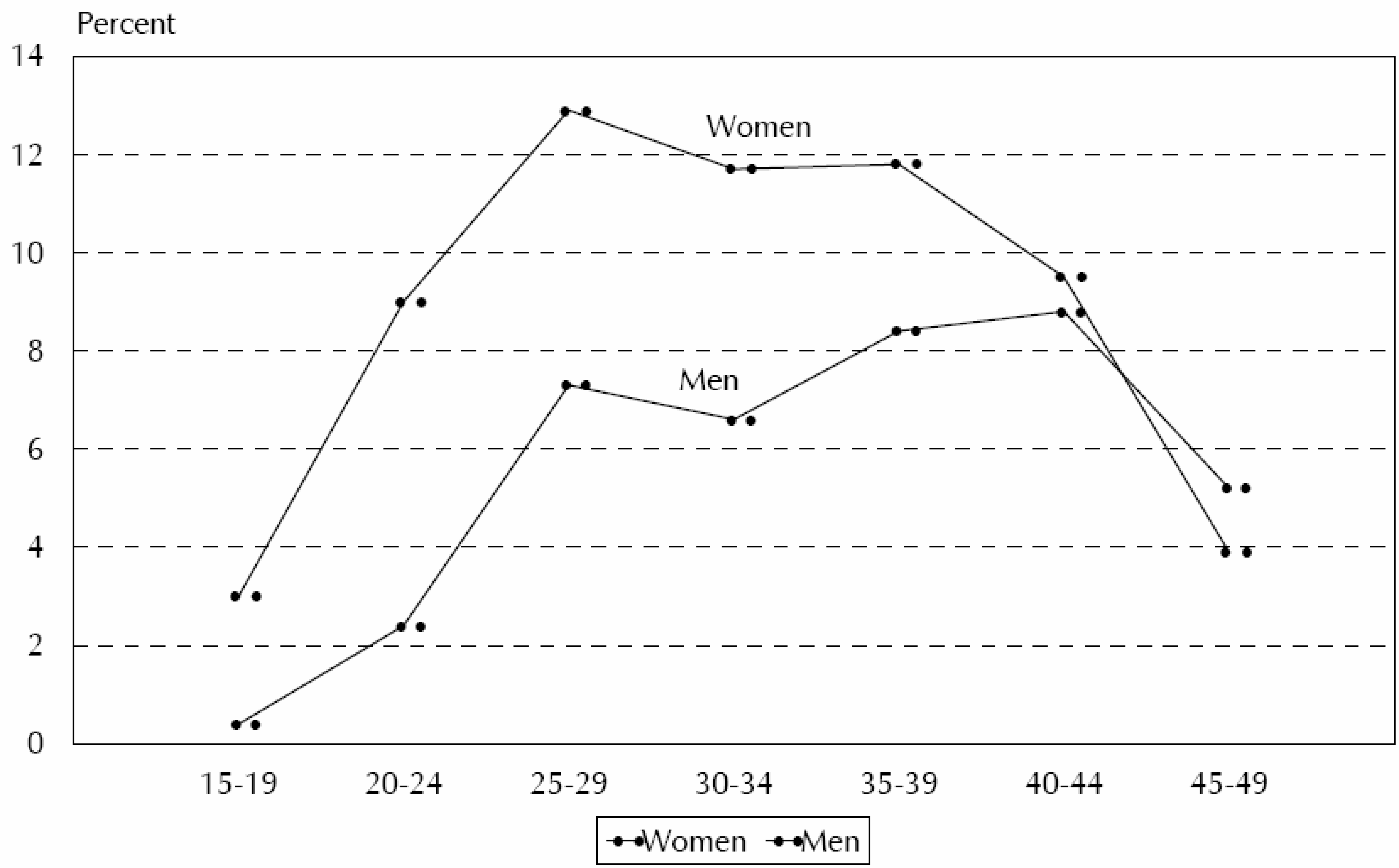
Economics 172
Issues in African Economic Development

Professor Ted Miguel
Department of Economics
University of California, Berkeley

Economics 172
Issues in African Economic Development

Lecture 10 – February 15, 2007

Figure 13.1 HIV Prevalence by Age Group and Sex



KDHS 2003

Counting HIV+ people in Kenya

- Based on antenatal clinic survey data, the official UNAIDS estimate of HIV+ adults in Kenya by late 2001 was 15.0%

- The 2003 Kenya Demographic and Health Survey (DHS) tried to survey a representative subsample of population. 73.4% agreed to be tested
  - This data indicates that “only” 6.7% of Kenyan 15-49 year olds tested are HIV+!

- Which of the two numbers is closer to the truth?

