# Modelling long term duration of usual behaviour

## A zero-inflated negative binomial approach

Pierre Walthery & Jonathan Gershuny

$40^{th}$ IATUR Conference, Budapest


Department of Sociology
UNIVERSITY OF OXFORD

# The problem

- Time diaries offer a window into the daily life of respondents...
- But a limited one: either
  - One/two days
  - A full week on a limited range of activities (paid work)
- Some behaviours are part of longer-term cycles/routines
  - Going swimming once a week
- And may occur occasionaly
- Policy relevance of some of these (ie physical exercise)
- How can we capture the latter with the former?

# Previous work

- Nutrition & synthetic data (Tooze et al 2006; 2010)
  - Two stages models
  - Synthetic dataset
- Gershuny et al (2012)
  - Long term prediction of usual behaviour
  - Two stage model
  - Stylised predictors
  - Day is finite
  - Probability of participation results from not participating in similar activities

# The study

- Five datasets
  - 2000 & 2015 UK TUS
  - Nine country studies
  - Home Online
  - US Innovation Panel
- 48,000 days
- Eleven activities
- Three models
  - Socio-demographics
  - Habit ie stylised frequency of specific activity
  - Saturated model (did you do it last week?)

# Activities

- Sleep, rest, self-care
- Paid work,study
- Caring
- Housework, cooking, DIY
- Eating at home
- Volunteering, helping
- Shopping
- Cultural consumption & leisure+
- Socialising & outings+
- Exercising+
- Travel

