Statistical Issues and Methods in Infertility Research

Vanja Dukic, PhD.

http://health.bsd.uchicago.edu/dukic/dukic.research.html

Dept. of Health Studies, University of Chicago

Based on joint work with:

G. Buck-Louis, P. Haegerty, T. Louis, C. Lynch, L. Ryan,

L. Schieve, E. Schisterman

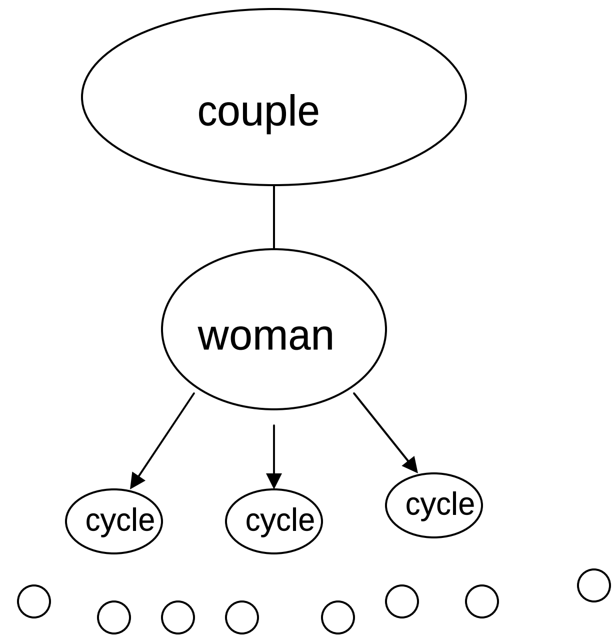
Main statistical issues: overview

- General structure in infertility data
- Study design
- Unit of analysis
- Time scale, confounders
- Multiple adjustments and interactions
- Correlated data methods
- Further methodology development

