



Max-Planck-Institut für demografische Forschung  
Max Planck Institute for Demographic Research  
Konrad-Zuse-Strasse 1 · D-18057 Rostock · GERMANY  
Tel +49 (0) 3 81 20 81 - 0; Fax +49 (0) 3 81 20 81 - 202;  
<http://www.demogr.mpg.de>

---

MPIDR WORKING PAPER WP 2013-006  
MAY 2013

**The dangers of conditioning on the time  
of occurrence of one demographic  
process in the analysis of another**

Jan M. Hoem ([jan.hoem@sociology.su.se](mailto:jan.hoem@sociology.su.se))

---

© Copyright is held by the author.

Working papers of the Max Planck Institute for Demographic Research receive only limited review. Views or opinions expressed in working papers are attributable to the authors and do not necessarily reflect those of the Institute.

**Author:**

Jan M. Hoem  
Professor Emeritus of Demometry,  
Demography Unit, Stockholm University, S-106 91 Stockholm.

**The dangers of conditioning on the time of occurrence of one demographic process  
in the analysis of another**

Words: 6844.

Short title: Conditioning on one out of two interacting behavioural processes.

Keywords: Parallel processes; migration and childbearing; marriage formation and childbearing; marriage formation and home purchase; estimation bias produced by conditioning; observational design in event-history analysis.

## **Abstract**

Demographers and others often study the interaction between two types of individual-level behaviour, such as migration and childbearing. Unfortunately, one can get estimation bias if one compares childbearing before and after migration from data confined to migrants, say, as is sometimes done to see whether international migration disrupts fertility. Similar issues can arise in comparisons of union formation before and after first birth, marriage formation before and after home purchase, as well as in other comparisons of behaviour before and after an index event if one confines the study only to those who have experienced the index event. We point out that if one does this, then the study of behaviour before the index event actually conditions on the later arrival of the index event. This amounts to anticipatory analysis, which normally produces an estimation bias. In this paper we discuss this issue, provide a mathematical and graphical representation of it, and show how one can avoid estimation bias and get a meaningful (unconditional) comparison of behaviour before and after the index event provided the data contain enough information for both sub-periods.

## **1. Introduction**

Demographers often study individual-level processes that develop in parallel and in interaction with each other. Some such pairs of interacting processes are migration and childbearing, marriage formation and childbearing, and marriage formation and home purchase; one can easily think of many more examples. In this paper we address the analytic danger involved in comparing fertility before and after geographic migration, marriage formation before and after entry into motherhood, marriage formation before and after the purchase of a home, and potential similar issues in mortality research. Such a conditional before-and-after analysis can be problematic, for there is a fundamental asymmetry between the analysis for the time *before* and the analysis for the time *after* the occurrence of an index

event when data are confined to those who experience that event. To be concrete we use the connection between childbearing and migration as our main vehicle of presentation, and show in particular that the analysis of childbearing shortly *before* an observed case of migration will underestimate the fertility of migrants in the sending population if the presence of a child hampers out-migration, while by contrast an assessment of the corresponding rate *after* migration correctly estimates the fertility of in-migrants in the population at the destination (unless some other problem interferes). The intensity estimate for the time *before* the index event turns out to be biased downwards in this case, but in other cases the bias is the reverse. For instance, the estimated rate of marriage formation shortly before a home purchase will be biased upward if only home-buyers are considered and if the home-purchase intensity is higher after marriage formation than before it. Only if the transition rates on the second behaviour (migration, say) are the same before and after the behaviour in focus (childbearing) can one be sure in general that there will be no such bias. Unfortunately one can rarely know or demonstrate whether a non-differentiability of this nature is present in the data.

To handle general issues of this kind we employ a device that is now standard in several disciplines, and that represent individual behaviour as transitions between discrete states in a time-continuous Markov chain, such as the transition from childlessness to motherhood. We have already used the name “intensity” for a theoretical rate of transition between two such states and thus have a brief and neutral term to denote what in the general theory of event-history analysis is often called the “instantaneous transition rate” (a name which is longer) or “hazard” or “risk” (which sound threatening).

Studying behaviour (say childbearing) before and after an event on a different process (such as migration) involves conditioning on a time of occurrence on the second process. With such conditioning, time is centred on the index event (the geographical move) and the intensity of the other kind of event (childbearing) is estimated for times before and after the index event, perhaps counting occurrence times formally as negative when they occur *before* the index event and positive when they

