A brief introduction to epidemiology

Introduction to Medicine
Imm 230

Rita Popat, PhD
Health Research and Policy
Division of Epidemiology
April 22, 2008
Outline

Part I: Epidemiology

- What is epidemiology?
- Basic concepts
- Applications using diabetes and other diseases as examples

Part II: How to Read a Clinical Paper
Definition of Epidemiology

“Epidemiology is the study of the distribution and determinants of health-related states or events in populations and the application of this study to control health problems”

- Last JM
Aims of Epidemiology

- Determine risk factors of various diseases
- Identify segments of the population with highest risk to target prevention and intervention opportunities
- Evaluate the effectiveness of health programs and services in improving health of the population
Clinical epidemiology

Epidemiology/clinical research informs all of the following:

- What diseases should you be looking for in patients?
- Which patients should you screen for disease?
- How are diseases are diagnosed?
- What does a disease diagnosis mean?
- What conditions cause the disease (risk factors)?
- How to prevent disease in your patients?
- How to treat diseases in your patients?
- What is disease prognosis?
- Public health policy/standard of practice.
Disease classification
Case definition or phenotype

General

Phenotypic classification
Presence or absence of clinical or pathologic manifestations

Pathogenetic classification
Underlying pathophysiologic process

Etiologic classification
Underlying causal mechanisms

Specific
Why is disease classification important?

Disease heterogeneity
- Clinical heterogeneity influences diagnosis, management, prognosis
- Etiologic heterogeneity (risk factors may vary)

Diabetes

- TYPE 1 Diabetes Mellitus (T1DM)
- TYPE 2 Diabetes Mellitus (T2DM)
Risk factors

Factors that influence the risk or occurrence of disease are called *risk factors*

(a) demographics (age, sex, race)
(b) genetic factors
(c) physical and biologic agents (drugs, chemical agents)
(d) life-style factors (smoking, diet, exercise)
Figure 1: Algorithm for classification of types of clinical research
Descriptive epidemiology

- Frequency of disease
  - Prevalence: number of people in the population that have the disease at a given point in time
  - Incidence: Rate at which new cases of disease appear in the population (# of new cases / persons at risk)

Note: Prevalence = Incidence x duration

- Who gets the disease? Frequency of disease by location, ethnicity, age, gender, and time period.
Descriptive epidemiology of T1DM

- One of the most common childhood illnesses
  - World-wide estimated 50,000 new cases annually
  - In Caucasian populations: 1-3 per 1000 children by the age of 20 years

- Prevalence in age group 1-15 yrs:
  .05%-0.3%
  - Approximately 123,000 persons aged 0-19 years in the U.S. who currently have T1DM.
Incidence of Type 1 diabetes (ages 0-14 years) in the area

- American non-Hispanic whites are about 1.5 times more likely to develop TIDM as African-Americans or Hispanics
- Asian populations have the lowest rates

Why the geographic and ethnic variation?
  a. Difference in susceptibility genes?
  b. Differences in prevalence of causative environmental factors?
  c. Combination of a and b?
Descriptive epidemiology of T2DM (or NIDDM)

- A pandemic?
  - Affects more than 16 million Americans and 135 million people worldwide

- Prevalence varies in different countries and among different racial and ethnic groups
  - In 2025, estimated worldwide prevalence among adults 5.4%; 300 million adults are estimated to have T2DM
  - The majority of cases of T2DM in the future will occur in developing countries (e.g., India and China)
Diabetes mellitus: A local epidemic

**Figure.** Prevalence of Obesity and Diagnosed Diabetes Among US Adults, 1991 and 2001 (JAMA 2003; 289: 76-79) Data from BRFSS
Descriptive epidemiology of T2DM (or NIDDM)

- The prevalence of T2DM varies greatly both in members of the same ethnic group living in different environments, as well as in different ethnic groups living in the same location.