Table 13.4 HIV prevalence by selected socioeconomic characteristics

Percentage HIV positive among women and men age 15-49 who were tested, by socioeconomic characteristics, Kenya 2003

	Women		Men		Total	
Socioeconomic characteristic	Percent HIV positive	Number	Percent HIV positive	Number	Percent HIV positive	Number
Residence						
Urban	12.3	779	7.5	716	10.0	1,495
Rural	7.5	2,372	3.6	2,135	5.6	4,507
Education						
No education	4.4	396	2.7	156	3.9	552
Primary incomplete	9.3	1,052	3.4	982	6.4	2,034
Primary complete	10.6	784	5.9	660	8.5	1,444
Secondary+	8.2	918	5.2	1,053	6.6	1,972
Employment						
Currently working	9.6	1,844	5.9	2,007	7.6	3,851
Not currently working	7.4	1,307	1.5	844	5.1	2,151
Wealth quintile						
Lowest	3.9	505	3.4	431	3.6	937
Second	8.5	580	4.2	501	6.5	1,082
Middle	7.1	597	2.2	528	4.8	1,125
Fourth	9.7	663	4.3	624	7.1	1,287
Highest	12.2	806	7.3	765	9.8	1,571

Table 13.4 HIV prevalence by selected socioeconomic characteristics

Percentage HIV positive among women and men age 15-49 who were tested, by socioeconomic characteristics, Kenya 2003

Socioeconomic characteristic	Women		Men		Total	
	Percent HIV positive	Number	Percent HIV positive	Number	Percent HIV positive	Number
Residence						
Urban	12.3	779	7.5	716	10.0	1,495
Rural	7.5	2,372	3.6	2,135	5.6	4,507
Education						
No education	4.4	396	2.7	156	3.9	552
Primary incomplete	9.3	1,052	3.4	982	6.4	2,034
Primary complete	10.6	784	5.9	660	8.5	1,444
Secondary+	8.2	918	5.2	1,053	6.6	1,972
Employment						
Currently working	9.6	1,844	5.9	2,007	7.6	3,851
Not currently working	7.4	1,307	1.5	844	5.1	2,151
Wealth quintile						
Lowest	3.9	505	3.4	431	3.6	937
Second	8.5	580	4.2	501	6.5	1,082
Middle	7.1	597	2.2	528	4.8	1,125
Fourth	9.7	663	4.3	624	7.1	1,287
Highest	12.2	806	7.3	765	9.8	1,571

Table 13.4 HIV prevalence by selected socioeconomic characteristics

Percentage HIV positive among women and men age 15-49 who were tested, by socioeconomic characteristics, Kenya 2003

Socioeconomic characteristic	Women		Men		Total	
	Percent HIV positive	Number	Percent HIV positive	Number	Percent HIV positive	Number
Residence						
Urban	12.3	779	7.5	716	10.0	1,495
Rural	7.5	2,372	3.6	2,135	5.6	4,507
Education						
No education	4.4	396	2.7	156	3.9	552
Primary incomplete	9.3	1,052	3.4	982	6.4	2,034
Primary complete	10.6	784	5.9	660	8.5	1,444
Secondary+	8.2	918	5.2	1,053	6.6	1,972
Employment						
Currently working	9.6	1,844	5.9	2,007	7.6	3,851
Not currently working	7.4	1,307	1.5	844	5.1	2,151
Wealth quintile						
Lowest	3.9	505	3.4	431	3.6	937
Second	8.5	580	4.2	501	6.5	1,082
Middle	7.1	597	2.2	528	4.8	1,125
Fourth	9.7	663	4.3	624	7.1	1,287
Highest	12.2	806	7.3	765	9.8	1,571

Table 13.4 HIV prevalence by selected socioeconomic characteristics

Percentage HIV positive among women and men age 15-49 who were tested, by socioeconomic characteristics, Kenya 2003

Socioeconomic characteristic	Women		Men		Total	
	Percent HIV positive	Number	Percent HIV positive	Number	Percent HIV positive	Number
Residence						
Urban	12.3	779	7.5	716	10.0	1,495
Rural	7.5	2,372	3.6	2,135	5.6	4,507
Education						
No education	4.4	396	2.7	156	3.9	552
Primary incomplete	9.3	1,052	3.4	982	6.4	2,034
Primary complete	10.6	784	5.9	660	8.5	1,444
Secondary+	8.2	918	5.2	1,053	6.6	1,972
Employment						
Currently working	9.6	1,844	5.9	2,007	7.6	3,851
Not currently working	7.4	1,307	1.5	844	5.1	2,151
Wealth quintile						
Lowest	3.9	505	3.4	431	3.6	937
Second	8.5	580	4.2	501	6.5	1,082
Middle	7.1	597	2.2	528	4.8	1,125
Fourth	9.7	663	4.3	624	7.1	1,287
Highest	12.2	806	7.3	765	9.8	1,571

Key questions in the study of HIV/AIDS

(0) Measuring the extent of the problem (today)

Key questions in the study of HIV/AIDS

(0) Measuring the extent of the problem (today)

(1) What impact does HIV/AIDS have on economic development in Africa?

Key questions in the study of HIV/AIDS

(0) Measuring the extent of the problem (today)

(1) What impact does HIV/AIDS have on economic development in Africa?

-- Labor productivity / labor turnover

Key questions in the study of HIV/AIDS

(0) Measuring the extent of the problem (today)

(1) What impact does HIV/AIDS have on economic development in Africa?

-- Labor productivity / labor turnover

-- Human capital accumulation (orphans)

-- Investment and savings (as time horizons change)

Key questions in the study of HIV/AIDS

(0) Measuring the extent of the problem (today)

(1) What impact does HIV/AIDS have on economic development in Africa?

(2) Why does HIV/AIDS continue to spread in Africa?

Key questions in the study of HIV/AIDS

(0) Measuring the extent of the problem (today)

(1) What impact does HIV/AIDS have on economic development in Africa?

(2) Why does HIV/AIDS continue to spread in Africa?

(3) What can / should public policy do about HIV/AIDS?

Econometric method: difference-in-differences

- How do we evaluate the impact of a program or “event” if we do not have a randomized experimental design?

