# Strategies for data analysis: Cohort study

Postgraduate Research Training in Reproductive Health

Faculty of Medicine, University of Yaounde

Pierre Marie Tebeu (M.D)

pmtebeu@yahoo.fr

# Specific objectives

- Identify similar parameters for inclusion in a cohort
- Identify the study groups of the cohort study
- Identify the event to be assessed
- Calculate the risk, the risk ratio and the Incidence rate ratio
- Calculate the odds ratio for cohort study
- Identify sub-groups for analysis
- Interpret the findings

# What is a cohort?

- a group of people having approximately the same age
- a band of warriors originally a unit of a Roman Legion
- a company of companions or supporters
- a group of animals of the same species, identified by a common characteristic

# What is :
# a study of a cohort? a cohort study?

- Cohort study: Comparable study of two or more groups of the same cohort studied over a period of time as part of a scientific or medical investigation
- Study of a cohort: a group of animals of the same species / a group of persons identified by a common characteristic, which are studied over a period of time as part of a scientific or medical investigation

# Relation
## a study of a cohort # a cohort study

the cohort study involves the study of the cohorts

( each group of the study is a cohort)

# The purpose of a cohort study

- For each group of a cohort, identify the event, the time to event, the incidence, the intensity of the event
- Then to compare different groups of a cohort according to the event

# Example of a cohort study
## (Tebeu et al. GRELL 2004)

- Question: Is adjuvant external radiotherapy really associated to risk of death in stage I endometrial cancer patients?

# Example of a cohort study
## (Tebeu et al. GRELL 2004)

- Common parameter to be in the cohort :

  Operated for stage I endometrial cancer

- Element of exposure: Adjuvant radiotherapy

- Number of groups in the cohort : 3

# Items of analysis

- Identify baseline characteristics of different groups of the cohort study
- Identify the event in different groups
- Calculate the risk, incidence rate, among groups

# Baseline characteristics

(Tebeu et al: GRELL 2004, Montpellier)

**Table 1 Patient and tumor characteristics of women with endometrial cancer according to type of adjuvant therapy**

| Characteristics | Radiotherapy | | | Total |
|---|---|---|---|---|
| | No | 'Brachytherapy' | 'External radiotherapy+/- Brachy' | |
| | N=92 (%) | N=94 (%) | N=111 (%) | N=297 (%) |
| Mean age | 67.0 (38-91) | 64.8 (35-90) | 63.2 (33-81) | 64.9 (33-91) |
| Period of diagnosis | | | | |
| 1980-87 | 30(32.6) | 54(57.4) | 49(44.1) | 133(44.8) |
| 1988-96 | 62(67.4) | 40(42.6) | 62(55.9) | 164(55.2) |
| Invasion of myometrium | | | | |
| < 50% | 68(73.9) | 80 (85.1) | 35 (31.5) | 183 (61.6) |
| >= 50% | 24 (26.1) | 14 (14.9) | 76 (68.5) | 114 (38.4) |
| Differentiation | | | | |
| Good | 66 (71.7) | 64 (68.1) | 50 (45.0) | 180 (60.6) |
| Moderate | 15 (16.3) | 25 (26.6) | 31 (27.9) | 71 (23.9) |
| Poor / undifferentiated | 10 (10.9) | 5 (5.3) | 24 (21.6) | 39 (13.1) |
| Unknown | 1 (1.1) | 0 (0) | 6 (5.4) | 7 (2.4) |
| Staging/differentiation | | | | |
| StageIbG1-2 | 61(66.3) | 75(79.8) | 26(23.4) | 162(54.5) |
| StageIbG3+Ic | 31(33.7) | 19(20.2) | 85(76.6) | 135(45.5) |

# Risk=proportion of people who develop the event over a period of time.

(Tebeu et al: Montpellier , GRELL2004)

Table2: Number of deaths, and risk of death from endometrial cancer at 5 years.

| | N | Number of specific deaths | Number of specific deaths | Risk of specific death |
|---|---|---|---|---|
| None | 92 | 13 | 5 | 5/92=0.054=5,4% |
| Brachytherapy only | 94 | 8 | 5 | 5/94=0.053=5,3% |
| External +/- Brachy | 111 | 21 | 16 | 16/111=0.144=14.4% |