Genetic and environmental factors?
Figure 1: Algorithm for classification of types of clinical research
### Some important contributions of epidemiology

<table>
<thead>
<tr>
<th>Risk factor</th>
<th>Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smoking</td>
<td>Lung cancer</td>
</tr>
<tr>
<td>DES</td>
<td>Vaginal adenocarcinoma</td>
</tr>
<tr>
<td>Asbestos</td>
<td>Mesothelioma</td>
</tr>
<tr>
<td>OCs</td>
<td>Thromboembolic stroke</td>
</tr>
<tr>
<td>Pesticides</td>
<td>Parkinson’s disease</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Protective agents</th>
<th>Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Folate</td>
<td>Neural tube defects</td>
</tr>
<tr>
<td>Low dose</td>
<td>Stroke</td>
</tr>
<tr>
<td>aspirin</td>
<td></td>
</tr>
<tr>
<td>Smoking</td>
<td>Parkinson’s disease</td>
</tr>
</tbody>
</table>
Cohort Studies

Target population

Exposed

Disease-free cohort

Not Exposed

TIME

Disease

Disease-free

Disease

Disease-free
Measures of association in a cohort study

- Relative Risk/Relative Rate (Risk Ratio)
  *Usual application: Search for causes*

- Absolute difference:
  e.g., Attributable risk in exposed and Population Attributable risk
  *Usual application: Primary prevention impact*
Relative Risk (RR)

\[
RR = \frac{\frac{a}{a + b}}{\frac{c}{c + d}} = \frac{I_e}{I_o}
\]

<table>
<thead>
<tr>
<th></th>
<th>Diseased</th>
<th>Nondiseased</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposure +</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td></td>
<td>a + b</td>
<td></td>
</tr>
<tr>
<td>Exposure -</td>
<td>c</td>
<td>d</td>
</tr>
<tr>
<td></td>
<td>c + d</td>
<td></td>
</tr>
</tbody>
</table>

Incidence (probability) of disease among exposed

Incidence (probability) of disease among unexposed
Relative Rate (RR)

<table>
<thead>
<tr>
<th></th>
<th>Diseased</th>
<th>Person yrs at risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposure +</td>
<td>a</td>
<td>X</td>
</tr>
<tr>
<td>Exposure -</td>
<td>c</td>
<td>Y</td>
</tr>
</tbody>
</table>

\[ RR = \frac{\frac{a}{X}}{\frac{c}{Y}} = \frac{I_e}{I_o} \]

Incidence (rate) of disease among exposed

Incidence (rate) of disease among unexposed
Example: *Body Mass index and risk of clinical Type 2 diabetes in women*

- **Design:** Nurses’ Health Study cohort established in 1976 consists of 121,700 female registered nurses aged 30-55 years who responded to a mail questionnaire about their medical history and health behaviors

  - 114,824 women who did not have DM, CHD, stroke, or cancer in 1976 were followed up for 14 years
  - Height and weight were ascertained in 1976
  - Biennial follow-up questionnaires
  - 2204 cases of T2DM were ascertained in 1990 (1.49 million person-years of follow-up)
  - Other hypothesis investigated: associations of BMI at 18 years and weight gain (since entry into the study) with risk of T2DM

Example: *Body Mass index and risk of clinical Type 2 diabetes in women*

<table>
<thead>
<tr>
<th>BMI</th>
<th>T2DM cases</th>
<th>Person yrs</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;22 kg/m²</td>
<td>55</td>
<td>466,052</td>
</tr>
<tr>
<td>22-22.9 kg/m²</td>
<td>71</td>
<td>194,433</td>
</tr>
</tbody>
</table>

\[
RR = \frac{71/194,433}{55/466,052} = \frac{36.5 \text{ per}100,000}{11.8 \text{ per}100,000} = 3.1
\]

Interpretation: Women with BMI of 22-22.9 kg/m² have a 3-fold increased risk of developing T2DM compared to women with BMI<22.0 kg/m²
Example: **Body Mass index and risk of clinical Type 2 diabetes in women**