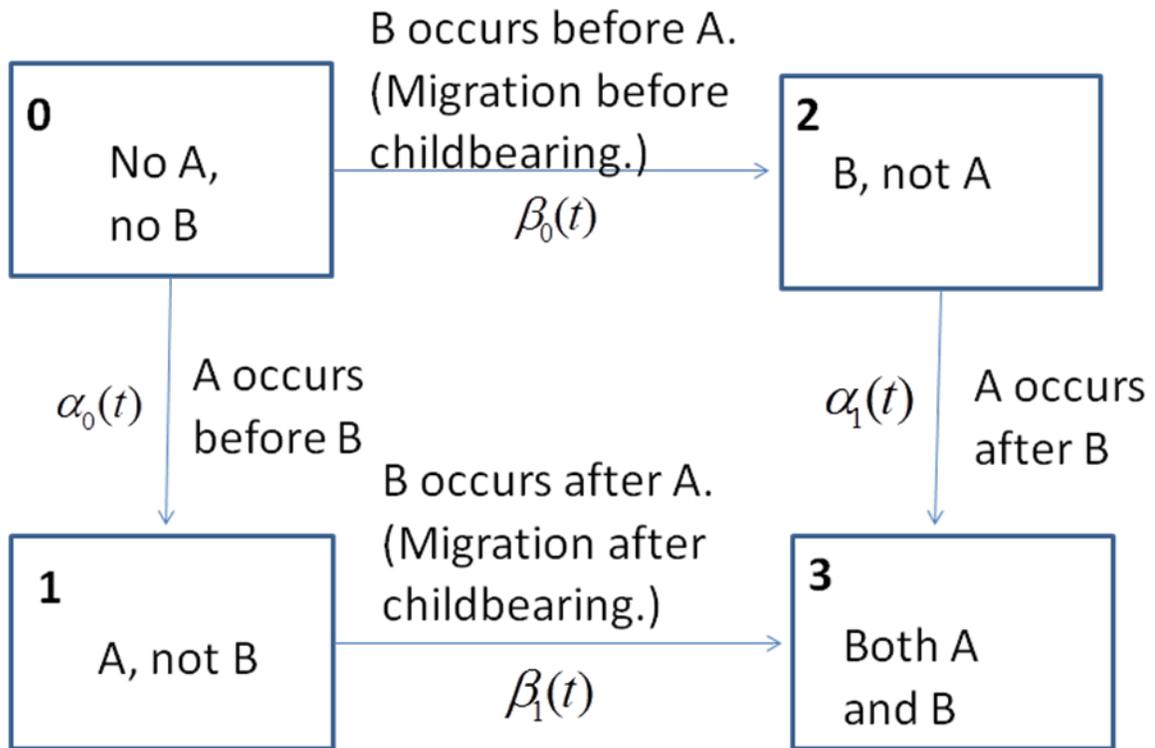
appear *after* it, with the time of occurrence of the index event as the origin (zero-point) for a conditional pre-and-post-event analysis. Anticipating later generalization we shall call the behaviour to be analyzed ‘Process A’ (childbearing) and use the name ‘Process B’ for the behaviour on which one conditions (migration). The two processes play asymmetric roles in the statistical theory. [If it is hard to remember the respective roles of the two processes, it may help the user who knows Scandinavian or German to keep track of them if we note that B is the first letter in the verb ‘betinge/beting/bedingen’, which means to ‘condition’. A is of course the first letter in the word ‘analyze/analysere/analysieren’. These mnemotechnical rules have helped us during the work with this paper.] The question is how comparing childbearing intensities before and after migration in individual-level data confined to immigrants can produce a bias, and what the bias is. The purpose of the present paper is to warn against pitfalls involved in analyses of this nature, which often are motivated by the consideration of plans or intentions. We address the question whether one can safely disregard the biases if, for example, people buy a home *because* they intend to marry, if they marry *because* they intend to have children, or if they change their criminal behaviour *because* they intend to marry. Correspondingly, one may ask whether the intention to have a child after migration removes any bias in the fertility estimate for periods before migration. Unfortunately, introducing intentionality does not provide a solution to the biasing issue. While intentions may play an important role in the behaviour observed, it turns out that they cannot be used to remove estimation bias unless one possibly has direct individual-level information about intentions. Of course planning ahead is part of human nature, but considerable care must be taken to avoid the pitfalls of anticipatory analysis in the absence of real information about intentions because the analyst cannot distinguish between estimation biases and real behaviour.

The present paper contributes to the already rich literature on observational design in event-history analysis; see for instance Aalen et al. (1980), Courgeau and Lelièvre (1992, Chapters 5-6), Andersen et al. (1993), Blossfeld and Rohwer (2002, Chapter 6), and just recently Keiding (2013). In this tradition, Hoem and Kreyenfeld (2006) may be among the authors who most strictly have rejected

the practice of including unmeasured plans and unmeasured intentions to motivate conditional pre-and-post-event analysis. In the present paper we spell out the reasons why in a novel manner and illuminate them with further examples. We add to the design literature by warning against conditioning on one behaviour in the analysis of another. We are not aware that anyone has addressed these issues in any generality before.

