SOA Research INSTITUTE
SOCIETAL PURPOSE

Applying Fairness Frameworks to Long-Term Care Insurance: Actuarial Considerations for AI and a Pricing Example

MAY | 2026

Aging and Retirement

Actuarial Innovation and Technology

Applying Fairness Frameworks to Long-Term Care Insurance

Actuarial Considerations for AI and a Pricing Example

AUTHORS Hong Beng Lim, ASA, PhD

Mengyi Xu, FSA, FIAA, PhD

Kenneth Q. Zhou, FSA, ACIA, PhD

CONTENTS

Applying Fairness Frameworks to Long-Term Care Insurance

Actuarial Considerations for AI and a Pricing Example

Regulatory Context

In the United States, insurance is primarily regulated at the state level. Several states have introduced or expanded expectations related to fairness and nondiscrimination, particularly as insurers adopt advanced analytics and artificial intelligence in underwriting and pricing. These requirements may include demonstrating that models and processes do not produce unfairly discriminatory outcomes across protected classes, even when such characteristics are not explicitly used as inputs. As a result, insurers must increasingly incorporate formal testing and governance to assess and document fairness in their pricing and related decision-making processes.

Executive Summary

Population aging and rising demand for long-term care are creating new challenges for insurers, regulators, and policyholders alike, including the sustainability of Long-Term Care Insurance (LTCI) products and growing scrutiny of potential unintended disparities across demographic groups. As the need for more long-term care financing grows, insurers are increasingly adopting data-driven approaches to pricing and risk assessment, including in the context of LTCI. At the same time, expectations around fairness and transparency in insurance practices have intensified. Traditional actuarial methods, while grounded in sound statistical principles, may potentially face challenges in the future if model outcomes reveal unintended, potential disparities across demographic groups.

To help avoid potential disparities across groups, this report examines how fairness-oriented modeling techniques, widely discussed in data science and emerging actuarial research, can be adapted to LTCI pricing. The analysis follows the recent work from the Society of Actuaries Research Institute (SOA) that highlights the evolving regulatory and methodological landscape around algorithmic fairness (Schraub et al., 2024). Rather than proposing prescriptive solutions, this report provides a structured overview of fairness methodologies and demonstrates one post-processing method for LTCI pricing, selected for illustration as it can be applied without altering the underlying actual model. This choice is illustrative only and does not imply a preference for this method over other fairness approaches, which may also be applicable in different settings.

The study uses data from the University of Michigan Health and Retirement Study (HRS) to evaluate potential fairness adjustments in an LTCI pricing framework. The goal is to inform the actuarial community about fairness approaches discussed in recent research and to illustrate, through a practical example, how a fairness adjustment developed in other modeling contexts could be incorporated into LTCI pricing workflows, if such considerations become relevant. The findings are intended to support future discussions on how fairness methods can complement traditional actuarial methods and inform ongoing research in insurance design, without implying that existing practices are inadequate or unfair.

Part I reviews the conceptual foundations of fairness in actuarial science and methodological categories that have been discussed. The discussion outlines how each category addresses potential bias at different stages of the modeling process and evaluates their suitability for hypothetical LTCI pricing applications.

Special attention is given to the unique characteristics of LTCI products, including their multi-state nature, long-term horizons, and dependence on demographic and health transitions. These features introduce distinct challenges for implementing potential fairness adjustments compared with short-term products such as auto or home insurance.

Fair pricing methodologies are generally categorized by the stage at which fairness adjustments are applied relative to model fitting: pre-processing (before fitting), in-processing (during fitting), or post-processing (after fitting). Among the three methodological categories, post-processing methods are identified as the most transparent and adaptable for the hypothetical LTCI pricing setting, as they allow potential fairness adjustments to be applied directly to modeling outcomes without re-estimating the underlying model.

Part II illustrates how fairness concepts can be applied in a hypothetical LTCI pricing setting using a post-processing framework. Based on data from the HRS, the case study models health transitions among the healthy, disabled, and deceased states within a multi-state actuarial model. These transition estimates form the foundation for pricing representative LTCI products.

The case study then applies the reweighting method of Lindholm et al. (2022)—as an example of a post-processing methodology – to adjust model outputs to remove unintended associations. This approach introduces potential fairness adjustments at the pricing stage without changing the underlying model. Several scenarios are examined by varying outcomes, reference distributions, and product features such as death benefit and payment structure.

The study's results show that the reweighting method can reduce potential unintended pricing disparities between demographic groups and that these results are robust across the various scenarios considered. The findings also highlight practical challenges, including data limitations, model assumptions, and predictive precision limitations. Importantly, this study demonstrates the feasibility of incorporating potential fairness adjustments with minimal disruption to actuarial workflows.

Section 1 Introduction

The aging of the population, especially in developed countries, is reshaping the long-term care landscape. By 2030, one in six people worldwide will be aged 60 or older, up from one in eight in 2020; the total number of older adults is also projected to reach 1.4 billion, up from 1 billion in 2020 (World Health Organization, 2025). With this demographic shift comes a growing need for long-term care, as nearly 70% of adults aged 65 and above are expected to require some form of support during their lifetime, and about 20% will need care for more than five years (Administration for Community Living, 2020a).

The difficulty of planning for large and uncertain long-term care expenses, particularly in later life, has increased financial pressure on individuals and families. These challenges have also created demand for sustainable insurance solutions that can help manage the costs associated with aging. Long-term care insurance (LTCI) has therefore become an important financial protection tool, allowing policyholders to transfer part of the care-related cost risk to insurers while maintaining autonomy, choice, and dignity in care.

As LTCI products evolve, the industry faces new expectations related to fairness in insurance practices. Regulatory frameworks already prohibit the use of certain characteristics in actuarial pricing, such as race in the United States (Campbell et al., 2020) and sex in the European Union (European Commission, 2012). However, as pricing models become increasingly data-driven and incorporate a wider range of socioeconomic and behavioral information, ensuring that outcomes remain free of unintended bias has become a growing concern.

Simply excluding protected variables does not guarantee that pricing outcomes are fair or compliant. Patterns present in historical data or in correlated variables can reproduce potential inequities even when prohibited factors are excluded. The SOA explored these issues in its 2024 report on non-discriminatory AI use in insurance (Schraub et al., 2024), while the Casualty Actuarial Society (CAS) published a companion report in 2024 that outlines practical considerations for promoting fairness in pricing models (Leong et al., 2024).

The actuarial community increasingly recognizes that fairness in insurance is both complex and context specific. An SOA study in 2022 on avoiding unfair bias in AI applications discussed the need to balance regulatory compliance, actuarial soundness, and social responsibility (Smith et al., 2022). A subsequent SOA report in 2023 examined fairness concepts in life insurance and noted that fairness may take different forms across underwriting, pricing, and claims processes (Schilling, 2023). Together, these studies suggest that fairness can be considered from several perspectives, including those of policyholders, insurers, and society.

For long-term products such as LTCI, where pricing depends on multi-state health transitions and long time horizons, fairness considerations are further shaped by mortality dynamics, health conditions, and socioeconomic differences. These factors influence both claim patterns and access to coverage, making it difficult to distinguish between legitimate risk differentiation and potential unfair disparities. The CAS 2022 report on defining discrimination in insurance noted that definitions vary across jurisdictions and industry practices, emphasizing the importance of clear concepts and consistent interpretation when evaluating fairness in actuarial work (Chibanda, 2022).

This report contributes to this broader discussion by examining how fairness-oriented methods developed in recent theoretical academic research can be applied in the context of LTCI pricing methodologies. Existing approaches to fairness in actuarial modeling are typically designed for products with a single short-term outcome, such as auto or health claims. In contrast, LTCI relies on models that capture multiple health states and transition intensities, which together determine the frequency and duration of benefits. Extending fairness concepts developed for short-term products to an LTCI framework requires careful consideration.

The objective of this report is twofold. The first is to summarize and compare methodological approaches to fairness in insurance pricing and to evaluate their relevance to long-term care insurance. The second is to demonstrate how these ideas can be translated into a data-driven framework that supports transparent and replicable analyses of LTCI pricing methodologies. The framework presented here is intended to help illustrate how newer fairness methods may interact with traditional actuarial techniques while maintaining predictive accuracy and consistency with established pricing practices.

Aligned with this objective, the report is organized into two main parts. Part I reviews existing fairness methodologies from actuarial and data science literature and outlines their potentials and limitations when applied to multi-state LTCI models. Part II develops a post-processing framework that applies potential fairness adjustments to transition rates or hypothetical premiums derived from an LTCI model. Using data from the HRS, the report demonstrates how this framework can be used in an actuarially consistent pricing outcome.

Section 2 Part I: Fairness Frameworks and Actuarial Considerations

This part of the report reviews recent developments related to fairness and their relevance to LTCI. The intent is to provide an overview of fairness concepts most commonly used in the fair pricing literature, highlight how these ideas may intersect with LTCI pricing methodologies, and identify the unique challenges that may arise when applying fairness considerations in practice. The discussion is intended to inform practitioners, regulators, and researchers about both the potential opportunities and potential practical limitations of incorporating these fairness principles into LTCI pricing methods.

2.1 DEFINITIONS

This subsection introduces the key terminology and conceptual framework used throughout the report. It first defines the notation used to describe data, models, and outcomes. It then summarizes commonly discussed notions of fairness and explains how fairness can be evaluated within insurance processes. Lastly, it outlines a commonly used methodological categorization—into pre-processing, in-processing, and post-processing—of fairness-oriented adjustments in actuarial pricing models.

2.1.1 NOTATION AND SETUP

The following notation describes a few key elements used to assess fairness within actuarial modeling and pricing applications.

- Y: the response variable representing the incurred loss or the present value of future claims.
- $\hat{Y}$: the predicted value or model-based premium estimate, i.e., the best estimate of the expected loss. derived from the pricing model.
- $\mathbf{Z}$: a vector of observable characteristics used to explain variation in expected losses.
- S: the attribute used for fairness evaluation, set to be one-dimensional and categorical in this study.

Together, $(Y, \mathbf{Z}, S)$ define the data structure for fairness assessment. The goal is to determine whether differences in $\hat{Y}$ across levels of S are consistent with actuarially justified variations in $\mathbf{Z}$ or indicate potential bias in model estimation.

In an LTCI pricing context, Y may represent the present value of future benefit payments, and $\hat{Y}$ the model-predicted premium. The covariate set

$$\mathbf{Z} = (\text{Age}, \text{Education}, \text{Job Physicality}, \text{Marital Status}, \text{Health Indicators}, \cdots)$$

captures observable factors that drive claim incidence and duration. The tested attribute S might represent any other attribute, which could be a sensitive attribute or legally protected class. While there are a variety of fairness definitions, for this study, fairness evaluation examines whether two applicants with similar $\mathbf{Z}$ values but different S categories are charged comparable premiums, and whether any systematic differences in $\widehat{\mathrm{Y}}$ can be explained by valid actuarial risk factors rather than by correlation between S and $\mathbf{Z}$.

2.1.2 NOTIONS OF FAIRNESS

Notions of fairness can be broadly divided into two categories: group fairness and individual fairness. This paper discusses each of these in turn.

Notions of **group fairness**, broadly speaking, require the predictions to be equal in some sense for the groups of interest. To introduce the three most commonly used notions of group fairness, consider a setting where a bank approves loans only if it believes that the applicant can repay in full. In this case, the outcome Y is a binary variable, which takes value 1 if the applicant can repay in full, and 0 otherwise. The bank's decision is represented by the prediction $\hat{Y}$, which is also binary and takes the value of 1 if the loan is approved, and 0 if it is denied. Suppose further that the bank wishes to assess whether its loan approval decisions are fair with respect to a specific applicant attribute, denoted by S. The three notions of group fairness are as follows:

- **Demographic parity:** Also known as *independence*, it requires that the prediction be independent of the attribute of interest, i.e., $\hat{Y} \perp S$, where $\perp$ denotes independence. In this example, the requirement is for loan approval rates to be equal across the groups defined by S.
- **Equalized odds:** Also known as *separation*, it requires that, conditional on the value of the outcome, the predictions be independent of the attribute of interest, i.e., $\hat{Y} \perp S|Y$. In this example, the requirement is that, among applicants who can repay in full, the approval rates are equal across groups defined by S, and likewise among applicants who are unable to repay in full.
- **Predictive parity:** Also known as *sufficiency*, it requires that, conditional on the value of the predictions, the outcome variable to be independent of the attribute of interest, i.e., $Y \perp S|\hat{Y}$. In this example, the requirement is that, among applicants who have been approved for a loan, the proportion who are truly capable of repaying it in full is equal across groups defined by S, and that the same holds among applicants who have not been approved for a loan.

The notions of equalized odds and predictive parity require conditioning on Y and $\hat{Y}$ respectively. This conditioning makes sense when the outcome and predictions relate to a classification problem: that is, when both Y and $\hat{Y}$ are categorical variables. As illustrated by this example, such a setting is common in the machine learning literature. However, in the actuarial context, Y and $\hat{Y}$, representing the actual incurred loss and premium respectively, are continuous in nature, making these notions of fairness far more cumbersome to utilize and less relevant. Thus, this paper focuses discussion on the actuarial context, and on the notion of demographic parity.

The requirement that the premiums are independent of the attribute of interest, i.e., $\hat{Y} \perp S$ is usually known as **strong demographic parity**. It other words, it requires that the distribution of the premiums be equal for each of the groups of interest. However, since equality in distribution is not necessarily easy to enforce, one may also consider a more practical notion known as **weak demographic parity**. This notion only requires that the *average* premiums be equal for each group of interest, i.e., for $\mathbb{E}\left[\hat{Y}\middle|S=s\right]$ be constant in s.

Notably, weak demographic parity is difficult to apply in the actuarial context, as it can be overly broad and may lead to unintended cross-subsidization, whereby one group subsidizes another. A commonly considered refinement is **conditional demographic parity**, which modifies weak demographic parity by requiring conditioning on additional covariates. For instance, a health insurer may seek equality of average premiums across categories of a given attribute by conditioning on smoking status, i.e., that average premiums be equal across categories among smokers, and likewise among nonsmokers. Nevertheless, it remains challenging to utilize this notion of fairness, as the choice of which covariates or risk classes to condition on is inherently subjective.

Turn to the notions of **individual fairness** first coined by Dwork et al. (2012). Broadly speaking, individual fairness requires that individuals with similar characteristics be given similar treatment: see Definition 2.1 in their paper for a precise definition. The difficulty in utilizing this notion of fairness has been in defining the "similarity" metric. Given this difficulty, most existing research into individual fairness has been in proposing intuitive measures of fairness which happen to fall under this categorization, rather than proposing an appropriate metric.

In particular, the familiar notion of **actuarial fairness** constitutes a form of individual fairness. It says that a pricing methodology is **actuarially fair** if it charges the same premium for all individuals sharing the same expected loss "unfairly discriminatory" if

"... allowing for practical limitations, there are premium differences that do not correspond to expected losses and average expenses or if there are expected average cost differences that are not reflected in premium differences." (Williams, quoted in Hoy and Ruse (2005)) "

Mathematically speaking, suppose there are two risks Y_1 and Y_2, with respective premiums $\hat{Y}_1$ and $\hat{Y}_2$. Actuarial fairness requires that, provided that the expected values of the two risks are equal, their premiums should also be equal. Another notion, defined by Lindholm et al. (2022), is what this paper shall call **unintentional indirect association**.

To illustrate, consider a health insurer operating in a jurisdiction where regulations prohibit using a particular attribute in ratemaking. As an extreme example, suppose the insurer's portfolio contains only two insureds: a younger individual and an older individual, each belonging to different groups defined by the particular attribute. In this case, age is perfectly confounded with the particular attribute, such that pricing based on age alone implicitly reflects information about the particular attribute. Although such perfect confounding is unlikely in practice, the association between age and the particular attribute can nevertheless allow information from the latter to be partially reflected in age-based premiums. This unintended reflection of information from variables excluded from the model is the phenomenon of interest in Lindholm et al. (2022). For more mathematical details on this notion of fairness, please see Appendix A.1. Section 2.3.3 describes this phenomenon more rigorously and outlines their proposed procedure to remove this unintended association.

2.1.3 CATEGORIES OF METHODOLOGIES

Methodologies that address fairness in predictive modeling can be organized into three broad categories, according to the stage at which adjustments are applied: (1) pre-processing, (2) in-processing, and (3) post-processing. Each category represents a distinct point of intervention within the workflow:

1. **Pre-Processing:** Methods that adjust or transform the input data before model estimation. These include removing or re-balancing information correlated with the tested attributes or redefining rating variables to mitigate indirect dependencies. Such approaches are relatively straightforward to implement but may reduce the predictive power and accuracy of the model if applied without careful consideration of context.
2. **In-Processing:** Methods that incorporate fairness constraints or objectives directly within model estimation process. These may modify the model's loss function or optimization routine to balance predictive accuracy and fairness metrics. In-processing approaches tend to be model-specific, computationally intensive, and highly context-dependent, but can provide strong control over bias propagation.
3. **Post-Processing:** Methods that adjust model outputs, such as rates, amounts, or premiums, after model fitting. Post-processing typically applies fairness constraints to final predictions or pricing rates while maintaining the core predictive structure of the underlying model. These approaches are flexible and transparent, allowing regulators and practitioners to evaluate potential fairness adjustments independently of model development.