<table>
<thead>
<tr>
<th>Body Mass Index</th>
<th>Cases</th>
<th>Person-Years of Follow-up</th>
<th>Age-Standardized Incidence Rate*</th>
<th>Age-Adjusted Relative Risk (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>kg/m²</td>
<td></td>
<td></td>
<td>%</td>
<td></td>
</tr>
<tr>
<td>≥22.0</td>
<td>55</td>
<td>466 052</td>
<td>13.0</td>
<td>1.0 (reference)</td>
</tr>
<tr>
<td>22.0–22.9</td>
<td>71</td>
<td>194 433</td>
<td>37.4</td>
<td>2.9 (2.0 to 4.1)</td>
</tr>
<tr>
<td>23.0–23.9</td>
<td>88</td>
<td>156 770</td>
<td>54.9</td>
<td>4.3 (3.1 to 5.8)</td>
</tr>
<tr>
<td>24.0–24.9</td>
<td>94</td>
<td>142 392</td>
<td>62.9</td>
<td>5.0 (3.6 to 6.6)</td>
</tr>
<tr>
<td>25.0–26.9</td>
<td>227</td>
<td>198 484</td>
<td>103.5</td>
<td>8.1 (6.2 to 10.5)</td>
</tr>
<tr>
<td>27.0–28.9</td>
<td>267</td>
<td>119 662</td>
<td>200.4</td>
<td>15.8 (12.7 to 19.8)</td>
</tr>
<tr>
<td>29.0–30.9</td>
<td>329</td>
<td>84 880</td>
<td>354.5</td>
<td>27.6 (22.7 to 33.5)</td>
</tr>
<tr>
<td>31.0–32.9</td>
<td>263</td>
<td>47 119</td>
<td>521.2</td>
<td>40.3 (33.7 to 48.3)</td>
</tr>
<tr>
<td>33.0–34.9</td>
<td>224</td>
<td>29 885</td>
<td>703.6</td>
<td>54.0 (45.6 to 64.0)</td>
</tr>
<tr>
<td>≥35.0</td>
<td>579</td>
<td>46 636</td>
<td>1190.5</td>
<td>93.2 (81.4 to 106.6)</td>
</tr>
</tbody>
</table>

* Rate per 100 000 persons standardized to the age distribution of length of follow-up in the cohort.

Advantages/Limitations: Cohort Studies

- **Advantages:**
  - Allows you to measure true rates and risks of disease for the exposed and the unexposed groups.
  - Temporality is correct (easier to infer cause and effect).
  - Can be used to study multiple outcomes.
  - Prevents bias in the ascertainment of exposure that may occur after a person develops a disease.

- **Disadvantages:**
  - Can be lengthy and costly! More than 50 years for Framingham.
  - Loss to follow-up is a problem (especially if non-random).
  - Selection Bias: Participation may be associated with exposure status for some exposures.
Case-Control Studies

Target population

Disease (Cases)

Exposed in past

Not exposed

Exposed

Not Exposed

No Disease (Controls)
Measure of Association in case-control studies: Odds Ratio (OR)

<table>
<thead>
<tr>
<th>Exposure (E+)</th>
<th>Disease (D+) Cases</th>
<th>No disease (D-) Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td></td>
<td>c</td>
<td>d</td>
</tr>
</tbody>
</table>

Exposure (E+)

| a+c          | b+d               |

Odds of exposure among cases = \( \frac{P(E|D)}{P(E^{-}|D)} \)

Odds of exposure among controls = \( \frac{P(E|D^{-})}{P(E^{-}|D^{-})} \)

\[
\text{OR} = \frac{a}{a+c} \cdot \frac{b}{b+d} = \frac{ad}{bc}
\]
Rare disease assumption: Odds ratio approximates the Relative Risk

2×2 Contingency table set up from a cohort study

<table>
<thead>
<tr>
<th></th>
<th>Diseased</th>
<th>Non-diseased</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposure +</td>
<td>a</td>
<td>b</td>
<td>a + b</td>
</tr>
<tr>
<td>Exposure -</td>
<td>c</td>
<td>d</td>
<td>c + d</td>
</tr>
</tbody>
</table>

\[
RR = \frac{P(D \mid E)}{P(D \mid E^-)} = \frac{\frac{a}{a + b}}{\frac{c}{c + d}} = \frac{\frac{a}{(a + b)}}{\frac{c}{(c + d)}} = \frac{a}{b} = \frac{ad}{bc}
\]
**Example: Use of cod liver oil during the first year of life among cases of type 1 diabetes and randomly selected control subjects in Norway**

- **Rationale:**
  - Immunomodulatory effects of Vitamin D (e.g., vit D found to prevent autoimmune diabetes in nonobese diabetic mice, and evidence from 2 epidemiologic studies)
  - Cod liver oil is an important source of vit D and long chain n-3 fatty acids (antiinflammatory effects) in the Norwegian population

- **Design:** Population-based case-control study to test hypothesis that cod liver oil or other vitamin D supplements taken during pregnancy or by the child during the first yr of life lowers the risk of T1DM
  - Other risk factors evaluated: duration of exclusive breastfeeding, age at introduction of solid foods, maternal education, maternal smoking during pregnancy
Example: *Use of cod liver oil during the first year of life among cases of type 1 diabetes and randomly selected control subjects in Norway*

