



TI 2002-120/3

Tinbergen Institute Discussion Paper

# Disentangling the Age, Period, and Cohort Effects using a Modeling Approach

*France Portrait<sup>1,2,5</sup>*

*Rob Alessie<sup>1,3,4,5</sup>*

*Dorly Deeg<sup>1,2</sup>*

*<sup>1</sup> Vrije Universiteit Amsterdam, <sup>2</sup> Medical Center, <sup>3</sup> Faculty of Economics and Business Administration, <sup>4</sup> Utrecht School of Economics, <sup>5</sup> Tinbergen Institute.*

**Tinbergen Institute**

The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam, and Vrije Universiteit Amsterdam.

**Tinbergen Institute Amsterdam**

Roetersstraat 31

1018 WB Amsterdam

The Netherlands

Tel.: +31(0)20 551 3500

Fax: +31(0)20 551 3555

**Tinbergen Institute Rotterdam**

Burg. Oudlaan 50

3062 PA Amsterdam

The Netherlands

Tel.: +31(0)10 408 8900

Fax: +31(0)10 408 9031

Please send questions and/or remarks of non-scientific nature to [wdriessen@few.eur.nl](mailto:wdriessen@few.eur.nl).

Most TI discussion papers can be downloaded at <http://www.tinbergen.nl>.

Disentangling the Age, Period, and Cohort  
effects using a modeling approach:  
An application to trends in functional  
limitations at older ages.

France Portrait\*

Rob Alessie †

Dorly Deeg ‡§

October, 2002; revised August, 2003

---

\*Longitudinal Aging Study Amsterdam, Institute for Research in Extramural Medicine, Vrije Universiteit Medical Center; Department of Economics Free University; and Tinbergen Institute

†Utrecht school of Economics, Utrecht University

‡Longitudinal Aging Study Amsterdam, Institute for Research in Extramural Medicine, Vrije Universiteit Medical Center

§The Longitudinal Aging Study Amsterdam (LASA) is primarily supported by a long term grant from the Dutch ministry of Health, Welfare and Sports.

*Address for correspondence:* F. Portrait, LASA, VU University Medical Center, Van der Boechorststraat 7, NL-1081 BT Amsterdam, The Netherlands.

Tel: 00-31-204446155

Email: fportrait@feweb.vu.nl

## **Abstract**

Disentangling age, period, and cohort effects in explaining health trends is crucial to assess future prevalence of health disorders. The identification problem – year of birth plus age is calendar year – is tackled by modeling cohort and period effects using lifetime macro-indicators. The method also reveals mechanisms underlying cohort and period effects. Specific attention is paid to the impact of unobserved heterogeneity and to selective attrition. We apply the graphical, two-factors, and modeling approaches on trends in functional limitations and compare results. We argue that the modeling approach is highly appropriate. The prevalence of functional limitations is found to increase in the nineteen-nineties due to adverse period effects.

# 1 Introduction

Assessing and understanding trends in health status at older ages is an important public health issue as health is highly associated with the well-being of older populations, with the total needs for and use of health care services, and eventually with mortality. Observed differences in the prevalence of health disorders across time may be the results of interactions between age, period, and cohort (APC) effects. Understanding the role of the three effects in governing health trends is crucial to get insights into what future changes in the prevalence of health disorders at older ages might be expected. However, research that aims at separating the role of age, period, and cohort factors in explaining trends faces a serious identification problem. Indeed, age, period, and cohort effects are perfectly linearly related as year of birth plus age equals calendar year. The main objective of this study is to present an approach to separate age, period, and cohort effects. The identification problem is tackled by modeling cohort and period effects using lifetime macro-indicators. To be more specific, the dummies for calendar year and year of birth – that usually characterize period and cohort effects – are replaced by better proxies for the underlying period and cohort processes. This method has been proposed (but not applied) by Nijdegger (1981) and Heckman and Robb (1985). To our knowledge, this approach is innovative in the literature on health trends and has been seldom applied in other areas (for a recent economic application of the method, see Kapteyn et al. 2000).

Let us first examine why age, period, and cohort factors should be included in studies explaining health trends. Age effects refer to changes associated with aging – such as physiological, socio-economical – and should therefore be included in analyses on health trends. A cohort consists of individuals born in a specific year. Cohort effects relate to the impact of macro conditions that different birth-cohorts have experienced during their life course and that may act upon health status at older ages. First, there is some evidence that the probability of suffering from chronic diseases at older ages is related to

the socio-economic and living conditions as well as to exposure to infectious diseases in early life (Manton et al. 1997; van de Mheen et al. 1998; Blackwell et al. 2001). Second, birth-cohorts that grew up during times of war or depression and/or that suffered from malnutrition or poverty in infancy or early adult life may experience a worse health status at older ages than other cohorts (van Poppel et al. 1996). On the other hand, the birth cohorts with higher socio-economic levels and/or healthier life style (lower lifetime exposure to smoking and drinking, more physical exercise, more balanced eating patterns) may benefit at older ages from a better health status than other generations (Manton et al. 1997; Joung et al. 2000). Consequently, cohort factors can not be ignored in analysis on health trends. Period effects refer to macro events that shape (the path of) health status of all cohorts at the time of occurrence. It is important to realize that, because period effects affect the entire population, only period effects that occur during the period of time considered in the study (i.e. the sample period) can be identified: affected cases (observations after the event took place) and non-affected cases (observations prior to the event) are needed to measure the impact of an event on health status. The period effects that took place before the sample period are not identifiable and are actually confounded in the cohort effects. Most often, sample periods are relatively short. Still, processes such as the availability of new medicines or medical technologies, restrictions in the supply of health care facilities, or the growing labor market participation of middle-aged females (an essential source of informal care for disabled older individuals) – that could affect the average health status of the population – may develop during the sample period and should be accounted for in studies on health trends.

Consequently, all three factors – age, period, and cohort – are potential determinants of health status at older ages, and no effect can a priori be excluded from analysis on health trends. Therefore the crucial question is: how can we go beyond the basic identification problem and disentangle the three effects? Studies that deal with the identification problem generally use

dummies to proxy the three effects. The only way out to handle the perfect collinearity is to restrict in some ways the parameters of the model. To our knowledge, in the epidemiological literature, the identification problem is tackled by making some arbitrary assumptions on the nature of the cohort and period effects (Mason and Fienberg 1985; te Grotenhuis et al. 1998). We will be more explicit on that in section 2. Heckman and Robb (1985) state that dummies are “crude proxies for a variety of causal variables. The approximation is so crude that it creates a problem of its own. ... At a minimum one should look for better proxies for the underlying unobserved variables.”

All this above motivates our approach to tackle the APC identification problem. We follow the suggestion of Heckman and Robb (1985) to go beyond the use of dummies and search for better proxies for the underlying processes. We use our expertise as social scientists to model the cohort and period effects. To be more specific, concerning cohort effects, several macro indicators could be used to grasp the role on health of hygienic and socio-economic conditions during infancy and young adulthood. Examples are the percentage of deaths due to infectious diseases or the average real Gross National Product per capita during the first year of life of the respondent. With respect to period effects, changes in the average number of nursing days in hospitals or of beds in residential homes per inhabitant may for instance proxy the restrictions in availability in acute and long-term care facilities that could lead to a deterioration of the average health status of the population.

The advantages of the modeling method are several. Not only the linearity constraint linking age, period, and cohort variables is removed by substituting period and cohort dummies by better proxies. The modeling approach also sheds light on the underlying factors and mechanisms that lead to observed cohort or period differences in health status.

We apply the modeling approach on trends in self-reports on functional limitations at older ages. Panel data methods are highly suitable methods for

dealing with (APC) health issues. The analyses of the current paper are performed on longitudinal data. However, panel data, specifically on older populations, may suffer from selective attrition due to mortality or refusals. In addition, unobserved components such as genetic endowment or lifestyle might induce the observed correlation between health outcomes and health determinants, like socioeconomic factors. Our analyses are corrected for selective attrition and for the impact of unobserved individual effects. In addition, note that we explicitly model cohort *and* period effects. This has not been done before in the APC literature. Finally, we compare the results of the modeling approach with results of more traditional APC analyses, namely the graphical and the two-factors methods.

The remainder of this paper is organized as follows. Section 2 reviews the literature on APC analysis on health trends. In section 3, the method and important empirical issues such as the impact of unobserved heterogeneity and selective attrition are extensively discussed. We conclude section 3 by briefly exposing two more traditional APC methods. Section 4 presents the data set and discusses the variables used in the empirical part. Section 5 reports the results of our analyses. Section 6 discusses the results and concludes.

## 2 Existing literature

To our knowledge, the large majority of APC analyses on health trends use dummies to proxy the age, period, and cohort effects and impose some restrictions on the parameters to get rid of the perfect collinearity between the three sets of dummies. First, a substantial number of epidemiological studies ignore a priori period effects<sup>1</sup> (e.g. Alwin and McCammon 2001; Reynolds et al. 1998). However, period effects may not be negligible (see

---

<sup>1</sup>We should be aware of the fact that ignoring a factor that should be included in the analyses largely contaminates the estimation of the remaining parameters. In econometric textbooks, this is usually referred to as biases due to omission of relevant variables.



section 1) and – even if they could be assumed to be of marginal interest in the context under study – we want to be able to test this hypothesis. Therefore we do not exclude a priori period effects from our analyses.

A second branch of the literature assesses the presence of age, cohort, and period effects by estimating successively three two-factors models (in each of which one effect is ignored<sup>1</sup>). By comparing the results on unexplained variances and/or significance of the parameters, researchers evaluate the extent to which cohort and period effects are of influence in the study (see e.g. Hsu et al. 2001). Hoeymans et al. (1997) follow a similar, more refined, approach. Instead of estimating the two-factors models on the same pooled data set, they estimate the two-factors models using different data set designs in which one factor is – by construction – held constant. By looking at the consistency of the results in longitudinal models (where the cohort variable is fixed), in cross-sectional studies (where the period variable is fixed), in time-series analyses (where age is held constant), they show that the deterioration in functional status of Dutch men aged 65-85 between 1990 and 1995 was mostly due to an age effect. No cohort or period effects could be detected. However, in the first model, their parameters estimates can be interpreted as either age or period differences; in the second model, they are the results of interactions between age and cohort effects, and, in the last model, they measure a cohort-period effect. Interpreting the results is still a subjective matter.

A third type of studies gets rid of the perfect correlation between the APC dummies by using dummy categories of different spans<sup>2</sup> or more precise information – like the exact date of birth or of interview – to model the three effects (te Grotenhuis et al. 1998). However, identification of the APC parameters in studies following these approaches is not based on the explana-

---

<sup>2</sup>For instance, a model includes dummies for each year of age, for each calendar year, and for each five successive years of birth. This boils down to assuming similar cohort effects for each birth period of five years. Such a restriction is sufficient to estimate the APC model. However this restriction can not be tested and requires a strong justification.

tory power of the effects, but on their statistical characterization. To give one example, assume that cohort effects are proxied by the year of birth, period effects by the year when the interview took place, and age effects by a continuous variable (i.e. the variable is not an integer and can take values like 49.4) or a linear spline measuring the exact number of years lived. Then, the age coefficient is identified based on the variation in health status within one year of age and, therefore, measures the effect of age on the variation in health status *within* one specific year of age. In other words, the most important factor is not so much the year of birth but at which moment of the year a respondent is born. Consequently, in this context, it is highly plausible that the age effects are found to be minimal or not statistically significant. Results on APC analyses that follow this kind of approaches need to be interpreted with caution, and, to our understanding, do not actually measure real age, cohort, or period effects.