## 2. Graphical representation

Following a tradition going back to Fix and Neyman (1951) and Sverdrup (1965), we represent the two behaviours that act in interaction in a simplified form by the transition diagram in Figure 1, where as usual rectangles stand for statuses and arrows represent transitions between them. Let us concentrate on the first birth (Event A). The first-birth intensity at age  $t$  is  $\alpha_0(t)$  in the region of origin and is  $\alpha_1(t)$  for the same type of individual in the region of destination (after the migration). Correspondingly  $\beta_0(t)$  is the out-migration intensity at age  $t$  for a childless individual (in the region of origin) and  $\beta_1(t)$  is the corresponding out-migration intensity for a similar parent (in the same region). For clarity we have written the intensities beside the transitions that they are associated with in Figure 1. We allow the intensities for each process to depend on what has happened in the other process up to the current moment (age  $t$ ), in the sense that each intensity may change character after the occurrence of an event on the other process. Comparing  $\alpha_0(t)$  and  $\alpha_1(t)$  is a solution to the issue of comparison of childbearing before and after (potential) migration.



**Figure 1. Two processes developing in interaction.**

← **Figure 1 in about here** →

The classical way of collecting data to analyze the intensities is of course to follow a sample of individuals from some starting time and to observe all transitions between states in the state space until censoring, allowing all transitions to be possible. Normally such a procedure allows the analyst to study childbearing before and after migration without distortion by comparing (estimates of)  $\alpha_0$  and  $\alpha_1$ , as was done by Kulu (2005, 2006), Nedoluzhko and Andersson (2007), Hoem and Nedoluzhko (2008), and surely many others. It is unproblematic in principle to collect the data and carry out such a before-and-after comparison if the populations of origin and destination are parts of the same overarching population and the migration is internal to the latter. However, complete observation may be impossible if the individuals live in a different population before than after migration, as will normally be the case for international migration. It is then tempting to base one's analysis on the data for in-migrants exclusively and to compare childbearing information for the migrants before and after they

migrate. With this observational scheme, childbearing information is only available conditional on the occurrence of migration by the same individuals. The pool of *potential* migrants will include both those who eventually migrated and those who did not. If migration and childbearing are interrelated, then in general there must be a problem with an analysis of childbearing solely among those who did migrate, for this cannot provide a portrait of the fertility of those who could have migrated, but did not. To spell it out, suppose that the childbearing counts and exposures (before and after migration) are only available for individuals who actually experience migration; for individuals who do not, childbearing occurrences and exposures are not observed. Estimating pre-migration fertility from data for the migrants alone must therefore produce fertility estimates that are lower than those in the overall pool of potential migrants if those who have given birth are the less likely to migrate. Below we display a formula which in principle makes it possible to assess how much lower the estimates will be (Formula 3 in Section 4).

Essentially, studying fertility in the sending population from data for out-migrants only is an event-history version of conditioning on the outcome (here: migration), a procedure that practitioners usually are advised to handle with great care in any type of analysis and normally to avoid if possible. Our purpose is to address issues concerning the analysis of such conditional data.

It is tempting (and apparently is quite common practice, as the literature for our examples below shows) to treat the time of the conditioning event as an origin (“time zero”). Times that appear before the origin are naturally taken to be negative and times that appear after the origin are taken to be positive. We will show

(a) that for periods *after* the conditional origin, intensity estimators of childbearing correctly provide unbiased estimates of the childbearing intensity  $\alpha_1(t)$ ,

(b) while for periods *before* but close to the conditional origin, the rates are estimators of a quantity  $\alpha_0(t)\beta_1(t)/\beta_0(t)$ , which in general may differ from  $\alpha_0(t)$ .

Item (b) shows that there is little distortion of  $\alpha_0(t)$  at times shortly before migration if the migration intensity is largely independent of whether the first birth has occurred or not (in which case  $\beta_0(t) \approx \beta_1(t)$ ). Otherwise one must count on the possibility that the conditioning procedure may produce systematic distortion in the estimates of  $\alpha_0(t)$  produced for the time shortly before migration, as in Formula (5) in Section 4 below. In particular, if the out-migration intensity is lower for parents than for non-parents, such computations will underestimate the intensity  $\alpha_0(t)$  at times shortly before migration, since the estimate of  $\alpha_0(t)$  is systematically smaller than the true value  $\alpha_0(t)$ . If parenthood hampers out-migration (i.e., if  $\beta_1(t) < \beta_0(t)$ ), this will be the case even if real fertility is the same in the regions of origin and destination for the subpopulation observed.

To facilitate reading, mathematical proofs of these statements are deferred to Section 4. In that general theory, our statements about the situation shortly before the conditioning event will show up as limits of estimation biases as process time approaches the time of occurrence of the conditioning event from below. In Formula (3) in the general theory in Section 4, we will display the estimation bias further back from the migratory event as well. In this argument we have assumed implicitly that all individuals are subject to the same set of intensities. This is of course a simplification, and we will consider various alternative specifications as we move along in our account. Before we get to that, however, we note that the same kind of issues arises whenever two interacting demographic behaviours or processes are studied in a conditional before-and-after schema. Section 3 contains a number of motivating illustrations, which we have already hinted at in our introduction.

