# Estimating Disability Prevalence and Incidence in Italy Using a Multistate Model: Merging Information from Different Sources

Enrico Roma[a], Rossella Miglio[a],
[a] University of Bologna, Department of Statistical Science ”Paolo Fortunati”, Italy.

## 1. Introduction

Disability profoundly affects people’s lives well beyond health. It is associated with social exclusion and elevated mortality (Officer & Posarac, 2011; Kuper et al., 2024; Smythe & Kuper, 2024). For ageing societies, accurately quantifying and projecting the disability burden is central to planning and equity. Policymakers need interpretable and operational indicators, together with projections that anticipate care needs and guide resource allocation. This challenge is acute in Italy, one of Europe’s fastest-ageing countries (EUROSTAT, 2020), where estimates of the numbers of people with limitations are critical to health and social-care planning. Relying on prevalence can be misleading because prevalence conflates disability incidence, recovery and mortality. To inform policy, we must explicitly model these flows.

The Global Activity Limitation Indicator (GALI) provides a pragmatic and widely adopted measure of disability (Robine, Jagger, Euro-REVES Group, 2003), aligned with the International Classification of Functioning, Disability and Health. It asks respondents whether they were limited in their activities by health problems in the previous six months. In this paper we analyze GALI and its transitions using two complementary data sources: the Survey of Health, Ageing and Retirement in Europe (SHARE) and ISTAT’s “Aspetti della vita Quotidiana” (AVQ) household survey. SHARE is a panel that follows part of the individuals over time and allows estimation of age dependent transition intensities between health states, while AVQ is cross-sectional and provides large sample sizes and regional representativeness but no follow up. Both surveys collect the GALI along with demographic and socio-economic variables and target noninstitutionalized adults living in private households.

To disentangle incidence, recovery and mortality we model the respondent’s health status as a continuous-time Markov process with three states: non-disabled, disabled and dead. This framework links age-specific hazards to transition probabilities and to model-based prevalences in a coherent way (Cox and Miller, 1965), under the assumption of stable population and no period effect (Roma and Miglio, 2025). We extend the influential method of van der Gaag et al. (2015), which assumes an illness–death model and derives a point estimate of the disability incidence rate by deterministically solving the equation linking hazards to prevalence. Our approach allows for recovery (transitions back to the non-disabled state), makes more efficient use of the available information, and propagates uncertainty rather than relying on plug-in estimates.

Our objective is to estimate a survey-integrated multistate disability model for Italy that leverages all available information from SHARE and AVQ and supports projections of disability under alternative scenarios.

## 2. Data

SHARE is a multidisciplinary survey conducted across 27 European countries and Israel (Börsch-Supan et al., 2013; Bergmann et al., 2019). We analyze waves from 1 (2003–2004) to 9 (2022) (SHARE-ERIC, 2024), excluding wave 3 because it is largely retrospective and omits the outcomes of interest. We focus on the nationally representative Italian subsample and use the SHARE imputed data, which provide a dichotomized GALI indicator alongside sex, age, and education level. Mortality timing is measured as the reported month of death. To complement these longitudinal data, we use AVQ, restricting to the 2015 wave for computational tractability under the maintained assumption of no period effects in disability (ISTAT, 2015). From AVQ we harmonize the same variables plus geographic area, standardize education to ISCED-97, and consider the age classes [45,55), [55,60), [60,65), [65,75), [75, +∞), noting that available AVQ reports grouped ages only. Based on the sampling designs, we treat both sources as targeting the same population: non-institutionalized adults living in private households.

## 3. Model

Let $Y(x) \in \{0,1,2\}$ be the health status at the age $x \in R^{+¿¿}$, with $Y(x)=0$ indicating non-disabled (nD), $Y(x)=1$ disabled (D), $Y(x)=2$ dead individuals. The full set of dichotomous covariates, defined from sex (2 levels), education (4 levels) and ISTAT geographical area (5 levels), is indicated by $z \in \{0,1\}^{p}$. If the geographical area is not considered, we use $z_{-r} \in \{0,1\}^{p-5}$. We assume a continuous-time illness-death with recovery Markov model, as in Figure 1. Hazards are smooth function of age, chosen as time measure and covariates enter under a proportional-hazard structure.