These three methodological classes collectively form a practical foundation for evaluating and applying fairness interventions in actuarial pricing. In later sections, representative examples from each category are reviewed in more detail with specific reference to LTCI applications.

2.2 WHY LTCI IS UNIQUE

There are several ways in which the setting of long-term care insurance is unique, relative to other insurance products on which the fairness methods discussed herein may be applied. One distinguishing feature is the need to use multi-state models to coherently price such products. LTCI reimburses insureds for long-term care services rendered or provides a pre-set cash amount for each day the insureds meet the benefit trigger. Typically, such payments will continue for as long as the insured is in a functionally disabled state but pause or terminate when the insured recovers sufficiently or dies. Further, premiums are typically paid only while the insured is healthy and pause while the insured is functionally disabled. Multi-state models are hence needed to calculate the actuarial present values of both the benefits received and premiums paid.

Thus, unlike in the standard regression framework, the price does not correspond immediately to any single output of the multi-state model, complicating the use of the fair pricing methods discussed herein. From the fitted model, the cumulative hazard function for any given individual can be obtained; these are then used to calculate transition probabilities, which finally are used to inform an expected present value calculation; the premium is obtained as a solution from this calculation. The complex dependence of the premium on the model output clearly poses an issue for fair pricing methods that rely on the fitted regression model having the price as an immediate output. Thus, much care has to be taken to apply the fair pricing methods discussed herein to extend to this much more intricate setting.

Finally, the time horizon also distinguishes LTCI from other insurance products. While short-term insurance products such as health and property/casualty insurance tend to be renewed over time, premiums are charged on an ongoing basis, and insurers generally have more flexibility to incorporate changes to better reflect the insured's current expected loss. In contrast, LTCI policies often span decades, with the price dictated far in advance, and much more care has to be taken in handling time-dependent factors.

2.3 REPRESENTATIVE METHODS

This subsection presents three representative examples for incorporating fairness into actuarial modeling, drawn from the methodological categories of pre-processing, in-processing, and post-processing. The objective is to clarify how each category of techniques addresses fairness considerations and to identify the most practical and transparent approach for LTCI applications.

2.3.1 PRE-PROCESSING: REMOVING BIAS IN DATA

Of the three modes of processing, pre-processing takes place the earliest in the process—before the model is trained. Specifically, such methods usually transform the input data, specifically how the response variable and covariates are distributed with respect to the tested attribute. The transformation is designed to encourage any model trained on the transformed data to better fulfill some intended notion of fairness.

An example of a pre-processing fair pricing methodology is that of Lindholm et al. (2024). Recall that strong demographic parity requires that $\hat{Y}$—predictions, or premiums in the actuarial context—be independent of the attribute with respect to which fairness is desired, S. If the input covariates $\mathbf{Z}$ were independent of S, then the predictions or premiums—as functions of $\mathbf{Z}$—would also be independent of S. Thus, this pre-processing methodology functions by artificially modifying values of $\mathbf{Z}$ such that their modified values are independent of S, and this is achieved using optimal transport methods.

For instance, consider the previous health insurance example, but instead suppose that there are two younger policyholders of the same age from one group defined by the particular attribute, and two older policyholders of the same age from another group. Because age and the particular attribute are not independent, demographic parity is not guaranteed. The methodology of Lindholm et al. (2024) addresses this by artificially adjusting the policyholders' ages such that each group has one younger and one older policyholder. This makes age independent of the particular attribute, i.e., $\mathbf{Z}$ is independent of S, which guarantees demographic parity in the predictions. For more complex distributions, the solution is less straightforward, but the methodology operates similarly: one first identifies a candidate distribution in which $\mathbf{Z}$ is independent of S, and then transforms the observed data to match this distribution, with the transformation mapping identified via optimal transport methods. A more technical definition of this procedure is provided in Appendix A.2.

There are several advantages of using pre-processing methods. First of all, such methods aim to avoid potential unwanted bias in results by reducing potential unwanted bias in the data, hence mitigating any potential downstream issues, particularly when the dataset is used for purposes other than its initial intended purpose. More and more data is becoming available, and with how various modelers are scraping data from various sources for various purposes, it is becoming increasingly likely that available data could be used for different purposes than it was originally intended. As a result, to limit any potential harm by any downstream-trained models, it may be desirable to first perform pre-processing. Doing so could enable more data to be safely released for public use.

Nevertheless, there are several challenges to applying pre-processing methods. First, such methods distort the original distribution of the data and may mask important relationships of how covariates relate to the response variable. At the same time, notions of fairness often contradict each other. Transforming the data to satisfy one notion of fairness misdirects users to think that the data is universally fair and can be used for any purpose, when in fact it is fair only under that one notion of fairness. Finally, since such transformation tends to only take $(\mathbf{Z}, S)$ into account, Y is often omitted from the process, and hence a priori one is unable to gauge the effects a given transformation has on the predictive power of a subsequently trained model.

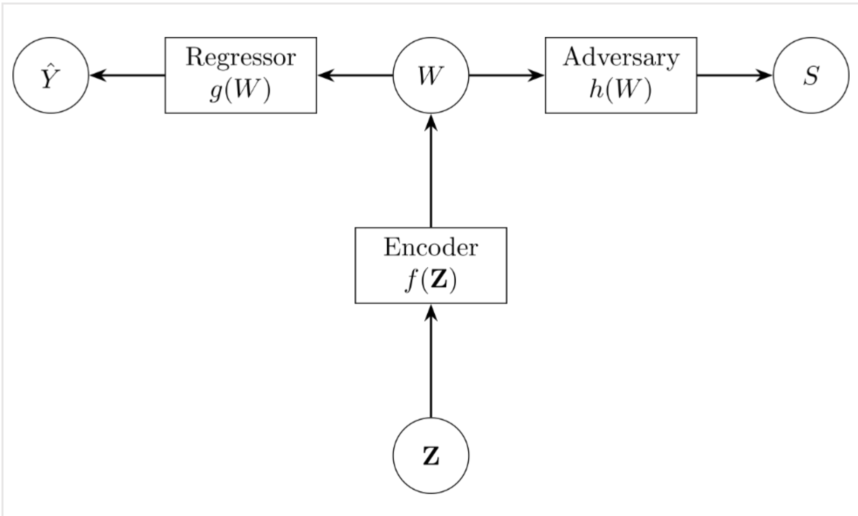
Some characteristics of LTCI also make pre-processing difficult to apply. First, pre-processing methods are not straightforward to apply to LTCI data due to the need to model transition rates in the multi-state model. Given that the price is a complicated function of the transition rates, it is *a priori* unclear whether transformations on the input covariates are sufficient to enforce the desired notion of fairness with respect to the prices. Secondly, pre-processing methods require complete data. On the other hand, existing LTCI data such as the HRS tend to be longitudinal surveys, thus missing data is likely prevalent either in certain variables or for specific subjects. Hence, the missing data issue has to be carefully handled before pre-processing methods can be applied.

2.3.2 IN-PROCESSING: BEING UNBIASED IN MODELING

Unlike pre-processing methods, in-processing methods are integrated into the model training process. Such methods encourage particular notions of fairness through mechanisms such as incorporating fairness metrics into the loss function, or by integrating fairness constraints into the model design.

One example of an in-processing methodology is that of Beutel et al. (2017) illustrated in Figure 1. This model sets up an intermediate representation of the input data, W, which is debiased during the training process. To achieve this, an adversarial learning framework is used to encourage the removal of any predictive value for S from W; in the ideal case, the two will become independent of each other. This independence from S in the ideal case carries over to any function of W. Thus, this model generates predictions as a function of W, written as $\hat{Y} = g(W)$ for some function g. Although perfect independence between W and S is rarely achieved in practice, adversarial learning still serves to encourage demographic parity. At the same time, to ensure functions of W remain sufficiently flexible for the prediction of Y, the prediction accuracy is explicitly included in the loss function used to train this network. Additional technical details of this procedure are provided in Appendix A.3.

Figure 1
THE ADVERSARIAL LEARNING FRAMEWORK OF BEUTEL ET AL. (2017)



The primary strength of in-processing methods is that the fairness objectives are built directly into the model, either through the loss function, the model design, or both; this direct approach enables model accuracy to be preserved.

However, this primary strength also serves as an obstacle to effectively utilizing such methods in the LTCI context. Given the complexity of the multi-state model, it is a priori unclear how one measures accuracy. Hence, careful theoretical justification is needed to identify a suitable choice. Further, if the object with respect to which accuracy is measured is too complex, e.g., a computed premium, each gradient step may become too computationally costly, hence the model may become infeasible to estimate.

2.3.3 POST-PROCESSING: ACHIEVING FAIRNESS IN PRICING

Post-processing methods operate after model estimation and adjust predicted values to remove undesirable dependencies from the tested attributes without re-estimating the underlying model. Among the three methodological approaches, they offer the most direct path to achieving fairness while maintaining the integrity of the estimation process. This approach is particularly appealing for actuarial applications where existing pricing frameworks must remain transparent and interpretable.

A representative example is the method proposed by Lindholm et al. (2022). In their setup, the true or best-estimate price of a policy is expressed as

$$\mu(\mathbf{z}, s) = \mathbb{E}[\, Y \mid \mathbf{Z} = \mathbf{z}, S = s \,].$$

The objective is then to derive a pricing function based on $\mu(\mathbf{Z}, S)$ that avoids both direct discrimination and unintentional indirect association (see Section 2.1.2 for the description of this notion) between $\mathbf{Z}$ and S. If the attribute S is completely omitted, then the outcome is an *unaware* or S-blind price,

$$\mu(\mathbf{z}) = \mathbb{E}[\, Y \mid \mathbf{Z} = \mathbf{z} \,],$$

which may still exhibit unintentional indirect associations if $\mathbf{Z}$ remains correlated with S. Lindholm et al. (2022) find that $\mu(\mathbf{z})$ indirectly factors in effects of S by implicitly making a "best guess" for the value of S based on the value of $\mathbf{Z}$. More precisely, this arises from the tower property of expectations:

$$\mathbb{E}[\,Y \mid \mathbf{Z} = \mathbf{z}\,] = \int_{S} \mathbb{E}[\,Y \mid \mathbf{Z} = \mathbf{z}, S = s\,]\, d\mathbb{P}(\,S = s \mid \mathbf{Z} = \mathbf{z}\,).$$

To address this, Lindholm et al. (2022) define the price as

$$h^{*}(\mathbf{z}) = \int \mu(\mathbf{z}, s)\ d\mathbb{P}^{*}(s),$$

where $\mathbb{P}^{*}(s)$ is a chosen distribution of the attribute S that is independent of $\mathbf{Z}$.

Intuitively speaking, $\mathbb{P}^{*}(s)$ replaces the empirical distribution of S given $\mathbf{Z}$, thereby removing any unintended influence of S that may persist through correlations with $\mathbf{Z}$. The resulting price $h^{*}(\mathbf{z})$ preserves the relationship between premiums and legitimate risk factors in $\mathbf{Z}$ while ensuring that pricing outcomes are not directly or indirectly influenced by the variable S.

From an actuarial perspective, this framework provides a structured way to isolate and remove unintentional indirect associations. The standard pricing function $\mu(\mathbf{z}, s)$ represents the best estimate of expected losses when all available variables are used, while the pricing function $h^{*}(\mathbf{z})$ explicitly corrects for residual dependence between $\mathbf{Z}$ and S, yielding a fair premium structure without altering the model estimation itself.

Revisiting the example for unintentional indirect association in Section 2.1.2, suppose now that the portfolio is much larger, such that age and the particular attribute are still correlated but no longer perfectly confounded. In this context, the methodology of Lindholm et al. (2022) could operate as follows:

- Estimate a pricing model that includes both age and the particular attribute as covariates.
- Compute marginal proportions of each group defined by the particular attribute, independent of age.
- For any given age x, obtain the group-specific prices at age x, and define the adjusted price as a weighted average of these prices, with weights given by the marginal proportions of the particular attribute.

For LTCI applications, post-processing methods such as that of Lindholm et al. (2022) are appealing because they can be integrated into existing actuarial workflows. The adjustments can be applied directly to pricing outcomes without redesigning complex actuarial models or re-estimating model parameters. However, successful implementation requires a sound understanding of LTCI pricing processes, population characteristics, and fairness objectives specific to this product line.

2.4 CONCLUDING REMARKS OF PART I

Part I of this report reviewed key fairness concepts, terminology, and three methodological approaches for addressing fairness in actuarial modeling: pre-processing, in-processing, and post-processing. While pre- and in-processing methods offer valuable theoretical insights and features, they present practical challenges when applied to LTCI. Post-processing methods, in contrast, provide a flexible and transparent framework that can be integrated into existing pricing workflows without altering model estimation.

The remainder of this report focuses on implementing a post-processing approach, given its capacity to apply potential fairness adjustments directly to pricing outcomes while preserving actuarial soundness and interpretability. This choice reflects the practical challenges identified throughout Part I, including the nature of LTCI, the data limitation, and the need to balance fairness with accuracy and feasibility. Specifically, it applies the reweighting framework of Lindholm et al. (2022) and evaluates its applicability to LTCI using the HRS dataset.

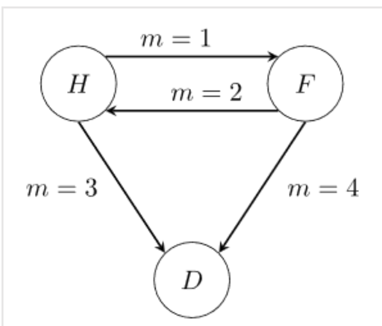
Section 3 Part II: Applying Post-Processing to LTCI Pricing

3.1 DATA OVERVIEW AND PREPARATION

LTCI benefits are paid when the policyholder becomes disabled and cease upon recovery to a healthy state or death. Accordingly, a three-state Markov process is commonly used to model and price LTCI policies (e.g., Fong et al., 2015b; Li et al., 2017). This paper adopts this framework to model transitions among three health states: healthy (H), functionally disabled (F), and dead (D). Health states are determined based on activities of daily living (ADLs) and cognitive function. The six ADLs include walking across a room, dressing, bathing, eating, getting in and out of bed, and using the toilet, while cognitive function is assessed using the Langa–Weir classification method (Langa et al., 2020). Consistent with the eligibility criteria of most LTCI policies, for this study, functional disability is defined as needing help with two or more ADLs or being cognitively impaired (Administration for Community Living, 2020b). Figure 2 illustrates the possible transitions among these health states.

Figure 2

THE THREE-STATE HEALTH STATE TRANSITION MODEL



H means healthy; F means functionally disabled or simply disabled; D means dead; m represents the type of transitions.

3.1.1 DATASET DESCRIPTION

The model for this study has been fit to the publicly available data from University of Michigan Health and Retirement Study (HRS). HRS is a biennial panel survey since 1992 of initially non-institutionalized Americans aged 50 and older. The extensive range of variables collected by HRS and its longitudinal design make this dataset particularly well-suited for modeling health state transitions and examining their associations with various covariates, relating to both the individual respondent and the household they live in. This paper analyzes data from 1998 to 2020 due to inconsistencies in survey questions on functional limitations before 1998 (Fong et al., 2015a). Note that while the HRS is not collected as an insurance dataset, it should still reveal information on transition rate patterns from the surveyed population. These estimated transition rates can be then used to price hypothetical LTCI products, assuming that the product is bought at the age that the participant entered the study.

To support the later proposed modeling approach, two datasets are constructed. The first dataset records the timing of health state transitions. While the date of death is directly provided by HRS, dates for other transitions have to be inferred. Specifically, the HRS follows respondents through biennial interviews, during which the interview date and contemporaneous information about each respondent are recorded. Non-death transitions can only be when a respondent's observed state changes between successive interviews. Because the exact timing of such transitions is unavailable, they are assumed to occur at the midpoint between two consecutive interviews.

The second dataset captures covariate information collected when individuals first entered the study, i.e., their initial interview, mirroring the underwriting process by recording baseline characteristics only at entry. These covariates include factors likely to influence disability and mortality, such as health behaviors (e.g., drinking and smoking), socioeconomic variables (e.g., education), as well as demographic attributes relevant to fairness considerations.

Table 1 lists the covariates included in this study, all of which are based on self-reported data. The dataset includes three categorical attributes (Attribute 1–3), each defined over a small number of groups. Financial variables are measured at the household level. Non-housing wealth includes both tangible assets (e.g., vehicles, businesses) and financial assets (e.g., retirement accounts, stocks, bonds, savings) net of non-mortgage debt. Household income reflects total income for the previous calendar year, including respondent and spouse earnings, pensions and annuities, Supplemental Security Income and Social Security Disability, Social Security retirement benefits, unemployment and workers' compensation, other government transfers, household capital income, and other income sources.