- **Diabetes registry:** all new cases of TIDM diagnosed in children aged <15 years have been registered in Norway since 1989
- **Cases:** all children in the diabetes registry with a date of diagnosis b/w 1997-2000 who were born b/w Jan ‘85 and Dec ‘99 were eligible (n=801)
- **Controls:** Children (n=3000) randomly selected from the national population registry who were born in the same time period were eligible
- **Exposure ascertainment:** Mailed questionnaire sent to all eligible subjects
  - **Response rates:** 73% cases and 56% controls

Use of cod liver oil during the first year of life among cases of type 1 diabetes and randomly selected control subjects in Norway

<table>
<thead>
<tr>
<th></th>
<th>Cases</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>197</td>
<td>777</td>
</tr>
<tr>
<td>No</td>
<td>318</td>
<td>834</td>
</tr>
</tbody>
</table>

\[
OR = \frac{(197)(834)}{(318)(777)} = 0.66
\]

OR = 0.66, 95% CI (0.52-0.82).
Interpretation: Using cod liver oil during the first year of life reduces the risk of type 1 diabetes by 34%

Biomarkers of Endothelial Dysfunction and Risk of Type 2 Diabetes Mellitus

- Nested case-control design
  - a case-control study nested within a cohort study

Nurses’ Health Study (cohort)
- \( t_0: 1976 \)
- total \( n=121,700 \)
- bld samples \( n=32,826 \) in 1989-1990

Cases
- \( n=737 \) incident T2DM

Controls \( n=785 \)
- (matched on age, race, fasting status)
Do these data suggest that risk of T2DM increases as levels of E-selectin increase?

Yes, the risk of T2DM increases as E-selectin levels increase (aka dose-response effect).

Advantages and Limitations: Case-Control Studies

- **Advantages:**
  - Cheap and fast
  - Great for rare diseases

- **Disadvantages:**
  - Exposure estimates are subject to recall bias (those with the disease are searching for reasons why they got sick and may be more likely to report an exposure) and interviewer bias (interviewer may prompt a positive response in cases).
  - Temporality is a problem (did exposure cause disease or disease cause exposure?)
### Putative Risk factors for T1DM

<table>
<thead>
<tr>
<th>Genetic factors</th>
<th>TIDM risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>(i) Family history</td>
<td></td>
</tr>
<tr>
<td>(surrogate measure of the combination of genetic and environmental exposures shared by family members)</td>
<td></td>
</tr>
<tr>
<td>(ii) Candidate genes</td>
<td></td>
</tr>
<tr>
<td><em>(HLA-DR and DQ regions)</em></td>
<td></td>
</tr>
<tr>
<td>Environmental factors</td>
<td></td>
</tr>
<tr>
<td>(iii) Viruses</td>
<td></td>
</tr>
<tr>
<td><em>(herpes viruses, mumps, rubella, retroviruses)</em></td>
<td></td>
</tr>
<tr>
<td>(iv) Routine childhood immunization</td>
<td></td>
</tr>
<tr>
<td>(v) Dietary factors</td>
<td></td>
</tr>
<tr>
<td><em>Breast feeding</em></td>
<td></td>
</tr>
<tr>
<td><em>Nitrates, nitrites, or nitrosamines</em></td>
<td></td>
</tr>
</tbody>
</table>

- ↑
- ↓
- ?
Putative Risk factors for T2DM: Genetic factors

(i) Family history

2-6 fold ↑ risk if a parent/sibling has T2DM

(genetic+nongenetic factors)

(ii) Twin studies

↑ concordance in MZ twins

(iii) Genes

(T2DM is a complex genetic disorder)

(iv) Race/ethnicity

![Prevalence of T2DM (%)](chart)

- Non-Indian
- Half-Indian
- Full-Indian

Extent of Pima Indian heritage
## Putative Risk factors for T2DM: Non-Genetic factors

<table>
<thead>
<tr>
<th>Factor</th>
<th>T2DM risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>(i) Obesity</td>
<td></td>
</tr>
<tr>
<td><em>Body Mass Index</em></td>
<td>↑</td>
</tr>
<tr>
<td><em>Waist-to-hip ratio</em></td>
<td>↑</td>
</tr>
<tr>
<td>(ii) Physical Activity</td>
<td>↓</td>
</tr>
<tr>
<td>(iii) Gestational diabetes</td>
<td>↑</td>
</tr>
<tr>
<td>(iv) Low birth weight</td>
<td>↑</td>
</tr>
<tr>
<td><em>(reduced beta-cell mass)</em></td>
<td></td>
</tr>
<tr>
<td>(v) Inflammation</td>
<td>↑</td>
</tr>
<tr>
<td>(vi) Dietary factors</td>
<td>difficult to summarize!!</td>
</tr>
</tbody>
</table>
Modifiable factors and T2DM

Percentage of type 2 diabetes that is potentially preventable by life-style modifications.

