cell (ID) that is present in both datasets

All cells of the second spreadsheet (Sheet1) – this is where the formula will search for values

In Sheet1, column B or the second column, contains the new values we want to insert into the cohort spreadsheet

False indicates that we want to find an exact match.

Vlookup

- **Task: insert a column after the one you just created (LVSD_BL) that includes the discharge EF data available in Sheet 1**

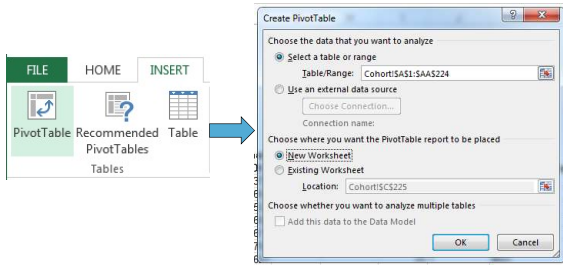 Final formula: =VLOOKUP(A2,Sheet1!A$2:B$51,2,FALSE)

  Note that in Sheet1 there are only 51 patients with a follow-up EF whereas there are 223 patients in the cohort. Therefore, when this formula is applied to other cells, many will show "#N/A" which means not available.

KAISER PERMANENTE® **thrive**

70

Pivot Tables

- **Task: Display gender by COPD status in a separate worksheet**

FILE   HOME   INSERT

PivotTable Recommended Table
            PivotTables

Tables

Create PivotTable

Choose the data that you want to analyze
- ◉ Select a table or range
  - Table/Range: Cohort!$A$1:$AA$224
- ○ Use an external data source
  - Choose Connection...
  - Connection name:

Choose where you want the PivotTable report to be placed
- ◉ New Worksheet
- ○ Existing Worksheet
  - Location: Cohort!$C$225

Choose whether you want to analyze multiple tables
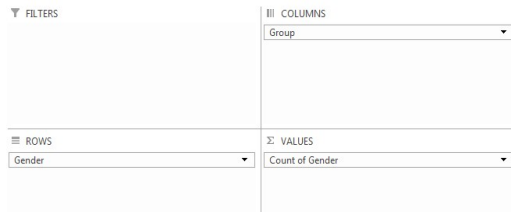- ☐ Add this data to the Data Model

OK   Cancel

KAISER PERMANENTE® **thrive**

71

Pivot Tables

- **Task: Display gender by COPD status in a separate worksheet**

Drag fields between areas below:

▼ FILTERS

▐▐▐ COLUMNS
Group ▾

≡ ROWS
Gender ▾

Σ VALUES
Count of Gender ▾

KAISER PERMANENTE® **thrive**

72

Pivot Tables

▪ **Task: Display gender by COPD status in a separate worksheet**

Count of Gender	Group		
Gender	COPD	NoCOPD	Grand Total
F	23	45	68
M	63	92	155
Grand Total	86	137	223

KAISER PERMANENTE. **thrive**

73

Pivot Tables

▪ **Task: Display mean metoprolol equivalent dose at discharge by COPD status**

KAISER PERMANENTE. **thrive**

74

Pivot Tables

▪ **Task: Display mean metoprolol equivalent dose at discharge by COPD status**

Drag fields between areas below:

▼ FILTERS	▌ COLUMNS
	Group ▼

≡ ROWS	Σ VALUES
	Count of Metoprolol Equivalent_discharge ▼

KAISER PERMANENTE. **thrive**

75

Pivot Tables

- Task: Display mean metoprolol equivalent dose at discharge by COPD status

Pivot Tables

- Task: Display mean metoprolol equivalent dose at discharge by COPD status

Average of Metoprolol Equivalent_discharge	Group		
	COPD	NoCOPD	Grand Total
Total	97.06790123	107.4074074	103.5300926

Pivot Tables

- Task: display mean SBP by beta blocker agent at discharge, then limit to just patients receiving metoprolol

Pivot Tables

- **Task: display mean SBP by beta blocker agent at discharge, then limit to just patients receiving metoprolol**

Pivot Tables

- **Task: display mean SBP by beta blocker agent at discharge, then limit to just patients receiving metoprolol**

Pivot Tables

- **Task: display mean SBP by beta blocker agent at discharge, then limit to just patients receiving metoprolol**

| Average of SBP	
Beta blocker_discharge	Total
Atenolol	124.6
Carvedilol	110.1538462
Metoprolol	115.6195652
NONE	88
Toprol XL	120
(blank)	105.5
Grand Total	115.4084507

Select the dropdown arrow to view filter options

Pivot Tables

- **Task: display mean SBP by beta blocker agent at discharge, then limit to just patients receiving metoprolol**

Average of SBP	
Beta blocker_discharge	Total
Metoprolol	115.6195652
Toprol XL	120
Grand Total	115.6432432

QUESTIONS?
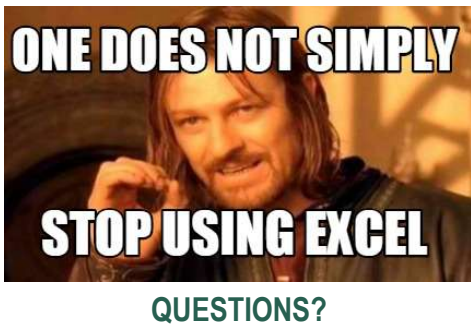
Now…work on your data collection tool

- **If time allows during the session…**

  - Split into groups of 2-3
  - Revise your data collection tool employing principles discussed today
  - Review and provide feedback to your partner(s) about their tool