Integrating Frailty and Physical Resilience to Improve Risk Estimation in Aging Populations

by

Nathan Kyle Smith

Submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy

at

Dalhousie University
Halifax, Nova Scotia
April 2024

Table of Contents

List of Tables

List of Figures

Abstract

Background

Physical resilience, broadly defined as the ability to resist or recover from health stressors, is a relatively new concept in the field of aging, yet, anticipated as a potentially game-changing idea. In particular, physical resilience has significant potential to complement the widely used concept of frailty. To date, conceptual and methodological difficulties have hampered such potential. Frailty and physical resilience are broad and closely related concepts with multiple interpretations, many approaches to measure physical resilience have been proposed, and empirical studies have rarely investigated frailty and physical resilience together.

Project Aim

Enrich our understanding of the relationship between frailty and physical resilience in the health of aging populations by providing a novel, integrated framework and a concurrent empirical investigation of the two concepts in longitudinal population data.

Specific Objectives

1. Guided by the integrated framework, operationalize multiple specific measures of frailty and physical resilience and provide a descriptive analysis of each measure.
2. Evaluate the relationship between frailty, physical resilience, and mortality.
3. Evaluate the relationship between frailty, physical resilience, and acute functional recovery.

Methods

This study uses repeated measurements of a 41-item frailty index (FI) from waves 3-13 of the Health and Retirement Study in the United States to operationalize three indicators of resilience: the frailty-disease mismatch (FM), the rate of aging (RoA), and a dynamical indicator of resilience (DIOR-FI). This study evaluates each measure using descriptive statistics and logistic regression models estimating the probability of 2-year all-cause mortality (n=27,744) and full functional recovery after myocardial infarction (n=1,905).

Results

Resilience indicators generally had low agreement (kappa ≤0.24) and moderate to weak correlation (Pearson r ≤0.53). All indicators showed statistically significant associations with mortality and recovery. Despite requiring careful interpretation, FM shows the greatest promise for adding predictive ability beyond age, sex, and frailty.

Conclusion

This study demonstrates key insights and lessons learned for future research on frailty and physical resilience. With further refinement of the methods proposed in this thesis, the combination of population data (for estimating FM) and routinely collected health data (for estimating RoA and DIOR-FI) offer promising opportunities to improve risk estimation in aging populations.

List of Abbreviations Used

ADL	Activities of Daily Living
AHEAD	Asset and Health Dynamics Among the Oldest Old (cohort)
AIC	Akaike Information Criterion
AUC	Area Under the Curve
CFS	Clinical Frailty Scale
DIOR	Dynamical Indicator of Resilience
DIOR-FI	Frailty Index Dynamical Indicator of Resilience
FI	Frailty Index
FM	Frailty-Disease Mismatch
HRS	Health and Retirement Study
IADL	Instrumental Activities of Daily Living
ICC	Intraclass Correlation Coefficient
IQR	Interquartile Range
ISRF-FPR	Integrated Stress Response Framework for Frailty and Physical Resilience
LCA	Latent Class Analysis
MI	Myocardial Infarction
PCA	Principal Component Analysis
PRIFOR	Physical Resilience Instrument for Older Adults
Q1	First quartile (25^{th} percentile)
Q3	Third quartile (75^{th} percentile)
RoA	Rate of Aging
SAVE	Scale of Aging and Vigor Epidemiology (frailty scale)
SBBP	Short Physical Performance Battery
SD	Standard Deviation

SF-36 PFS	Short Form 36 Survey Physical Function Subscale
SHARE	Survey of Health, Aging, and Retirement in Europe

Acknowledgements

I would like to express my sincerest gratitude to my supervisors, Dr. Yukiko Asada and Dr. Susan Kirkland, as well as my committee members, Dr. Michel Grignon and Dr. Cindy Feng, for their unwavering patience and diligent efforts in supporting the development of this thesis. I would also like to extend my thanks to Dr. George Kephart and Dr. Finlay Maguire for providing insightful feedback during the final stages of review. Lastly, I would like to thank all the supportive faculty and staff in the Department of Community Health and Epidemiology.

Funding support for this thesis was provided by the Dalhousie Faculty of Medicine 2022 Graduate Studentship.

Chapter 1: Introduction

Today, older adults are living longer than previous generations. As a result, an increasing number of individuals are reaching advanced ages (80+), driving unique healthcare needs and considerations. Increased general vulnerability at these advanced ages means that treating diseases individually is no longer sufficient. As the limits of human longevity continue to be tested, we approach the brink of what is likely to be the next epidemiological transition: from chronic disease to frailty. Broadly considered as an aging-related state of increased vulnerability to stressors, frailty has emerged as a key concept in understanding the health of older adults. This vulnerability results in an increased risk of a negative health outcome (e.g., hospitalization, death, etc.) after encountering a stressor (e.g., fall, infection, minor surgery, etc.). Thus, a frail individual is an individual who is likely to end up being hospitalized or dying after a seemingly banal event that would not have triggered a serious response in a less frail individual. Without intervention, the typical progression a frail individual will experience is functional decline, disability, loss of independence, and, eventually, death. Morley et al. describe this as the frailty cascade (1). With our population living longer, more adults than ever experience and succumb to frailty. In Canada, as elsewhere globally, an aging population is changing the needs of its healthcare system, and addressing the complex health needs of this population has become a priority. Better understanding of aging-related vulnerability is key to address the needs of this growing population.

To date, measures of frailty have been widely and effectively used in healthcare settings and health research for patient risk stratification, prognosis, and risk assessment (2). Knowledge of frailty has improved our ability to predict adverse events, yet unexplained heterogeneity in outcomes remains. In terms of prediction of adverse outcomes, the best performing measures of frailty typically have an area under the receiving operating characteristic curve (AUC, a measure of discrimination) between 0.6 to 0.8, depending on the outcome, length of follow-up, and population (3–6). Though variation exists, values in this range are often described as poor to good discriminatory accuracy (7). The need for a deeper understanding of an older individual's risk of adverse outcomes has recently been highlighted by Andrew et al. (8). Noting heterogeneous outcomes in COVID-19 severity among frail individuals in long term care facilities, Andrew et al.

call for moving beyond a simple understanding and grading of vulnerability and suggest that understanding resilience is key.

The concept of resilience, broadly defined as the ability to resist or recover from stress, has a long history of application in various fields such as psychology and engineering but has only recently been applied to physical health and aging. Specifically, a new construct termed physical resilience has been proposed as a whole-person level characteristic which determines an individual's ability to resist or recover from functional decline following a stressor (9). This concept has particular relevance to understanding heterogeneous outcomes in frail individuals: an important feature in the frailty cascade is the potential to recover rather than progress down the cascade (1). Being able to identify who is most likely to recover would improve estimation of risk associated with frailty and, thus, would help reduce unexplained heterogeneity in outcomes.

The concept of physical resilience has clear potential to complement and expand upon what frailty offers; it has been proposed to assist in clinical decision making, developing care models, and identifying preventative strategies (10). Recently, there has been an increasing interest in how the concepts of physical resilience and frailty fit together (11–14). However, despite numerous editorials and conceptual articles, very few empirical studies have measured both frailty and physical resilience simultaneously, and to the best of my knowledge, none have provided a comparison of their ability to predict adverse outcomes. This leaves many questions unanswered, and the complementary potential of physical resilience and frailty has yet to be realized. To address this gap in the literature, both conceptual and methodological challenges need to be addressed. For example, frailty and physical resilience are broad and closely related concepts with multiple interpretations. This leads to a lack of conceptual clarity between the two concepts, which is further exacerbated by the existence of additional similar aging-related concepts (e.g., intrinsic capacity, biological aging, etc.). The lack of conceptual clarity then leads to a frequent disconnect between empirical work and underlying theory. In terms of methodological issues, owing to the relative novelty of the concept of physical resilience, empirical measurement approaches have yet to be firmly established. For example, given that physical resilience is a multidimensional construct, researchers have used various operationalizations in their empirical work. Furthermore, not only are empirical studies evaluating both frailty and physical resilience lacking, but there is also a lack of studies comparing these different measurement approaches for physical resilience. Thus, not much is known about how these different approaches relate to one

another, nor their relative performance. Understanding the individual and combined performance of frailty and physical resilience to predict health outcomes has significant potential to improve the identification of at-risk groups and individuals, which, in turn, can aid clinical decision making and resource planning and allocation.

With the goal of better understanding the relationship between frailty, physical resilience, and health risk in aging populations, this thesis aims to address these challenges by providing a novel integrated conceptual framework and a concurrent empirical analysis to evaluate the relationship between multiple measures of frailty, physical resilience, and two key outcomes: mortality and recovery. To accomplish this goal, this thesis will proceed as follows: Chapter 2 reviews existing conceptual and empirical work. Chapter 3 synthesizes the information from the review into an integrated framework to bridge the gap between concepts and measurement approaches. Chapter 4 states the specific objectives of the empirical analysis. Chapter 5 describes the operationalization of the measurement approaches highlighted in the integrated framework and the empirical methods to investigate the individual and combined effects of frailty and physical resilience on estimates of mortality and recovery. Chapter 6 presents the results of the empirical investigation. Chapter 7 concludes with the discussion, contextualizing the results and providing recommendations for future work.

Chapter 2: Literature Review

This chapter discusses concepts and measurement of both frailty and physical resilience. Given that concepts and measurement of frailty are well established but those of physical resilience are not, this chapter begins with a brief overview of frailty (section 2.1), then, proceeds with a review of the conceptual literature on physical resilience (section 2.2), followed by a review of studies employing an empirical measurement of physical resilience (section 2.3). The literature review explores the relationship between frailty and physical resilience where applicable. Table 1 provides a glossary of key terms introduced in the review of the conceptual literature on physical resilience (section 2.2).

Table 1. Glossary of Key Terms Introduced in the Conceptual Review

Term	Definition
Actualization	In the context of stress response, actualization refers to the observed (realized), post-stress response. An observed stress response is the actualization (or realization) of the pre-stressor potential to respond.
Frailty	A state of low reserve across multiple organ systems which leads to vulnerability to stressors. Frail individuals are more likely to experience decline after encountering a stressor.
Intrinsic capacity	"The composite of all the physical and mental capacities of an individual" (World Health Organization). Intrinsic capacity takes a positive ability perspective, focuses on specific domains, and emphasizes longitudinal monitoring. A person with high intrinsic capacity is less likely to experience decline after encountering a stressor.
Physical resilience	The ability to resist or recover from the negative health effects of a stressor. Physical resilience is constrained by the level of reserve but not wholly determined by it. Physical resilience can be conceptualized at the overall, whole-person level, or at the level of specific organs/organ systems.
Potential	In the context of stress response, potential refers to what can be estimated prior to experiencing a stressor. Pre-stress reserve represents one's potential to respond to a stressor. This potential can only be realized after experiencing a stressor.
(Physiologic) Reserve	A term used to refer to the resources the body has to deal with incoming stress. Reserve can be conceptualized at the overall,

Term	Definition
	whole-person level, or at the level of specific organs/organ systems. Exact definitions vary, and the literature sometimes uses synonyms such as “functional reserves” or “reserve capacity”. Reserve is the central linking construct between aging-related concepts.
Robustness	The ability to resist decline after experiencing a health stressor. Robustness can be considered to be the opposite of frailty. Robustness can also be considered as resistance to a stressor and quantified as the time to and peak magnitude of perturbation.
Stressors	Anything that poses a challenge to the body. This includes relatively minor stressors that can cause a system to reach a tipping point, resulting in a state change to a worse health state (e.g., exercise or drinking alcohol). Additionally, the state changes themselves (often entailing a major health event) can further be considered a stressor on the body (e.g., major health events such as stroke or hip fracture, or clinical stressors such as surgery or chemotherapy).
Stressors, acute	Stressors that are short-lived, usually have a higher intensity/severity, and have more immediate effects on the body (e.g., stroke, and hip fracture, and surgery).
Stressors, chronic	Stressors that are repeated or extended over time, are typically lower in intensity, and have effects on the body that occur over a longer period (e.g., dialysis and chemotherapy).
Tipping Point	A threshold that determines how much stress a system can take before a state change occurs. If the stress threshold is exceeded, the tipping point is reached, typically leading to a worse health state. Stress thresholds are determined by reserve. A frail individual is considered to be highly prone to experiencing tipping point events and transitioning to a worse health state.
Vulnerability	Higher risk of experiencing a negative effect from a stressor. Vulnerable individuals are more likely to experience declines in health after encountering a stressor.

2.1 An Overview of Frailty

2.1.1 Concepts of Frailty

The term “frailty” was first introduced in 1979 in the demography literature by Vaupel et al. (15) as a way to account for unexplained heterogeneity in mortality risk among individuals of the same

chronological age. Today, the term has been generalized to explain differences in risk of adverse outcomes between individuals with similar exposures (including age) (16). Frailty is conceptualized as a state of low physiologic reserve which leads to increased vulnerability to stressors, and thus, higher risk of adverse outcomes. Physiologic reserve can broadly be considered to represent the physiologic resources an individual has to handle incoming stressors. Frailty can be considered as an inherently dynamic concept because it relies on the idea of stress response. However, conventional measurement approaches are static, as they are not estimated in relation to time or stressors (measurement of frailty is described in section 2.1.2).

Despite this relatively agreed-upon conceptual definition of frailty, there has been much debate over how best to operationalize this concept. Much of this debate has focused on two conceptual perspectives: the phenotype model (17) and the deficit accumulation model (18). The phenotype model is based on the idea that sarcopenia (the progressive age-related loss of skeletal muscle mass (19)) is the central impairment on which the multifactorial causal pathways of frailty converge and, in turn, this impairment perpetuates the progression of frailty (17). This idea is reflected by common phenotypic criteria that capture strength, physical activity, and energy levels. On the other hand, the deficit accumulation model states that frailty results from the accumulation of age-related health deficits (18). Deficits can include signs and symptoms, diseases, disabilities, or abnormalities in medical/laboratory tests (e.g., high cholesterol) (20). Rockwood and Mitnitski (21) suggest that deficit accumulation is the basis for loss of physiologic reserve and, thus, is indistinguishable from loss of reserve; as more deficits accumulate, both the level of frailty and risk of adverse outcomes increase.

Both models are based on the presence or absence of impairments. However, the phenotype model views frailty as a pre-disability syndrome that may manifest in the absence of clinical disease (22). In contrast, the cumulative deficit model views frailty as a state in which both clinical disease and disability can be incorporated (i.e., as deficits) (18). Thus, it is possible that the two approaches identify different types of vulnerability (22).

2.1.2 Measurement of Frailty

Several approaches to measure frailty have been well characterized and firmly established (23), but this review will only focus on two: the frailty phenotype, representing the phenotype model,

and the frailty index (FI), representing the cumulative deficit model (24). Although much variation exists, many other frailty measures can be considered variations of these two. The frailty phenotype is defined by five criteria: unintentional weight loss, exhaustion, weak grip strength, slow walking speed, and low physical activity (25). This discrete measure classifies an individual as frail if they meet three or more of these criteria. In contrast, the frailty index is a continuous measure defined as the proportion of deficits an individual has accumulated. A minimum of 30 deficits are recommended, and the specific deficits used do not matter as long as the deficits cover several domains, such as self-assessed health status, cognition, chronic conditions, function, and physical performance measures (20). The interchangeable nature of the deficits included in frailty index makes it very flexible, allowing construction in numerous different data sources. The deficits chosen should cover a wide range of bodily systems, be related to health, and increase in prevalence with age (20). The frailty index ranges from 0 (indicating no deficits) to 1 (indicating all deficits), but it has been consistently observed that the empirical upper limit for human survival is approximately 0.70 (26). Interestingly, the frailty index has been proposed as a measure of biological age[1] (30) and is not limited to use in older adults (22). In fact, multiple studies have demonstrated that after accounting for chronological age, the frailty index has the strongest independent relationship with mortality compared to other estimators of biological age, such as those estimated by a battery of biomarkers including the Klemera and Doubal method, DNA methylation, and telomere length (31–33). Thus, biological aging can be thought of as the cumulative loss of reserve, which can be quantified by the accumulation of deficits (i.e., the frailty index). The corresponding loss in reserve across multiple systems eventually leads to a state where the body has an impaired ability to effectively compensate for stressors. Thus, the individual is at increased risk of adverse outcomes, including hospitalization and death.

It has been demonstrated that these two approaches identify different groups of individuals as robust and frail (5,34). In fact, one study found the prevalence of frailty to be drastically different depending on the approach used: 3.6% using the frailty phenotype, and 34% using the frailty index (34). Similarly, a study of eight different frailty scales (including the frailty phenotype and frailty

[1] Similar to frailty, the concept of biological age (also referred to as functional or physiological age) was introduced to account for the observed heterogeneity in health status among individuals of the same chronological age (27). Approaches to the estimation of biological age vary widely, but typically involve the measurement of multiple biomarkers related to aging, chronic disease, or mortality, and can include a range of both functional and molecular indicators (28,29).

index) found that the scales identify different but overlapping groups of individuals (3). This discrepancy highlights that frailty may not be a unified concept; different approaches are tailored to different purposes. For example, the frailty phenotype, focusing on signs and symptoms, may be limited to older adults with no disability but, among them, is better at predicting disability (22). The frailty index, on the other hand, provides predictive information among a broader range of individuals regardless of age or functional status. Despite their differences, frailty measures display similar empirical characteristics (3). Cesari et al. advocate that the frailty phenotype and the frailty index should be considered complementary, not competing alternatives or substitutes, as they are so often perceived (22). The value of this view has been demonstrated in empirical studies of risk estimation: Kulminski et al. showed that though the frailty index was more predictive of death than the frailty phenotype, classifying individuals by simultaneously using the frailty phenotype and the frailty index may be a promising option to improve mortality prediction (35). Similarly, Watanabe et al. found that individuals identified to be frail by two instruments (the Frailty Screening Index and the Kihon Checklist, which are screening tools modeled after the frailty phenotype and frailty index, respectively) had a higher mortality risk than those only identified as frail by one (36). These findings illustrate that while having multiple operationalizations of the same concept may be initially challenging, there are advantages of capturing different facets of the same concept. They can be used in different applications or in the same application to improve risk estimation. This is an important insight to keep in mind when assessing different approaches to physical resilience: not only may physical resilience be complementary to frailty, but multiple measures of physical resilience may also be complementary to each other.

