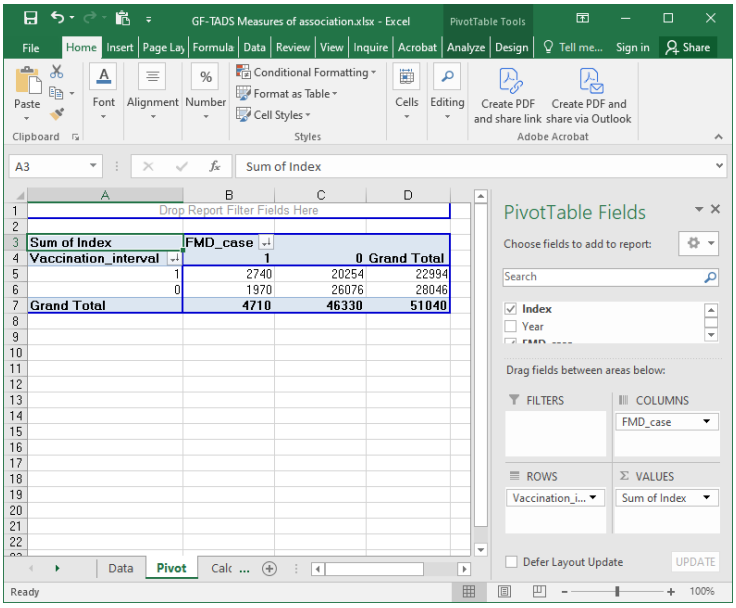
Measures of association

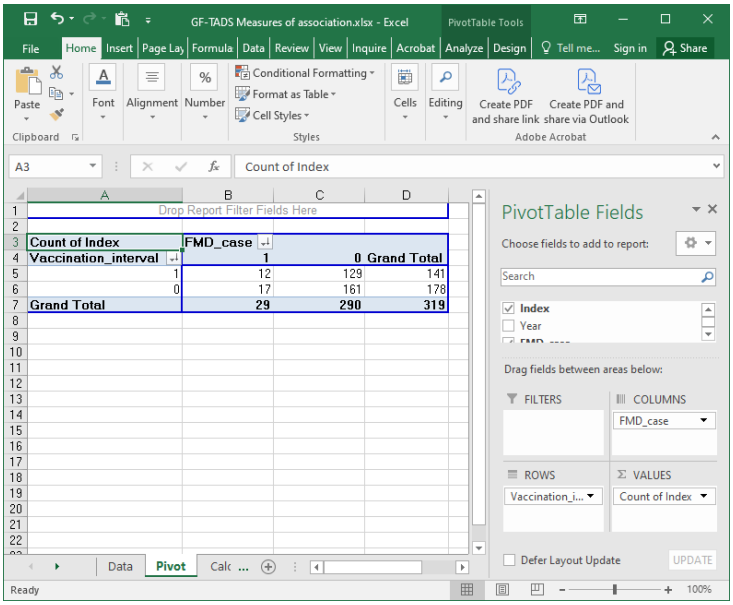
Data were extracted from WOAH reports concerning reported FMD outbreaks in a southern African country. A case-control study was performed to investigate risk factors for FMD detection and reporting. Risk factor information was collected from government records and spatial data sources. The objective of these evaluations was to determine the factors significantly associated with the reporting of FMD in effort to improve control in the region.

Risk factor data have been saved as "GF-TADS Measures of association.xlsx". Columns related to the quantitative data have been hidden and only variables that have been dichotomized will be analyzed as part of this exercise. The data are saved on the "Data" worksheet. A pivot table has already been generated that will help you answer the following questions. A worksheet named "Calculations" has also been provided that can be used to perform the formal statistical analysis. Please contact a lecturer if anything is unclear concerning the provided data or the subsequent questions.