Issues of the same nature arise in any setting where interacting behaviours are studied.

### **3. Illuminating examples**

#### **3a. Marriage and first birth**

Replicating a pattern noted in conditional before-and-after studies, Hoem and Kreyenfeld (2006) showed that conditioning on the time of first birth produces a cusp in the marriage intensity around the time of first birth in their German data, but that this is an artefact of the method of analysis and that it disappears and other patterns appear when childbearing (taken as Process B in our general theory) and marriage formation (Process A) are studied as two interacting processes, each in its own right. Conditioning on one process (B) and studying the other behaviour (A), conditional of the first, in some manner corresponds to taking Process B to be the cause of Process A in a situation where both processes are probably jointly determined by other features that are not observed by the analyst, such as the decision process of the individual or the couple. We would regard it as better practice, therefore, to study A and B together rather than to condition on one of them and run the risk of distortion.

#### **3b. The disruption of childbearing due to migration**

In the case of migrant fertility, some of the main existing theories revolve around issues of fertility 'disruption'. The fertility disruption hypothesis (e.g. Ford 1990; Hill and Johnson 2004; Kulu 2005; Roig Vila and Castro Martín 2007) refers to a temporary discontinuity in childbearing among migrants that has no counterpart for non-migrants. This hypothesis holds that in the period just before a case of migration, during the migration process itself, and for some time following migration, childbearing is likely to be postponed or possibly forgone among migrants. Fertility disruption may occur as a result of spousal separation surrounding migration, or because of a difficult transition period in terms of time needed to prepare for migration, to settle into the destination society, and to find basic housing and employment. The theory further suggests that following temporary fertility disruption due to these causes, at least some of the fertility that was postponed may be made up during a period of fertility catch-up. If the analyst wants to test the disruption hypothesis, then some analysts hold that the

natural focus is on the fertility of migrants and not on the fertility of those who could have migrated, but did not. They therefore do not regard as relevant our warning regarding potentially biased estimates of fertility confined to the pool of migrants. In a test of the disruption hypothesis, what they see as relevant is an examination of fertility behaviour among migrants as they approach the time of migration as well as following migration. If for example, it is found that the recorded fertility of people who actually migrate is lowered during the period approaching migration and in the period shortly after migration, but that it rebounds some time after migration, this would be a pattern that these analysts see as consistent with the predictions of the disruption hypothesis.

While we admit that the mechanisms said to produce fertility disruption in this argument seem plausible to us, and while we have no quarrel with the analysis of fertility *after* migration, we consider the practice of estimating fertility for the time leading up to the migration as problematic. Using intensities is a way of handling the *potential* for behaviour. The behaviour need not materialize for everyone in the population segments that are subject to such a potential. We have already argued that the estimation of pre-migration fertility for manifest migrants is likely to under-estimate fertility in the sending population and therefore is likely to run the risk of appearing to confirm pre-migration childbearing disruption even where there is none. Our point is that the analyst in reality cannot know whether what looks like confirmation is produced by estimation bias (analytic distortion).

As we have noted already, the approach that a study of fertility disruption is confined to migrants alone, is caused by the difficulty or impossibility of collecting data for the sending population in international migration. We do not believe that confining the analysis to migrants before migration is common practice in studies of *internal* migration, like that of Hoem and Nedoluzhko (2008), and we fail to see the fundamental difference between the two situations beyond the problem of data availability. Again the best practice would be to study migration and childbearing together rather than

one conditional on the other. If data are constricted to migrants alone, the analyst had better refrain from studying pre-migration fertility to avoid uncontrolled estimation bias.

### **3c. Union formation and home purchase. Intentionality**

The literature contains some contributions (Feijten and Mulder 2002, Mulder and Wagner 1998, 2001, Holland 2011) that at least for descriptive purposes (Mulder, private communication) advocate the inclusion of (unobserved) plans and intentions in discussions of the conditional before-and-after analyses that we focus on in this paper. The methodological issues connected with the association between marriage formation and home purchase are similar to those explicated already; one might even argue more strongly that marriage formation and home purchase are planned together as intentional behaviour. If intentionality is stressed, it may then appear (i) that the post-purchase marriage intensity is larger than the pre-purchase intensity (i.e., that  $\alpha_1 > \alpha_0$ ) and (ii) that the purchase intensity is larger after than before marriage formation (i.e., that  $\beta_1 > \beta_0$ ). This fits with an argument about intentional behaviour in house purchasing and in marriage formation, an argument that we find pretty convincing in principle. Unfortunately, for a period close to the conditioning event the conditional pre-and-post-event approach tends to overestimate  $\alpha_1$  when  $\beta_1 > \beta_0$  and to overestimate  $\beta_1$  when  $\alpha_1 > \alpha_0$ , as we have shown above. (Thus, incidentally, the conditional pre-and-post event procedure does not always lead to an under-estimation of the A-process intensities before the index event.) The problem is that the conditional pre-and-post-event approach does not allow the user to distinguish between on the one hand the bias produced by anticipatory analysis and on the other hand the effects of intentional behaviour, however convincing the latter may be in theoretical argumentation. The bias of the estimates remains if one only includes homebuyers in the empirical analysis. The only sure way to account for intentionality is to include explicit information about it in the data collection. Failing such information, the analyst is better advised to investigate both processes simultaneously based on data on all transitions in the state

