Cohort Studies and Relative Risks

Richard Ssekitoleko
Yale University
Cohort studies and Relative risks

Richard Ssekitoleko
Objectives

• Define a cohort study and the steps for the study
• Understand the populations in a cohort study
• Understand timing in a cohort study and the difference between retrospective, prospective and ambi-directional cohort studies
• Understand the selection of the cohort population and the collection of exposure and outcome data
• Understand the sources of bias in a cohort study
• Understand the calculation and interpretation of the relative risk
• Understand use of the new-castle Ottawa quality assessment score for cohort studies
Hierarchy of Evidence

Study type

- Observational
  - Descriptive
  - Ecological
  - Cross-sectional
  - Case-control
  - **Cohort**

- Interventional
  - Experiment
  - Randomised Controlled Trial
Study designs and the basic principles

- Randomised controlled trial

- **Cohort study** (Also called Longitudinal, follow up or Incidence study)

- Case-control study

- Cross-sectional study

- Ecological study
Cohort studies

• Observational study type

• Steps in a cohort study
  • Define the target population
  • Get a sample of the target population
  • Identify the exposure status of the sampled members
  • Follow up the members over time to identify new (incident cases) of the disease (outcome)
  • Compare the risk of the outcome in those who are exposed at baseline to those who are not exposed (risk ratio)
Populations studied in Cohort studies

• Open/Dynamic cohort
  • Individuals may enter or leave at anytime
  • losses may occur
  • Defined by changeable characteristic
  • Measure incidence rate

• Fixed Cohort
  • Irrevocable event
  • Does not gain members/Losses may occur
  • Measure incidence rate

• Closed cohort
  • Irrevocable event
  • Does not gain members; no losses occur
  • Measure cumulative incidence
Timing of cohort studies

- Events in a cohort study defined by 3 terms
  - Prospective/ concurrent: Meaning to look forward in time
  - Retrospective: Meaning to look back in time
  - Ambidirectional: Meaning to look both ways

Figure 2: Schematic diagram of concurrent, retrospective, and ambidirectional cohort studies  Grimes et al. Lancet 2002;359:341-45
Derivation and validation of a universal vital assessment (UVA) score: a tool for predicting mortality in adult hospitalised patients in sub-Saharan Africa

Christopher C Moore,1 Riley Hazard,2 Kacie J Saulters,3 John Ainsworth,4 Susan A Adakun,5 Abdallah Amir,6 Ben Andrews,7 Mary Auma,6 Tim Baker,6 Patrick Banura,9 John A Crump,10 Martin P Grobusch,11 Michaèla A M Huson,11 Shevin T Jacob,12 Olamide D Jarrett,13 John Kellett,14 Shabir Lakhi,15 Albert Majwala,6 Martin Opio,16 Matthew P Rubach,17 Jamie Rylance,18 W Michael Scheld,1 John Schieffelin,15 Richard Ssekitoleko,5 India Wheeler,16 Laura E Barnes20

BMJ Glob Health
Retrospective cohort example

- Derivation and validation of a universal vital assessment score
- Methods: Pooled data from hospital based cohort studies from 2009 to 2015
- Analysis involved 5573 patients
- Exposure: Baseline UVA score
- Outcome: Inpatient mortality 996(17.3%)
- Temporal association between exposure and outcome clear
- By time study occurred both exposures and outcomes had occurred
- Lots of missing data with imputation → Information bias
• 2829 (50.8%) were female.

• Median (IQR) age was 36 (27–49) years

• The UVA score included points for temperature, heart and respiratory rates, systolic blood pressure, oxygen saturation, GCS and HIV serostatus.

• The UVA score had an area under the receiver operating characteristic curve (AUC) of 0.77 (95% CI 0.75 to 0.79)

• UVA score Outperformed other scoring systems (MEWS and qSOFA)