We have some additional comments regarding the use of dummies to proxy APC effects. First, we already mentioned in section 1 that dummies are very poor proxies of the unobserved underlying cohort, period, and age effects (Heckman and Robb 1985). For instance, the year of birth that summarizes cohort effects in APC analyses is a very crude measure of all the lifetime (economic, political, socio-cultural, technological, environmental) macro-influences that a specific cohort faced and that may affect his or her current health status. Additionally, dummies are not informative about the causal mechanisms underlying cohort and period effects. Moreover, some past conditions may have a positive effect (like the introduction of a new treatment) on health status whereas others (like the exposition to poverty or war) are likely to have a negative influence. Both positive and negative effects are confounded in the dummy coefficients. All above motivates our approach to tackle the APC identification problem by modeling the cohort and period effects using lifetime macro-indicators of the conditions in which cohorts spent their life.

## 3 Modeling approach

### 3.1 Model

Longitudinal data methods are highly suitable for dealing with (APC) health issues. Assume a panel data set that includes  $I$  respondents at baseline and  $T$  waves. After pooling the data of the  $T$  waves together, we can specify for the health status indicator ( $H$ ) the following model:

$$H_i^t = \alpha + A_i^t\beta + C_i\gamma + P_i^t\delta + x_i^t\zeta + v_i^t \quad (1)$$

where:

$i$ : individual index, running from  $1 \dots I$ ,

$t$ : time index in year, running from  $1 \dots T$ ,

$A_i^t$ : age of individual  $i$  at  $t$ ,

$C_i$ : year of birth of individual  $i$ , running from  $1 \dots C$ ,

$P_i^t$ : calendar year  $t$  during which respondent  $i$  is interviewed,

$x_i^t$ : vector of characteristics of individual  $i$  at period  $t$ ,

$v_i^t$ : error term.

$\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ , and  $\zeta$  are the parameters to be estimated<sup>3</sup>. Clearly, model (1) is not identified as the age, period, and cohort covariates are perfectly linearly correlated, namely calendar time equals year of birth plus age in full year. As we already stated in section 2, many studies solve the identification problem by relying on some – rather arbitrary and non-testable – assumptions. We

---

<sup>3</sup>Note that our approach to deal with the identification problem still holds if we allow for more flexible relations between health status and age, cohort, period effects. For instance, full sets of dummies, a polynomial or spline approach, or interaction variables can be used to model the APC effects. Note however that we do not allow for interaction effects between age, cohort, and period factors in our empirical study.

have argued that this may actually result in a wrong interpretation of the data. We propose here a different approach.

Cohort and period effects are the results of (lifetime) exposure to macro-factors. Denote the macro-factors responsible for cohort and period effects by  $CV$  and  $PV$  respectively. By definition, variables  $CV$  are constant for each cohort and variables  $PV$  are fixed within a specific period of time. We overcome the identification problem by modeling the cohort effects and/or period effects using lifetime macro-indicators. If we model the cohort effects by using the  $CV$  variables, equation (1) becomes (using the same notation as before):

$$H_i^t = \alpha + A_i^t \beta + \sum_{k=1}^{K_c} CV_{i,k} \gamma_k^* + P_i^t \delta + x_i^{t'} \zeta + v_i^t \quad (2)$$

where  $K_c$  refers to the total number of cohort variables included in the model and  $\gamma^*$  accounts for the parameters to be estimated associated with the included cohort variables  $CV$ . Identification of the model rests upon the assumption that the macro-indicators reflecting cohort effects  $CV$  do not depend linearly on the variable year of birth. Similarly, we could adjust equation (1) by modeling the period effects:

$$H_i^t = \alpha + A_i^t \beta + C_i \gamma + \sum_{k=1}^{K_p} PV_{k,i}^t \delta_k^* + x_i^{t'} \zeta + v_i^t \quad (3)$$

where  $K_p$  refers to the total number of period variables included in the model and  $\delta^*$  accounts for the parameters to be estimated associated with the included period variables  $PV$ . It is easy to see that modeling both cohort and period effects will also solve the identification problem:

$$H_i^t = \alpha + A_i^t \beta + \sum_{k=1}^{K_c} CV_{i,k} \gamma_k^* + \sum_{k=1}^{K_p} PV_{k,i}^t \delta_k^* + x_i^{t'} \zeta + v_i^t \quad (4)$$

Actually, rewriting (1) in (2), (3), or (4) boils down to putting some restrictions on the model. Cohort and period effects are not allowed to freely evolve across generations and time periods anymore, but are explained by changes in cohort and period variables included in the model. This is why we are able to identify all parameters of our model – as long as the number of cohort and period variables is not too large. To maintain identification of models (2), (3) and (4),  $K_c$  should be smaller than the total number of cohorts minus 2 and  $K_p$  must be smaller than the total number of periods minus 2 (Kapteyn et al. 2000). The joint significance of  $CV$  and of  $PV$  can be examined by means of a Wald test. An examination of the signs and magnitude of the parameters of the macro-variables informs us on the mechanisms explaining the observed differences between cohorts.

However, one needs to be cautious in attaching a causal interpretation to the coefficients corresponding to the macro-effects reflecting cohort or period effects. With respect to, e.g., period effects, no reliable figures on all possibly pertinent aspects of health care that may affect health status are available, and, as a consequence, changes in market conditions may be incorrectly identified by our proxies. It may be possible that the associations found between health status and cohort or period variables are as a matter of fact “spurious”. The cohort and period variables may indeed reflect a trend that is not captured by the included regressors. With respect to cohort effects, one might wish to add to a model (including age dummies, cohort variables, and dummies for the year of interview) the variable “year of birth” to address the “spurious regression” (trend correlation) problem. If the cohort variables  $CV$  are still significant, one could then attach a causal interpretation to the macro-variables coefficients, running from the cohort macro-indicators to health outcomes. However, by adding the variable “year of birth”, the model is clearly not identified anymore. We address the “spurious regression” problem by adding a quadratic “year of birth” term and see whether the macro-indicators still partly explain the health variable. In a similar way, we include a quadratic “year of interview” to address the possible

“spurious correlation” issue between health outcomes and period variables.

A remaining important issue is to assess whether we are able to explain most – if not all – of the cohort and/or period effects. This can be tested using a Wald test (Kapteyn et al. 2000). For instance, to test the validity of our cohort variables, model (2) can be tested against a general model including age dummy ( $A_i^t$ ), period dummy ( $P_i^t$ ), cohort variables ( $CV$ ), and an arbitrary set of  $(C - K_c - 2)$  cohort dummies. The specification test is on the joint significance of the additional cohort dummies. If the parameters of these cohort dummies are not jointly different from zero, no cohort effects are left unexplained in our specification. We can then claim that the cohort effects are appropriately described by the macro-indicators included in the model.

### 3.2 Empirical specification

We adjust specifications (2), (3), and (4) in several ways, namely to take into account the impact of unobserved individual effects and to correct for possible selective attrition.

**Unobserved individual effects** The error term  $v_i^t$  is likely to be correlated with the right-hand-side variables. In order to get more insight into this correlation, we write  $v_i^t$  as:

$$v_i^t = c_i + u_i^t$$

where  $u_i^t$  is a white noise error term (i.e. independent over time and across individuals). The term  $c_i$  reflects time-constant unobserved characteristics such as genetic factors or inherent frailty. The latter factors are clearly determinants of health and may be correlated with demographic and socio-economic characteristics of the individual  $X_i^t$  ( $X_i^t$ , in contrast to  $x_i^t$ , include APC variables):  $E(c_i | X_i^1, \dots, X_i^T) \neq 0$ . The right-hand variables in our health model are assumed to be *strictly exogenous* conditional on the unobserved effect  $c_i$ ,

i.e:

$$E(u_i^t | X_i^1, \dots, X_i^T, c_i) = 0 \quad (5)$$

The assumption of strict exogeneity implies that the explanatory variables in each time period are uncorrelated with the idiosyncratic error term  $u_i^t$ .

Common ways to deal with the correlation between  $c_i$  and the right-hand side variables include (i) extending the model with equations for the endogenous control variables, (ii) adopting a fixed-effect approach, and (iii) capturing the correlation between  $c_i$  and the regressors through the use of instrumental variables or by including individual specific averages (see for instance Mundlak 1978, Hausman and Taylor 1981, Chamberlain 1983). Extending equations (2), (3), or (4) with separate equations for the possible endogenous variables may not be convenient in our context as it may relate to a relatively large number of elements. One drawback of the fixed effect approach is that the time-invariant regressors – like all cohort variables, year of birth, or sex – drop out of the model. Consequently, this approach is incompatible with the object of this study. Therefore, we opt for a Mundlak approach (1978).

The Mundlak approach (1978) boils down to include in, e.g. model (4), individual specific averages for the time-varying variables. Mundlak showed that estimation of this extended model and of the fixed effects model yields exactly the same results for the coefficients corresponding to the time-varying variables (in the context of both a balanced or unbalanced panel data set). Mundlak’s approach implies that we have to add the individual specific averages of the age and period variables  $A_i^t$  and  $PV_{k,i}^t$ . However, in a balanced panel, the average age variable is perfectly correlated with the cohort dummy “year of birth” and the average period variable is perfect collinear with the constant term. Given the aim of our study, we typically do not want to include these terms. Consequently, in our context, the Mundlak’s approach boils down to the estimation of (for equation 4):

$$H_i^t = \alpha + A_i^t \beta + \sum_{k=1}^{K_c} CV_{i,k} \gamma_k^* + \sum_{k=1}^{K_p} PV_{k,i}^t \delta_k^* + x_i^t \zeta + \bar{x}_i \pi + \omega_i + u_i^t \quad (6)$$

where  $\pi$  refers to parameters to be estimated. It is important to note that the remaining individual effect  $\omega_i$  and the included regressors are at present assumed to be uncorrelated.

**Attrition** Longitudinal data sets – especially on older individuals – are likely to suffer from attrition due to mortality or frailty. Health disorders and mortality are clearly related and it is moreover conceivable that there exist factors that are usually not observed – for instance genetic factors, inherent frailty, or shocks – that relate to both. In other words, the average health status of those who leave the sample and of those who stay in the sample are likely to be different – even after we control for observed explanatory variables. As a result, an initially random sample may end up as a selective sample where the relatively healthy individuals are over- or under-represented. This will result in incorrect parameter estimates of the explanatory variables.

