

**Looking within the household:**  
Impacts for individual outcomes using  
aggregate household data

Vincenzo Di Maro

# Topic

How to infer outcomes of a specific member/group (for ex. caloric intake of children) when only aggregate household data is available

## Plan is to:

- **motivate** the study
- **describe the methodology** [Chesher, 1997] and the data used [Progresa, wave nov 1999]
- show how this approach was applied in a **policy evaluation framework**
- discuss what can be learned from the **results** with a special focus on **heterogeneity of the program impact** and **asymmetric information** held by **respondents**
- **next steps** of this research

# Motivation

Development of **methods for inferring individual outcomes** using household data is **important** because:

- **household budget surveys less expensive and less labour intensive**
- **individual surveys might be too intrusive** (eating habits)
- information collected in **household budget surveys is more comparable between countries**

**In the policy evaluation context:**

Most interventions **target a specific group within the household**; for example, children and/or women. But **data** might be collected at a level of aggregation that **does not allow to study directly the impact** of the program on many relevant dimensions. This is the case of **Progresa**.

# Methodology: how household data can be individualized

Engle *et al.* (1986), “relation between weather and electricity sales”

Chesher (1997), “sex, age-nutrient intake relationships”

We want to estimate average individual caloric intake, then we start from a **model of household nutrient consumption:**

The **average nutrient consumption of each member  $p$**  within the household is modeled as function of individual  $x_p$  and household  $z$  characteristics:

$$E\left[c_p \mid x, z\right] = f(x_p, z)$$

**Average total household nutrient consumption  $c$**  will be then the sum of these individual nutrient consumptions:

$$E\left[c \mid x, z\right] = \sum_{p=1}^P f(x_p, z)$$

We want to estimate functions  $f(x_p, z)$ . We will do it exploiting the **moment condition**

$$E \left[ \left\{ c - \sum_{p=1}^P f(x_p, z) \right\} g(x, z) \mid x, z \right] = 0$$

When **only household totals** are recorded only (indirect) **methods such as this** are available.

**Assumption on functions**  $f(x_p, z)$ :

multiplicatively separable functions of individual and household characteristics

$$f(x_p, z) = f(x_p)u(z)$$

# Estimation

We allow for **different functions for male and females**

$$f(x_p) = S_p f_M(\text{age}_p) + (1 - S_p) f_F(\text{age}_p)$$

**A simplification: problem is discretized**

$f_M(\text{age}_p)$  and  $f_F(\text{age}_p)$  are approximated by step functions with points of increase at integer years of age

Then we can write the **model for expected household consumption** as:

$$E[c | x, z] = \left( \beta_0 + n'_M \beta^M + n'_F \beta^F \right) \exp(z' \gamma)$$

where  $n_S$  are counts of household members at sex S at each integer year of age

In this form we are interested in the **estimation of  $\beta^M$  and  $\beta^F$** , that are the average intakes at each age for sex S, or  $\beta^S = (\beta_0^S, \dots, \beta_{97}^S)$

**Consistent estimators of  $\gamma, \beta^M$  and  $\beta^F$**  can be obtained by **non-linear least squares methods**, but it is likely that the **estimated calorie-age profile** will exhibit **too much variability** (or more variation that is plausible in the underlying nutrient-age relationship)

This happens because the estimator above will fit a curve **only taking into account goodness-of-fit**

Since we are interested also in **smoothness of the estimated relationship** we employ a **roughness penalty approach**: it tries to compromise between two, often conflicting, aims in curve fitting, goodness-of-fit and smoothness.

*(Example of estimation without smoothing provided below)*

### **More details:**

- on the approach in general: Green and Silverman (1994)
- implementation when the functions are discretized: Chesher (1997)

This **procedure seems to work in practice**: Naska, Vasdekis and Trichopoulou, *Public Health Nutrition*, 2001 compares individualised (with the approach above) household budget survey (HBS) with individual nutrition surveys (INS)

# Implementation and data issues

## Outcome is: household daily caloric intake

Calories derived converting household food consumptions

Whole range of ages recorded exploited: **age 0 to 97**

## Household characteristics are **omitted** for the moment

*Chesher (1997) finds that introducing household characteristics has little effect on the smoothness and shape of the estimated energy-intake profile*

*However, we will show some results allowing for education of head and partner, if time permits*

## Progresa (now Oportunidades)

The program **has multiple objectives**, primarily improving educational, health and nutritional status of poor families, **particularly of children and their mother.**

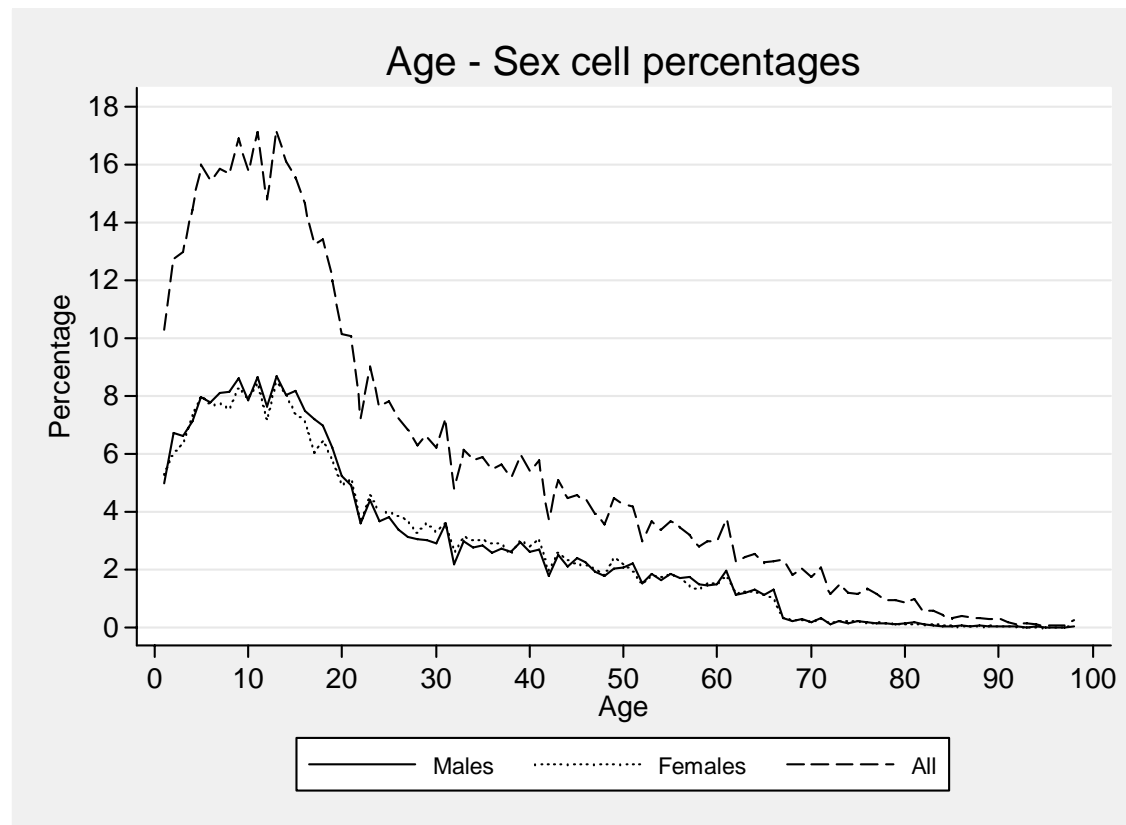
Adoption of an **experimental design in the early stage** of the implementation:  
localities randomized