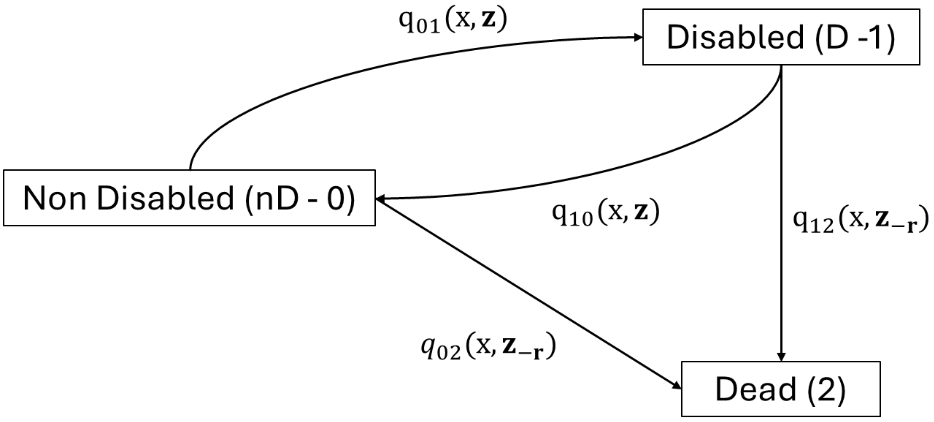

*Figure 1- Continuous-time Markov model with transition intensities*

The generator of the continuous-time Markov model is a matrix $Q(x/z)$ with elements $[q_{jk}(x/z_k)]$, $j,k=0,1,2$, $q_{jj}(x/z_j) = -\sum_{k \neq j} q_{jk}(x/z_k)$, $z_k = ¿\ z$ if $k \neq 2$ and $z_{-r}$ otherwise. For $q_{01}(x \vee z)$and $q_{10}(x \vee z)$, we model the log transition hazards as $\log q_{jk}(x/z) = \sum_{h=1}^{H} \beta_{jk,h} b_h(x) + \gamma_{jk}^{T} z$, where $\beta_{jk,h}, \gamma_{jk} \in R$ are the coefficients to be estimated and $b_h(x)$ are cubic B-spline basis. We let $H=6$, with knots automatically placed at percentiles of the distribution of age. Note that $\gamma_{jk}^{T} = (\gamma_{-r,jk}^{T}, \gamma_{r,jk}^{T})$ where $\gamma_{r,jk}^{T}$ are the specific coefficient for the geographical area and we imposed $\sum_{r=1}^{5} \gamma_{r,jk}^{T} = 0$. We will use $\Gamma_{-r}$ and $\Gamma$ to indicate the vector of all $\gamma_{-r,jk}^{T}$ and $\gamma_{r,jk}^{T}$ coefficients.

Let $P(x, x+\delta)$ denote the $3 \times 3$ transition probability matrix over the age interval ¿, with $\delta \geq 0$. For a non-homogeneous process (see Cox and Miller, 1965), it holds that

$$\frac{\partial}{\partial x} P(x, x+\delta) = Q(x) P(x, x+\delta), P(x,x) = I.$$

We approximate $Q(\cdot)$ as piecewise constant over a grid of age with 2 years intervals (average time interval between two SHARE waves and finer than spline knots). To obtain transition probabilities from age $x$ to $x+\delta$ we traverse the intervals intersecting $[x, x+\delta)$, applying the matrix exponential over each segment's duration. Multiplying these segment-wise transitions in chronological order yields the overall transition matrix. Spline bases are evaluated at the lower bound of the age interval (Machado et al., 2021).

## 4. Estimation

SHARE provides interval-censored panel data. For each respondent $i$, we observe health state $y_{i,l}\left(x_{i,l}\right)$, for$l=0,\ldots,L-1$, with no information on the path between interviews. The likelihood contribution of the $i$-th respondent for interval $\left[x_{i,l},x_{i,l+1}\right]$ is the scalar entry $p_{jk}\left(x_{i,l},x_{i,l+1}\middle|z_{-r,i}\right.$¿ of the transition probability matrix. Under the first-order Markov property, the joint probability of an individual's observed sequence factors into the product of interval transition probabilities. The individual contribution to the SHARE likelihood is therefore

$$L^{SHARE}\left(\beta,\Gamma_{-r}\right)=\prod_{i=1}^{N}L_i^{SHARE}\left(\beta,\Gamma_{-r}\right)=\prod_{i=1}^{N}\prod_{l=0}^{L_i-1}\left[p_i\left(x_{i,l},x_{i,l+1}\vee z_{i,-r,};\beta,\Gamma_{-r}\right)\right].$$

Where $\beta,\Gamma_{-r}$ are vectors of parameters including all transitions, $L_i$ is the number of observations of the same respondent. Time of death is known exactly and the likelihood entry modified according to Machado et al., 2021. Note that SHARE does not permit the estimation of the geographical area specific parameters.