Our technique for testing and correcting for attrition bias follows the approach of Wooldridge (2002, chapter 17, sections 17.7.2 and 17.7.3) in a linear panel data model with unobserved heterogeneity. A simple test for selective attrition is proposed by Wooldridge (2002, p. 581)<sup>4</sup>. He suggests to include, e.g. in model (4), a selection indicator, say  $s_i^{t+1}$ , equal to one if respondent  $i$  participates to the study at  $(t + 1)$  and to zero if not. Under the null hypothesis – i.e. absence of selective attrition –, the coefficient of the selection variable  $s_i^{t+1}$  should not be significant.

Correcting for attrition bias is more complicated. We extend the method presented by Wooldridge (2002) in section 17.7.3 for a fixed effects approach to a random effects approach. First, note that the method presented here

---

<sup>4</sup>Verbeek and Nijman (1992) present a similar approach.



treats attrition as an absorbing state, implying that respondents that leave the sample at  $t$  do not re-enter the sample at  $\tau > t$ <sup>5</sup>. Consider the following panel data model:

$$H_i^t = X_i^t \theta + c_i + u_i^t \quad (7)$$

where  $X_i^t$  refers to all included right-hand side variables (including age, cohort and time dummies but excluding the household specific averages),  $c_i$  is a time-constant unobserved component, and  $u_i^t$  is a white noise error term. Conditional on  $s_i^{t-1} = 1$ , write a (reduced form) selection equation for  $t \geq 2$  as:

$$s_i^t = 1 \left[ z_i^t, \eta_t + \mu_i^t \right], \mu_i^t \mid \{z_i^t, s_i^{t-1} = 1\} \sim \text{Normal}(0, 1) \quad (8)$$

$z_i^t$  must contain variables observed in time  $t$  for all individuals with  $s_i^{t-1} = 1$ .  $z_i^t$  may, for instance, include the variables in  $X_i^{t-1}$ . We will also include some exclusion restrictions (see section 4.2). Equation (8) says that selection does not depend on  $X_i^t$  once  $z_i^t$  has been controlled for.

In order to estimate the model, we have to make some assumptions about:

- the relationship between the individual effect  $c_i$  on the one hand and  $X_i = (X_i^1, \dots, X_i^T)'$  ( $X$ -variables in all waves) and  $\mu_i^t$  (error term selection equation) on the other hand; we make the following assumption (cf. assumption 17.7c of Wooldridge, 2002, page 583):

$$E(c_i \mid \bar{X}_i, \mu_i^t) = \bar{X}_i' \pi + \xi_t \mu_i^t \quad (9)$$

where  $\bar{X}_i$  are the sample individual averages of  $X_i^t$ . This assumption is basically an adapted version of the “Mundlak” approach.

- the relationship between the error terms  $u_i^t$  and  $\mu_i^t$ . We make the following assumptions (cf. equation 17.60 of Wooldridge, 2002, page 586):

$$\begin{aligned} E(u_i^t \mid c_i, X_i, z_i^t, \mu_i^t, s_i^{t-1}) &= E(u_i^t \mid \mu_i^t) = \rho_t \mu_i^t \\ E(u_i^\tau \mid c_i, X_i, z_i^t, \mu_i^t, s_i^{t-1}) &= E(u_i^\tau \mid \mu_i^t) = 0 \text{ if } \tau \neq t \end{aligned} \quad (10)$$

---

<sup>5</sup>In our empirical study, attrition is indeed an absorbing state.

Equations (7), (9), and (10) imply that:

$$\mathbb{E}(H_i^t | X_i, \mu_i^t) = X_i^t \theta + \bar{X}_i' \pi + \phi_t \mu_i^t \quad (11)$$

where  $\phi_t = \xi_t + \rho_t$ . If we condition on  $s_i^t = 1$  instead of on  $\mu_i^t$  (because  $s_i^{t-1} = 1$  when  $s_i^t = 1$ , we do not have to condition on  $s_i^{t-1} = 1$ ), we get (note that attrition is an absorbing state and that the data set has  $T$  waves):

$$\begin{aligned} \mathbb{E}(H_i^1 | X_i) &= X_i^1 \theta + \bar{X}_i' \pi = W_i^1 \Theta \\ \mathbb{E}(H_i^2 | X_i, s_i^2 = 1) &= X_i^2 \theta + \bar{X}_i' \pi + \phi_2 \lambda(z_i^2, \eta_2) = W_i^2 \Theta \\ \mathbb{E}(H_i^3 | X_i, s_i^3 = 1) &= X_i^3 \theta + \bar{X}_i' \pi + \phi_3 \lambda(z_i^3, \eta_3) = W_i^3 \Theta \\ &\vdots \\ \mathbb{E}(H_i^T | X_i, s_i^T = 1) &= X_i^T \theta + \bar{X}_i' \pi + \phi_T \lambda(z_i^T, \eta_T) = W_i^T \Theta \end{aligned} \quad (12)$$

where  $\lambda(z_i^2, \eta_2)$ ,  $\lambda(z_i^3, \eta_3)$ , and  $\lambda(z_i^T, \eta_T)$  are the inverse Mills ratio associated with the sample selection equations (8) for  $t = 2, 3$  and  $T$ ,

$\Theta = (\theta', \pi', \phi_2, \phi_3, \dots, \phi_T)'$ ,  $W_i^1 = (X_i^1, \bar{X}_i', 0, 0, \dots, 0)'$ ,  $W_i^2 = (X_i^2, \bar{X}_i', \lambda(z_i^2, \eta_2), 0, \dots, 0)'$ ,  $W_i^3 = (X_i^3, \bar{X}_i', 0, \lambda(z_i^3, \eta_3))', \dots, 0$ , and  $W_i^T = (X_i^T, \bar{X}_i', 0, 0, \dots, \lambda(z_i^T, \eta_T))'$ . It now follows that pooled OLS of  $H_i^t$  on  $X_i^t$ ,  $\bar{X}_i$ ,  $d2_i \hat{\lambda}_i^2$ ,  $d3_i \hat{\lambda}_i^3$ ,  $\dots$ ,  $dT_i \hat{\lambda}_i^T$  – where  $d2_i$ ,  $d3_i$ , and  $dT_i$  are dummy variables indicating whether individual  $i$  participated in periods 2, 3, and  $T$  respectively, and  $\hat{\lambda}_i^2$ ,  $\hat{\lambda}_i^3$ , and  $\hat{\lambda}_i^T$  are the Mill ratio's computed after estimation of the selection equations (8) associated with  $t = 2, 3$ , and  $T$  – yield consistent estimates for  $\Theta$  (see Wooldridge 2002 for further details). The selection equations (8) are estimated using a probit specification.

**Final empirical specification** Consequently, the final empirical estimation is given by (for e.g. equation (4) and using the same notation as before)<sup>6</sup>:

---

<sup>6</sup>For similar reasons as mentioned above, we exclude from the individual averages the APC variables.

$$\begin{aligned}
H_i^t = & \alpha + A_i^t\beta + \sum_{k=1}^{K_c} CV_{k,i}\gamma_k^* + \sum_{k=1}^{K_p} PV_{k,i}^t\delta_k^* + x_i^t\zeta + \bar{x}_i'\pi + \phi_2 d2_i \hat{\lambda}_i^2 + \phi_3 d3_i \hat{\lambda}_i^3 + \dots \\
& + \phi_T dT_i \hat{\lambda}_i^T + \omega_i + u_i^t
\end{aligned} \tag{13}$$

Since we use a two-step estimation procedure, we have to correct the standard errors resulting from our analyses. We do that using the formulae of Wooldridge (2002, section 12.5). For a detailed exposition of the computation of the standard errors, see Appendix 1. Note that the obtained standard errors are robust to the presence of heteroscedasticity and autocorrelation. STATA is used to perform the calculations.

### 3.3 Traditional APC approaches:

In addition to the modeling approach, we perform two more traditional APC analyses: first, a graphical approach (see, for instance, Reynolds et al. 1998) and second, an approach in which two-factors models are estimated. We estimate successively three two-factors models, in which one factor is in turn arbitrarily excluded from the analyses. By comparing the goodness of fit of the three models and the sign and variances of the three effects, the researcher is able to decide whether age, period, and/or cohort effects are present (see, for instance, Hoeymans et al. 1997, Hsu et al. 2001). Although we somewhat disclaim APC methods in which identification of the APC parameters is based on some statistical assumptions, we perform these APC analysis so as to compare results of the graphical approach, the two-factors approach, and the modeling approach.

The last part of the paper is devoted to the application, and the comparison, of the graphical, two-factors, and modeling approaches on trends in self-reports on functional limitations of older individuals. Before presenting the results of the analyses, section 4 briefly discusses the data and measures used in the empirical part.

## 4 Data and Measures

### 4.1 Data

The analyses in the current study are conducted using data from the Longitudinal Aging Study Amsterdam (LASA), an ongoing multidisciplinary research project. The design and purposes of the LASA study are described in detail elsewhere (Deeg and Westendorp de Serière 1994, Deeg et al. 1998). The LASA study follows 3,107 non-institutionalized and institutionalized (3% at wave I) individuals aged 55-85 at baseline during the period 1992-1999 in three subsequent observations of three years apart. Advantages of the data for this study are its longitudinal character, the fact that it is nationally representative of the Dutch older population, as well as the availability of a broader range of health measures and of a detailed information on predictors of changes in health status. Physical, emotional, cognitive, and social functioning are assessed by objective and subjective instruments. The respondents were living at baseline in 11 different municipalities in the West, North-East, and South of the Netherlands. Respondents are submitted to a complete face-to-face interview or, if they refused for this, to a short telephone interview. The respondents were aged 55-85 in 1992-1993, 58-88 in 1995-1996, and 61-91 in 1998-1999. The objective of the paper is to examine differences between cohorts in the nineteen-nineties. Given the current availability of data, conclusions on differences between cohorts can be drawn for ages 61-85. Table 1 below summarizes the attrition in LASA.

< Insert Table 1 about here. >

Table 1 indicates that about 24.5% of the initially selected respondents died between wave I and III and that approximately 8.5% of them refused or were too frail to participate in the study any longer. In addition, about 14.5% of the respondents were submitted to a telephone interview. These respondents are also excluded from the study as no sufficient information is available on

them to perform the analyses. As a result of the sample selection, we suspect that healthier individuals have a higher probability to remain in the sample (Portrait 2000; Deeg et al. 2002). In other words, the sample attrition is presumably endogeneous.

## 4.2 Measures:

Summary measures on the included variables are provided in Table 2a, Figure 1, and Table 2b. Table 2a reports on health, demographic and socioeconomic variables. Figure 1 reports on the cohort macro-variables. Table 2b reports on the period macro-variables.

**Health status** We illustrate the three methods by applying them on trends in self-reports on functional status of older individuals, a health measure that has been used in several other health trend studies (e.g. Allaire et al. 1999; Hoeymans et al. 1998; Pitkala et al. 2001). Functional limitations restrict the ability, for instance, to move, bend, grasp, see, or hear and are measured in the current study by self-reports on mobility activities in daily life. These self-reports include the ability of respondents to: (1) cut one's own toenails, (2) walk up and down a 15-steps staircase without stopping, and (3) make use of private or public transportation (van Sonsbeek 1988). Note that the choice of these three items has been done step-wise: in the LASA pilot study, nine items were used to measure functional ability and the above mentioned three items were the most consistent ones to describe functional ability (Kriegsman 1997; Smits 1997). The score takes on value 0, 1, 2 when a test item is performed without any difficulty, with difficulties or only with help respectively. A score equal to 3 is given to the respondent when the activity can not be performed. These response categories are widely used (Wilson 1981). The total score is obtained by summing the three activity scores and, consequently, ranges from 0 (all activities could be performed without any difficulties) to 9 (no activity could be performed). Table 2a shows some results on the variable on functional limitations at wave I.