• UVA score could help with triage decisions in the study settings
<table>
<thead>
<tr>
<th>Variable</th>
<th>Overall (n=5573)</th>
<th>Survived (n=4607)</th>
<th>Died in-hospital (n=966)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female, n (%)</td>
<td>2829 (50.8)</td>
<td>2323 (50.4)</td>
<td>506 (52.4)</td>
<td>0.27</td>
</tr>
<tr>
<td>Age (years), median (IQR)</td>
<td>36 (27–49)</td>
<td>36 (26–50)</td>
<td>36 (29–46)</td>
<td>0.20</td>
</tr>
<tr>
<td>Temperature (°C), median (IQR)</td>
<td>37.4 (36.3–38.5)</td>
<td>37.4 (36.3–38.5)</td>
<td>37.6 (36.2–38.4)</td>
<td>0.36</td>
</tr>
<tr>
<td>Heart rate (bpm), median (IQR)</td>
<td>100 (85–120)</td>
<td>100 (84–116)</td>
<td>110 (92–128)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Respiratory rate (bpm), median (IQR)</td>
<td>26 (22–32)</td>
<td>24 (20–32)</td>
<td>30 (24–40)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>SBP (mm Hg), median (IQR)</td>
<td>100 (90–120)</td>
<td>100 (90–120)</td>
<td>90 (80–110)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>DBP (mm Hg), median (IQR)</td>
<td>62 (55–80)</td>
<td>65 (60–80)</td>
<td>60 (50–70)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Oxygen saturation (%), median (IQR)</td>
<td>96 (94–98)</td>
<td>96 (94–98)</td>
<td>96 (93–98)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>GCS score, median (IQR)</td>
<td>15 (15–15)</td>
<td>15 (15–15)</td>
<td>15 (13–15)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>HIV-infected, n (%)</td>
<td>2122 (38.1)</td>
<td>1537 (33.4)</td>
<td>585 (60.1)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>CD4 (cells/μL), median (IQR)</td>
<td>72 (23–156)</td>
<td>79 (28–175)</td>
<td>46 (14–112)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>WBC (10^3/μL), median (IQR)</td>
<td>6.0 (3.7–9.7)</td>
<td>6.1 (3.9–9.7)</td>
<td>5.6 (3.1–9.7)</td>
<td>0.003</td>
</tr>
<tr>
<td>Haemoglobin (g/dL), median (IQR)</td>
<td>10.1 (7.8–12.2)</td>
<td>10.4 (8.1–12.5)</td>
<td>8.7 (6.9–11.2)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Platelets (10^3/μL), median (IQR)</td>
<td>180 (105–270)</td>
<td>188 (113–275)</td>
<td>159 (61–256)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Lactate (mmol/L), median (IQR)</td>
<td>3.5 (2.4–4.9)</td>
<td>3.3 (2.4–4.5)</td>
<td>4.2 (2.7–6.1)</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

bpm, beats per minute; bpm, breaths per minute; CD, cluster of differentiation; DBP, diastolic blood pressure; GCS, Glasgow Coma Scale; SBP, systolic blood pressure; WBC, white blood cell concentration.
Hypoglycemia at admission is associated with in-hospital mortality in Ugandan patients with severe sepsis

Richard Ssekitoleko, MBChB, MMed; Shevin T. Jacob, MD, MPH; Patrick Banura, MBChB, MPH; Relana Pinkerton, PhD; David B. Meya, MBChB, MMed; Steven J. Reynolds, MD, MPH; Nathan Kenya-Mugisha, MBChB; Harriet Mayanja-Kizza, MBChB, MS; Rose Muhindo, MBChB, MMed; Sanjay Bhagani, MBBS; W. Michael Scheld, MD; Christopher C. Moore, MD, FACP

Crit Care Med 2011 Vol. 39, No. 10
Prospective cohort example

- Prospective observational study on patients with Sepsis in 3 Ugandan hospitals.
- Analysis involved 418 admitted patients
- Exposure: Admission blood glucose concentration
- Outcome: In hospital mortality 113(27%)
- Measure of association: Hazard ratio
- Results: Significantly higher rates of mortality in patients with hypoglycemia: HR 95% CI 1.9(1.1-3.3)
Table 3. Univariate predictors of survival meeting ≤0.30 criteria and final multivariate model results using Cox regression