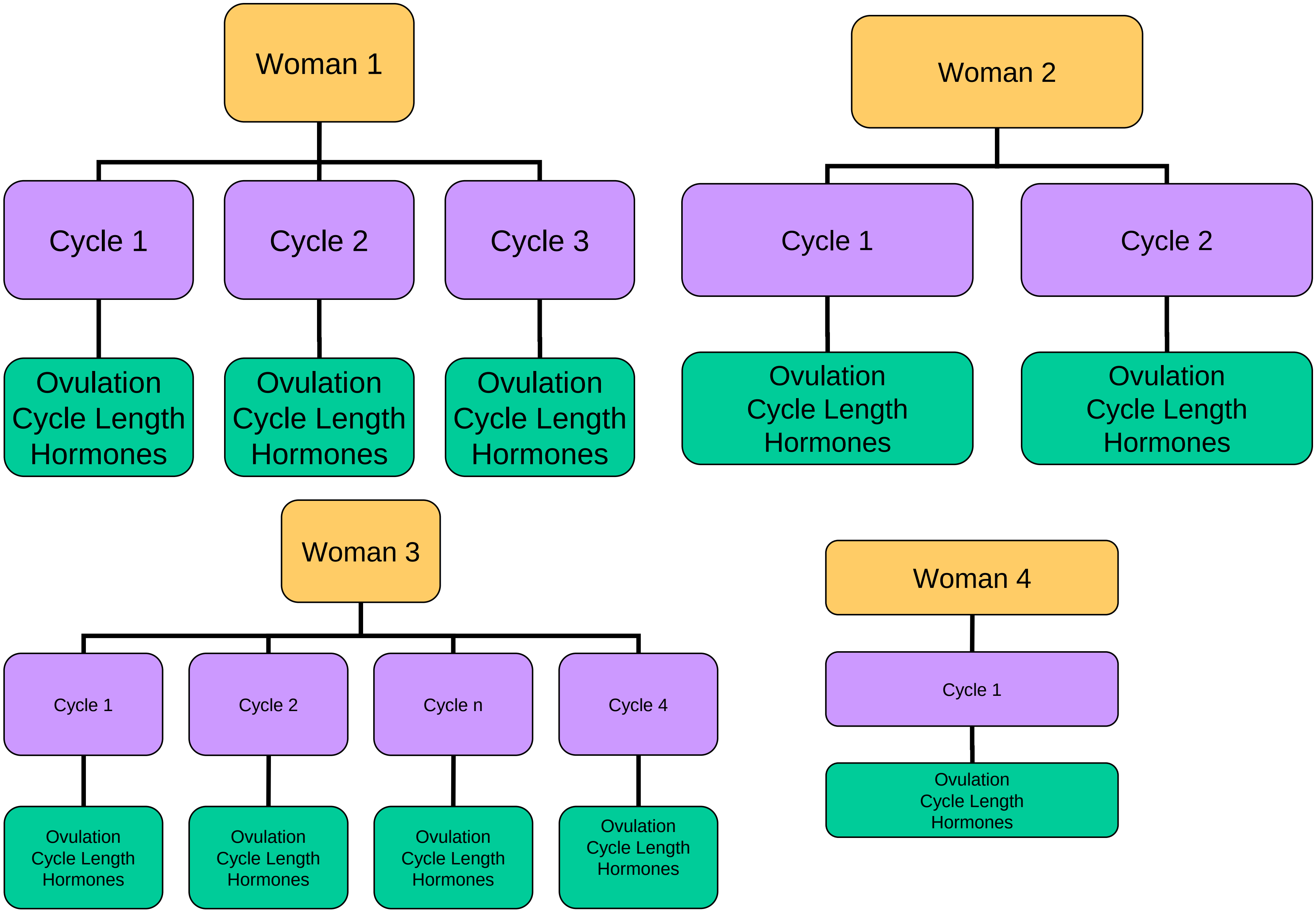
Structure in infertility data

Hierarchical (multi-level) nature of data – clustering at the level of:

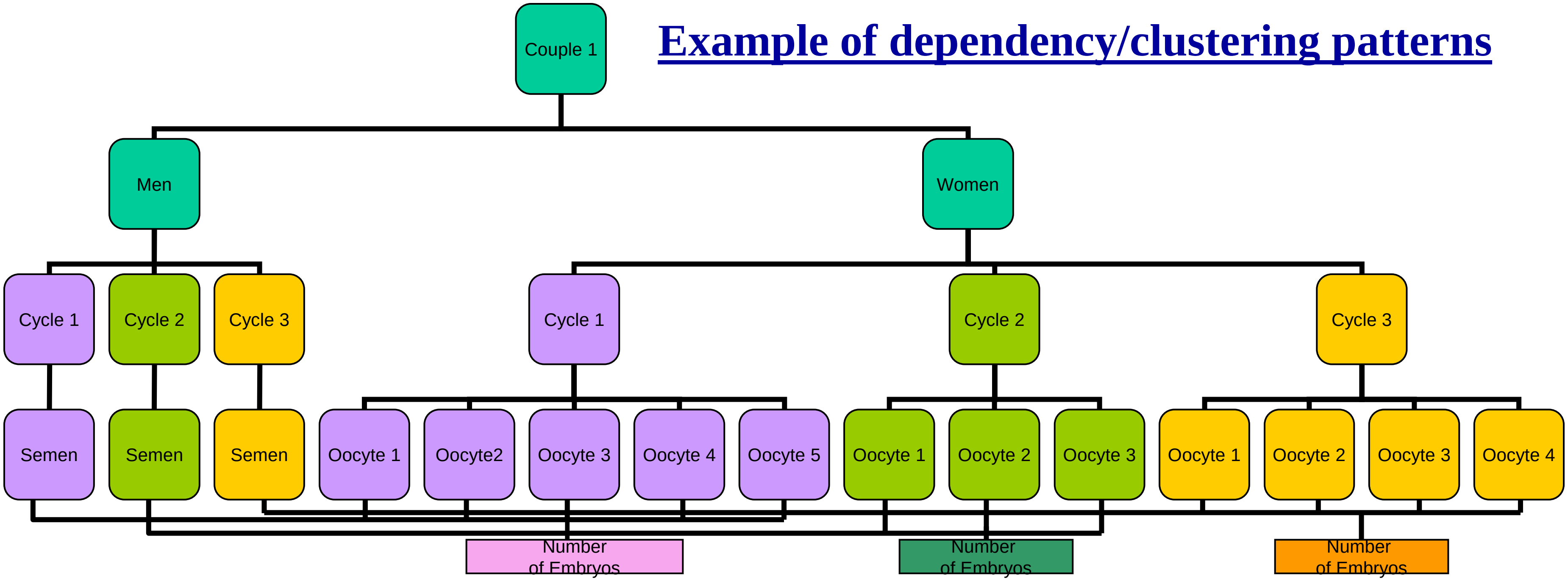
- Couples
- Women
- Menstrual cycles
- Days
- Oocytes…