**Age, Cohort, and Period variables** Age effects are characterized by means of piecewise linear splines (with 4 knots equally spaced over the range of age, namely 62.9, 69.4, 76.6, and 83.8 years of age). Cohort effects are first characterized by the year of birth of the respondent recorded as linear splines with 8 knots equally spaced over the range of years of birth, namely 1911, 1914, 1917, 1920, 1924, 1927, 1930, 1933. Age and cohort effects are modeled by a spline characterization for two main reasons: (1) first using linear splines with a sufficient number of knots is a very flexible way of modeling age and cohort effects – even if it is a little less flexible than using full sets of dummies<sup>7</sup>, (2) it saves some degrees of freedom in the estimation of the parameters. The LASA respondents are born between 1907 and 1937 and reach ages 55 till 91 during the study. If a “dummy” approach was followed, 30 cohort- and 36 age-parameters should be estimated. The number of parameters to be identified reduces to 14 using a “spline” approach. Period effects are first proxied by a set of five dummies (the years during which respondents were interviewed, the first year of interview (1992) excluded to avoid perfect correlation). Note that including a spline characterization of age and cohort effects and a set of dummies for period effects in specification (1) do not artificially remove the identification problem: the sum of the spline cohort variables (characterizing year of birth) plus the sum of the spline age variables (in full years) equals the sum of the dummy variables characterizing year of interview. Table 2a reports some summary statistics on the age, year of birth, and year of interview of the respondents at wave I.

In the modeling approach, cohort and period variables are replaced by macro-variables characterizing cohort and period effects. Before starting this section, one should be aware of the fact that we are limited in our choice of cohort variables. Especially with older individuals, historical data over a long period of time are required to construct cohort variables. These data are not always available. In the Netherlands, the main source for historical macro-data is Statistics Netherlands and, therefore, our macro-data are from

---

<sup>7</sup>In the sensitivity analyses, the cohort effects are characterized by a set of 30 dummies.

Statistics Netherlands. 19 variables are used to capture most characteristics of the epidemiological, hygienic, socioeconomic lifetime conditions of the respondents that may explain cohort and period differences at older ages. With respect to cohort variables, we include:

- the number of survivors after one year of life per 10 newborns
- the percentage of deceased individuals due to infectious diseases and due to tuberculosis at birth of the respondent,
- a variable indicating the average attained level of education of fathers at birth of the respondent,
- the average number of children per female at birth of the respondent,
- the average level of attained education level of children when the respondent was 14 years of age,
- the number of children in primary school when the respondent was 7 years of age, and
- the real Gross National Product (G.N.P.) per capita during pregnancy, during the first year of life, between ages 2 and 7, between ages 7 and 14 (both periods are characterized by growth of the children), and between ages 14 and 20 (at the entrance on the job market).

By including these variables, we aim at investigating for instance whether early exposure to (infectious) diseases and whether the economic conditions in which children grew up affect the functional status at older ages. Figure 1 provides information on these cohort variables and shows that the trends are as expected (with respect to G.N.P., only the trends during the first year of life are shown): increasing number of survivors at one year of age, a steady decline of the number of deaths due to infectious diseases and/or tuberculosis with the exception of the years around the first World War, a slight increase of the average education level of the father and of the children, a small drop in the fertility rate, and a relatively stable G.N.P. from 1907 till 1925 that starts rising afterwards.

With respect to period variables, a set of 7 variables is included to characterize the changes in availability of care services that occurred in the Netherlands

during the nineteen-nineties. We use:

- the number of hospital beds per 1,000 inhabitants,
- the number of nursing days in hospitals per 1,000 inhabitants,
- the average duration of stays in hospitals,
- the number of persons in residential homes per 1,000 individuals aged 65 and above
- the number of nursing days in nursing homes per 1,000 inhabitants aged 65 and above,
- the number of workers in home care organizations per 1,000 individuals aged 65 and above, and
- the proportion of middle-aged females participating to the labor market.

Table 2b provides information on the period variables included in the study. It shows a decreasing availability of care services over time and an increasing female labor participation. Using these variables, we aim at investigating whether these restrictions affected the average health status of older individuals in the nineteen-nineties. For instance, we can assess the effects on self-reports on functional limitations of spending shorter times in hospitals or of the decrease in the rate of institutionalization among older individuals.

**Independent variables** Additional demographic and socio-economic covariates are: gender (1 = “male”, 2 = “female”), attained education level of the respondent (three dummies ranging from “low education level” till “high education level”), household real net monthly income in 1,000 euro, whether the longest job was manual labor or not (0 = “no”, 1 = “yes”), occupational prestige of the longest job according to Sixma and Ultee (1983) (ranging from 0 = “never had job” till 87 = “high prestige”), degree of urbanization of the municipality in which the respondent lives (categorical variable ranging from 1 = “low” till 10 = “high”), place of residence (two dummies for “North-East”, and “South”, with reference category “West”), partner status (0 = “no partner”, 1 = “partner”), frequency of church attendance (categorical variable ranging from 1 = “yearly or less” till 3 = “weekly or more”), and whether



the respondent experienced a significant event (war, poverty, death of the parents, divorce, drinking problems of one of the parents) during childhood (0 = “no”, 1 = “yes”).

**Exclusion restrictions** In the selection equations (8), we include – in addition to the demographic and socioeconomic variables of the second-step equations – some exclusion restrictions. First it is important to note the difficulty of finding exclusion restrictions in our context. Indeed we have to come up with variables that explain attrition due to mortality, frailty, or refusals and that do not explain functional limitations outcomes. One way to circumvent this issue is to include a full set of cohort dummies (respectively a full set of period dummies) in the selection equations associated with the second-step equations in which cohort effects (respectively period effects) are modeled using lifetime macro-variables. As additional exclusion variables, we interact sex with year of interview.

< Insert Table 2a, Figure 1, and Table 2b about here. >

## 5 Results

We analyze trends in self-reports on functional limitations of older individuals applying successively the graphical, the two-factors, and the modeling approaches.

### 5.1 A graphical approach

Figure 2 displays the prevalence of self-reported functional limitations at the three measurement points (1992-1993, 1995-1996, and 1998-1999) for 10 cohort categories of 3-year-of-birth span. On each line, the average year of birth of the cohort followed is shown. For instance, on the first line, 1936 relates to the year of birth of individuals born between 1935 and 1937. These respondents were aged 55-57 in the first wave, 58-60 in the second

wave, and 61-63 in the third wave. On each line, the cohort factor is held constant while age and calendar time evolve simultaneously. The trends that we observe are therefore due to age *and* period effects. Figure 2 reveals a clear age-period effect: the prevalence of functional limitations increases exponentially between age 55 and age 91, with a steeper increase from age 68. The vertical difference between lines measures the cohort-time effect (age factor is held constant). No noticeable cohort-period effect could be detected, as the ordering of the average scores in the three waves changes as age increases. Clearly, a graphical approach does not help to disentangle the age, period, and cohort effects. Other methods are therefore required. Note however that the graphical analysis gives a first understanding of the data and that the results of the graphical analyses may be used to confirm the results of subsequent analytical methods.

< Insert Figure 2 about here. >

## 5.2 The two-factors approach

Table 3 reports the results of the APC two-factors analyses on self-reported functional limitations. Model AP is an age-period model. Model AC is an age-cohort model. Model PC is a period-cohort model. The three models are estimated using Random Effects techniques. Gender and selection dummies<sup>8</sup> are included to the three specifications to make the results more comparable with the results of the modeling approach.

< Insert Table 3 about here. >

Results of models AP and AC are compared first. The age parameters are jointly significant in both specifications ( $\chi^2(5) = 886.93$  in model AC and

---

<sup>8</sup>We make here a distinction between attrition due to mortality and attrition due to other reasons.

$\chi^2(5) = 1121.4$  in model AP). In both specifications, the number of functional limitations rises significantly with increasing age. The slopes are similar; however the levels are somewhat different with the AC age parameters greater than the AP age parameters. Models AC and PC are then compared to find whether cohort effects are present in the study. The cohort effects are significant in both specifications ( $\chi^2(9) = 66.16$  in model AC and  $\chi^2(9) = 661.5$  in model PC). However, cohort effects are much less significant in the AC specification than in the PC specification. Moreover, the cohort coefficients are of opposite signs in the two specifications. Consequently, no confident conclusion can be drawn on the cohort effects. Finally, models PC and AP are compared to assess the presence of period effects. The period variables are jointly significant in the model PC ( $\chi^2(9) = 744.1$ ) and in the model AP as well ( $\chi^2(9) = 72.1$ ). The period variables are much less significant in the model AP than in the model PC. However, in contrast with the cohort parameters, the coefficients of the period variables are qualitatively consistent in models AP and PC: they are positive and (more or less) increasing with time. The  $R^2$  statistics are higher in models AP and AC than in model PC. They are almost equal in the models including age effects. This finding stresses the importance of including age effects in studies on trends in self-reports on functional limitations. Finally, females report on average more functional limitations than males and the mortality dummy is significant in all specifications, which confirms the need to correct for selective attrition.

In summary, the conclusions that can be drawn are that (i) age effects can not be excluded from analyses on trends in self-reports on functional limitations, (ii) there is some evidence of increasing period effects, (iii) no confident conclusion can be drawn with respect to cohort effects. The results of the previous analyses are incomplete since we can not make any strong conclusions concerning cohort effects; moreover, no information is gained on the mechanisms underlying the cohort and period effects.

### 5.3 The modeling approach

Table 4 reports the results of the modeling analyses. The table displays the parameter estimates of three specifications (sex and the Mill ratio's correcting for selective attrition are included in all specifications), namely:

- (a) A specification including age characterized by a linear spline, a full set of time dummies (the first one excluded), and a set of cohort variables,
- (b) A specification including age characterized by a linear spline, a set of cohort and period variables,
- (c) A specification including age characterized by a linear spline, a set of cohort and period variables, as well as demographic and socio-economic factors,

By including demographic and socio-economic risk factors in specification (c) only, we are able to investigate whether the cohort and period effects observed in specifications (a) and (b) remain after we correct for individual characteristics.

Table 4a reports the  $\chi^2$ -statistics of the joint significance of the age, period, and cohort parameters. Table 4b reports the  $\chi^2$ -statistics to test whether the cohort and period effects are correctly specified. We follow the approach of Kapteyn et al. (2000). An arbitrary set of  $(C - K_c - 2)$  cohort dummies is included in specification (a). In specifications (b) and (c), we include  $(C - K_c - 1)$  cohort dummies<sup>9</sup>. Likewise, an arbitrary set of  $(P - K_p - 1)$  period dummies is included in specifications (b) and (c). The joint significance of the additional cohort and period dummies is tested using a Wald test. If the models are correctly specified, the parameters of the added (cohort and period) dummies are not significantly different from 0.

