Estimation subject to reporting delays and incomplete event adjudication with applications to disability insurance

Oliver Lunding Sandqvist
PFA Pension and University of Copenhagen

Insurance Data Science Conference, Stockholm

June 17, 2024

Based on paper with **Kristian Buchardt** (AP Pension) and **Christian Furrer** (University of Copenhagen).

Overview

Problem outline

- **Goal**: Estimate individual pricing and reserving models for disability insurance products.
  - Lack of steady-state makes aggregate reserves problematic.
- Represent disability insurance schemes using multistate model.
  - Capture a priori known structure of payments and model intertemporal dependencies.
- Hazard rates characterize conditional distribution (needed for reserving) and can be estimated with censored data.
- **Problem**: Biased sampling due to reporting delays (IBNR) and incomplete event adjudication (RBNS).

  $\Rightarrow$ Fitting model directly to observed data leads to severe bias!

Where we are going...

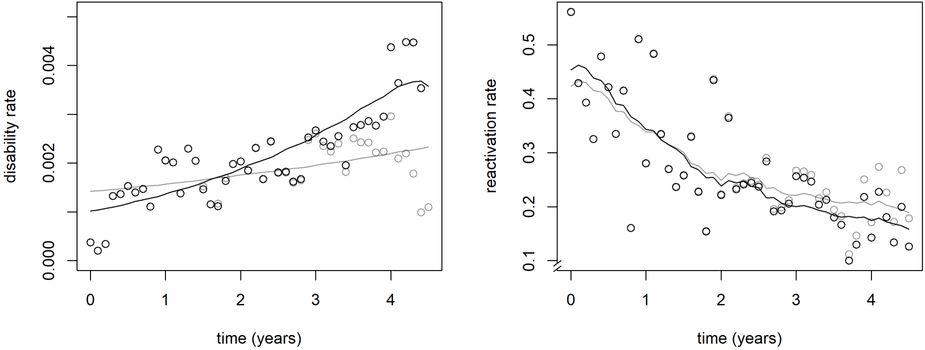
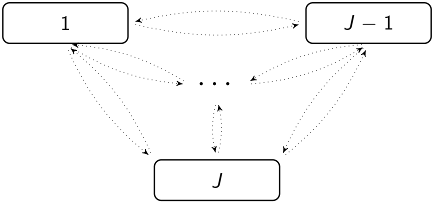


Figure 1: Fitted rates (lines) and occurrence-exposure rates (points) for the proposed method (black) and the naive method (gray). Disability rates are shown on the left and reactivation rates on the right.

Classic multi-state model

- State process $\{Y(s)\}_{s\geq 0}$: When insurance events occur.
- State space for Y:



- Counting process representation
  $N_{jk}(t) = \#\{s \leq t : Y(s-) = j, Y(s) = k\}$.
- Marked point process representation $(T_m, Y_m)_{m\geq 1}$.

Classic multi-state estimation

- Statistical model: Parametric intensity of $N_{jk}(s)$:

$$s \mapsto \mu_{jk}(s, \theta), \quad \theta \in \mathbb{R}^d.$$

- Maximum likelihood estimator (MLE) with discretization $0 = t_0 < t_1 < \cdots < t_I = t$:
  - Occurrences and exposures:

$$O_{jk}(t_i) = N_{jk}(t_{i+1}) - N_{jk}(t_i),$$
$$E_j(t_i) = \int_{t_i}^{t_{i+1}} 1\{Y(s) = j\} \, \mathrm{d}s.$$

  - Input $(O_{jk}(t_i))_{j,k,i}$ as independent Poisson observations with mean $(\mu_{jk}(t_i, \theta) E_j(t_i))_{j,k,i}$ (see Lindsey (1995)).

Research problem

Problem: $\{Y(s)\}_{s \leq t}$ not available at time t due to reporting and processing delays $\Rightarrow$ MLE cannot be used directly.

Focus of presentation:

- Illustration of the problem.
- Outline of mathematical approach & literature.
- Overview of results.
- Data applications.

