Using Time Use Data to Parameterize Models for the Spread of Close-contact Infectious Diseases

Emilio Zagheni[1] Piero Manfredi[2] Francesco C. Billari[3]

[1]Department of Demography - UC Berkeley

[2]Dipart. di Statistica e Matematica Applicata all'Economia - University of Pisa

[3]"Carlo F. Dondena" Centre for Research on Social Dynamics - Bocconi University

POLYMOD Meeting, Antwerp, June 8th-11th 2008

Why time use data

- Time of exposure of susceptibles to infectious people is a reasonable indicator of the probability of transmission of certain diseases
- Time use data have been collected in a lot of countries and can be used for comparative purposes. They are immediately available and inexpensive or free.
- Time use surveys offer good quality data that have not been explored for epidemiological purposes yet
- Time use surveys can be used to test the results of contact surveys against an independent source

Why time use data

- Time of exposure of susceptibles to infectious people is a reasonable indicator of the probability of transmission of certain diseases
- Time use data have been collected in a lot of countries and can be used for comparative purposes. They are immediately available and inexpensive or free.
- Time use surveys offer good quality data that have not been explored for epidemiological purposes yet
- Time use surveys can be used to test the results of contact surveys against an independent source

Why time use data

- Time of exposure of susceptibles to infectious people is a reasonable indicator of the probability of transmission of certain diseases
- Time use data have been collected in a lot of countries and can be used for comparative purposes. They are immediately available and inexpensive or free.
- Time use surveys offer good quality data that have not been explored for epidemiological purposes yet
- Time use surveys can be used to test the results of contact surveys against an independent source

Why time use data

- Time of exposure of susceptibles to infectious people is a reasonable indicator of the probability of transmission of certain diseases
- Time use data have been collected in a lot of countries and can be used for comparative purposes. They are immediately available and inexpensive or free.
- Time use surveys offer good quality data that have not been explored for epidemiological purposes yet
- Time use surveys can be used to test the results of contact surveys against an independent source

Outline

- Data
- Summary of the time use methodology and some empirical results for the United States

  See: Zagheni E, Billari FC, Manfredi P, Melegaro A, Mossong, J and Edmunds WJ. *(Forthcoming)*. Using time use data to parameterize models for the spread of close-contact infectious diseases. *American Journal of Epidemiology*
- New empirical estimates of time of exposure matrices for Italy
- Comparison of time of exposure vs contact matrices in terms of fit to serological data

Outline

- Data
- Summary of the time use methodology and some empirical results for the United States

  See: Zagheni E, Billari FC, Manfredi P, Melegaro A, Mossong, J and Edmunds WJ. *(Forthcoming)*. Using time use data to parameterize models for the spread of close-contact infectious diseases. *American Journal of Epidemiology*

- New empirical estimates of time of exposure matrices for Italy
- Comparison of time of exposure vs contact matrices in terms of fit to serological data

Outline

- Data
- Summary of the time use methodology and some empirical results for the United States

  See: Zagheni E, Billari FC, Manfredi P, Melegaro A, Mossong, J and Edmunds WJ. *(Forthcoming)*. Using time use data to parameterize models for the spread of close-contact infectious diseases. *American Journal of Epidemiology*
- New empirical estimates of time of exposure matrices for Italy
- Comparison of time of exposure vs contact matrices in terms of fit to serological data

Outline

- Data
- Summary of the time use methodology and some empirical results for the United States

  See: Zagheni E, Billari FC, Manfredi P, Melegaro A, Mossong, J and Edmunds WJ. *(Forthcoming)*. Using time use data to parameterize models for the spread of close-contact infectious diseases. *American Journal of Epidemiology*
- New empirical estimates of time of exposure matrices for Italy
- Comparison of time of exposure vs contact matrices in terms of fit to serological data