Table 1

LIST OF VARIABLES. C DENOTES CATEGORICAL VARIABLES, AND N DENOTES NUMERICAL VARIABLES

Demographic	Health	Finance
Age (N)	Self-rated health (C)	Non-housing wealth (N)
Attribute 1 (C)	Ever had diabetes (C)	Household income (N)
Attribute 2 (C)	Ever had lung disease (C)	
Attribute 3 (C)	Ever had heart problems (C)	
Years of education (C)	Ever had stroke (C)	
Census region (C)	Body mass index (BMI) (N)	
Marital status (C)	Ever drinks any alcohol (C)	
No. of household members (N)	Ever smoked cigarettes (C)	

The final step is to join the covariate dataset with the transition dataset. Note that the transition dataset contains multiple entries for each individual in the covariate dataset. Hence, the two datasets are joined by appending the relevant covariate information for each individual to the entries in the transition dataset, by using individual ID as the identifier.

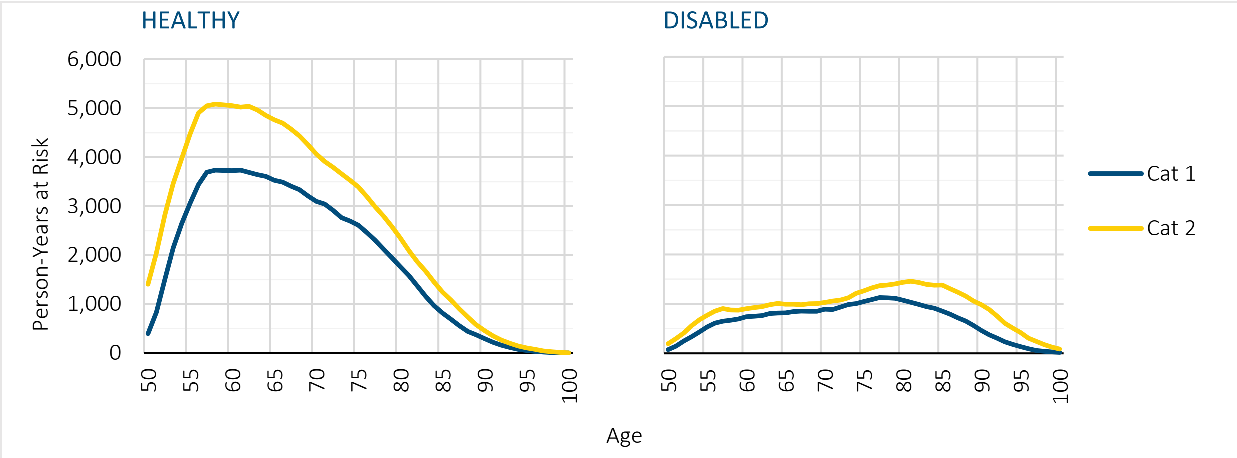
3.1.2 EXPLORATORY ANALYSIS

Figure 3 to Figure 6 provide an overview of the person-years at risk and the number of health state transitions in the sample. Note that ages begin at 50 due to the minimum age at entry for the HRS survey. As shown in Figure 3, there are differences in representation across categories of Attribute 1—more Category 2 than Category 1—particularly in the healthy state. The vast majority of person-years are spent in the healthy state, which gradually declines after age 60. In contrast, person-years in the disabled state increase steadily with age, peaking in the late 70s for Category 1 and the early 80s for Category 2.

Figure 4 presents person-years at risk by Attribute 2, indicating that Category 1 comprises the majority of the sample.

Figure 3

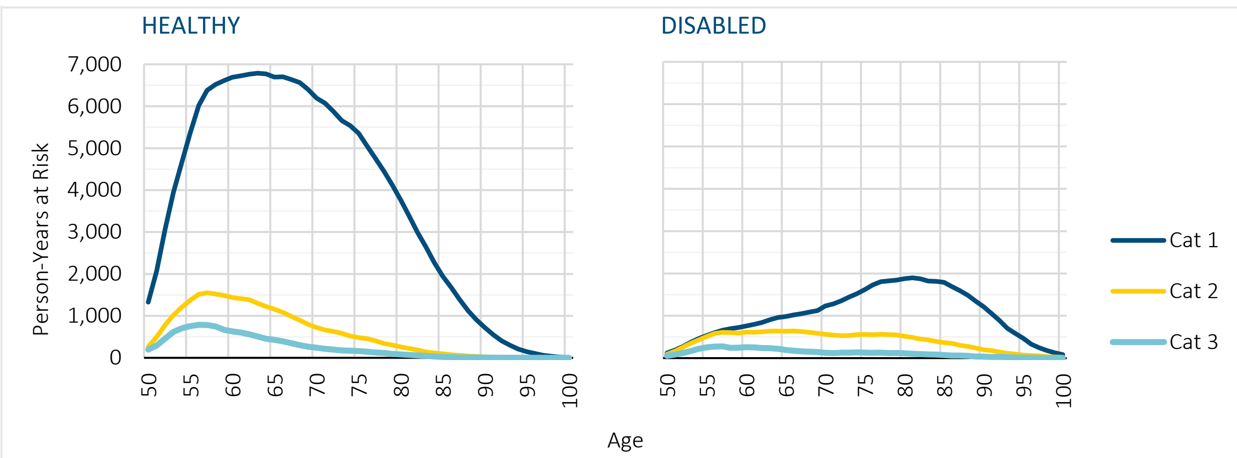
TOTAL PERSON-YEARS AT RISK IN HEALTHY AND DISABLED STATES BY ATTRIBUTE 1



Cat denotes category; this notation is used throughout.

Figure 4

TOTAL PERSON-YEARS AT RISK IN HEALTHY AND DISABLED STATES BY ATTRIBUTE 2



Regarding the number of health state transitions, Figure 5 and Figure 6 show that within each category of Attribute 1 and Attribute 2, transitions into disability are the most frequent before advanced ages, followed by recoveries. After approximately age 85, transitions from the disabled to the dead state become the most common across categories of Attribute 1 and for Category 1 of Attribute 2. For Category 2 of Attribute 2, this shift occurs earlier, around age 80. For Category 3 of Attribute 2, the pattern is less distinct due to the smaller exposure to risk in this group. Because the transition counts are influenced by the amount of time individuals spend at risk in each state, the next step is to examine crude transition rates—computed as the number of transitions divided by the corresponding person-years at risk—for a more comparable assessment across categories.

Figure 5

AGE-SPECIFIC COUNTS OF HEALTH STATE TRANSITIONS BY ATTRIBUTE 1

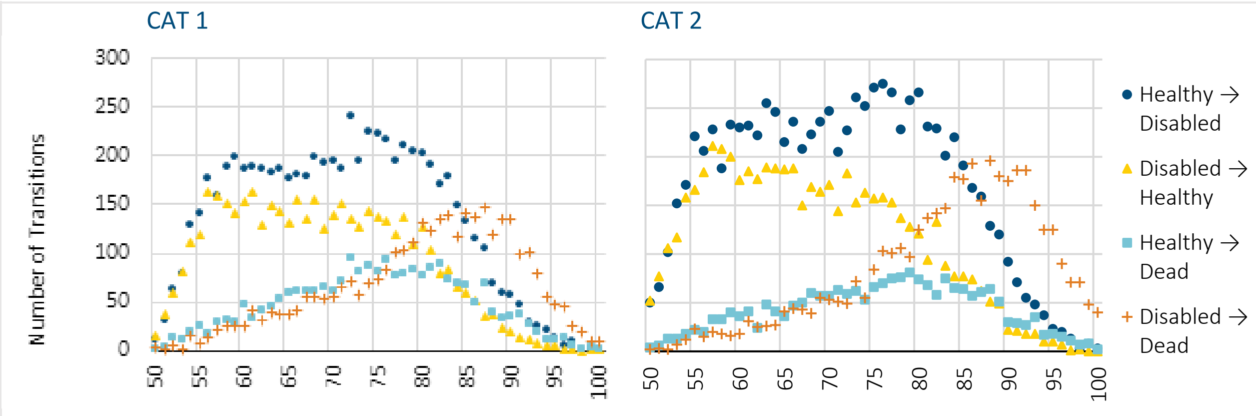


Figure 6

AGE-SPECIFIC COUNTS OF HEALTH STATE TRANSITIONS BY ATTRIBUTE 2

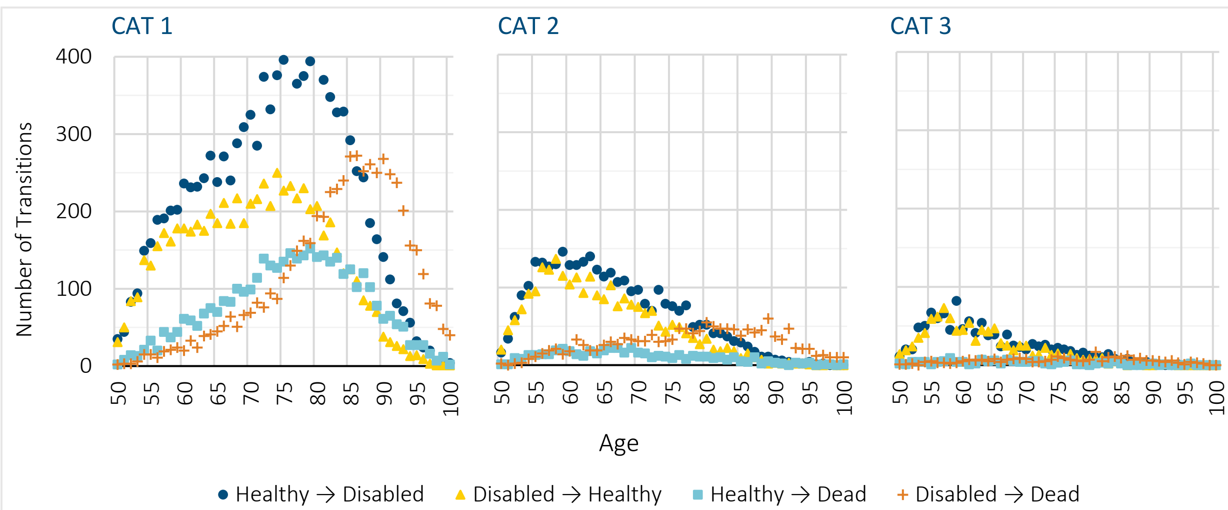


Figure 7 presents the crude transition rates by Attribute 1. For both disability and recovery, there is no substantial difference across categories, although Category 1 exhibits slightly higher recovery rates after age 80. In contrast, mortality shows a pronounced difference across categories, with Category 1 experiencing considerably higher mortality rates, particularly from the healthy state.

Figure 7
AGE-SPECIFIC CRUDE TRANSITION RATES BY ATTRIBUTE 1

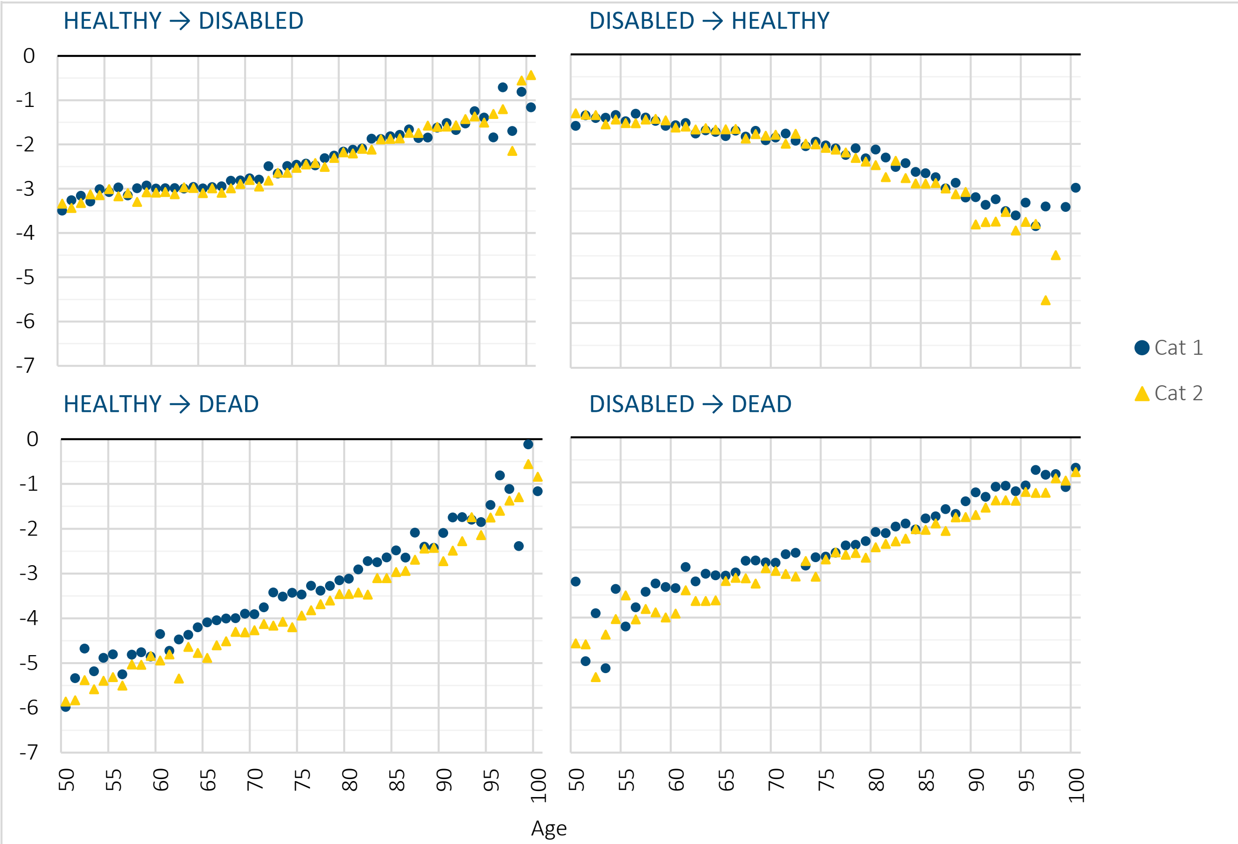
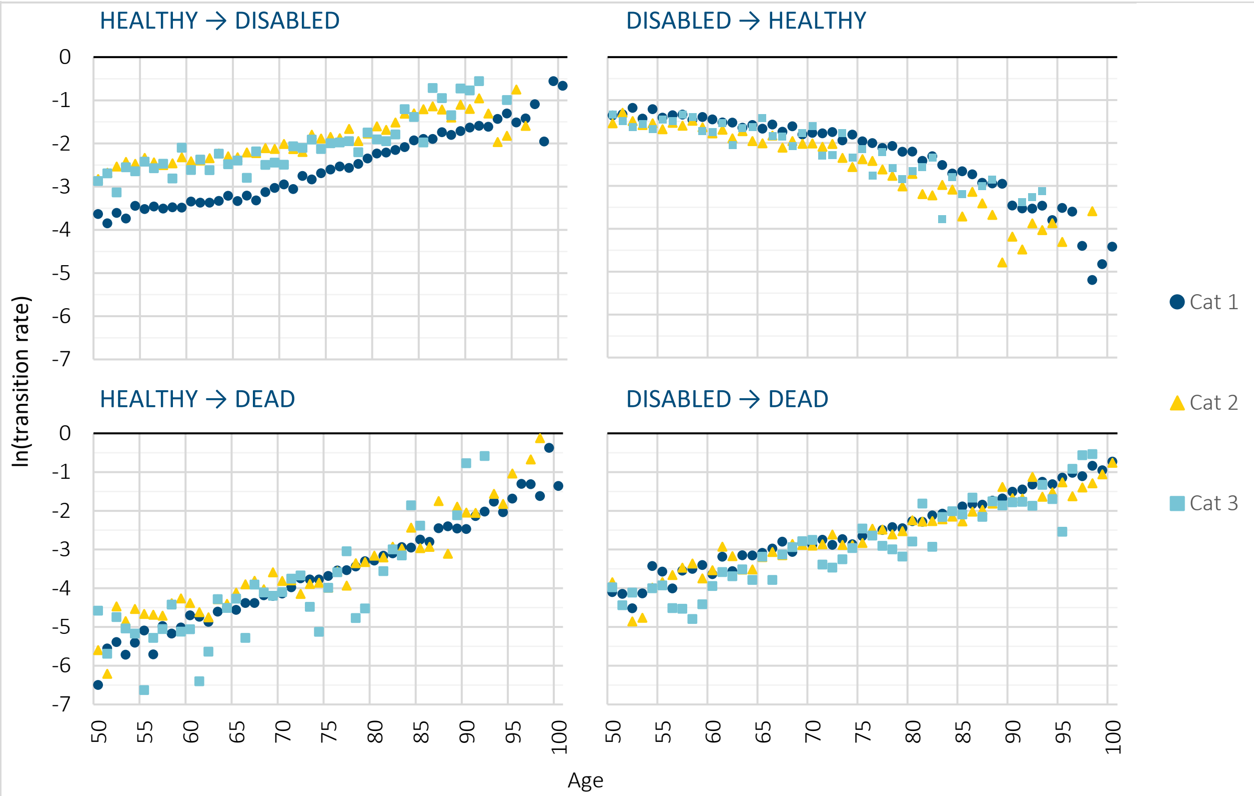


Figure 8 displays the crude transition rates by Attribute 2. The incidence of disability is notably lower for Category 1 compared with the other categories. Recovery rates are slightly higher for Category 1, though the difference is smaller than that observed for disability. Mortality patterns show greater variability for Category 3, likely due to smaller sample size; however, despite this volatility, this group generally exhibits lower mortality from both the healthy and disabled states. Category 2 demonstrates higher mortality from the healthy state, while mortality from the disabled state is similar between Categories 1 and 2.