Classic multi-state model

Main example: Disability insurance with reactivation.

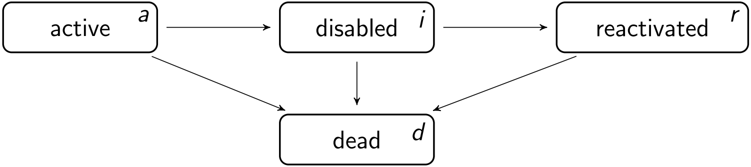
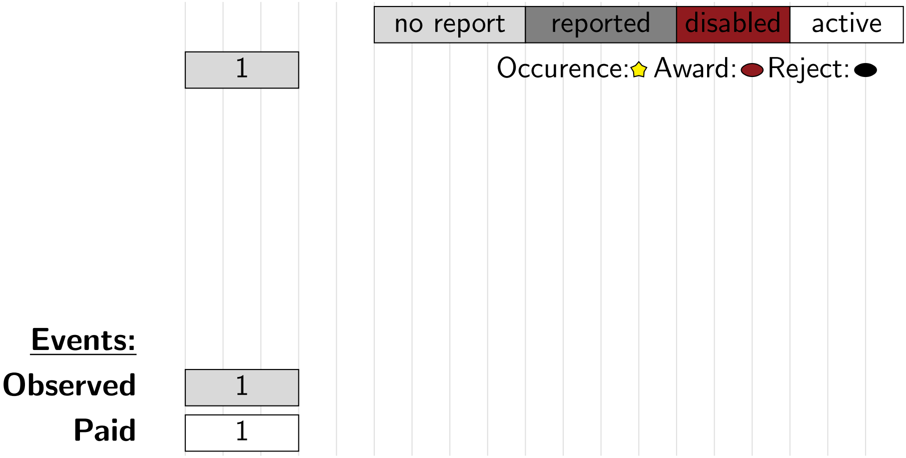


Figure 2: State space for state process Y.

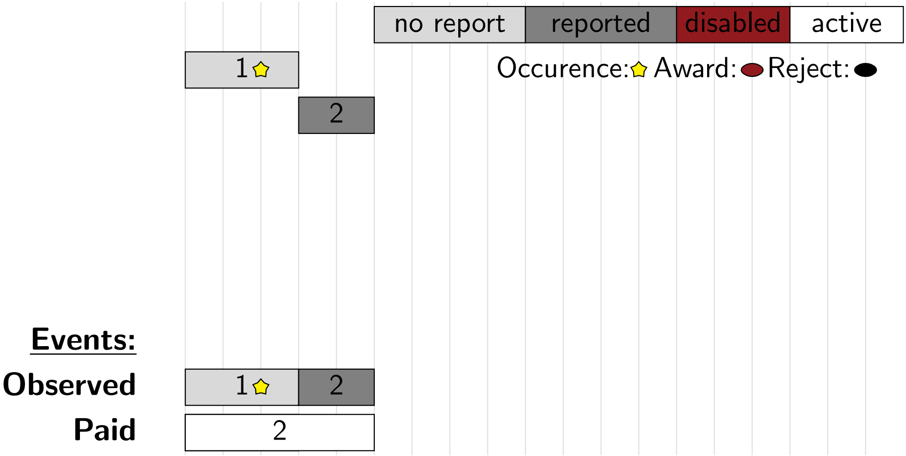
Available information

- Example: Timeline for one insured.



