What is Event History Analysis?

Methods @ Manchester
11th of March 2010

Dr. Vanessa Gash, CCSR/Social Statistics

Introduction

- Aim to offer a broad overview of event history analysis (EHA).
- I will introduce the key concepts behind the analysis of change in events.
- I hope to finish the talk with a practical example of research that applies EHA.

- In a nutshell: EHA is a technique that allows researchers to study the social processes that lead to the occurrence of *an event*.
- Event is a ***change*** from one state to another and is measured as a categorical/discrete dependent variable.

- Origins in the bio-medical sphere
- Different terminologies for the one statistical technique:

Event history analysis, survival analysis, duration analysis, failure time analysis and hazard analysis…

- concerns the change in $y(t)$, or $y(it)$
- change within $y(t)$ also termed a transition

Examples of changes in Y(it):

- Transitions across labour market status, from unemployed to employment
- Transitions out of marriage to divorce/separation
- Transitions from poverty to financial security
- From one occupational or social grouping to another..
- From happiness to unhappiness
- Recidivism, what factors predict further criminality?

- What is crucial here is the timing..

- Unit of analysis need not be the individual, could be: household, country, firm..etc
- The different values of y(t) are termed its state space.
- Careful consideration required in estimation of the timing of an event transition, conceptually and practically.

- Basic EHA can be done with info on the state space of y(t) and the timing of the event however, in practice, a common aim of EHA is to approximate a causal model of outcome.

- In particular researchers using EHA tend to be interested in the impact of key covariates on the risk of a transition.

Analysing change, data structure

- Format of the longitudinal dataset.

- Best research design for EHA (according to Paul Allison) is prospective.

- Many EHA done on retrospective data.

Limitations of retrospective data

- recall error
- recall bias (social desirability)
- Recall accuracy tends to be high for marital and fertility history
- Recall tends to be low for attitudes and values, unemployment or layoff, and subjective indicators (Dex 1991).

Analysing change, data structure

Examples of short-term and long-term retrospective questionnaires

- SHARE (SHARELIFE RETROSPECTIVE SURVEY)

http://www.share-project.org/sharelife/

- BHPS

http://www.iser.essex.ac.uk/survey/bhps

- Other examples can be found at:
http://www.esds.ac.uk/longitudinal/Introduction.asp

Analysing change, data structure

- SHARELIFE focuses on people's life histories.
- 30,000 people across 14 European countries took part.
- The respondents are representative for the European population aged 50+ in:

  Denmark, Sweden

  Austria, France, Germany, Switzerland, Belgium, and the Netherlands

  Spain, Italy and Greece

  Czech Republic and Poland.

Analysing change, data structure

Respondents shown a ‘Life History Calendar’ on a screen that lists all the years of the respondent’s life, from birth to the present.

Information on event that occurred in these years is then collected of respondents.

There is a row for each of the different areas of the respondents life.

The calendar can search for national and world events that have occured during the respondent’s life (which may help them determine better when other events in their life happened).

Analysing change, data structure

In the BHPS, as in many employment focused datasets, retrospective data is collected on the different economic activities of the previous year.

Analysing change, data structure

An EHA requires the following 3/4 variables:

1- Identifier variable. Might be a personal id variable identifying the individual in a panel dataset or it could be an id of households.

2- Categorical variable measuring the changes in the event of interest overtime: Y(t)

3- Time variables that identify when a transition occurs for your unit of analysis

4- Covariates that may or may not change with time, used as explanatory factors of the transition being analysed.