Figure 8
AGE-SPECIFIC CRUDE TRANSITION RATES BY ATTRIBUTE 2



3.2 MODELING MULTI-STATE TRANSITIONS

3.2.1 MULTI-STATE MODEL FRAMEWORK

To examine how transition intensities vary with individual characteristics, prior studies commonly assume that the transition intensity follows a log-linear function of covariates:

$$\ln \lambda_{k,m}(t) = \beta_m' \cdot z_k + \gamma_m' \cdot \omega_k(t), \quad (1)$$

where $m = 1, 2, 3, 4$ denotes the four possible health state transitions. For the k^{th} individual, $\boldsymbol{z_k}$ denotes the vector of static covariates, and $\boldsymbol{\omega_k}(t)$ denotes the vector of time-varying covariates at time t. The model parameters, $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma})'$, are then estimated using the maximum likelihood estimation (MLE) method. The likelihood function is presented in Appendix B.

3.2.2 A UNIFYING FRAMEWORK FOR MULTI-STATE TRANSITION MODELS

In practice, the form of (1) can be rather limiting; a more general form given by

$$\ln \lambda_{k,m}(t) = f_m\big(z_k, \omega_k(t)\big), \quad (2)$$

is hence desired. In wider actuarial practice, machine learning techniques are used to model such forms, with techniques such as generalized additive models, gradient boosting machines, random forests and occasionally neural networks.

While in principle one can simply replace the log-linear model (1) with the general form (2) in the likelihood, the main obstacle to successful estimation is to actually maximize the likelihood, or otherwise incorporate the desired functional form. Existing solutions to this problem have taken disparate forms, such as tree-splitting using custom losses (Hothorn et al., 2006; Ishwaran et al., 2008), piecewise approximation networks (Cottin et al., 2022), and the Poisson-survival connection (Wang et al., 2022). It is possible that the lack of a unifying framework across these multiple solutions, and hence the need to learn a new framework for each machine learning technique, is the reason behind why LTCI modeling remains largely confined to log-linear models.

Towards facilitating wider usage by the LTCI actuarial community, this research offers a unifying framework based on the Poisson-survival connection used in Wang et al. (2022). The researchers' insight is that this framework is much more general than just neural networks, but in fact extends to any regression technique which can utilize the Poisson likelihood loss, including tree-based methods and generalized additive models. Thus, by simply changing the loss function used, and taking an appropriate transformation of the observations, this framework will work with any off-the-shelf implementation, such as TensorFlow or the R implementation of other models. At the same time, it converts the relatively more niche setting of LTCI pricing to the much more familiar setting of Poisson regression. This allows the established expertise and best practices from the Poisson regression setting to be utilized in LTCI pricing.

The details of this unifying framework are relegated to Appendix C for the interested reader. Here, only the implications of this framework will be described. The framework allows one to, without loss of information, express the estimation of transition rates as a set of Poisson regression problems, with a separate regression corresponding to each type of transition. Thus, provided the data is correctly structured, one can bypass the need to set up the complex multi-state model likelihood, and instead utilize the much more familiar Poisson regression framework in estimating transition rates.

This paper focuses on the setting of pricing an LTCI product based on information at underwriting. The sole time-varying covariate is hence age, which is taken to be age last birthday. Non-age covariates are assumed to be static as of policy inception; these two combine to form the covariate set for Poisson regression. The length of time spent in a given age forms the exposure; the log-exposure hence serves as an offset variable in this Poisson regression procedure.

In principle, one could use any method of choice to model the transition rates, including boosting, random forest, generalized additive models and neural networks. For simplicity, this paper focuses on the log-linear model.

3.2.3 DATA PREPARATION AND VARIABLE SELECTION

Given the choice to utilize log-linear models, the desired models can simply be estimated using the GLM routine in the R programming language, using the log-link and the Poisson family of distributions. Provided that the data is correctly structured, estimation is straightforward using the formula and offset interface native to R.

Before the covariate dataset can be used, the missing values in the data need to be addressed. This study employs the following 2-step approach towards this end:

- Variables with extensive missing values are first identified and removed.
- Once this is done, all individuals with incomplete entries are removed.

The full covariate dataset consists of 30,561 individuals. Of the variables with missing entries, job physicality makes up by far the largest number with 4,253 entries (about 14%), whereas the next largest is only missing for 583 entries (about 2%). Utilizing this covariate would mean the need to remove a large number of entries, each of which is mostly otherwise complete. For variables with such a large proportion of missing values, another major concern is whether the missingness is ignorable, i.e., whether the missingness is related to other covariates.

Hence, this variable is removed from the covariate set, leaving only 1,018 incomplete entries. Since this number is relatively much smaller, all incomplete entries are removed, resulting in a complete dataset. Accordingly, all entries in the transition dataset pertaining to the removed individuals are also removed.

Towards developing a parsimonious model, the model is fitted with all the covariates to identify simplifications which do not impact the model fit materially. Below are the changes that the researchers identify:

- Marital status is simplified into 5 categories: Married, Separated, Partnered, Widowed, Never.
- Years of education is simplified into the following eight categories: did not complete elementary school (0–5 yrs), completed elementary but not middle school (6–7 yrs), completed middle school (8 yrs), some high school (9–11 yrs), completed high school (12 yrs), some college (13–15 yrs), completed college (16 yrs), and postgraduate (17+ yrs).
- Entries with census region of 5 are removed from the dataset as there are too few entries with this region (17 out of approximately 30,000 records).

This study also explored the possibility of an interaction term between Attributes 2 and 3. It was found that in general, the interaction effect was not particularly material. Since including the interaction effect would also complicate interpretation, the study proceeds with only the main effects for these two variables. The researchers also confirm that the removal of job physicality results in a model with better fit: see Appendix D for details.

Finally, log-transformation on the financial variables is considered because they exhibit a large skew. In particular, in place of raw total household income, it is first transformed using the function $\ln(1+x)$ before including it as a covariate. The case of total non-housing wealth is more complicated as it includes negative values, hence precluding a simple transformation. However, letting V denote the wealth, the researchers have found that transforming it using $\ln(1+|V|)$ gives an approximately normal distribution. In view of this, and to incorporate the possible specific effects of negative wealth, this study includes the following three terms as covariates: $\ln(1+|V|)$, $\operatorname{sgn}(V)$, and $\operatorname{sgn}(V)\cdot\ln(1+|V|)$.

3.2.4 MODEL VALIDATION: SELECT ESTIMATED COEFFICIENTS

Towards validating the fitted models and ensuring that the signs of their coefficients comport with intuition, the coefficients of these models are displayed for each of the four transitions, for selected variables; coefficients for the remainder of variables are displayed in Appendix E.

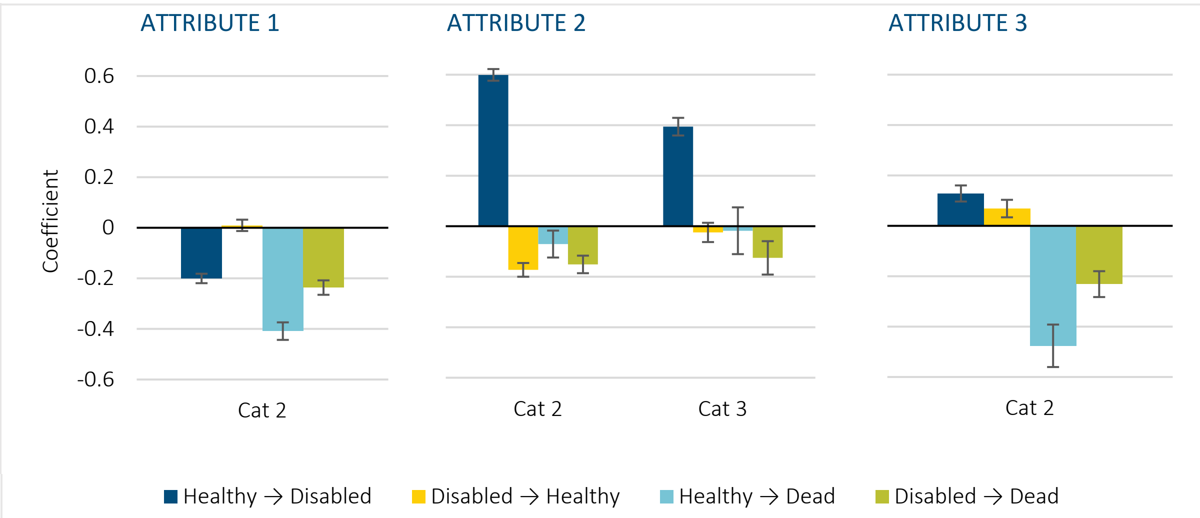
First estimated coefficients for categorical variables are displayed. Recall that this study estimates four sets of coefficients, one for each transition. As a log-linear model is being used for the transition rates, coefficients for categorical variables—when exponentiated—represent the ratio between transition rates for that particular category and the baseline category, holding all else equal. Take for instance the coefficient for Category 2 of Attribute 2 (relative to the baseline Category 1) of –0.20041, for the transition of healthy → disabled. One can interpret this coefficient as saying that the transition rate for a healthy person becoming disabled in Category 2 is exp(-0.2004) = 0.8184 times, i.e., 18.16% less than that for an otherwise identical individual in Category 1.

The results for the demographic dimensions corresponding to Attributes 1—3 are shown in Figure 9. For such variables, whose categories do not have any inherent ordering, the coefficients are displayed using side-by-side bar graphs; black lines denoting the one-standard-error margin for each of the estimated coefficients are also displayed.

For these three attributes, the baseline levels used are chosen to be the most common category of each. Relative to Category 1 of Attribute 1, Category 2 exhibits significantly lower rates of becoming disabled, as well as lower mortality. For Attribute 3, relative to Category 1, Category 2 shows higher rates of both becoming disabled and recovering, whereas their mortality is significantly lower.

As for Attribute 2—the attribute of interest in this example—relative to Category 1, Category 2 has markedly higher rates of becoming disabled, and significantly lower rates of recovery. The corresponding differences between Category 3 and Category 1 are smaller. Both Category 2 and Category 3 appear to have lower rates of mortality than Category 1.

Figure 9
COEFFICIENTS FOR ATTRIBUTES 1—3



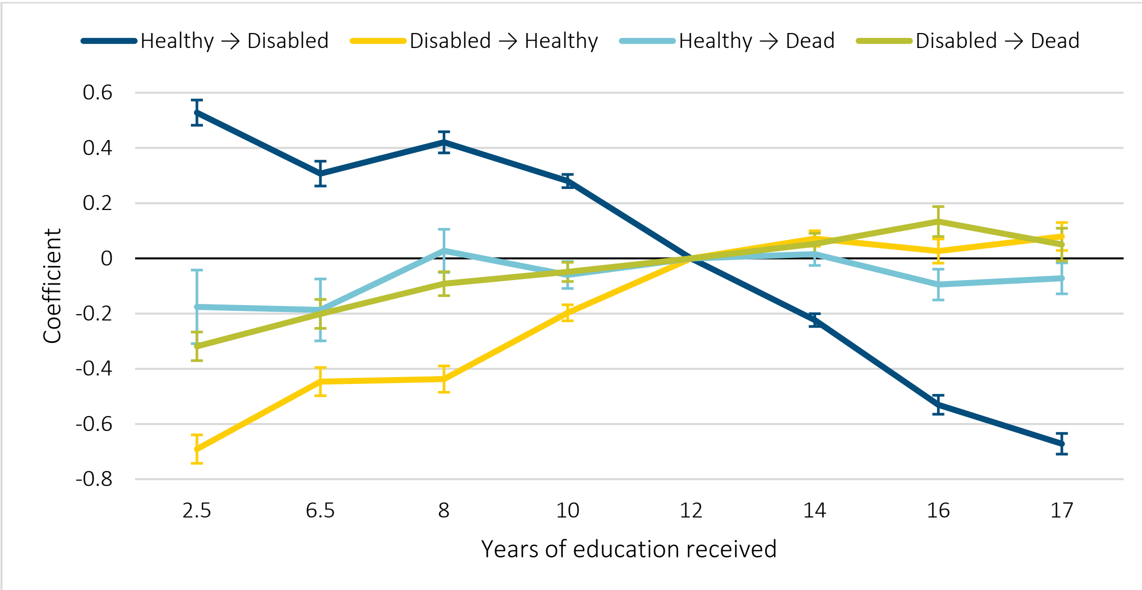
The reference categories are Category 1 for Attribute 1, Category 1 for Attribute 2, and Category 1 for Attribute 3.

The plot for years of education received (EDYRS) is shown in Figure 10. Although this variable is utilized as a categorical variable, its categories are ordered and can be interpreted as a continuous variable, hence a stacked line graph is instead used to display the coefficients. For this type of graph, the level of the line denotes the value of the estimated coefficient, whereas a 1-standard-error margin is provided alongside the estimated value as before.

The reference level used is EDYRS=12 (completed high school), and the midpoint of the number of years is used for plotting on the x-axis. As expected, the likelihood of becoming disabled decreases with increasing years of education received, whereas the opposite is observed for the likelihood of recovering from disability. Puzzlingly, mortality—both from the healthy and disabled states—appears to actually increase with education. It is interesting to note this observation is not confined to this study: multiple other studies have noted the same phenomenon (e.g., Hoffmann, 2011; Xu et al., 2019).

Figure 10

COEFFICIENTS FOR YEARS OF EDUCATION RECEIVED

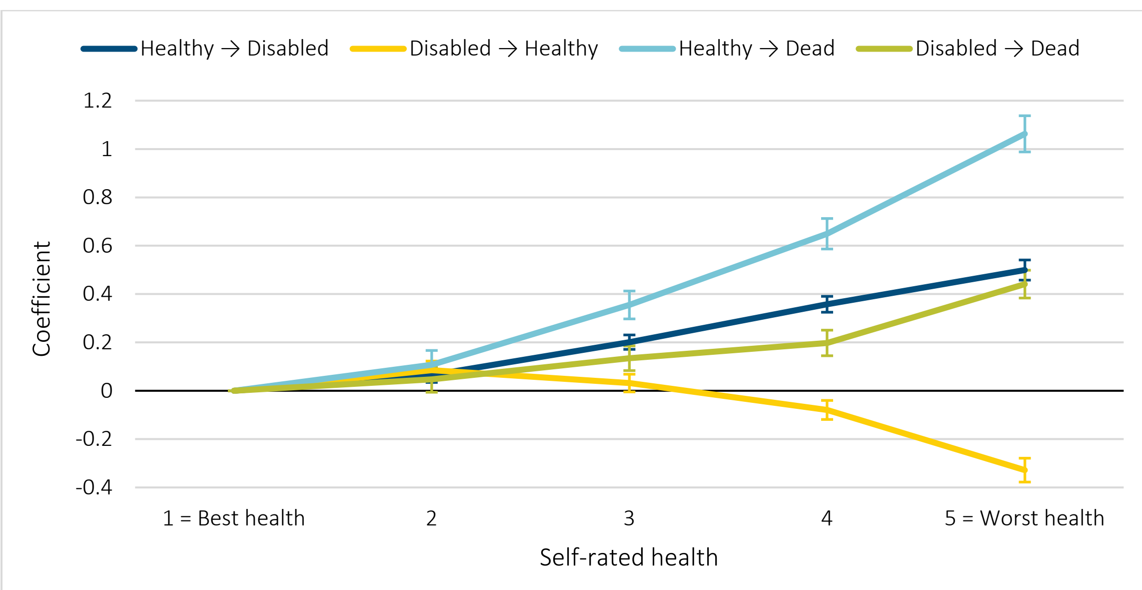


Years of education is grouped into the eight categories: did not complete elementary school (0–5 years), completed elementary but not middle school (6–7 years), completed middle school (8 years), some high school (9–11 years), completed high school (12 years), some college (13–15 years), completed college (16 years), and postgraduate (17+ years). The reference category is 12 years of education.

The last categorical variable considered for the purposes of validation is self-rated health. This variable ranges from 1 to 5, with 1 representing the best health, and 5 the worst. This study takes 1 as the baseline. Relative to this baseline, as expected, those with lower self-rated health exhibit higher rates for transitioning to both disability and mortality, and lower rates of recovery.

Figure 11

COEFFICIENTS FOR SELF-RATED HEALTH



1 = Excellent; 2 = Very good; 3 = Good; 4 = Fair; 5 = Poor. The reference category is 1.