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Survived</th>
<th>Died</th>
<th>Hazard Ratio (95% Confidence Interval)</th>
<th>p</th>
<th>Multivariate</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Admission glucose concentration, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Euglycemia (4.4–6.1 mmol/L)</td>
<td>113 (80.7)</td>
<td>27 (19.3)</td>
<td>2.0 (1.2–3.6)</td>
<td>.013</td>
<td>1.9 (1.1–3.3)</td>
<td>.03</td>
</tr>
<tr>
<td>Hypoglycemia (&lt;4.4 mmol/L)</td>
<td>44 (64.7)</td>
<td>24 (35.3)</td>
<td>1.5 (0.96–2.4)</td>
<td>.08</td>
<td>1.6 (0.97–2.5)</td>
<td>.07</td>
</tr>
<tr>
<td>Hyperglycemia (&gt;6.1 mmol/L)</td>
<td>148 (70.5)</td>
<td>62 (29.5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMS, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No AMS</td>
<td>274 (77.6)</td>
<td>79 (22.4)</td>
<td>2.5 (1.6–3.7)</td>
<td>&lt;.001</td>
<td>2.2 (1.5–3.4)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>AMS</td>
<td>31 (47.7)</td>
<td>34 (52.3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White blood cell count, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥4,000 to ≤12,000 cells/μL</td>
<td>162 (79.8)</td>
<td>41 (20.2)</td>
<td>1.7 (1.1–2.4)</td>
<td>.01</td>
<td>1.7 (1.1–2.5)</td>
<td>.013</td>
</tr>
<tr>
<td>&lt;4,000 or &gt;12,000 cells/μL</td>
<td>136 (67.0)</td>
<td>67 (33.0)</td>
<td>1.7 (0.96–2.4)</td>
<td>.08</td>
<td>1.6 (0.97–2.5)</td>
<td>.07</td>
</tr>
<tr>
<td>Heart rate, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤90 beats/min</td>
<td>18 (85.7)</td>
<td>3 (14.3)</td>
<td>1.9 (0.60–6.0)</td>
<td>.27</td>
<td>—</td>
<td>Not significant</td>
</tr>
<tr>
<td>&gt;90 beats/min</td>
<td>286 (72.4)</td>
<td>109 (27.6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bacteremia or fungemia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>248 (74.7)</td>
<td>84 (25.3)</td>
<td>1.3 (0.88–2.0)</td>
<td>.18</td>
<td>—</td>
<td>Not significant</td>
</tr>
<tr>
<td>Positive</td>
<td>57 (66.3)</td>
<td>29 (33.7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Platelets, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥100,000 cells/μL</td>
<td>245 (79.5)</td>
<td>63 (20.5)</td>
<td>2.4 (1.6–3.5)</td>
<td>&lt;.001</td>
<td>1.8 (1.2–2.7)</td>
<td>.007</td>
</tr>
<tr>
<td>&lt;100,000 cells/μL</td>
<td>48 (52.2)</td>
<td>44 (47.8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hospital site, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mulago or Masaka</td>
<td>249 (77.3)</td>
<td>73 (22.7)</td>
<td>2.4 (1.6–3.5)</td>
<td>&lt;.001</td>
<td>1.8 (1.2–2.9)</td>
<td>.004</td>
</tr>
<tr>
<td>Mbarara</td>
<td>56 (58.3)</td>
<td>40 (41.7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
• Temporal association between admission blood glucose and mortality clear

• Researchers identified baseline exposures and then followed up patients (Prospective observational study)
Selecting the cohort population

• Based on study hypothesis
  • Guided by the exposure to be studied e.g smokers vs non smokers

• May be population based cohort based on common exposures e.g smoking, alcohol consumption, exercise and common chronic illnesses

• May be exposure based e.g occupational groups such as road builders
Collection of Exposure and Outcome data

• Study outcomes have not occurred at the beginning of the cohort follow up period.

• Exposures of interest may vary during the study period.
  • May be present at the beginning
  • May occur during the study
  • May stop during the study period

• Temporal association between exposure and outcome is clear
Study population

• At the beginning of follow up all cohort members should be alive not have the outcome of interest

• All members should be at risk of getting the outcome of interest
  • E.g In a study of women involving an outcome of uterine ca, one cannot include women who had a hysterectomy at baseline
Comparison populations in cohort studies

**Cohort Studies**

- **Single cohort**
  - Members of a single population are classified by levels of exposure
  - Comparison group is unexposed or less exposed group
  - Need to account for confounding factors
  - Common exposures: alcohol, smoking, exercise

- **Double cohort**
  - Involves an exposed population and an unexposed population
  - Comparison group may be the general population
Sources of exposure information

- Data may be collected routinely during follow up period

- Sources of data
  - Participant interviews
  - Monitoring data from home or workplace
  - Laboratory monitoring
  - Medical records
Outcome data in a cohort study

- Reports of symptoms and signs
- Medical assessment results
- Medical records
- Disease registry results
- Medical examination results
- Death certificates
Follow up in a cohort study

• All participants need to be tracked throughout the study
  • To get their true outcome
  • To get their person time contribution to the study

• Loss to follow up is a form of bias and reduces validity of results
  • Decreased sample size reducing ability of the study to detect an association if present
  • Those lost to follow up may differ in important ways from those who stay

• May occur due to death, change of residence, migration or participant decision to stop taking part in the study
Minimizing loss to follow up in a cohort study