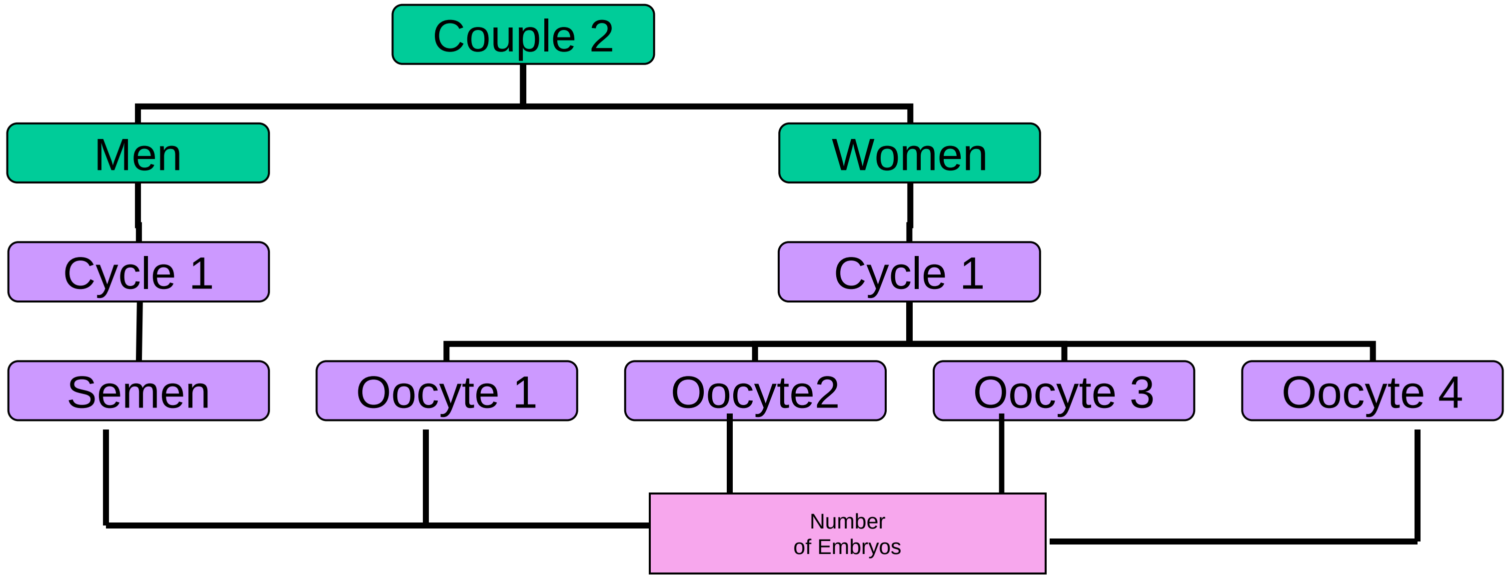


Example of dependency/clustering patterns



Example of dependency/clustering patterns





Dependency in pregnancy data

- Multi-level clustering induces complex dependency pattern among pregnancy outcomes
- This dependency has to be considered when designing a study, and accounted for in data analysis
- Ignoring this dependency can lead to loss of power, inefficiency of analysis and bias in results

Design issues in Infertility Treatment Studies

Design issues:

How best to **design** studies to answer the questions of interest

-Prospective Studies?

-Retrospective Studies?

What is the right **unit of analysis**:

Couple

Woman/Man

Cycle?

Oocyte/Embryo

What (time-varying) **covariates** should be measured?

The timely question: is ART harmful?

ART procedures increased 20 times since 1986

Number of live ART infants increased 100 times since 1986

Are there adverse effects?