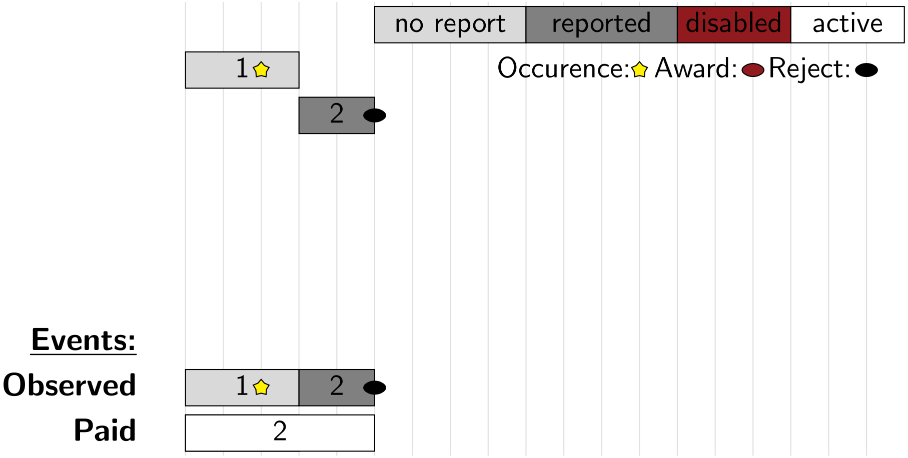
Available information

- Example: Timeline for one insured.



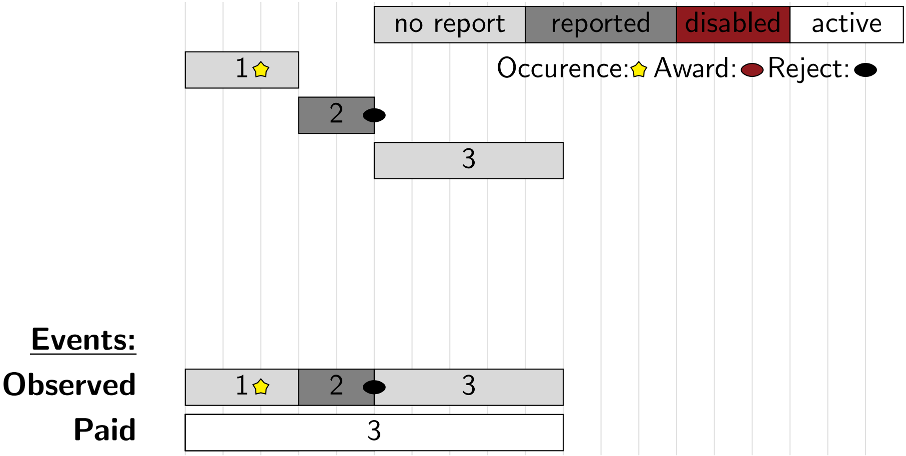
Available information

- Example: Timeline for one insured.



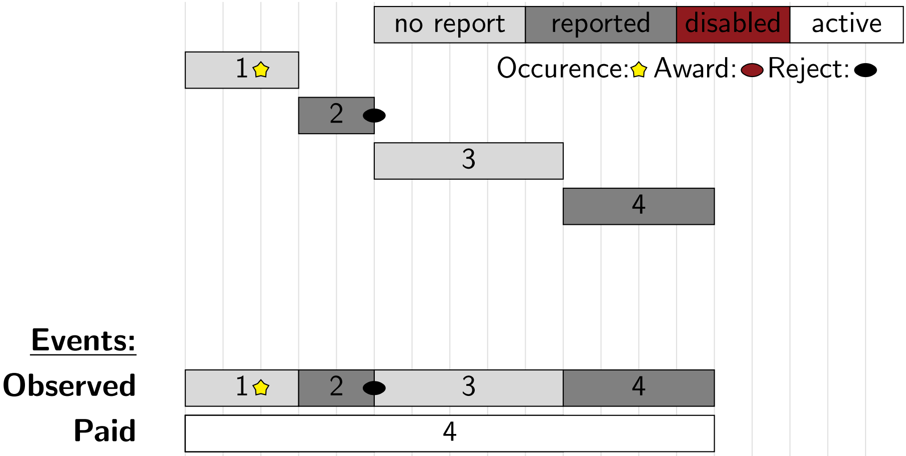
Available information

- Example: Timeline for one insured.