**Data** is the evaluation sample, **wave nov 1999**

2 years of exposure to the program

# Samples

	Treat	Control	
Poor	8889	5990	14879
Non Poor	2365	1638	4003
	11254	7628	18882





Patterns:

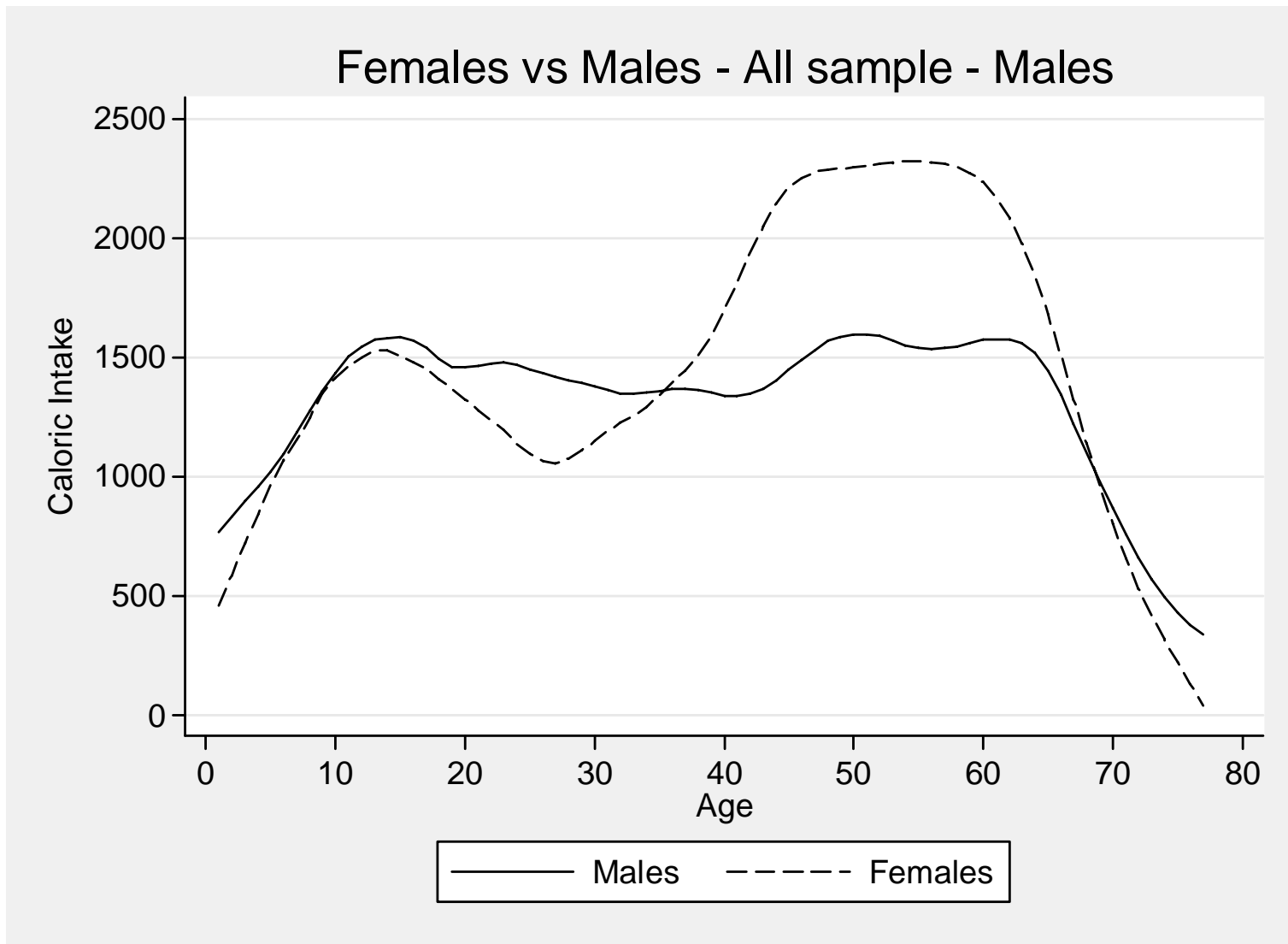
- puberty peak

- not far from recommended values (for ex. 890 kcal for age 0-5; around 2000 kcal for adults)



Patterns:

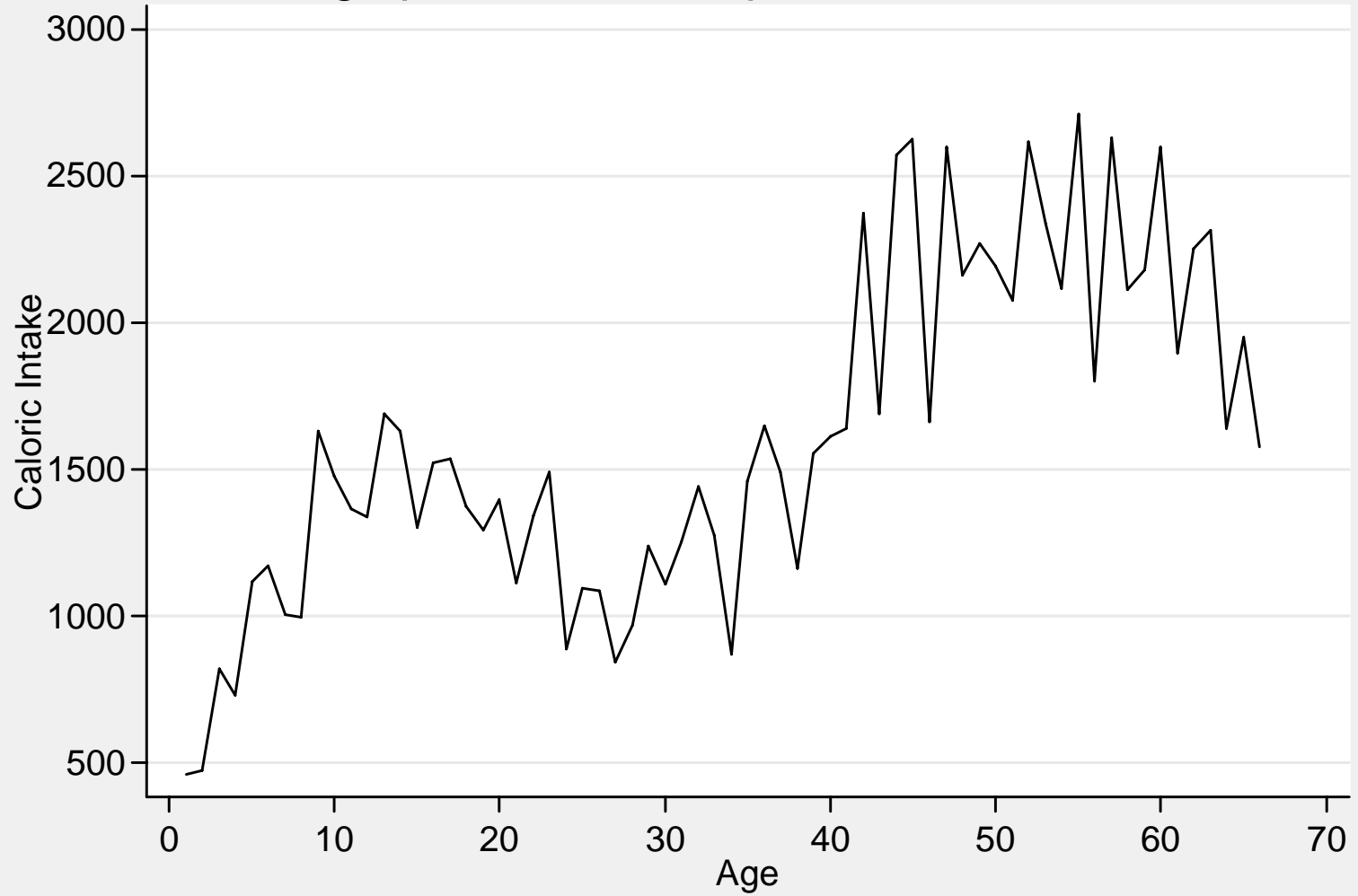
- puberty peak, then steady afterwards
- too low intake with respect to a benchmark (for ex. recommended daily intake for a male age 15-19 is 3000 kcal)



