

What are Longitudinal Data?

NIH RESEARCH MATTERS

April 27, 2021

Lack of sleep in middle age may increase dementia risk

At a Glance

- People who slept six hours or less per night in their 50s and 60s were more likely to develop dementia later in life.
- The findings suggest that inadequate sleep duration could increase dementia risk and emphasize the importance of good sleep habits.

What would a study **need** to look like to conclude this?

The Design of a Longitudinal Study

- ▶ Can we conclude this by taking a sample of elderly individuals directly?

The Design of a Longitudinal Study

- ▶ Can we conclude this by taking a sample of elderly individuals directly?
- ▶ **No.** How do we determine how much they slept 20 years prior?

The Design of a Longitudinal Study

- ▶ Can we conclude this by taking a sample of elderly individuals directly?
- ▶ **No.** How do we determine how much they slept 20 years prior?
- ▶ Can we conclude this by taking a sample of middle aged individuals directly?

The Design of a Longitudinal Study

- ▶ Can we conclude this by taking a sample of elderly individuals directly?
- ▶ **No.** How do we determine how much they slept 20 years prior?
- ▶ Can we conclude this by taking a sample of middle aged individuals directly?
- ▶ **No.** How do we determine who will develop dementia later on?

The Design of a Longitudinal Study

- ▶ Can we conclude this by taking a sample of elderly individuals directly?
- ▶ **No.** How do we determine how much they slept 20 years prior?
- ▶ Can we conclude this by taking a sample of middle aged individuals directly?
- ▶ **No.** How do we determine who will develop dementia later on?
- ▶ Can we conclude this by taking independent samples of middle aged individuals *and* elderly individuals?

The Design of a Longitudinal Study

- ▶ Can we conclude this by taking a sample of elderly individuals directly?
- ▶ **No.** How do we determine how much they slept 20 years prior?
- ▶ Can we conclude this by taking a sample of middle aged individuals directly?
- ▶ **No.** How do we determine who will develop dementia later on?
- ▶ Can we conclude this by taking independent samples of middle aged individuals *and* elderly individuals?
- ▶ **No.** How do we pair the individuals?

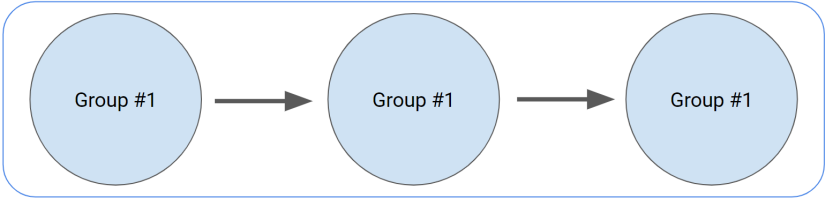
The Design of a Longitudinal Study

We would *need* to be able to follow individuals, starting when they are middle-aged, recording information like how often they sleep, and continue following them until the onset of dementia.

This is a longitudinal study.

Longitudinal Studies

Longitudinal Study



Cross-Sectional Study

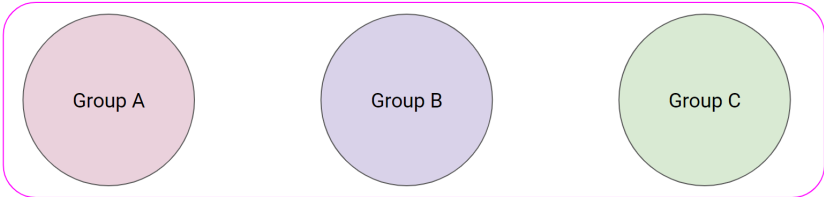


Figure 1

What is a Longitudinal Study?

A research study in which **subjects are followed over time**. Typically this involves **repeated measurements of the same variables**.

Longitudinal studies differ from **cross-sectional** studies and **time series** studies.

Why Do We Care?

Uses for Longitudinal Studies

- ▶ To detect *changes* in outcomes, both at the population and individual level.

Uses for Longitudinal Studies

- ▶ To detect *changes* in outcomes, both at the population and individual level.
- ▶ **Longitudinal effects** as compared to **cohort effects**.

Uses for Longitudinal Studies

- ▶ To detect *changes* in outcomes, both at the population and individual level.
- ▶ **Longitudinal effects** as compared to **cohort effects**.
- ▶ Correctly ascertain the exposures.