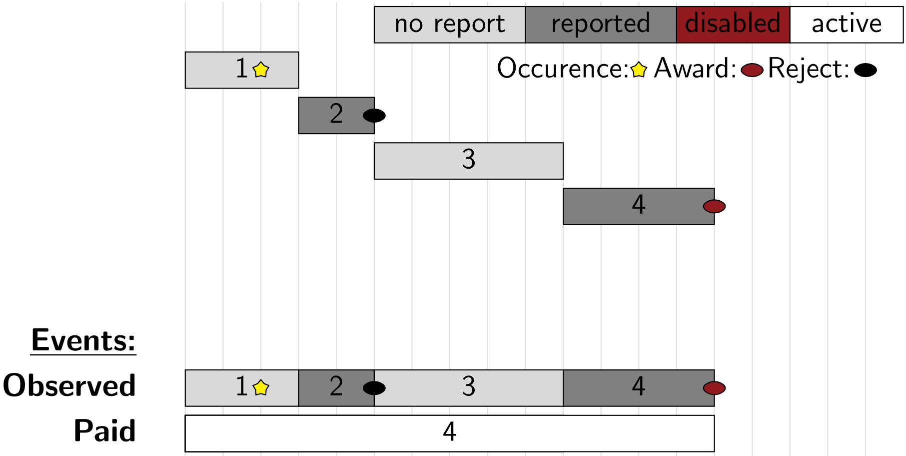
Available information

- Example: Timeline for one insured.



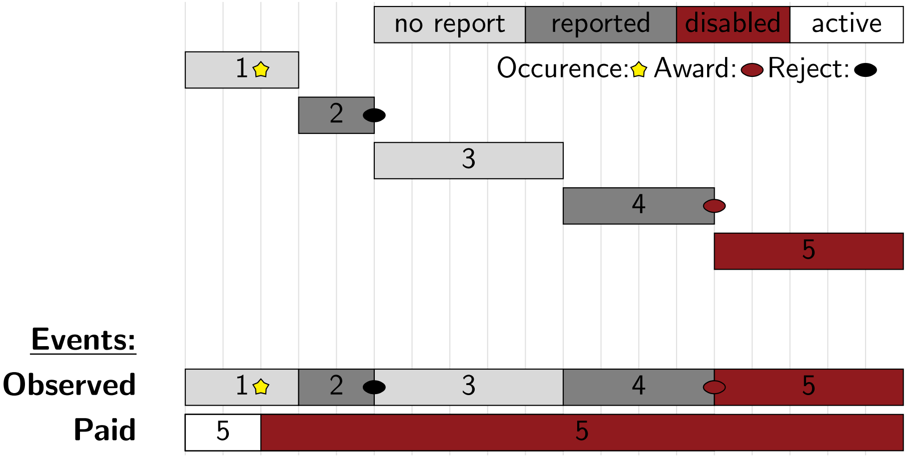
Available information

- Example: Timeline for one insured.



Available information

- Example: Timeline for one insured.



How to estimate?

- **Question**: How to estimate θ for $\mu_{jk}(s, \theta)$ based on observed information?
- **Naive approach**: Use old data (backcensoring).
- **Our approach**: Derive estimators $\hat{\theta}$ under IBNR and RBNS contaminated data for multi-state models.

Pros and cons over naive approach

Advantages

- More efficient use of data $\Rightarrow$ less estimation risk.
- Use new data faster $\Rightarrow$ capture new trends.
- Estimates of IBNR and RBNS applies to reserving.

Disadvantages

- Additional model elements $\Rightarrow$ added estimation and model risk.
- Requires detailed data.
- Slightly more complicated to implement.

Contributions

Contributions:

- Handle reporting delays for general multi-state model.
    - Insurance literature: Antonio & Plat (2014) and Bücher & Rosenstock (2024) for Marked Poisson process; Badescu et al. (2016,2019) for Marked Cox process.
- Handle incomplete event adjudication for hazard estimation, general multi-state model, and dynamical conditioning in adjudication probabilities.
    - Literature: Cook & Kosorok (2004), Bladt & Furrer (2023).
- Simultaneous treatment of reporting delays and incomplete event adjudication.
- Large-sample properties of the estimators.