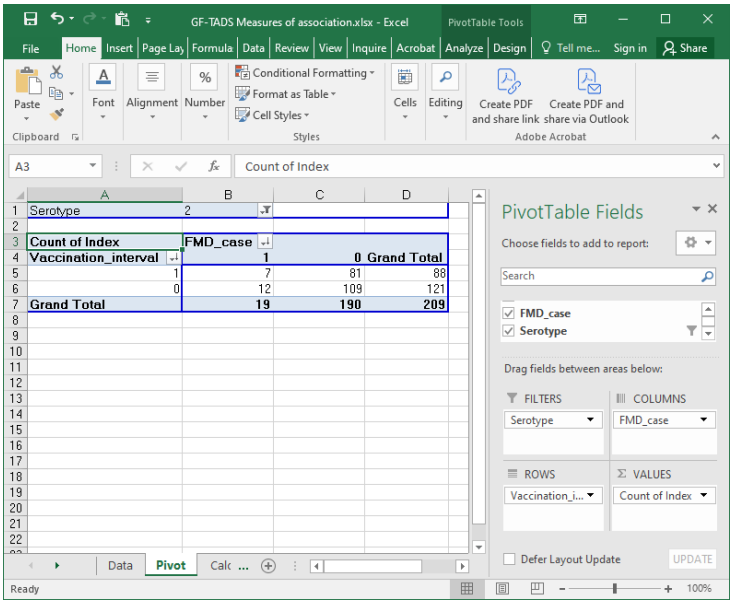
Pivot tables are powerful tools that can be used to create 2x2 tables for the calculation of important epidemiological measures. Instructions are provided here to help you create 2x2 tables consistent with the presentation in the lecture and guide you in answering questions. It is possible to enter formulas underneath the pivot table that calculate measures of association automatically as you update the data in the pivot table. Go to the worksheet named "Pivot" to access the pivot table. The table should currently be empty. Drag the variable "FMD_case" and drop it as a column field. Drag the variable "Vaccination_interval" and drop it as a row field. Drag "Index" and drop as a value field. The result should be the same as the image below.



This is a base 2x2 table but the default approach by Excel is not consistent with what epidemiologists expect for the table. Double-click on the cell with "Sum of Farm_ID" and change the function to **count**. (Note that Excel typically sorts 0-1 but epidemiologists want the tables sorted 1-0 so I already made that change within the data.)



The table is now in the correct format to perform the necessary calculations to answer the following questions. Please be sure that the format of all tables are similar whenever you evaluate a new exposure factor (positions of "1" and "0"). From the table above, we see that there were 319 total locations with 141 having longer than median vaccination intervals. There were also 29 FMD affected locations. The "filter" field is used to evaluate subsets of the data. For example, the table below has been filtered and only shows the results for SAT2 outbreaks.



Introduction

The variable for the outcome (disease) of interest for these data is "**FMD_case**".
Estimate the association of the following 11 factors with FMD outbreak reporting:
Vaccination_interval (column H), Vaccination_proportion (column K), Vaccine_match (column N), Cattle_population (column Q), Cattle_inspection (column T), Movements_in (column W), Movements_out (column Z), Human_population (column AC), Game_reserve_distance (column AG), Roads_distance (column AJ), Rivers_distance (column AM).