Econometric method: difference-in-differences

- How do we evaluate the impact of a program or “event” if we do not have a randomized experimental design?

- A common and powerful approach utilizes data on the same individual (or household, firm, village, etc.) over time. This is called “panel data” (or “longitudinal data”)

Econometric method: difference-in-differences

- How do we evaluate the impact of a program or “event” if we do not have a randomized experimental design?

- A common and powerful approach utilizes data on the same individual (or household, firm, village, etc.) over time. This is called “panel data” (or “longitudinal data”)

- What assumptions allow us to estimate causal impacts of a program/event using panel data?

Treatment effects and omitted variable bias

(1) $Y_i = a + bT_i + cX_i + e_i$

(2) $E(Y_i \mid T_i=1) - E(Y_i \mid T_i=0)$

$$= [a + b + cE(X_i \mid T_i=1) + E(e_i \mid T_i=1)]$$
$$- [a + 0 + cE(X_i \mid T_i=0) + E(e_i \mid T_i=0)]$$

$$= b + c\,[E(X_i \mid T_i=1) - E(X_i \mid T_i=0)]$$

True effect "Omitted variable/selection bias" term

Extend the data to two time periods (pre, post)

(1) $Y_{it} = a + bT_{it} + cX_{it} + e_{it}$

$t = 0$: Pre-program/pre-event period

$t = 1$: Post-program/post-event period

Extend the data to two time periods (pre, post)

(1) $$Y_{it} = a + bT_{it} + cX_{it} + e_{it}$$

$t = 0$: Pre-program/pre-event period

$t = 1$: Post-program/post-event period

For concreteness, let the "event" here be an individual falling ill with AIDS, and the outcome be their productivity at work (e.g., how many kilos of tea they pick per day)

Extend the data to two time periods (pre, post)

$$(1) \qquad Y_{it} = a + bT_{it} + cX_{it} + e_{it}$$

$t = 0$: Pre-program/pre-event period

$t = 1$: Post-program/post-event period

In period $t = 1$, the treatment effect estimate is (as before):

$$(2) \qquad E(Y_{i1} \mid T_{i1}=1) - E(Y_{i1} \mid T_{i1}=0)$$

$$= [a + b + cE(X_{i1} \mid T_{i1}=1) + E(e_{i1} \mid T_{i1}=1)]$$

$$- \ [a + 0 + cE(X_{i1} \mid T_{i1}=0) + E(e_{i1} \mid T_{i1}=0)]$$

$$= b + c\,[E(X_{i1} \mid T_{i1}=1) - E(X_{i1} \mid T_{i1}=0)]$$

True effect *Omitted variable bias term for $t=1$*

Extend the data to two time periods (pre, post)

In period $t = 0$, before the event, the difference between the two groups is the following, where we know who will later become ill in period $t = 1$ ($T_{i1}=1$):

(3) $E(Y_{i0} \mid T_{i1}=1) - E(Y_{i0} \mid T_{i1}=0)$

Extend the data to two time periods (pre, post)

In period $t = 0$, before the event, the difference between the two groups is the following, where we know who will later become ill in period $t = 1$ $(T_{i1}=1)$:

$$(3) \qquad E(Y_{i0} \mid T_{i1}=1) - E(Y_{i0} \mid T_{i1}=0)$$

$$= [a + 0 + cE(X_{i0} \mid T_{i1}=1) + E(e_{i0} \mid T_{i1}=1)]$$

$$- \ [a + 0 + cE(X_{i0} \mid T_{i1}=0) + E(e_{i0} \mid T_{i1}=0)]$$

$$= c\,[E(X_{i0} \mid T_{i1}=1) - E(X_{i0} \mid T_{i1}=0)]$$

Omitted variable bias term for t=0

Extend the data to two time periods (pre, post)

In period $t = 0$, before the event, the difference between the two groups is the following, where we know who will later become ill in period $t = 1$ $(T_{i1}=1)$:

$$(3) \qquad E(Y_{i0} \mid T_{i1}=1) - E(Y_{i0} \mid T_{i1}=0)$$

$$= [a + \boxed{0} + cE(X_{i0} \mid T_{i1}=1) + E(e_{i0} \mid T_{i1}=1)]$$

$$- \; [a + 0 + cE(X_{i0} \mid T_{i1}=0) + E(e_{i0} \mid T_{i1}=0)]$$

$$= c\,[E(X_{i0} \mid T_{i1}=1) - E(X_{i0} \mid T_{i1}=0)]$$

Omitted variable bias term for t=0

Taking the difference-in-differences

- The difference-in-differences (DD) estimator takes the differences between equations (2) and (3) to eliminate omitted variable bias and deliver the true effect:

  Equation (2) – Equation (3)

$$= \{E(Y_{i1} \mid T_{i1}=1) - E(Y_{i1} \mid T_{i1}=0)\}$$
$$- \{E(Y_{i0} \mid T_{i1}=1) - E(Y_{i0} \mid T_{i1}=0)\}$$

Taking the difference-in-differences

- The difference-in-differences (DD) estimator takes the differences between equations (2) and (3) to eliminate omitted variable bias and deliver the true effect:

Equation (2) – Equation (3)

$$= \{E(Y_{i1} \mid T_{i1}=1) - E(Y_{i1} \mid T_{i1}=0)\}$$
$$- \{E(Y_{i0} \mid T_{i1}=1) - E(Y_{i0} \mid T_{i1}=0)\}$$

$$= b + c\,[E(X_{i1} \mid T_{i1}=1) - E(X_{i1} \mid T_{i1}=0)]$$
$$- c\,[E(X_{i0} \mid T_{i1}=1) - E(X_{i0} \mid T_{i1}=0)]$$

Taking the difference-in-differences

- The difference-in-differences (DD) estimator takes the differences between equations (2) and (3) to eliminate omitted variable bias and deliver the true effect:

Equation (2) – Equation (3)

$$= \{E(Y_{i1} \mid T_{i1}=1) - E(Y_{i1} \mid T_{i1}=0)\}$$
$$- \{E(Y_{i0} \mid T_{i1}=1) - E(Y_{i0} \mid T_{i1}=0)\}$$

$$= b + c\,[E(X_{i1} \mid T_{i1}=1) - E(X_{i1} \mid T_{i1}=0)]$$
$$- c\,[E(X_{i0} \mid T_{i1}=1) - E(X_{i0} \mid T_{i1}=0)]$$

= *True effect*

+ {(*OVB term for t=1*) – (*OVB term for t=0*)}

Taking the difference-in-differences

- The difference-in-differences (DD) estimator takes the differences between equations (2) and (3) to eliminate omitted variable bias and deliver the true effect:

Equation (2) – Equation (3)

$$= \{E(Y_{i1} \mid T_{i1}=1) - E(Y_{i1} \mid T_{i1}=0)\}$$
$$- \{E(Y_{i0} \mid T_{i1}=1) - E(Y_{i0} \mid T_{i1}=0)\}$$

$$= b + c\,[E(X_{i1} \mid T_{i1}=1) - E(X_{i1} \mid T_{i1}=0)]$$
$$- c\,[E(X_{i0} \mid T_{i1}=1) - E(X_{i0} \mid T_{i1}=0)]$$

= *True effect* =0?

+ {(*OVB term for t=1*) – (*OVB term for t=0*)}

Dealing with omitted variable bias

- When is omitted variable bias not a problem?

  1) Collect information on X

  2) The omitted variable does not affect the outcome

  3) The omitted variables are not correlated with the explanatory variable of interest (here, T) – <u>randomization</u>

Dealing with omitted variable bias

- When is omitted variable bias not a problem?

  1) Collect information on X

  2) The omitted variable does not affect the outcome

  3) The omitted variables are not correlated with the explanatory variable of interest (here, T) – randomization

  4) If we have panel data and the extent of omitted variable bias is constant over time, then a difference-in-differences (DD) estimate delivers the true effect

When is DD is a sound approach?

4) If we have panel data and the extent of omitted variable bias is constant over time, then a difference-in-differences (DD) estimate delivers the true effect

- When is this likely to hold?

When is DD is a sound approach?

4) If we have panel data and the extent of omitted variable bias is constant over time, then a <u>difference-in-differences (DD)</u> estimate delivers the true effect

- When is this likely to hold?

  Basically, when unobserved factors that could affect the outcome or individuals are “stable” over the study period

When is DD is a sound approach?

4) If we have panel data and the extent of omitted variable bias is constant over time, then a <u>difference-in-differences (DD)</u> estimate delivers the true effect

- When is this likely to hold?

  Basically, when unobserved factors that could affect the outcome or individuals are “stable” over the study period

- For concreteness, let the “event” here be an individual falling ill with AIDS, and the outcome be their productivity at work (e.g., how many kilos of tea they pick per day)

  -- If labor conditions, rules, effort, etc. are stable over the study period, then DD seems a reasonable approach

Key questions in the study of HIV/AIDS

(0) Measuring the extent of the problem (today)

(1) What impact does HIV/AIDS have on economic development in Africa?

- -- Labor productivity / labor turnover
- -- Human capital accumulation (orphans)
- -- Investment and savings (as time horizons change)

HIV/AIDS and labor productivity in Kenya

- Fox et al (2004) study the impact of HIV/AIDS illness on labor productivity on tea plantations in Kenya, 1997-2002

HIV/AIDS and labor productivity in Kenya

- Fox et al (2004) study the impact of HIV/AIDS illness on labor productivity on tea plantations in Kenya, 1997-2002
- They compare the labor productivity – the kilograms of tea leaves picked per day – of workers who became sick with HIV/AIDS during the study period to workers who remained healthy
  - The total sample is 271 workers. 54 died or retired due to HIV/AIDS

HIV/AIDS and labor productivity in Kenya

- Fox et al (2004) study the impact of HIV/AIDS illness on labor productivity on tea plantations in Kenya, 1997-2002
- They compare the labor productivity – the kilograms of tea leaves picked per day – of workers who became sick with HIV/AIDS during the study period to workers who remained healthy
  - The total sample is 271 workers. 54 died or retired due to HIV/AIDS. In the statistical analysis, these are the "treated" group ($T_{i1}=1$).

Table I Study population

Parameter	Cases		Comparison pluckers		
	Value	SD	Value	SD	P-value
n	54		217		
Age (mean)*	35.74	7.26	37.33	8.17	0.21
Years of service (mean)*	5.15	3.37	6.20	2.42	0.06
Sex (proportion male)	61%		71%		0.16

* Dates computed as date on last day of observation.