< Insert Tables 4, 4a, and 4b about here. >

---

<sup>9</sup>We gain one additional degree of freedom as the period effects are modeled using macro-indicators, and not using a full set of dummies.

Before discussing the results, note that we have also performed the following analyses for males and females separately. However, this did not provide any important additional insights. Moreover, the number of cases was also approximately divided by two, which obviously made the parameter estimates less efficient. Therefore, we did not pursue these analyses any further. In the following, all analyses are performed for males and females jointly.

First of all, to decide whether we should correct for selective attrition, we have performed the test on selective attrition suggested by Wooldridge (2002, p.581) in all specifications (full results available on request by the authors). The selection dummy  $s_i^{t+1}$  was negative and statistically significant (t-value equals to 9.92 in specification (a), 10.0 in (b) and 9.2 in (c)), showing that respondents who remain in the LASA study report on average less functional limitations than the attriters. This indicates that, to get correct parameter estimates, one needs to control for selective attrition. We do that using the techniques derived in section 3.2 to correct for selective attrition in the context of a random effects linear model with unobserved heterogeneity<sup>10</sup>.

As first-step estimations, selection equations (8) were estimated including explanatory variables of specifications (a), (b), and (c), augmented by a set of exclusion restrictions. We do not show the full results on the selection equations as we are not interested in the parameters estimates (results available on request by the authors). In the selection equations associated with specification (a), we have used the interaction variables between period dummies and age parameters as exclusion restrictions (in total 5 restrictions in selection equation for wave II and 4 restrictions in selection equation for wave III<sup>11</sup>). In the selection equations associated with specifications (b) and (c),

---

<sup>10</sup>We have also estimated the attrition bias fixed effects model of Wooldridge (2002). In this model, the cohort effects are subsumed in the individual effects and the time effects are modeled by means of macro-indicators. In this way, we were able to check whether our approach and that of Wooldridge yielded similar estimates for the age coefficients. From our sensitivity analysis, this appeared to be the case (results are available upon request).

<sup>11</sup>In wave III, respondents are older than 62.9 and the first interaction variable age\*period should be removed.

we have used a full set of period dummies and the interaction variables between period dummies and age parameters as exclusion restrictions (in total 6 restrictions in selection equation for wave II and 5 restrictions in selection equation for wave III). Table 4c reports the joint significance of the exclusion variables used in the selection equations. We did not include in any selection equations (8) a full set of cohort dummies as the dummies were not jointly significant in any specification.

< Insert Table 4c about here. >

The table shows that the exclusion restrictions are more satisfying in the second selection equation than in the first.

As second-step estimations, the health specifications, in which the Mill ratio's associated with selection equations (8) are included, are estimated. Results are shown in Table 4. As expected, the Mill ratio's associated with the second and third measurement are both negative and statistically significant in all specifications. This confirms again the need to correct for selective attrition. With respect to the effect of unobserved heterogeneity, note that we exclude from the analyses the "Mundlak" variables (i.e. the individual specific averages of the time-varying variables<sup>12</sup>) when they were not statistically significant. In this case, they do not account for a significant correlation between independent variables and error term and they may obscure the interpretation of the covariates because of a high correlation between the covariates and the individual specific averages. Nevertheless, we end up with one significant "Mundlak" variable (namely average income), which shows the need to correct for possible correlation between the unobserved individual effects and the time-varying right-hand side variables.

A significant increasing age effect ( $\chi^2(5) = 207$ ) is found in specification (a): the older the individuals are, the more functional limitations they report

---

<sup>12</sup>Two time-varying regressors are included in our analyses, namely real net monthly income and partner status.

(more specifically: after approximately age 63). Specifications (b) and (c) show the existence of strong age effects as well (see Table 4a). The spline parameters are to a large extent similar in all specifications. However, in specification (c), age appears to have a significant increasing effect on functional limitations at age 69 and older, that is to say a little bit later than without correcting for demographic and socio-economic characteristics.

In preliminary analyses, we have estimated specification (a) including step-wise all cohort variables described in the section on measures. The variable that had the highest explanatory power (t-value equals to 1.4) was the variable indicating the number of deaths due to tuberculosis in the first year of the respondent<sup>13</sup>. Given the trends in exposure to tuberculosis, more recent cohorts will report less functional limitations than older cohorts, all other things equal. All other cohort variables had no or an even lower explanatory power. This indicates that early exposure to poor hygienic conditions – and not to poor socio-economic conditions – affects the prevalence of functional status at older ages. However, cohort effects do not play an important part in explaining the prevalence of self-reported functional limitations at older ages.

To validate this result, we have estimated an additional specification including the splines variables for age, a full set of cohort dummies, and the period variables included in specification (b). Again we have found no evidence of cohort effects ( $\chi^2(29) = 29.29$ ). To check whether the “tuberculosis” cohort variable should nevertheless be included in our model, we have re-estimated specification (c) without including the cohort variable and assessed whether the age and period parameters changed. The age parameters remained very similar: only the slope at younger ages (age 69-77) was a little bit steeper in the specification excluding the “tuberculosis” variable. The size of the period effects decreased slightly. We have also re-estimated specification (c)

---

<sup>13</sup>We also have re-estimated specification (a) including “year of birth”<sup>2</sup> to correct for “spurious” regression (see section 3.1). The estimates result did not change and the variable “year of birth”<sup>2</sup> was not significant.

in which the period variables were replaced by period dummies. The pattern was still increasing, but the slope was less steep when the cohort variable was included. Therefore, we conclude that, even if the cohort variable is not strongly statistically significant, it should be included in the analyses as the age and period parameters are not fully insensitive to it. Finally, to test whether the cohort effects are correctly specified, we have re-estimated all specifications in which 27 or 28 cohort dummies (see Table 4b) were included. We could accept the hypothesis of correct specification of the cohort effects.

With respect to period effects, a significant almost linearly increasing trend is shown in specification (a) ( $\chi^2(5) = 21$ ): individuals report more functional limitations at the end of the nineteen-nineties, after correction for age and cohort effects. Note that, as with the cohort variables, we have estimated in preliminary analyses specification (b) including step-wise all 7 period variables of section 4.2 and looked at the explanatory power of each period variable. Table 2b shows that the changes in period variables are to a large extent linear. To avoid “spurious” correlation between self-reported functional status and the period variables, we have included in the preliminary analyses a variable indicating quadratic “year of interview”. This variable appeared to be statistically significant in all regressions and was therefore included in all subsequent analyses<sup>14</sup>. The preliminary analyses showed that the increased prevalence of self-reported functional limitations was related to the restrictions in hospital and home care services, as well as to the increase in the percentage of middle-aged working females. We could not find any significant effects on functional status of the supply reductions in the care provided in residential and nursing homes. To maintain identification of the model, the number of period variables should be smaller than  $T - 1$ . We decided to include in the final model three period variables, namely one to proxy the restrictions in hospital services, one to proxy the restrictions in

<sup>14</sup>The quadratic “year of interview” variable, when included without any other period variables, indicates a significant increase in self-reports on functional limitations in the nineteen-nineties.



home care services, and one to proxy the restrictions in informal care. The variables that are presented in Table 4 have jointly the highest explanatory power. The three variables point at a significant increase in self-reported functional limitations due to restrictions in the availability of acute, home, and informal care services<sup>15</sup>. Note furthermore that the quadratic “year of interview” is negative and significant in specifications (b) and (c), pointing at a decrease in self-reported functional limitations in the nineteen-nineties after correction for the impact of the restrictions in acute, home, and informal care. To check whether we have explained most of the period effects, we have estimated specifications (b) and (c) after inclusion of one additional time dummy (see Table 4b). The parameter of the additional time dummy is not significantly different from 0 in both specifications. Therefore it can be concluded that all period effects are explained. It is interesting to see that the age, cohort, and gender parameters are similar in specifications (a) and (b), which indicates again that the period effects are correctly modeled. After correction for individual characteristics, the parameters associated with the period variables are still jointly statistically significant (see Table 4a). The period effects persist after we correct for demographic and socio-economic variables as the parameters of the period variables in specifications (b) and (c) are to a large extent similar. We have run one additional regression to get insight into the direction of the period effects after correction for demographic and socio-economic characteristics. Specification (c) in which the two period variables are replaced by five period dummies was estimated again. The parameters of the period dummies still showed a significant increasing trend ( $\chi^2(5) = 21.0$ ).

With respect to demographic and socio-economic variables, females report significantly more functional limitations than males in all specifications. We

---

<sup>15</sup>Note that the variable indicating the percentage working females is not significant in specification (b). When the variable was separately included in specification (b), the parameter was significant. This may be explained considering the high correlation between the hospital and working female variables (equal to -0.97).

find strong socioeconomic effects on self-reported functional ability. First, medium educated respondents report significantly more functional limitations than lower educated respondents. Second, individuals with higher incomes in the long-term report less functional disorders. No effect of variation of incomes in the short-term is shown<sup>16</sup>. Furthermore, respondents for whom the longest job was a job with a high prestige are less functional disabled at older ages than others. We did not find any effect of “physical job” after we controlled for other socio-economic variables. We have excluded this variable in the final estimations. We find some effect of the region (respondents in North-East report more functional limitations than in West), but not of the degree of urbanization of the municipality in which the respondent lives (this variable is also excluded from the final analyses). Strong negative effects on functional status of having experienced a significant event during childhood – such as war, poverty, death of the parents, divorce, or drinking problems of one of the parents – emerge. Finally, we find a strong positive effect of partner status on self-reported functional limitations: having currently a partner decreases the probability of suffering from functional disorders. We did not find any significant effect of the frequency of church attendance on self-reported functional limitations. This variable was also excluded from the final estimations.

## 6 Discussion and Conclusions

The paper aims at presenting a method to disentangle age, period, and cohort effects in explaining health trends. The identification problem – year of birth plus age is calendar year – is tackled by modeling cohort and period effects using lifetime macro-indicators. The method also reveals mechanisms underlying cohort and period effects. Specific attention is paid to selective attrition and to unobserved heterogeneity. We successively applied the graph-

---

<sup>16</sup>The parameters associated with the averages over time are interpreted as long-term effects and those related to time-varying factors as short-term effects.

ical, two-factors, and modeling methods on trends in self-reported functional limitations.

Let us first examine how the results from our modeling approach compare to the preliminary graphical conclusions of sections 5.1. First, the age-period trends shown in figure 2 are confirmed by our modeling analysis: we find increasing age and period effects for our measure for functional limitations. Second, figure 2 does not indicate any increasing cohort-period effects. Our modeling analyses do not produce evidence of clear cohort effects as well. Thus, it can be stated that the graphical and modeling results nicely confirm as well as complement each other.