2.2 Review and Synthesis of the Conceptual Literature on Physical Resilience

To understand the concept of physical resilience and its relation to frailty, it is helpful to trace how the concept has evolved over time. This section, therefore, reviews and synthesizes the conceptual literature on physical resilience in chronological order. This section focuses on conceptual development, but because the concept has evolved hand in hand with empirical, methodological development, it also briefly introduces landmark methodological contributions. The empirical literature will be reviewed closely in section 2.3.

2.2.1 A Brief History of the Physical Resilience Literature

Over the past 20 years, the concept of physical resilience has attracted increasing attention in the aging literature, as evidenced by the growing number of articles with physical resilience in the title and abstract. Figure 1 shows this trend (see Appendix A for PubMed search terms). The history of the concept of physical resilience can be characterized by 13 landmark publications (indicated in numbers in parentheses in Figure 1) and broken down into three periods. Years from 2011 to 2017 represent the foundational period when the conceptual groundwork for physical resilience was laid out. Years from 2017 to 2020 represent a period of proliferation of methodological innovation as well as an increasing interest in the relationship between frailty and physical resilience. Year 2017 is both the end of the conceptual foundational period (period 1) and the beginning of the methodological innovation period (period 2) because a landmark study by Gijzel et al. (37) published in 2017 made both conceptual and methodological contributions, bridging the two periods. Lastly, years from 2021 to 2023 represent the period of the first major prospective clinical studies specifically designed to investigate physical resilience with updated conceptual frameworks. This period also observed a shift in discourse where the concept of intrinsic capacity (described in section 2.2.4) was brought into the discussion surrounding frailty and physical resilience.

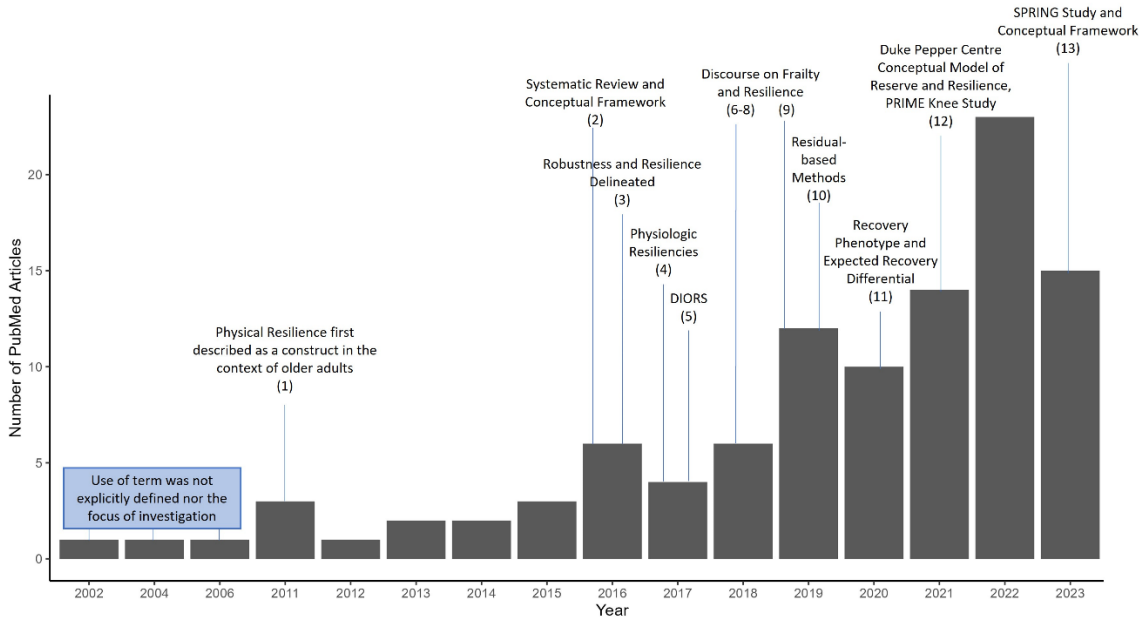


Figure 1. Timeline of Physical Resilience Literature

The number of PubMed articles with physical resilience in the title/abstract as of January 3rd, 2024. See Appendix A for PubMed search terms. Numbers in parentheses indicate 13 landmark publications: (1) Resnick et al. 2011; (2) Whitson et al. 2016; (3) Ukraintseva et al. 2016; (4) Hadley et al. 2017; (5) Gijzel et al. 2017; (6) Whitson et al. 2018; (7) Kuchel 2018; (8) Varadhan et al. 2018; (9) Rikkert and Melis 2019; (10) Wu et al. 2019; (11) Colon-Emeric et al. 2020; (12) Whitson et al. 2021; and (13) Walston et al. 2023.

2.2.2 The Foundational Period (2011 – 2017)

In 2011, Resnick et al. (landmark publication #1) first described the term “physical resilience” in the context of the physical health of older adults as "the ability to recover or optimize function in the face of age-related losses or disease" (p.644) (38). Subsequently, in 2016, Whitson et al. (landmark publication #2) published the first major influential effort to clarify the concept of physical resilience informed by a systematic review (9). This seminal work paved the way for subsequent studies by proposing a working definition and a conceptual model of physical resilience, identifying three potential measurement approaches to physical resilience, and specifying three key future research areas.

Whitson et al. defined physical resilience as “a characteristic at the whole person level which determines an individual’s ability to resist functional decline or recover physical health following a stressor” (p.493) (9). Of note, this definition conceptualizes physical resilience as a characteristic

(i.e., “an ability or capacity that can change over time” (p. 491)) rather than a trait (i.e., “a relatively fixed characteristic, part of one’s nature” (p. 491)) or a trajectory (i.e., “change in symptoms or function over time” (p. 491)). The conceptual model of physical resilience proposed by Whitson et al. postulated that physical resilience influences the outcome after a stressor and that physical resilience is influenced by both external and internal factors (Whitson et al. (9), Figure 3). The external factors include environment and life experiences, whereas the internal factors include genetics, psychosocial factors, and what Whitson et al. called “physiologic reserve” defined as “the potential capacity of a cell, tissue, or organ system to function beyond its basal level in response to alterations in physiologic demands” (p.492) (9). The level of physiologic reserve across organ systems constrains physical resilience at the whole person level, and physiologic reserve can be measured by performing stress tests on the system(s) involved. This concept of physiologic reserve became central in subsequent work aiming to further clarify the concept of physical resilience, as discussed below. In addition, Whitson et al. identified three potential measurement approaches as indicators of physical resilience: phenotypes (e.g., frailty, fatiguability), age discrepancy (i.e., biological vs. chronological age), and functional trajectories after a stressor. For the latter, they distinguished “resistant trajectories” (i.e., no change in function after a stressor) from “resilient trajectories” (i.e., initial decline then recovery in function after a stressor) (p. 493). Furthermore, Whitson et al. specified three key future research areas: the measurement of physical resilience; the examination of physiologic reserve between and across systems; and effective interventions to optimize resilience.

Building upon Whitson et al.’s distinction between resistant and resilient trajectories, Ukraintseva et al. (landmark publication #3) further clarified the concept of resilience and differentiated it from the concept of robustness (39). Ukraintseva et al. considered a two-stage response to a stressor: deviation from baseline (resistance) and return to baseline (recovery). They referred to the former, the ability to resist deviation from the baseline state, as “robustness,” and the latter, the ability to quickly and completely recover after a deviation from the baseline state, as “resilience.” By understanding these concepts of robustness and resilience as abilities, Ukraintseva et al. suggested that the ability to resist and the ability to recover may be different and perhaps have different underlying mechanisms. In response to Ukraintseva et al., Whitson et al. (40) clarified that they considered robustness to be the opposite of frailty. This is a key point that was further elaborated later in the second period, 2017-2020.

A report from the National Institutes of Aging (US) workshop, "Measures of physiologic resiliencies in human aging," held in 2015 and published in 2017 by Hadley et al. (landmark publication #4) (41), corresponds to one of the three key future research areas proposed by Whitson et al., the examination of physiologic reserve between and across systems. The "physiologic resiliencies" in Hadley et al. align with the concept of resistance to and recovery from stressors discussed by Ukraintseva et al. (40) but specifically applied to lower levels, such as cells, tissues, organs, or organ systems, rather than the whole person. These physiologic resiliencies in aggregate can influence whole person physical resilience. Hadley et al. conceptualized resilience, whatever the level, as a homeostatic stress response to a stressor that occurs in four sequential phases as illustrated in Figure 2: 1) a pre-stressor baseline, 2) time to and peak magnitude of perturbation, 3) time course of recovery, and 4) stabilization after recovery (i.e., completeness of recovery). Phase 2 in Hadley et al. corresponds to robustness discussed by Ukraintseva et al., and phases 3 and 4 in Hadley et al. correspond to what Ukraintseva et al. called resilience.

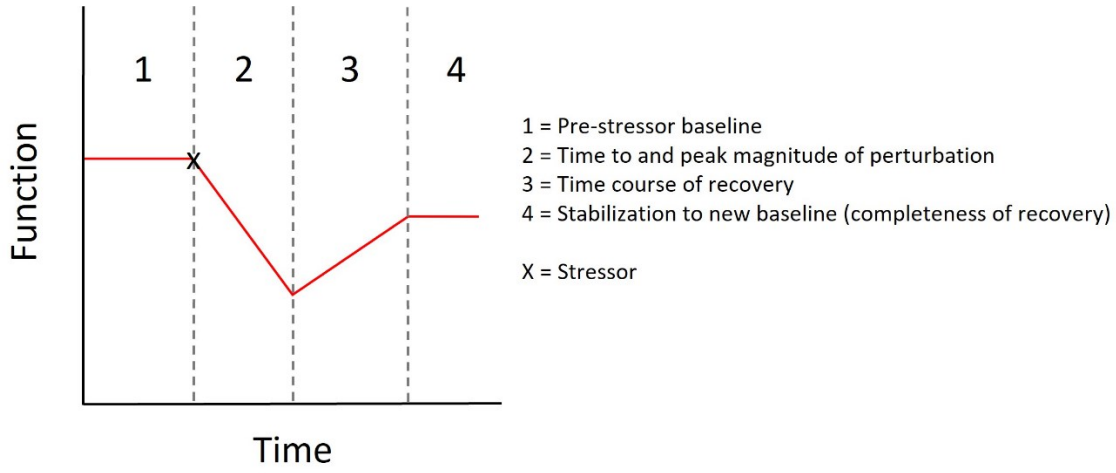


Figure 2. Four Sequential Phases of a Homeostatic Stress Response to a Stressor

Four sequential phases of a homeostatic stress response to a stressor as described by Hadley et al. 2017.

Importantly, Hadley et al. (41) suggested that physiologic reserve determines threshold stress levels and, in turn, influences the stress response trajectory. If a stress level exceeded the threshold determined by reserve, a perturbation and/or state change would occur, at which point the system could not easily return to its basal (prestress) state. Thus, reserve influences phase 2

in Hadley et al., time to and peak magnitude of perturbation (i.e., robustness in Ukraintseva et al. (39)), rather than recovery phases of 3 and 4.

The final major contribution in the foundational period is the application of complex dynamical systems theory to resilience in humans, operationalized as dynamical indicators of resilience (DIORs) in empirical investigation. This contribution was first made in an empirical paper by Gijzel et al. (37) in 2017 (landmark publication #5) and later expanded in a conceptual paper by Scheffer et al. (42) in 2018. The idea behind DIORs is that changes in the dynamics of a system may indicate proximity to a "tipping point," defined as an abrupt state change of the system that occurs past a certain threshold (42). Identification of proximity to a tipping point may indicate an opportunity for intervention before a critical transition to a worse state occurs. In this understanding, "critical slowing down" could serve as an early warning of a tipping point: slow recovery from small perturbations (e.g., a decrease in blood pressure) may indicate loss of resilience, and a critical tipping point (e.g., syncope) may be near (42). Moreover, small perturbations at multiple subsystem levels (e.g., organs, organ systems) themselves, before encountering any stressor, may indicate resilience at the whole person level. Thus, whole-person resilience can be seen as an emergent property resulting from the underling network of specific physiologic resiliencies.

Importantly, Scheffer et al. argued that what they called functional reserves critically determine the ability of the system to respond to stressors. Functional reserves discussed by Scheffer et al. are equivalent to physiologic reserves discussed by Whitson et al. 2016 and Hadley et al. 2017. Taken together, the views by Whitson et al. 2016 (9), Ukraintseva et al. 2016 (39), Hadley et al. 2017 (41), and Scheffer et al. 2018 (42) converge: reserves determine stress thresholds (Whitson et al. 2016(9), Hadley et al. 2017 (41)) and tipping points (Scheffer et al. 2018 (42)), and the level of stress thresholds (which determine the ability to resist stress) can be characterized as robustness (Ukraintseva et al. 2016 (39)).

Taken together, these foundational ideas set the scene for further conceptual and empirical development of physical resilience.

Key Takeaways from the Foundational Period

1. Physical resilience is understood through stress response.

2. Stress response can be decomposed into four sequential phases: 1) pre-stressor baseline, 2) resistance/decline, 3) recovery, and 4) stabilization to new baseline.
3. Resistance is characterized as robustness. This distinction between resilience (recovery) and robustness (resistance) implies different underlying mechanisms.
4. Physiologic reserve influences resistance (robustness) by determining stress thresholds, at which tipping points occur.
5. Whole person physical resilience is comprised of a network of specific physiologic resiliencies.
6. Dynamical systems theory suggests that the dynamics of a system, before experiencing a tipping point, can indicate the resilience of a system.

2.2.3 Methodological Innovation and Discourse on Frailty and Resilience (2017-2020)

The next period is marked by prominent discourse on the relationship between frailty and resilience, along with the proliferation of methodological innovation. The methodological work will be fully explored in section 2.3 but is included briefly here to provide a complete picture of the historical overview.

Methodological Milestones

This period observed the following three methodological milestone studies. In 2017, Gijzel et al. (37) (landmark publication #5) published the first application of dynamical indicators of resilience (DIOR). In 2019, Wu et al. (43) (landmark publication #10) first introduced residual-based methods to estimate physical resilience by conceptualizing physical resilience as adaptation to cumulative stress. In 2020, Colon-Emeric et al. (44) (landmark publication #11) proposed two innovative methods to quantify recovery: the recovery phenotype and the expected recovery differential. Section 2.3 will describe in detail these notable milestones.

Conceptual Discourse on Frailty and Resilience

While earlier discourse focused on the distinction between frailty and physical resilience and examined whether one concept should be abandoned for the other (e.g., (45,46)), more recent conceptual discourse has affirmed differences between these two concepts and advocated for their combined use. In 2018 and 2019, a number of leading experts published commentaries and

editorials on the relationship between frailty and physical resilience. Their titles eloquently captured the gist of the recent conceptual discourse: “Physical resilience: Not simply the opposite of frailty” by Whitson et al. 2018 (11) (landmark publication #6), “Frailty and resilience as outcome measures in clinical trials and geriatric care: Are we getting any closer?” by Kuchel 2018 (13) (landmark publication #7), “Can a link be found between physical resilience and frailty in older adults by studying dynamical systems?” by Varadhan et al. 2018 (12) (landmark publication #8), and “Rerouting geriatric medicine by complementing static frailty measures with dynamic resilience indicators of recovery potential” by Rikkert and Melis 2019 (14) (landmark publication #9). A common view shared by these articles is that physical resilience has significant potential to complement measures of frailty by adding a dynamic element that frailty lacks.

Rikkert and Melis contributed to the conceptual discourse in this second period by revisiting the idea of tipping points. They defined tipping points as “the points in time that separate a more healthy condition from an acute, but in principle reversible disease condition and malfunction of the human’s subsystems or organ dysfunction” (14) (p.2) and considered frailty as being highly prone to these tipping points. In other words, frail individuals are more likely to experience such a change in health after encountering a stressor. In addition, Rikkert and Melis reinforced the idea that dynamical indicators at the subsystem level could indicate proximity to and recovery from these tipping points. Given that frailty is often conceptualized as a state of low physiologic reserve across multiple organ systems (see section 2.1.1), Rikkert and Melis’ view aligns with Hadley et al.’s view that reserve influences the stress response trajectory by determining threshold stress levels (i.e., tipping points). Whitson et al. 2018 further reinforced this view by connecting the concepts of frailty, robustness, and resilience: “If the spectrum from robustness to frailty reflects the amount of physiologic potential one has to react to stressors, physical resilience refers to the actualization of that potential” (11) (p.1460). In addition to connecting the three concepts of frailty, robustness, and resilience, Whitson et al. made an important contribution by distinguishing between “potential” and “actualization”. Potential is what can be estimated prior to a stressor. Actualization is what can be observed post-stressor. In other words, one’s potential to recover can be estimated by pre-stressor reserve, while one’s actualized recovery can only be observed post-stressor.

This potential-actualization distinction by Whitson et al. can be related back to Hadley’s four phases of stress response (section 2.2.2). The level of reserve across multiple organ systems

determines where an individual is on the spectrum from robust to frail (i.e., frailty represents low reserve, and robustness represents high reserve). The level of reserve can be estimated prior to experiencing a stressor and, thus, can represent an individual's pre-stressor potential to respond. This corresponds to phase 1 of stress response (Figure 3). Everything after phase 1 represents the post-stressor realization of this potential. The level of pre-stress reserve determines stress thresholds and, thus, determines whether decline (or resistance) will occur after encountering a stressor (phase 2). Decline (or resistance) is a necessary precursor to recovery (phase 3) and subsequent stabilization (phase 4). Thus, the level of reserve (pre-stressor) constrains the level of physical resilience (post-stressor) by determining stress thresholds.