• Explain need to follow up with participants at start

• Get contact details for participant, friends, relatives or physician

• Maintain regular follow up (Mail, phone or personal contact)

• Follow up on non responses and disappearances promptly

• Offer incentives for follow up e.g transport refund
Analysis in a cohort study

• Need to calculate incidence in the exposed and unexposed groups

• May calculate cumulative incidence or incidence density rate depending on the available information

• Comparing incidence in the exposed and unexposed groups will enable estimation of the relative risk
Risk Ratio (relative risk)

\[
\text{Risk Ratio (RR)} = \frac{\text{RISK of outcome occurrence in exposed}}{\text{RISK of outcome occurrence in unexposed}}
\]

- RR > 1 suggests exposure predisposes to outcome
- RR < 1 suggests exposure protects against outcome
- RR = 1 is null and indicates no association between exposure and outcome
Interpreting the relative risk

• Gives the strength of association between the exposure and outcome

• May not be causal

• Could be explained by random error, confounding or bias

• May represent the cumulative incidence ratio or the incidence density ratio depending on how it is calculated
Risk Ratio Calculations

If, after follow up, the following is seen:

<table>
<thead>
<tr>
<th>Exposure</th>
<th>Disease</th>
<th>Yes</th>
<th>No</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>d₁</td>
<td>h₁</td>
<td>n₁</td>
<td>d₁+h₁</td>
</tr>
<tr>
<td>No</td>
<td>d₀</td>
<td>h₀</td>
<td>n₀</td>
<td>d₀+h₀</td>
</tr>
<tr>
<td>Total</td>
<td>d=d₁+d₀</td>
<td>h=h₁+h₀</td>
<td>n=d+h</td>
<td></td>
</tr>
</tbody>
</table>
Simple Cumulative incidence example

- The table below summarizes a population of 1000 subjects with respect to a particular disease D broken down by sex

<table>
<thead>
<tr>
<th>Disease</th>
<th>D</th>
<th>D</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men</td>
<td>140</td>
<td>60</td>
<td>200</td>
</tr>
<tr>
<td>Women</td>
<td>180</td>
<td>620</td>
<td>800</td>
</tr>
<tr>
<td>Total</td>
<td>320</td>
<td>680</td>
<td>1000</td>
</tr>
</tbody>
</table>

- What is the relative risk of getting the disease associated with being a man as opposed to being a woman?

Relative risk = Risk of disease in men / Risk in women = (140/200) / (180/800) = 0.7 / 0.225 = 3.1
Odds ratios and risk ratios

• How do you interpret the relative risk?
The risk of getting the disease in males is 3.1 times the risk of getting the disease in females

• What is the odds ratio for the disease among men as opposed to women?
Odds of the disease in men: Odds = Risk of disease in men / risk of no disease in men = (140/200)/(60/200) = 0.7/0.3 = 2.3
Odds of the disease among women: Probability of disease in women / Probability of no disease = (180/800)/(620/800) = 0.29
The odds ratio for disease associated with being a man as opposed to a woman. Odds ratio = Odds in men / Odds in women = 2.3 / 0.29 = 7.93

• In which type of study is the odds ratio the preferred measure of association?

• Compare the risk ratio to the odds ratio. What do you conclude?
Incidence density

• Person-time at risk
  • Length of time for each individual that they are in the population at risk
  • Sum of person time for each individual during their stay in study is the total person-time

• When a person is no longer at risk, they no longer contribute to person time e.g. when they get the outcome

• Incidence density
  • Rate of occurrence of new cases of disease during person time of observation in a population at risk of getting the disease
  • Numerator = Number of new cases of disease
  • Denominator = Total person time of observation in population at risk

• A rate and the units are Inverse time (1/time)
Incidence density ratio

- Incidence density ratio = Incidence density in exposed group / Incidence density in unexposed group

<table>
<thead>
<tr>
<th></th>
<th>Total person time of observation</th>
<th>Number of persons with outcome</th>
<th>Incidence density</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed group</td>
<td>A</td>
<td>C</td>
<td>C/A</td>
</tr>
<tr>
<td>Unexposed group</td>
<td>B</td>
<td>D</td>
<td>B/D</td>
</tr>
</tbody>
</table>

- Incidence density ratio = \((C/A)/(B/D)\)
Incidence rate ratio example

A study examined mortality among homeless shelter residents in New York City from 1987 to 1994. There were 15 deaths observed among women aged 25-34, with 728 person-years of observation. Among men aged 25-34, 31 deaths were observed, with 1988 person-years of observation. (Am J Public Health. 1999 Apr;89(4):529-34).