T2DM: Low-risk definition
- body mass index < 25 kg/m²
- physical activity equivalent to > 30 min per day of brisk walking
- Good diet (e.g., low in saturated and trans-fat, fiber)
- moderate alcoholic consumption
- Non-smoking

Source: Willet WC. Science, 2002
Causation in human health & disease

- Association does not prove causation
  - If a putative risk factor and the occurrence of an outcome are strongly associated with each other it does not provide evidence that the risk factor causes the disease, only implies that it is correlated with outcome
  - Non-causal explanations may cause a spurious association – study biases (measurement error, selection bias, confounding, sampling error)
Causation in human health & disease

Several aspects of the risk factor-disease association must be examined

- Temporality
- Strength of the association
- Dose-response relationship
- Consistency of findings
- Biologic plausibility
Risk of T2DM with physical activity in the Nurses Health Study

Relative Risk

0.8
0.6
0.4

0-2.0
2.1-4.6
4.7-10.4
10.5-21.7
>21.7

Metabolic equivalent task (MET) hours/week

RRs adjusted for age, smoking, HTN, menopausal status, h/o HRT, family h/o DM, alcohol consumption

p for trend < 0.001
Study validity

- **INTERNAL VALIDITY**
  
  Do we believe the results?

- **EXTERNAL VALIDITY**
  
  Can the results be applied to the target population i.e., beyond the subjects in the study?
Threats to Internal validity:
Non-causal explanations due to study biases

- Confounding
- Selection bias
- Misclassification bias (measurement error)
Confounding

Confounding is defined as a distortion of an exposure-outcome association brought about by the association of another factor with both the outcome and exposure.

Exposure (e.g., physical activity)

Confounding factor (e.g., BMI)

Outcome (e.g., T2DM)
Methods for controlling confounding

- Adjusted for confounders in the statistical analysis

- Matching
  - Matching cases and controls on the confounding factor so that there are no differences in confounder distribution between study groups from the start of the study
Relationship of physical activity vs BMI with T2DM in women.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Inactive*</th>
<th>Active†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Women, No. (%)</td>
<td>24,942 (65.8)</td>
<td>12,936 (34.2)</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
</tr>
<tr>
<td>With diabetes mellitus</td>
<td>985 (3.9)</td>
<td>376 (2.9)</td>
</tr>
<tr>
<td>Adjusted hazard ratio (95% CI)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>1.00</td>
<td>0.73 (0.65-0.83)</td>
</tr>
<tr>
<td>Multivariate‡</td>
<td>1.00</td>
<td>0.85 (0.75-0.97)</td>
</tr>
<tr>
<td>Multivariate§</td>
<td>1.00</td>
<td>0.91 (0.80-1.03)</td>
</tr>
</tbody>
</table>

Abbreviation: CI, confidence interval.
*Energy expenditure was less than 1000 kcal/wk.
†Energy expenditure was at least 1000 kcal/wk.
‡Adjusted for age, family history of diabetes, alcohol use, smoking status, hormone therapy use, hypertension, high cholesterol, dietary factors, and randomized Women’s Health Study treatment groups.
§Adjusted for aforementioned covariates and body mass index.

Weinstein et al. JAMA 292:1188-94
Selection bias

- A form of sampling bias due to systematic differences between those who are selected for study (or agree to participate) and those who are not selected (or refuse to participate).

- Occurs when the subjects selected for a study are not representative of the underlying population in one or more characteristics that affect the association of interest.
Selection bias
Going back to the cod-liver example!

- The response rate among controls was only 50%.
- Non-respondents are generally “less healthy”
- So, non-respondents likely to have a lower prevalence of cod liver or vitamin D use compared to respondents
- How would this selection bias affect the observed measure of association?

- The observed prevalence of cod-liver use among controls = 777 / (834+777)=48%
- Suppose the prevalence of cod-liver use among non-respondents = 30%
- So, the “actual” prevalence of cod-liver use among controls = (48%+30%)/2=39%
- The new numbers (shown in red) based on this prevalence estimate…
### Selection bias
Going back to the cod-liver example!