-increased risk for decrements in gestation and birth size (Australia)

-birth defects (Bergh *et al.* 1999; Hansen *et al.* 2002)

-developmental disabilities (Stromberg *et al.* 2002)

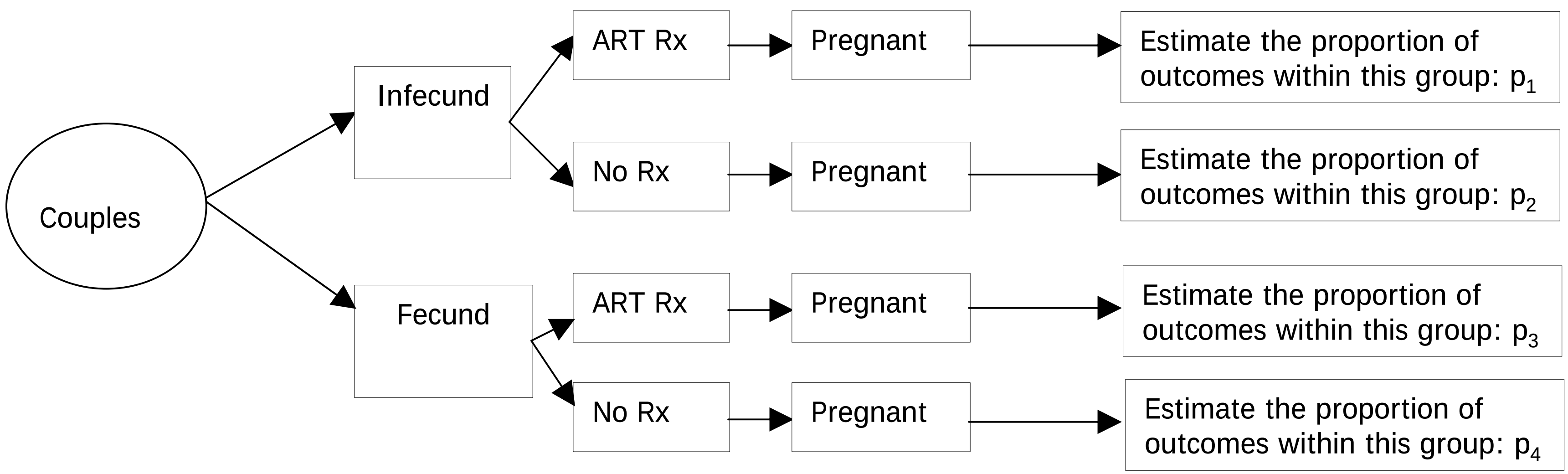
- Jackson *el al.* 2004 compared IVF infants to spontaneously conceived and found (after adjusting for age and parity) 2 fold increase in risk of:

-perinatal mortality

-preterm delivery

-low birth weight and SGA

The main problem with the question… we need a randomized clinical trial



Cannot estimate p_2 and (almost) p_3 and thus can't differentiate absolute ART treatment effects from the underlying fecundity impairments!

Design: prospective vs retrospective studies

Prospective studies

pros: can capture timings and ordering of exposures

can capture couple interaction

cons: more expensive

Retrospective studies:

pros: cheaper

more convenient for rare outcomes

cons: potential bias

imprecision

timing and correlation of exposure patterns poor

Design: “unit of analysis” and “time scale”

Deciding on the right **unit of analysis**:

-must be objective-driven and planned in advance

(couple, woman, man, cycle, embryo, oocyte…)

Deciding on the right **time scale**

-calendar time

-cycle time

No switching between units and scales should be done after the study has begun: loss of power may occur, and confounders may emerge

Further decisions about what covariates to record depend on both the time scale and unit of analysis

Analysis of Infertility Treatment Studies

Analysis: Once designed, how to best to analyze the data? Three main issues to pay attention to:

- Multiple adjustment : main effects ("mean model") needs to be well specified (the "Goldilocks method"):
  - All predictor variables of interest need to be included
  - Only those predictor variables whose presence in the model can be theoretically justified should be left in the model
- Interactions:
  - Interactions among variables need to be included if the effect of one variable changes depending on the value of another
- correlated outcomes:
  - Avoid the correlation by using only single pregnancy (first, last, random)
  - Use multiple pregnancies and ignore the correlation
  - Use multiple pregnancies and model the correlation explicitly
  - Use multiple pregnancies and use previous outcome as the predictor for the next