**Summary:** reasonable profile for females; values are too low for males

These **differences** will be **explored further** below

Calorie - Age profile - All sample - Females - No smoothing



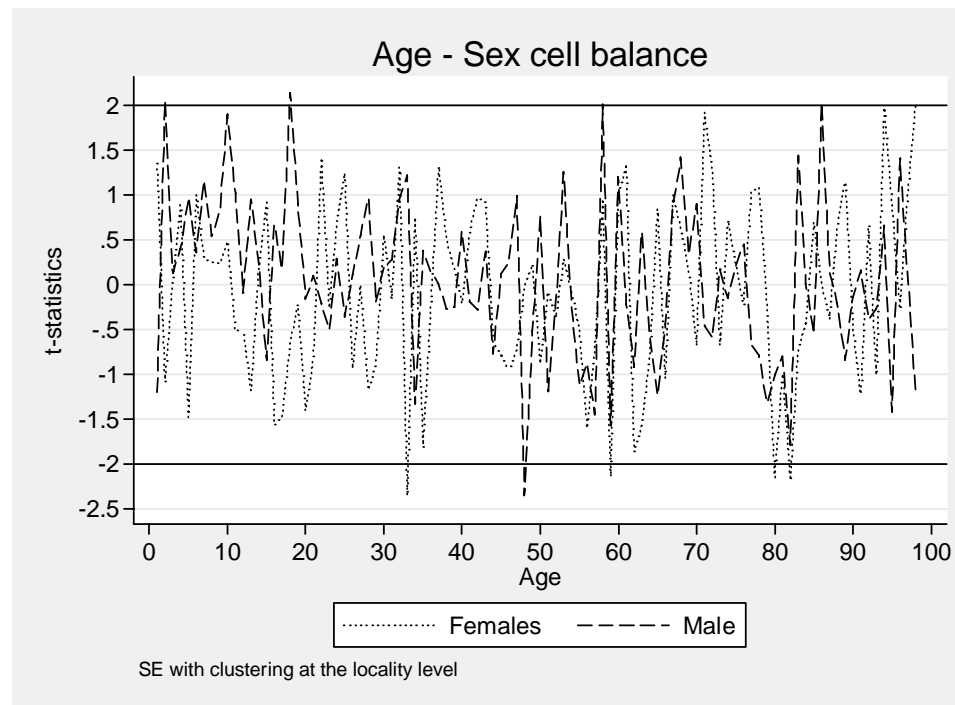
# Program Impact and Heterogeneity

In general we want to estimate the **Treatment effect on the treated, TT**

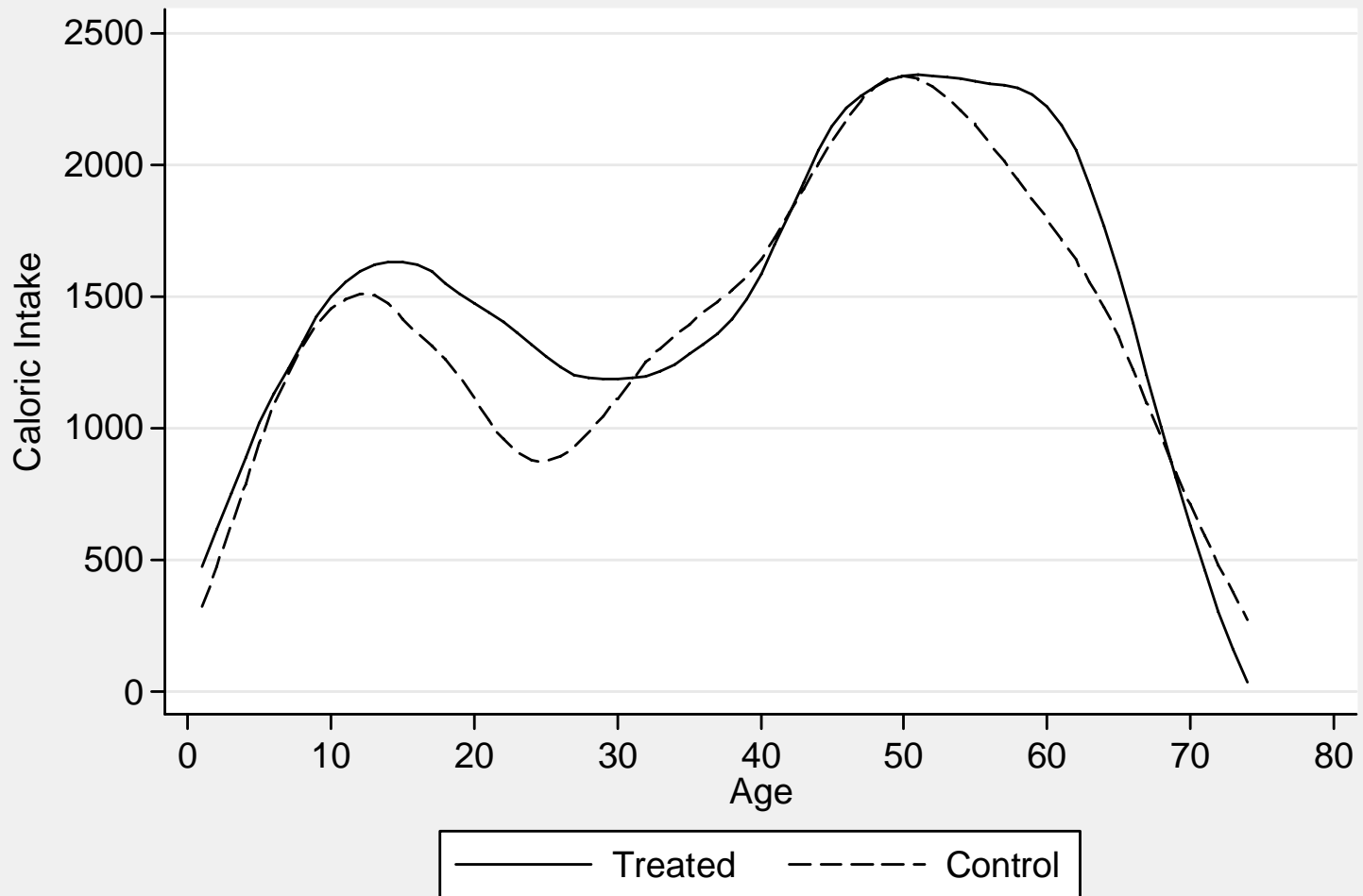
We repeat the estimation on the **samples: “Treat and Poor” and “Control and Poor”**  
*using the same smoothing parameter*

**Heterogeneity** comes into the picture since we **allow for a different impact by sex and age, AgeSexTT**

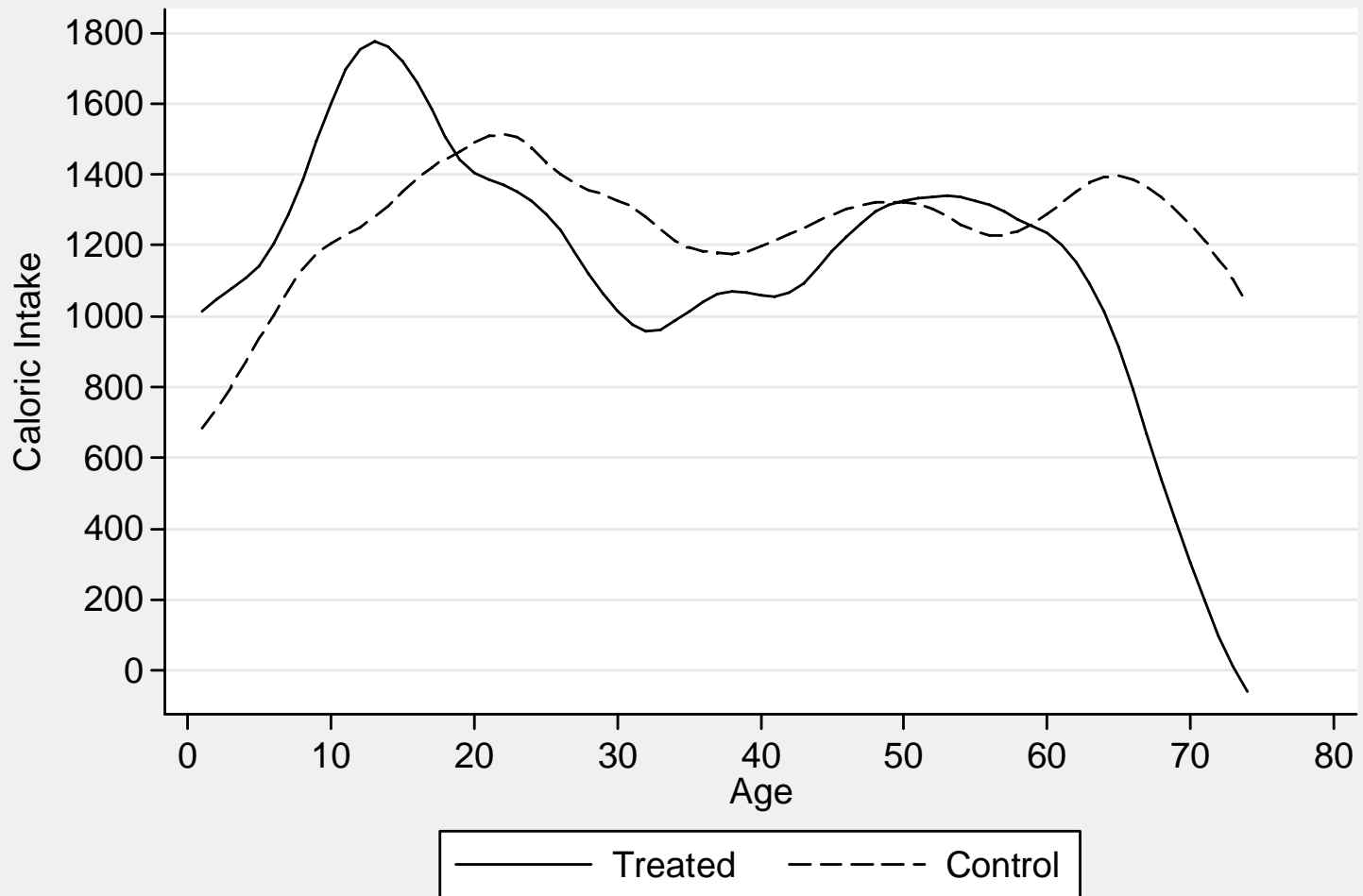
**Randomization was successful**, Behrman and Todd (1999)