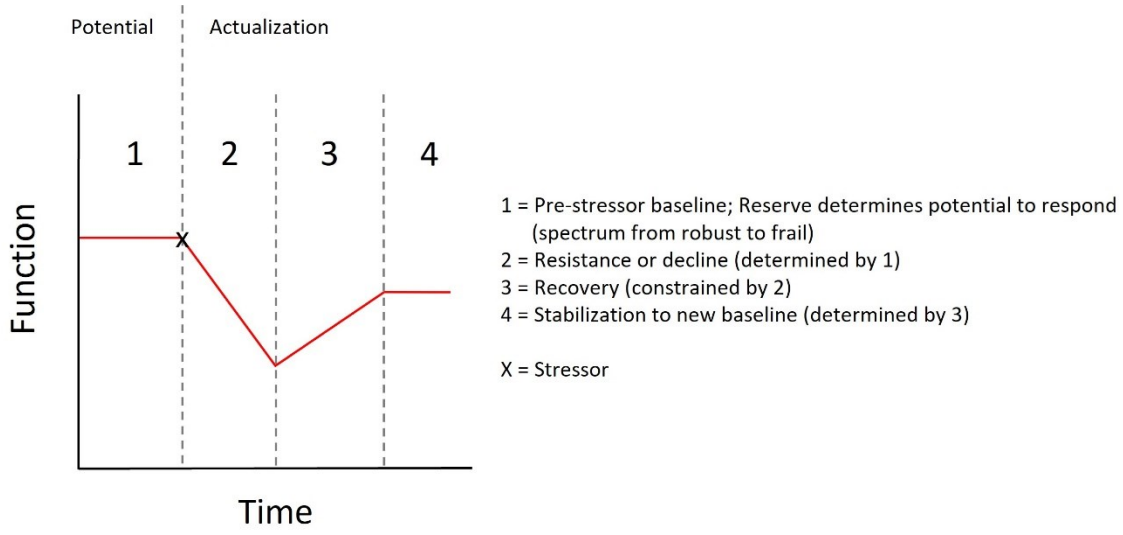


Figure 3. Phases of Stress Response Reflecting Ideas of Potential and Actualization

This figure explains the four phases of stress response described in Hadley et al. by incorporating the distinction between "potential" and "actualization" by Whitson et al. Reserve determines where a person is on the scale from robust to frail and further determines an individual's pre-stressor potential to respond to a stressor. This potential can only be realized after encountering a stressor, thus, phases 2-4 encompass the realization of that potential.

Since proposed by Whitson et al. 2018, the idea of resilience as the actualization of physiologic potential to react to stressors has gained traction in other articles (for examples, see (47–49)). The post-stressor actualization of physical resilience is what studies of physical resilience aim to predict. Frailty is a key component of understanding the pre-stressor potential to respond to stressors. However, as currently implemented, frailty lacks a dynamic element that may prove useful in estimating this potential.

Key Takeaways from the Period of Methodological Innovation and Discourse on Frailty and Resilience

1. Methodological innovations proliferated, including dynamical indicators of resilience (DIORs), residual-based methods, and advanced measures of recovery.
2. Interests grew as to how frailty and resilience related to each other, and the conceptual discourse in this period built upon the understanding of reserve and robustness from the foundational period.
3. Robustness and frailty can be considered opposite ends of the reserve spectrum.
4. The level of reserve influences the stress response trajectory by determining stress-response thresholds. A robust person has high reserve (and thus high stress thresholds) and is much less likely to experience decline after encountering a stressor. Conversely, a frail person has low reserve (and thus low stress thresholds) and is much more likely to experience decline after encountering a stressor.
5. The robustness-frailty reserve spectrum reflects the pre-stressor potential to respond to stressors, whereas physical resilience is the post-stressor realization of the potential.
6. Reserve (pre-stressor potential) constrains the level of resilience (post-stressor realization).

2.2.4 The **First Major Clinical Studies and Discourse Shift (2021-2023)**

The most recent and final period is marked by the initiation of two major clinical studies of physical resilience and their guiding conceptual frameworks, showing the cutting-edge development of physical resilience on the clinical front. The introduction of the concept of intrinsic capacity also shifted the conceptual discourse during this period.

In 2021, Whitson et al. (landmark publication #12) described their ongoing clinical study, the Physical Resilience Indicators and Mechanisms in the Elderly Knee Replacement (PRIME-KNEE) study, as a template for physical resilience research among older adults (48). The aim of the ongoing PRIME-KNEE study is to validate clinical tests and biomarkers that predict resilience to total knee arthroplasty. The guiding framework for the PRIME-KNEE study is called the Duke University Pepper Center Conceptual Model of Resilience (Figure 1, Whitson et al. (48)). This framework considers resilience as a dynamic response to a stressor. The level of pre-stress reserve is comprised of multiple domains (cognitive, psychological, and physical) and influences the

dynamic stress response (resilience). The range of potential responses lead to a range of potential outcomes regarding survival, independence, quality of life, and morbidity. The entire stress response trajectory plays out in the broad context of environmental factors. Each point in the framework (pre-stress reserve, stressors, resilience (dynamic response), and environment) is a potential opportunity to intervene, and such an opportunity can be before, during, and after the stressor. Of note, this framework considers resilience as a dynamic response rather than a characteristic as in the original conceptualization of resilience by the same research group (Whitson et al. 2016, landmark publication #2). This shift reinforces their "actualization of potential" idea discussed in Whitson et al. 2018 (landmark publication #6) and clearly differentiates what can be measured pre- vs. post-stressor; Resilience can only be realized after a stressor. This framework has been discussed within and the beyond the context of the PRIME-KNEE study in recent published articles (e.g., see (50,51)).

In 2023, Walston et al. (52) (landmark publication #13) published a conceptual framework of physical resilience for their clinical study, the Study of Physical Resilience and Aging (SPRING). This study has three sub-studies representing three different major clinical stressors: bone marrow transplant, dialysis initiation, and total knee replacement. Walston et al. developed their own terminologies: physiologic resilience capacity, representing pre-stressor estimation of the ability to recovery, and resilience phenotypes, representing the observed post-stressor trajectories. Walston et al. propose that physiologic resilience capacity is not directly quantifiable but can be described through a combination of static surrogate measures and dynamic stimulation tests. This physiologic resilience capacity in Walston et al. is conceptually equivalent to what Whitson et al. (48) call reserve (i.e., pre-stressor potential to react to stressors), however, it suggests an integrative measurement approach to quantify it. Similarly, the resilience phenotypes in Walston et al. are equivalent to what Whitson et al. calls resilience (i.e., post-stressor actualization of the potential). Walston et al.'s framework focuses specifically on clinical stressors and presents a link from stressor to physiologic resilience capacity, to resilience phenotypes, to clinical outcomes. They hypothesize that pre-stressor resilience capacity is influenced by such factors as age, disease, health behaviours, and psychosocial elements and can be estimated through a combination of static surrogate measures (of which they include frailty), and stimulus response measures (see section 2.3.2). Walston et al.'s study plans to examine post-stressor resilience phenotypes for one year after the stressor. In summary, these two ongoing clinical studies will provide valuable

information to characterize resilience following clinical stressors and determine what factors influence pre-stressor reserve, which in turn influences resilience.

In addition to these two major clinical studies of physical resilience, the most recent period has observed a shift in the discourse on frailty and resilience by introducing the concept of intrinsic capacity (10,49,53,54). These recent articles describe frailty, physical resilience, and intrinsic capacity as different tools for different purposes, but the introduction of intrinsic capacity can further enrich the discourse on frailty and resilience. In the *World Report on Ageing and Health* in 2015, the World Health Organization defined intrinsic capacity as "the composite of all the physical and mental capacities of an individual" (55) (p.28), and in 2018, Cesari et al. suggested intrinsic capacity concerns the following five domains: locomotion, vitality, cognition, psychological, and sensory (56). Intrinsic capacity is another concept for which the construct of reserve plays a critical role, and some authors in recent articles suggest that intrinsic capacity represents an evolution of the concept of frailty (57). Intrinsic capacity represents a global level of reserve indicated by the aforementioned domains, while frailty represents a state of low reserve across multiple physiologic systems. Despite being similarly based on reserve, intrinsic capacity focuses on the monitoring of reserves over time for prevention of premature aging and promotion of healthy aging, while frailty focuses on health deficits and is best suited to cross-sectional risk assessment (10,49). By bringing the concept of intrinsic capacity into the discourse on frailty and resilience, recent articles emphasize the added value of longitudinal assessment of reserves to the typical cross-sectional assessment of studies of frailty and the typical short-term assessment, immediately before and after a stressor, of studies of resilience.

Key Takeaways from the Period of the First Major Clinical Trials and a Discourse Shift

1. Landmark prospective clinical studies of physical resilience are underway building upon the conceptual discourse on physical resilience over 20 years.
2. The discourse on frailty and physical resilience is recently being expanded to include intrinsic capacity. The concept of reserve plays a critical role in all three concepts.
3. The introduction of intrinsic capacity to the discourse highlights the importance of longitudinal monitoring of reserves.

This final period concludes the historical overview of the concept of physical resilience. The next section will discuss how physical resilience has been empirically operationalized to date.

2.3 Review and Synthesis of the Empirical Literature on Physical Resilience

Specific empirical operationalizations of physical resilience vary widely but can be broadly described by their general measurement approach. The first categorization we can make is whether a measure of physical resilience is defined by: a system, a stressor, and an outcome (or state) (10). For example, we could define the system as the whole person, the stressor as a myocardial infarction, and the outcome as physical function, measured by the Short Form 36 Physical Function Subscale (SF-36 PFS). Measures of physical resilience that apply this useful "triad" of system, stressor, and outcome, can be further grouped into four general categories of measurement approaches: 1) measures of post-stressor change, 2) stimulus response measures, 3) residual-based measures, and 4) dynamical indicators of resilience.[2] This section first reviews each of these four major categories. This section then briefly reviews measures of physical resilience that do not apply the triad of stressor, system, and outcome, including self-reported measures, static surrogates, proxy measures, and aggregate indicators. The section concludes with a summary and comparison of the four major categories of measures.

2.3.1 Measures of Post-Stressor Change

Measures of post-stressor change directly quantify the response to a defined stressor using multiple observations over time. This approach is by far the most common, and though no gold standard measures currently exist, measuring a functional trajectory following a well-defined stressor has been suggested as a promising candidate (9). However, such a measure can take many forms, and significant variation exists in the empirical literature. Measures have been operationalized using acute, chronic, and unspecified stressors. An acute stressor is one that is short-lived and usually has a higher intensity/severity, while a chronic stressor occurs over a longer period and is typically lower in intensity. For example, a heart attack could be considered

[2] Note that some specific interpretations or operationalizations of physical resilience may not fit nicely into only one of these categories, and these categories may be further refined as literature expands. However, for now, such categorization can aid in understanding the different approaches that have been taken.

an acute stressor with more immediate effects, while dialysis could be considered a chronic stressor with more long-term effects. This distinction is important as it determines what kind of post-stressor response can be observed.

Acute Stressors

Operationalization of response to acute stressors ranges from simple metrics of recovery using two time points to complex statistical approaches to quantifying recovery trajectories across multiple outcomes. As an example of simple metrics, Calle et al. (58) operationalized two measures of recovery after orthopedic surgery (hip fracture, hip replacement, and knee replacement) and stroke (ischemic or hemorrhagic): absolute and relative functional gain.[3] Calle et al. defined absolute functional gain as the difference between the Barthel Index score (which measures functional independence) at discharge and admission, and relative functional gain as the percentage of lost function recovered. Results were stratified by the category of stressor (orthopedic or stroke), but the exact type and severity were not taken into account, which may contribute to variation in the amount of recovery. They examined the relationship between recovery and frailty-related factors (e.g., delirium, ability to walk, sarcopenia) but did not use an overall measure of frailty.

In contrast to this simple approach, Colón-Emeric et al. (59) used a latent growth mixture model to determine common recovery patterns across 10 functional outcomes over 12 months in patients with hip fracture. Colón-Emeric et al. categorized individuals as high, medium, and low resilience based on the common recovery patterns across all outcomes. Colón-Emeric et al. (44) expanded upon this initial work by describing two approaches to quantify physical resilience: the "recovery phenotype" and the "expected recovery differential".

The recovery phenotype observes clinical events or trajectories and uses statistical methods (e.g., latent class analysis (LCA) or principal component analysis (PCA)) to summarize recovery patterns across multiple variables (e.g., time series of different functional outcomes for LCA and multiple clinical classification variables for PCA). Colón-Emeric et al. presented the latent class analysis approach described above for hip fracture as one method to determine the recovery phenotype.

[3] Note that while Calle et al. were studying recovery, they did not use the term "physical resilience" in the publication.

They further described an approach to use PCA in the case of using multiple clinical classification variables that cannot be captured as a time series (e.g., survival). They provided the example of pneumonia as the stressor, where the outcomes used to determine the recovery phenotype included length of hospital stay, intensive care unit admission, death within 28 days, and discharge location. This is a unique application which provides an opportunity to incorporate multiple relevant outcomes that may not fit to a conventional times series analysis. Importantly, in both applications (hip fracture and pneumonia) the authors considered stressor characteristics related to severity such as anesthesia type, duration of surgery, partial vs total arthroplasty, and laboratory findings.

Colón-Emeric et al. defined an alternative approach, the expected recovery differential, as the expected vs. observed recovery based on a population-derived model.[4] In their empirical demonstration, the recovery phenotype typically identified the healthiest individuals as the most resilient, and the least healthy as least resilient. However, the expected recovery differential identified healthy individuals who had worse recovery than expected and unhealthy individuals who had better recovery than expected (based on the population-derived model). Given these results, Colón-Emeric et al. suggested the recovery phenotype is useful for characterizing complex recovery patterns, while the expected recovery differential is useful for exploring the biological mechanisms underlying physical resilience (44). These methods represent two unique approaches that can be tailored for different applications and used for multiple outcomes to ascertain an overall level of physical resilience.

Chronic Stressors

Operationalization of response to chronic stressors typically takes the form of a simple categorization of a trajectory after the initiation of a new chronic stressor (hereafter referred to as an "incident chronic stressor"). As an example, Hladek et al. used phenotypic trajectories of four variables over 12 months following incident hemodialysis: physical function, mental health, vitality (defined as energy or fatigue), and general health (60). They characterized three simple phenotypes for each variable: improving, stable, and declining. Physical function and vitality were

[4] Though the expected recovery differential is a residual-based method, the key reason why it is included here is because it is defined by a singular specific stressor and time course of recovery. Thus, these characteristics fit better with measures of post-stressor change rather than the hallmarks that define the residual-based methods described in section 2.3.3.

found to be independently associated with mortality after accounting for baseline covariates including age, sex, race, BMI, comorbidity, and several disease-specific biomarkers (60). A similar study by Presley et al. (61) defined resilience as maintenance or improvement in disability scores over eight months in a sample of patients newly diagnosed with stage 4 lung cancer (i.e., an incident chronic stressor) who were just beginning treatment. Importantly, neither study observed a baseline trajectory prior to initiation of the chronic stressor. Thus, it is impossible to know whether the patients were stable, improving, or declining prior to the stressor. It would likely be more informative to determine if the trajectory changes after initiation of the chronic stressor rather than just observing the post-stressor trajectory.

Unspecified Stressors

In addition to operationalizing responses to acute and chronic stressors, a few studies also operationalized the response to unspecified stressors. Pedone et al. (62) determined physical resilience by examining the change in physical function after a nonspecific stressor, defined as a self-reported "major health event". Pedone et al. categorized individuals as resilient (maintained function after event), non-resilient (declined after the event), decliners (declined in absence of event), and controls (no event, no decline). They found that resilient individuals were similar to controls in terms of change in Activities of Daily Limitations (ADLs) and mortality over time. This is an interesting way to broaden the applicability of a stress response measure, by including a non-specific stressor that allows estimation in a broader population than a specific stressor, and to categorize and compare individuals who did not experience a health event. However, this approach does not consider recovery. Additionally, the subjectivity of a self-reported major health event means that there could be significant variation in the type, intensity, and duration of the reported events that is unaccounted for.

Duan-porter et al. (63) conducted another study with an unspecified stressor. Their study population was a cohort of older overweight cancer survivors at least five years after their diagnoses. They defined decline as a drop of 13 or more points of the SF-36 physical function subscale, resistance as a lack of any decline, and resilience as regaining at least 50% of lost function over a two-year period. Decline, resistance, and resilience were measured without any reference to a defined stressor. Duan-porter et al. found that most older cancer survivors exhibit resilience, which is unsurprising given that the less resilient individuals would likely have a higher

risk of death. Of note, while this study did not explicitly define the stressor, by using this cohort, it is possible to consider cancer as the stressor. However, Duan-porter et al. did not explicitly conceptualize it as such, and relevant details such as remission status, duration of treatment, cancer type and stage were missing (the study only reported time from diagnosis and type of treatment). Furthermore, they noted in the discussion that those who experienced more decline may have experienced more severe stressors leading to the decline. Thus, for the current discussion, I consider this study as having an undefined stressor. This example highlights the need for explicit conceptual consideration of the three defining features of physical resilience (stressor, system, outcome) when operationalizing a measure.

Summary of Measures of Post-Stressor Change

Despite the variation in the health events used as stressors, all studies reviewed in this subsection operationalized resilience at the whole person level, and most of these studies used measures of physical function (e.g., SF-36 physical function subscale) as the outcome. All studies were conducted over a relatively short term, with three years after an incident stressor as the longest follow-up period. Over the observation period, studies typically employed three to four measurements, with a minimum of two and a maximum of eight. Lastly, all cited studies above used clinical samples except Pedone et al. (62). See Appendix B for a literature review table of all cited empirical studies using a measure of physical resilience between 2016 to 2023.

