Dagu workshop at Gondar University May 14-18, 2018

Advanced Epidemiology

Lars Åke Persson

COHORT STUDY

Design and inferences

Suppose that we want to study
the association between breast-feeding
and diarrhoeal disease

Three alternative approaches:
1) A cross-sectional study
2) A cohort study
3) A case-referent study

IMCH, Lars Åke Persson

The Cohort Design

The cohort study

- We follow two groups of children, one of which has been breast fed for 6 months, the other group not
- We register if diarrhoea occurred during a defined period, i.e. 6-11.99 mo.

	DD Yes	DD No	To-tal
BF Yes	50	150	200
BF No	150	50	200

The cohort study

$I_{BF} = 50/200 = 0.25$

$I_{Non\ BF} = 150/200 = 0.75$

RR = 0.75/0.25 = 3

Interpretation?

	DD Yes	DD No	To-tal
BF Yes	50	150	200
BF No	150	50	200

Cohort: stratified analysis

Suppose that we have to take water quality into account

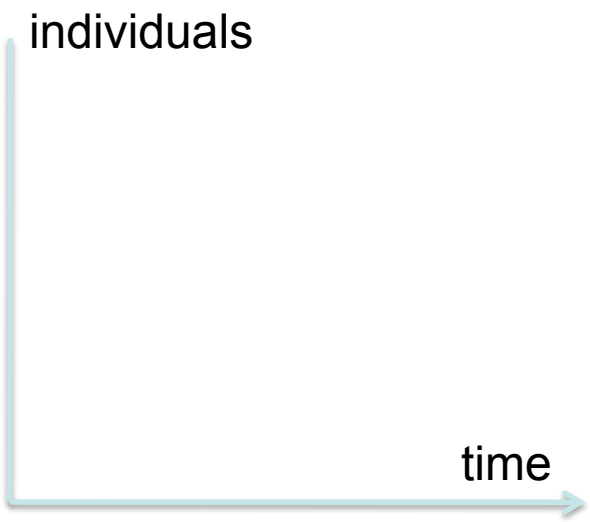
BF	Water	DD Yes	DD No	Total	Incidence	
Yes	Good	5	45	50	0.10	0.25
	Bad	45	105	150	0.30	
No	Good	75	45	120	0.625	0.75
	Bad	75	5	80	0.938	
Total	Good	80	90	170	0.47	0.50
	Bad	120	110	230	0.52	

What do you want to show?

