Biostatistics workshop series: Analytical Elements

Dr Cameron Hurst
cphurst@gmail.com

CEU, ACRO and DAMASAC, Khon Kaen University

July 9, 2013

What we will cover....

Conventions

The conventions I will use:

Note:.....

Things to note will occur in a green box

Pitfalls:.....

Common mistakes and things to watch out for will occur in a red box

asic elements of a study

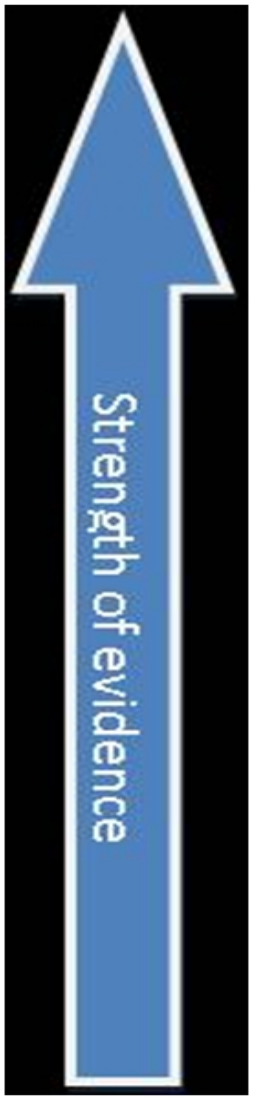
Target population

Should be revision for you

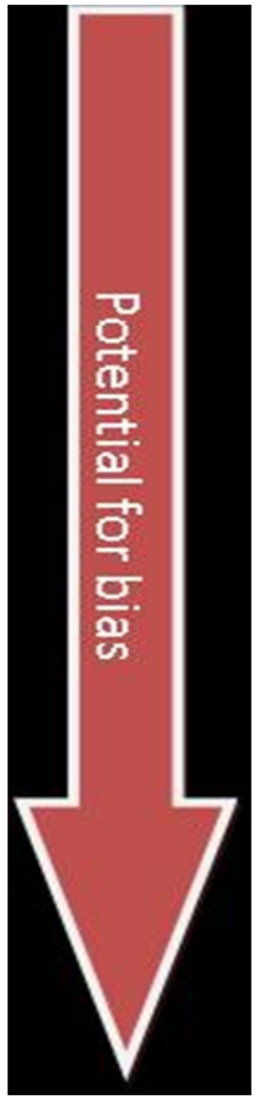
The **TARGET POPULATION** is the group of subjects (e.g. Patients) we want to make statements (inferences) about

The **SAMPLE** is our group of subjects that we hope (through good design) will represent the target population (an unrepresentative sample means our infernecs may not be accurate

Study designs



Study design
Randomized Controlled Trial
Quasi-experiment
Cohort(Longitudinal)
Cross-sectional
Case-Control
Ecological
Case series
Case study



Motivating examples

We will consider four examples (coinciding with the four main dataset you are using from the DAMUS website):

1. The DMHT (Diabetes and Hypertension) dataset;
2. The EMS (Emergency Medical System) dataset;
3. The PCTC (Prospective Cohort of Thai Children) dataset; and
4. TNCS: Thai Nurse Cohort Study.

To keep things simple, we will focus on cross-sectional aspects of all of the datasets (even from the PCTC and TNCS)

To keep it relevant, I will choose examples that are really close to your hearts

The study variables

The study variables can be split into three groups:

1. The **Outcome(s)**: The dependant (clinical endpoints) that is of main interest in our study

The **Effects** which can subdivided into:

3. The **Study Effect**: The main explantory variable (risk factor or intervention) we are interested in (research hypothesis)
4. The **Covariates**: Other variables that may be important to the outcome, OR the outcome-study effect relationship (more later)

Example: TNCS study

For example:

- Our **TARGET POPULATION** is **Thai nurses**
- Our **SAMPLE**: is the **Nurses that returned the mail questionaire in the inital sample of the baseline cohort**
- We might be interested in the **OUTCOME: Intention to leave the nursing profession (yes/no)**
- Our **STUDY EFFECT** might be **Exposure or Risk of exposure to TB (Yes/No)**
- Other **COVARIATES** we might consider are: **Age, Gender, Years experience, Type of nurse, Sector etc**

Example: DMHT

Another example:

- The target population (in our case) are Diabetics in Thailand
- Target sample is T2DM patients (samples) at one of the 598 participating hospitals
- Our otucome might be achievement of all three (A1C, BP and LDL-C) clinical objectives (yes/no)
- Our study effect is Hypertension (yes/no)
- Covariates (we will consider) are: Age, Gender, DM Duration, Smoking

Different types of explantory (X) variables

We have already discussed the **Study effect**, and we have discussed before different types of covariates, but I would like to discuss them again (in detail).