Analysing change, data structure

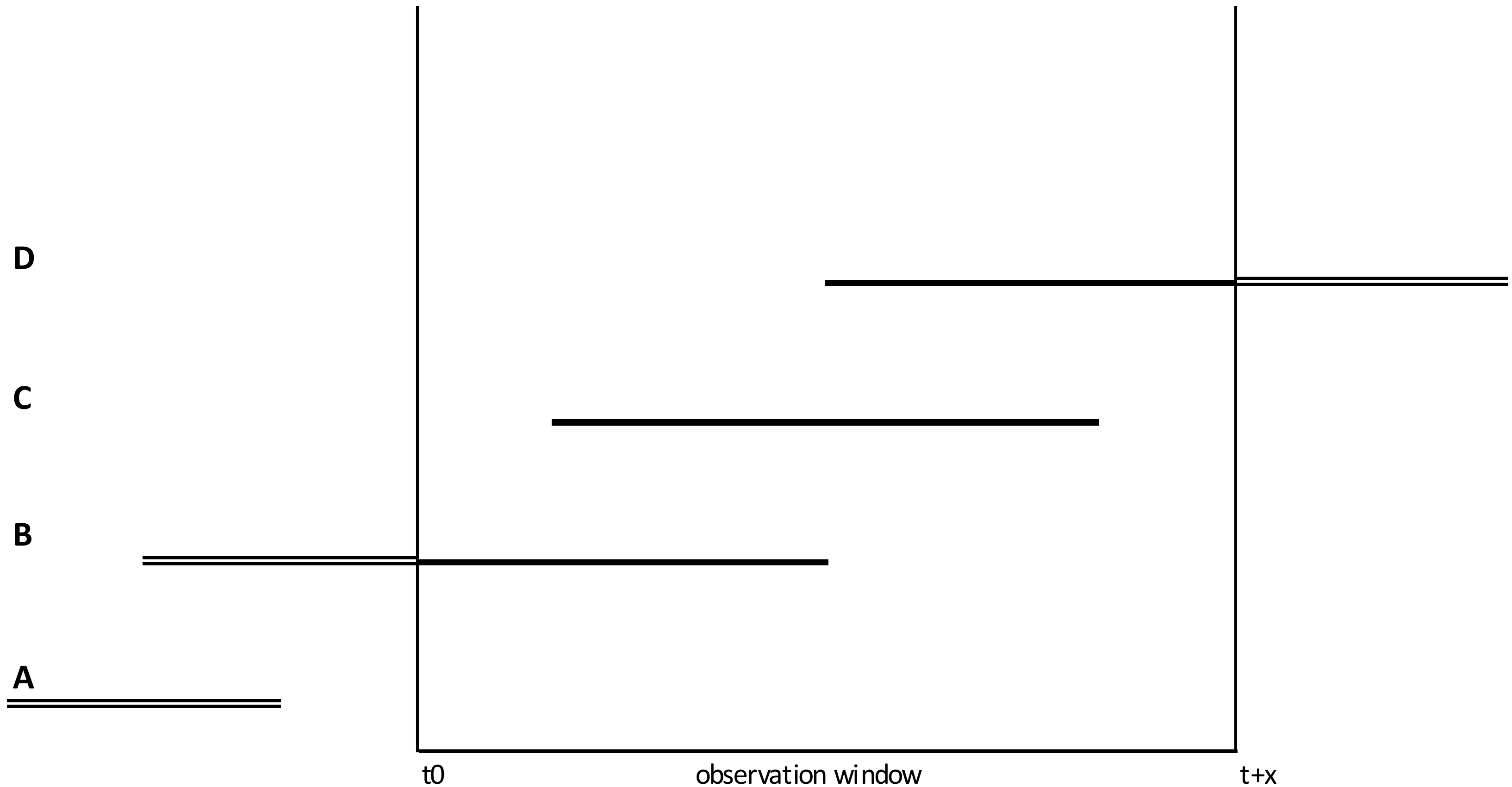


Figure 1 shows different types of censoring in the observation window t0 to t+x

Analysing change, data structure

Episode A is fully censored on the left. Worst case scenario.

Episode B is censored only on the left. We don't know when the event started, but we know they are in state j at t0 and onwards.

Episode C, is our ideal type, having complete information.

Episode D, is right-censored, and is easy to deal with in EHA.

Analysing change, data structure

Informative versus non-informative random censoring.

While EHA can deal with censored data (except for Episode A in the previous example) in principle the censoring should be non-informative.