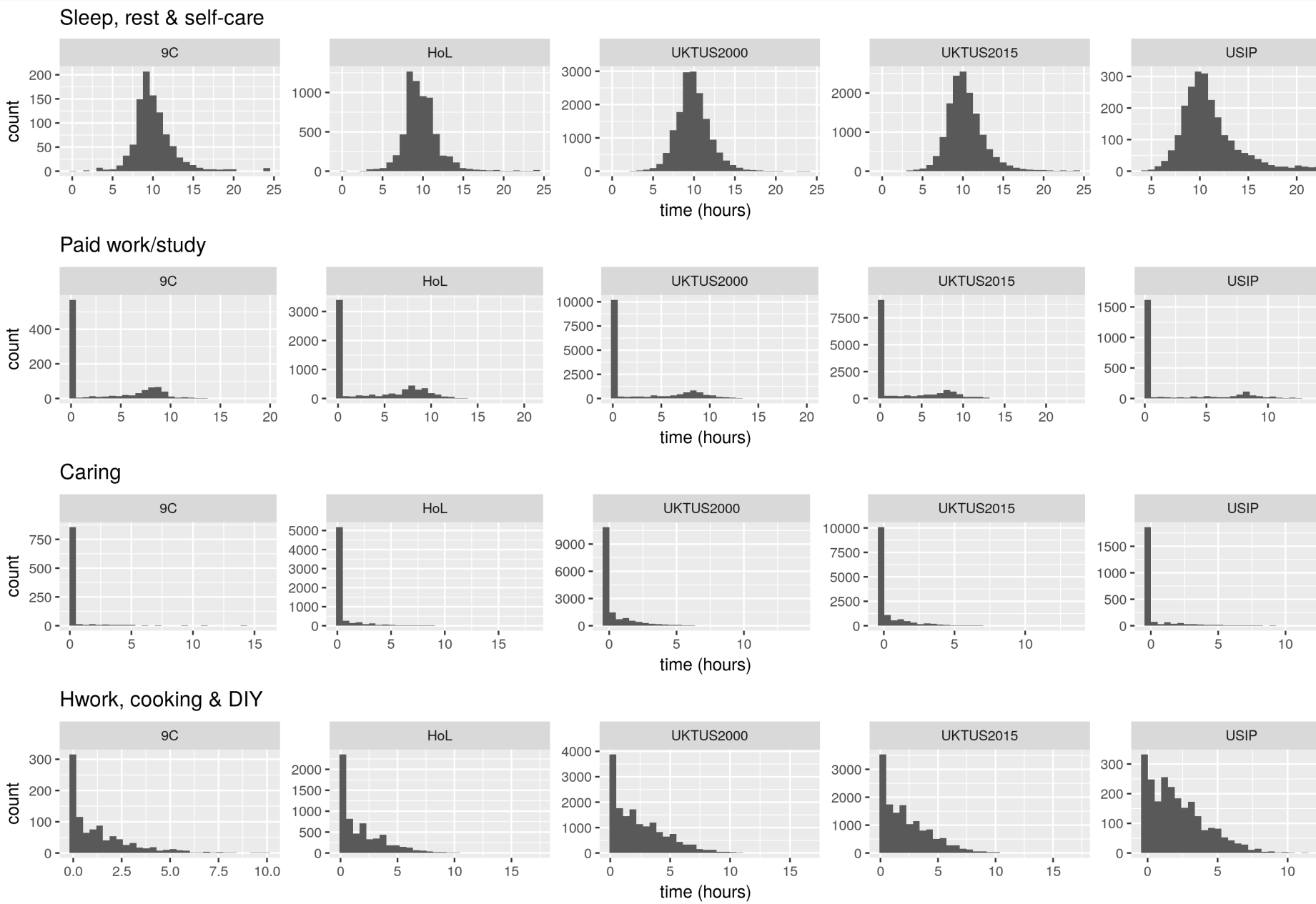
- Harmonised accross the five datasets

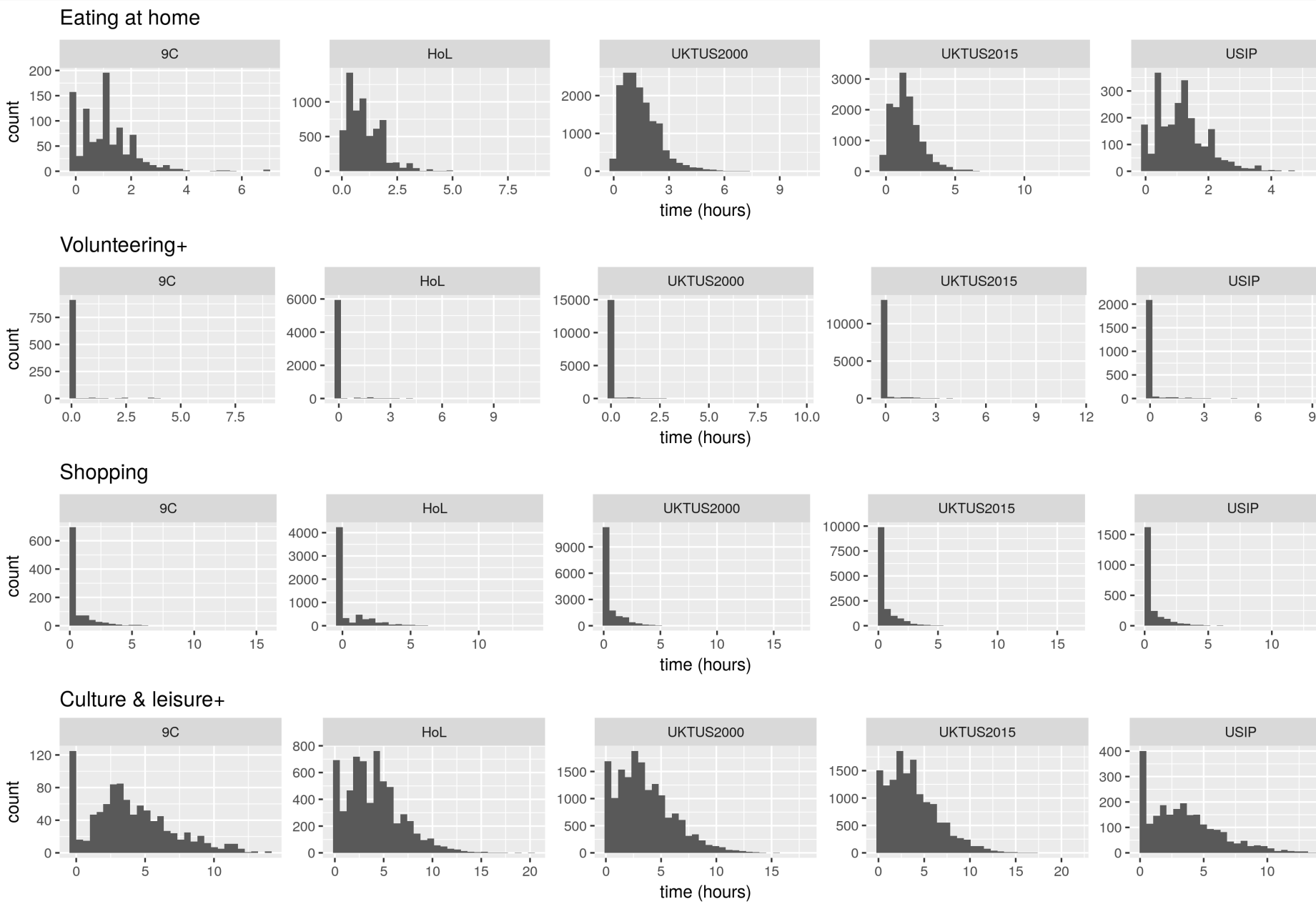
# Variables

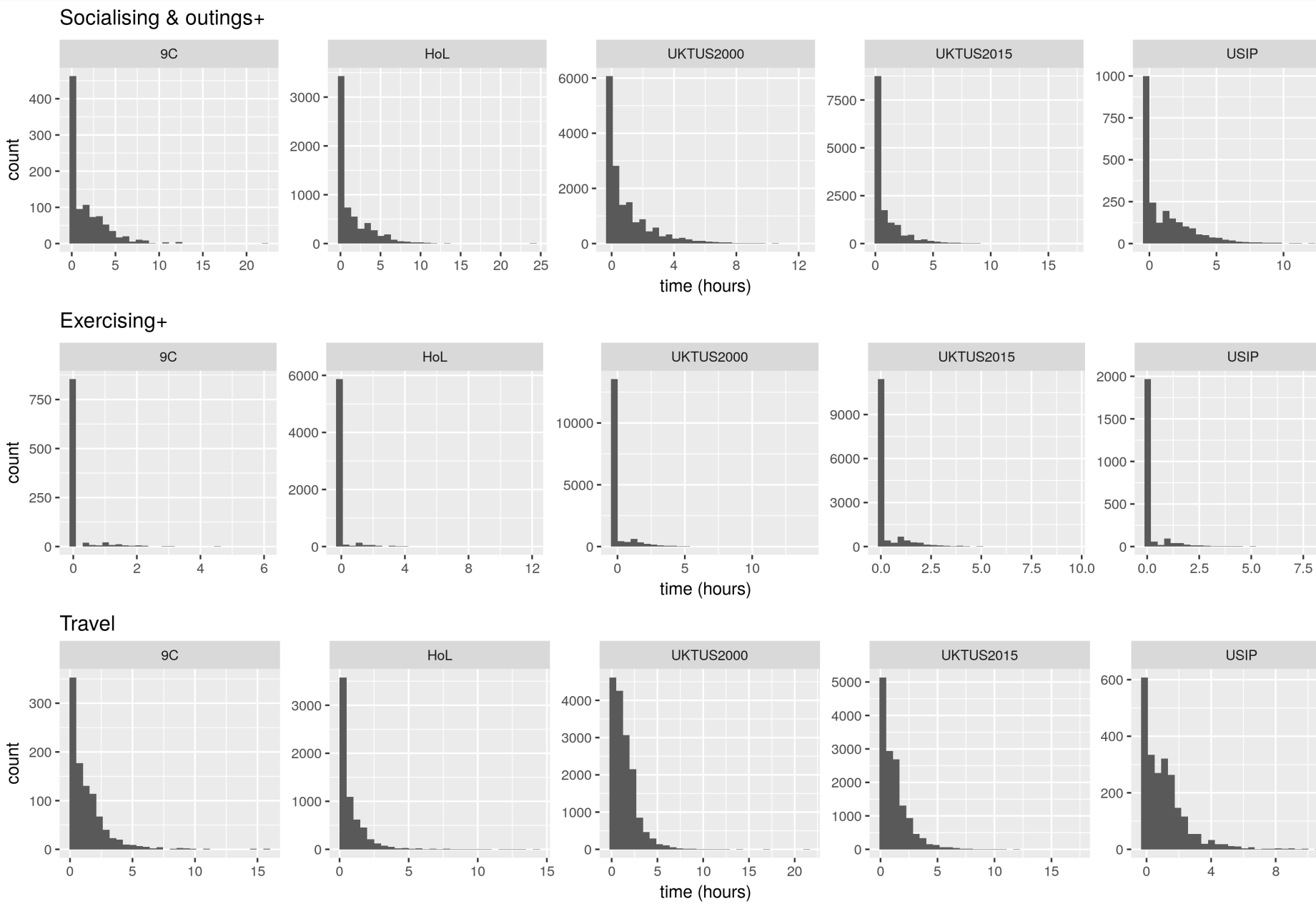
1. Socio-demographic ie sex, age, marital status whether in paid work, education
2. Stylised indicator ie How often do you...?
    - Cultural consumption
    - Volunteering/helping
    - Leisure-time physical activity
    - Did you do it last week? (UK TUS 2015)
3. Saturated models