Time use (TU) data

Data collected through time-structured diaries

Time, pm	**What were you doing?** Record your main activity for each 10-minute period from 04.00 pm to 07.00 am! Only one main activity on each line! Distinguish between travel and the activity that is the reason for travelling. Do not forget the mode of transportation. Distinguish between first and second job, if any.	**What else were you doing?** Record the most important parallel activity.	**Were you alone or together with somebody you know?** Mark "yes" by crossing: Alone	Children up to 9 living in your household	Other household members	Other persons that you know	
04.00-04.10	Bus from work to the day care centre	Planned a birthday party for my son	☒	☐	☐	☐	
04.10-04.20	-"-	-"-	☒	☐	☐	☐	
04.20-04.30	Talked with the child minder	Helped the children dressing	☐	☒	☐	☒	
04.30-04.40	Went to the grocery, on foot	Talked with my children	☐	☒	☐	☐	
04.40-04.50	Bought food for my family and my neighbour		☐	☒	☐	☐	
04.50-05.00	Went home on foot		☐	☒	☐	☐	
05.00-05.10	Delivered food to my neighbour		☐	☒	☐	☒	
05.10-05.20	Put own food in fridge		☒	☐	☐	☐	
05.20-05.30	Cooked supper	Listened to the radio	☒	☐	☐	☐	
05.30-05.40	-"-	-"-	☐	☐	☒	☐	
05.40-05.50	Had supper	Talked with my family	☐	☒	☒	☐	
05.50-06.00	-"-	-"-	☐	☒	☒	☐	
06.00-06.10	-"-	-"-	☐	☒	☒	☐	
06.10-06.20	Cleared the table		☐	☐	☒	☐	
06.20-06.30	Had a rest		☒	☐	☐	☐	
06.30-06.40	Watched TV with my children	Knitted	☐	☒	☒	☐	
06.40-06.50	-"-	-"-	☐	☒	☒	☐	
06.50-07.00	-"-	-"-	☐	☒	☒	☐	

Data - before POLYMOD surveys

Time use data:

- American Time Use Study (ATUS) 2003; Activity Pattern Survey of California Children 1989-1990 and Activity Pattern Survey of Californians 1987-1988

Seroprevalence data for varicella:

- National Health and Nutrition Examination Surveys (NHANES III) 1988-1994; Behavioral Risk Factor Surveillance System (BRFSS) 1991

A richer collection of data sets for Italy

Time use data

- Italian Time Use Survey 2002-2003 (collected by ISTAT)

Social Contacts data

- Italian Contact Survey - available through Polymod

Seroprevalence data

- Serological survey for VZV and parvovirus B19 - available through Polymod

Assumptions

- Time of exposure between people of different age groups is a relevant epidemiological variable
- Proportionate time mixing (PTM) at the level of single activity/location and time slot.
  The idea is that when we consider a specific location and small time intervals, people of different age groups divide their time with people of other age groups proportionately to their relative presence in the location during the time interval considered

Assumptions

- Time of exposure between people of different age groups is a relevant epidemiological variable
- Proportionate time mixing (PTM) at the level of single activity/location and time slot.
  The idea is that when we consider a specific location and small time intervals, people of different age groups divide their time with people of other age groups proportionately to their relative presence in the location during the time interval considered

How to compute the *ij*th element of the exposure matrix

- **Step 1**: Divide the 24-hours day into 1440 time slots each of which consists of one minute and define

$$T_i^{hz}$$

as the number of people belonging to the age group i that are in the location h during the time slot z. It is equal to the number of minutes spent by the population in the age group i, in the location h and during that particular time slot z of the day considered: it is a measure of person-minutes.

- **Step 2:** compute the time of exposure of people in the age group i to people in the age group j, in the location h and for the time slot z is:

$$T_i^{hz} \times \frac{T_j^{hz}}{\sum_{k=1}^{n} T_k^{hz}}$$

How to compute the *ij*th element of the exposure matrix

- **Step 1**: Divide the 24-hours day into 1440 time slots each of which consists of one minute and define

$$T_i^{hz}$$

as the number of people belonging to the age group i that are in the location h during the time slot z. It is equal to the number of minutes spent by the population in the age group i, in the location h and during that particular time slot z of the day considered: it is a measure of person-minutes.