Non-informative censorship: if individual A is censored at t+2 because they drop out of the sample, individual A should ideally be similar to individual B who does not drop out of the sample.

What does a *prepared* dataset look like?

Example of a single episode event history dataset

pid	sex	employment status	Date began	Date Ended	Event
1	female	1	1991	1992	1
2	female	1	1991	1995	1
3	female	2	1991	2001	1
4	female	3	1991	2002	0
5	female	1	1991	1997	0
6	female	2	1991	1993	1
7	female	3	1991	1999	1
8	female	1	1991	2000	1
9	female	2	1991	2002	0
10	female	1	1991	2002	0

Analysing change, data structure

Example of a multi-episode event history dataset

pid	sex	mlstat	employment status	Date began	Date Ended	Event 1	Event 2
1	female	never ma	Temp	1991	1992		
2	male	never ma	Ed	1992	1995		
2	male	never ma	Temp	1995	2001		
2	male	never ma	Temp	2001	2002		
3	female	widowed	Perm	1991	1991		
3	female	widowed	Unemp	1992	1992		
3	female	widowed	Retired	1993	1993		
4	female	widowed	Perm	1991	1997		
4	female	widowed	Temp	1997	2002		
5	female	married	Perm	1991	2002		

Strengths of EHA over standard regression

- Important to recognise why you are using EHA models, rather than linear regression for instance.
- In the past not uncommon for duration type data to be analysed using standard linear regression techniques.
- Problem lies with censored data, and while previously people included a dichotomous variable in a regression to indicate censorship, this would no longer get past peer review.

- Strengths of EHA
  - Ability to examine the underlying causal mechanisms behind event occurrence
  - One can control for censored data
  - One can examine the impact of time varying covariates on outcome.

Basic statistical concepts behind EHA

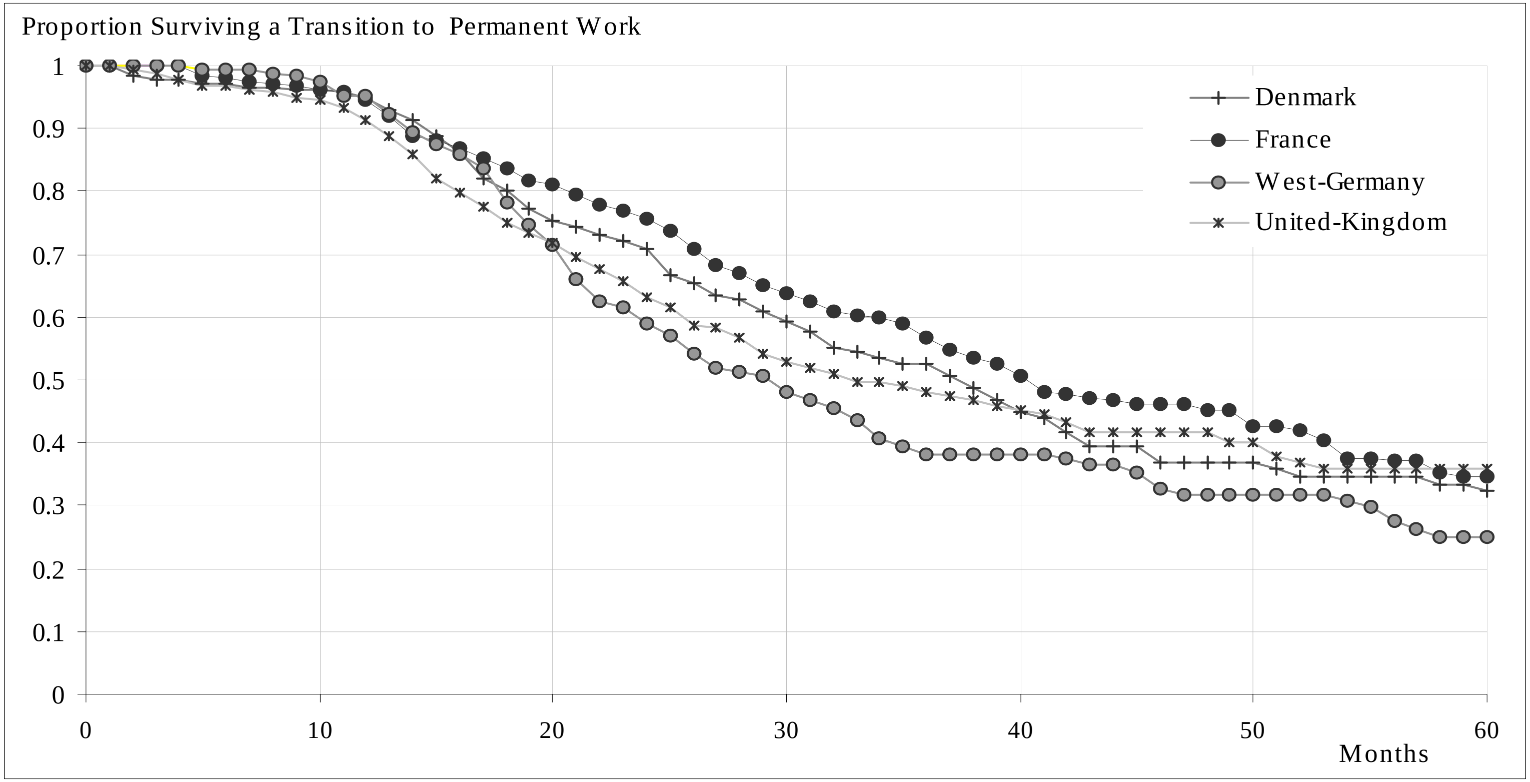
- EHA statistical methods analyse the occurrence of events.
- An event can be defined as a qualitative change in the unit of analysis from state **j** to **k**.
- An example would be: Examining exits from unemployment, an event in this instance could be the transition from unemployment (state j) to employment (state k).