Table I Study population

	Cases $T_{i1}=1$ AIDS		Comparison pluckers $T_{i1}=0$ No AIDS		
Parameter	Value	SD	Value	SD	P-value
n	54		217		
Age (mean)*	35.74	7.26	37.33	8.17	0.21
Years of service (mean)*	5.15	3.37	6.20	2.42	0.06
Sex (proportion male)	61%		71%		0.16

* Dates computed as date on last day of observation.

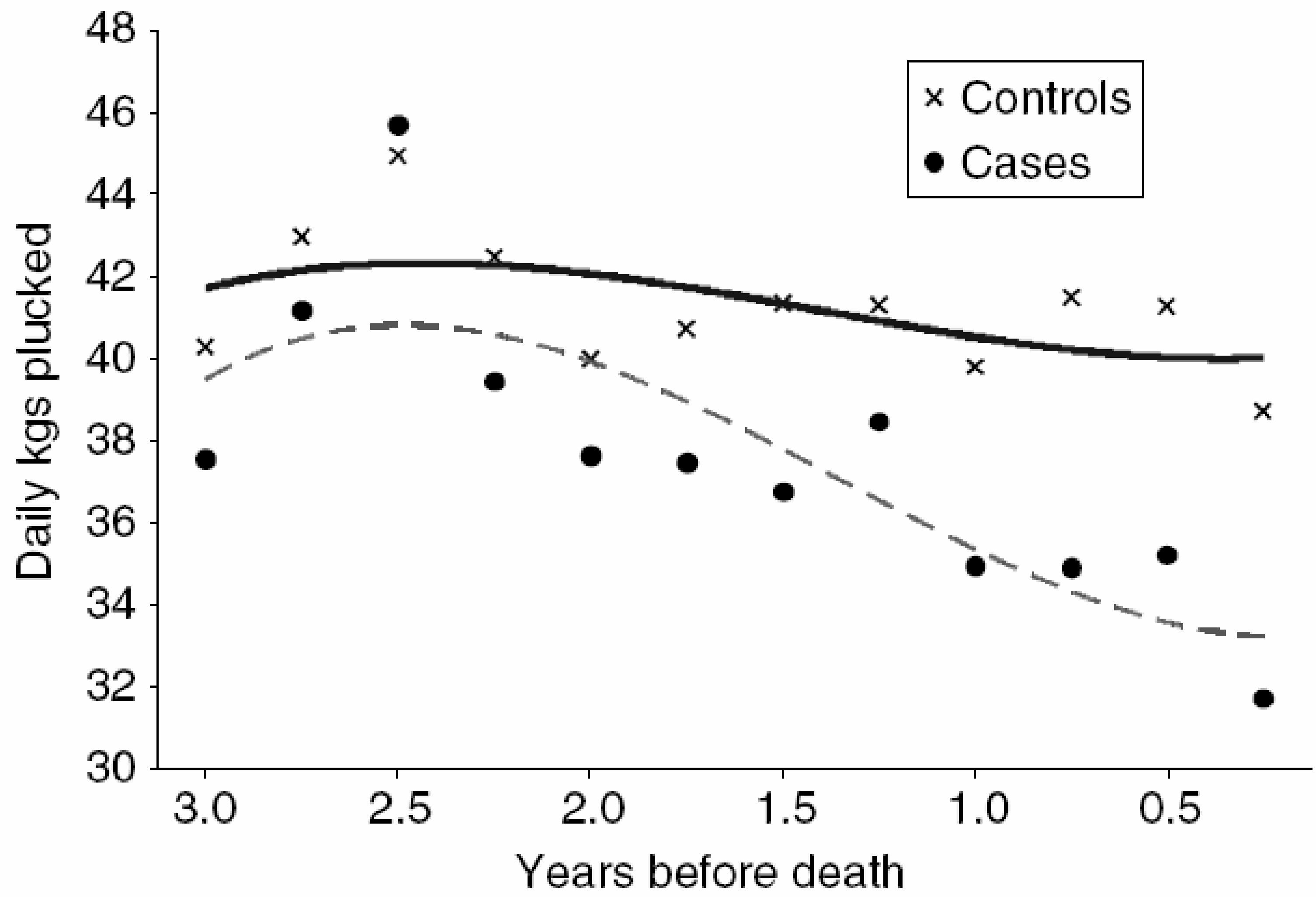


Figure 1 Mean kilograms of tea plucked per day on days of plucking for cases and controls (univariate analysis – curves are trend lines fit using polynomial regression for each group. Note that vertical access scale begins at 30 kg/day).