1. Using the Pivot Table, create a 2x2 table for the **Vaccination_interval** variable (row variable) against **FMD_case** as the outcome (column variable). Copy the data from the four central cells of the Pivot Table and Paste Special (values) into the **Observed** table on the **Calculations** worksheet. Calculate the odds ratio (OR) comparing the likelihood of FMD-positive locations for this variable. The numerator should be when the variable = 1 (e.g., **Vaccination_interval** = 1) compared to (the denominator) when the variable is coded "0". Enter the appropriate formula on the "Calculations" worksheet in the orange shaded area to the right of the OR label so that it is automatically calculated for the data presented in the 2x2 table to the left (**Observed** table). The calculation of the OR is the point estimate of the association between **Vaccination_interval** and the FMD status of the location. It is also necessary to calculate a measure of the expected variability around this estimate – also referred to as the precision of the estimate. This is typically performed by calculating 95% confidence intervals (CI). Notice that the 95% confidence interval for the OR is calculated on the worksheet now that you have entered the formula for the OR. Epidemiological measures of association should also be statistically tested to determine if the evidence is strong enough to suggest the presence of a causal association. An appropriate statistical test for data presented in 2x2 tables is the chi-square test. The chi-square test statistic and its associated P value are also presented on the **Calculations** worksheet. The CI and chi-square calculations are updated whenever the data in the **Observed** 2x2 table are changed. The chi-square statistic is calculated based on the formal comparison between the observed and the expected 2x2 tables. When the data are changed in the **Observed** table the data will also be changed in the **Expected** table as the first step of calculating the chi-square statistic. Your OR calculation will also be updated; the data only need to be changed in the **Observed** table and the spreadsheet will perform all necessary calculations. Formally present each calculated OR for the 11 variables listed above with their corresponding 95% confidence interval and chi-square test P value in the table below.

Exposure variable	No. of locations exposed	Odds ratio	95% CI		P value
			Lower	Upper	
Vaccination_interval (column H)	141	0.88	0.41	1.91	0.748
Vaccination_proportion (column K)	159	1.09	0.51	2.33	0.832
Vaccine_match (column N)	55	1.00	0.36	2.81	1
Cattle_population (column Q)	157	5.67	2.10	15.3	0.00015
Cattle_inspection (column T)	131	2.19	1.01	4.75	0.044
Movements_in (column W)	70	2.02	0.89	4.57	0.087
Movements_out (column Z)	78	2.03	0.92	4.52	0.077
Human_population (column AC)	158	0.81	0.38	1.75	0.595
Game_reserve_distance (column AG)	158	0.19	0.07	0.50	0.0003
Roads_distance (column AJ)	158	2.06	0.93	4.59	0.071
Rivers_distance (column AM)	159	1.48	0.68	3.20	0.321

2. Which of the evaluated risk factors were significantly associated with FMD outbreak detection?

Increasing cattle population was significantly associated with a higher odds of reported FMD outbreaks (OR = 5.67, P < 0.001).

Increasing distance to the game reserve fence was significantly associated with lower odds of a reported FMD outbreak (OR = 0.19, P < 0.001).

Increasing cattle inspection intensity was significantly associated with a higher odds of a reported FMD outbreak.

3. Please rank the relative importance of the 11 risk factors from smallest to largest P-value. Is this ranking different than what you expected based on the expert opinion exercise?

	Exposure variable	No. of locations exposed	Odds ratio	95% CI		P value
				Lower	Upper	
1	Cattle_population (column Q)	157	5.67	2.1	15.3	0.00015
2	Game_reserve_distance (column AG)	158	0.19	0.07	0.5	0.0003
3	Cattle_inspection (column T)	131	2.19	1.01	4.75	0.044
4	Roads_distance (column AJ)	158	2.06	0.93	4.59	0.071
5	Movements_out (column Z)	78	2.03	0.92	4.52	0.077
6	Movements_in (column W)	70	2.02	0.89	4.57	0.087
7	Rivers_distance (column AM)	159	1.48	0.68	3.2	0.321
8	Human_population (column AC)	158	0.81	0.38	1.75	0.595
9	Vaccination_interval (column H)	141	0.88	0.41	1.91	0.748
10	Vaccination_proportion (column K)	159	1.09	0.51	2.33	0.832
11	Vaccine_match (column N)	55	1	0.36	2.81	1

Data analysis	**Expert opinion**
Cattle_population	Cattle population
Game_reserve_distance	Vac coverage
Cattle_inspection	Cattle inspection
Roads_distance	Game reserve
Movements_out	Vac interval
Movements_in	Vaccine match
Rivers_distance	Movement in
Human_population	Movement out
Vaccination_interval	Rivers
Vaccination_proportion	Roads
Vaccine_match	Human population