The results of the modeling approach and of the two-factors technique do not correspond in all respects. Both approaches confirm the existence of age and period effects. However, using the Two-factors technique, we are not able to make any statement regarding cohort effects whereas the modeling approach shows that cohort effects do not play any important part in the prevalence of self-reported functional limitations at older ages. It is important to realize that the two-factors models PC and AC are rejected by our modeling approach, as the latter shows that age and period effects could not be omitted from the analyses. Note that, in many empirical studies, period effects are a priori excluded from the analyses (see section 2). Our results show that this is incorrect. In the current study, this will lead to erroneous reading of the data. Figure 3 displays the age parameters of models AP, AC, and of the modeling approach. The slopes are very similar in the three specifications. However, the levels are different: the two-factors model AC overestimates the age effects as compared with the modeling and AP approach. The model AC is strongly rejected by the modeling approach which argues for non-negligible period effects and no cohort effects. As expected, ignoring cohort effects does not strongly affect the age parameters: model AP proposes a similar age profile to the one of the modeling approach. Note furthermore that the models correct for selective attrition in two different ways: model AP uses attrition dummies and our modeling approach follows

the approach proposed by Wooldridge (2002).

Period parameters of the model AP and of the modeling approach are to a certain extent comparable. Model PC and the modeling approach show strongly different period patterns: ignoring (increasing) age effects in model PC considerably affects the period parameters, which show, then, a steeper increase. Moreover, when ignoring age effects in model PC, the cohort effects display a different pattern: more recent cohorts report more functional limitations than older cohorts. This is rejected by the modeling approach. Clearly, the Two-factors approach does not fully help to separate and measure the age, period, and cohort effects. Moreover, it may lead to incorrect interpretation of the data. Finally, no insights are gained on the mechanisms underlying cohort and period effects. It can therefore be concluded that the modeling approach is highly appropriate to disentangle the age, period, and cohort effects in trends in health status and that the graphical approach can be used preliminary to give a first understanding of the data.

< Insert Figure 3 about here. >

The general conclusion of our empirical study is that the prevalence of functional limitations at older ages grew during the nineteen-nineties in the Netherlands. This increase is explained by adverse period effects – that persist after we correct for demographic and socio-economic variables. The first obvious advantage of the modeling approach is that we are able to make valid statements regarding the three effects without making any (statistical) assumption on the presence of age, period, and cohort effects. Including additional information in the model (namely on the macro-factors likely to engender cohort and/or period effects) allows to unravel the identification problem.

The only cohort variable which is to some extent associated with the prevalence of self-reported functional limitations is the number of deaths due to tuberculosis during the first year of age of respondents. All other cohort vari-

ables have no or a lower explanatory power. This supports the assumption that early exposure to poor hygienic conditions – and not to poor socio-economic conditions – affects the most the prevalence of functional status at older ages. One may be surprised by the minor role played by cohort effects in explaining self-reports on functional ability. This result may be related to the subjective nature of our measurement instrument. Further research will include more objective measures of functional ability and of health status in general. With respect to period effects, a negative effect on the prevalence of functional status of the restrictions in health care that the Netherlands experienced in the nineteen-nineties is demonstrated (taken into account the possibility of “spurious” regression between self-reported functional status and period variables). We are able to explain the period effects as a result of reductions in the availability of health care services: the decreasing average number of nursing days in hospitals, the decreasing number of home care workers per 1000 individuals aged 65 and over, and the increase in the percentage of middle-aged working females – an essential source of informal care for older individuals. One explanation may be that individuals experience (higher level of) functional limitations for a longer period of time as they are waiting for e.g. surgeries, home care or informal care. Patients may also be discharged from hospitals earlier which may result in a deterioration of their health status. When they are at home, the diminution in hospital care can not be fully compensated by home and informal care services, which may lead to a further decline in functional status. In addition, our results show that the prevalence of self-reported functional limitations of older individuals would have decreased in the nineteen-nineties if we could rule out the impact of the restrictions in acute, home, and informal care. All this above demonstrates the second advantage of the modeling approach. By applying this method, information is gained on the mechanisms explaining the cohort and period effects. All this allows us to conclude that the macro-period variables that we include in the study explain the period effects in self-reported functional limitations in the nineteen-nineties in the Netherlands. Neverthe-

less, further research on this topic may include more refined measures of the restrictions on the health care market.

The conclusion of our empirical study is that the increased prevalence in self-reports on functional limitations of Dutch older individuals in the nineteen-nineties is mainly caused by adverse period effects, and not by cohort effects, and that these period effects do not vanish after we correct for individual characteristics at older ages. These period effects are mainly due to restrictions in acute, home care and informal care facilities faced by the Netherlands in the nineteen-nineties.

## References

- Alwin DF, McCammon RJ. 2001. Aging, Cohorts, and Verbal Ability. *Journal of Gerontology: SOCIAL SCIENCES* **56B**: S151-S161.
- Allaire SH, Lavalley MP, Evans SR, O'Connor GT, Kelly-Hayes M, Meenan RF, Levy D, Felson DT. 1999. Evidence for Decline in Disability and Improved Health Among Persons Aged 55 to 70 Years: The Framingham Heart Study. *American Journal of Public Health* **89**: 1678-1683.
- Blackwell DL, Hayward MD, Crimmins EM. 2001. Does childhood health affect chronic morbidity in later life? *Social Sciences and Medicine* **52**: 1269-1284.
- Chamberlain G. 1987. Panel data. In Z. Griliches and M. Intriligator, eds., *Handbook of Econometrics*. Amsterdam: North Holland.
- Deeg DJH, Westendorp de Serière M. 1994. *Autonomy and well-being in the aging population, report from the Longitudinal Aging Study Amsterdam 1992-1993*, VU University Press, Amsterdam.
- Deeg DJH, Beekman ATF, Kriegsman DMW, Westendorp de Serière M. 1998. *Autonomy and well-being in the aging population 2, report from the Longitudinal Aging Study Amsterdam 1992-1996*, VU University Press, Amsterdam.
- Deeg DJH, van Tilburg T, Smit JH, de Leeuw ED. 2002. Attrition in the Longitudinal Aging Study Amsterdam: The effect of differential inclusion in side studies. *Journal of Clinical Epidemiology* **55**: 319-328.
- te Grotenhuis M, Lammers J, Pelzer B, Scheepers P. 1998. The (im)possibility to separate age, period, and cohort effects: an application to the church attendance in the Netherlands between 1970 and 1995 (In Dutch). *Sociologische Gids* **45**: 8-28 .
- Hausman JA, Taylor WE. 1981. Panel Data and Unobservable Individual Effects. *Econometrica* **49**: 1377-1398.

- Heckman J, Robb R. 1985. Using longitudinal data to estimate Age, Period, and Cohort Effects in Earning Equations. In *Cohort Analysis in Social Research Beyond the Identification Problem*, W. Mason and S. Fienberg (eds), New-York: Springer-Verlag New-York Inc.
- Hoeymans N, Feskens EJM, van den Bos GAM, Kromhout D. 1997. Age, Time, and Cohort effects on Functional Status and Self-Rated Health in Elderly men. *American Journal of Public Health* **87**: 1620-1625.
- Hsu H, Lew-Ting C, Wu S. 2001. Age, Period, and Cohort Effects on the Attitude Toward Supporting Parents in Taiwan. *The Gerontologist* **41**: 742-750.
- Joung IMA, Kunst AE, van Imhoff E, Mackenbach JP. 2000. Education, Aging, and Health. To what extent can the rise in educational level relieve the future health care burden associated with population aging in the Netherlands? *Journal of Clinical Epidemiology* **53(9)**: 955-963.
- Kapteyn A, Alessie R, Lusardi A. 2000. Explaining the wealth holding of different cohorts: Productivity Growth and Social Security. *Research Memorandum, Tinbergen Instituut, Free University Amsterdam*.
- Kriegsman DMW, Deeg DJH, van Eijk TM, Penninx BWJH, Boeke AJP. 1997. Do disease specific characteristics add to the explanation of mobility limitations in patients with different chronic diseases? A study in the Netherlands. *Journal of Epidemiology and Community Health* **51**: 676-685.
- Manton KG, Stallard E, Corder LS. 1997. Changes in the age dependence of mortality and disability: cohort and other determinants. *Demography* **34**: 135-157.
- Mason WH, Fienberg SE. 1985. *Cohort analysis in social research: Beyond the identification problem*. New-york: Springer-Verlag.
- van de Mheen H, Stronks K, Looman C, Mackenbach J. 1998. Does childhood socioeconomic status influence adult health through behavioral factors? *International Journal of Epidemiology* **17**: 431-437.



- Mundlak Y. 1978. On the Pooling of Time Series and Cross Section Data. *Econometrica* **46**: 69-85.
- Nijdegger CN. 1981. On being caught up in time. *Human Development* **24**: 1-12.
- Pitkala KH, Valvanna J, Kulp S, Stranberg TE, Tilvis RS. 2001. Secular trends in Self-reported Functioning, Need for Assistance and Attitudes towards Life: 10-Year Differences of three Older cohorts. *Journal of the American Geriatrics Society* **49**: 596-600.
- van Poppel F, Tabeau E, Willekens F. 1996. Trends and sex-differentials in Dutch mortality since 1850: Insights from a cohort- and period-perspective. *GENUS* **LII**: 107-134.
- Portrait F. 2000. Long-term care services for the Dutch elderly – An investigation into the process of utilization. *Thesis Free University, Amsterdam*.
- Reynolds SL, Crimmins EM, Saito YS. 1998. Cohort differences in Disability and Disease Presence. *The Gerontologist* **38**: 578-590.
- Sixma H, Ultee WC. 1983. Occupational prestige score for the Netherlands in the eighties (in Dutch). *Mens & Maatschappij* **58**: 360-382.
- Smits CHM, Deeg DJH, Jonker C. 1997. Cognitive and emotional predictors of disablement in older adults. *Journal of Aging and Health* **9**: 204-221.
- van Sonsbeek JLA. 1988. Methodological and substantial aspects of the OECD questionnaire regarding long-term limitations in physical functioning (In Dutch). *Maandbericht Gezondheid (Central Bureau of Statistics), Holland* **88**: 4-17.
- Verbeek M, Nijman T. 1992. Testing for Selectivity Bias in Panel Data Models. *International Economic Review* **33**: 681-703.
- Wilson RW, McNeil JM. 1981. Preliminary analysis of OECD disability on the pretest of the post census disability survey. *Revue Epidémiologie et Santé Publique*, **29**: 469-475.

Wooldridge JM. 1995. Selection Corrections for Panel Data Models under Conditional Mean Independence Assumptions. *Journal of Econometrics* **68**: 115-132.

Wooldridge JM. 2002. Econometric Analysis of Cross Section and Panel Data. *The MIT Press Cambridge, Massachusetts; London, England*.