A key consideration for these studies is whether stressor severity was accounted for. Stressors of higher severity are more likely to result in less recovery. Thus, if unaccounted for, the resilience measures may be confounded by the stressor severity. Most studies described above tried to account for stressor severity by including additional measures of severity or by restricting the selection criteria. As illustrated by the samples used in these studies, accounting for stressor severity is more difficult to address in population data where such detailed severity information may not be available. Therefore, measures of post-stressor change may be more feasible in clinical settings.

Quantifying recovery has strong face validity for determining physical resilience and may be the best candidate for a gold standard (9). After all, the most recent conceptual frameworks for the ongoing clinical studies clearly differentiate between pre-stressor potential, and post-stressor

realization: resilience can only be realized after encountering a stressor. However, as noted by Whitson et al. (9), past recovery may be unreliable to predict future recovery. Observing a resilient trajectory does not necessarily mean the individual will demonstrate a resilient response again in the future. Recovery is influenced by both intrinsic (e.g., level of reserve) and extrinsic (e.g., nature of the stressor, care provided, rehabilitation, etc.) factors that may or may not be observed. In addition, it is possible that a stressor may alter an individual's ability to recover in the future by depleting reserves (9). Thus, recovery measures may be useful to validate new measures of physical resilience but may have limited potential for predicting future recovery. Further work is needed to determine what impact stressors have on reserve and if past recovery is an adequate predictor of future recovery. Furthermore, a major limitation to these approaches is that they can only be applied to a small subset of the population who have experienced a specific health stressor. The use of a self-reported "major health event" by Pedone et al. (62) attempts to overcome this limitation, but this approach comes with its own limitations, such as the self-reported nature of events, and still limits application to those who have experienced *some* event. Lastly, the empirical implementation of measures of post-stressor change, to date, has limited consideration of frailty.

2.3.2 Stimulus Response Measures

Stimulus response measures involve deliberately eliciting a response with a precise external stimulus and quantifying the response. Examples of this include monitoring heart rate during an exercise test or monitoring blood pressure during an orthostatic challenge (e.g., going from sitting to standing) (47). As a recent example, Koivunen et al. (64) studied several hemodynamic and postural sway indices after an orthostatic challenge and found that change in diastolic blood pressure and sway root mean square showed a significant hazard ratio for mortality in men. Additionally, the ongoing PRIME-KNEE (48) and SPRING (52) clinical studies, as discussed in section 2.2.4, are both evaluating specific stimulus response measures across different bodily systems, including the immune system, adrenal system, and cardiovascular system. The advantage of this approach is that the stressor can be clearly defined and quantified, overcoming one of the main difficulties in measures of post-stressor change. Additionally, stimulus response tests can be performed prior to experiencing a real-life stressor, making them widely applicable in the clinical setting and potentially useful for predicting future events. However, a major drawback to this approach is that subjecting frail individuals to a stressor may impose unnecessary risk and it may

be infeasible to design practical tests that are safe for frail older adults to complete (47). Stimulus response measures have great potential for practical use in clinical settings, however, they require special tests and/or equipment and potentially pose risks to frail individuals. Thus, these measures may not be feasible in all individuals (i.e., frail) or settings (i.e., population). The key difference between measures of post-stressor change and stimulus response measures is that stimulus response is something that can be specifically tested using a minor elicited stressor (e.g., going from sitting to standing) before any major real-word stressors (e.g., heart attack) are encountered. In contrast, measures of post-stressor change reflect actual changes in health after experiencing such real-world stressors that pose a risk to health.

2.3.3 Residual-based Measures

The next category of physical resilience measures considers an individual's ability to adapt to cumulative stress as a manifestation of physical resilience. In contrast to measures of post-stressor change that use incident stressors and within-person functional change over time, residual-based measures consider the impacts of cumulative stressors on health by using the residuals of population-based models. By utilizing between-individual information, residual-based measures have the advantage of not requiring repeated measures (i.e., can be estimated in a cross-section).

Wu et al. (43) were the first to conceptualize physical resilience as adaptation to cumulative stress. Wu et al. operationalized physical resilience as the mismatch between an individual's observed and expected level of frailty, based on the residual of a population-derived linear regression model. This method was previously described by Sanders et al. (65), but Wu et al. were the first to apply it to the concept of physical resilience. Hereafter, Wu et al.'s method is referred to as the "frailty-disease mismatch method". Their regression model used age, sex, chronic disease, and disease burden to predict frailty at a single timepoint. They considered those with higher frailty than expected to be "premature frailers", those with similar expected and observed values to be "expected agers", and those with lower frailty than expected to be "adapters" (i.e., are better able to adapt to the cumulative stress on their body; resilient). Wu et al. measured frailty with the Scale of Aging and Vigor Epidemiology, a modified version of the frailty phenotype designed to provide greater differentiation of the healthiest individuals (66). In a study population of initially well-functioning older adults, Wu et al. examined years of able life (defined as the number of

disability free years), years of healthy life (defined as the number of years reporting good or better self-rated health), years of healthy and able life (defined as the number of years reporting good health and having no disability), disability, hospitalization, mortality, and survival to 90 years as validation outcomes. All validation outcomes followed the expected gradient, with adapters having the best outcomes, and premature frailers having the worst. Subsequently, they followed this work with a simplified approach that determined the mismatch without requiring a regression model (67). Rather, they grouped participants by the number of diseases they had and classified each individual as an adapter, expected ager, or premature frailer, based on the category-specific mean and standard deviation of frailty. The results showed the simplified approach was less precise as it did not account for potential confounders such as demographic characteristics as in the regression approach. However, the results displayed acceptable agreement with the original approach, and the validation outcomes again followed the expected gradient. Wu et al. suggested the simplified method could make quantification more accessible and timely, potentially enabling clinical adoption, though they recommended the original method be used for research (67). The frailty-disease mismatch method is similar to a chronological/biological age mismatch, proposed as a possible measurement approach by Whitson et al. (9) (see section 2.2.2). A chronological-biological age mismatch could be understood as using chronological age as a proxy for cumulative stress and is consistent with the frailty-disease mismatch method described above. However, the frailty-disease mismatch method used age, chronic disease, and disease burden to capture exposure to cumulative stress while accounting for sex differences. Thus, if the frailty-disease mismatch method were used to estimate an individual's expected FI (a measure of biological age), it would be similar to, yet more comprehensive than, a simple chronological/biological age mismatch.

Zhang et al. (68) followed up on the frailty-disease mismatch method by applying to a different outcome measure. They used a similar linear regression (i.e., including age, sex, and disease as predictors) with the short physical performance battery (SPPB) instead of the SAVE frailty scale. Similar to Wu et al. (43), Zhang et al. considered those with a better (worse) observed SPPB compared to expected SPPB as resilient (non-resilient). Their results supported those found by Wu et al.: continuous, binary, and four-category physical resilience, as quantified by residuals from a linear model, all demonstrated lower hazard for all cause mortality as physical resilience increased. Zhang et al. provided both unadjusted and adjusted models, the latter included age,

sex, smoking status, marital status, race/ethnicity, BMI, and care status as covariates. Both models demonstrated similar results.

Milman et al. recently developed another noteworthy residual-based measure called the Frailty Resilience Score (FRS) (69). Rather than being a measure of physical resilience, they referred to the FRS as a measure of "resilience to frailty" (69). Their method is similar to the frailty-disease mismatch method of Wu et al. (43) in that the FRS is based on the expected (predicted) vs. observed level of frailty. The main difference is that instead of conditioning on disease burden in the regression model, the FRS conditions on a polygenic risk score[5] that indicates elevated risk of frailty. Additionally, Milman et al. used a longitudinal mixed effect model for repeated measures of FI, and the FRS was defined as the average of the residuals (based on fixed effects only) in the earliest three interviews (in contrast to the single point-estimate of Wu et al.). Milman et al. found that one standard deviation decrease in the FRS, which indicates lower frailty than expected, resulted in a 38% reduction in the hazard of mortality (69). Though qualitatively different, conditioning on a polygenic risk score effectively serves the same purpose as conditioning on a level of disease (or other cumulative stress): both increase an individual's expected level of frailty, allowing researchers to identify a mismatch which indicates a level of resilience.

Lastly, Sotos-Prieto et al. applied a similar approach to longitudinal data (70). Though not exactly a residual-based measure in the same sense as the frailty-disease mismatch method or the FRS, Sotos-Prieto et al.'s approach similarly operationalizes physical resilience as the difference between expected and observed accumulation of deficits based on a population model. Specifically, they operationalized physical resilience as accumulating fewer deficits than expected over a 3.2 year follow up, with the expected increase being 0.74 per year, cited from previous research in their cohort (70). This expected number was the cohort-average slope for age estimated in a linear mixed effects model adjusting for education, diet, smoking, alcohol consumption, physical activity, sedentary behaviour, and body mass index (71). Sotos-Prieto et al. limited the resilience analysis to those with an FI above the cohort median, with the rationale that many deficits these individuals had accumulated must be chronic stressors. Thus, these

[5] A polygenic risk score determines an individual's risk of a specific disease based on their genetic profile.

individuals were the unhealthiest 50% of the cohort and had higher exposure to chronic stress than the excluded half of the cohort.

Of these individuals, Sotos-Prieto et al. further selected those who reported exposure to acute stressors in the year before the end of follow up. They defined these acute stressors as hospitalization, unintentional weight loss, or hip fracture. Of these individuals who had experienced more chronic and acute stressors (by their definitions), Sotos-Prieto et al. considered individuals who accumulated fewer deficits than the entire cohort average of 0.74 per year to be resilient. In cases where the individuals maintained or even improved their FI, Sotos-Prieto et al. considered them to be "over-resilient". Though Sotos-Prieto et al.'s approach mirrors that of Wu et al. in concept (i.e., expected versus observed frailty), it does not use individual-level information to determine the expected value (i.e., it uses the population average as the "expected" value for everyone, ignoring any subgroup heterogeneity), and importantly, it limits the cohort to those above the median proportion of deficits accumulated, and having experienced a specific set of acute stressors. By adjusting for disease burden in their model, Wu et al. estimated physical resilience for everyone in their sample, without limiting the analysis to those with a specific level of stress: even individuals with no disease burden (i.e., cumulative stress) can be considered resilient or non-resilient based on their expected vs observed level of frailty.

Despite taking a different approach than measures of post-stressor change, residual-based measures fit with the idea of adaptation to cumulative stress on the body. However, it is unclear how residual-based measures relate to measures of post-stressor change. A clear advantage of residual-based measures is that they do not require the observation of incident acute stressors and can be estimated without multiple data collections. Importantly, capturing stressor severity is a challenge for residual-based measures as for measures of post-stressor change. Wu et al. attempted to capture severity by including self-rated health and the number of medications. This is a reasonable strategy with the data available and could be improved further with additional data, if available, such as length of time with each condition. This strategy could yield a reasonable approximation in absence of true gradings of severity. Despite this shared challenge of capturing stressor severity, residual-based measures are more widely applicable than measures of post-stressor change and may be better suited to population studies, such as nationally representative cross-sectional studies. Conversely, measures of post-stressor change are better suited to clinical studies with repeated measures on a small, clinical sample. The cited studies support this idea, as

the studies employing residual-based measures used population-based samples, while most studies employing measures of post-stressor change used clinical samples. These different measures need to be further examined and characterized in the same population to assess how they relate to one another and how their relative performance compares.

2.3.4 Dynamical Indicators of Resilience (DIORs)

The final category of physical resilience measures is Dynamical Indicators of Resilience (DIORs). As discussed in section 2.2.2, a key concept that supports DIORs is critical slowing down, which suggests that slow recovery from small perturbations may indicate loss of resilience and that a critical tipping point is near (42). Thus, the idea behind DIORs is that microrecoveries from stochastic deviations can indicate the resilience of the subsystem being measured and can be extrapolated to indicate resilience at the whole person level (42). DIORs are distinct from all other measures of physical resilience in that they are not conditional on a certain stressor (like measures of stimulus response and post-stressor change), nor on a level of cumulative stress (like residual based measures). Rather, they consider the variability from baseline to represent micro-recoveries from minor stochastic stressors.

The two most prominent DIORs are variance and cross-correlation of time series measurements, with increasing values indicating diminished resilience. An increase in variance of the specified outcome measure suggests loss of dynamic regulation ability (i.e., the system is less able to maintain stability in the face of stochastic microstressors). An increase in cross-correlation among multiple measures suggests loss of independence among interconnected subsystems, such that a less resilient subsystem is less able to deal with a stressor, and thus the stressor has a more prominent impact on the connected subsystems (e.g., measures of cardiac and renal function showing similar fluctuations). DIORs typically involve many repeated measurements of a specific parameter within a brief period. Gijzel et al. (37) measured self-rated physical, mental, and social health over the course of 100 days in a small sample of 22 institutionalized older adults. They evaluated variance, temporal autocorrelation, and cross-correlation as DIORs and found that cross-correlation and variance of all three domains were associated with baseline frailty (measured by the frailty index). This study used a small, non-generalizable sample as proof-of-concept and demonstrated promising preliminary results for these measures as DIORs.

In a subsequent study, Gijzel et al. (72) examined a time series of postural balance over 30 seconds in a sample of high functioning older adults. They found that lower variance and temporal autocorrelation of mediolateral displacement, but not anteroposterior displacement, was associated with higher physical activity among hikers compared non-hikers. They supported this result with reference to studies suggesting that aging-related postural instability starts in the mediolateral direction, making it a more sensitive measure. Additionally, variance was independently associated with a successful aging index[6] at one year post measurement. In the most recent paper, the same research group, Gijzel et al. (73), measured variability in several outcomes (heart rate, physical activity, life satisfaction, anxiety and discomfort) in a geriatric in-patient population with acute illness. Variability in life satisfaction and variability in anxiety independently predicted three-month recovery after accounting for frailty (increased area under the curve (AUC) from 0.70 to 0.79). The authors defined three-month recovery as a binary indicator where "good recovery" was defined by remaining to live independently, not readmitted to hospital, and did not develop a new ADL difficulty by thee months post admission.

Gijzel et al. (73) provides initial evidence that the microrecovery approach to resilience can modestly improve recovery prediction at three months when measured alongside frailty. Interestingly, only variability in psychological variables (life satisfaction and anxiety) was associated with recovery. In contrast, a study by Kolk et al. (74) found that higher variability in fear of falling was associated with both more decline and more functional recovery in a study of acutely hospitalized older adults. This finding violates the expectation of higher variability correlating with worse recovery. In addition, variation in step count, pain, and fatigue were not associated with recovery. Together these studies show that the choice of DIOR variable is important, and in some cases, variability shows the opposite association than what is expected.

The potential opposite association is further illustrated in a study by Rector et al. (75). They demonstrated that using physical activity variability as a DIOR did not behave as expected in a sample of a geriatric inpatient population: higher variance in physical activity was associated with better ADL function and frailty scores rather than worse as hypothesized. Physical activity may have been a poor choice as any activity in this inpatient setting is likely a good sign. This makes

[6] The successful aging index consisted of eight indicators covering four domains including active engagement with life, personal resources, physical function, and emotional function (72).

intuitive sense as a stable low activity individual would likely be in worse shape than a high-variance medium-high activity individual. Rector et al. underscored the importance of evaluating assumptions of variables when examining resilience: can we reasonably assume that higher variability in a specific variable reflects loss of homeostatic regulation? A recent study by Lucas et al. (76) supports the finding of the association between high physical activity variability and better outcomes: step count variability was positively correlated with measures of physical function in older adults receiving hemodialysis. Lucas et al. thus suggest that this could be used as a novel measure of physical resilience.

Though not explicitly examining DIORs, studies by Zhu et al. (77) and Rouch et al. (78) found that blood pressure variability was independently associated with frailty, potentially providing a rationale for investigating blood pressure variability as a DIOR. Taken together, DIORs offer a measure that could be universally applied regardless of disease state or whether an individual has already experienced a specific health event. However, more work is needed to characterize appropriate variables and their relationship with different outcomes.

There are interesting parallels between this dynamical indicator approach and recent work studying the intra-individual variability of whole-person-level functioning and frailty in the long-term. A recent study by Stolz et al. (79) highlights the potential for investigating whole-person-level variation by analyzing frailty index fluctuations over 12 years in the Survey of Health, Aging and Retirement in Europe (SHARE). They found a non-negligible level of variability, with individuals fluctuating by an average of 2 to 2.5 deficits (0.04 to 0.05 FI) over the follow-up period. Both FI and FI instability increased with age, and was higher among women, individuals with low socioeconomic status, and those who died. Similar results were noted by Lin and Kelley-Moore (80) who examined the intra-individual variability in functional limitations and cognitive impairment. These results suggest that relying on average health trajectories may mask valuable information. This point has a marked resemblance to dynamical indicators of resilience, albeit, on a larger scale (i.e., reflecting actual recoveries rather than microrecoveries). This non-negligible, previously untapped heterogeneity provides an opportunity to explore the idea of longer-term macro-scale DIORs.