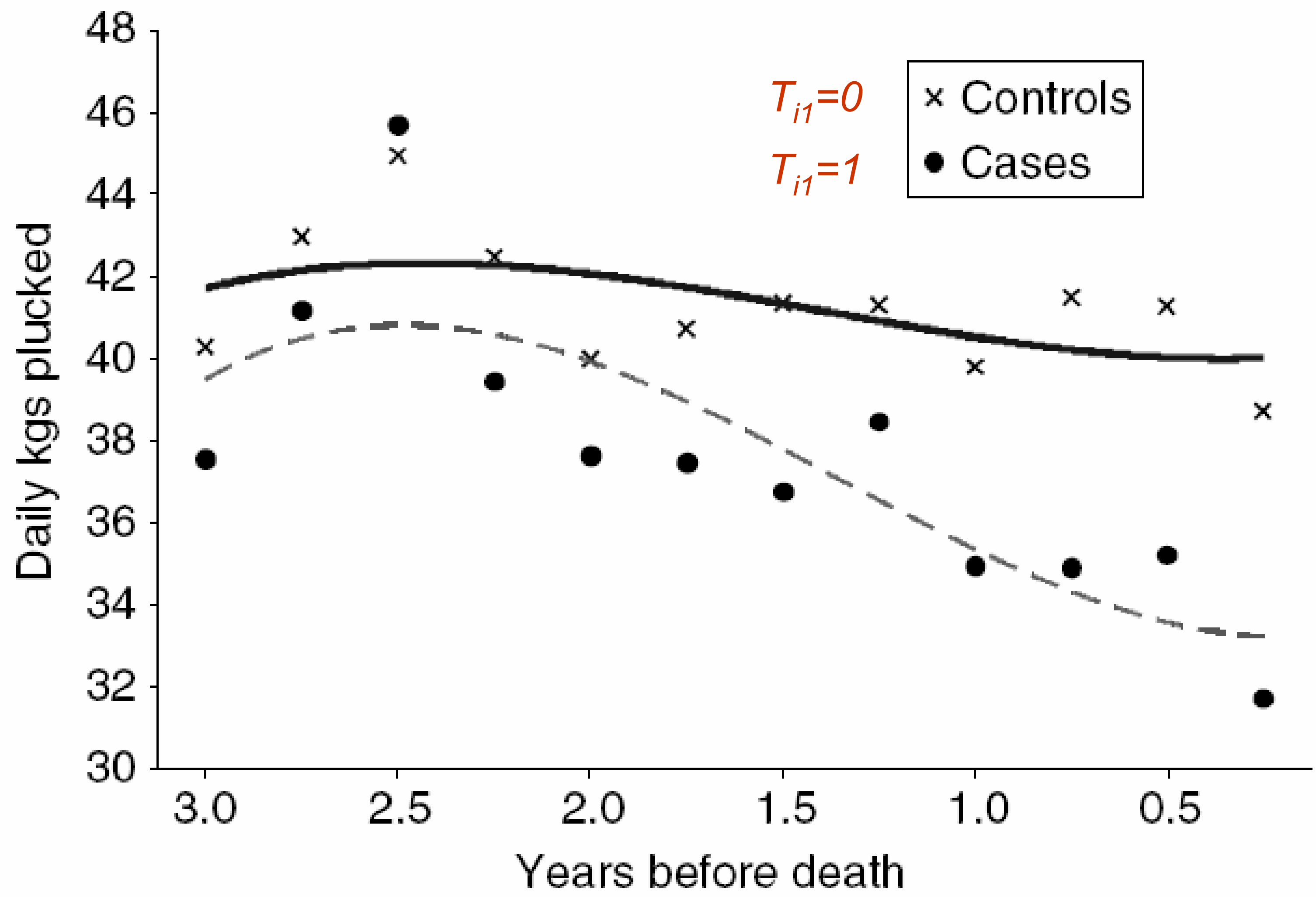


Figure 1 Mean kilograms of tea plucked per day on days of plucking for cases and controls (univariate analysis – curves are trend lines fit using polynomial regression for each group. Note that vertical access scale begins at 30 kg/day).