Finally, the coefficients for age are displayed, along with the contribution to the likelihood for each fitted coefficient. The coefficient indicates—once exponentiated—the ratio of the subsequent integer age relative to the previous integer age. For instance, the coefficient of 0.0482 for the transition of healthy to disabled can be interpreted as saying that, for an otherwise identical person, a person with age $x + 1$ has transition rate of healthy to disabled which is exp(0.0482) = 1.0493times, i.e., 4.93% more than that of a person with age x. Note that this ratio compounds: if instead the person is aged x +5, then the person's transition rate will be 1.0493^5 = 1.2725 times that of a person aged x.

The table displays coefficients for each of the four possible transitions. It can be seen that the signs of the coefficients make intuitive sense, being positive for the transitions of healthy to disabled (disability), healthy to dead (healthy mortality), and disabled to dead (disabled mortality), while being negative for the transition of disabled to healthy (recovery), indicating deteriorating health with increasing age. The contribution to the likelihood is also very large. A contribution of 4 or above indicates significance, and it can be seen that the contributions for age far exceed this threshold. While not reported here, the contributions for other continuous variables are generally 20 or below, making age by far the variable with the largest explanatory power among the other continuous variables.

Table 2
TABLE OF COEFFICIENTS FOR AGE

Variable	Healthy → Disabled	Disabled → Healthy	Healthy → Dead	Disabled → Dead
Age	0.0482	-0.0436	0.0848	0.0725
	(1177.17)	(800.72)	(987.08)	(1166.23)

The contribution to the likelihood of each coefficient, i.e., the reduction in likelihood in the final model were the covariate to be removed, is given below in parentheses.

3.3 A HYPOTHETICAL LTCI PRICING EXERCISE

This section demonstrates how the post-processing method of Lindholm et al. (2022) can be applied to LTCI pricing using the transition rates estimated from the HRS-based multi-state model. The illustration serves as a baseline example linking the modeling framework developed earlier in Part II with the fairness methodology introduced in Part I.

3.3.1 SETTING AND ASSUMPTIONS

This subsection outlines the settings and assumptions for this hypothetical illustration of these potential fairness adjustments in LTCI pricing. The analysis focuses on how potential fairness adjustments can be implemented at the transition-rate level to evaluate their impact on pricing outcomes across demographic groups.

Although the HRS data do not represent an insured population, this illustration treats the transition rates estimated from this study's model as if they describe the actual experience of a representative LTCI portfolio. To approximate a realistic insured population, this analysis includes only individuals aged 50–80 at issue (on an age-last-birthday basis) and excludes those with disqualifying health conditions. Specifically, eligible individuals must not be functionally disabled at issue and must not report severe heart or lung conditions, diabetes, or stroke. This restriction yields a relatively healthy baseline population suitable for LTCI pricing analysis.

This illustration considers a simplified LTCI product with a single lump-sum premium paid at policy inception and an annual benefit of $1 at the beginning of each year the insured remains in the disabled state. It assumes that each individual in the baseline insured population represents an insured whose coverage begins immediately after the individual's previous birthday. It also assumes that this insured's transition rates are given by those predicted for this same individual, with a terminal age of 110 assumed for the transition rates.

Let $J_x \in \{H, F, D\}$ denote the health state at age x. The lump-sum premium for a policy issued to an individual aged x is given by

$$\text{Lump sum premium} = \sum_{t=0}^{110-x} v^t \Pr(J_{x+t} = F \mid J_x = H),$$

where the discount factor v is set to 1.03^{-1}. The multi-year transition probability, $Pr(J_{x+t} = F \mid J_x = H)$, is obtained from the one-year transition probabilities via the Chapman–Kolmogorov equations. The one-year transition probabilities, in turn, are computed from the transition intensities using the matrix exponential, which solves the Kolmogorov forward equations.

This illustration further assumes that Attribute 2 is chosen as the attribute of interest S, with three categories: denoted as Category 1, Category 2, and Category 3. All other covariates, denoted collectively by $\boldsymbol{z}$ are drawn from the HRS dataset as described in Section 3.1. To evaluate fairness in pricing, this illustration computes three sets of hypothetical premiums under different modeling assumptions. See Section 2.3.3 for the definitions of each of these prices:

- **Best-estimate price:** Transition rates $\hat{\lambda}_m(\boldsymbol{z}, x, s)$, for transition type $m = 1, \dots, 4$, are estimated using all available covariates $\boldsymbol{z}$, the policyholder's age x, and their attribute s, based on the fitted multi-state GLM model:

  $$\hat{\lambda}_m(\boldsymbol{z}, x, s) = \exp\left(\hat{f}_m^{(\text{Best})}(\boldsymbol{z}, x, s)\right),$$

  where $\hat{f}_m^{(\text{Best})}(\cdot)$ denotes the fitted GLM predictor function that includes the attribute of interest as an explanatory variable.

- ***S*-blind price**: Transition rates $\hat{\lambda}_m(\boldsymbol{z}, x)$ are re-estimated after removing s from the covariate set, which assumes the insurer is unaware of the attribute s:

  $$\hat{\lambda}_m(\boldsymbol{z}, x) = \exp\left(\hat{f}_m^{(\text{-}s)}(\boldsymbol{z}, x)\right),$$

  where $\hat{f}_m^{(\text{-}s)}(\cdot)$ denotes the re-fitted GLM predictor function excluding the attribute of interest from both data and estimation.

- **Adjusted price:** Transition rates are adjusted using the post-processing method of Lindholm et al. (2022); see Section 2.3.3. The adjusted transition rate for an individual with characteristics $(\boldsymbol{z}, x)$ is defined as

  $$\lambda_m^*(\boldsymbol{z}, x) = \sum_s \hat{\lambda}_m(\boldsymbol{z}, x, s)\, \mathbb{P}^*(S = s),$$

  where $\mathbb{P}^*(S = s)$ denotes a reference distribution of the attribute of interest that is independent of $\boldsymbol{z}$ and x, and is set equal to the observed proportions of each category in the full covariate dataset. Note that this example does not condition on age or time at entry for these proportions.

In this illustration, the potential fairness adjustment is applied to transition rates (λ) rather than to premiums directly, so that potential fairness is introduced at the stage where risk intensities are modeled. Alternative choices of the response variable Y and of $\mathbb{P}^*$ will be discussed in Section 3.4.

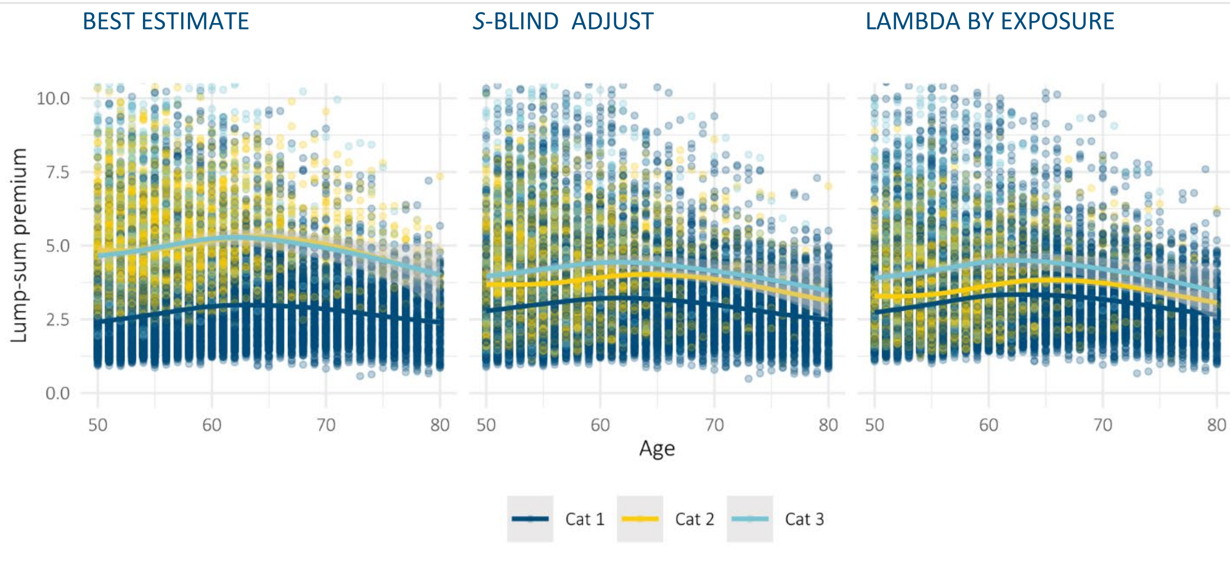
3.3.2 RESULTS AND DISCUSSION

The results of the hypothetical pricing exercise are shown in Figure 12. Each point corresponds to the lump-sum premium charged to an individual. For individuals of the same age and category of Attribute 2, premiums may differ because of other personal characteristics used in pricing. To facilitate comparison across categories, this illustration includes three smooth curves in each panel—one for each category—showing how premiums generally change with age.

Across categories of Attribute 2 and modeling approaches, the smoothed lump-sum premiums change relatively little with age. This occurs because two opposing forces largely offset each other. As people age, they are more likely to experience disability and less likely to recover, which tends to increase expected benefit payments. At the same time, older individuals face higher mortality risk, which shortens the expected time spent in a disabled state and reduces total expected payments.

Figure 12

HYPOTHETICAL LUMP-SUM PREMIUMS BY ATTRIBUTE 2



Each panel represents a different pricing model specification. Each point represents a single insured in the portfolio of interest. There are three smoothed lines in each panel, one for each category. Each smoothed line is generated using a generalized additive model, regressing the lump-sum premium on age.

Figure 12 visually assesses how each set of prices is situated in terms of demographic parity. Recall from Figure 5 that, relative to Category 1 of Attribute 2, the model estimates substantially higher disability rates for Categories 2 and 3, while recovery rates are either similar or noticeably lower. Given that the best-estimate prices precisely use these modeled rates to calculate prices, it is thus not surprising that the gap across categories is largest for this set of prices. The graph shows that prices are much higher for Categories 2 and 3 relative to Category 1.

When Attribute 2 is omitted from the model, the resulting hypothetical attribute-blind prices show that the gap in average prices persists, albeit now much smaller. This suggests that the attribute of interest S is highly informative for the transition rates across categories. While direct effects of the attribute of interest are no longer present in these sets of prices, indirect effects—resulting from indirect inference of the attribute through other covariates—are still present, and likely a contributing factor to the gap persisting.

The methodology of Lindholm et al. (2022) ensures that such inference of the attribute S is no longer present in the estimated transition rates. Nevertheless, Lindholm et al. (2024) show that the absence of unintentional indirect association and notions of group fairness in general do not imply one another, i.e., preventing indirect inference does not guarantee equal average premiums across categories. Hence one does not expect the gap to be completely eliminated. Indeed, the researchers observe in this hypothetical illustration that the gap persists in a smaller form. Since the reweighting method eliminates indirect effects of the attribute of interest, the researchers interpret this as reflecting average differences due solely to the direct effects of all other covariates.

Nevertheless, the elimination of indirect effects in this hypothetical illustration does seem to have markedly reduced the gap between Categories 1 and 2, although the gap between Categories 1 and 3 appears to be largely unaffected. This suggests that in this hypothetical illustration, for Category 2, the attribute-blind prices contain strong indirect effects of the attribute through the other covariates, which the adjusted prices eliminate. It also suggests in this hypothetical illustration that indirect effects appear to be less pronounced for Category 3.

3.4 ROBUSTNESS TEST: ALTERNATIVE SETTINGS FOR POST-PROCESSING FOR LTCI

3.4.1 CHOOSING $\mathbb{P}^*$

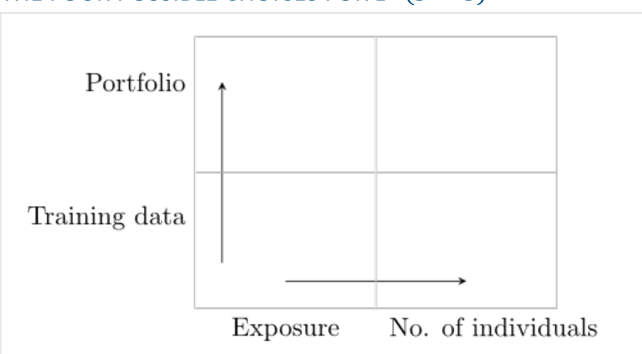
When choosing the marginals $\mathbb{P}^*(S = s)$ for the application of Lindholm et al. (2022), two dimensions may be considered:

- By exposure (i.e., number of person-years at risk) vs. by number of individuals, and
- Training data vs. portfolio of interest.

The combination of these dimensions yields four possible formulations for $\mathbb{P}^*(S = s)$, as illustrated in Figure 13.

Figure 13

THE FOUR POSSIBLE CHOICES FOR $\mathbb{P}^*(S = s)$



Arrows indicate increasing application/outcome orientation.

Recall that the methodology of Lindholm et al. (2022) reweights outcomes by replacing $\mathbb{P}(S = s \mid X = x)$ with $\mathbb{P}^*(S = s)$. Hence, a choice is more data/estimation-oriented if it is a closer reflection of the original $\mathbb{P}(S = s \mid \boldsymbol{X} = \boldsymbol{x})$ replaced, whereas it is more application/outcome-oriented if it is better tailored to the application or outcome of interest. Put another way, in this hypothetical pricing exercise, the choice is data/estimation-oriented if it better reflects the data on which transition rates are estimated, whereas it is more application-oriented if it relates to the portfolio on which fairness is desired, and more outcome oriented if it relates to the computed premiums.

The first dimension concerns the choice between exposure and number of individuals. Since the framework breaks down the multi-state estimation problem into its constituent Poisson regressions, the $\mathbb{P}(S = s \mid \boldsymbol{X} = \boldsymbol{x})$ associated with each regression serve as the weights underlying the (S-blind) transition rates. The datasets used to estimate these transition rates are weighted by person-years at risk or exposure. Hence, defining $\mathbb{P}^*(S = s)$ as the proportion of person-years belonging to each category of the attribute of interest is more data/estimation-oriented. Since exposure depends on an individual's initial state, the proportion of person-years belonging to each category, $\mathbb{P}^*(S = s)$, would also vary accordingly under this formulation.

On the other hand, the number of distinct computed premiums depends only on the number of individuals in the portfolio, not the total exposure. If the goal is to use these prices as input to the Lindholm et al. (2022) adjustment, the weights should not depend on the initial state. In this case, $\mathbb{P}^*(S = s)$ is defined as the proportion of individuals belonging to each category of the attribute of interest, rather than the proportion of person-years. Such a formulation has the additional conceptual advantage that the resulting weights are independent of each individual's health trajectory, unlike the exposure-based formulation.

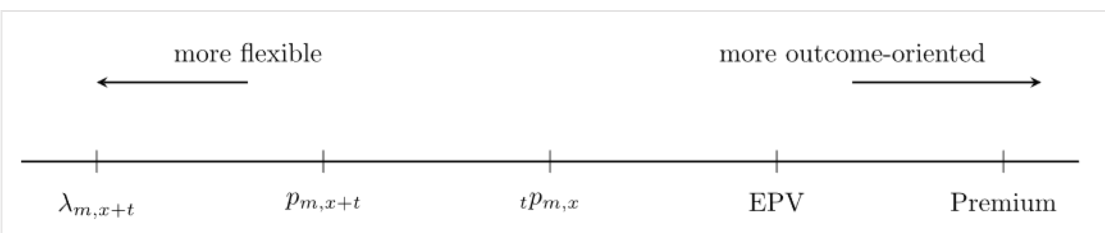
The second dimension concerns whether the proportion of individuals belonging to each category, $\mathbb{P}^*(S = s)$, is derived from the training data or from the portfolio. The former is clearly more data/estimation-oriented, as it relies on all available data used for model fitting. The latter, however, is preferable when the primary goal is to eliminate potential unintentional indirect associations within the specific portfolio of interest, making it more application/outcome-oriented.

In this hypothetical pricing exercise, the researchers experimented with all the $\mathbb{P}^*$ formulations described above. As the resulting prices are largely similar, detailed comparisons are omitted for brevity.

3.4.2 CHOOSING THE RESPONSE VARIABLE Y

The researchers consider several possible definitions for the response variable Y, guided by the intermediate quantities that arise when deriving long-term care insurance (LTCI) prices from the estimated transition intensities. Figure 14 illustrates the spectrum of potential choices for Y. Variables toward the left of the spectrum are more flexible, as they serve as foundational quantities from which those on the right are derived. In contrast, variables toward the right are increasingly outcome-oriented, as they lie closer to the premium, which represents the final output of the pricing.