space, as we have explained already. In other cases where this is not possible, then the analyst should rather abstain from a conditional before-and-after study.

These problems will not go away even if even if one argues that the sequence of the two processes may largely depend on chance (i.e., whether one buys the house shortly before or shortly after marriage may depend on the housing market and be purely accidental). The interference of chance should be handled explicitly and not by conditioning on the outcome in one of the interacting processes.

### **3d. Potential application to the analysis of mortality**

There is an interesting parallel to these issues in mortality analysis, and perhaps the lack of activity in studying pre-event behaviour without any information beyond the time (and perhaps cause) of death is indicative of difficulties that must be kept in mind: In studies of cause-specific mortality one can assume for sure that an individual who dies from a given chronic disease (say cancer) has been sick with the disease for a period toward the end of his or her life, and this cannot fail to have influenced behaviour on other dimensions. With the possible exception of Andersson and Sobolev (2001) we are not aware of any attempt to take such unobserved disease periods into consideration in mortality studies.

## **4. General theory**

For our general argument we let A and B be any two interacting concurrent processes on the individual level. In a Markov-chain formulation the intensities are  $\alpha_0(t)$  for the A-transition where the event A occurs at a time when B has not occurred,  $\alpha_1(t)$  for the occurrence of A *after* B,  $\beta_0(t)$  is the intensity of the occurrence of B at a time when A has not occurred, and  $\beta_1(t)$  is the intensity of the occurrence of B after A has occurred. When the four statuses are numbered from 0 to 3 as indicated in Figure 1, then  $\alpha_0(t)$  is the intensity for the transition from State 0 to 1 at process time  $t$ ,  $\alpha_1(t)$  is the

intensity for the transition from State 2 to 3 at process time  $t$ ,  $\beta_0(t)$  is the intensity for the transition from state 0 to 2 at process time  $t$ , and  $\beta_1(t)$  is the corresponding intensity for the transition from State 1 to State 3, as indicated in the figure. A proper before-and-after analysis of the A process consists in comparing (an estimate of)  $\alpha_0(t)$  with (an estimate of)  $\alpha_1(t)$ .

To pick up the effect of conditioning on an occurrence of the process B in the analysis of process A, we define a conditional intensity  $r_A(s|t)$  by the relation

$$\begin{aligned} r_A(s|t)ds &= P\{\text{event A occurs in } (s, s+ds) \mid \text{A has not occurred before time } s \text{ and B first occurs at time } t\} \\ &= \frac{P\{A \text{ first occurs in } (s, s+ds) \text{ and B first occurs in } (t, t+dt)\}}{P\{A \text{ does not occur before time } s \text{ and B first occurs in } (t, t+dt)\}}. \end{aligned} \quad (1)$$

Then  $r_A(s|t)$ , or an approximation to it, is what is estimated by the computations at the positive and negative time axis as described above, depending on whether  $s>t$  or  $s<t$ . Corresponding to the positive and negative times we now treat the cases  $t<s$  and  $t>s$  separately, as follows.

First consider positive occurrence times ( $s>t$ ). The transitions in Figure 1 that are involved in the numerator in (1) above, amount to the following scenario:

$$\text{“Stay in State 0 up to time } t, \text{ then move to State 2 in the interval } (t, t+dt), \quad (2)$$

and subsequently stay in State 2 through time  $s$ ”.

The corresponding probability is  $p_{00}(0,t)\beta_0(t)dt p_{22}(t,s)$ , where in each instance  $p_{ij}(x,y)$  is the probability of moving to State  $j$  by time  $y$  for someone who is in State  $i$  at time  $x$ . This gives us an expression for the numerator in the long fraction in (1). A corresponding expression for the denominator consists of the elements in the first row of (2), followed by a transition from State 2 to State 3 in the interval  $(s, s+ds)$ . Therefore, the denominator is  $p_{00}(0,t)\beta_0(t)dt p_{22}(t,s) \alpha_1(s)ds$ .