- All the standard approaches to EHA are probabilistic or stochastic: the times that an event occurs is assumed to be the outcome of a random process.
- Therefore Y(t) is a random variable with a probability distribution.
- The assumed distribution of Y(t) is absolutely central to your model choice in EHA.

Probability distributions

Figure showing differences in the survivor function of fixed-term contract workers from different countries (DK, FR, West-Germany and the UK).

Cox regression based Test for the Equality of survival Curves confirmed the statistically sig. diff of entering perm work by country.



Probability distributions

Table showing differences in the survivor function of fixed-term contract workers.

	DENMARK	FRANCE	WEST GERMANY	UNITED-KINGDOM
	Transition to Permanent Contract Employment			
Time(months)	**S(f)a**	**S(f)a**	**S(f)a**	**S(f)a**
0	1.000	1.000	1.000	1.000
12	0.9474	0.9454	0.9534	0.9144
36	0.5245	0.5684	0.3829	0.4808
48	0.3689	0.4511	0.3186	0.4172

Probability distributions

Key statistical concept in EHA: The Hazard Function

The Hazard function aims to quantify the probability that an event will occur in the small interval between t and a change (Δ) in t , conditional on the individual surviving to time t.

Once the person has made a transition from **state j** to **state k** they are no longer at risk, so need to be removed from our calculation of the hazard of the event occurring.

$$h(t)jk = \lim_{\Delta t \to 0} \frac{Prob\ (t,\ t + \Delta t)}{\Delta t}$$

- Useful to think of the hazard function as a property of individuals.

Probability distributions

The Hazard Function, practical conceptualisation

Suppose we have a sample of 1,000 single women, that we observe over a period of a month, and after one month, 25 women enter a relationship.

The time constant hazard for the women of entering a relationship is: 25/1000= 0.025

And the reciprocal, the prob. of it not occurring is: 1/0.025 = 40. That is the amount of time the average woman can expect to remain single is 40 months.

Another example:

Study of how cyclists risks of injury varies overtime.

We start with a sample of 500 cyclists (our **risk set**) and observe that 33 of them get injured over a two year period.