Estimator construction (simplified)

- First estimate conditional reporting delay and adjudication probabilities.
- Estimator $\hat{\theta}$ can then (approximately) be based on observed occurrences $O_{jk}^{\text{obs}}(t_i)$ and exposures $E_j^{\text{obs}}(t_i)$ after modifying as follows:

$$O_{jk}(t_i) \leftarrow O_{jk}^{\text{obs}}(t_i) \times \hat{P}(\text{Confirm } O_{jk}^{\text{obs}}(t_i) \mid \mathcal{F}_t^{\text{obs}}),$$

$$E_{jk}(t_i) \leftarrow E_j^{\text{obs}}(t_i) \times \hat{P}(\text{Reporting delay} < t - t_i \mid \{Y(s)\}_{s<t_i}, Y(t_i) = k).$$

- Large-sample properties: Consistency, asymptotic normality, and Efron's simple nonparametric bootstrap is valid.

Simulation study

400 samples of size $n = 1500$ with $t = 5$.

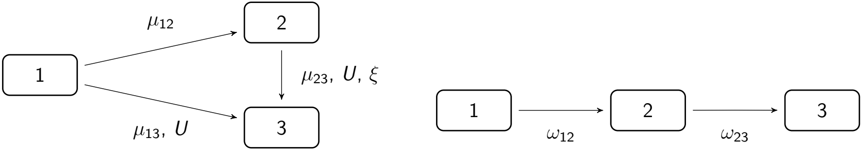


Figure 3: Event history model (left) and adjudication model (right). Symbols U and ξ indicate the presence of reporting delays and adjudication processes, respectively.

Setup:

- Moderately large transition rates.
- Reporting delays with mean 1.
- Confirm 40% of jumps.

Simulation results

	Proposed method		Oracle		Approximation		Naive	
Parameter	Bias	SD	Bias	SD	Bias	SD	Bias	SD
$\theta_1 = \log 0.15$	-.004	.067	-.008	.031	-.010	.067	-.010	.066
$\theta_2 = 0.1$	-.000	.020	-.001	.020	-.006	.020	-.006	.021
$\theta_3 = 0.4$	.003	.078	.003	.078	-.002	.078	-.000	.079
$\theta_4 = \log 0.1$	.003	.084	.001	.083	.012	.091	-.051	.082
$\theta_5 = 0.03$	.000	.012	-.000	.013	-.006	.016	-.015	.014
$\theta_6 = -0.3$	-.000	.094	-.001	.088	.007	.094	-.007	.090
$\theta_7 = -0.3$	-.011	.066	-.011	.054	-.012	.066	.148	.069

Table 1: Bias and empirical standard deviation (SD) of the estimator $\hat{\boldsymbol{\theta}}_n$ based on 400 simulations of size $n = 1500$.

Overall:

- Bias: Oracle=Proposed method<Approximation≪Naive.
- SD: Oracle<Proposed method=Approximation=Naive.

Application to real data: Model

Disability insurance data.

- Disability exposure and occurrences.
- Reactivation exposure and occurrences.
- Disability reporting delays.
- Adjudications.

Time window $[0, t]$ is $[31/01/2015, 01/09/2019]$.

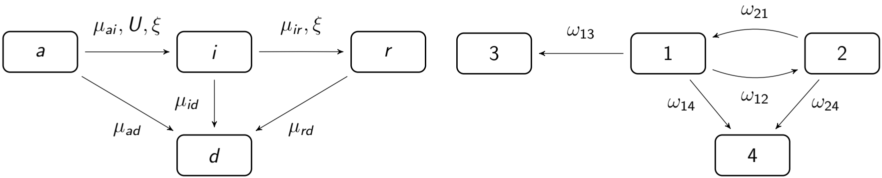


Figure 4: Event history model (left) and adjudication model (right). For events, active is a, disabled is i, reactivated is r, and dead is d. For adjudications, active report is 1, inactive report is 2, adjudicated is 3, and dead is 4.