- **Step 2:** compute the time of exposure of people in the age group i to people in the age group j, in the location h and for the time slot z is:

$$T_i^{hz} \times \frac{T_j^{hz}}{\sum_{k=1}^{n} T_k^{hz}}$$

- **Step 3:** Sum up the matrices for all 1440 time slots and all the L locations considered:

$$\sum_{h=1}^{L}\sum_{z=1}^{1440}\left(T_i^{hz} \times \frac{T_j^{hz}}{\sum_{k=1}^{n} T_k^{hz}}\right)$$

- **Step 4:** Add the matrix of time of exposure between household members (if you can obtain it directly from the survey data)
- **Step 5:** Divide the elements of the exposure matrix by the population size of the respective age groups to find the average daily time of exposure between age groups

- **Step 3:** Sum up the matrices for all 1440 time slots and all the L locations considered:

$$\sum_{h=1}^{L}\sum_{z=1}^{1440}\left(T_i^{hz} \times \frac{T_j^{hz}}{\sum_{k=1}^{n} T_k^{hz}}\right)$$

- **Step 4:** Add the matrix of time of exposure between household members (if you can obtain it directly from the survey data)
- **Step 5:** Divide the elements of the exposure matrix by the population size of the respective age groups to find the average daily time of exposure between age groups

- **Step 3:** Sum up the matrices for all 1440 time slots and all the L locations considered:

$$\sum_{h=1}^{L}\sum_{z=1}^{1440}(T_i^{hz} \times \frac{T_j^{hz}}{\sum_{k=1}^{n} T_k^{hz}})$$

- **Step 4:** Add the matrix of time of exposure between household members (if you can obtain it directly from the survey data)
- **Step 5:** Divide the elements of the exposure matrix by the population size of the respective age groups to find the average daily time of exposure between age groups

Example

- Consider the location 'pub' between 8:00pm and 8:01pm and three age groups:
  - 20 people in the age group 15-19 are there during the time slot
  - 50 people in the age group 20-24 are there during the time slot
  - 30 people in the age group 25-29 are there during the time slot
- Overall, for the specific location and time slot:
- 15-19 accounts for 20/100= 20% of the person-minutes
- 20-24 accounts for 50/100= 50% of the person-minutes
- 25-29 accounts for 30/100= 30% of the person-minutes

$$\begin{bmatrix} 20 \times 0.2 & 20 \times 0.5 & 20 \times 0.3 \\ 50 \times 0.2 & 50 \times 0.5 & 50 \times 0.3 \\ 30 \times 0.2 & 30 \times 0.5 & 30 \times 0.3 \end{bmatrix}$$

Example

- Consider the location 'pub' between 8:00pm and 8:01pm and three age groups:
  - 20 people in the age group 15-19 are there during the time slot
  - 50 people in the age group 20-24 are there during the time slot
  - 30 people in the age group 25-29 are there during the time slot
- Overall, for the specific location and time slot:
- 15-19 accounts for $20/100 = 20\%$ of the person-minutes
- 20-24 accounts for $50/100 = 50\%$ of the person-minutes
- 25-29 accounts for $30/100 = 30\%$ of the person-minutes

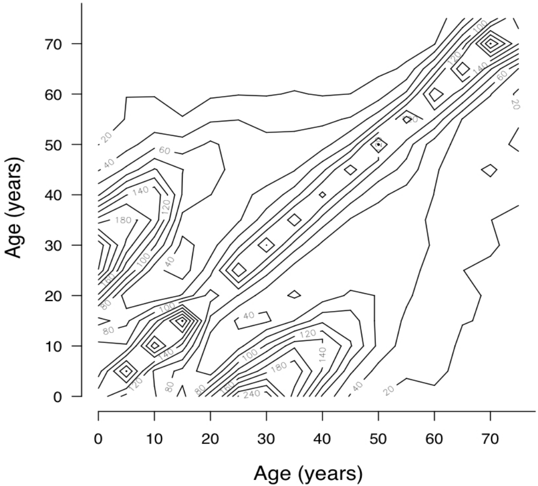
$$\begin{bmatrix} 20 \times 0.2 & 20 \times 0.5 & 20 \times 0.3 \\ 50 \times 0.2 & 50 \times 0.5 & 50 \times 0.3 \\ 30 \times 0.2 & 30 \times 0.5 & 30 \times 0.3 \end{bmatrix}$$