We interview them first in January, and interview them at 6 month intervals to establish whether any of them have been injured whilst cycling.

Probability distributions

3- The Hazard Function

months	N of cyclists who get injured	N of cyclists at risk of injury	Estimated Risk of Injury
0	0	500	0.000
6	3	500	0.006
12	11	497	0.022
18	5	486	0.010
24	14	481	0.029
Total injured	33		

Note:

The denominator changes to the N of cyclists ***who remain at risk*** of injury at each time point.

What happens to the risk of injury? Does it increase? Is it constant overtime?

- **RECOMMENDED FURTHER READINGS:**

- **Allison, P.D. 1984. *Event History Analysis: Regression for Longitudinal Event Data*. Beverley Hills: Sage.
- Blossfeld, H-P, A. Hamerle, and K. U. Mayer. 1989. *Event History Analysis*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Blossfeld, H-P and G. Rohwer. 2002. *Techniques of Event History Modelling: New Approaches to Causal Analysis, 2nd Edition*. Mawah, NJ: Lawrence Erlbaum Associates
- **Blossfeld, H-P, Golsch, K. and G. Rohwer. 2007. *Event History Analysis with Stata.* Mawah, NJ: Lawrence Erlbaum Associates
- **Castilla, E. (2007) *Dynamic Analysis in the Social Sciences.* London: Elsevier
- Cleves, M., W. W. Gould, and R. Gutierrez. 2004. *An Introduction to Survival Analysis Using Stata, Revised edition*. College Station, Texas: Stata Press.
- Jenkins, S. (1995) ‘Easy Estimation Methods for Discrete-Time Duration Models’, *Oxford Bulletin of Economics and Statistics*. 57 (1) 129-138.
- Halpin, B. 1998. "Unified BHPS work-life histories : Combining multiple sources into a user-friendly format," *Bulletin de Methodologie Sociologique* 60: 34-79.
- Tuma, N. 1994. “Event History Analysis”. In *Analysing Social and Political Change* Edited by A. Dale and Richard B. Davies. London: Sage.

Practical Example

- From: Cooke, L.P and Gash, V (forthcoming) "*Wives' Part-time Employment and Marital Stability in Great Britain, West Germany and the United States*". Sociology.

Principle RQ: Does part-time work enhance marital stability?

Do different institutional contexts vary female employment effects on divorce risk?

We select couples as they enter into first marriage (1985-1995) for GSOEP and PSID and (1992-2000) for the BHPS and then follow them until 1997 (PSID), 2000 (GSOEP) and 2007 (BHPS).

These criteria yield an analytic sample of:

- West Germany 559 couples, 4473 couple-years
- UK 666 couples, 4174 couple-years
- US 502 couples, 2535 couple-years

Theories Concerning Divorce Risk

- **H1: Independence hypothesis:** predicts **wife's employment will increase the risk of divorce** (Becker et al., 1977). Wives with an independent income have less incentive to work out marital problems.
- H1b: In all countries we should observe the **smallest divorce risk** among couples where the wife remains out of the labour force, a somewhat **greater risk** among couples where she works part-time, and the **greatest risk** among full-time dual-earner couples.

- **H2. Flexibility hypothesis: wives' employment (over more recent time periods) should predict greater marital stability**. As gendered division of paid and unpaid labour within the family is a high-risk strategy leaving households vulnerable to economic downturns Oppenheimer (1997).
- H2b. In all countries we should observe the **smallest divorce risk** among couples where the wife remains in full-time employment, a somewhat **lower risk** among couples where she works part-time, and the **greatest risk** in couples where the wife is economically inactive.

Recent Findings

- Women’s employment hours increases divorce risk among Dutch (Kalmijn et al. 2004), Swedish (Henz & Jonsson 2003), Hungarian (Bukodi & Robert 2003) and UK (Chan and Halpin 2002) couples.

- Cooke (2004) finds among a recent cohort of West German first-married couples, women’s employment hours only significantly increase divorce risk after birth of first child.