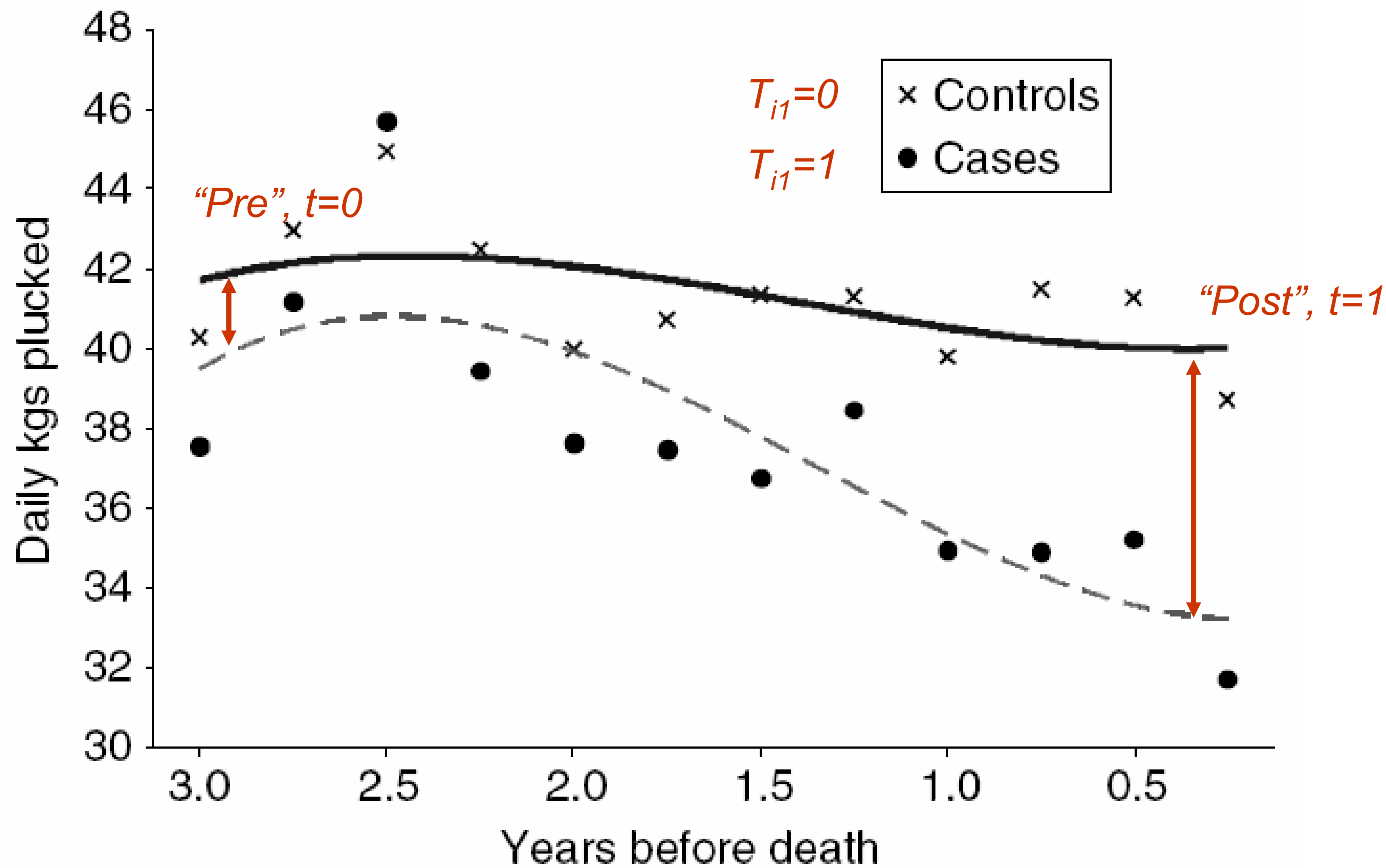


Figure 1 Mean kilograms of tea plucked per day on days of plucking for cases and controls (univariate analysis – curves are trend lines fit using polynomial regression for each group. Note that vertical access scale begins at 30 kg/day).

Table 2 Adjusted differences between cases and comparison pluckers on days plucking at 6-month intervals prior to AIDS-related termination*

Years before termination	Difference†	Percentage difference‡	SE	*P*-value
3.0 years	−1.689	−4%	2.732	0.536
2.5 years	0.466	1%	2.224	0.834
2.0 years	2.400	6%	1.956	0.220
1.5 years	4.113	10%	1.871	0.028
1.0 years	5.605	13%	1.940	0.004
0.5 years	6.876	16%	2.191	0.002
Near termination	7.927	19%	2.684	0.003

* The final regression model included age, a dummy variable for matched group, the variables for time and a dummy variable to indicate pluckers who went on to an AIDS-related termination.
† Difference in kilograms.
‡ Expressed as a per cent of the average kilograms plucked by comparison pluckers, 41.

Table 2 Adjusted differences between cases and comparison pluckers on days plucking at 6-month intervals prior to AIDS-related termination*

	Years before termination	Difference†	Percentage difference‡	SE	*P*-value
"Pre" $t=0$	3.0 years	−1.689	−4%	2.732	0.536
	2.5 years	0.466	1%	2.224	0.834
	2.0 years	2.400	6%	1.956	0.220
	1.5 years	4.113	10%	1.871	0.028
	1.0 years	5.605	13%	1.940	0.004
"Post" $t=1$	0.5 years	6.876	16%	2.191	0.002
	Near termination	7.927	19%	2.684	0.003

* The final regression model included age, a dummy variable for matched group, the variables for time and a dummy variable to indicate pluckers who went on to an AIDS-related termination.

† Difference in kilograms.

‡ Expressed as a per cent of the average kilograms plucked by comparison pluckers, 41.

Table 2 Adjusted differences between cases and comparison pluckers on days plucking at 6-month intervals prior to AIDS-related termination*

	Years before termination	Difference†	Percentage difference‡	SE	*P*-value	
"Pre" $t=0$	3.0 years	−1.689	−4%	2.732	0.536	$Diff_0$ $= +1\%$
	2.5 years	0.466	1%	2.224	0.834	
	2.0 years	2.400	6%	1.956	0.220	
	1.5 years	4.113	10%	1.871	0.028	
	1.0 years	5.605	13%	1.940	0.004	
"Post" $t=1$	0.5 years	6.876	16%	2.191	0.002	$Diff_1$ $= -17\%$
	Near termination	7.927	19%	2.684	0.003	

* The final regression model included age, a dummy variable for matched group, the variables for time and a dummy variable to indicate pluckers who went on to an AIDS-related termination.
† Difference in kilograms.
‡ Expressed as a per cent of the average kilograms plucked by comparison pluckers, 41.