## Appendix I: Computation of the standard errors of the two-step estimates

In our paper, we consider the following two-steps model:

$$H_i^t = W_i^t(\hat{\eta}_2, \hat{\eta}_3, \dots, \hat{\eta}_T)' \Theta + \epsilon_i^t \quad (14)$$

where  $W_i^t$  is defined in the section on attrition,  $(\hat{\eta}_2, \hat{\eta}_3, \dots, \hat{\eta}_T)$  are obtained by running the  $T - 1$  Probit selection equations (8), and  $\epsilon_i^t = \omega_i + u_i^t$ . One obtains an estimate for the parameter vector  $\Theta$  by running OLS on equation (14). The score vector  $q_i^t(\Theta, \hat{\eta})$  for observation  $i, t$  corresponding to equation (14), where  $\hat{\eta} = (\hat{\eta}_2, \hat{\eta}_3, \dots, \hat{\eta}_T)$ , is equal to (in case of OLS one minimizes  $0.5 \sum_{i=1}^I \sum_{t=1}^T s_i^t \epsilon_i^t(\Theta, \hat{\eta})^2$  with respect to  $\Theta$ ):

$$q_i^t(\Theta, \hat{\eta}) = s_i^t W_i^t(\hat{\eta}) \epsilon_i^t(\Theta, \hat{\eta}) \quad (15)$$

As starting point, we take equation (12.35) of Wooldridge (2002):

$$I^{-0.5} \sum_{i=1}^I q_i^t(\Theta_o; \hat{\eta}) = I^{-0.5} \sum_{i=1}^I q_i^t(\Theta_o; \eta^*) + F_o \sqrt{I}(\hat{\eta} - \eta^*) + o_p(1)$$

where  $\Theta_o$  is the true parameter  $\Theta$ ,  $\eta^*$  is the true parameter of our selection model,  $F_o$  is the  $K \times J$  - derivatives with respect to  $\eta$  of the score vector (15) ( $K$  is the dimension the parameter vector  $\Theta$  and  $J$  is the dimension of the parameter vector  $\eta$ ). Therefore,  $F_o$  is defined by:

$$F_o \equiv E \left[ \nabla_{\eta} q_i^t(\Theta_o; \eta^*) \right]$$

Wooldridge (2002) shows that the variance-covariance  $V$  matrix of  $\Theta$  can be estimated as follows (see equation 12.41):

$$\hat{V}(\hat{\Theta}) = \hat{A}^{-1} \hat{D} \hat{A}^{-1}$$

where  $g_i(\hat{\Theta}; \hat{\eta}) = \sum_{t=1}^T q_i^t(\hat{\Theta}; \hat{\eta}) + \hat{F}_o r_i(\hat{\eta})$ ;  $r_i = (r_i^1, \dots, r_i^T)'$  with  $r_i^t$  equals to minus the inverse of the average estimated Hessian times the estimated score of the probit log-likelihood function for individual  $i$ ;  $\hat{D} = I^{-1} \sum_{i=1}^I g_i g_i'$ ;

$\hat{A} = I^{-1} \sum_{i=1}^I \sum_{t=1}^T s_i^t \hat{H}_i^t$ , where  $H_i^t$  is the Hessian associated with equation (14). The most difficult task is to obtain a formula for  $\hat{F}_o$ : see lemma below.

**Lemma**

$$\hat{F}_o = (\hat{F}_2 | \hat{F}_3 | \cdots | \hat{F}_T) \quad (16)$$

where  $(\lambda_i^t = \lambda(z_i^t, \hat{\eta}_t))$  and  $\lambda_i^t$ , its derivative w.r.t  $z_i^t, \eta_t$ :

$$\begin{aligned} \hat{F}_2 &= I^{-1} \sum_{i=1}^I \hat{\phi}_2 s_i^2 \hat{\lambda}_i^2 W_i^2 z_i^2, = -I^{-1} \sum_{i=1}^I \hat{\phi}_2 s_i^2 \hat{\lambda}_i^2 (z_i^2, \hat{\eta}_2 + \hat{\lambda}_i^2) W_i^2 z_i^2, \\ \hat{F}_3 &= I^{-1} \sum_{i=1}^I \hat{\phi}_3 s_i^3 \hat{\lambda}_i^3 W_i^3 z_i^3, = -I^{-1} \sum_{i=1}^I \hat{\phi}_3 s_i^3 \hat{\lambda}_i^3 (z_i^3, \hat{\eta}_3 + \hat{\lambda}_i^3) W_i^3 z_i^3, \\ &\vdots \\ \hat{F}_T &= I^{-1} \sum_{i=1}^I \hat{\phi}_T s_i^T \hat{\lambda}_i^T W_i^T z_i^T, = -I^{-1} \sum_{i=1}^I \hat{\phi}_T s_i^T \hat{\lambda}_i^T (z_i^T, \hat{\eta}_T + \hat{\lambda}_i^T) W_i^T z_i^T, \end{aligned}$$

**Proof**

From Wooldridge (1995), equation (A3), we know that

$$\hat{F} = I^{-1} \sum_{i=1}^I \sum_{t=1}^T \frac{\partial q_i^t(\hat{\Theta}, \hat{\eta})}{\partial \eta'}$$

First, we have to compute the derivative of the score vector (cf. equation 15) with respect to  $\eta$ . Obviously (no Mill's ratio's appear in equation 12-1),

$$\frac{\partial q_i^1}{\partial \eta_2'} = 0; \frac{\partial q_i^1}{\partial \eta_3'} = 0; \cdots; \frac{\partial q_i^1}{\partial \eta_T'} = 0$$

For wave 2, the derivatives are equal to  $(\lambda_i^2 = \lambda(z_i^2, \hat{\eta}_2))$  and  $\lambda_i^2$  its derivative w.r.t  $z_i^2, \eta_2$ ) (see equation 12-2):

$$\begin{aligned} \frac{\partial q_i^2}{\partial \eta_2'} &= -s_i^2 W_i^2 \frac{\partial \epsilon_i^2}{\partial \eta_2'} = s_i^2 W_i^2 z_i^2, \hat{\phi}_2 \hat{\lambda}_i^2, = -s_i^2 W_i^2 z_i^2, \hat{\phi}_2 \hat{\lambda}_i^2 (z_i^2, \hat{\eta}_2 + \hat{\lambda}_i^2) \\ \frac{\partial q_i^2}{\partial \eta_3'} &= 0; \cdots; \frac{\partial q_i^2}{\partial \eta_T'} = 0 \end{aligned}$$

For wave 3, the derivatives are equal to  $(\lambda_i^3 = \lambda(z_i^3, \hat{\eta}_3))$  and  $\lambda_i^3$  its derivative w.r.t  $z_i^3, \eta_3$ ) (see equation 12-3):

$$\begin{aligned}\frac{\partial q_i^3}{\partial \eta_3'} &= -s_i^3 W_i^3 \frac{\partial \epsilon_i^3}{\partial \eta_3'} = s_i^3 W_i^3 z_i^3 \hat{\phi}_3 \hat{\lambda}_i^3, = -s_i^3 W_i^3 z_i^3 \hat{\phi}_3 \hat{\lambda}_i^3 (z_i^{3'} \hat{\eta}_3 + \hat{\lambda}_i^3) \\ \frac{\partial q_i^3}{\partial \eta_2'} &= 0; \dots; \frac{\partial q_i^3}{\partial \eta_T'} = 0\end{aligned}$$

Similarly, for wave T, the derivatives are equal to  $(\lambda_i^T = \lambda(z_i^T, \hat{\eta}_T))$  and  $\lambda_i^T$ , its derivative w.r.t  $(z_i^T, \eta_3^T)$  (see equation 12-T):

$$\begin{aligned}\frac{\partial q_i^T}{\partial \eta_T'} &= -s_i^T W_i^T \frac{\partial \epsilon_i^T}{\partial \eta_T'} = s_i^T W_i^T z_i^T \hat{\phi}_T \hat{\lambda}_i^T, = -s_i^T W_i^T z_i^T \hat{\phi}_T \hat{\lambda}_i^T (z_i^{T'} \hat{\eta}_T + \hat{\lambda}_i^T) \\ \frac{\partial q_i^T}{\partial \eta_2'} &= 0; \frac{\partial q_i^T}{\partial \eta_3'} = 0; \dots\end{aligned}$$

Having obtained these derivatives, it is easy to establish the result mentioned in the lemma.

Table 1: Pattern of attrition in the LASA study

	Wave I	Wave II	Wave III
Number of individuals	3,107	2,297	1,874
Deceased	-	417	344
Too frail / Refusals	-	145	125
Respondents with a telephone interview	-	243	202

Table 2a: Descriptive statistics: Health, Demographic and Socioeconomic factors; Wave I

Variables	Outcome	(%)
Number of respondents*		2,991
Self reports on Functional Limitations	0	58.6
	1-3	25.0
	4-6	9.2
	7-9	7.2
Age	55-60	16
	60-65	17.5
	65-70	17
	70-75	15.3
	75-80	18
	80-85	16.2
Year of birth	1908-12	18.2
	1913-17	18.5
	1918-22	15.1
	1923-27	15.8
	1928-32	16.7
	1933-37	15.7
Year of interview	1992	33.9
	1993	66.1
Female		51.2
Attained education level	Low	43.9
	Medium	42.2
	High	13.9
Net monthly income (in Euro)	< 625	22.3
	625-852	22.4
	853-1080	16.7
	1081-1477	18.9
	1478-1932	10.4
	> 1933	9.5
Type of longest job	Not manual	73.0
Occupational prestige longest job	Mean	27.2
Degree of urbanization	Low (score 1-3)	12.7
	Medium (score 4-6)	27.0
	High (score 7-10)	60.3
Place of residence	North-East	30.7
	South	23.9
	West	45.4
Partner status	No partner	33.5
Church attendance	≤ Yearly	47.1
	Monthly	13.7
	≥ Weekly)	40.2
Significant event during childhood	No	72.4

\*: After exclusion of missing values.

Table 2b: Descriptive statistics: Period macro-variables

Variables	1992	1993	1995	1996	1998	1999
Number hospital beds per 1,000	4.2	4.1	3.9	3.8	3.7	3.6
Number nursing days in hospitals per 1,000	1.1	1.08	1.02	1.0	0.94	0.88
Average duration of hospital stay (in days)	3.02	2.95	2.86	2.71	2.59	2.40
Number of resid. homes dwellers per 1,000 65+	127	124	119	117	108	107
Number nursing days in nursing homes per 65+	9.62	9.66	9.75	9.76	9.70	9.57
Number of home care workers per 1,000 65+	0.025	0.025	0.024	0.023	0.023	0.023
% middle-aged working females	34.76	34.81	41.44	42.31	45.8	47.07



**Table 3:** Results on two-factors APC approach on Self-reported Functional Limitations

	Model AP		Model AC		Model PC	
	Par.	T-stat.	Par.	T-stat.	Par.	T-stat.
Constant	-1.60	-1.5	-361.94	-2.4	337.17	2.3
Selection dummies						
Mortality	1.282	12.9	1.326	13.2	1.195	12.0
Other reason	0.197	1.5	0.202	1.5	0.131	0.9
Other reason	0.326	1.7	0.349	1.9	0.256	1.3
Age spline						
[55, 62.9)	0.006	0.4	0.061	3.5		
[62.9, 69.4)	0.032	2.6	0.075	5.6		
[69.4, 76.6)	0.107	9.1	0.159	11.2		
[76.6, 83.8)	0.21	16.1	0.287	18.6		
[83.8)	0.368	15.8	0.435	19.1		
Period dummies						
1993	0.236	3.3			0.349	4.8
1995	0.213	3.2			0.529	7.9
1996	0.427	5.9			0.886	12.5
1998	0.188	2.6			0.842	12.4
1999	0.609	7.4			1.418	18.8
Cohort spline						
[1908, 1911)			0.186	2.4	-0.176	-2.3
[1911, 1914)			0.016	0.2	-0.294	-3.9
[1914, 1917)			0.227	3.0	-0.024	-0.3
[1917, 1922)			-0.086	-1.1	-0.283	-3.7
[1922, 1925)			0.163	2.2	-0.005	-0.1
[1925, 1928)			-0.072	-1.0	-0.182	-2.4
[1928, 1931)			0.078	1.0	-0.01	-0.1
[1931, 1934)			0.038	0.5	-0.042	-0.5
[1934, 1937)			0.042	0.5	-0.021	-0.3
Demographic Variables						
Female	1.026	13.5	1.038	13.6	1.029	13.5
R <sup>2</sup>	0.2871		0.2880		0.2795	