# We need to talk about distributions

- Skewed, positive, zero-inflated
- Count based model but...
  - Overdispersed ($\sigma^2 > \mu$)
- Negative binomial distribution
  - Additional term $\alpha$ accounting for over dispersion
  - Logit link
- Two flavours:
  - Zero truncated vs zero inflated
  - Assumption about p
    - $wp_i^a = 1 - (1 - p_i^a)^7)$
  - Zeros are produced by a mixture distribution
    - Natural vs structural


Sleep, rest & self-care
Paid work/study
Caring
Hwork, cooking & DIY
9C
HoL
UKTUS2000
UKTUS2015
USIP
count
time (hours)


Eating at home
9C
HoL
UKTUS2000
UKTUS2015
USIP
count
time (hours)
Volunteering+
9C
HoL
UKTUS2000
UKTUS2015
USIP
count
time (hours)
Shopping
9C
HoL
UKTUS2000
UKTUS2015
USIP
count
time (hours)
Culture & leisure+
9C
HoL
UKTUS2000
UKTUS2015
USIP
count
time (hours)


Socialising & outings+
9C
HoL
UKTUS2000
UKTUS2015
USIP
count
time (hours)
Exercising+
9C
HoL
UKTUS2000
UKTUS2015
USIP
count
time (hours)
Travel
9C
HoL
UKTUS2000
UKTUS2015
USIP
count
time (hours)

# Mean daily probabilities of participation in activities

| | 9C | HoL | UKTUS2000 | UKTUS2015 | USIP |
|---|---|---|---|---|---|
| Sleep, rest, self-care | 1 | 1 | 1 | 1 | 1 |
| Paid work,study | 0.42 | 0.47 | 0.37 | 0.37 | 0.31 |
| Caring | 0.13 | 0.22 | 0.32 | 0.30 | 0.19 |
| Hwork, cooking, DIY | 0.67 | 0.74 | 0.89 | 0.90 | 0.86 |
| Eating at home | 0.84 | 0.91 | 0.98 | 0.96 | 0.92 |
| Volunteering+ | 0.06 | 0.06 | 0.07 | 0.08 | 0.09 |
| Shopping | 0.37 | 0.33 | 0.45 | 0.46 | 0.41 |
| Culture & leisure+ | 0.87 | 0.91 | 0.94 | 0.93 | 0.86 |
| Socialising & outings+ | 0.58 | 0.56 | 0.62 | 0.56 | 0.57 |
| Exercising+ | 0.12 | 0.07 | 0.16 | 0.22 | 0.15 |
| Travel | 0.75 | 0.59 | 0.87 | 0.80 | 0.74 |