A covariate can be one of three types of variables:

1. An **INDEPENDANT RISK FACTOR**: Something that explains the outcome in it's own right (and does NOT interfer with the study effect)
2. A **CONFOUNDER**: a covariate that (if we don't account for it in the model) will change (diminish or enhance) the reltionship between the outcome and the study effect
3. An **EFFECT MODIFIER**: where the nature of the relationships between the outcome and study effect changes with the LEVEL of the effect modifying variable

Confounders vs Effect modifiers

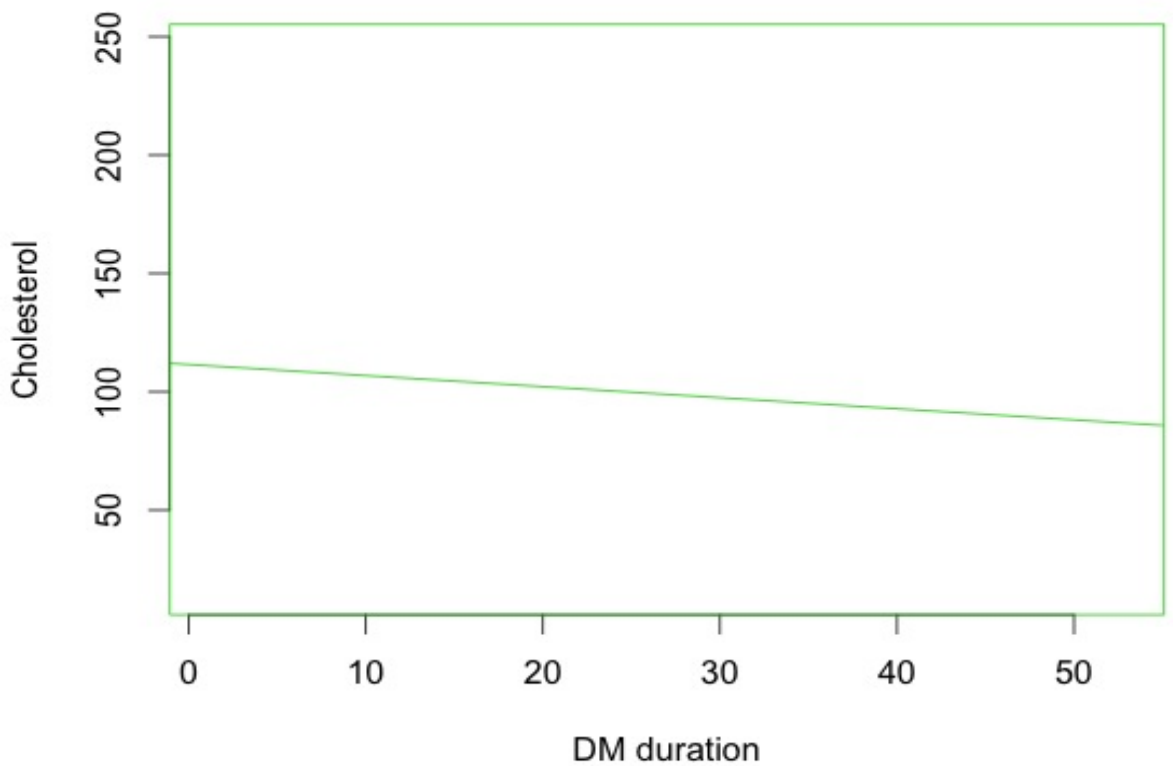
Understandable, people get confused between the difference between **CONFOUNDERS** and **EFFECT MODIFIERS** so I am going to spend the rest of this session going into detail about the difference.

HINT: For **CONFOUNDERS** think about **PRESENCE** and for **EFFECT MODIFIERS** think about **LEVEL**

Now let's consider some examples....