<table>
<thead>
<tr>
<th></th>
<th>Cases</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>(197)</td>
<td>(777)</td>
</tr>
<tr>
<td></td>
<td>197</td>
<td>628</td>
</tr>
<tr>
<td>No</td>
<td>(318)</td>
<td>(834)</td>
</tr>
<tr>
<td></td>
<td>318</td>
<td>983</td>
</tr>
</tbody>
</table>

**Observed estimate**

$$OR = \frac{(197)(834)}{(318)(777)} = 0.66$$

**Estimate corrected for selection bias**

$$OR = \frac{(197)(983)}{(318)(628)} = 0.97$$

Information bias (measurement error)

Information bias results from either imperfect definitions of study variables (dependent or independent) or flawed data collection procedures (i.e., flawed measurements)

Erroneous classification of an individual (e.g., disease status) or/and attribute (e.g., exposure or genotype) into a category other than that to which it should be assigned.
Information bias → misclassification

Gold-standard

<table>
<thead>
<tr>
<th></th>
<th>Outcome +</th>
<th>Outcome -</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outcome +</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td></td>
<td>TP</td>
<td>FP</td>
</tr>
<tr>
<td>Outcome -</td>
<td>c</td>
<td>d</td>
</tr>
<tr>
<td></td>
<td>FN</td>
<td>TN</td>
</tr>
<tr>
<td></td>
<td>a+c</td>
<td>b+d</td>
</tr>
</tbody>
</table>

Sensitivity = \( \frac{a}{a+c} \)
Specificity = \( \frac{d}{b+d} \)

Can do the same for exposure misclassification
Information bias
Going back to the cod-liver example!

• Assume the cases report their cod-liver exposure accurately (I.e., sensitivity and specificity ~ 100%)
• Suppose, controls have imperfect recall of their exposure:
  • they may over-report or
  • under-report

If controls over-report cod-liver use, then the observed estimate of 0.66 is an overestimate and the actual OR is > 0.66...PROBLEM 😞

If controls under-report cod-liver use, then the observed estimate of 0.66 is an overestimate and the actual OR is <0.66......no problem! 😊

Part II: How to Read a Clinical Paper

Systematically critique a research paper

- Identify the study design, methods for sample selection, exposure and outcome measurements
- Evaluate internal validity of study findings: non-causal relationships and positive features of causation
Relationship of physical activity vs BMI with T2DM in women.

**Context:** Physical inactivity and body mass index (BMI) are established independent risk factors in the development of type 2 diabetes; however, their comparative importance and joint relationship with diabetes are unclear.

**Objective:** To examine the relative contributions and joint association of physical activity and BMI with diabetes.

Weinstein et al. JAMA 292:1188-94
Relationship of physical activity vs BMI with T2DM in women.

**Design, Setting, and Participants:**

- Prospective cohort study of 37,878 women free of cardiovascular disease, cancer, and diabetes with 6.9 years of mean follow-up.
- Weight, height, and recreational activities were reported at study entry. Normal weight was defined as a BMI of less than 25; overweight, 25 to less than 30; and obese, 30 or higher. Active was defined as expending more than 1000 kcal on recreational activities per week.

**Main Outcome Measure:** Incident type 2 diabetes, defined as a new self-reported diagnosis of diabetes.

Weinstein et al. JAMA 292:1188-94
Relationship of physical activity vs BMI with T2DM in women

What is the study design? [Section C]

What is the appropriate measure of association?

- Predictors: BMI and physical activity (continuous measures but categorized for statistical analysis)
- Outcome: Incident type 2 diabetes (dichotomous)
- Measure of association?

Weinstein et al. JAMA 292:1188-94
Table 1. Characteristics of Women by Body Mass Index Category at Baseline*