<table>
<thead>
<tr>
<th></th>
<th>Death</th>
<th>Person-time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Women</td>
<td>15</td>
<td>728</td>
</tr>
<tr>
<td>Men</td>
<td>31</td>
<td>1988</td>
</tr>
</tbody>
</table>
The measure of relative risk appropriate for this data is the Incidence density ratio.
   - Incidence density = number of new cases / total person time at risk

The relative risk of mortality among women aged 18-24 compared to men aged 18-24 is the incidence density ratio and is given by:
   \[
   \frac{\text{Incidence density women}}{\text{Incidence density men}} = \frac{15/728}{31/1988} = 1.32
   \]

**Interpretation**
   - The rate of mortality among women was 1.32 times the rate of mortality among men in New York City between 1987 and 1994.

**What is the difference between the Incidence density ratio and the cumulative incidence ratio?**

**How do we get the person time?**
Limitations of cohort studies

• Measurement error (A form of information bias)
  • Commonly errors in exposure measurement
  • Errors in outcome assessment (People may die from competing risks, actual onset of the disease may be missed)

• Confounding- Occurs when a factor is causally associated with both the outcome and exposure under study

• Selection bias (To the different groups and loss to follow up)

• Loss to follow-up (A form of selection bias)
  • If it is related to the exposure or outcome of interest
  • May be differential or non-differential
Loss to follow-up

- A problem with cohort studies is loss to follow-up

- Loss to follow-up may be non-differential i.e. not related to exposure and outcome

- Or differential i.e. is related to exposure and/or outcome. e.g. subjects with poor education who contract HIV die very quickly and do not present to health centres or hospitals. Affects the measure of effect
Cohort Studies

Advantages

• Clear temporal relationship: between exposure and outcome (Compare cross sectional studies)
• Good for rare exposures
• Can evaluate multiple effects of an exposure
• Can minimise biases in exposure measurement
• Directly measures disease incidence or risk

Disadvantages

• Usually expensive and Time consuming (Prospective)
• Poor information on exposures and other key variables (Retrospective)
• Inefficient for disease with long induction and latent periods(Prospective)
• Bias/ confounding
• Changes over time can affect exposure and disease classification
Critical review for cohort studies

**NEWCASTLE-OTTAWA QUALITY ASSESSMENT SCALE COHORT STUDIES**

Note: A study can be awarded a maximum of one star for each numbered item within the Selection and Outcome categories.
A maximum of two stars can be given for Comparability.

### Selection
1. Representativeness of the exposed cohort
   - a) truly representative of the average _______ (describe) in the community ★
   - b) somewhat representative of the average _______ in the community ★
   - c) selected group of users e.g. nurses, volunteers
   - d) no description of the derivation of the cohort ★
2. Selection of the non exposed cohort
   - a) drawn from the same community as the exposed cohort ★
   - b) drawn from a different source
   - c) no description of the derivation of the non exposed cohort
3. Ascertaining of exposure
   - a) secure record (eg. surgical records) ★
   - b) structured interview ★
   - c) written self report ★
   - d) no description ★
4. Demonstration that outcome of interest was not present at start of study
   - a) yes ★
   - b) no ★

### Comparability
1. Comparability of cohorts on the basis of the design or analysis
   - a) study controls for _______ (select the most important factor) ★
   - b) study controls for any additional factor ★ (This criteria could be modified to indicate specific control for a second important factor.)

### Outcome
1. Assessment of outcome
   - a) independent blind assessment ★
   - b) record linkage ★
   - c) self report ★
   - d) no description ★
2. Was follow-up long enough for outcomes to occur
   - a) yes (select an adequate follow up period for outcome of interest) ★
   - b) no ★
3. Adequacy of follow up of cohorts
   - a) complete follow up - all subjects accounted for ★
   - b) subjects lost to follow up unlikely to introduce bias - small number lost - > ___% (select an adequate %) follow up, or description provided of those lost ★
   - c) follow up rate < ___% (select an adequate %) and no description of those lost ★
   - d) no statement ★
References

• Essentials of Epidemiology in Public health: Ann Aschengrau and George R Seage
• Introduction to the field of Statistics: David S Moore, George P McCabe and Bruce A Craig.
Acknowledgements

• Jack Colford: School of Public Health. University of California, Berkeley
• Michael Bates: School of Public Health. University of California, Berkeley.
• Sarah Lewis: Department of Social Medicine. University of Bristol UK.