Figure 14
SPECTRUM FOR THE RESPONSE VARIABLE Y

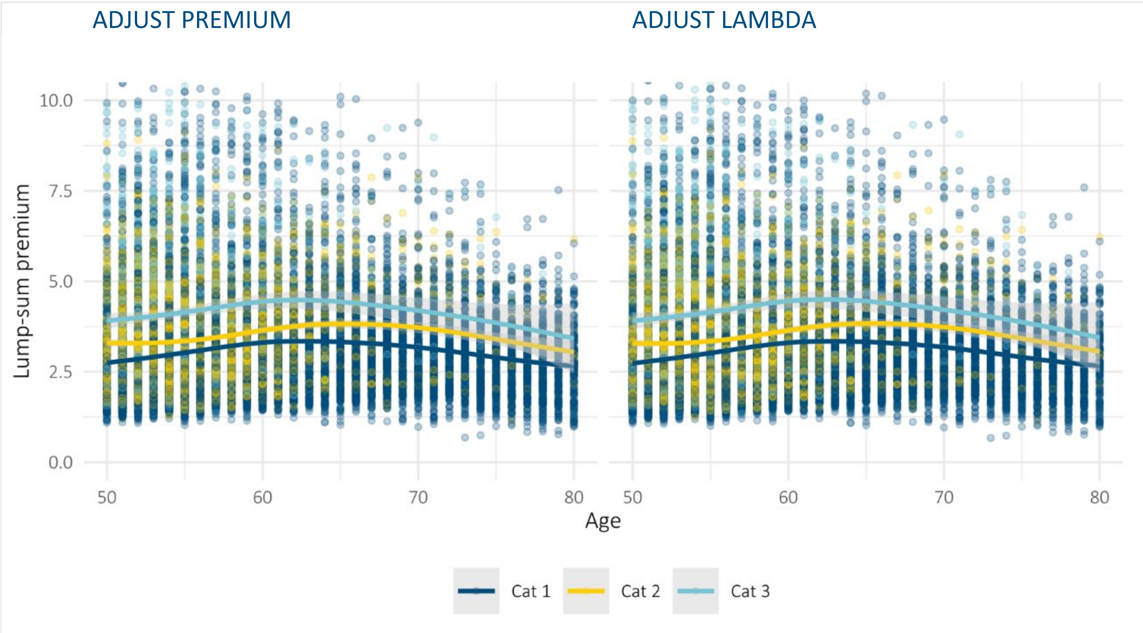


The term $\lambda_{m,x+t}$ denotes the transition intensity for transition type m at age $x + t$. The term $p_{m,x+t}$ represents the one-year transition probability, while ${}_tp_{m,x}$ denotes the t-year transition probability. The symbol EPV refers to the expected present value.

Figure 15 compares hypothetical premiums obtained under two different adjustment methods. In the left panel, the response variable is the premium—the most outcome-oriented choice in the spectrum of response variables—and the weights are based on the number of individuals in the portfolio, which is likewise the most outcome-oriented among the four possible weighting schemes. In contrast, the right panel uses the transition intensity as the response variable—the most flexible in the spectrum—and weights based on the exposure in the training data, the least outcome-oriented choice. In other words, the two panels represent opposite ends of the spectrum in terms of both response variable and weighting scheme. Yet, the resulting adjusted premiums are virtually identical, indicating that the adjustment is robust to these modeling choices.

Figure 15

HYPOTHETICAL LUMP-SUM PREMIUMS BY ATTRIBUTE 2



In the left panel, the response variable is the premium, and the weights are based on the number of individuals in the portfolio. In the right panel, the response variable is the transition intensity, and the weights are based on the exposure in the training data. Each point represents a single insured in the portfolio of interest. There are three smoothed lines in each panel, one for each category. Each smoothed line is generated using a generalized additive model, regressing the lump-sum premium on age.

3.4.3 CHOOSING LTCI PRODUCTS

To make the simplified LTCI product more realistic, it is extended to include additional policy features commonly found in practice. Specifically, this hypothetical illustration considers LTCI policies that charge a level annual premium, payable only when the policyholder is in the healthy state. The level annual premium for a policy issued to an individual aged x is given by

$$\text{Level annual premium} = \frac{\sum_{t=0}^{110-x} v^t \Pr(J_{x+t} = F \mid J_x = H)}{\sum_{t=0}^{110-x} v^t \Pr(J_{x+t} = H \mid J_x = H)}.$$

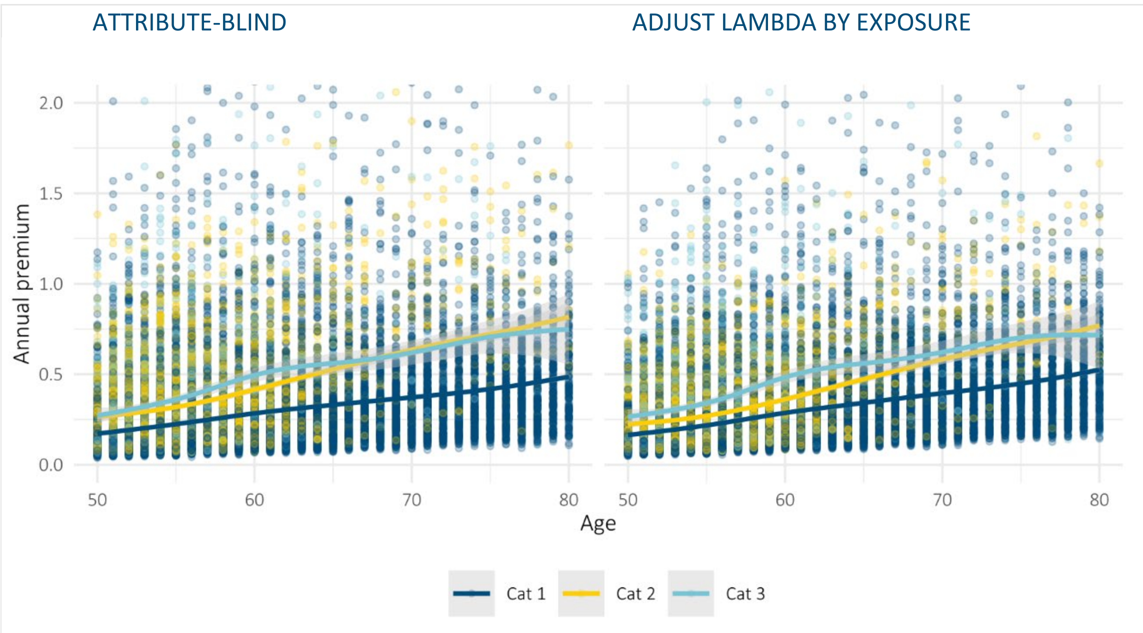
This hypothetical illustration also considers policies that include a death benefit rider, under which a lump-sum amount is paid at the end of the year of death. The death benefit is defined as the total amount of premiums paid up to that point, without interest accumulation. The lump-sum premium for policies that refund premiums upon death can be calculated as the standard lump-sum premium, divided by one minus the expected present value of a life insurance policy which pays $1 at the end of the year of death. This can be expressed as

$$\text{Lump sum premium} = \frac{\sum_{t=0}^{110-x} v^t \Pr(J_{x+t} = F \mid J_x = H)}{1 - \sum_{t=0}^{110-x} v^{t+1}[\Pr(J_{x+t+1} = D \mid J_x = H) - \Pr(J_{x+t} = D \mid J_x = H)]}.$$

The derivation of the corresponding annual premium for policies including the death benefit rider is more involved and is therefore provided in Appendix F.

Figure 16 presents the attribute-blind and adjusted prices for the hypothetical LTCI policies with level annual premiums, where the adjustment is applied to the transition intensities and based on exposure. Figures 17 and 18 display corresponding results for hypothetical policies that include a death benefit rider, showing outcomes for both lump-sum and level annual premium structures, respectively. The overall patterns remain consistent with those observed for the baseline hypothetical policy (lump-sum premium without a death benefit rider) after applying the adjustment based on the method of Lindholm et al. (2022). Specifically, in this hypothetical illustration, the adjustment narrows the gap between Categories 1 and 2 of Attribute 2, while its impact on Category 3 remains limited.

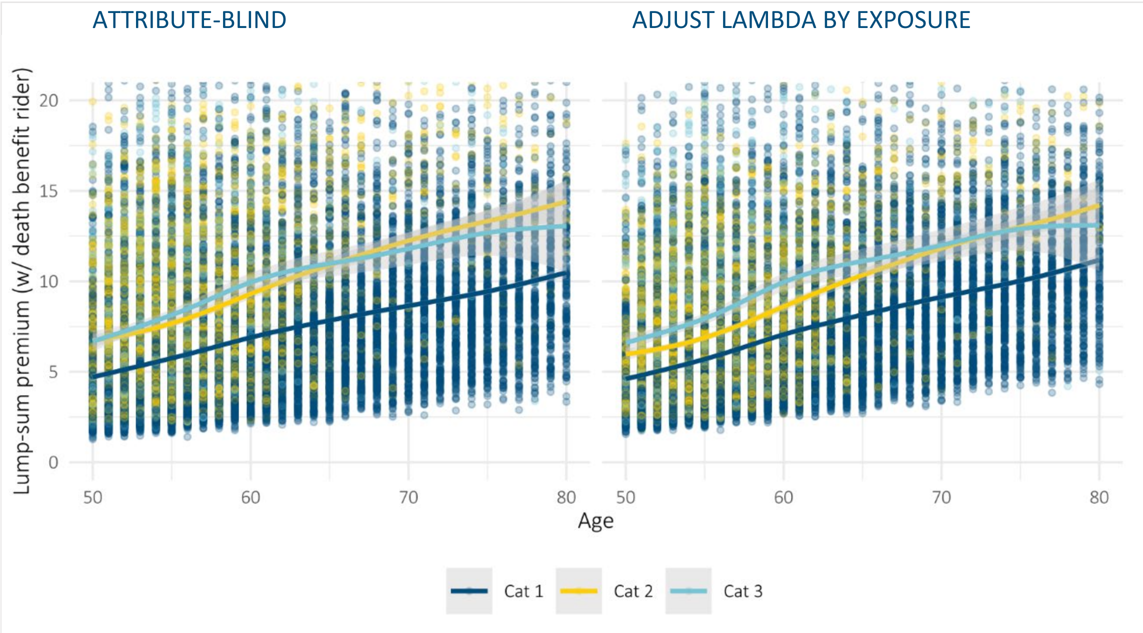
Figure 16
HYPOTHETICAL ANNUAL PREMIUMS BY ATTRIBUTE 2



Each panel represents a different pricing model specification. Each point represents a single insured in the portfolio of interest. There are three smoothed lines in each panel, one for each category. Each smoothed line is generated using a generalized additive model, regressing the lump-sum premium on age.

Figure 17

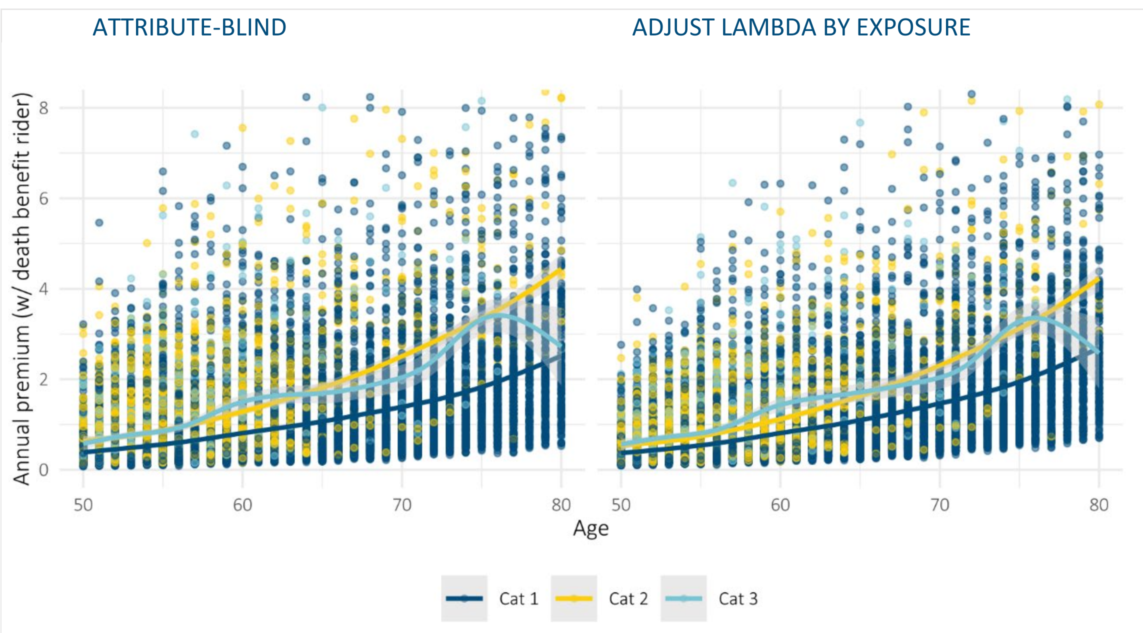
HYPOTHETICAL LUMP-SUM PREMIUMS FOR LTCI POLICIES THAT RETURN PREMIUMS UPON DEATH



Each panel represents a different pricing model specification. Each point represents a single insured in the portfolio of interest. There are three smoothed lines in each panel, one for each category. Each smoothed line is generated using a generalized additive model, regressing the lump-sum premium on age.

Figure 18

HYPOTHETICAL ANNUAL PREMIUMS FOR LTCI POLICIES THAT RETURN PREMIUMS UPON DEATH



Each panel represents a different pricing model specification. Each point represents a single insured in the portfolio of interest. There are three smoothed lines in each panel, one for each category. Each smoothed line is generated using a generalized additive model, regressing the lump-sum premium on age.

3.5 CONCLUDING REMARKS OF PART II

Part II implemented the post-processing approach introduced in Part I to examine how potential fairness adjustments can be applied to LTCI pricing. Using the HRS dataset, a three-state transition model was estimated to capture disability, recovery, and mortality dynamics across demographic groups. These transition intensities were then used to price hypothetical LTCI products under different policy designs. Lindholm et al.'s reweighting method was applied to adjust pricing outcomes, with sensitivity tests conducted for alternative choices of outcomes, reference distributions, and product features.

The findings indicate that post-processing may reduce potential systematic pricing disparities between demographic groups while maintaining actuarial consistency and transparency. The analysis also highlights practical considerations such as data quality, model assumptions, and the trade-off between predictive accuracy and potential fairness adjustment. Overall, the results demonstrate the feasibility of incorporating fairness-oriented post-processing methods within standard actuarial workflows. These findings provide a foundation for pricing methodologies and portfolio analysis in LTCI.

Section 4 Discussion

This section discusses practical implications and limitations of this study. Results indicate that using a post-processing approach enables modularization of potential fairness adjustments. This is crucial as actuarial teams often have best practices and established workflows which have been established over many working cycles and are sensitive to disruptions. An approach that allows the integration of fairness outcomes without requiring substantial changes to upstream processes, such as data cleaning and modeling, is therefore ideal. Having potential fairness adjustments occur towards the very end also preserves the flexibility of insurers to design new products.

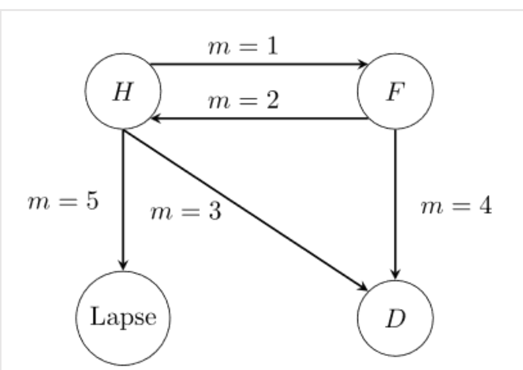
The Poisson-based framework used in this study appears flexible enough to accept any alternate multi-state framework. In particular, if one wishes to incorporate lapses into the pricing, they may be included as an additional absorbing state into the multi-state model (see Figure 19). In the absence of death benefits, one can essentially view the lapse transition rate as an addition onto healthy mortality rate. Further, the approach for modeling lapses may be tailored to whether it is recorded in the training data:

- If lapse data is unavailable, setting it equal to some assumed rate—e.g., from historical studies—would enable entering it into the pricing formula without explicitly modeling it in the estimation process.
- If the exact time of lapse is recorded, the estimation process may incorporate lapses as a fourth possible state. While covariates may be incorporated, it is recommended to keep the dependence as simple as possible due to lapses being much rarer compared to deaths.

Being a longitudinal survey on the general public, the HRS dataset used in this study does not reflect the demographic makeup of the insured population of any LTCI product. Thus, relationships estimated from the model used in this study reflect those which apply to the population at large and not necessarily the insured population underlying any given product or the broader LTCI industry. Thus, an insurer replicating this study using their own data would better reflect real situations in LTCI.

Figure 19

THE HEALTH STATE TRANSITION INCORPORATING LAPSE



Policyholders are assumed not to lapse if they are functionally disabled (i.e., receiving benefits).

Section 5 Acknowledgments

The authors' deepest gratitude goes to those without whose efforts this project could not have come to fruition: the volunteers who generously shared their wisdom, insights, advice, guidance, and review of this study prior to publication. Any opinions expressed may not reflect their opinions nor those of their employers. Any errors belong to the authors alone.