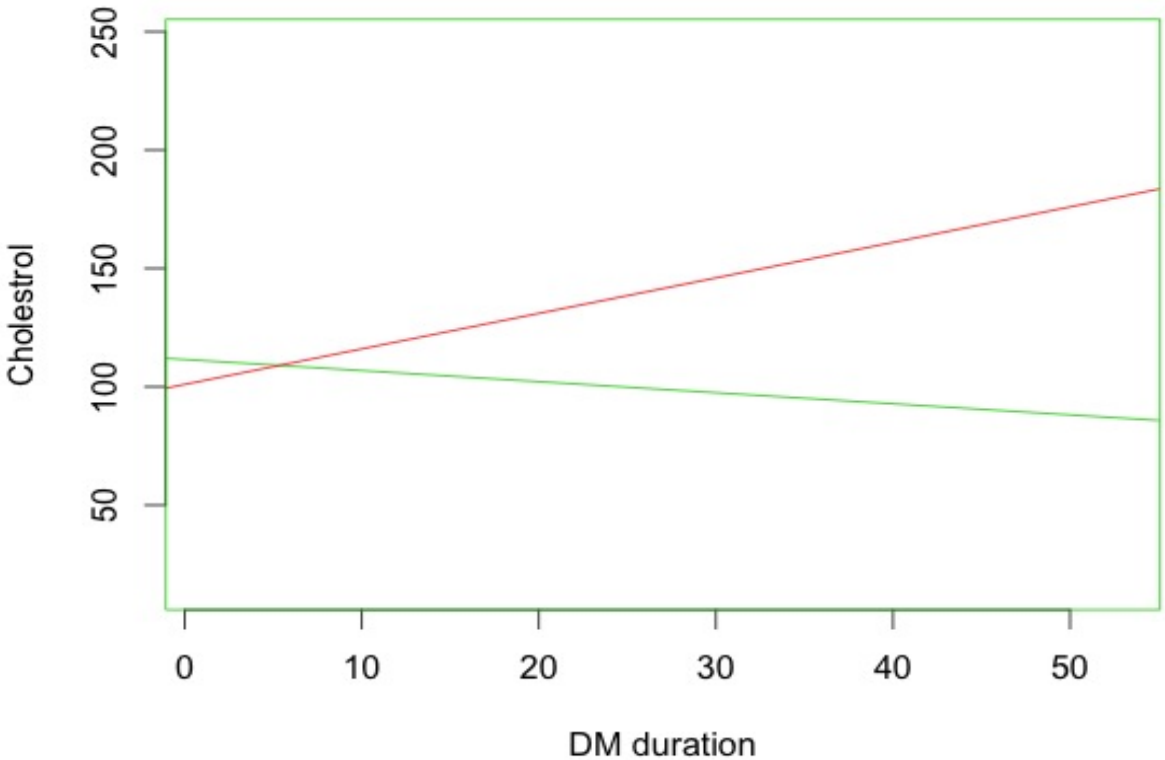
DMHT: Cholesterol vs Duration of DM

Let's consider the relationship between Cholesterol level (Y) against the duration of diabetes.