Application to real data: Results

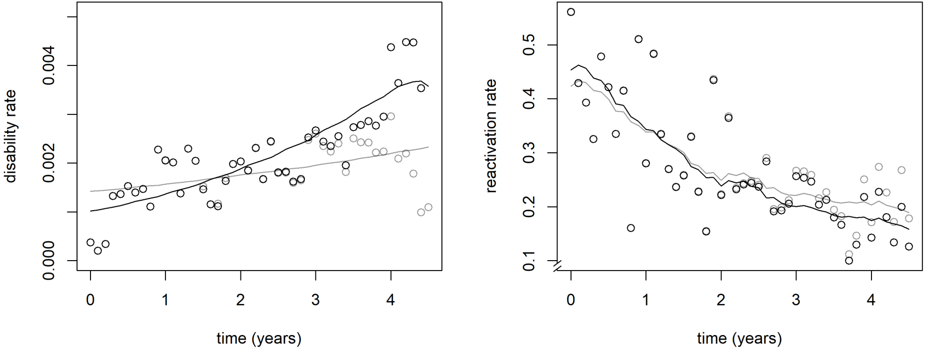


Figure 5: Fitted rates (lines) and occurrence-exposure rates (points) for the proposed method (black) and the naive method (gray). Disability rates are shown on the left and reactivation rates on the right.

Related paper: Reserving

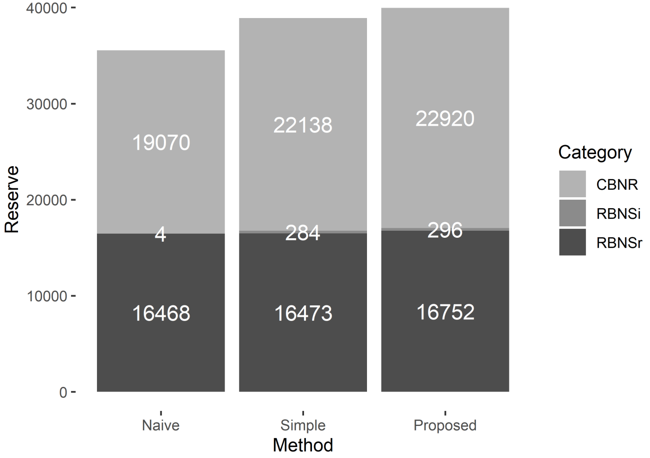


Figure 6: Portfolio level reserve decomposed by category.

References I

[1] J.K. Lindsey (1995).
Fitting parametric counting processes by using log-linear models.
Journal of the Royal Statistical Society: Series C, 44(2):201–212.

[2] T.D. Cook and M.R. Kosorok (2004).
Analysis of time-to-event data with incomplete event adjudication.
Journal of the american statistical association, 99(468):1140–1152.

[3] K. Antonio and R. Plat (2014).
Micro-level stochastic loss reserving for general insurance.
Scandinavian Actuarial Journal, 2014(7):649–669.

References II

[4] A.L. Badescu, X.S. Lin, and D. Tang (2016).
A marked Cox model for the number of IBNR claims: Theory.
Insurance: Mathematics and Economics, 69:29–37.

[5] A.L. Badescu, X.S. Lin, and D. Tang (2019).
A Marked Cox Model for the Number of IBNR Claims: Estimation and Application.

[6] M. Bladt and C. Furrer (2024).
Expert Kaplan–Meier estimation.
Scandinavian Actuarial Journal, 2024(1):1–27.

References III

[7] A. Büche and A. Rosenstock (2024).
Combined modelling of micro-level outstanding claim counts and individual claim frequencies in non-life insurance.
European Actuarial Journal, 2024:1–33.

Thank you for your attention!