**Table 4:** Results on APC modeling approach on Self-reported Functional Limitations

	Spec. (a)		Spec. (b)		Spec. (c)	
	Par.	T-stat.	Par.	T-stat.	Par.	T-stat.
Constant	-1.47	-1.4	21.45	3.5	20.5	3.2
Selection dummies						
Mill ratio (t=2)	-0.94	-2.5	-0.90	-2.4	-1.05	-2.7
Mill ratio (t=3)	-1.24	-2.4	-1.25	-2.4	-1.02	-1.9
Age spline						
[55, 62.9)	0.004	0.2	0.004	0.3	0.002	0.1
[62.9, 69.4)	0.05	2.6	0.05	2.6	0.03	1.5
[69.4, 76.6)	0.10	4.4	0.10	4.4	0.08	3.6
[76.6, 83.8)	0.23	10.6	0.23	10.6	0.21	9.6
[83.8)	0.37	5.7	0.38	5.8	0.37	5.6
Period dummies						
1993	0.24	2.9				
1995	0.36	2.3				
1996	0.61	3.7				
1998	0.28	1.4				
1999	0.65	3.7				
Cohort variables						
Death tub.	0.004	1.4	0.004	1.5	0.004	1.5
Period variables						
Hosp. nursing days			-22.50	-4.0	-21.51	-3.5
Home care workers			-2.12	-3.2	-2.07	-2.9
% working females			0.06	1.3	0.09	2.5
(year of interview) <sup>2</sup>			-0.06	-3.8	-0.07	-3.4
Dem. & Soc-eco. Variables						
Female	0.79	11.1	0.80	11.1	0.50	6.4
Medium education					0.21	2.1
High education					-0.11	-1.1
Real income (in 1,000 Euro)					0.06	0.9
Real income					-0.71	-6.6
Prestige					-0.005	-2.6
North-East					0.16	1.8
South					0.11	1.1
Childhood event					0.30	3.5
Partner status					-0.22	-2.3
R <sup>2</sup>	0.2602		0.2601		0.2793	
Number of observations	6,973		6,973		6,276	

Table 4a:  $\chi^2$ -values for joint significance of the age / cohort / period parameters

	Spec. (a)	Spec. (b)	Spec. (c)
Age parameters	<i>207</i>	<i>207</i>	<i>180</i>
Period parameters	<i>21</i>	<i>17</i>	<i>14</i>
Cohort parameters	2.0	2.2	2.1

(Figures in italic are significant at 1%)

Table 4b:  $\chi^2$ -values for model misspecification

	Spec. (a)	Spec. (b)	Spec. (c)
Cohort effects	$\chi^2_{(27)} = 26.9$ (0.47)	$\chi^2_{(28)} = 24.9$ (0.58)	$\chi^2_{(28)} = 25.6$ (0.55)
Period effects	n.a.	$\chi^2_{(1)} = 0.44$ (0.44)	$\chi^2_{(1)} = 2.37$ (0.12)

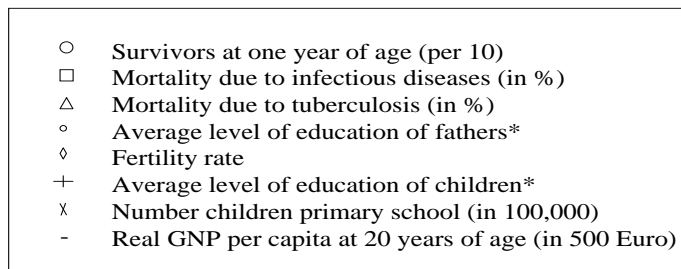
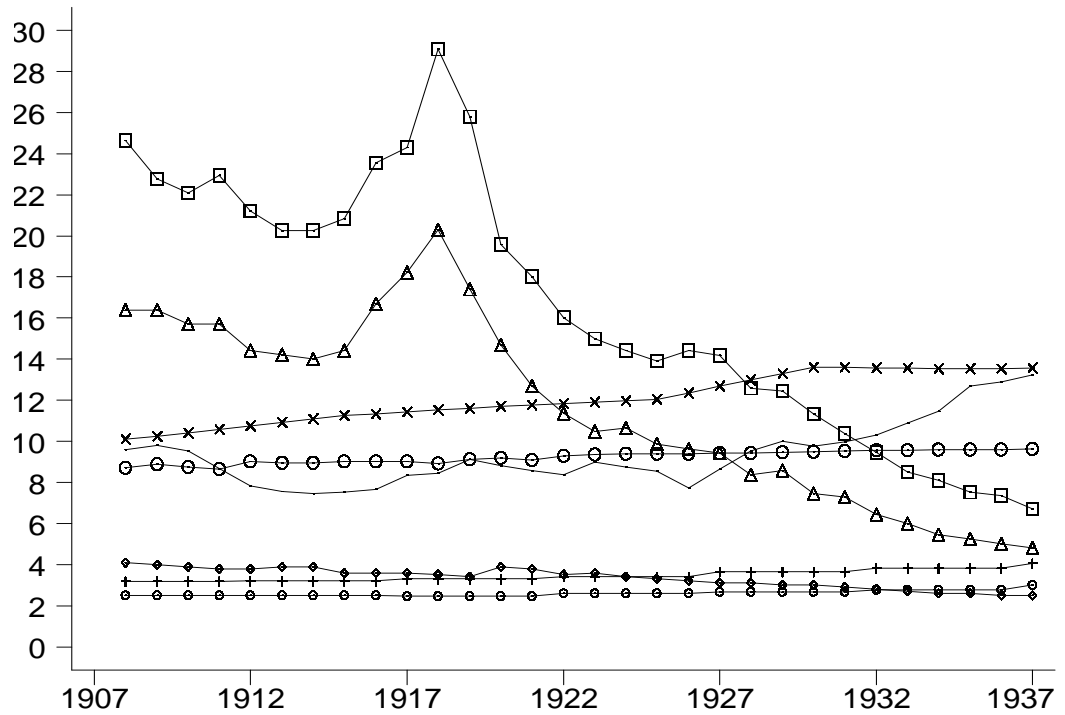
(Figures in brackets are the probabilities of acceptance of  $H_0$ : models are correctly specified.)

Table 4c: Wald test on the exclusion restrictions

	Spec. (a)	Spec. (b)	Spec. (c)
First selection equation	$\chi^2_{(5)} = 4.2$ (0.51)	$\chi^2_{(6)} = 4.9$ (0.54)	$\chi^2_{(6)} = 6.9$ (0.32)
Second selection equation	$\chi^2_{(4)} = 9.1$ (0.06)	$\chi^2_{(5)} = 9.70$ (0.08)	$\chi^2_{(5)} = 11.9$ (0.03)

(Figures in brackets are the probabilities of acceptance of  $H_0$ : Parameters are not different from 0.)

Figure 1: Descriptive statistics: Cohort macro-variables



\*: calculation based on an education indicator with scores ranging from 1 (primary education not complete) till 9 (university 2nd grade)

Figure 2: Age / Period / Cohort effects in Functional Limitations

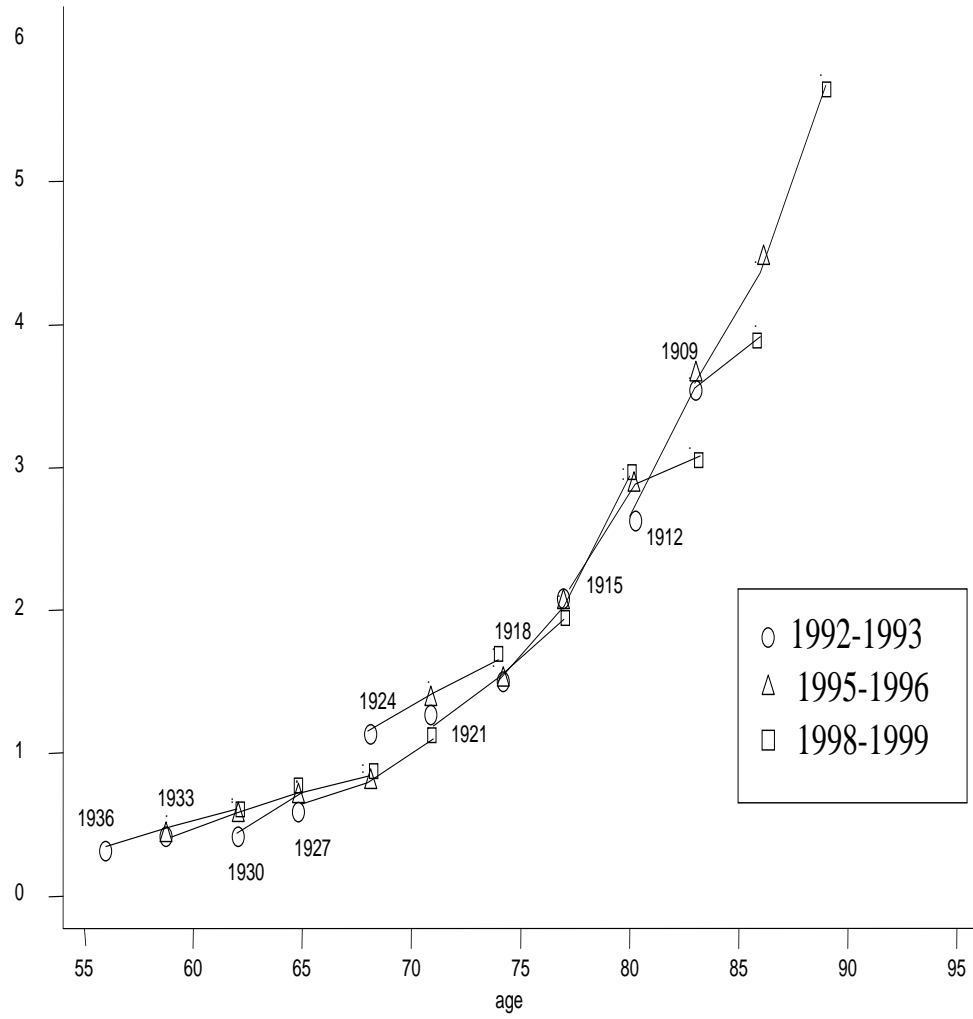


Figure 3: Age parameters in models AP, AC, and of the modeling approach