Rather than looking at single physiologic measures over the course of days or weeks (as done with typical short-term DIORs), we may be able to gain further insights by taking a long-term, macro-

level approach to measuring physical resilience by analyzing the intra-individual variability of health deficit accumulation and recovery over the course of years. A typical DIOR observes small, reversible deviations in the baseline level of a system-specific marker, which are then extrapolated to the subsystem or whole person to indicate a certain level of physical resilience. An individual's frailty index trend line could be considered as a composite, whole-person baseline, and rather than microrecoveries as deviations, we could observe full recoveries and declines as deviations from that baseline. Compared to the typical DIORs focusing on microrecoveries at a subsystem for a short time, such an approach would have the advantage of more accurately reflecting resilience at the whole-person level as there is no need for extrapolation. At the same time, this approach loses the advantage of the typical DIORs, which can be estimated before actual declines in health or function have taken place. Furthermore, as mentioned in the conceptual review (section 2.2), DIORs are based on the concept of critical slowing down: greater variability may indicate that a critical tipping point is near. If we extrapolate this idea to the whole person, deviations at the whole person level could indicate that a whole person tipping point, a transition to disability, dependence, or death, is near. All DIOR papers cited in this sub-section studied a clinical population, except Gijzel et al. (72), which examined postural balance over time among hikers. It would be worthwhile to explore the concept of critical slowing down in large-scale, longitudinal population-based studies.

DIORs are the final measurement approach that can be described using the system-stressor-outcome triad. The next two sub-sections briefly review measures that do not use the system-stressor-outcome triad. Specifically, section 2.3.5 reviews self-reported measures of physical resilience, and section 2.3.6 reviews measures that have been suggested or used in place of well-specified measures of physical resilience.

2.3.5 Self-reported Measures

Self-reported measures have also been used to study physical resilience. In contrast to the four categories of measures reviewed above, which objectively observe some sort of response to a defined stressor, self-reported measures directly ask patients who suffered from a stressor about their subjective experience related to recovery, with the aim of capturing an individual's ability to

physically recover from stressors. Resnick et al. (38) developed the Physical Resilience Scale[7] questionnaire and evaluated its reliability and validity against general resilience questionnaires used in psychology literature. The results showed some support for the reliability and validity of the measure, but importantly they did not evaluate its performance against other measures of physical resilience or recovery. Park et al. (81) subsequently used the Physical Resilience Scale to determine the relationship between frailty, osteoarthritic symptoms, physical resilience, and disability. Using the Tilburg Frailty Indicator, a self-reported questionnaire assessing multidimensional frailty (82), they concluded that frailty is a mediator of the relationship between symptoms and disability and that physical resilience is an effect modifier of the relationship between symptoms and frailty as well as between symptoms and disability independent of frailty. This is an interesting finding that highlights the potential of concurrently measuring frailty and physical resilience. However, this study compared no additional measures of physical resilience, and it is unknown how this questionnaire relates to other measures of physical resilience. Additionally, the Physical Resilience Scale does not appear to have been validated in the language or study population in which Park et al. employed it (Korean). The question Park et al. tried to answer (how do frailty and physical resilience relate to important aging-related outcomes?) is an excellent one with significant implications. Replication of these results with additional measures of physical resilience will help elucidate the roles of frailty and physical resilience.

A newer questionnaire developed by Hu et al. (83), the Physical Resilience Instrument for Older Adults (PRIFOR)[8] was recently evaluated in a population of older adults admitted to medical wards with a Clinical Frailty Scale (CFS)[9] rating between 4 (vulnerable) and 6 (moderately frail) (83). Hu et al. examined the predictive validity of PRIFOR using the EQ-5D (a health-related quality of life measure), the Clinical Frailty Scale, and the Katz Activities of Daily Living (ADL) scale. They found that PRIFOR was only associated with the Clinical Frailty Scale at one month after discharge and suggested PRIFOR could be used to predict recovery from frailty. Items from both questionnaires

[7] The Physical Resilience Scale is a 15-item questionnaire that asks questions related to recovery following acute events or illnesses. Examples of items include “I was determined to recover”, and “I accepted the new challenges”, with each item allowing a binary yes/no response for a total score out of 15 (38).

[8] PRIFOR is a 16-item questionnaire with responses on a 5-point Likert scale. Items cover three categories: positive thinking, coping and adjustment, and belief/hopeful mindset (83).

[9] Originally published in 2005 as a 7-point scale, the CFS is assessed by clinical judgment and focuses heavily on function while considering comorbidities and their management (84). In its current iteration, the CFS 2.0, the scale has 9 designations to allow for further discrimination between states, ranging from 1 being “very fit” and 9 being “terminally ill” (85).

(PRIFOR and the Physical Resilience Scale) reflect an individual's attitudes and perceptions related to recovery. This is a very different approach to measuring resilience than those defined by a stressor, system, and outcome. Though more development and validation are required, well-developed questionnaire approaches have significant potential for clinical use as they could be completed at a single timepoint and provide an indication of physical resilience prior to experiencing a stressor. This thesis focuses on the physical resilience measurement approaches that can be defined by a system, a stressor, and an outcome, but future work exploring the concurrent investigations of self-reported measures and the approaches using the system-stressor-outcome triad may offer valuable insight.

2.3.6 Static Surrogate Measures, Proxy Measures, and Aggregate Indicators

This subsection introduces measures that are not considered measures of physical resilience in themselves but may be useful when well-defined measures are not available or feasible.

Static Surrogate Measures

In the absence of well-defined measures of physical resilience, static surrogates (e.g., point estimates of frailty and physical function), which are easily obtained at a single timepoint and assumed to be correlated with resilience, may coarsely represent the level of physical resilience. Walston et al. provides examples of static surrogates in clinical studies: general static surrogates (e.g., phenotypic frailty, SF-36) and specific measures relevant to the specified stressor (e.g., Knee Injury and Osteoarthritis Outcome Score for knee replacement) (52). Walston et al. used these static surrogates in combination with stimulus response measures to estimate ability to recover before encountering a stressor (52).

Proxy Measures

Coarse proxy measures have been used to approximate physical resilience in longitudinal data without having to implement more complex and specific measures. These measures focus on development of disease and subsequent survival. Arbeev et al. used the age at onset of "unhealthy life" (defined as the first occurrence of a major complex disease) as a proxy for robustness (i.e., resistance to decline), and survival following onset as a proxy for resilience (86). Similarly, Galvin et al. (87) used the avoidance of disease at age 65+ as a proxy for robustness, and

survival to extreme ages as a proxy for resilience. Another study by Galvin et al. (88) defined resilience as survival following the onset of cardiovascular disease. Proxy measures such as these have the advantage of being simple and easy to implement with longitudinal data. However, they are very coarse approximations and are determined ex-post, leaving them with no ability to inform intervention efforts or clinical decision making. For example, if survival was used as a proxy for robustness, the dead were deemed non-resilient, but an intervention opportunity for them was lost after their deaths. The usefulness of these proxies appears to be for the assessment of the validity of a new measure of physical resilience. For example, Arbeev et al. used proxy measures to validate a new measure of "physiological dysregulation", an aggregate indicator of robustness and resilience (86).

Aggregate Indicators

Ukraintseva et al. suggested that composite indices combining several biomarkers can be used to indicate resilience (89). They provided two specific examples, the frailty index and an index of physiological dysregulation. The index of physiological dysregulation describes the average deviation from a normal physiological baseline across multiple physiologic measures, reflecting deterioration of homeostatic mechanisms across different physiological systems.[10]

2.3.7 Summary of Physical Resilience Measurement Approaches

There is a significant diversity in empirical methods representing different conceptual perspectives of physical resilience. Measures of post-stressor change, stimulus-response measures, residual-based measures, and DIORs all use the system-stressor-outcome triad and are tightly connected to concepts of physical resilience. Hence, they represent four promising approaches to capturing an individual's ability to cope with physical stressors.

Measures of post-stressor change represent post-stressor actualization of resilience; they allow direct quantification of response after an incident chronic or acute stressor. The drawbacks to measures of post-stressor change are their narrow scope, and the difficulty in accounting for stressor severity. An incident stressor needs to be observed; thus, these measures can only be

[10] Physiological dysregulation shares similarities to the population-level expected vs observed nature of residual-based measures, but the key difference is that it does not condition on any sort of stress.

estimated in a small portion of a population who have experienced a specific stressor, and thus, measures of post-stressor change are better suited to clinical rather than population-based studies. Second, if the goal is to predict a resilient response, it is unclear whether observing the response to an incident stressor may indicate future response. The reason for this is that stressors could potentially reduce reserve and subsequently reduce future ability to respond to stressors. For example, if a stroke resulted in some permanent loss of function, the individual may not be able to recover from future stressors.

Stimulus response measures have the major advantage of having no variation in stressor severity and may indicate an individual's ability to respond prior to experiencing a real-world stressor. However, stimulus response measures require specialized tests and equipment, which may not be feasible for all individuals or outside of clinical settings. Residual-based measures and DIORs have the advantage of being non-specific and widely applicable, regardless of the setting or health of the individual, without the need to observe an incident stressor. Residual-based measures can be estimated using a single timepoint and are the most universally applicable. However, residual-based estimates are based on a population average, so estimates may change depending on the reference population. DIORs require further investigation to find appropriate response variables.

2.4 Complementing ongoing Clinical Studies with Longitudinal Population Data

In the clinical setting, measures of physical resilience have the potential to be used as tools to guide individual patient care or as clinical outcomes to evaluate the effectiveness of new treatments/interventions. The current ongoing clinical studies, such as PRIME-KNEE by Whitson et al. (49) and SPRING by Walston et al. (53) introduced in section 2.2.4 and discussed in section 2.3.2, are poised to advance the field by enhancing the understanding of these different types of measurement approaches in the context of clinical stressors. The expected advance likely includes better characterizing the post-stressor actualization of the ability to respond to major clinical stressors (i.e., measures of post-stressor change), better characterizing the pre-stressor potential to respond, including stimulus response, subsystem DIORs and static surrogates, as well as understanding relationships between these measures within and across the pre- and post-stressor period.

Contrary to the excellent ongoing clinical studies, population studies, to date, are lacking such integrative advancement that investigates multiple approaches at once. In addition, no

longitudinal or population-based frameworks exist to support such integrative measurement. This is an important gap in the literature as clarifying physical resilience in population settings can serve different functions to complement the ongoing clinical studies. In population settings, measures of physical resilience can be used to inform public health policies and programs that target groups of individuals. For example, population approaches to physical resilience can identify high-risk groups, evaluate effectiveness of policy interventions, and monitor the health of populations over time. Furthermore, population approaches can describe resilience in a broader range of (often healthier) individuals than clinical studies.

To address this gap, this thesis aims to complement the ongoing, relatively short-term, clinical work with longitudinal population-level work that simultaneously investigates multiple measurement approaches to frailty and physical resilience. Measures of post-stressor change, residual-based measures, and DIORs can be applied to longitudinal population data, however, stimulus response measures require specific clinical tests and are not currently available in population data. Thus, this thesis will not consider stimulus response measures further. The other three approaches can be understood as complementary and are derived from different conceptualization of the stressor (Table 2).

Table 2. Three Measurement Approaches to Physical Resilience in Population Data

Stressor	System	Outcome
Post-stressor change (Resilience as recovery)		
Incident acute or chronic	Typically whole person	Functional measures*
Residual-based (Resilience as adaptation)		
Existing/cumulative	Whole person	Frailty, Functional measures
DIOR (Resilience as stability)		
"micro-stressors": stochastic deviations in the absence of major stressors	Whole person or subsystem	Biomarkers, health status

*Functional measures include any functional output of the system under study. For example, at the whole person level, a functional outcome could be mobility or ADLs.

Building on the conceptual and empirical reviews, the next chapter proposes the Integrated Stress Response Framework for Frailty and Physical Resilience, specifically designed to guide longitudinal analyses in population data.

Chapter 3: An Integrated Stress Response Framework for Frailty and Physical Resilience

Building on the conceptual and empirical literature reviews in Chapter 2, this chapter proposes the Integrated Stress Response Framework for Frailty and Physical Resilience (ISRF-FPR). This framework describes how frailty and physical resilience relate to each other in response to acute and chronic stressors and emphasizes the importance of longitudinal impacts of these stressors by explicitly incorporating cumulative stress. This framework is specifically for longitudinal, population-level analyses. This framework is flexible, integrating various concepts and methods proposed in the literature. The concept of physiologic reserve plays a central role in the framework.

3.1 The Framework

The ISRF-FPR builds on the standard understanding of physiologic reserve and physical resilience in the literature: physiologic reserve (or "reserve" for short) represents the pre-stressor potential to react to a stressor, whereas physical resilience (or "resilience" for short) is post-stressor realization of such potential (see section 2.2.3) (Figure 4). Stressors can be acute (i.e., short duration, typically with higher severity and immediate effects) or chronic (i.e., longer duration, typically with lower severity and longer-term effects) (see section 2.3.1). Physiologic reserve serves as the focal point of this framework; an individual's level of reserve determines where they are on the spectrum from robust (high reserve) to frail (low reserve). Aggregation of the level of reserve in each sub-system (e.g., organs/organ systems) determines the level of reserve of the whole person. Frailty is typically understood as a low level of physiologic reserve across multiple bodily systems; thus, the ISRF-FPR considers frailty to be one specific approach to capture the global level of reserve. It is well established that frailty influences an individual's ability to respond to stressors. For example, a frail individual (i.e., a person with low reserve) is more likely to have a poor outcome (e.g., loss of function) after encountering an acute or chronic stressor compared to a non-frail, robust individual (i.e., a person with higher reserve). Thus, the level of reserve constrains the level of physical resilience. However, the level of physical resilience is not wholly determined by the level of reserve. While frail individuals are typically non-resilient and robust individuals are typically resilient, there are numerous observations that frail individuals are

unexpectedly resilient and robust individuals are unexpectedly non-resilient, resulting in surprising health outcomes.

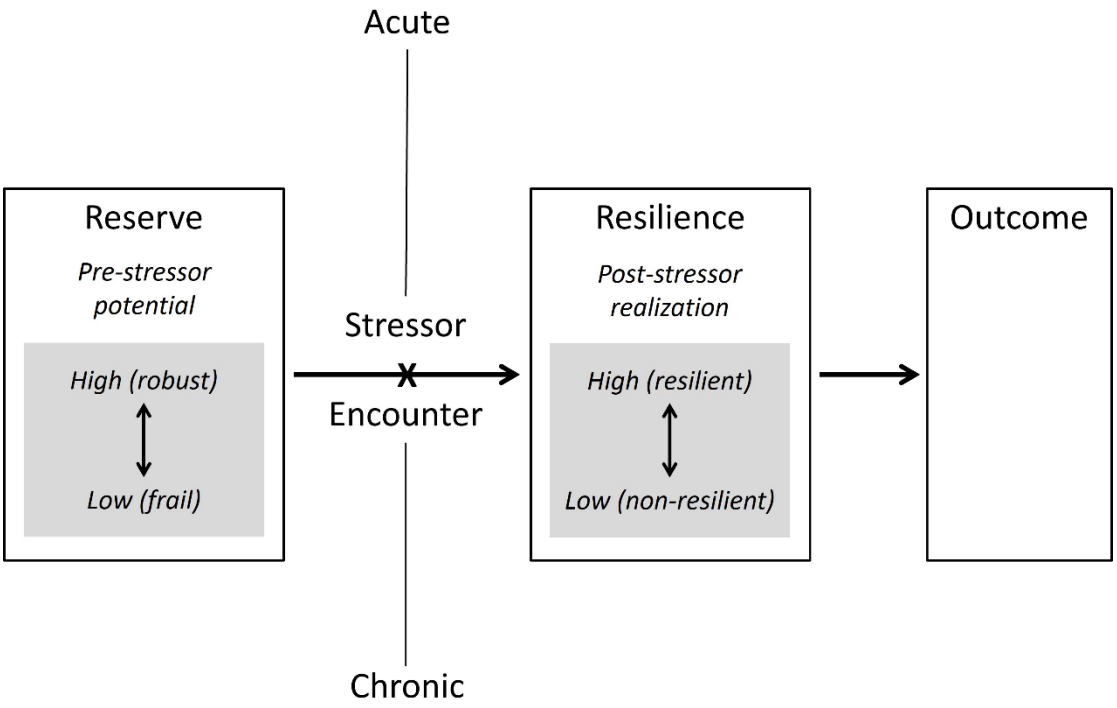


Figure 4. Base Framework Representing Single Stressor Encounter

Reserve represents the spectrum from robust to frail and determines an individual's pre-stressor potential to respond. Resilience represents the post-stressor realization of that potential.

Thus far, the framework is similar to those proposed by Whitson et al. (48) and Walston et al. (52), used to guide the PRIME-KNEE and SPRING studies, respectively (see section 2.2.4). These frameworks focus on a single, relatively short-term stressor. However, the recent inclusion of intrinsic capacity in the conceptual discourse underscores the importance of longitudinal assessment (see section 2.2.4). Additionally, focusing on a single stressor encounter does not allow for the perspective of physical resilience as adaptation. To fill this gap, the ISRF-FPR expands upon the scope of these existing frameworks by explicitly acknowledging the longitudinal cycle of stress. Specifically, an individual's pre-stressor reserve is determined by multiple factors, including age, sex, and, importantly, cumulative stress that the body has experienced over the life course, up until that point in time (Figure 5). Newly encountered acute and chronic stressors that

challenge the body can contribute to this cumulative stress and, in turn, potentially "use up" a portion of reserve, rendering the body more susceptible to future stress. The dashed lines in Figure 5 show a cyclic understanding of stress response.