Dividing out factors common to the numerator and denominator in (1) thus gives  $r_A(s|t) = \alpha_1(s)$  when  $s>t$ . This shows that *the estimand of the procedure used after the index time of occurrence of an event on process B is the post-B transition intensity for the event A*. This mirrors general results for all kinds

of counting processes (Andersen et al., 1993) and is normally taken for granted without further reflection. In the migration-and-childbearing situation it means that at least as far as the analysis of fertility *after* in-migration is concerned, it can be carried out with no specific danger of estimation bias, as has been exploited by Andersson (2004), Andersson and Scott (2005), Kulu and Vikat (2007), Milewski (2008), and innumerable others.

Next consider the situation when  $s < t$  (corresponding to negative occurrence times). Then the transitions from Figure 1 that correspond to the numerator in (1) are as follows:

“Stay in State 0 until time  $s$ , and then follow one of the following disjoint alternatives:  
 either (i) also stay on in State 0 until time  $t$ , and then move to State 2 before time  $t+dt$ ,  
 or (ii) move to State 1 by time  $t$  and then move from State 1 to State 3 before time  $t+dt$ ”.

The probability for such a scenario is  $p_{00}(0, s)[p_{00}(s, t)\beta_0(t)dt + p_{01}(s, t)\beta_1(t)dt]$ . The transitions involved in the denominator in (1) are as follows:

“Stay in State 0 until time  $s$ , then move to State 1 by time  $s+ds$ , then stay there until time  $t$ , and subsequently move to State 3 by time  $t+dt$ ”.

The corresponding probability is  $p_{00}(0, s)\alpha_0(s)ds p_{11}(s, t)\beta_1(t)dt$ . When we combine the latter two formulas and remove factors that cancel, we get

$$r_A(s | t) = \frac{p_{11}(s, t)\beta_1(t)}{p_{00}(s, t)\beta_0(t) + p_{01}(s, t)\beta_1(t)} \alpha_0(s). \quad (3)$$

Estimating the transition rate for the Process A for times before the index occurrence of event B therefore does not provide an unbiased estimator for the pre-B transition rate  $\alpha_0(s)$  of A except in special cases (to which we return below). Instead we get the correct transition rate  $\alpha_0(s)$  multiplied by a distortion factor

$$\varphi_A(s, t) = \frac{p_{11}(s, t)\beta_1(t)}{p_{00}(s, t)\beta_0(t) + p_{01}(s, t)\beta_1(t)}. \quad (4)$$

For values of  $s < t$  close to the time  $t$  of occurrence of B, the distortion factor is close to

$$\lim_{s \uparrow t} \varphi_A(s, t) = \beta_1(t) / \beta_0(t), \quad (5)$$

which shows that there is little distortion of  $\alpha_0(s)$  at negative times close to the index event if the risk of an occurrence of B is largely independent of whether A has occurred or not (in which case  $\beta_0(t) \approx \beta_1(t)$ ). Otherwise one must count on the possibility that *there can be considerable systematic distortion in the estimates of  $\alpha_0(s)$  produced for times before the occurrence of the (conditioning) index event on the B process*, and one should probably avoid making such estimates. In particular, if  $\beta_1(t) < \beta_0(t)$ , such computations will underestimate the intensity  $\alpha_0(s)$  at times just before  $t$ .

Let us pursue the idea of non-differential intensities for B, which means that  $\beta_i(t)$  is independent of  $i$ , i.e.,  $\beta_i(t) = \beta(t)$  for the same function  $\beta(t)$  for both values of  $i$ . In this case

$$p_{11}(s, t) = \exp\{-\int_s^t \beta(u) du\} \text{ and } p_{0j}(s, t) = \bar{p}_{0j}(s, t) \exp\{-\int_s^t \beta(u) du\} \text{ for } j = 0, 1,$$

where  $\bar{p}_{0j}(s, t)$  is the value that one will get for  $p_{0j}(s, t)$  when  $\beta(t) \equiv 0$ . (Thus  $\bar{p}_{0j}(s, t)$  is what is sometimes called a partial probability; see Hoem 1969.) Therefore,  $\exp\{-\int_s^t \beta(u) du\} \beta(t)$  is a common factor in the numerator and denominator of the distortion factor in (4), which means that it cancels and leaves just 1 in the numerator. In the denominator,  $\bar{p}_{00}(s, t) = \exp\{-\int_s^t \alpha_0(u) du\}$  and  $\bar{p}_{01}(s, t) = \int_s^t \bar{p}_{00}(s, u) \alpha_0(u) du = 1 - \exp\{-\int_s^t \alpha_0(u) du\}$ , so the whole denominator is also identically equal to 1. [In fact, we must have  $\bar{p}_{00}(s, t) + \bar{p}_{01}(s, t) \equiv 1$  because from State 0 there is nowhere else to go than to one of the States 0 and 1 between times  $s$  and  $t$  when  $\beta_0(t) \equiv \beta_1(t) \equiv 0$ .] Thus  $\varphi_A(s, t) \equiv 1$  and there is no distortion when  $\beta_0(t) \equiv \beta_1(t)$ , i.e., when there is no differentiability in the B process. [In statistical terminology going back to a fundamental paper by Schweder (1970), B is then said to be locally

independent of A. Note that when B is locally independent, the non-distortion result holds and we get no estimation bias before the B event. Establishing *whether* B is locally independent is normally likely to be a project in itself, however.]