Example

- Consider the location 'pub' between 8:00pm and 8:01pm and three age groups:
    - 20 people in the age group 15-19 are there during the time slot
    - 50 people in the age group 20-24 are there during the time slot
    - 30 people in the age group 25-29 are there during the time slot
- Overall, for the specific location and time slot:
- 15-19 accounts for $20/100 = 20\%$ of the person-minutes
- 20-24 accounts for $50/100 = 50\%$ of the person-minutes
- 25-29 accounts for $30/100 = 30\%$ of the person-minutes

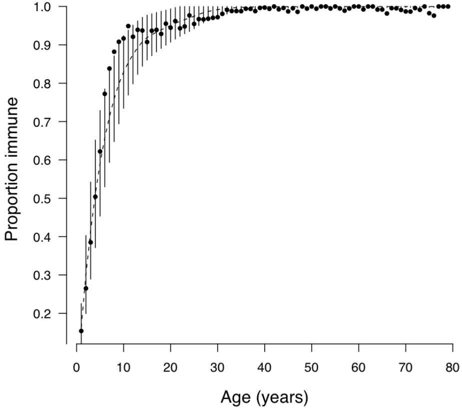
$$\begin{bmatrix} 20 \times 0.2 & 20 \times 0.5 & 20 \times 0.3 \\ 50 \times 0.2 & 50 \times 0.5 & 50 \times 0.3 \\ 30 \times 0.2 & 30 \times 0.5 & 30 \times 0.3 \end{bmatrix}$$

USA Time of exposure matrix

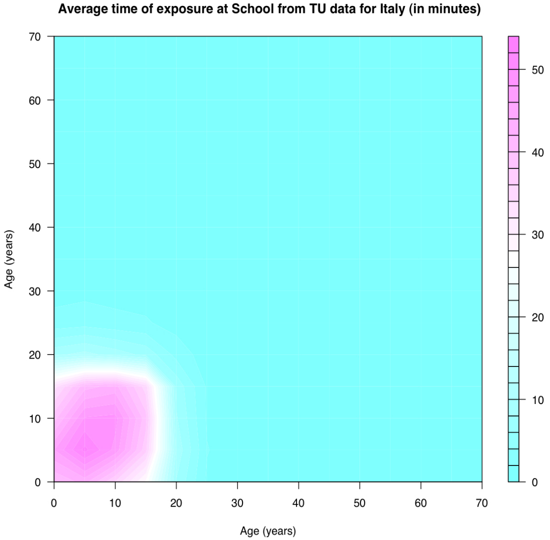


VZV: fit to serological data based on USA Time Use data

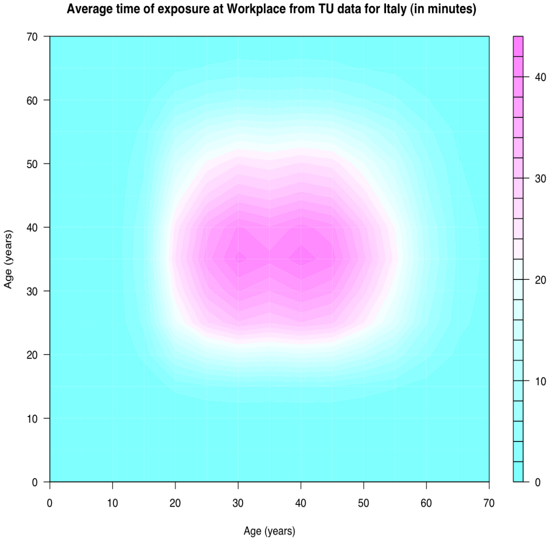
Assumption: the number of potentially infectious events between people of different age groups is proportional to their average time of exposure (see Wallinga et al., 2006)