Project Oversight Group members:

Henry Chen, ASA, MAAA

Robert Eaton, FSA, MAAA

Rachel Erne, FSA, CERA, MAAA

Patricia Fay, FSA, MAAA

Sean Hilton, FSA, MAAA

Min Ji, FSA, FIA, MAAA

Eileen Luxton, FSA, FCIA

David Schaub, FSA, CERA, MAAA, AQ

At the Society of Actuaries Research Institute:

Lisa A. Schilling, FSA, EA, FCA, MAAA, Director, Practice Research

Barbara Scott, Senior Research Administrator

Appendix A Details of Fair Pricing Methods

A.1 UNINTENTIONAL INDIRECT ASSOCIATION

First, the authors clarify the notion of unintentional indirect association proposed in Lindholm et al. (2022). They noticed that even when S is omitted from a given model, certain values of $\mathbf{X}$ may result in

$$\mathbb{E}[Y|\mathbf{X}=\mathbf{x}] \approx \mathbb{E}[Y|\mathbf{X}=\mathbf{x}, S=s]$$

for some value of s. That is, conditioning on $\mathbf{X}$ may inadvertently use information on S in the model. They find that this is due to the tower property of expectation:

$$\mathbb{E}[Y|\mathbf{X}=\mathbf{x}] = \int_s \mathbb{E}[Y|\mathbf{X}=\mathbf{x}, S=s]\,\mathrm{d}P(S=s|\mathbf{X}=\mathbf{x}).$$

Thus, unintentional indirect association is due to the association between S and $\mathbf{X}$. They also define the notion of *a pricing function which does not utilize unintentional indirect association*, π, as one which is agnostic to the conditional distribution $S|\mathbf{X}$. More precisely, for probability measures $\mathbb{P}$ and $\mathbb{Q}$ with identical marginals for $\mathbf{X}$, S and the conditional distribution $Y|\mathbf{X}, S$ but possibly differing in their conditional distributions $S|\mathbf{X}$, π should produce prices

$$\boldsymbol{\pi}(\mathbf{X}, \mathbb{P}) = \boldsymbol{\pi}(\mathbf{X}, \mathbb{Q}).$$

A.2 THE PRE-PROCESSING METHODOLOGY OF LINDHOLM ET AL. (2024)

The authors now provide the details of the pre-processing methodology of Lindholm et al. (2024). Given the observed empirical distribution $\mathbb{P}$, the methodology functions by first identifying a candidate distribution $\mathbb{P}^{\perp}$ such that

- the marginal of $\mathbf{X}$ under $\mathbb{P}$ and $\mathbb{P}^{\perp}$ are in some sense similar,
- the marginal of S under $\mathbb{P}$ and $\mathbb{P}^{\perp}$ are identical, and
- $\mathbf{X} \perp S$ under $\mathbb{P}^{\perp}$.

Then, an optimal transport map r is identified which transforms $\mathbb{P}$ to $\mathbb{P}^{\perp}$. Then, letting $\mathbf{X}^{\perp} := r(\mathbf{X})$, the methodology outputs the data $(\mathbf{X}^{\perp}, Y)$. Given that $\mathbf{X}^{\perp} \perp S$, so will any function $\hat{Y} = g(\mathbf{X}^{\perp}) \perp S$, hence by definition any model trained on $(\mathbf{X}^{\perp}, Y)$ will satisfy strong demographic parity.

A.3 THE IN-PROCESSING METHODOLOGY OF BEUTEL ET AL. (2017)

Here the authors provide the details of the methodology of Beutel et al. (2017). Their model trains a representation Z towards two aims:

- It is informative for the response variable Y.
- It is free of information on the attribute S.

They thus propose a neural network with architecture as in Figure 1. The objective function of the model is given as

$$\min_{f,g} \max_{h} \left\{ L_Y\big(Y, g(Z)\big) + \lambda\, L_S\big(S, h(Z)\big) \right\}, \tag{3}$$

where L_Y is a loss function for Y, which may be mean-squared error in the regression setting or likelihood loss in the GLM setting, whereas L_S refers to cross-entropy loss. Each training step towards achieving the desired optimization is composed of the following two alternating steps:

- An adversarial gradient step to maximize (3) with respect to h and
- A model gradient step to minimize (3) with respect to (f, g).

By the alternating actions of the adversary and the model, ideally one obtains $Z \perp S$, while ensuring Z remains maximally predictive for both Y. As in the case of pre-processing, $Z \perp S$ implies $\hat{Y} = g(Z) \perp S$, i.e., demographic parity is achieved. In practice, neither demographic parity nor perfect accuracy is achievable, hence λ serves as a tuning parameter to balance between the two loss functions.

Appendix B Log-Likelihood Function for Estimating Transition Rates

This appendix formulates the log-likelihood for estimating transition rates in a multi-state model. Suppose there are a total of K individuals, M transition types, and I interview occasions. Recall that the HRS conducts follow-up interviews with respondents every two years. Let $t_{k,i}$ denote the time of the i^{th} interview for the k^{th} individual, and $\hat{t}_{k,i}$ the time of transition occurring between the i^{th} and the $(i+1)^{\text{th}}$ interviews, if any. Define two indicator variables: $T_{k,m,i} = 1$ if transition type m is observed between the i^{th} and $(i+1)^{\text{th}}$ interviews; $T_{k,m,i} = 0$ otherwise. $R_{k,m}(t) = 1$ if the k^{th} individual is exposed to the risk of transition type m at time t, $R_{k,m}(t) = 0$ otherwise.

Based on these definitions, the log-likelihood function is given by

$$l(\theta) = \sum_{k=1}^{K} \sum_{m=1}^{M} \sum_{i=1}^{I-1} l_{k,m,i}(\theta), \tag{4}$$

where

$$l_{k,m,i}(\theta) = T_{k,m,j} \ln \lambda_{k,m}\left(\hat{t}_{k,i}\right) - R_{k,m}\left(t_{k,i}\right) \int_{t_{k,i}}^{\min\{\hat{t}_{k,i}, t_{k,i+1}\}} \lambda_{k,m}(u)\, \mathrm{d}u - R_{k,m}\left(\hat{t}_{k,i}\right) \int_{\min\{\hat{t}_{k,i}, t_{k,i+1}\}}^{t_{k,i+1}} \lambda_{k,m}(u)\, \mathrm{d}u.$$

Appendix C Details of Unified Framework

Here the authors provide mathematical details on how the multi-state models can be viewed as independent Poisson regressions. To illustrate this, consider instead the simplified setting of a single transition for a single individual. As before, M types of transitions are possible, each with respective constant transition rate λ_m. Let T be the time of transition and let δ_m be an indicator for if the transition is of type m. Also, let $\boldsymbol{\theta}$ denote the model parameters. Then the likelihood is given by

$$L(\theta) = \exp\left(-T\sum_{m=1}^{M}\lambda_m\right)\prod_{m=1}^{M}\lambda_m^{\delta_m}.$$

Note that

$$L(\theta) \propto \prod_{m=1}^{M}\frac{\exp(-T\cdot\lambda_m)\,(T\cdot\lambda_m)^{\delta_m}}{\delta_m!},$$

which is equivalent to the likelihood of S independent Poisson processes with intensity λ_m and length T, and each having λ_m occurrences.

Note that the assumption of λ_m being constant is not a major limitation in the model given by Equation (2) because in practice, most time-varying covariates are piecewise constant. Hence, one can simply break up the model into the intervals on which the transition rates are constant.

This paper focuses on pricing an LTCI product based on information at underwriting. Thus, the focus is on only the single time-varying covariate of age x, which is taken to be age last birthday; non-age covariates are assumed to be static after policy inception and are denoted by $\boldsymbol{z}$. Further, each age x is associated with a partial exposure τ_x. The likelihood is then given by

$$\begin{aligned} L(\theta) &= \prod_{m\in M}\prod_{x\leq T}\exp\big(-\tau_x\lambda_m(x,\boldsymbol{z})\big)\big(\lambda_m(x,\boldsymbol{z})\big)^{\delta_m} \\ &\propto \prod_{s\in S}\prod_{x\leq T}\frac{\exp\big(-\tau_x\lambda_m(x,\boldsymbol{z})\big)\big(\tau_x\lambda_m(x,\boldsymbol{z})\big)^{\delta_m}}{\delta_m!}, \end{aligned}$$

where

$$M = \begin{cases}\{1,3\}, & \text{current state is } H,\\ \{2,4\}, & \text{current state is } F.\end{cases}$$

This allows using the Poisson regression model of choice, e.g., boosting, random forest, generalized additive models or neural networks to model $\lambda_m(x,\boldsymbol{z})$, with τ_x serving as an offset to $\ln\lambda_m$. To facilitate estimation under this framework, entries in the transition dataset are structured such that the age x remains constant throughout the time period spanned by the entry.

Appendix D Validating Removal of Job Physicality

Consider the model fitted for each of the four transitions. In a likelihood model, the deviance measures how much the current likelihood differs from the ideal case, where every prediction is perfectly accurate. The null deviance measures the deviance of the intercept-only model, whereas the residual deviance measures the deviance of the currently fitted model. The pseudo-R^2 is then measured as

$$R^2_{\text{pseudo}} = 1 - \frac{\text{Deviance}_{\text{residual}}}{\text{Deviance}_{\text{null}}}$$

Given that the data volume differs between the case with and without job physicality, the pseudo-R^2 allows a standardized comparison between the two.

Table 3

LIKELIHOODS OF MODEL FIT, WITH AND WITHOUT JOB PHYSICALITY

	With RxJPHYSTOT			Without RxJPHYSTOT		
Transition	Null	Residual	Pseudo-R^2	Null	Residual	Pseudo-R^2
Healthy → Disabled	98,818	90,343	0.0858	112,803	102,948	0.0874
Disabled → Healthy	51,588	48,086	0.0679	61,361	56,812	0.0741
Healthy → Dead	41,139	36,215	0.1197	46,267	40,753	0.1192
Disabled → Dead	41,177	37,177	0.0971	53,874	48,713	0.0958
Total	232,722	211,821	0.0898	274,305	249,226	0.0914

Table 3 shows the comparison. One can see that when comparing the two model-by-model, the pseudo-R^2 is sometimes higher for the model without job physicality and sometimes higher for the model with. However, overall, it is higher for the model without job physicality. Since this also yields the dataset with higher data volume, the model without job physicality was chosen for this study.

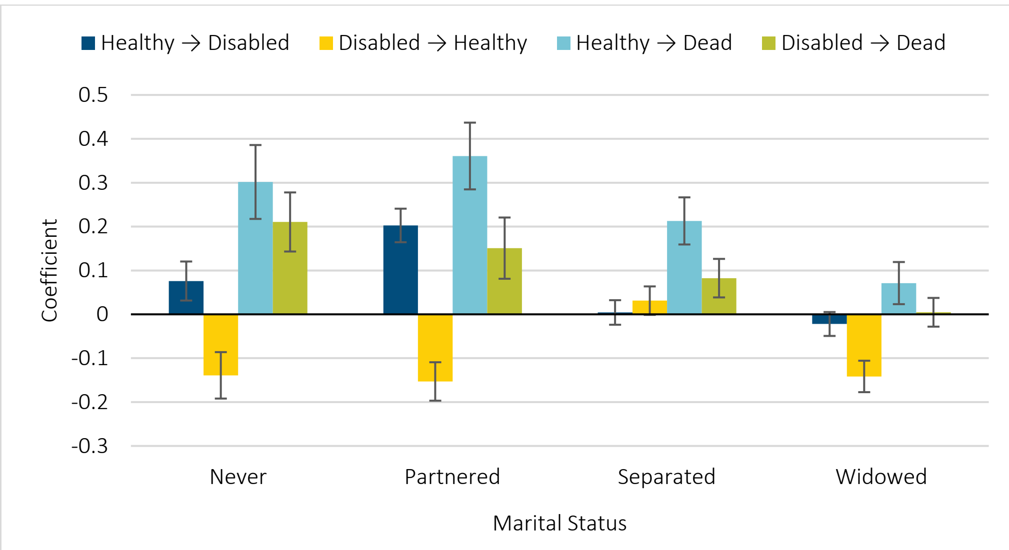
Appendix E Other Estimated Coefficients

This section reports coefficients in the fitted models which are not displayed in section 3.2.4. These include the categorical variables of marital status (MSTAT), census region (CENREG), labor force participation (LBRF), and indicators for various diseases or lifestyle habits. Coefficients for other continuous variables are also reported, including non-housing wealth, body mass index (BMI), income, and number of members of the household.

For the variable of marital status (MSTAT), this study uses as the baseline the most common category of MARRIED. As expected, shows that the baseline has largely the lowest rates of transitioning from healthy to disabled, as well as the highest rates of recovery; its rates of mortality are also largely the lowest. PARTNERED appears to have the largest difference overall with MARRIED, followed by NEVER. Of the remaining two, SEPARATED has the larger difference in terms of both healthy and disabled mortality, whereas WIDOWED has the larger difference in terms of having lower recovery rate.

Figure 20

COEFFICIENTS FOR MARITAL STATUS

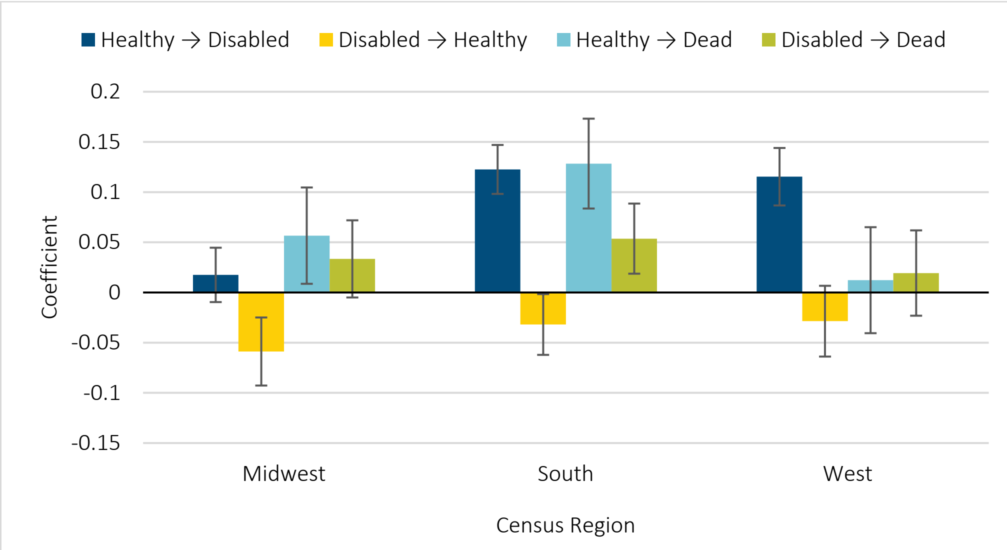


The reference category is married.

For census region (CENREG), the census region of Northeast is taken as the baseline. The researchers find that coefficients tend to be quite small, and many differences tend not to be particularly significant. The region with the largest difference with the baseline appears to be the South. Rates of transitioning to disabled, as well as healthy and disabled mortality are significantly higher for the South than for other regions. This is followed by the West, which is only significantly higher in rates of transitioning to disabled, whereas the Midwest appears to be the most similar to the Northeast.

Figure 21

COEFFICIENTS FOR CENSUS REGION

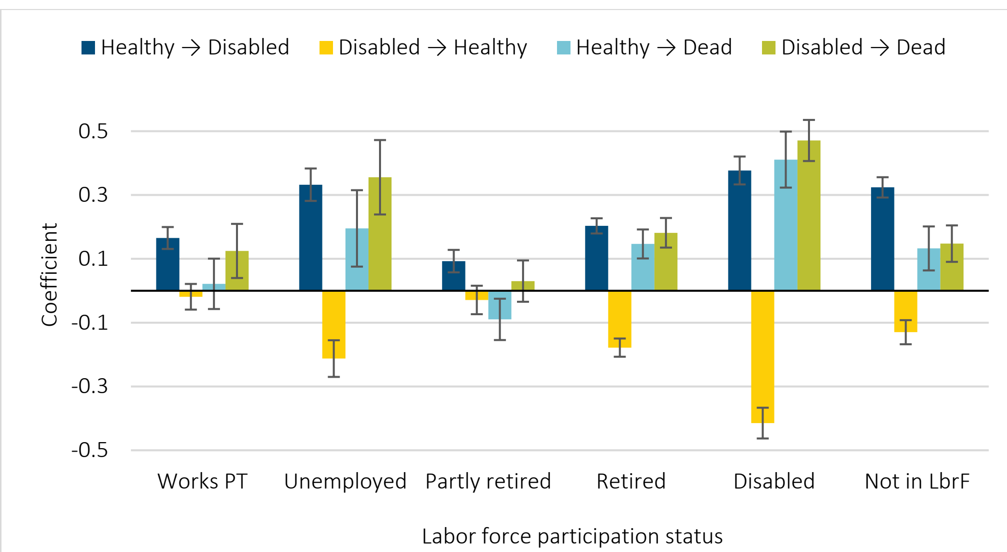


The reference category is Northeast.

Turning now to the variable of labor force participation (LBRF), the baseline is those working full-time. Relative to this baseline, those who are disabled exhibit the largest differences, with significantly higher rates of transitioning to disabled as well as both rates of mortality, whereas recovery rates are significantly lower. This seems to be followed by those who are unemployed, those who are retired or not in the labor force, and then those who are working part-time, and finally those who are partially retired.