These arguments carry over to a situation where the intensities depend on a (possibly multidimensional) covariate, say  $X$ , which most likely will include individual age (Milewski 2008, Andersson 2004, Andersson and Scott 2005, Kulu 2006 and elsewhere, Kulu and Vikat 2007, Nedoluzhko and Andersson 2007, and many others. Subdividing the in-migrants according to their region of origin is a special case.) In this case the functions that define the distortion factor will also depend on  $X$  and so will the distortion itself, which we can now call  $\varphi_A(s, t, X)$ . If the ( $X$ -dependent) intensities  $\beta_0(t, X)$  and  $\beta_1(t, X)$  are coinciding functions of  $X$  and  $t$  (the local-independence case), then  $\varphi_A(s, t, X) \equiv 1$  and there is no estimation bias. If the  $\beta$  intensities differ from each other, then there may be noticeable estimation bias.

This covers the case of *observed* heterogeneity, in that  $X$  is taken to represent observed characteristics of the individuals. Alternatively or additionally, the intensities may be subject to unobserved heterogeneity which varies from individual to individual and which is structured in a manner which makes individuals who have higher childbearing intensities also have lower migration intensities and conversely lets individuals with higher migration intensities also have lower fertility. This may be quite plausible and is a different way of obtaining the effect we have sought to cover so far by our specification of the relationships between the basic intensities. (For a particularly clear specification of this nature, see Kulu 2006.) For the analyst who is confronted with data collected conditionally on the occurrence of a migratory event it is a problem, however, that (as far as we know) all methods that can be used to reveal the existence and effect of unobserved heterogeneity require data both for individuals who migrate and for those who do not. If data are not available for migrants and non-migrants alike, the analyst has no way of determining whether the relations between the fertility

observed before and after migration is caused by features of reality or by the method of analysis. This unresolved feature (produced by the presence of heterogeneity) will appear in similar form in all our illuminating examples, and instead of repeating it in each instance we have disregarded both observed and unobserved heterogeneity in the rest of our discussion in this paper.

As we have noted before, the accounts above contain very simple specifications of the behaviour in question. Another way of making them a bit more realistic is to let the post-B intensity for the process A depend on the duration  $d$  since the occurrence of B and thus be a function  $\alpha_1(s, d)$ . Similarly one can let the post-A intensity for B be some function  $\beta_1(t, d)$  that depends on the duration  $d$  since the occurrence of A, as in work by Kulu (2005) and elsewhere. Duration-dependent intensities make formulas such as those for distortion factors considerably more complex, but the principles remain the same, in fact estimates of the intensities at times on the positive axis of process time are unbiased while estimates of the intensities at negative process times are biased except in special cases.

We can also include births of orders after the first and more migration events than the single one used in the arguments above, in the spirit of Figure 6.2 in the book by Courgeau and Lelièvre (1992). Again such specifications can be handled by simple extensions of our arguments; in fact multiple birth orders are handled most easily by treating each birth order separately, as demographers are prone to do. (See for instance Kulu 2006.) In each case the structure of the outcome is the same, and in particular A-process estimates are unbiased when the B-process is locally independent. In practice, however, it is probably at least as hard to establish such local independence as it is to study the two interacting processes A and B, each in its own right.

## 5. Conclusions

In the analysis of two interacting behaviours it is always best to study all transition intensities involved and to use data that make this possible. In general this will allow the analyst to compare pre-

index-time with post-index-time behaviour without the risk of estimation biases. Otherwise, conditioning on one behaviour in the analysis of another will normally introduce distortions in the analysis unless the process on which the analyst conditions (Process B) has non-differential transition intensities with respect to the behaviour in focus (Process A) [i.e., unless B is locally independent of A]. In practice the conditional before-and-after type of analysis that is frequently carried out is likely to produce a distortion in the pre-B-event intensity estimate for the Process A in most cases and should therefore be avoided. In particular, it is good practice to refrain from trying to study fertility before migration on the basis of data confined to in-migrants or to study any behaviour A before event B only on the basis of those who experience B. Normally the analyst has little control over estimation biases in periods before the conditioning event and cannot safely distinguish between behavioural intentions and such biases. Intentions are best documented (if it can be done at all) from direct observation. If complete data are available, the unbiased procedure is to analyze all transition intensities. Studying each process conditionally on event-occurrence in the other (as is sometimes done) cannot be recommended as an alternative to analyzing all intensities in their own right, because a conditional both-ways analysis will just suffer from estimation biases in both directions.