Crude model: DuraDM significant effect[p<0.01]:
$LDL - C = 111.5 + -0.5 DuraDM$

DMHT: Now let's adjust for age



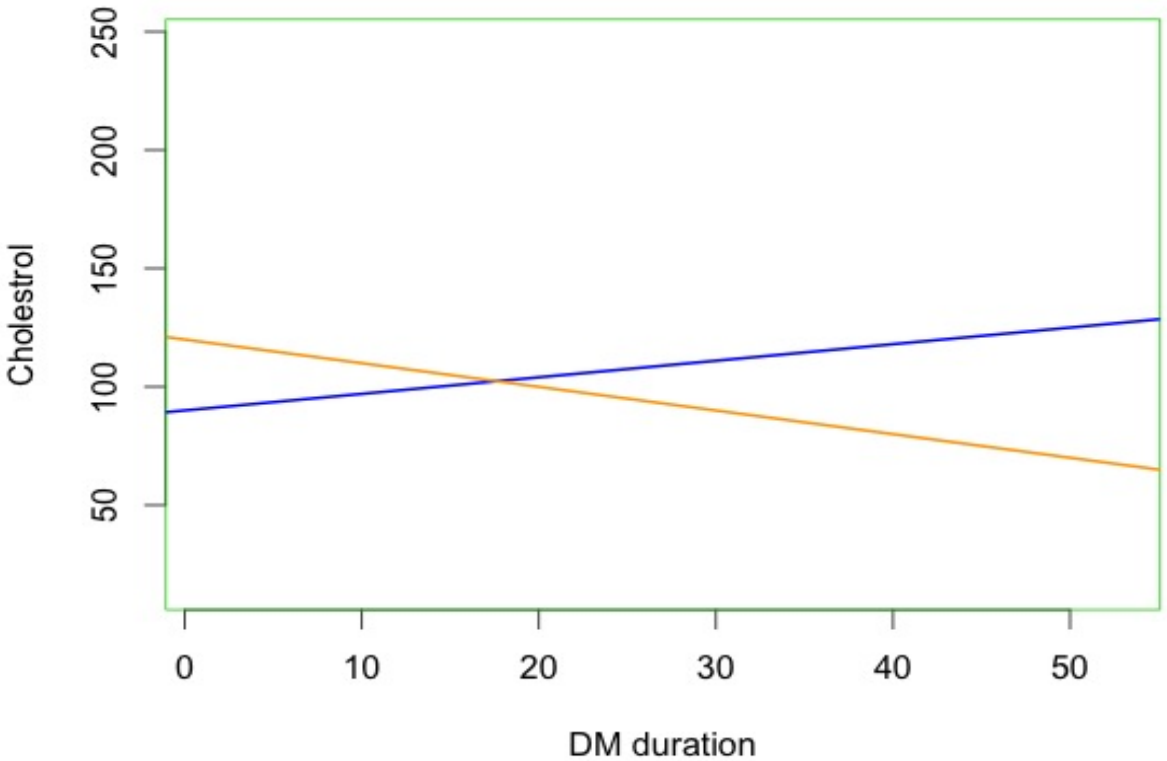
Model1 (green): $LDL - C = 111.5 + -0.5 DuraDM$

Model2 (red): $LDL - C = 111.5 + -0.5 DuraDM + \beta_{Age} Age$

Is Age a confounder or and effect modifier????

Think: **LEVEL** or **PRESENCE**

LDLC vs Dura: Males Vs Females



Females: $LDLC = 120 + -1 DuraDM$

Males: $LDLC = 90 + 0.7 DuraDM$

Is Gender a confounder or an effect modifier???:

Think: **LEVEL** or **PRESENCE**

Your turn: PCTC

- Outcome: Cognative ability (CogAb: low / not low)
- Study effect: Reared by grand parents (Grand parents: no/yes)
- Covariate: Rurality (Rural: urban/rural)

As it is a binary outcome we use binary logistic regression and our measure of association will be ORs

Effect	OR_{Crude}	$OR_{adjusted}$	OR_{Rural}	OR_{Urban}
Grand parents	1.5	2.5*	1.95	2.05
Rurality	2*	2.3*	-	-

* means significant at 0.05 level

Is RURALITY: (a) an Independant risk facor; (b) a confounder; or (c) an effect modifier??

Your turn: TNCS

- Outcome: Intention to Leave (IoL: yes / no)
- Study effect: Exposed to TB (TB: no/ yes)
- Covariate: Age (young/old)

As it is a binary outcome we will (again) use binary logistic regression and our measure of association will be ORs

Effect	OR_{Crude}	$OR_{adjusted}$	OR_{Young}	OR_{Old}
TB	10*	9.5*	25	1.5
Age	5*	4*	-	-

* means significant at 0.05 level

Is Age: (a) an Independant risk facor; (b) a confounder; or (c) an effect modifier??

Your turn: EMS

- Outcome: Dead on arrival for cardiac arrest patients (DoA: yes / no)
- Study effect: CPR administered (CPR: yes / no)
- Covariate: Response time (RT: $<$ 8 mins / $>$8 mins)

Again binary outcome, so ORs
HINT: A tricky one.

Effect	OR_{Crude}	$OR_{adjusted}$	$OR_{<8mins}$	$OR_{>8mins}$
CPR	10*	2	1.2	20
RT	10*	8*	-	-

* means significant at 0.05 level

Is RT: (a) an Independant risk facor; (b) a confounder; or (c) an effect modifier??

THANK-YOU!!

Questions??

YOUR TURN