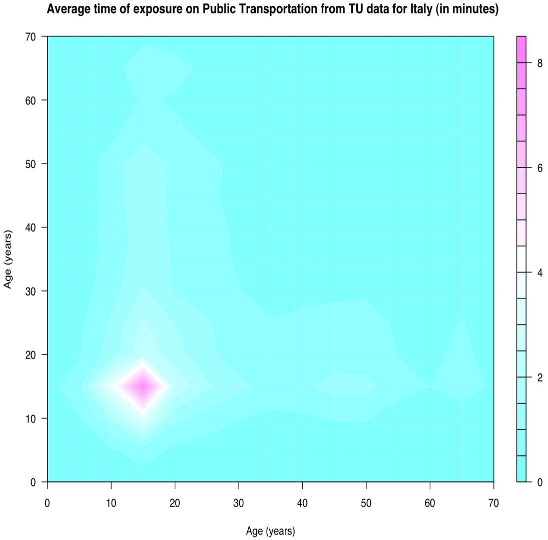
Time of exposure matrices for Italy: School



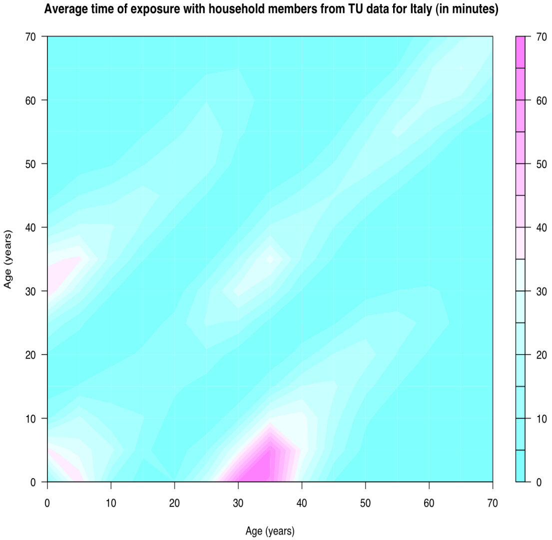
Workplace


Average time of exposure at Workplace from TU data for Italy (in minutes)
Age (years)
Age (years)
0
10
20
30
40
50
60
70
0
10
20
30
40

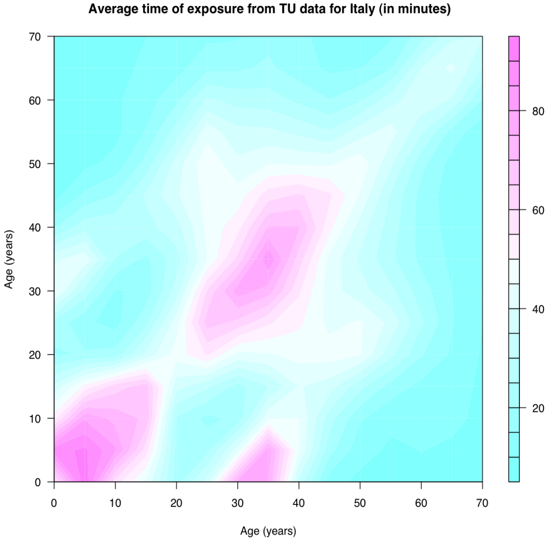
Public Transportation



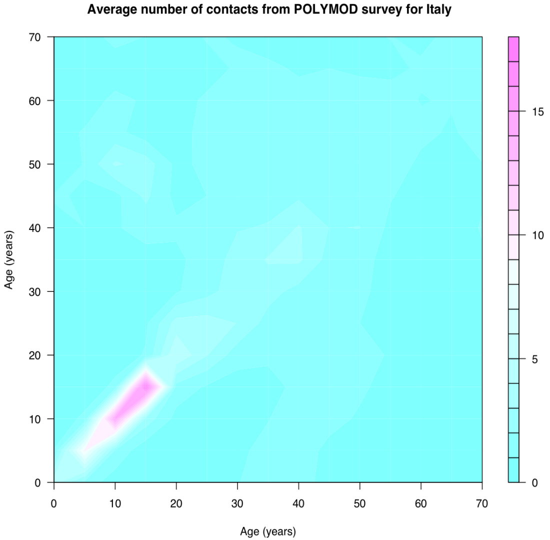
Time of exposure to household members



Average time of exposure



Average number of contacts



Data Source: Mossong et al. (2008)