### **Acknowledgements**

We are grateful to Hill Kulu, Laurent Toulemon, and Gerda Neyer, all of whom have given us very useful advice and help during our work on this paper. Comments from Gunnar Andersson and Clara Mulder, are also much appreciated. Furthermore, we are grateful to the referees and the editor of the present journal whose comments have led us to sharpen and extend our presentation at several points.

## References

- Aalen, O.O., Borgan, Ø., Keiding, N. 1987. Interaction between life history events: Nonparametric analysis of prospective and retrospective data in the presence of censoring. *Scandinavian Journal of Statistics* 7, 177-190.
- Andersen, P.K., Borgan, Ø., Gill, R.D. & Keiding, N. 1993. *Statistical Models Based on Counting Processes*. New York: Springer, 767 pp.
- Andersson, Gunnar and Boris Sobolev. 2001. Small effects of selective migration and selective survival in retrospective studies of fertility. Max Planck Institute of Demographic Research, Rostock, *Working Paper* 2001-031; <http://www.demogr.mpg.de/Papers/Working/wp-2001-031.pdf>
- Blossfeld, H.-P. & Rohwer, G. (2002). *Techniques of Event History Modeling*. Mahwah, NJ: Erlbaum.
- Cassan, Francine, Laurent Toulemon, and Annie Kensey. 2000. L'histoire familiale des hommes détenus, *Insee première* 706, 4pp., [http://www.insee.fr/fr/ffc/docs\\_ffc/ip706.pdf](http://www.insee.fr/fr/ffc/docs_ffc/ip706.pdf).
- Courgeau, Daniel and Eva Lelièvre. 1992. *Event History Analysis in Demography*. Oxford: Clarendon Press.
- Feijten, Peteke and Clara H. Mulder. 2002. The timing of household events and housing events in the Netherlands: A longitudinal perspective. *Housing Studies* 17, 773-792.
- Fix, Evelyn and Jerzy Neyman. 1951). A simple stochastic model of recovery, relapse, death and loss of patients. *Human Biology* 33(3), 205–241.
- Ford, K. 1990. Duration of residence in the United States and the fertility of U.S. immigrants. *International Migration Review* 24(1), 34-68.
- Hill, L. E. and H. P. Johnson. 2004. Fertility changes among immigrants: generations, neighborhoods, and personal characteristics. *Social Science Quarterly* 85(3), 811-826.
- Hoem, Jan M. 1969. Purged and partial Markov chains. *Skandinavisk Aktuarietidskrift* 52, 147-155.

- Hoem, Jan M. and Michaela Kreyenfeld. 2006. Anticipatory analysis and its alternatives in life-course research. Part 2: Two interacting processes. *Demographic Research* 15(17), 485-498.
- Hoem, Jan M. and Lesia Nedulezhko. 2008. Marriage formation as a process intermediary between migration and childbearing. *Demographic Research* 18 (21), 611-628.
- Keiding, N. (2013, forthcoming). Event history analysis: Local dependence and cross-sectional sampling. In: Neyer, G., Andersson, G., Kulu, H., Bernardi, L., Bühler, Ch. (eds.): *The Demography of Europe*. Springer Verlag.
- Holland, Jennifer. 2011. Home and where the heart is. Marriage timing and home purchase. *European Journal of Population* 27(1), 1-25.
- Kulu, Hill. 2005. Migration and fertility: competing hypotheses re-examined. *European Journal of Population* 21, 51-87.
- Kulu, Hill. 2006. Fertility of internal migrants: comparison between Austria and Poland. *Population, Space and Place* 12, 147-170.
- Kulu, Hill and Nadja Milewski. 2007. Family change and migration in the life course: an introduction. *Demographic Research* 17(19), 567-590.
- Kulu, Hill and Andres Vikat. 2007. Fertility differences by housing type: the effect of housing conditions or of selective moves? *Demographic Research* 17(26), 775-802.
- Milewski, Nadja. 2008. *Fertility of Immigrants and Their Descendants in West Germany — An Event-history Approach*. Universität Rostock, Dissertation.
- Mulder, Clara and Michael Wagner. 1998. First time home ownership in the family life course. *Urban studies* 35, 687-713.
- Mulder, Clara and Michael Wagner. 2001. The connections between family formation and first-time home ownership in the context of West Germany and the Netherlands. *European Journal of Population* 17, 137-164.

- Roig Vila, M. and Teresa Castro Martín. 2007. Childbearing patterns of foreign women in a new immigration country: the case of Spain, *Population* 62(3), 351–379.
- Schweder, Tore. 1970. Composable Markov processes. *Journal of Applied Probability* 7,400-410.
- Sverdrup, Erling. 1965. Estimates and test procedures in connection with stochastic models for deaths, recoveries and transfers between different states of health. *Skandinavisk Aktuarietidskrift* 1965(3-4):184–211.