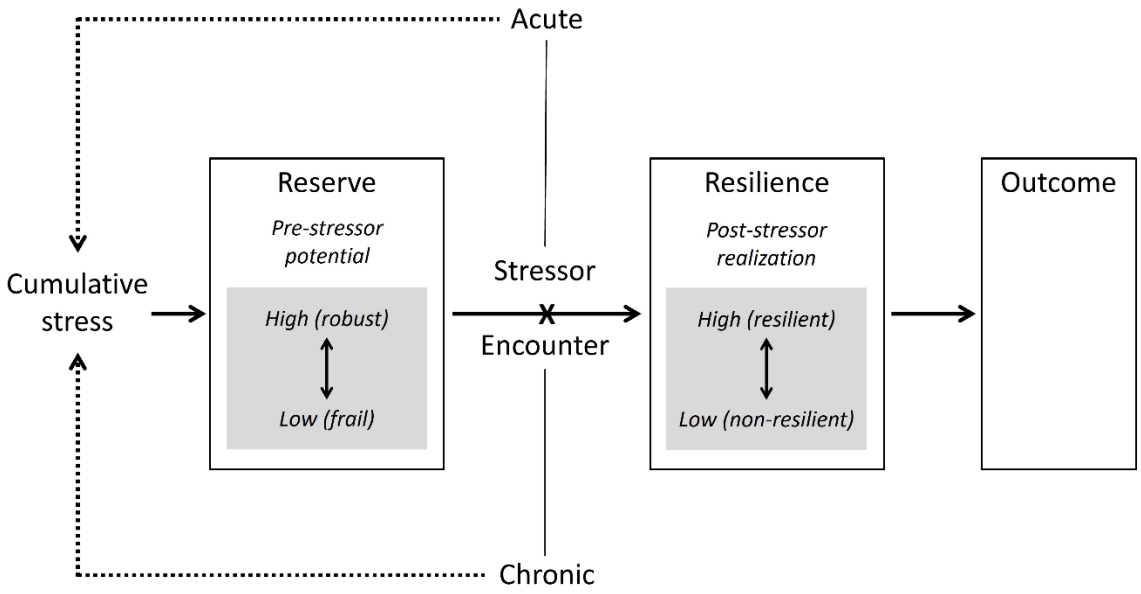


Figure 5. Longitudinal Cycle of Stress

Acute and chronic stressors accumulate over the life course and impact the level of reserve, which, in turn, affects the ability to respond to new stressors.

Importantly, to what extent acute and chronic stressors contribute to cumulative stress, which, in turn, impacts future reserve, is likely mediated by the level of reserve and the level of resilience at the time of the encounter with the stressor (Figure 6). The same stressor would therefore have less impact on future reserve for robust and/or resilient individuals than frail and/or non-resilient individuals. The dashed lines in Figure 6 show a cyclic understanding of stress and its relationship with reserve and resilience.

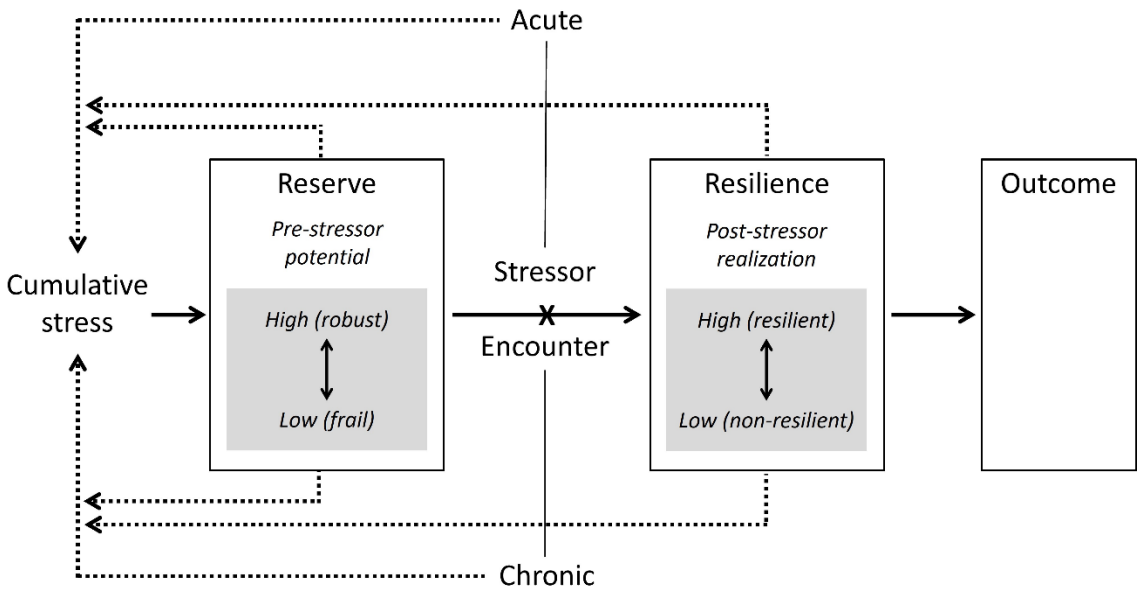


Figure 6. Impact of Reserve and Resilience on Longitudinal Cycle of Stress

The level of reserve and resilience at the time of the stressor encounter influences the level of reserve in the long term, and thus, the potential to respond to future stressors.

The ISRF-FPR next maps promising measurement approaches in the existing empirical literature (Figure 7). These measurement approaches correspond to the three concepts of physical resilience: resistance and recovery, adaptation, and stability. The concept of resilience as resistance and recovery corresponds to measures of post-stressor change; the concept of resilience as adaptation corresponds to residual-based measures conditioning on cumulative stress; and the concept of resilience as stability corresponds to stochastic deviations in the absence of defined stressors. In addition, the concept of resilience as stability also fits another measurement approach, longitudinal trajectories, as discussed in section 2.3.4. Longitudinal trajectory measures complement measures of stability (i.e., how much an individual varies around their mean trajectory), rather than being a measure of physical resilience per se (79,80). However, longitudinal trajectories have strong face validity as a dynamic surrogate indicator of physical resilience and embody the potential importance of longitudinal monitoring.

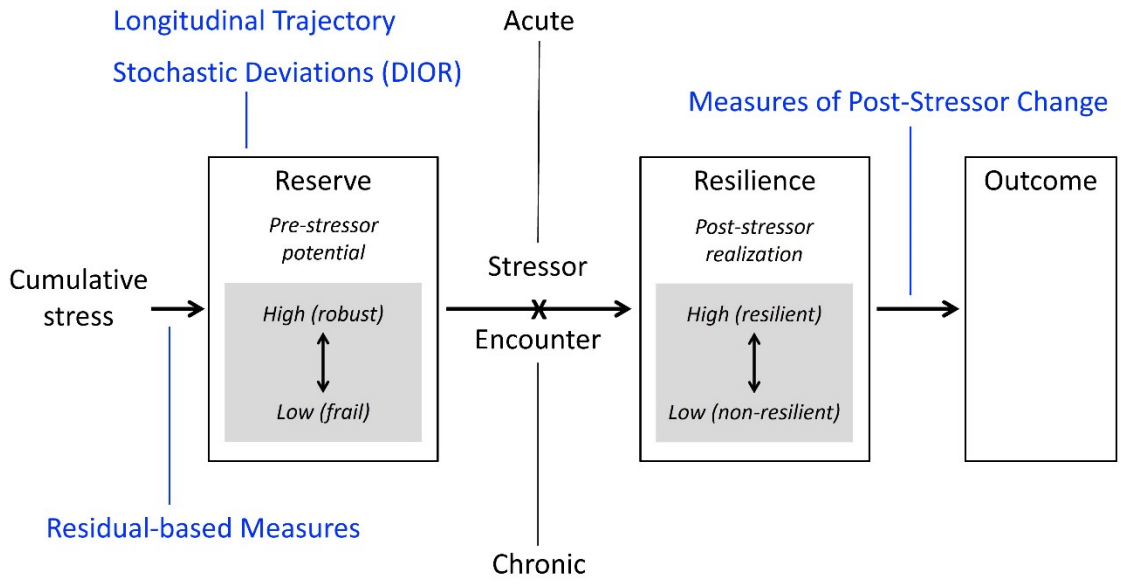


Figure 7. Measurement Approaches to Resilience Using Longitudinal Population Data

The blue text maps measurement approaches from the empirical review onto the Integrated Stress Response Framework for Frailty and Physical Resilience. A key element of this figure is what side of the stressor encounter these measurement approaches fall on. Longitudinal trajectory, DIORs, and residual-based measures represent the pre-stressor potential to respond: they are indicators of resilience that can be estimated without the observation of a specific incident stressor. Measures of post-stressor change are direct measures of resilience after the observation of a specific incident stressor.

Importantly, all of these measures, except measures of post-stressor change, are located on the left-hand side of the figure, representing the pre-stressor potential. These pre-stressor measures are *indicators* of resilience, as opposed to direct *measures* of resilience. They do not require the observation of a specific incident stressor. Thus, they can be estimated in the entire population, rather than a clinical population, and can predict future stress response.

The ISRF-FPR is flexible as it integrates multiple concepts and methods proposed in the literature when existing frameworks typically focus on a single stressor and do not consider alternative conceptualizations of physical resilience, such as resilience as adaptation. The framework thus far primarily focuses on frailty as an approach to capture the level of reserve at the whole person level, but lower-level applications are possible. For example, instead of operationalizing frailty as a global measure of reserve, one could operationalize it as a specific kind of reserve of a particular body system, or even multiple specific kinds of reserve, which would allow the assessment of cross-correlation of DIORs across multiple body systems.

Now returning to the original purpose of the thesis with the ISRF-FPR laid out, the objective is to see if indicators of resilience can complement static measures of frailty to improve risk estimation in older adults.

Chapter 4: Research Objectives

The overall goal of this project is to elucidate the relationship between frailty, physical resilience, and risk in aging populations by providing the first comparative empirical evaluation of multiple measures of frailty and physical resilience. To meet this goal, I address three specific objectives:

1) Guided by the ISRF-FPR operationalize multiple specific measures of frailty and physical resilience and describe the distributions of each measure.
2) Evaluate how frailty and physical resilience relate to mortality, by exploring independent associations, potential effect modifications and the discriminatory ability of the selected measures.
3) Evaluate how frailty and physical resilience relate to acute functional recovery, by exploring independent associations, potential effect modification, and the discriminatory ability of the selected measures.

Chapter 5: Methods

5.1 Operationalization of the ISRF-FPR and Overall Analytic Approach

Though the operationalization of the ISRF-FPR, introduced in Chapter 3, could take many forms, the frailty index (FI) serves as the basis of the current investigation. As previously mentioned in section 2.1.1, Rockwood and Mitnitski suggested that deficit accumulation is the basis for loss of physiologic reserve and, thus, is indistinguishable from loss of reserve (21). Additionally, the FI is a continuous variable and is constructed in such a manner that it can increase or decrease as deficits are accumulated or recovered over time, and past research has consistently demonstrated its strong relationship with mortality. Thus, the FI is an excellent choice to operationalize pre-stressor reserve.

Based on this operationalization of reserve, I use a mixed effects growth curve modelling approach to operationalize three additional pre-stressor indicators of resilience:

1. **Rate of Aging (RoA)**: Defined as the average rate of deficit accumulation in the polynomial growth model.
2. **Dynamical Indicator of Resilience (DIOR-FI)**: Defined as the variability of residuals around estimated FI trajectories.
3. **Frailty-Disease Mismatch (FM)**: Defined as a point estimate of the expected vs. observed FI based on the population average model (fixed effects only), accounting for stressors.

These three measures correspond to the three identified pre-stressor measurement approaches in the ISRF-FPR: RoA is a measure of longitudinal trajectory, DIOR-FI is a measure of stochastic deviations (around estimated FI trajectories), and FM is a residual-based measure. Figure 8 below shows these specific measures mapped on to the framework in place of the general measurement approach categories. Furthermore, these three approaches represent resilience conceptualized as stability (RoA and DIOR-FI), and resilience conceptualized as adaptation (FM).

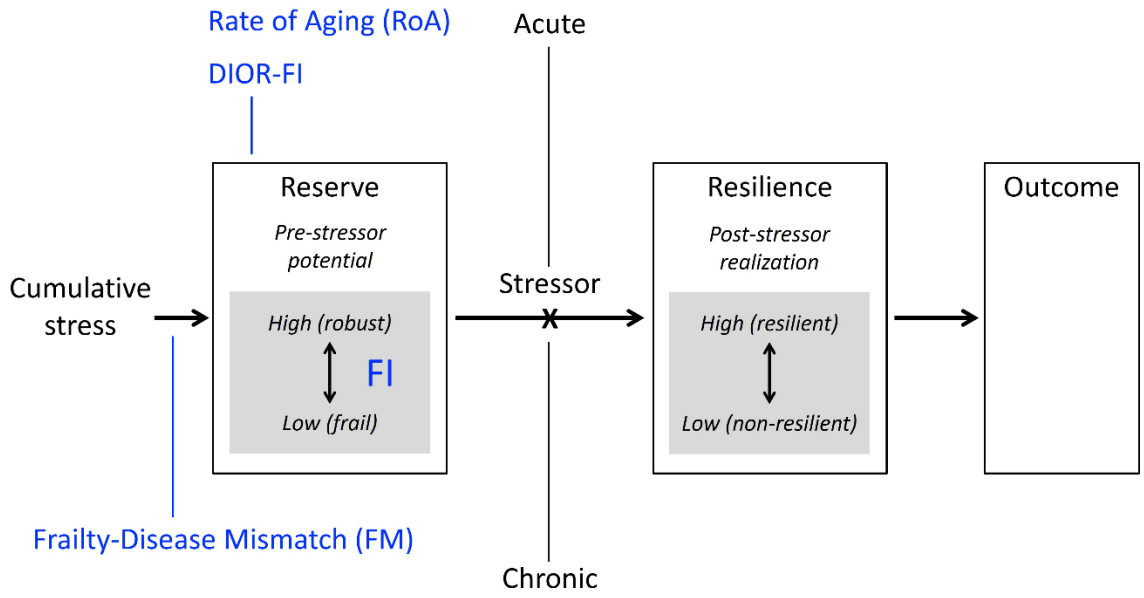


Figure 8. Operationalization of Pre-stressor Indicators of Resilience

This figure shows the central operationalization of reserve (frailty index – FI), and the specific operationalized measures of each general measurement approach category: RoA is a longitudinal trajectory approach, DIOR-FI is a longitudinal stochastic deviations approach, and the FM is a residual-based measure (point-estimate). Together, these measures based on the FI represent resilience as stability (RoA and DIOR-FI), and resilience as adaptation (FM).

I specifically chose to operationalize and evaluate pre-stressor indicators of resilience as these are the universal measures: they can be estimated in anyone without the requirement of having to observe a specific stressor. Additionally, as discussed in Chapters 2 and 3, experiencing a stressor may "use up" a portion of reserve. Thus, in addition to limiting the population to those who have experienced a specific stressor, measures of post-stressor change may not reliably reflect future ability to respond to stress.

Overall Analytic Approach

This thesis follows a two-stage modelling approach. Stage one estimates the pre-stressor indicators using information derived from a mixed effect growth curve model and provides a descriptive analysis of each measure (objective 1). Using the estimated measures from Stage 1 as independent variables in logistic regression models, Stage 2 estimates their effect on two relevant outcome measures: mortality and functional recovery (objectives 2 and 3, respectively).

Note that this is only one operationalization of the framework, and many different approaches could be employed. The aim of this empirical work is to explore how we can leverage longitudinal population data to estimate resilience, indicated by pre-stressor potential, reserve.

5.2 Data and Study Population

This study employs a longitudinal cohort study design, using data from the Health and Retirement Study (HRS). HRS is well suited for this study as it is one of the largest and longest running health and aging population-based surveys in the world, offering a nationally representative sample of the non-institutionalized population over the age of 50 in the United States. HRS was established as a research platform to understand the aging process and the impacts of social and policy changes at the national level and has become the model for a network of similar studies around the globe, producing harmonized data on the Gateway to Global Aging Data Platform (90). HRS collects a breadth of information in four broad topic areas: income and wealth, health and healthcare services, work and retirement, and family connections (90). HRS has collected information on participants every two years since the first wave in 1992 and includes a detailed exit survey and follow-up protocol which provides near-complete mortality capture and has been validated using records from the National Death Index (91).

The original HRS target population for the first wave included all adults residing in households in the contiguous United States, born between 1931 and 1941 (aged 51-61 at enrollment). In wave two, the original HRS cohort merged with the Asset and Health Dynamics Among the Oldest Old (AHEAD) cohort which includes individuals born before 1924 (aged 70+ at enrollment). In wave four, HRS added two new cohorts to provide a complete representative sample of the entire 50+ age range, including individuals born between 1942 and 1947 (War Babies cohort), and those born between 1924 and 1930 (Children of Depression cohort) to fill the previous cohort gap. Since wave four, HRS has maintained this representative coverage by replenishing the sample every six years (three waves) with the addition of the following six-year birth cohort: Early Baby Boomers (born between 1948 and 1953) were added in 2004, Middle Baby Boomers (born between 1954 and 1959) were added in 2010, Late Baby Boomers (born between 1960 and 1965) were added in 2016, and most recently, Early Generation X (born between 1966 and 1971) were added in 2022.

HRS employs a multi-stage area probability sample design, with oversampling of Black and Hispanic respondents as well as respondents from the state of Florida. Sample weights are included in datasets to account for differential probability of selection and non-response in each wave (92). HRS defines the observational unit as a household financial unit with at least one age-eligible member of the recruited cohort. A household financial unit consists of related individuals living in the same dwelling (e.g., a respondent and their spouse). If the age-eligible recruited household member has a spouse, HRS also recruits the spouse regardless of their birth cohort. In the event there is more than one financial unit in the same household (i.e., unrelated individuals living together), one financial unit is randomly selected for inclusion (93). Institutionalized populations, including incarcerated individuals or those in nursing homes or long-term care facilities are excluded at recruitment (92). However, recruited individuals continue to be followed if they transition into a nursing home or long-term care facility (90). Historically, HRS typically conducted baseline interviews in person with follow-up interviews conducted over the phone. In 2006 HRS began a mixed follow-up procedure in which half of the sample completed the core questionnaire over the phone while the other half completed an in-person interview for additional data collection, such as physical measures and collection of biological specimens. With the release of the wave 15 data (2020), HRS has collected and published publicly available longitudinal information on more than 42,000 individuals.