Taking the difference-in-differences

- The difference-in-differences (DD) estimator takes the differences between equations (2) and (3) to eliminate omitted variable bias and deliver the true effect:

  Equation (2) – Equation (3)

$$= \{E(Y_{i1} \mid T_{i1}=1) - E(Y_{i1} \mid T_{i1}=0)\}$$
$$- \{E(Y_{i0} \mid T_{i1}=1) - E(Y_{i0} \mid T_{i1}=0)\}$$

Taking the difference-in-differences

- The difference-in-differences (DD) estimator takes the differences between equations (2) and (3) to eliminate omitted variable bias and deliver the true effect:

  Equation (2) – Equation (3)

$$= \{E(Y_{i1} \mid T_{i1}=1) - E(Y_{i1} \mid T_{i1}=0)\}$$
$$- \{E(Y_{i0} \mid T_{i1}=1) - E(Y_{i0} \mid T_{i1}=0)\}$$

$$= Diff_1 - Diff_0 = (-17\%) - (+1\%) = -18\%$$

Interpreting Fox et al.'s (2004) results

- Including absenteeism, final year income drops 35%

Interpreting Fox et al.'s (2004) results

- Including absenteeism, final year income drops 35%

- Some controls are also likely ill with HIV/AIDS, leading the estimated T–C difference to be a lower bound. The decline over time in C productivity may be AIDS related

- Sick workers have family member "helpers". So estimates are again likely to be lower bounds

Interpreting Fox et al.'s (2004) results

- Including absenteeism, final year income drops 35%
- Some controls are also likely ill with HIV/AIDS, leading the estimated T–C difference to be a lower bound. The decline over time in C productivity may be AIDS related
- Sick workers have family member "helpers". So estimates are again likely to be lower bounds
- Is the assumption of no time-varying omitted variables reasonable? AIDS victims have higher absenteeism three years prior. What is the right "counterfactual"?

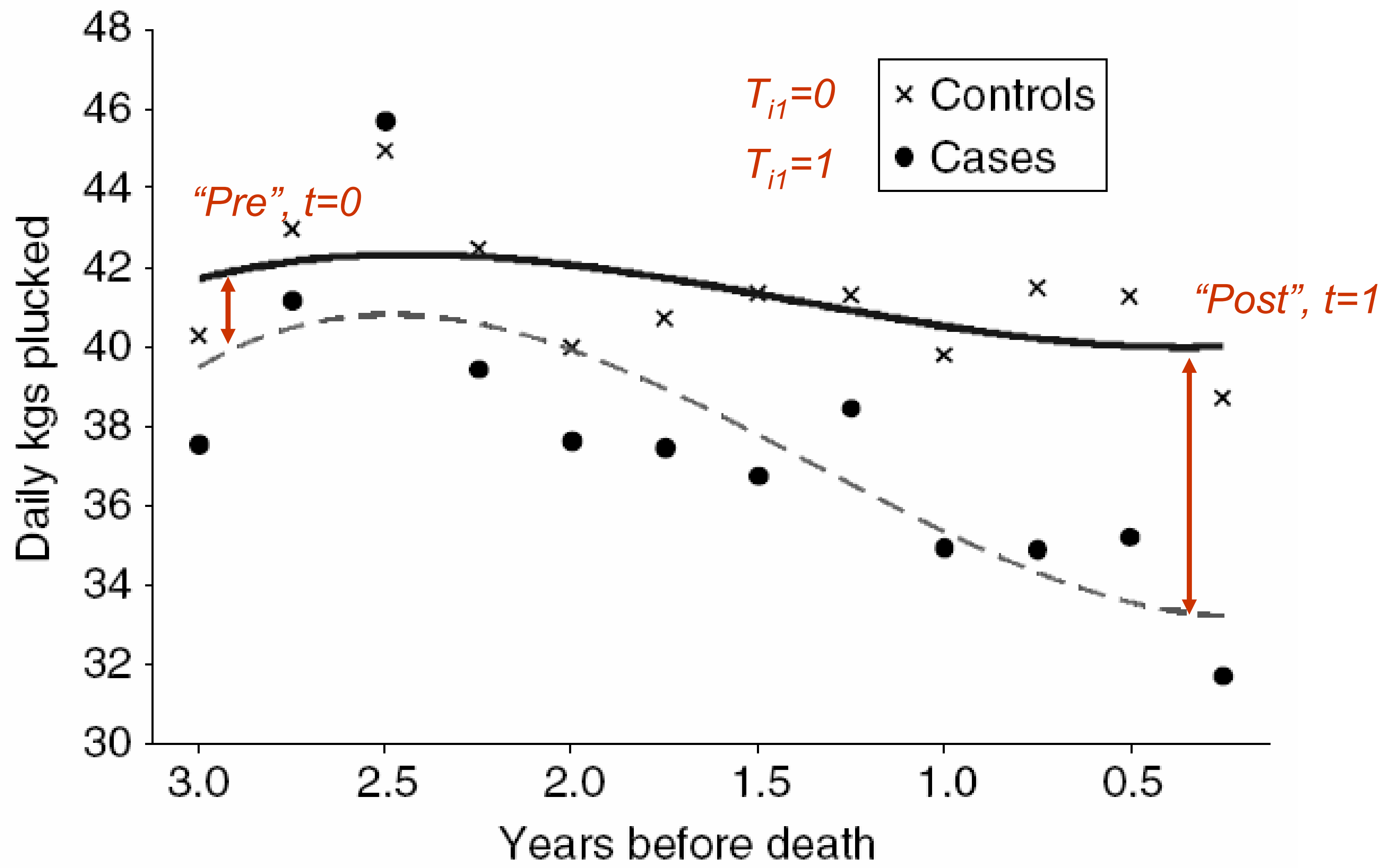


Figure 1 Mean kilograms of tea plucked per day on days of plucking for cases and controls (univariate analysis – curves are trend lines fit using polynomial regression for each group. Note that vertical access scale begins at 30 kg/day).

- For next time: continue the HIV/AIDS readings

Whiteboard #1

Whiteboard #2

Whiteboard #3

Whiteboard #4

Whiteboard #5

Map of Africa

Map of Africa