For AVQ, let $y_c$ be the number of GALI-limited out of $n_c$ respondents in cell $c$, defined by the combination of the age classes with limits $\left[L_c,U_c\right]$ and the $p$ covariates. The class average model prevalence is defined as

$$p_c\left(\beta,\Gamma_{-r},\Gamma_r\right)=\frac{1}{U_c-L_c}\int_{L_c}^{U_c}w\left(x/z;\beta,\Gamma_{-r},\Gamma_r\right)dx\,,$$

where $w\left(x/z;\beta,\Gamma_{-r},\Gamma_r\right)$ solves the Kolmogorov forward equation

$$\frac{dw(x/z)}{dx}=q_{01}(x,z)\left[1-w(x/z)\right]-q_{10}(x,z)w(x/z)+w(x/z)\left[1-w(x/z)\right]\left(q_{02}\left(x,z_{-r}\right)-q_{12}\left(x,z_{-r}\right)\right)$$

with $y_c \quad Bin\left(n_c,p_c\right)$. Regional heterogeneity enters by a unique set of parameters $\Gamma_r$ and by a logit shift at an anchor age $w(45/z)=logit^{-1}\left(w_0\left(45\vee z_{-r}\right)+\gamma_r\right)$. The parameter $w_0\left(45\vee z_{-r}\right)$ can be initialized for each combination of the covariates solving the local equilibrium equation for national level prevalences $\frac{dw\left(x=45/z_{-r}\right)}{dx}=0$ in $[0,1]$. The joint likelihood to be optimized is the product of the two survey likelihoods with a penalization to facilitate identifiability, using a Ridge penalty on geographical area specific coefficient.

## 3. Preliminary results

Incidence of disability raised steeply with age, while recovery declined approximately linearly toward zero at advanced ages, implying that late-life disability is effectively permanent. A clear socioeconomic gradient emerged: lower education is associated with higher incidence and lower recovery, with gender contrasts mirroring these incidence–recovery profiles and well-known mortality differentials. We observed a good agreement between predicted versus observed prevalence with higher uncertainty at higher ages due to a combination of lower sample sizes and broader age classes. Overall, joint estimation using SHARE and AVQ tightened uncertainty, yielding more precise incidence and recovery and enabled identification of the full transition-intensity matrix $Q(x/z)$, under the proportional assumption on the covariates.

## References

Bergmann, M., Kneip, T., De Luca, G., Scherpenzeel, A. (2019). Survey participation in the Survey of Health, Ageing and Retirement in Europe (SHARE), Wave 1–7. Based on Release 7.0.0. SHARE Working Paper Series, 41-2019. MEA, Max Planck Institute for Social Law and Social Policy, Munich, (DE). Working Paper

Börsch-Supan, A., Brandt, M., Hunkler, C., Kneip, T., Korbmacher, J., Malter, F., Schaan, B., Stuck, S., Zuber, S. (2013). Data Resource Profile: The Survey of Health, Ageing and

Retirement in Europe (SHARE). International Journal of Epidemiology, 42(4), pp. 992-1001. Journal Article

Cox, D.R., Miller, H.D. (1965). The Theory of Stochastic Processes. Chapman & Hall, London, (UK). Book

Eurostat. (2020). Ageing Europe – statistics on population developments. Eurostat Statistics Explained. Available at: https://ec.europa.eu/eurostat/statisticsexplained/index.php?title=Ageing_Europe_-_statistics_on_population_developments. Accessed: 28 January 2025. Web Article

Istituto Nazionale di Statistica (ISTAT). (2015). Aspetti della vita quotidiana (Aspects of Daily Life Survey). Data: 2013–2022. Available at: https://www.istat.it/microdati/aspetti-della-vita-quotidiana/. Accessed: 1 September 2025. Dataset

Kuper, H., et al. (2024). The association between disability and mortality: a mixed-methods study. The Lancet Public Health, 9(5), pp. e306-e315. Journal Article

Machado, R.J., van den Hout, A., Marra, G. (2021). Penalised maximum likelihood estimation in multi-state models for interval-censored data. Computational Statistics & Data Analysis, 153, pp. 107057. Journal Article

Officer, A., Posarac, A. (2011). World report on disability. World Health Organization, Geneva, (CH). Book

Robine, J., Jagger, C., Euro-REVES Group. (2003). Creating a coherent set of indicators to monitor health across Europe: the EuroREVES 2 project. European Journal of Public Health, 13(Suppl 3), pp. 6-14. Journal Article

Roma, E., Miglio, R. (2025). Disability trends in selected European countries: an Age-Period-Cohort analysis. Rivista Italiana di Economia Demografia e Statistica, pp. 199-210. Journal Article

SHARE-ERIC. (2024). Survey of Health, Ageing and Retirement in Europe (SHARE), Waves 1–9. Release version: 9.0.0. DOIs: 10.6103/SHARE.w1.900; 10.6103/SHARE.w2.900; 10.6103/SHARE.w4.900; 10.6103/SHARE.w5.900; 10.6103/SHARE.w6.900; 10.6103/SHARE.w7.900; 10.6103/SHARE.w8.900; 10.6103/SHARE.w9.900. Dataset

Smythe, T., Kuper, H. (2024). The association between disability and all-cause mortality in low-income and middle-income countries: a systematic review and meta-analysis. The Lancet Global Health, 12(5), pp. e756-e770. Journal Article

Van der Gaag, N., et al. (2015). A multistate model to project elderly disability in case of limited data. Demographic Research, 32, pp. 75-106. Journal Article

Van Imhoff, E. (1990). The exponential multidimensional demographic projection model. Mathematical Population Studies, 2(3), pp. 171-182. Journal Article