<table>
<thead>
<tr>
<th>Variable</th>
<th>&lt;25 (n = 19630)</th>
<th>25 to &lt;30 (n = 11700)</th>
<th>≥30 (n = 6548)</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Body mass index, mean (SD)</td>
<td>22.3 (1.7)</td>
<td>27.2 (1.4)</td>
<td>34.4 (4.1)</td>
<td>&lt;.001†</td>
</tr>
<tr>
<td>Physical activity, median (IQR), kcal/wk</td>
<td>650 (1163)</td>
<td>567 (1155)</td>
<td>393 (983)</td>
<td>&lt;.001†</td>
</tr>
<tr>
<td>Age, mean (SD), y</td>
<td>54.5 (7.1)</td>
<td>55.0 (7.1)</td>
<td>53.8 (6.5)</td>
<td>&lt;.001†</td>
</tr>
<tr>
<td>Years of follow-up, mean (SD)</td>
<td>7.0 (0.8)</td>
<td>6.9 (1.0)</td>
<td>6.6 (1.4)</td>
<td>&lt;.001†</td>
</tr>
<tr>
<td>Person-years of follow-up</td>
<td>136876</td>
<td>60690</td>
<td>43488</td>
<td>&lt;.001†</td>
</tr>
<tr>
<td>Family history of diabetes, No. (%)</td>
<td>4166 (21.2)</td>
<td>3126 (26.7)</td>
<td>2121 (32.4)</td>
<td>&lt;.001‡</td>
</tr>
<tr>
<td>Hormone therapy, No. (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>9015 (46.0)</td>
<td>5531 (47.4)</td>
<td>3498 (53.5)</td>
<td></td>
</tr>
<tr>
<td>Past</td>
<td>1775 (9.1)</td>
<td>1260 (10.8)</td>
<td>792 (12.1)</td>
<td>&lt;.001‡</td>
</tr>
<tr>
<td>Current</td>
<td>8803 (44.9)</td>
<td>4889 (41.9)</td>
<td>2243 (34.3)</td>
<td></td>
</tr>
<tr>
<td>Hypertension, No. (%)§</td>
<td>3144 (16.0)</td>
<td>3294 (28.2)</td>
<td>2910 (44.5)</td>
<td>&lt;.001‡</td>
</tr>
<tr>
<td>High cholesterol level, No. (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smoking status, No. (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>9895 (50.4)</td>
<td>5986 (51.2)</td>
<td>3463 (53.0)</td>
<td></td>
</tr>
<tr>
<td>Past</td>
<td>6934 (35.4)</td>
<td>4284 (36.7)</td>
<td>2347 (35.9)</td>
<td>&lt;.001‡</td>
</tr>
<tr>
<td>Current</td>
<td>2787 (14.2)</td>
<td>1417 (12.1)</td>
<td>730 (11.2)</td>
<td></td>
</tr>
<tr>
<td>Alcohol use, No. (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rare</td>
<td>7525 (38.3)</td>
<td>5422 (46.4)</td>
<td>3788 (57.9)</td>
<td></td>
</tr>
<tr>
<td>1-3 drinks/mo</td>
<td>2557 (13.0)</td>
<td>1584 (13.5)</td>
<td>864 (13.2)</td>
<td>&lt;.001‡</td>
</tr>
<tr>
<td>1-6 drinks/wk</td>
<td>6944 (35.4)</td>
<td>3652 (31.2)</td>
<td>1580 (24.1)</td>
<td></td>
</tr>
<tr>
<td>≥1 drink/d</td>
<td>2599 (13.2)</td>
<td>1039 (8.9)</td>
<td>314 (4.8)</td>
<td></td>
</tr>
<tr>
<td>Intake, mean (SD)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fruits and vegetables, servings/d</td>
<td>6.2 (3.6)</td>
<td>6.1 (3.5)</td>
<td>6.0 (3.9)</td>
<td>.01†</td>
</tr>
<tr>
<td>Fiber, g/d</td>
<td>19.4 (6.2)</td>
<td>18.9 (5.7)</td>
<td>18.0 (5.4)</td>
<td>&lt;.001†</td>
</tr>
<tr>
<td>Folate, μg/d</td>
<td>442 (230)</td>
<td>424 (221)</td>
<td>396 (210)</td>
<td>&lt;.001†</td>
</tr>
<tr>
<td>Saturated fat, g/d</td>
<td>19.1 (4.9)</td>
<td>19.8 (4.7)</td>
<td>21.0 (4.8)</td>
<td>&lt;.001†</td>
</tr>
<tr>
<td>Vitamin E, mg/d</td>
<td>6.8 (5.1)</td>
<td>6.6 (5.0)</td>
<td>6.2 (3.5)</td>
<td>&lt;.001†</td>
</tr>
</tbody>
</table>

Abbreviation: IQR, interquartile range.
*Percentages may not sum to 100 due to rounding.
†P Value, analysis of variance.
‡P Value, χ² test of trend.
§Hypertension was defined as self-reported high blood pressure diagnosed by a physician, a self-reported systolic blood pressure of at least 140 mm Hg, or diastolic blood pressure of at least 90 mm Hg.
||High cholesterol was defined as self-reported high cholesterol diagnosed by a physician, self-reported total cholesterol.

Relationship of physical activity vs BMI with T2DM in women.