Figure 22

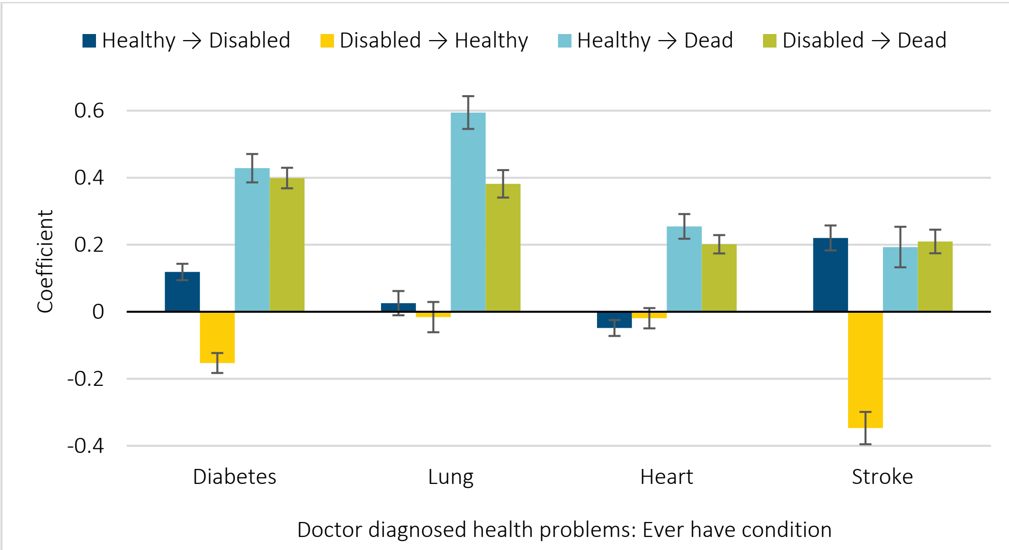
COEFFICIENTS FOR LABOR FORCE PARTICIPATION STATUS



PT = part-time. The reference category is full-time employment.

Next, consider the indicators for the presence of various diseases (Figure 23). These conditions are not mutually exclusive. Diseases that have a substantial impact on mortality include lung conditions, followed by diabetes, heart conditions, and stroke. Stroke has by far the strongest effect in reducing recovery and increasing disability rates, followed by diabetes.

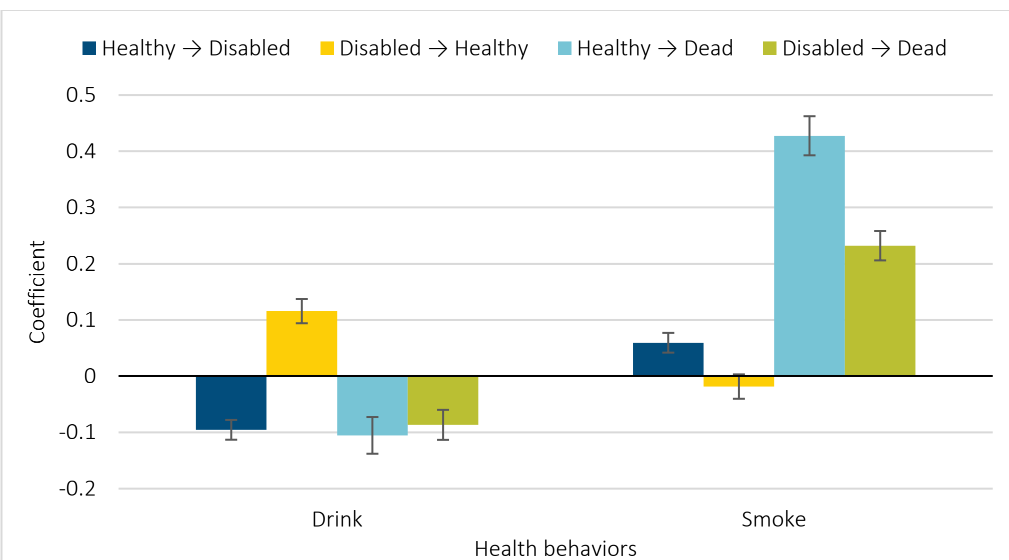
Figure 23
COEFFICIENTS FOR VARIOUS DISEASES BEING PRESENT



The health problems are not mutually exclusive. The reference category for each health problem is never diagnosed.

Figure 24 displays the coefficients related to health behaviors, including drinking and smoking. Relative to non-smokers, smokers exhibit higher disability and mortality rates and lower recovery rates, as expected. Interestingly, individuals who report drinking have lower disability and mortality rates and higher recovery rates than those who do not. This finding reflects the mixed results in the literature regarding the health implications of moderate alcohol consumption (see Yu et al., 2022 and references therein).

Figure 24
COEFFICIENTS FOR HEALTH BEHAVIORS



The remaining continuous coefficients and their contributions to the log-likelihood are displayed in Table 4. Age is by far the most important of these, especially in terms of its contribution to the log-likelihood; the coefficients are positive for both disability and mortality, and negative for rates of recovery, which intuitively makes sense as health deteriorates with age.

This is followed by non-housing wealth; the interaction effect is significant for both rates of disability and healthy mortality, but otherwise the most important covariate is $\ln(1+|W|)$. The sign of the coefficients largely suggests that health improves with increasing wealth. Total household income is much less important, with it only contributing significantly to likelihood for rates of disability.

Interestingly, while intuitively a higher BMI is indicative of higher obesity and worse health, the estimated coefficients suggest it is associated with *lower* mortality and *higher* rates of recovery. Finally, the number of members in the household is largely unimportant in any of the four transition models.

Table 4
TABLE OF CONTINUOUS COEFFICIENTS

Variable	Healthy → Disabled	Disabled → Healthy	Healthy → Dead	Disabled → Dead
Age	0.0482	-0.0436	0.0848	0.0725
	(1177.17)	(800.72)	(987.08)	(1166.23)
$\text{sgn}(W)$	0.1891	-0.0083	0.1975	0.0291
	(22.69)	(0.03)	(5.27)	(0.32)
$\ln(1+\|W\|)$	-0.0414	0.0275	-0.0327	-0.0111
	(65.98)	(20.88)	(9.13)	(2.42)
$\text{sgn}(W)\cdot\ln(1+\|W\|)$	-0.0274	0.0023	-0.0286	-0.0092
	(36.25)	(0.19)	(8.49)	(2.02)
BMI	0.0008	0.005	-0.0116	-0.0136
	(0.13)	(4.18)	(6.73)	(16.30)
$\ln(1+\text{Income})$	-0.0256	0.0038	-0.0134	0.0046
	(9.25)	(0.13)	(0.42)	(0.08)
Number of people living in the household	0.0032	-0.0122	-0.0129	0.0057
	(0.11)	(1.14)	(0.94)	(0.15)

The contribution to the likelihood of each coefficient, i.e., the reduction in likelihood in the final model were the covariate to be removed, is given below in parentheses. W stands for non-housing wealth.

Appendix F Deriving the Annual Premium for Policies with a Death Benefit Rider

A key step in deriving the annual premiums for policies that include a death benefit rider is to determine the distribution of the number of premiums paid by time t. Let N_t denote the number of premiums paid immediately after time t, where $t = 0, 1, \ldots, 110 - x$ for a policy issued to an individual aged x. Then, recursive formulas are used to obtain the joint distribution of (N_t, J_t), where J_t represents the health state at time t.

Since the policy is issued to healthy individuals, the initial conditions are given by

$$\Pr(N_0 = 1, X_0 = H) = 1, \Pr(N_0 = 1, X_0 = F) = 0, \Pr(N_0 = 1, X_0 = D) = 0.$$

For $t = 1, \ldots, 110 - x$, the following recursions hold:

$$\begin{aligned}
\Pr(N_t = k, X_t = H) &= \Pr(N_{t-1} = k-1, X_{t-1} = H)\ p_{t-1}^{HH} + \Pr(N_{t-1} = k-1, X_{t-1} = F)\ p_{t-1}^{FH}, k = 2,3, \ldots, t+1; \\
\Pr(N_t = k, X_t = F) &= \Pr(N_{t-1} = k, X_{t-1} = H)\ p_{t-1}^{HF} + \Pr(N_{t-1} = k, X_{t-1} = F)\ p_{t-1}^{FF}, k = 1,2, \ldots, t; \\
\Pr(N_t = k, X_t = D) &= \Pr(N_{t-1} = k, X_{t-1} = H)\ p_{t-1}^{HD} + \Pr(N_{t-1} = k, X_{t-1} = F)\ p_{t-1}^{FD}, k = 1,2, \ldots, t,
\end{aligned}$$

where $p_{t-1}^{ij} = \Pr(J_t = j \mid J_{t-1} = i)$.

Once the probabilities $\Pr(N_t = k, J_t = D)$are calculated, the expected present value of returned premiums can be expressed as

$$\text{EPV(returned premiums, 1 per year)} = \sum_{t=1}^{100-x} v^t \sum_{k=1}^{t} k \Pr(N_t = k, J_t = D).$$

The correctness of the probabilities $\Pr(N_t = k, J_t = D)$ can be verified by confirming that

$$\sum_{k=1}^{t} \Pr(N_t = k, J_t = D) = \Pr(J_t = D).$$

The level annual premium is given by

$$\text{Level annual premium} = \frac{\sum_{t=0}^{110-x} v^t \Pr(J_{x+t} = F \mid J_x = H)}{\sum_{t=0}^{110-x} v^t \Pr(J_{x+t} = H \mid J_x = H) - \sum_{t=1}^{100-x} v^t \sum_{k=1}^{t} k \Pr(N_t = k, J_{x+t} = D)}.$$

References

Administration for Community Living, 2020a. How much care will you need? (Web page). U.S. Department of Health and Human Services, Washington, D.C.

Administration for Community Living, 2020b. Receiving long-term care insurance benefits (Web page). U.S. Department of Health and Human Services, Washington, D.C.

Beutel, A., Chen, J., Zhao, Z., Chi, E.H., 2017. Data decisions and theoretical implications when adversarially learning fair representations (Preprint No. arXiv:1707.00075).

Campbell, E., Czajkowski, J., Mitchell, S., Nordman, E., Roland, C., Tetrault, P., 2020. Milestones in racial discrimination within the insurance sector (Technical Report). National Association of Insurance Commissioners and the Center for Insurance Policy and Research, Kansas City, MO.

Chibanda, K.F., 2022. Defining discrimination in insurance (Technical Report), Research Paper Series on Race & Insurance Pricing. Casualty Actuarial Society, Arlington, VA.

Cottin, A., Pecuchet, N., Zulian, M., Guilloux, A., Katsahian, S., 2022. IDNetwork: A deep illness-death network based on multi-state event history process for disease prognostication. Statistics in Medicine 41, 1573–1598. https://doi.org/10.1002/sim.9310

Dwork, C., Hardt, M., Pitassi, T., Reingold, O., Zemel, R., 2012. Fairness through awareness, in: Proceedings of the 3rd Innovations in Theoretical Computer Science Conference, ITCS '12. Association for Computing Machinery, Cambridge, MA, pp. 214–226. https://doi.org/10.1145/2090236.2090255

European Commission, 2012. Guidelines on the application of Council Directive 2004/113/EC to insurance, in the light of the judgment of the Court of Justice of the European Union in Case C-236/09 (Test-Achats) (Official Journal of the European Union). European Union.

Fong, J.H., Mitchell, O.S., Koh, B.S.K., 2015a. Disaggregating activities of daily living limitations for predicting nursing home admission. Health Services Research 50, 560–578. https://doi.org/10.1111/1475-6773.12235

Fong, J.H., Shao, A.W., Sherris, M., 2015b. Multistate actuarial models of functional disability. North American Actuarial Journal 19, 41–59. https://doi.org/10.1080/10920277.2014.978025

Hoffmann, R., 2011. Socioeconomic inequalities in old-age mortality: A comparison of Denmark and the USA. Social Science & Medicine, Part Special Issue: Analysing global health assistance 72, 1986–1992. https://doi.org/10.1016/j.socscimed.2011.04.019

Hothorn, T., Bühlmann, P., Dudoit, S., Molinaro, A., Van Der Laan, M.J., 2006. Survival ensembles. Biostatistics 7, 355–373. https://doi.org/10.1093/biostatistics/kxj011

Hoy, M., Ruse, M., 2005. Regulating Genetic Information in Insurance Markets. Risk Management and Insurance Review 8, 211–237. https://doi.org/10.1111/j.1540-6296.2005.00057.x

Ishwaran, H., Kogalur, U.B., Blackstone, E.H., Lauer, M.S., 2008. Random survival forests. Annals of Applied Statistics 2, 841–860.

Langa, K.M., Weir, D.R., Kabeto, M., Sonnega, A., 2020. Langa-Weir classification of cognitive function (1995 onward) (Survey Research Center). Institute for Social Research, University of Michigan.

Leong, J., Moncher, R., Jordan, K., 2024. A practical guide to navigating fairness in insurance pricing. (Technical Report), Research Paper Series on Race & Insurance Pricing. Casualty Actuarial Society, Arlington, VA.

Li, Z., Shao, A.W., Sherris, M., 2017. The impact of systematic trend and uncertainty on mortality and disability in a multistate latent factor model for transition rates. North American Actuarial Journal 21, 594–610. https://doi.org/10.1080/10920277.2017.1330157

Lindholm, M., Richman, R., Tsanakas, A., Wüthrich, M.V., 2024. What is fair? Proxy discrimination vs. demographic disparities in insurance pricing. Scandinavian Actuarial Journal 2024, 935–970. https://doi.org/10.1080/03461238.2024.2364741

Lindholm, M., Richman, R., Tsanakas, A., Wüthrich, M.V., 2022. Discrimination-free insurance pricing. ASTIN Bulletin 52, 55–89. https://doi.org/10.1017/asb.2021.23

Madras, D., Creager, E., Pitassi, T., Zemel, R., 2018. Learning adversarially fair and transferable representations, in: Dy, J., Krause, A. (Eds.), Proceedings of the 35th International Conference on Machine Learning, Proceedings of Machine Learning Research. PMLR, pp. 3384–3393.

Schilling, L.A., 2023. Challenges and opportunities with rethinking fairness metrics for life insurance processes: An actuarial perspective (Technical Report). Society of Actuaries Research Institute, Schaumburg, IL.

Schraub, D., Lang, J., Zhang, Z., Sayre, M.A., 2024. Comparison of regulatory framework for non-discriminatory AI usage in insurance (Technical Report). Society of Actuaries Research Institute, Chicago, IL.

Smith, L.T., Pirchalski, E., Golbin, I., 2022. Avoiding unfair bias in insurance applications of AI models (Technical Report). Society of Actuaries Research Institute, Schaumburg, IL.

Wang, Q., Hanewald, K., Wang, X., 2022. Multistate health transition modeling using neural networks. Journal of Risk and Insurance 89, 475–504. https://doi.org/10.1111/jori.12364

World Health Organization, 2025. Ageing and health (Fact sheets). World Health Organization, Geneva, Switzerland.

Xu, M., Sherris, M., Meyricke, R., 2019. Systematic mortality improvement trends and mortality heterogeneity: insights from individual-level HRS data. North American Actuarial Journal 23, 197–219. https://doi.org/10.1080/10920277.2018.1513369

Yu, D., Lu, B., Piggott, J., 2022. Alcohol consumption as a predictor of mortality and life expectancy: Evidence from older Chinese males. The Journal of the Economics of Ageing 22, 100368. https://doi.org/10.1016/j.jeoa.2022.100368

About The Society of Actuaries Research Institute

Serving as the research arm of the Society of Actuaries (SOA), the SOA Research Institute provides objective, data-driven research bringing together tried and true practices and future-focused approaches to address societal challenges and your business needs. The Institute provides trusted knowledge, extensive experience and new technologies to help effectively identify, predict and manage risks.

Representing the thousands of actuaries who help conduct critical research, the SOA Research Institute provides clarity and solutions on risks and societal challenges. The Institute connects actuaries, academics, employers, the insurance industry, regulators, research partners, foundations and research institutions, sponsors and non-governmental organizations, building an effective network which provides support, knowledge and expertise regarding the management of risk to benefit the industry and the public.

Managed by experienced actuaries and research experts from a broad range of industries, the SOA Research Institute creates, funds, develops and distributes research to elevate actuaries as leaders in measuring and managing risk. These efforts include studies, essay collections, webcasts, research papers, survey reports, and original research on topics impacting society.

Harnessing its peer-reviewed research, leading-edge technologies, new data tools and innovative practices, the Institute seeks to understand the underlying causes of risk and the possible outcomes. The Institute develops objective research spanning a variety of topics with its strategic research programs: aging and retirement; actuarial innovation and technology; mortality and longevity; diversity, equity and inclusion; health care cost trends; and catastrophe and climate risk. The Institute has a large volume of topical research available, including an expanding collection of international and market-specific research, experience studies, models and timely research.

Society of Actuaries Research Institute
8770 W Bryn Mawr Ave, Suite 1000
Chicago, IL 60631
www.SOA.org