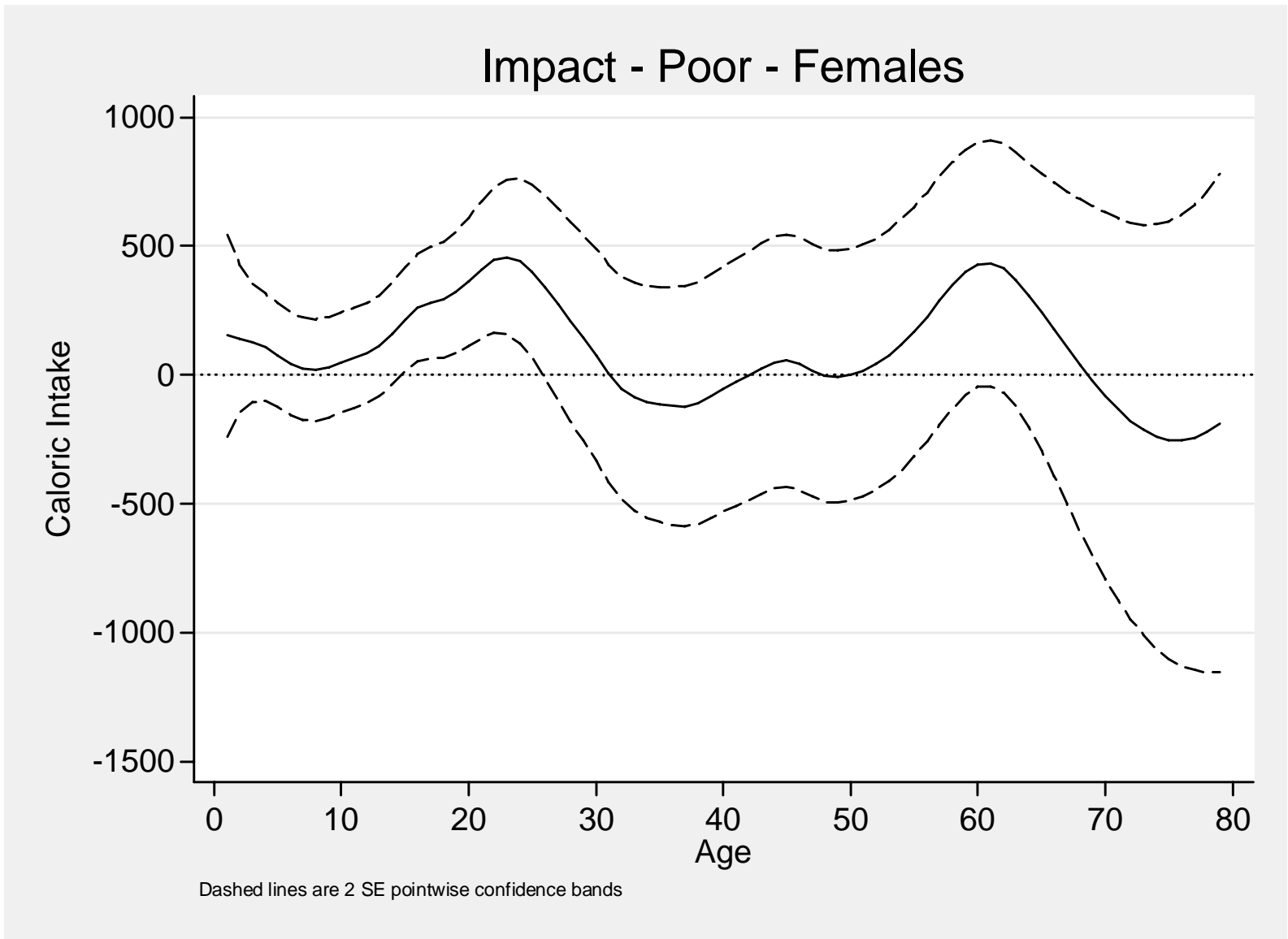
### Treated vs Control - Poor - Females



### Treated vs Control - Poor - Males

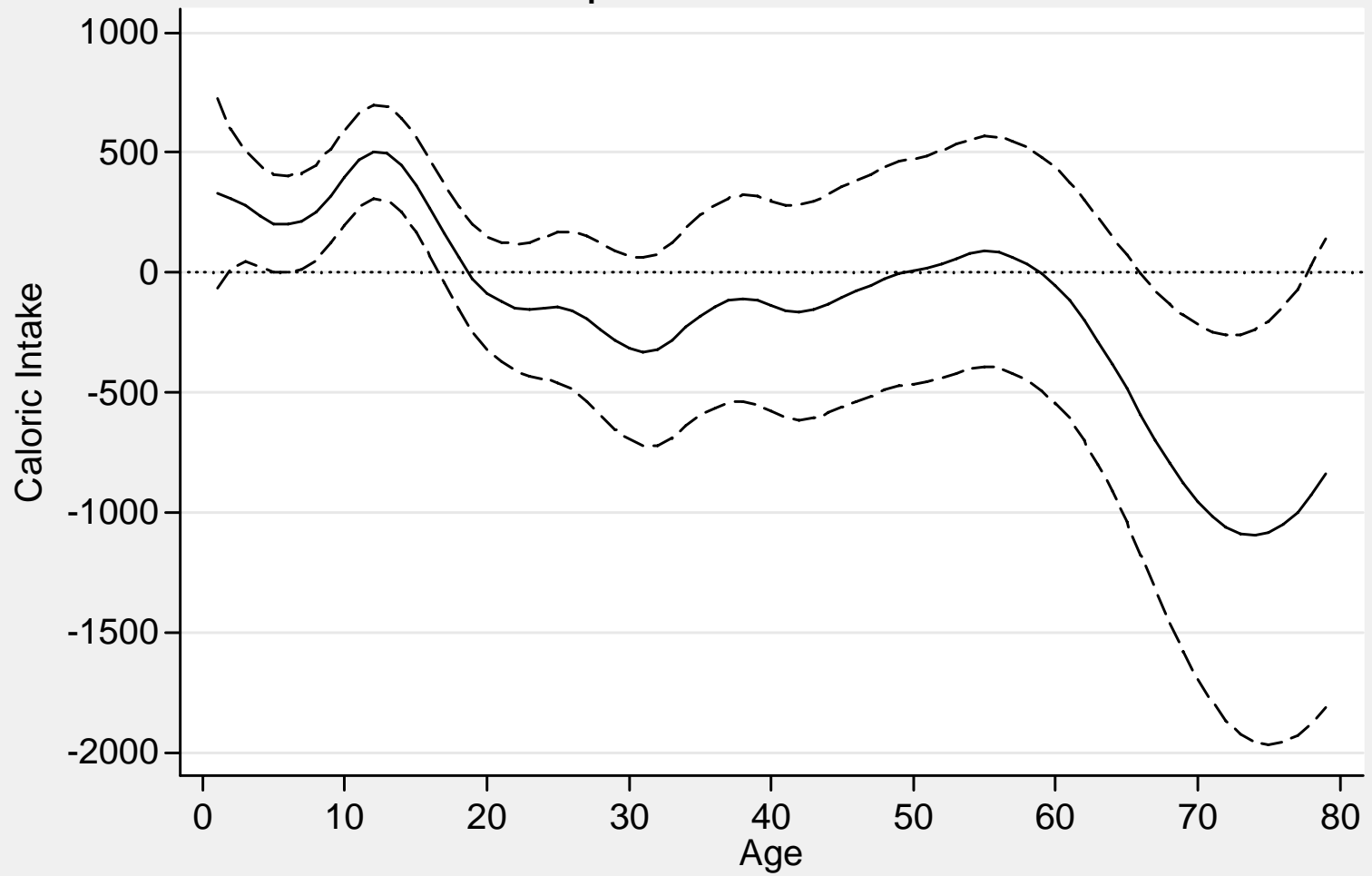


# Age Sex TT



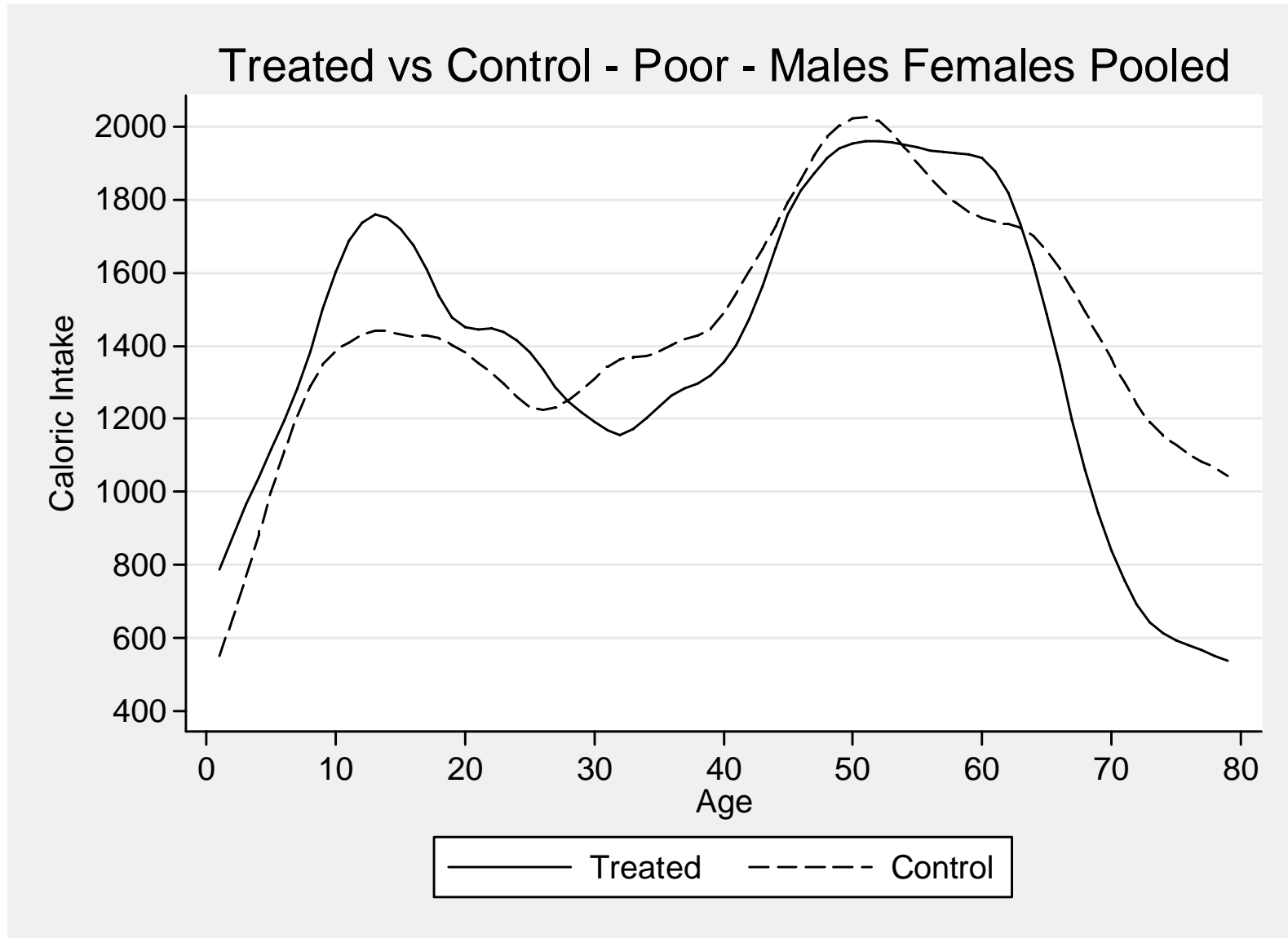
T

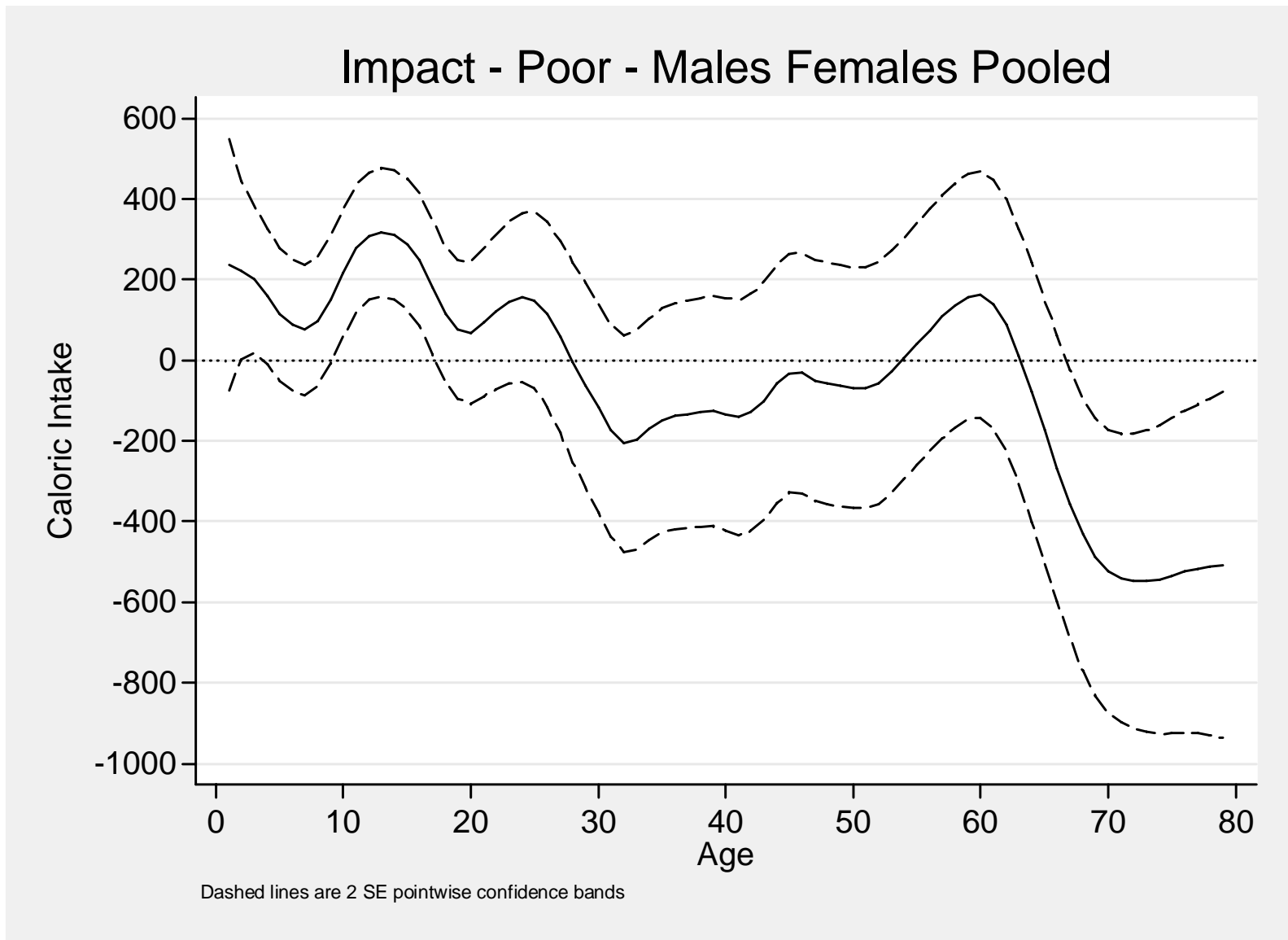
# Impact - Poor - Males



Dashed lines are 2 SE pointwise confidence bands

# Impact - one level of aggregation backwards (Age TT)





**ATT** estimated on the same sample is **203** (se 14.59) or **8.7%**

## To sum up:

The program by design wants to have an **impact within the household:**  
<..*nutritional status of poor families, particularly of children and their mother*>

Aggregate **household data does not allow** (directly) **to estimate impacts at the individual level**

We **looked within the household (estimation of the individual caloric intake)** and we found that the **program is achieving its objectives:**