Weinstein et al.
JAMA 292
:1188-94
Results: Relationship of physical activity vs BMI with T2DM in women

<table>
<thead>
<tr>
<th>Variable</th>
<th>Body Mass Index Category</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Normal Weight (n = 19,630)</td>
</tr>
<tr>
<td>No. (%) of women with diabetes mellitus</td>
<td>178 (0.9)</td>
</tr>
<tr>
<td>Hazard ratio (95% CI)</td>
<td></td>
</tr>
<tr>
<td>Age adjusted</td>
<td>1.00</td>
</tr>
<tr>
<td>Multivariate adjusted*</td>
<td>1.00</td>
</tr>
<tr>
<td>Multivariate adjusted†</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Abbreviation: CI, confidence interval.
*Adjusted for age, family history of diabetes, alcohol use, smoking status, hormone therapy use, hypertension, high cholesterol, dietary factors, and randomized Women's Health Study treatment groups.
†Adjusted for aforementioned covariates and physical activity.
Relationship of physical activity vs BMI with T2DM in women.

**Table 3. Hazard Ratios of Diabetes Mellitus by Physical Activity Category**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Inactive*</th>
<th>Active†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Women, No. (%)</td>
<td>24,942 (65.8)</td>
<td>12,936 (34.2)</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
</tr>
<tr>
<td>With diabetes mellitus</td>
<td>985 (3.9)</td>
<td>376 (2.9)</td>
</tr>
<tr>
<td>Adjusted hazard ratio (95% CI)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>1.00</td>
<td>0.73 (0.65-0.83)</td>
</tr>
<tr>
<td>Multivariate‡</td>
<td>1.00</td>
<td>0.85 (0.75-0.97)</td>
</tr>
<tr>
<td>Multivariate§</td>
<td>1.00</td>
<td>0.91 (0.80-1.03)</td>
</tr>
</tbody>
</table>

Abbreviation: CI, confidence interval.
*Energy expenditure was less than 1000 kcal/wk.
†Energy expenditure was at least 1000 kcal/wk.
‡Adjusted for age, family history of diabetes, alcohol use, smoking status, hormone therapy use, hypertension, high cholesterol, dietary factors, and randomized Women’s Health Study treatment groups.
§Adjusted for aforementioned covariates and body mass index.
Relationship of physical activity vs BMI with T2DM in women.

HRs adjusted for age, smoking, HTN, family h/o DM, ROH use, smoking HT use, etc

Weinstein et al. JAMA 292

Figure 1. Joint Relation of Body Mass Index vs Physical Activity in Quartiles With the Hazard Ratio of Incident Diabetes
Conclusions:

- Although BMI and physical inactivity are independent predictors of incident diabetes, the magnitude of the association with BMI was greater than with physical activity in combined analyses.

- These findings underscore the critical importance of adiposity as a determinant of diabetes.
SECTION B: General Methodologic Issues

- A priori hypothesis clearly stated?
- Source population identified?
- Inclusion criteria?
- Exclusion criteria?
- Number of excluded or refusal (before study) reported?
- Withdrawals reported (during study), explained, and reasonable?
- Withdrawals equal in groups?
- Sample size preplanned?
- Statistical analysis appropriate?
- Adjustment for multiple comparisons
- Adjustment for important variables?
- Results verifiable from raw data?
SECTION K: Specific Methodologic Issues

- Zero time identified?
- Baseline comparability reported?
- Same data collection in all groups?
- Important baseline variables measured, valid and reliable?
- Exposure adequately measured?
- Regular visits during follow-up?
- Co-exposures monitored?
- Duration of follow-up adequate?
- Outcomes defined, measurable, valid?
- Blind assessment of outcomes?
Internal validity- noncausal relationships

- Are the results likely to be affected by selection bias?
- Are the results likely to be affected by information bias?
- Are the results likely to be affected by confounding?
- Are the results likely to be affected by chance variation?
Internal validity - positive features of causation

- Is there a correct time relationship?
- Is the relationship strong?
- Is there a dose-response relationship?
- Are the results consistent within the study?
External validity- generalization of the results

- Can the results be applied to the eligible population?
- Can the results be applied to the source population?
- Can the results be applied to other relevant populations?
Comparison of the results with other evidence

- Are the results consistent with other evidence, particularly studies of similar or more powerful study design?

- Are the results plausible, in terms of a biological mechanism?

So, do you believe the study results and agree with the conclusions?
Contact information

Rita Popat, PhD
Clinical Assistant Professor
Office: HRP Redwood Building, Rm T209
Tel: 650-498-5206
E-mail: rpopat@stanford.edu