# Incidence Rate=proportion of people who develop the event during the unit of total observation time.

(Tebeu et al: Montpellier 2004)

- **Numbers at risk: 297( specific death)**

| Years | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| No radiotherapy | 92(4) | 88(1) | 87(3) | 84(4) | 80(4) | 76 |
| Brachy only | 94(2) | 92(3) | 89(3) | 86(0) | 86(2) | 84 |
| External/Brachy | 111 | 107 | 100 | 91 | 87 | 82 |

# Incidence Rate=proportion of people who develop the event during the unit of total observation time.

(Tebeu et al: Montpellier 2004)

- Years 0 1 2 3 4 5
- No radiotherapy 92(4) 88(1) 87(3) 84(4) 80(4) 76
- Number of death(D): 4+1+3+4+4= 16 deaths
- Total amount of observation time(T):
- T= (4x1)+(1x2)+(3x3)+(4x4)+(80x5)=431 patients.years
- Rate=D/T=16 deaths / 431 patients.years =0.0371=3.71%
- Rate=Incidence= 3.7 Death per 100 patients each year

# Differerence of Risk (DR)

- DR = attributabe risk
  = Risk in exposed - risk in unexposed
- The attributable risk is the excess of event due to the exposure to a specific condition

# Difference of risk ; attributable risk (DR) (Tebeu et al. GRELL 2004)

- **Risk of death for stage I if exposed to external radiotherapy : R(E+)=14.4%**
- **Risk of death for stage I if not exposed to external radiotherapy : R(E-)=14.4%**
- **The risk attributable to the external radiotherapy: DR= R(E+)- R(E-)= 14.4%-5.4%= 9%**
- **The excess death risk in exposed group is 9%**

# Incidence Rate Ratio(1)

| | Event | No event | Total |
|---|---|---|---|
| Exosed (E+) = | a | b | a+b |
| Unexposed(E-) = | c | d | c+d |

- Incidence rate Ratio= Ratio of Incidence Rates (involves the unit of observation time)

# Incidence Rate Ratio(2)

| | Even | No even | Total |
|---|---|---|---|
| Exposed (E+) = | a | b | a+b |
| Unexposed (E-) = | c | d | c+d |

- Incidence rate of Event in exposed group=IR(E+)= a/a+b,
- Incidence Rate of Event in non-exposed group= IR(E-)= c/c+d
- Incidence Rate Ratio of the event= IR(E+) : IR(E-)

# What is Odds ratio in cohort study (OR)?

(See details in case-control study)

| | Event | No event | Total |
|---|---|---|---|
| Exposed (E+) = | a | b | a+b |
| Unexposed (E-)= | c | d | c+d |

- Odds ratio is Ratio of the Odds of exposure for those presenting the event / Odd of exposure among those without the event
- Then (OR)= (a/c) : (b/d)= ad / bc
- OR in cohort study has the same significance as in case-control study

# Sub-group analysis

1. Sub-group analysis is a process of separate each group of the cohort in different subset for better analysis according to some characteristics
2. Tebeu et al. GRELL 2004, Each group of the cohort study was divided in two sub-group for better analysis
3. Sub-group Ib G1/2
4. Sub-group IbG3, Ic

# What are Confounders?

- Some baseline characteristics that can significantly influence the development of the event, but who were not considered as similar parameter of for the cohort at the entrance
- There is no strong frontier from sub-group to confounders

# The use of Confounders?

- Confounders are identified by their epidemiological impact (age, differentiation of the tumor, myometrial invasion……)
- They can also be identified at univariate analysis (by comparing the outcome in subset of patients presenting the characteristic or not)
- They are then using for more detailed analysis (multivariate analysis, but need software)

# Conclusion

- Cohort study can be conducted in a rural health center
- Analysis of data from cohort study can be easily performed in a setting with no existing calculator

# Useful links:

- Http://www.ccnmtl.columbia.edu/projects/episim/study2f.html
- http://www.gfmer.ch/Medical_education_En/PGC_Yaounde_2004.htm
- http://bmj.bmjjournals.com/collections/statsbk/8.shtml