**stronger impact** for **younger ages** (both male and females)

increased impact for **females** lasts **till age 30** (mothers)

## Asymmetric information regarding food consumptions?

- example of **additional dimensions** that can be studied only with this derived information at the individual level.
- **difference** in the profiles for **men and women**
  - not satisfied with the estimated calorie-age profile for men  
(caloric intake is unreasonably low)

This point seems worthwhile to be investigated further. Estimation? Data?

At the current stage **one (preliminary and tentative) story is:**

**Respondent** might **not have a good information** of household activities made by **specific age-sex groups within the family**

REF: **Boozer and Goldstein, WP, 2003** with dataset from Ghana

**Husbands and wives interviewed separately.** Survey asks each respondent to report their own expenditure, **the expenditure of their spouse (cross-reporting)**, and the expenditure of any other person in the household that was used for household consumption.

They find that **some components** of consumption are “**private**” in nature, and thus essentially **unobserved in the cross reports**

**Notice:** they exploit particular feature (**cross-reporting**) of Ghanaian dataset

Thanks to the **individualisation approach** we are able (try) to **study a similar issue** with a general household level survey (where there is **no cross-reporting**)

The sample we are using is “asymmetric” in terms of who responds

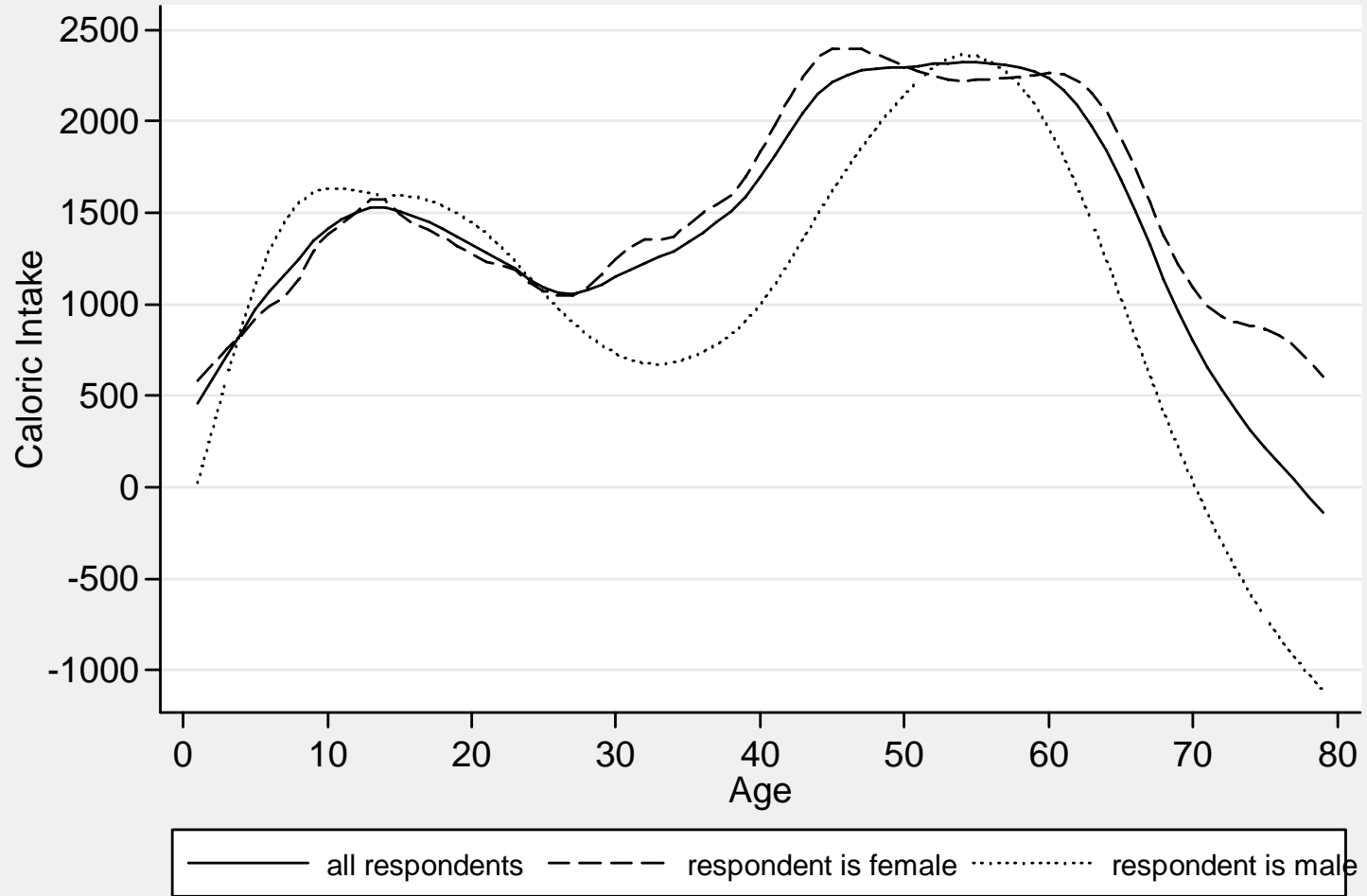
Respondent is most likely to be a woman: 85% of questionnaires