I used the RAND longitudinal file, randhrs1992_2020v1, as the primary data source, and merged additional longitudinal variables from Harmonized HRS D 1992-2021 and individual HRS exit files, wave 6 to 13 (1996 to 2016). All data is publicly available from the University of Michigan HRS data portal (https://hrsdata.isr.umich.edu/).

5.3 Analytical Sample

This study used longitudinal data from waves 3 to 14 of HRS to create two analytical samples: a mortality sample and a recovery sample (Figure 9). To be included in the mortality sample individuals must be aged 50 or older at the time of interview (e.g., an individual younger than 50 at recruitment in HRS is included in my mortality sample only when they turn 50), were interviewed between waves 3 and 13, have known vital status in wave 14, have a minimum of three FI estimates prior to death or survival to wave 14, and not be missing key covariates used in growth curve models (described in section 5.4.1). To be included in the recovery sample

individuals must be aged 50 or older at the time of interview, observed between waves 3 and 14, report a first-time myocardial infarction (MI) between waves 6 and 13, have a minimum of three FI estimates prior to their first-reported MI, have available pre- and post-MI information to determine whether the individual made a full recovery (i.e., pre-MI function, and post-MI function or post-MI death), and not be missing key covariates used in growth curve models. I chose these criteria because HRS includes small numbers of individuals below the target population age of 50 (i.e., spouses of respondents), and one of the unique measures in this study requires a minimum of three repeated measures (the DIOR-FI, described in section 5.4.4). Furthermore, I excluded waves 1 and 2 due to noncomparability of FI variables, and a substantial proportion of missing values for several FI variables, respectively (further described in Appendix C).

The final analytical sample for the mortality analysis consists of 27,744 individuals with a total of 190,553 repeated FI observations between waves 3 and 13. This results in an average of 6.87 observations per individual (standard deviation (SD) = 2.82, range = 3-11). The final analytical sample for the recovery analysis consists of 1,905 individuals with a total of 10,085 repeated FI observations between waves 3 and 12, resulting in an average of 5.29 repeated observations per individual (SD = 2.07, range = 3-10).

In addition to the main analyses, I conducted a sensitivity analysis using a restricted sample of individuals with three consecutive FI measurements prior to the event (incident MI) or outcome (death/survival to wave 14) of interest. The purpose of this sensitivity analysis was to check the robustness of the results to differential lengths of follow up across individuals and non-consecutive observations. Applying this criterion resulted in a restricted mortality sample of 23,644 and a restricted recovery sample of 1,839.

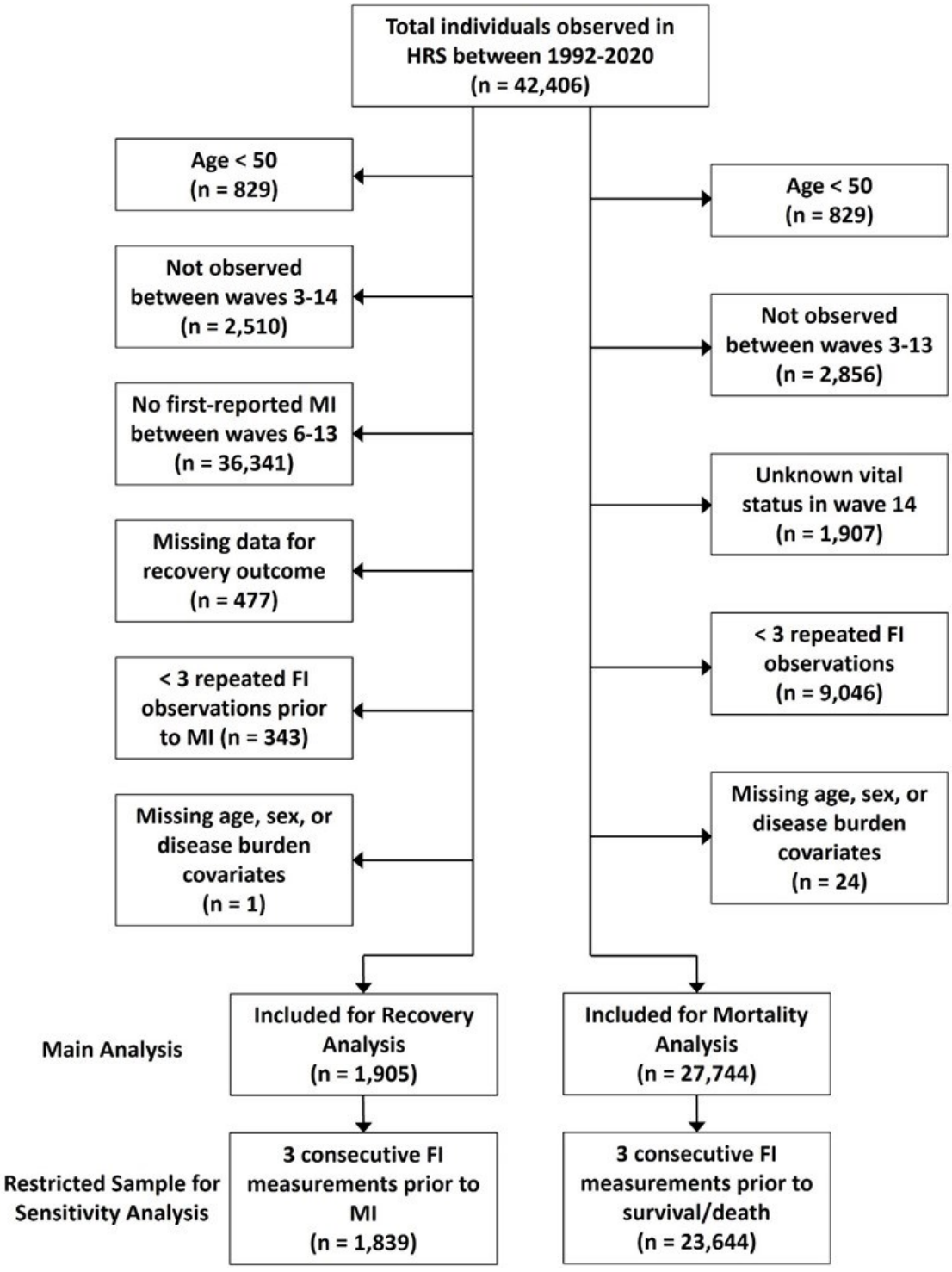


Figure 9. Analytical Sample Selection Flow Chart

Flowchart illustrating the step-by-step process for sample selection, including data exclusion, eligibility criteria and participant inclusion for the main and restricted analyses, respectively. 42,406 represents the number of individuals in the two main longitudinal analytic files: randhrs1992_2020v1 and Harmonized HRS D 1992-2021. MI stands for myocardial infarction. FI stands for frailty index.

5.4 Objective 1: Deriving Indicators of Resilience and Descriptive Analyses

This section describes the methodology used to meet objective 1: operationalize multiple specific measures of frailty and physical resilience and provide a descriptive analysis of each measure.

5.4.1 Mixed Effects Growth Curve Modelling

Growth curve modelling refers to a set of statistical methods used for analyzing repeated measures to estimate between-individual differences in within-individual patterns of change over time (94). In general, growth curve modelling can be accomplished via two approaches: mixed effect modelling or latent trajectory modelling, a form of structural equation modelling. This analysis used a mixed effect approach because this approach has been demonstrated to be more advantageous when dealing with complex data structures such as time-unstructured data (i.e., not all individuals are measured at the same time or at every wave) and multiple levels of nesting (95). HRS employs a rolling enrollment with individuals being measured at different ages and in different waves (e.g., not all early cohort members were recruited at ages 50-54, and not all individuals have observations in all waves). Additionally, HRS has a three-level nesting structure: repeated measures nested within individuals, and individuals nested within households.

Mixed effects models are models with a mixture of fixed and random effects. The term “fixed effects” refers to coefficients that are not allowed to vary by individual: everyone has the same coefficients. The term “random effects” refers to coefficients that are allowed to vary by individual, allowing for the estimation of individual growth curves.

Model selection including testing different functional forms of age (i.e., linear, quadratic, cubic), random effects (i.e., linear age only, or linear plus polynomial terms), random effect structures, and residual correlation structures. Nested models were compared using a likelihood ratio test, and non-nested models were compared using the Akaike information criterion (AIC, where lower values indicate better model fit). Models estimated for comparison purposes used maximum likelihood, while final models were estimated using restricted maximum likelihood. All growth curve models estimated in this thesis share the specifications outlined in Table 3. I found these parameters to produce the best fitting model with no convergence issues across samples. Appendix D provides additional modelling details.

Table 3. Common Growth Curve Model Specifications

Functional form of age	Cubic polynomial
Random effects	Intercept and slope of age (linear component only, not squared or cubed).
Random effect structure	General positive-definite matrix, with no additional structure, using a log-Cholesky parameterization. Random effects are not assumed to be independent, and the degree of covariance is estimated from data.
Residual correlation structure	Continuous autoregressive order-1 correlated residuals.
Estimation method	Restricted maximum likelihood (REML).

The mixed effect growth curve models consist of repeated FI observations nested within individuals. Given the household sampling design of HRS, I evaluated the household clustering effects in both the mortality and recovery samples to determine whether to include household as a third level in the models (i.e., individuals nested within households). To make this decision, I considered the design effect, model fit, and additional complexity of including the third level. The design effect uses the average cluster size and the intraclass correlation coefficient (ICC) to determine the magnitude of negative bias on the standard errors resulting from the non-independence within clusters.

The mortality sample includes 18,595 households with 27,744 individuals. With an average cluster size of 1.49 individuals per household, and an ICC of 0.54, the design effect for household clustering is 1.27. Previous literature suggested that it is not necessary to include a random effect unless the design effect is greater than two (96). Allowing effects to vary by household did however improve fit as evaluated by the AIC and likelihood ratio test, but the added complexity also led to convergence errors in some of the sensitivity and subgroup analyses. Considering these different factors, I decided to use the simpler two-level model for the main analysis but included a sensitivity analysis with the best-fitting household random effects included.

The recovery sample includes 1,839 households with 1,905 individuals.[11] With an average cluster size of 1.04, and an ICC of 0.63, the design effect for household clustering is negligible at 1.02. Thus, I did not include a household sensitivity analysis for the recovery sample. Appendix D provides further details on the calculation of the design effect.

I evaluated model assumptions and found that residuals were not normally distributed nor homoscedastic (in both samples). Given its distribution, the frailty index is often log transformed to improve model residuals. Transforming the dependent variable slightly improved but did not fix these violations. After careful consideration of the purpose of these models – to extract residuals and/or individual coefficients after explicitly accounting for specific theoretically informed variables – I kept the dependent variable on the original scale to favor interpretability over improving the residuals. I judged violations of these model assumptions as acceptable given the goal of the analysis and the large sample size.

Model Variables

Below I describe the variables used in growth models. Briefly, the dependent variable is the frailty index, the main independent variable is age, and the control covariates are sex, wave, and nine variables used to capture an individual's disease burden: whether the respondent has ever had any of seven diseases, self-rated health, and whether the respondent regularly uses prescription medication. These disease burden variables were chosen following Wu et al. (43) to operationalize the frailty-disease mismatch (described in section 5.4.3).

Frailty Index: a continuous, time varying measure of overall health at the time of each interview. Specifically, I constructed a 41-item Frailty Index for each individual in each wave following guidance from previously published work (20,97). The 41 items span eight domains including self-rated health, hearing and vision, activities of daily living (ADLs), instrumental activities of daily living (IADLs), receiving assistance with ADLs and IADLs (e.g., receives help bathing), other functional and mobility limitations (e.g., difficulty picking up a dime), medication use (e.g., takes medication for hypertension), signs and symptoms (e.g., urinary incontinence), and equipment

[11] Average household size was smaller in the recovery sample due to nature of selection. Everyone with a known vital status was included in mortality, while only individuals who had a heart attack were included in recovery. Having a household where both spouses had a heart attack was much less common than having a household where both members had known vital status.

use (e.g., wears hearing aid). I assessed all variables for suitability using criteria from Blodgett (97). Variables that meet the criteria are 1) associated with age, 2) do not saturate too early,[12] 3) are not too common (>80%) or too rare (<1%), and 4) have minimal missing data (<5%). Individuals must not be missing 20% or more of answers to the deficit questions for an FI to be calculated. In addition to these criteria from Blodgett, if any variables are too highly correlated (0.95 or greater), I only included one of the pair to avoid redundancy (O Theou, personal communication, June 10, 2022). Given this project calls for a longitudinal application, chosen variables must be comparable across waves, and the FI must be identically constructed across all waves (20). I excluded waves 1 and 2 due to limitations in creating a comparable frailty index across waves: wave 1 had several questions that were asked differently, leading to non-comparable responses, and wave 2 had a large proportion of missing on several variables.

Importantly, I have omitted chronic disease conditions and self-rated health from the 41-item frailty index given concerns over using these variables in both the left- and right-hand side of the growth curve model equation (i.e., both as independent variables, and components of the dependent variable, the frailty index). Given the interchangeable nature of components of the frailty index, I hypothesized that the inclusion or exclusion of these variables should not have a large impact on results. To confirm this, I also ran a sensitivity analysis with a 51-item frailty index which included self-rated health and chronic diseases. Furthermore, since the beginning of this project, a more recent article was published which acts as a guide to creating a frailty index, formalizing the criteria noted above and providing step-by-step instructions. This published example used the same data (HRS), from waves 5 to 12. The major difference in this published FI and the current FI is the inclusion of health service utilization variables. I did not consider these variables for inclusion in the FI given concerns over health care access and financial implications of the US health system. To see if this makes a difference I included a third, 56-item frailty index as a sensitivity analysis that is comparable to that published in Theou et al. (98). Appendix C provides further details on creation and comparison of all three frailty indices.

Age: a continuous, time-varying measure of the respondent's age in years at the date of each interview. I created this variable by taking the respondent's age in months at the end date of the

[12] For example, presbyopia would not be a good candidate deficit as it becomes nearly universal by age 55 (20).

interview and dividing it by twelve. The RAND codebook suggests using the age at the end date of the interview (rather than the beginning or middle) because if the interview took place over multiple dates, the bulk was usually done at the end. The original age in months variable was determined by subtracting the respondent's birthdate from the interview end date.

Sex: a binary, time-invariant indicator of the respondent's biological sex with two categories: male and female. Female is the reference category.

Wave: an indicator variable of the interview wave, ranging from 3 to 13. Wave 3 is the reference category. I included wave to account for any potential differences between interview waves (e.g., different interviewers or modes of interview) and/or period effects (e.g., the 2008 economic crisis). Including wave resulted in a better model fit, indicated by a significant likelihood ratio test, so it remained in the final adjusted model.

Disease variables: binary, time-varying indicators of whether the respondent has ever had any of the following conditions:

1. Arthritis or rheumatism
2. Hypertension or high blood pressure
3. Stroke or transient ischemic attack
4. Diabetes or high blood sugar
5. Cancer or malignant tumor of any kind except skin cancer
6. Chronic lung disease (excluding asthma, chronic bronchitis, and emphysema)[13]
7. Heart problems including heart attack, coronary heart disease, angina, congestive heart failure, or other heart problems

For each of these conditions, "No" is the reference category.

Self-rated health: a time-varying indicator of how a respondent views their overall health at the time of each interview, with response categories: excellent, very good, good, fair, and poor. Excellent is the reference category.

[13] Disease exclusions such as skin cancer in the cancer variable or emphysema in the chronic lung disease variable are present in the HRS data, not imposed by the author.

Regularly Takes Prescription Drugs: a binary, time-varying indicator of whether the respondent has regularly used prescription medication in the last two years prior to each interview. "No" is the reference category.

Model 1: Unadjusted (age-only)

I estimated age-based trajectories of the frailty index using a polynomial (cubic) specification. The models have two levels: up to 11 repeated measures of the FI (j), nested within individuals (i). Age in years at the time of the repeated measure, along with it's powers (i.e., age-squared and age-cubed) are the only predictor in this model. The model includes a random intercept and random slope for age (but not age-squared or age-cubed). I centred age at 50 so that the intercept (fixed and random) represents FI at the lower age limit in the sample.

$$\text{FI}_{ij} = \beta_0 + \beta_1 \times \text{age}_{ij} + \beta_2 \times \text{age}_{ij}^2 + \beta_3 \times \text{age}_{ij}^3 + b_{0i} + b_{1i} \times \text{age}_{ij} + \epsilon_{ij} \quad (1)$$

Where FI_{ij} is the FI of individual i at the jth repeated measurement occasion. β_0 is the fixed intercept, β_1 to β_3 are the fixed effects of age, age-squared, and age-cubed, respectively. b_{0i} is the random effect for the intercept, b_{1i} is the random effect for the slope of age, and ϵ_{ij} is the residual.

Model 2: Adjusted Model

The disease burden adjusted model expands upon the unadjusted model by adjusting for the fixed effects of the sex, wave, and the nine disease burden variables described above, represented by the addition of $C\delta$ below:

$$\text{FI}_{ij} = \beta_0 + \beta_1 \times \text{age}_{ij} + \beta_2 \times \text{age}_{ij}^2 + \beta_3 \times \text{age}_{ij}^3 + \delta\text{C}_{ij} + b_{0i} + b_{1i} \times \text{age}_{ij} + \epsilon_{ij} \quad (2)$$

Where FI_{ij} is the FI of individual i at the jth repeated measurement occasion. β_0 is the fixed intercept, β_1 to β_3 are the fixed effects of age, age-squared, and age-cubed, respectively. δC_{ij} represents the matrix of control covariates (C_{ij}) and their corresponding fixed effects (δ). b_{0i} is the random effect for the intercept, b_{1i} is the random effect for the slope of age, and ϵ_{ij} is the residual.