Fit to serological data

- We test the ability of time of exposure matrices and contact matrices to fit seroprofiles for VZV and parvovirus B19
- In order to make the matrices comparable, we consider proportions (of time of exposure or number of contacts by age): each row of the matrix sums to 1.
- We determine age-specific transmission parameters by multiplying each element of the contact matrix or time of exposure matrix by a scaling factor q. Estimate q by maximum likelihood.
- We also consider a linear combination of contact matrix and time of exposure matrix (two q parameters) and linear combinations of location-specific time of exposure matrices (school, public transportation, workplace, household exposure)

Fit to serological data

- We test the ability of time of exposure matrices and contact matrices to fit seroprofiles for VZV and parvovirus B19
- In order to make the matrices comparable, we consider proportions (of time of exposure or number of contacts by age): each row of the matrix sums to 1.
- We determine age-specific transmission parameters by multiplying each element of the contact matrix or time of exposure matrix by a scaling factor q. Estimate q by maximum likelihood.
- We also consider a linear combination of contact matrix and time of exposure matrix (two q parameters) and linear combinations of location-specific time of exposure matrices (school, public transportation, workplace, household exposure)

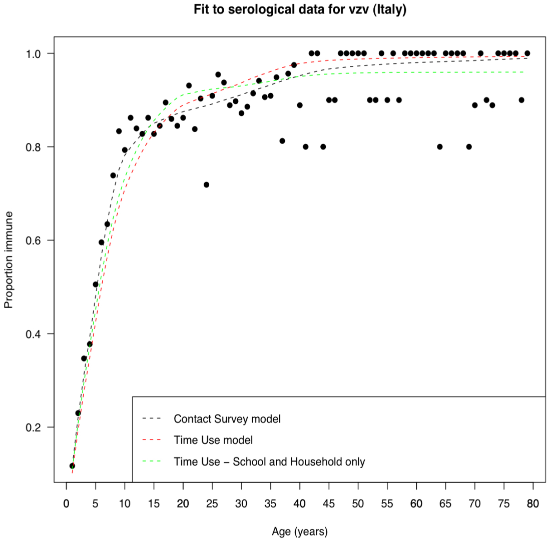
Fit to serological data

- We test the ability of time of exposure matrices and contact matrices to fit seroprofiles for VZV and parvovirus B19
- In order to make the matrices comparable, we consider proportions (of time of exposure or number of contacts by age): each row of the matrix sums to 1.
- We determine age-specific transmission parameters by multiplying each element of the contact matrix or time of exposure matrix by a scaling factor q. Estimate q by maximum likelihood.
- We also consider a linear combination of contact matrix and time of exposure matrix (two q parameters) and linear combinations of location-specific time of exposure matrices (school, public transportation, workplace, household exposure)

Fit to serological data

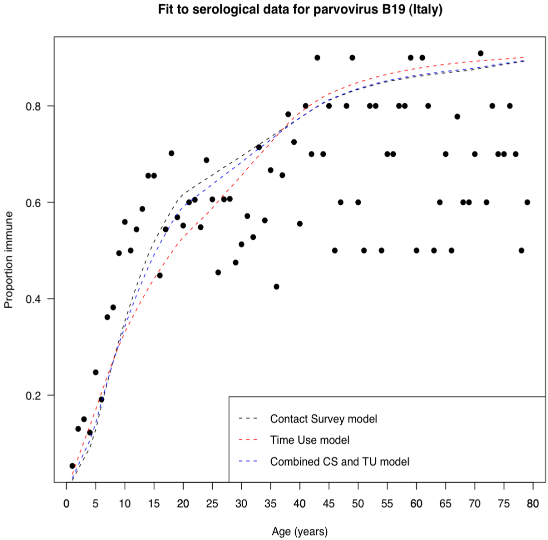
- We test the ability of time of exposure matrices and contact matrices to fit seroprofiles for VZV and parvovirus B19
- In order to make the matrices comparable, we consider proportions (of time of exposure or number of contacts by age): each row of the matrix sums to 1.
- We determine age-specific transmission parameters by multiplying each element of the contact matrix or time of exposure matrix by a scaling factor q. Estimate q by maximum likelihood.
- We also consider a linear combination of contact matrix and time of exposure matrix (two q parameters) and linear combinations of location-specific time of exposure matrices (school, public transportation, workplace, household exposure)