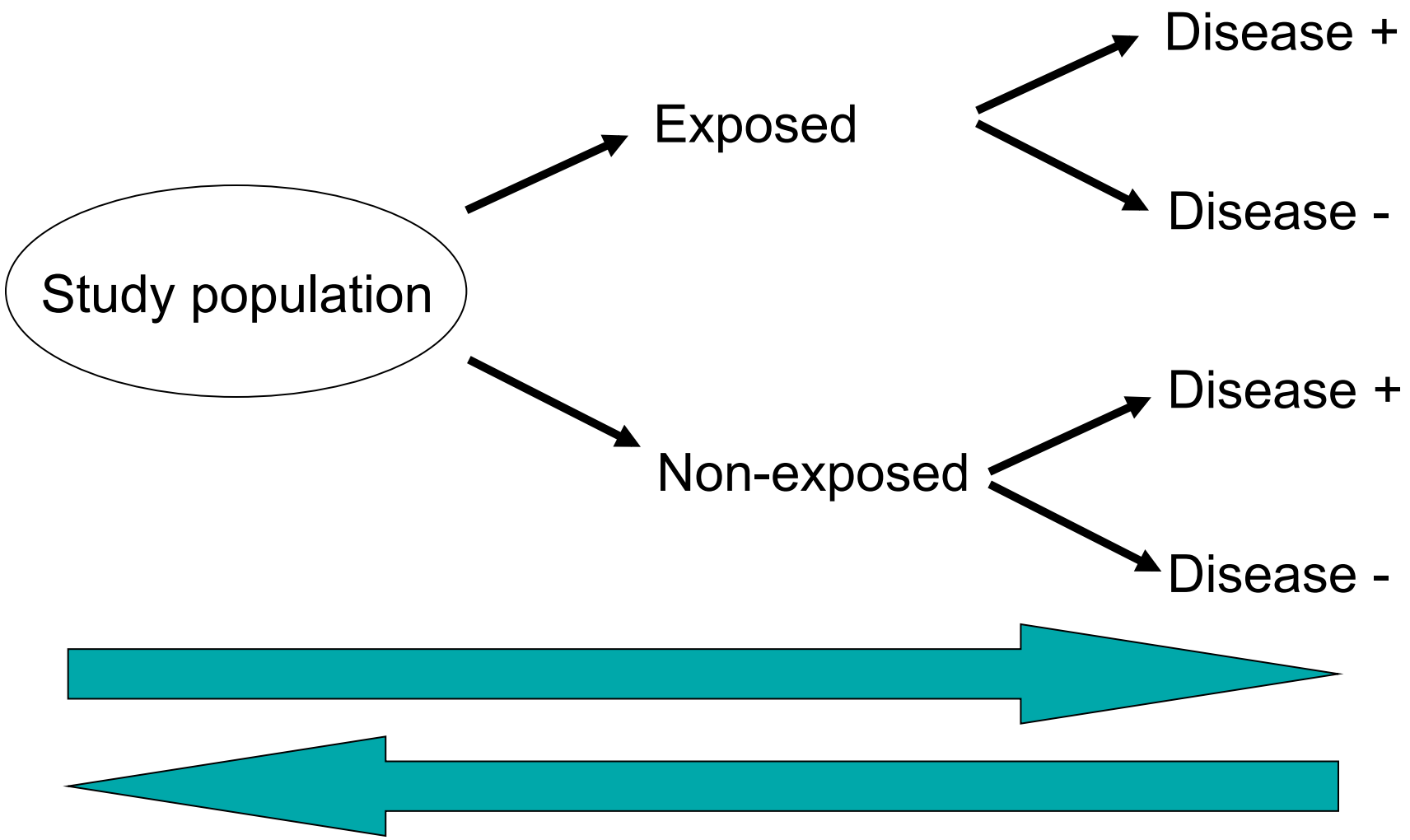
Objective	*Design*
Prevalence	Cross-sectional
Incidence	Cohort
Risk	Cohort; case-referent
Prognosis	Cohort
Treatment effect	RCT

Study base

The “study base” is a population of individuals, each carrying the burden of personal and group risk factors



Cohort study (prospective or retrospective)



Example cohort study for critical reading

Effect of expanded insecticide-treated bednet coverage on child survival in rural Kenya: a longitudinal study

Greg W Fegan, Abdisalan M Noor, Willis S Akhwale, Simon Cousens, Robert W Snow

Summary

Background The potential of insecticide-treated bednets (ITNs) to contribute to child survival has been well documented in randomised controlled trials. ITN coverage has increased rapidly in Kenya from 7% in 2004 to 67% in 2006. We aimed to assess the extent to which this investment has led to improvements in child survival.

Methods A dynamic cohort of about 3500 children aged 1–59 months were enumerated three times at yearly intervals in 72 rural clusters located in four districts of Kenya. The effect of ITN use on mortality was assessed with Poisson regression to take account of potential effect-modifying and confounding covariates.

Findings 100 children died over 2 years. Overall mortality rates were much the same in the first and second years of the study (14·5 per 1000 person-years in the first year and 15·4 per 1000 person-years in the second). After adjustment for age, time period, and a number of other possible confounding variables, ITN use was associated with a 44% reduction in mortality (mortality rate ratio 0·56, 95% CI 0·33–0·96; p=0·04). This level of protection corresponds to about seven deaths averted for every 1000 ITNs distributed.

Interpretation A combined approach of social marketing followed by mass free distribution of ITNs translated into child survival effects that are comparable with those seen in previous randomised controlled trials.

Lancet 2007; 370: 1035–39

See Editorial page 1007

See Comment page 1009

Malaria Public Health and Epidemiology Group, Centre for Geographic Medicine Research—Coast, Kenya Medical Research Institute/Wellcome Trust Research Programme, Nairobi, Kenya (G W Fegan PhD, A M Noor PhD, Prof R W Snow PhD); **Infectious Disease Epidemiology Unit, Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, UK** (G W Fegan, Prof S Cousens Dip Math Stat);

1. Why is this study done
2. What is main exposure?
3. What is “dynamic” cohort?
4. Cluster sampling, what is that?
5. How was confounding managed?
6. Effect modification?
7. Comment result

1. Why is this study done?

- Need to go from "efficacy" to program "effectiveness"
- Influence of contextual factors
- Effects in subgroups
- Cost and effectiveness

Background The potential of insecticide-treated bednets (ITNs) to contribute to child survival has been well documented in randomised controlled trials. ITN coverage has increased rapidly in Kenya from 7% in 2004 to 67% in 2006. We aimed to assess the extent to which this investment has led to improvements in child survival.