- Analysis of US mixed (Rogers 2004), with most recent evidence indicating the risk of employment associated with wives’ employment hours off-set when husbands increase their share of housework (Cooke 2006).

Divorce Risk: **Country Specific Hypotheses**

West Germany

Part-time predictive of more stable marriages, possibly stronger effect than in UK due to higher quality of part-time employment

United Kingdom

Part-time predictive of more stable marriages, but poor job quality might result in more modest effect than in W. Germany

United States

Strongest market pressures; part-time less desirable

Dependent variable: whether divorce occurs

Independent variables (all lagged by one time period):

woman's working time

full-time >30 hours a week

part-time <30 hours a week (reference housewife/inactive)

Interaction of part-time*children

Controls include: women's % of household earnings, household income, women's age at marriage, his and her education, children, and years since married.

Table 2 Relative risk of divorce from year of marriage in Great Britain, West Germany, and the United States

	West Germany				Great Britain				United States			
	Model 1		Model 2		Model 1		Model 2		Model 1		Model 2	
	Odds Ratio	*RSE*	*Odds Ratio*	*RSE*	*Odds Ratio*	*RSE*	*Odds Ratio*	*RSE*	*Odds Ratio*	*RSE*	*Odds Ratio*	*RSE*
Wife works part-time (<=30)	0.57*	0.15	0.50	0.29	0.74	0.27	0.85	0.31	1.38	0.50	3.34**	1.56
Wife works full-time (>30)	0.66	0.19	0.66	0.19	0.94	0.45	0.95	0.44	1.04	0.45	1.26	0.53
Ref: housewife, out of labour force												
Wife's per cent couple earnings	1.01+	0.01	1.01+	0.01	1.01	0.01	1.01	0.01	1.00	0.01	1.00	0.01
Husband non-employed	1.47+	0.34	1.47	0.34	2.74**	0.99	2.73**	1.14	1.71	1.03	1.63	0.95
Wife with tertiary education	0.59	0.21	0.59	0.21	0.56	0.22	0.56	0.26	1.17	0.29	1.16	0.30
Husband with tertiary education	0.71	0.17	0.71	0.18	0.40*	0.16	0.40*	0.17	0.39**	0.12	0.39**	0.12
Children (*0 = none*)	0.27***	0.05	0.27***	0.05	1.18	0.32	1.21	0.34	0.42***	0.11	0.69	0.20
Children*women part-time			1.17	0.72			0.85	0.61			0.24**	0.12
Log of total household income	0.80***	0.04	0.80***	0.04	1.10	0.28	1.10	0.28	0.79	0.12	0.79	0.11
Wife's age at marriage	1.00	0.02	1.00	0.02	0.94*	0.02	0.94*	0.03	0.92*	0.04	0.92*	0.04
Years since marriage	1.24**	0.10	1.24**	0.10	0.96	0.09	0.96	0.09	1.69**	0.29	1.78***	0.32
(Years since marriage)^2	0.99**	0.01	0.99**	0.01	0.99	0.01	0.99	0.01	0.96*	0.02	0.96**	0.02
Pseudo log-likelihood	*-777.65*		*-777.61*		*-418.18*		*-418.14*		*-357.93*		*-353.33*	
Wald chi-square	*96.40****		*96.64****		*33.55****		*33.45****		*56.36****		*61.86****	
n couple-years (couples)	*4,473 (559)*		*4,473 (559)*		*4,174 (666)*		*4,174 (666)*		*2,535 (502)*		*2,535 (502)*	

*$p < .05$. **$p < .01$. ***$p < .001$. (two-tailed tests)

Conclusions

Germany: Wives‘ part-time employment predicts lower risk of divorce

The UK: No effect of wives employment on risk of divorce

The US: Mothers‘ part-time employment significantly lowers divorce risk

Death of the Male Breadwinner?

Nowhere does specialization significantly enhance marital stability.

However, gender specialisation is still strong.

In DE and the UK, men who became unemployed had higher divorce risks.