VZV Italy: contact survey data fit the seroprofile better



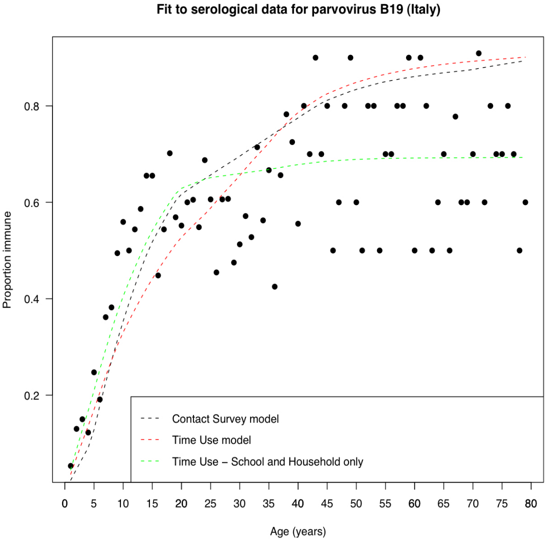
Deviance CS= 76; Deviance TU= 143; Deviance TU by loc.=95

Parvovirus B19: combining TU and CS data



Deviance CS= 251; Deviance TU= 266; Deviance combined= 248

B19: TU data by location fit the seroprofile better



Deviance CS= 251; Deviance TU: 266; Deviance TU by loc: 108

What do these results tell us?

- Contact matrices and time of exposure matrices give a slightly different representation of social interaction: qualitative similar, but different weight is given to interaction between children/teenagers and with household members
- This explains the different profile of immunity predicted by the the two different models
- Depending on the infectivity of the disease and the routes of transmission, one approach can be more suitable than the other one
- Linear combinations of time of exposure at school and with household members give a flexible model that can fit well different seroprofiles

What do these results tell us?

- Contact matrices and time of exposure matrices give a slightly different representation of social interaction: qualitative similar, but different weight is given to interaction between children/teenagers and with household members
- This explains the different profile of immunity predicted by the the two different models
- Depending on the infectivity of the disease and the routes of transmission, one approach can be more suitable than the other one
- Linear combinations of time of exposure at school and with household members give a flexible model that can fit well different seroprofiles

What do these results tell us?

- Contact matrices and time of exposure matrices give a slightly different representation of social interaction: qualitative similar, but different weight is given to interaction between children/teenagers and with household members
- This explains the different profile of immunity predicted by the the two different models
- Depending on the infectivity of the disease and the routes of transmission, one approach can be more suitable than the other one
- Linear combinations of time of exposure at school and with household members give a flexible model that can fit well different seroprofiles

What do these results tell us?

- Contact matrices and time of exposure matrices give a slightly different representation of social interaction: qualitative similar, but different weight is given to interaction between children/teenagers and with household members
- This explains the different profile of immunity predicted by the the two different models
- Depending on the infectivity of the disease and the routes of transmission, one approach can be more suitable than the other one
- Linear combinations of time of exposure at school and with household members give a flexible model that can fit well different seroprofiles

Conclusions

- Time use surveys are an interesting source of secondary data for epidemiological purposes
- Combining the information provided by time use surveys with contact surveys seems a promising task
- It may be extremely interesting to use both data sources to calibrate micro-simulation models and to assess the impact of certain public health interventions

Conclusions

- Time use surveys are an interesting source of secondary data for epidemiological purposes
- Combining the information provided by time use surveys with contact surveys seems a promising task
- It may be extremely interesting to use both data sources to calibrate micro-simulation models and to assess the impact of certain public health interventions

Conclusions

- Time use surveys are an interesting source of secondary data for epidemiological purposes
- Combining the information provided by time use surveys with contact surveys seems a promising task
- It may be extremely interesting to use both data sources to calibrate micro-simulation models and to assess the impact of certain public health interventions