Example: the impact of within-woman correlation

- Simple simulation: birth weight for 1 (1-p)% or 2 (p%) pregnancies per woman; exposure X (binary)
  - 2 scenarios:
    - X not changing over time
    - X randomly changing from pregnancy 1 to pregnancy 2 (either 0 or 1 with probability 0.5)
- True correlation between births **rho**
- Naïve standard error (correlation ignored) compared to the true one (based on rho)
- Simple linear regression model:

$$Y_{i,j} = b_0 + b_1 * X_{i,j} + e_{i,j}$$

Scenario 1: non-changing X

N	% with 2	corr.	Naive (x1000)	True (x1000)	Ratio
N=200	p=0.33	rho=0.4	15.04	18.02	0.834
		rho=0.8	15.04	21.01	0.716
	p=0.50	rho=0.4	13.33	16.89	0.789
		rho=0.8	13.33	20.44	0.652
	p=0.67	rho=0.4	11.98	15.82	0.757
		rho=0.8	11.98	19.66	0.609
N=400	p=0.33	rho=0.4	7.52	9.01	0.834
		rho=0.8	7.52	10.50	0.716
	p=0.50	rho=0.4	6.67	8.44	0.789
		rho=0.8	6.67	10.22	0.652
	p=0.67	rho=0.4	5.99	7.91	0.757
		rho=0.8	5.99	9.83	0.609

Scenario 2: randomly-changing X

Women	% with 2	corr	Naive (x1000)	True (x1000)	Ratio
N=200	p=0.33	rho=0.4	15.04	12.05	1.248
		rho=0.8	15.04	9.07	1.658
	p=0.50	rho=0.4	13.33	9.7	1.364
		rho=0.8	13.33	6.22	2.143
	p=0.67	rho=0.4	11.98	8.13	1.473
		rho=0.8	11.98	4.29	2.793
N=400	p=0.33	rho=0.4	7.52	6.03	1.248
		rho=0.8	7.52	4.53	1.658
	p=0.50	rho=0.4	6.67	4.89	1.36
		rho=0.8	6.67	3.11	2.143
	p=0.67	rho=0.4	5.99	4.07	1.473
		rho=0.8	5.99	2.14	2.793

Handling correlation in reproductive data

- Pretend it's not there
  - Incorrect inference (CIs, Hypothesis tests)
- Design it away by picking one outcome per woman
  - Inefficiency of estimates and loss of power to detect effects
- Adjust for it:
  - Robust Standard Errors (Huber-White adjustment)
    - Easy and approximately correct (Stata, SAS…)
    - Needs strong assumptions about missingness (MCAR)
    - No heterogeneity estimates/individual predictions
    - Potentially not as efficient as other methods
  - GEE (software: Stata, SAS, etc.)
  - Hierarchical or Bayesian models (software: SAS,WinBUGS, HLM, MIXOR, etc.)

Handling dependency in pregnancy data

- GEE (Stata or SAS,…)
  - specification of longitudinal correlation that weighs clustered data and makes estimation more efficient
  - Robust standard errors ("generalized" Huber-White clustering adjustment method) valid even if the longitudinal correlation was incorrectly specified
  - No heterogeneity and individual prediction

Bayesian Models
(aka mixed, multi-level, hierarchical…)

- Some can be fitted using SAS, WinBUGS, or custom software
- Allow for explicit modeling of heterogeneity of outcomes
- Allow more elaborate biological models (such as EU)
- Allow incorporation of prior information: useful for clinical prediction and patient counseling
- However, level-specific parameters (aka random effects) must be unrelated to the covariates in the model for inference to be valid (ie needs non-endogenous covariates)
- Extendable, can accommodate a variety of sources of information (including prospective and retrospective study combination, multi-center studies, and meta-analyses)

CPP Example
(Buck et al., 2005)

- Effects of smoking on SGA
- Women with adverse pregnancy outcomes are twice as likely to repeat such outcomes in comparison to women who had healthy ones.
- Traditionally, this dependency has been:
  - ignored, or
  - adjusted for via including Hx as a confounder
  - designed away by restricting the analysis to one pregnancy per woman

CPP Example, cont.
(Buck et al., 2005)

Sample

- 2,211 (17%) primigravid pregnant women with 2+ prospectively followed pregnancies

Exposures & Outcomes

- Exposures
  - Cigarette smoking during pregnancy
- Outcomes
  - Small-for-gestational age (SGA)

CPP Example, cont.
(Buck et al., 2005)