**Data:** [1] Nine Country Project, Home Online Project, 2000 and 2015 UK Time Use surveys, Understanding Society Innovation Panel Wave G

# Observed mean daily time in activities

| | 9C | HoL | UKTUS2000 | UKTUS2015 | USIP |
|---|---|---|---|---|---|
| Sleep, rest, self-care | 604 | 587 | 595 | 620 | 665 |
| Paid work,study | 431 | 434 | 419 | 397 | 407 |
| Caring | 241 | 176 | 99 | 109 | 174 |
| Hwork, cooking, DIY | 122 | 151 | 167 | 152 | 155 |
| Eating at home | 79 | 71 | 91 | 93 | 76 |
| Volunteering+ | 149 | 152 | 113 | 111 | 125 |
| Shopping | 98 | 109 | 72 | 72 | 84 |
| Culture & leisure+ | 271 | 266 | 227 | 242 | 242 |
| Socialising & outings+ | 171 | 163 | 102 | 94 | 137 |
| Exercising+ | 92 | 106 | 102 | 88 | 88 |
| Travel | 110 | 85 | 97 | 92 | 93 |

**Data:** [1] Nine Country Project, Home Online, 2000 and 2015 UK Time Use surveys, Understanding Society Innovation Panel Wave G

# Predicted weekly duration of activities, Home Online

| | *Observed* | OLS | Poisson | ZT NB | ZI NB |
|---|---|---|---|---|---|
| Sleep, rest, self-care | *4070* | 4074 | 4074 | 4074 | 4074 |
| Paid work,study | *1491* | 1622 | 1622 | 1622 | 1487 |
| Caring | *257* | 102 | 407 | 491 | 244 |
| Hwork, cooking, DIY | *771* | 1075 | 945 | 1152 | 774 |
| Eating at home | *443* | 463 | 479 | 512 | 460 |
| Volunteering+ | *61* | 178 | 332 | 556 | 61 |
| Shopping | *250* | 289 | 605 | 620 | 252 |
| Culture & leisure+ | *1659* | 1706 | 1801 | 1940 | 1671 |
| Socialising & outings+ | *653* | 1121 | 921 | 1131 | 645 |
| Exercising+ | *56* | 255 | 210 | 415 | 57 |
| Travel | *369* | 519 | 537 | 705 | 374 |
| Total | *10080* | 11404 | 11933 | 13218 | 10099 |

# Predicted weekly duration of activities

| | 9C | HoL | TUS 2000 | TUS 2015 | USIP |
|---|---|---|---|---|---|
| Sleep, rest, self-care | 4150 | 4074 | 4149 | 4351 | 4686 |
| Paid work,study | 1278 | 1487 | 1351 | 1248 | 1065 |
| Caring | 213 | 244 | 228 | 222 | 237 |
| Hwork, cooking, DIY | 545 | 774 | 999 | 924 | 895 |
| Eating at home | 465 | 460 | 602 | 611 | 488 |
| Volunteering+ | 63 | 61 | 54 | 61 | 73 |
| Shopping | 276 | 252 | 221 | 222 | 242 |
| Culture & leisure+ | 1700 | 1671 | 1438 | 1503 | 1409 |
| Socialising | 703 | 645 | 394 | 338 | 483 |
| Exercising+ | 88 | 57 | 99 | 128 | 94 |
| Travel | 593 | 374 | 601 | 527 | 493 |
| Total | 10074 | 10101 | 10137 | 10135 | 10164 |

Zero-inflated negative binomial-predictions weighted for non response Data: 9 Country Project, Home Online, 2000 and 2015 UK Time Use surveys, Understanding Society

# Conclusion

- Caveat: clustering of observations
- Negative binomial: a promising approach
- Empirical vs theory-driven probabilities
- Need to test the approach on more datasets
- Longer term predictions
- Next application: working time
- Synthetic dataset using day of the week and month?