2. What is main exposure?

At each cross-sectional survey, bednet use among surviving children was ascertained, including whether the child was reported as usually using a net, and if so, whether they used a net the previous night. For all those who were reported as usually using a net, whether this net had been treated with insecticide, and if so, whether it had been treated within 6 months, was also recorded.

3. What is “dynamic” cohort?

- Closed cohort: no one entering after start
- Open or dynamic cohort: may enter after start

4. Cluster sampling, what is that?

- Here 18 randomly selected communities within four sentinel districts
- Feasible sampling strategy to get a representative sample (e.g. DHS)
- Consequences for analysis: variation within and between clusters
- More about this in cluster course (third semester)

Methods A dynamic cohort of about 3500 children aged 1–59 months were enumerated three times at yearly intervals in 72 rural clusters located in four districts of Kenya. The effect of ITN use on mortality was assessed with Poisson regression to take account of potential effect-modifying and confounding covariates.

5. How was confounding managed?

- Multivariable analyses with potential confounding factors
- A factor may be a confounder and/or effect modifier

non-users, and Poisson regression was used to control for potential confounding of the association between ITN use and mortality, with a random effect term used to model variation between communities. To estimate the number of deaths averted as recommended[2,12] we applied the rate ratios obtained to the unexposed group.

6. Effect modification?

	Model A	Model B
Use ITN	0·51 (0·30–0·87; p=0·01)	0·56 (0·33–0·96; p=0·04)
District†		
Kisii	0·12 (0·06–0·24; p<0·0001)	..
Kwale	0·26 (0·16–0·41; p<0·0001)	..
Makueni	0·06 (0·02–0·17; p<0·0001)	..
Year (2005 *vs* 2006)	0·75 (0·50–1·13; p=0·17)	0·75 (0·50–1·13; p=0·18)
Sex (female)	0·68 (0·45–1·02; p=0·06)	0·66 (0·44–0·99; p=0·05)
Child's age‡		
6–11 months	0·96 (0·51–1·81; p=0·90)	0·94 (0·50–1·77; p=0·85)
1–2 years	0·87 (0·49–1·52; p=0·62)	0·84 (0·48–1·47; p=0·54)
2–3 years	0·28 (0·14–0·58; p=0·001)	0·28 (0·14–0·57; p=0·0003)
3–4 years	0·14 (0·06–0·35; p<0·0001)	0·14 (0·05–0·35; p<0·0001)
4–5 years	0·09 (0·03–0·30; p<0·0001)	0·09 (0·03–0·30; p<0·0001)
Mother had some education§	0·87 (0·44–1·72; p=0·73)	1·11 (0·61–2·03; p=0·73)
Mother >30 years old	0·88 (0·56–1·37; p=0·58)	0·83 (0·53–1·29; p=0·41)
Wealth quintile¶		
Second	1·11 (0·60–2·04; p=0·75)	1·12 (0·60–2·10; p=0·73)
Third	1·30 (0·72–2·34; p=0·39)	1·34 (0·72–2·54; p=0·34)
Fourth	1·18 (0·63–2·22; p=0·61)	1·30 (0·66–2·55; p=0·45)
Fifth	1·23 (063–2·37; p=0·55)	1·44 (0·71–2·94; p=0·31)
Tertile of distance to health facility‖		
Middle	1·17 (0·69–1·98; p=0·57)	1·07 (0·59–1·94; p=0·83)
Furthest	1·08 (0·63–1·86; p=0·78)	1·01 (0·52–1·96; p=0·97)
Quartiles of net coverage of each enumeration area**		
Second	1·01 (0·59–1·72; p=0·98)	1·25 (0·57–2·72; p=0·58)
Third	0·55 (0·28–1·01; p=0·08)	0·58 (0·24–1·41; p=0·23)
Fourth	1·32 (0·73–2·37; p=0·36)	1·11 (0·49–2·54; p=0·80)