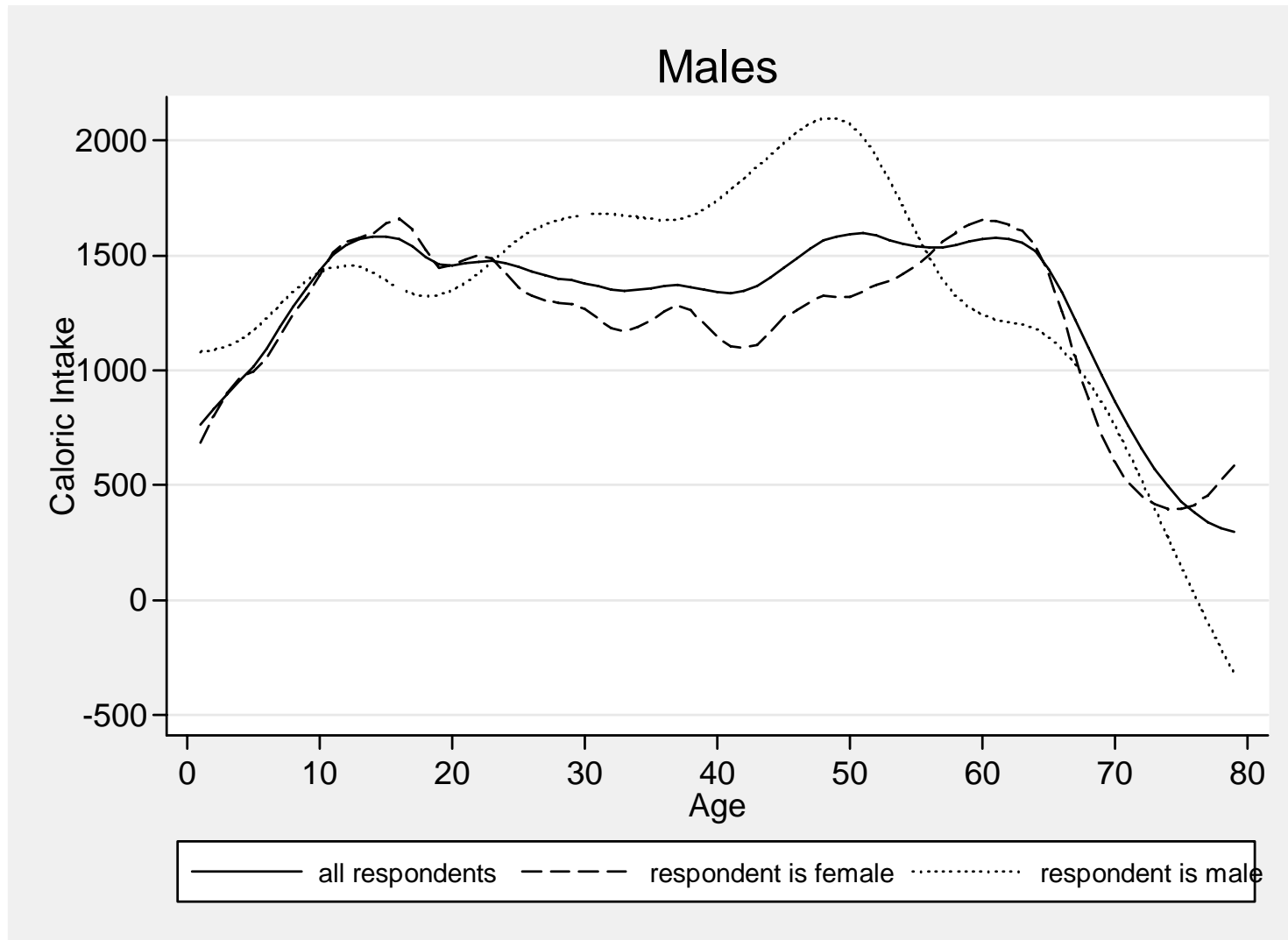
A **simple strategy** is employed to study whether there are “**private**” **components of consumptions**

**Estimation is repeated** in the samples:

- only male respondents
- only female respondents

# Females





- when **respondent is male** the **shape is different** (increasing for adults) and caloric intake is higher (somewhat closer to reasonable values) **for adults**.
- for **younger ages** it does **not seem substantially different**

**Conclusion:** (weak) evidence of **under-reporting of food intake** (caloric intake) of **other-sex adult members** in the household.

In particular (since **most of the respondents are females** in our sample):  
women seem to have **distorted information** on **food intake of male adults**

**Story is tentative** but it seems to **make a case for further research** on this issue

# Summary

We have tried to **look within the household...**

...using a **model of food consumptions** and a **semi-parametric approach**

...and data from the **Progresa** evaluation sample

so as to **estimate the individual caloric intake** (by age and sex)

Findings show that:

The **impact** of the program is **heterogenous**  
*stronger for younger ages and for mothers*

...and it is **in line** with the **focus of the intervention**

Some **components of consumption** might be **“private”** in nature.

## Next steps

Use of **dataset** with **individual consumptions** (only a subset of localities)

*Assess whether the individualisation works in this context*

Explore further the issue of **asymmetric information** held by **respondents**

*Suggestions?*

Disentangle the **impact** of the “**informational component**” of the program

*Program wants to change (improve) nutritional and health practices.*

*Some groups within the household are more likely to “get in contact” with the **informational component** of the program (pregnant and lactating women)*

*Estimate the **micronutrient-age profile** and assess whether the **quality of diet** is changing in a different way for different age- and sex- groups*

**End**

**Thank You!**