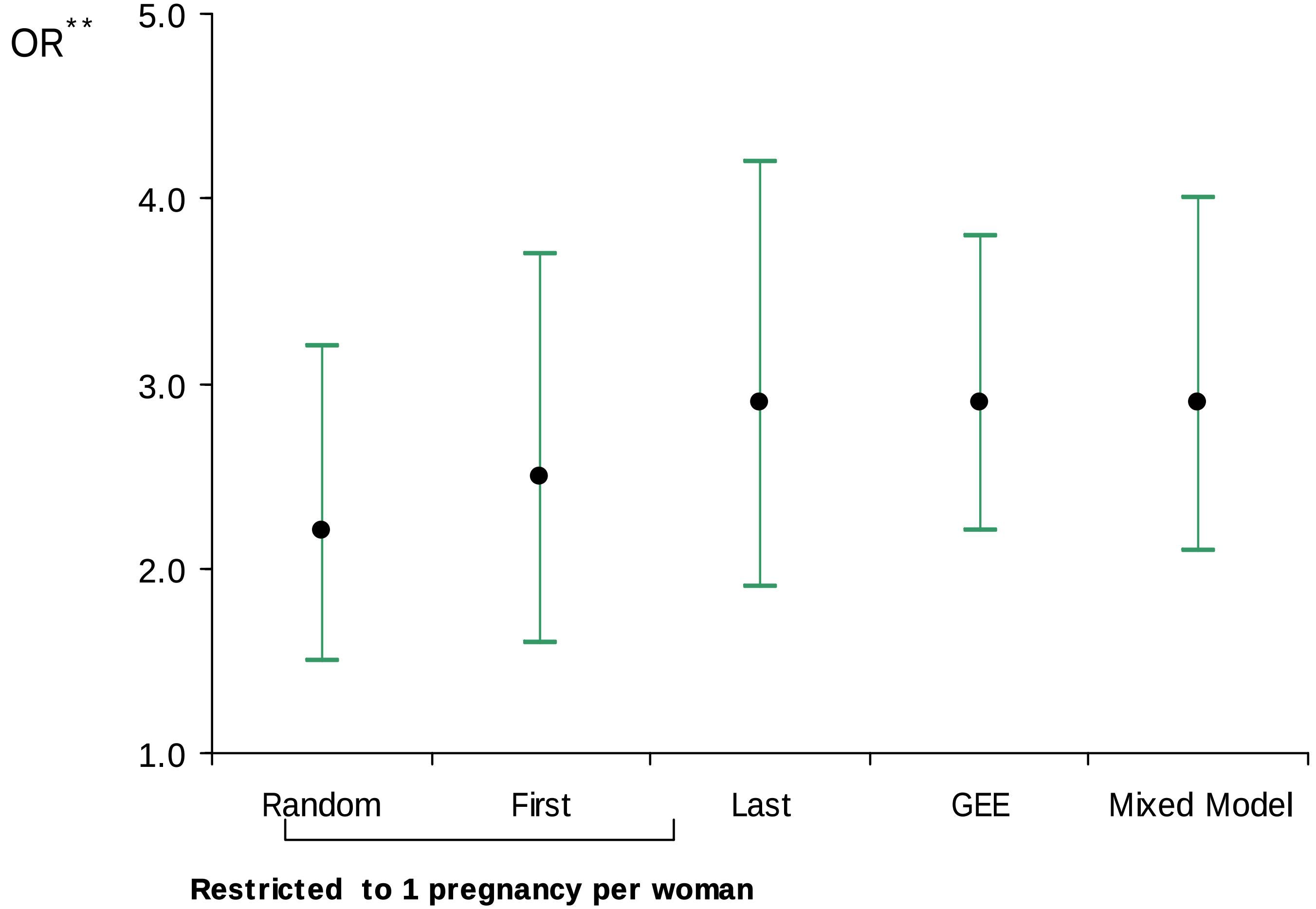
Single pregnancy

1. Random pregnancy
2. First pregnancy
3. Last pregnancy

Multiple pregnancies

4. GEE with working independence assumption
5. Bayesian (Mixed) model with random intercept

Smoking ≥ 1 pack/day and adjusted risk of SGA by Model*



* Adjusted for maternal race, smoking and pre-pregnancy weight; pregnancy interval; clinical site; income; & infant sex using mothers aged 20-24 years as the reference group.

** OR = Odds ratio and 95% confidence interval

Conclusions

- Reproductive outcomes are strongly correlated
- Single outcome models are inefficient and may overlook important effects
- Ignoring correlation can result in incorrect inference and hypothesis tests
- Adjusting for correlation needed
- Including prior outcome as a predictor may wash away some small effects of interest
- GEE analysis with independence working correlation and mixed models are useful approaches to modeling dependent fertility and pregnancy outcomes
- Bayesian models can be very useful