Data are rate ratio (95% CI; p value). *Cluster (ie, community) level random effects parameter (θ) in model A was 0·00 (95% CI 0–0; p=1·0), and in model B was 0·67 (95% CI 0·29–1·53; p<0·0001). †Reference group is Bondo. ‡Reference group is 1–5 month olds. §Taken from models with 97 rather than 100 deaths due to missing values of maternal education; all other parameter values remained essentially unchanged. ¶ Reference group is most poor—ie, first wealth quintile. ||Middle corresponds to a range of about 30 min to 1 h. Furthest corresponds to about 1–5 h. **Second ranges from 23% to 30%, third ranges from 30% to 35%, fourth from 35% to 62%.

Table 2: **Adjusted mortality rate ratios***

7. Comment result

- Size of effect
- Bias?
- Money well spent

The size of effect is particularly striking because it is greater than the pooled estimate of protection under conditions of efficacy trials.[1] We attempted to control for several important covariates associated with child survival, but there will be covariates we have not included and our definition of exposure might have an element of bias. Unlike carefully controlled trials, and given the ethical constraints of repeated investigations of cohorts of children not sleeping under nets, we cannot be completely certain about the reported use of the ITNs by children who died. A further possible source of bias is that some of those who had either out-migrated or were age-censored by reaching age 5 between the midpoint survey and the last cross-sectional survey could have been deaths. However, among this subset of children the proportion of ITN users at the time of the midpoint survey was the same as the overall population and any bias that such missed deaths could have introduced is likely to be non-differential. Further, given the advanced age of these children, the probability of death among such children has to be fairly small (table 2).

The national estimate of protection afforded through the expansion of the Kenyan ITN delivery programme might be more conservative than we have shown in districts that were able to achieve over 67% coverage. We are nevertheless confident that a substantial effect on child survival was achieved during the expansion phase of the ITN strategy and might have reduced by over a third the numbers of childhood deaths in high coverage districts in 2006. Donor agencies should regard this as money well spent and recognise that the challenge is now to maintain and increase funding to expand coverage further.

Cohort study

- To study incidence, natural history
- To analyse risk factors, relative risk
- Cause and effect, temporal sequence
- Retrospective cohorts often cheaper
- May look at several outcomes
- Confounding major problem
- Selection bias, loss to follow up may cause bias

Cohort studies

- Starts with the population at risk
- Measures characteristics at baseline
- Follow-up over time to measure incidence of disease
- Compare event rates in the cohort (among those who have and those who don’t have the characteristics of interest)
- Confounding may be a problem

Confounding

- An alternative explanation for observed association between exposure and disease or
- A mixing of effects – the association between exposure and outcome is distorted
- May distort the true association toward the null or away from the null

Confounding

- A variable can not be a confounding factor if it is located on the causal chain from exposure to outcome
- Confounding can be controlled in the design (matching, restriction) or in the analysis (stratified analysis, adjustment)
- Confounding factors are (other) risk factors for disease
- If comparing crude and adjusted measures of association you may judge whether confounding was present or not

Confounding

- Randomisation in trials – to control for known and unknown confounding factors
- There may be residual confounding – addressed in the discussion

Quiz 2

1. *Which of the following measures is used frequently as a denominator to calculate the incidence rate of a disease*?
   a) number of cases observed
   b) number of new cases observed
   c) number of asymptomatic cases
   d) person-years of observation
   e) persons lost to follow-up

2. *When a new treatment is developed that prevents death but does not produce recovery from the disease, the following will occur:*
   a) prevalence of the disease will decrease
   b) incidence of the disease will increase
   c) prevalence of the disease will increase
   d) incidence of the disease will decrease
   e) incidence and prevalence of the disease will decrease

3. *Which of the following is an advantage of a retrospective study?*
   a) There is little of no bias in assessment of exposure to the factor
   b) Multiple disease outcomes following a selected exposure can readily be studied
   c) Dependence on recall by subjects in the study is minimized
   d) It is possible to determine the true incidence rate of the disease
   e) It may be used to study etiology of a rare disease

4. *True statements concerning cohort studies include which of the following*:
   a) Cohort studies are longitudinal in design
   b) Subjects are selected on the basis of characteristics present before the onset of the condition being studied
   c) Subjects are observed over time to determine the frequency of occurrence of the condition under study
   d) They are primarily descriptive, rather than analytic