Uses for Longitudinal Studies

- ▶ To detect *changes* in outcomes, both at the population and individual level.
- ▶ **Longitudinal effects** as compared to **cohort effects**.
- ▶ Correctly ascertain the exposures.
- ▶ Understand different sources of variation

Uses for Longitudinal Studies

- ▶ To detect *changes* in outcomes, both at the population and individual level.
- ▶ **Longitudinal effects** as compared to **cohort effects**.
- ▶ Correctly ascertain the exposures.
- ▶ Understand different sources of variation
- ▶ **Between-** and **within-**subject variation.

Uses for Longitudinal Studies

- ▶ To detect *changes* in outcomes, both at the population and individual level.
- ▶ **Longitudinal effects** as compared to **cohort effects**.
- ▶ Correctly ascertain the exposures.
- ▶ Understand different sources of variation
- ▶ **Between-** and **within-**subject variation.
- ▶ To detect **time effects**, both directly and as interactions with other relevant factors.

Uses for Longitudinal Studies

Bottom line: There are many questions of interest which can only be answered using longitudinal data.

We should probably learn how to analyze it.

Why are Longitudinal Data Special?

Why are Longitudinal Data Special?

What makes longitudinal data more difficult to analyze?

Why are Longitudinal Data Special?

The data are **correlated**.

Everyone's favourite assumption (assume that X_1, \dots, X_n are iid...) will **not** hold.

Now what?

Welcome to STAT 437

Example Datasets

TLC Trial

ID	Treatment	W0	W1	W4	W6
1	P	30.8	26.9	25.8	23.8
2	A	26.5	14.8	19.5	21
3	A	25.8	23	19.1	23.2
⋮	⋮	⋮	⋮	⋮	⋮
98	A	29.4	22.1	25.3	4.1
99	A	21.9	7.6	10.8	13
100	A	20.7	8.1	25.7	12.3

- ▶ Is there a difference between **placebo** and **treatment**?
- ▶ How does the blood lead level **change over time** (in each group)?
- ▶ Is the **change** over time **equal** between treatment groups?

Sales Data

DATE	brand	prod	QTY	PROMO
2014-01-02	1	1	7	0
2014-01-02	1	2	3	0
2014-01-02	1	3	0	0
⋮	⋮	⋮	⋮	⋮
2018-12-31	4	8	1	1
2018-12-31	4	9	0	0
2018-12-31	4	10	3	1

- ▶ Are the **different brands comparable** in terms of overall sales?
- ▶ Are the **different products comparable**?
- ▶ Do **promotions increase** the quantity sold? If so, **by how much**?
- ▶ Do the effects of time, and promotion, **change by brand** or product?

Podcast Data

Rating	No. Reviews	Title	Date	...
4.9	6400	Dissect	2019-11-01	...
4.9	26300	The Adventure Zone	2019-11-01	...
4.8	3700	Song Exploder	2019-11-01	...
⋮	⋮	⋮	⋮	⋮
4.2	1100	Finding Fred	2019-12-01	...
3.9	648	Inside Frozen 2	2019-12-01	...
4.6	6400	Pop Culture Happy Hour	2019-12-01	...

- ▶ Can we **predict** the number of ratings that a podcast will receive over time?
- ▶ Can we **predict** the average rating value that a podcast will receive over time?

Stroke Data

year	Prop. (0,0)	Prop. (0,1)	Prop. (1,0)	Prop. (1,1)
1	57/344	17/72	17/79	5/23
2	27/287	8/55	9/62	4/18
3	23/260	8/47	5/53	3/14
⋮	⋮	⋮	⋮	⋮
8	10/129	1/15	5/23	1/4
9	17/119	3/14	4/18	0/3
10	13/102	1/11	2/14	0/3

- ▶ This is **time to event** data
- ▶ What is **probability of surviving** beyond some point?
- ▶ Does this **differ** if you previously had a stroke? If you **received treatment**?

Summary

Summary

- ▶ Longitudinal data occur when we take repeated measurements on the same individuals over time.
- ▶ Longitudinal data are required for answering questions about changes within an individual (compared to between individuals) and to capture time effects.
- ▶ Longitudinal data are challenging to work with because the data are correlated.