5.4.2 Frailty-Disease Mismatch (FM)

The frailty-disease mismatch (FM) is a residual-based measure of physical resilience that describes the difference between the expected and observed frailty of an individual based on a population model that accounts for age, sex, wave, and disease burden (indicative of an individual's level of cumulative stress) (43). Individuals who have lower frailty than expected are "adapters," as they have been better able to adapt to the cumulative stress of aging and disease burden (i.e., are more resilient). Conversely, those with higher-than-expected frailty are "premature frailers." I generated a point estimate of FM for each individual at each measurement occasion by subtracting the population predicted value from the observed FI value. I estimated the population predicted values for each individual by generating predictions for the adjusted model (equation 2) using the fixed effects only. Using only the fixed effects emulates the between individual model of Wu et al. (43), the original method authors, and mirrors the mixed effects approach for longitudinal data used by Milman et al. (69).

Given that the resulting distribution of FM is not normal, I used the 25th and 75th percentiles to determine cut points: individuals equal to or below the 25th percentile are the "adapters" (i.e., considered the most resilient), and individuals above the 75th percentile are the "premature frailers" (i.e., considered the least resilient). Individuals between the 25th and 75th percentiles are "expected agers" and serve as the reference group.

5.4.3 Rate of Aging (RoA)

The rate of aging describes an individual's rate of deficit accumulation. This is operationalized as the mean of an individual's derivatives across all repeated measurement occasions. This reflects an individual's average rate of aging across their non-linear trajectory estimated in the unadjusted, age-only model (equation 1). To estimate RoA, I extracted individual coefficients (fixed plus random) from the model using a post-estimation command from the nlme package, which uses an empirical best linear unbiased prediction approach to estimate random effects. I used these individual coefficients in the derivative equation of equation 1:

$$\mathrm{RoA}_{ij} = \beta_1 + 2\beta_2 \mathrm{age}_{ij} + 3\beta_3 \mathrm{age}_{ij}^{\,2} + b_{1i} \quad (3)$$

This equation finds the instantaneous slope of FI based on the respondent's age at each repeated measurement occasion.

I created two versions of RoA. The first version used the average of all derivatives at each interview for each individual. The RoA for individual i is the average over all j (repeated measures) for this individual, where n is the total number of repeated measures for individual i:

$$RoA_i = \frac{\sum_{j=1}^{n} RoA_{ij}}{n} \quad (4)$$

However, this first version was found to be problematic after comparing to the main sensitivity analysis using only three interviews per individual (described in section 5.7). Thus, I created the second version of RoA using the average of the last three derivatives only. The second version ensures all individuals have a comparable estimate, regardless of how many interviews they completed. The second version of RoA for individual i is the average over the last three j (repeated measures) for this individual:

$$RoA_i = \frac{\sum_{j=n-2}^{n} RoA_{ij}}{3} \quad (5)$$

Below I present both results, using the first and second versions of ROA, in the order of the chronological development of the project: results using the first version of RoA (all derivatives), the sensitivity analysis, then results from the second version of RoA (last three derivatives). The results from the second version of RoA are the primary results of interest, because the second version of RoA provides comparable estimates across individuals with different lengths of observation. See sections 5.7 and 6.1.9 for details.

To remain consistent with the FM, I categorized RoA by the 25th and 75th percentiles: "slow agers" are individuals equal to or below the 25th percentile, and "fast agers" are individuals above the 75th percentile. "Average agers" fall between the 25th and 75th percentiles and serve as the reference group.

5.4.4 A Dynamical Indicator of Resilience Based on the Frailty Index (DIOR-FI)

I used the long-term intra-individual variability of the frailty index as a longitudinal macro-scale dynamical indicator of resilience (DIOR). As individuals accumulate or recover from deficits over

time, they display unique patterns of variation around their mean trajectory. This provides dynamic health information that conventional analyses often overlook, and closely resembles the microrecovery approach employed by typical DIORs. Similar to what is observed with micro-scale DIORs, I hypothesized that given the same average frailty trend, high stability reflects a high level of resilience (Figure 10, top row), and low stability reflects a low level of resilience (Figure 10, bottom row).

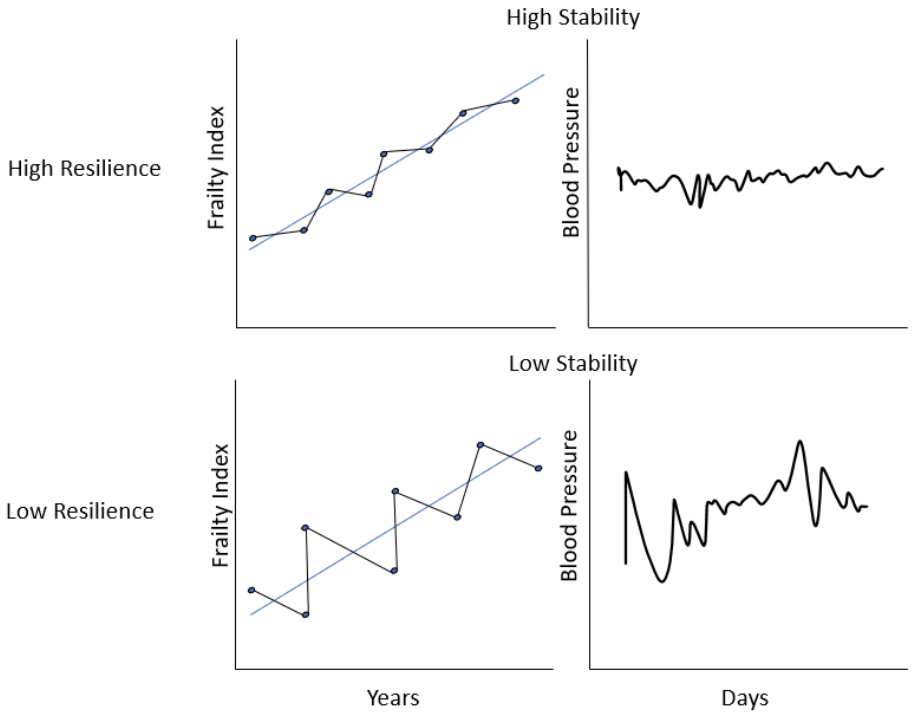


Figure 10. Longitudinal Frailty Index Instability as a Dynamical Indicator of Resilience

Depicted on the right is blood pressure measurement over the course of days representing a typical, short-term DIOR. The variation represents fluctuations around an average baseline value. Depicted on the left is the proposed longitudinal, macro-scale DIOR: Frailty index over the course of years. Variation represents fluctuations around an average trajectory (analogous to the baseline in a short-term DIOR).

The DIOR-FI is defined as the standard deviation of the raw within-individual residuals in the adjusted model (equation 2). In contrast to the population-average prediction residuals used for creating the FM (i.e., predictions using the fixed effects only), the DIOR uses the individual prediction residuals (i.e., predictions using both fixed and random effects).

Following the decisions made for RoA, as explained above, I also estimated a second DIOR-FI using only the last three residuals to provide comparable estimates across individuals with different lengths of follow-up. Similar to the results of RoA, below I present both results in the order of the chronological development of the project, but the results from the DIOR-FI using the last three residuals are the primary results of interest. See sections 5.7 and 6.1.9 for details.

I classified the DIOR following the previous resilience variables: individuals equal to or below the 25th percentile have high stability (resilient), and individuals above the 75th percentile have low stability (non-resilient). Individuals between the 25th and 75th percentiles have average stability and function as the reference group.

5.4.5 Descriptive Analysis

The descriptive analysis includes non-parametric univariate and bivariate descriptive statistics to describe the continuous distribution of each measure (FI, RoA, FM, DIOR-FI), as well as the distribution of other variables (age, sex, mortality, recovery, FI, RoA, FM, DIOR-FI) across the categories of each measure. I presented univariate descriptive statistics with the median and interquartile range (IQR) limits (i.e., the 25th and 75th percentiles) as the variables of interest are not normally distributed. I compared continuous variables across categories using the Wilcoxon rank sum test for two groups or the Kruskal-Wallis rank sum test for more than two groups. I calculated Pearson correlation coefficients for all pairwise continuous measures, and I used Cohen's Kappa (unweighted) to determine the agreement between the resilience categorizations.

Previous literature proposed 0.03 as the minimal important difference in FI at the conservative end (99,100). I used this value when making comparisons in the descriptive analyses. In addition to continuous FI, this study also used FI categories for descriptive analyses: less than or equal to 0.1 is non-frail, greater than 0.1 but less than or equal to 0.21 is vulnerable, greater than 0.21 but less than or equal to 0.45 is frail, and greater than 0.45 is most frail (34,101).

I used the last values for the point estimates (i.e., observed FI and estimated FM at the final interview) in the descriptive analysis to maintain consistency with the values used in the subsequent logistic models (described in section 5.6). For example, for an individual who died in wave 6, point estimates from wave 5 were used (i.e., the most recent FI and FM), along with the longitudinal information estimated from all three waves prior (i.e., RoA and DIOR-FI estimated

from waves 3-5). Similarly, for an individual who remained alive through the observation period (i.e., is still alive at wave 14), the point estimates from wave 13 were used, along with the longitudinal information estimated from all observed waves prior. The last/most recent value was chosen for point estimates to represent the most up to date (and informative) estimate of health before the index event (MI) or outcome (death/survival). In particular, the last observed frailty index represents the conventional risk assessment that we are trying to improve with indicators of resilience.

5.5 Summary of Independent Variables Carried Forward to Step 2

In total, this study evaluated four individual-level variables. These include the frailty index (FI) the rate of aging (RoA), the frailty-disease mismatch (FM) and a longitudinal dynamical indicator of the frailty index (DIOR-FI). Figure 11 shows a simplified illustration of these four measures. Furthermore, Table 4 summarizes the four measures and provides an example interpretation of each. I estimated all resilience variables (RoA, FM, and DIOR-FI) separately for each sample (mortality and recovery).

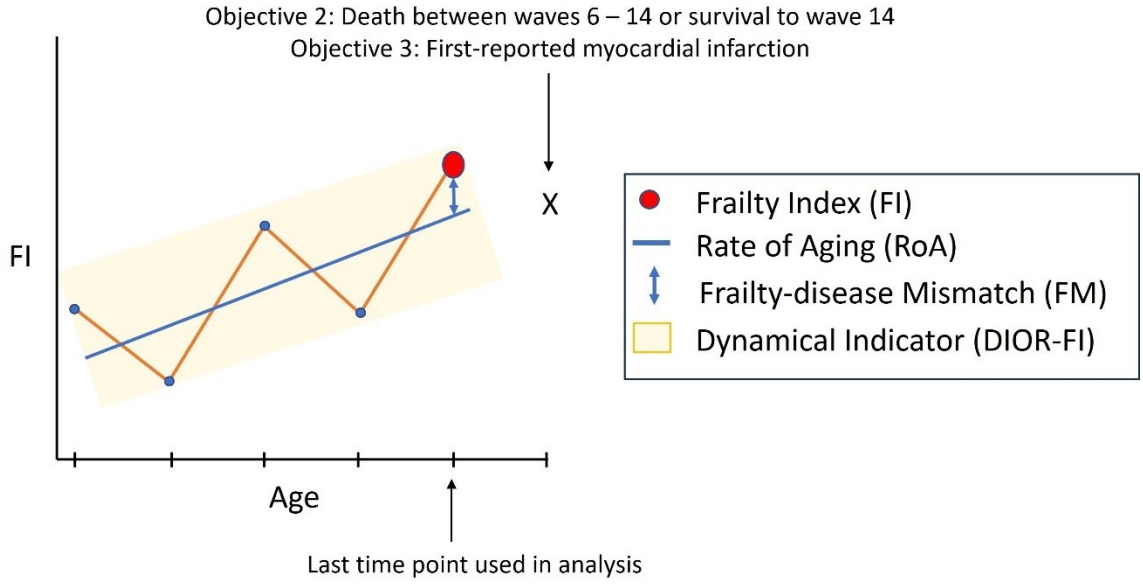


Figure 11. Illustration of Frailty and Resilience Measures in Relation to Mortality and Recovery

This figure shows the temporal relationship of all four independent variables of interest, and the outcomes for objectives 2 and 3.

Table 4. Summary of Frailty and Resilience Measures

Variable	Categorization	Interpretation
Frailty Index (FI)	Continuous FI is the primary measure, but descriptive analyses use categories: Non-frail < 0.1 Vulnerable (0.1, 0.21) Frail (0.22, 0.45) Most Frail >0.45	Frailty index at a single point in time. A threshold minimum important difference of 0.03 is used when comparing groups (99,100). This is equal to 1.23 deficits in the 41-item frailty index. I used the most recent observation in all analyses.
Frailty-Disease Mismatch (FM)	Categorized by IQR: Adapters (resilient) ≤ Q1 Expected Agers (reference) (Q1, Q3) Premature Frailers (non-resilient) > Q3	The differential between the observed frailty index and the expected frailty index given the respondent's level of disease burden at a single point in time. Those with higher frailty than expected are considered non-resilient, and those with lower frailty than expected are considered resilient. I used the most recent estimate prior to the event of interest (MI or death), or end of follow up.
Rate of Aging (RoA)	Categorized by IQR: Slow Agers (resilient) ≤ Q1 Average Agers (reference) (Q1, Q3) Fast Agers (non-resilient) > Q3	Unadjusted rate of deficit accumulation as measured by the frailty index. This estimate reflects individual i's average non-linear slope of frailty over the observation period. E.g., a rate of aging of 0.03 would mean a person's average slope over their observation period is equal to 0.03 FI per year. I estimated RoA using all observations prior to the event of interest (MI or death), or end of follow up.
Dynamical Indicator of Resilience (DIOR-FI)	Categorized by IQR: High Stability (resilient) ≤ Q1 Average Stability (reference) (Q1, Q3) Low Stability (non-resilient) > Q3	DIOR-FI is the standard deviation of the residual, which captures the intraindividual variability of the frailty index over the observation period. Higher numbers indicate more variability (less resilient) around their estimated trajectory, and lower numbers indicate lower variability (more resilient).

Note: Q1 is the 25th percentile, and Q3 is the 75th percentile. I used these cut points because no established cut points exist, and this makes sense from a distributional perspective as none of the measures were normally distributed.

5.6 Objectives 2 and 3: Estimating Effects of Frailty and Resilience on Mortality and Recovery

5.6.1 Mortality Models

I used logistic regression models to estimate the effects of the independent variables on the odds of two-year all-cause mortality. Details of the models are provided below.

Data

This analysis compares the last observed/estimated values in survivors vs decedents. This corresponds to wave 13 for survivors, and the wave before death for decedents (can be any wave between 6 and 13).

Measures

The outcome measure in the mortality models is all-cause mortality. I assigned a mortality value of 1 to individuals who died by any cause, and 0 to individuals who survived. For both survivors and decedents, there is approximately a 2-year window between the ascertainment of the independent variables and the reporting of the outcome (survival/death). The control variables are age at final interview and sex. The main independent variables of interest are the FI, FM at final interview, and the RoA and DIOR-FI estimated using all prior waves (see sections 5.4. and 5.5 for details).

Regression Models

I chose logistic regression (rather than a duration model) so that longitudinal variables (i.e., RoA and DIOR-FI) could be estimated and compared between survivors and decedents. I estimated a series of logistic regression models predicting mortality in four successive modelling steps: 1) I ran unadjusted models for each measure (FI, FM, RoA, DIOR-FI), 2) I accounted for the influence of FI by including it as a covariate in the resilience models. The adjustment aimed to isolate specific contribution of resilience beyond the effect attributable to frailty. 3) I further refined the resilience models by adjusting for age, sex and FI. This adjustment aimed to account for the potential influence of age and sex on the relationship between resilience and mortality, while still considering the effect of FI as a covariate. 4) I tested for interaction between the resilience

indicator and FI within the age-sex adjusted model. This allows for exploring if the relationship between resilience and mortality varies across different levels of FI, while considering the effect of age and sex.

I interpreted regression coefficients for each model and compared the discriminatory ability by the area under the receiver operating characteristic curve (AUC). To determine the added discriminatory value of each resilience indicator, I compared the AUC of the adjusted resilience models to the age-sex-adjusted FI model. Finally, I ran a combined model with all resilience indicators to determine the maximal increase in AUC.

5.6.2 Recovery Models

Similar to the mortality analysis, I used logistic regression models to estimate the effects of the independent variables on the odds of full functional recovery after an incident myocardial infarction. Details of the models are provided below.

Data

Conditioning on first-time MI events between waves 6 and 13, I used data from each individual's pre- and post-MI waves for the recovery analysis. Information from the pre-MI wave is used to predict functional recovery in the post-MI wave. I chose MI as the acute stressor because it is a common and well characterized major health event that can impact whole-person physical functioning. Restricting to first-reported MI events should reduce variability in pre-disposing factors that cannot be fully accounted for in the available data (e.g., the number of MI events prior to enrollment in HRS). Repeat heart attacks may be more likely to cause more damage than a single heart attack and, thus, be more difficult to fully recover from. In absence of other severity indicators, I restricted the analysis to first-reported MI events to reduce variability in severity.

Measures

The outcome measure in the recovery models is post-MI full functional recovery (see below for details). The control variables are pre-MI age and pre-MI physical function (see below for details). The main independent variables of interest are the pre-MI FI, FM, RoA, and DIOR-FI (see sections

5.4 and 5.5 for details). Sex was not included in the recovery models as it was not statistically significant.

Pre-MI physical function: To remain consistent with previous functional recovery literature, physical function was assessed using a modified SF-36 physical function subscale (SF-36 PFS). While HRS data does not include the SF-36 PFS, it does ask questions that closely resemble nine of the ten items of the physical function subscale (the missing dimension is “walking more than a mile” but there is a question elsewhere in the HRS questionnaire on ability to walk several blocks that I used as a proxy, see below), allowing a coarse approximation to be calculated (102). A key difference is that the SF-36 PFS asks participants to what degree each activity is limited by their health (three response categories: “limited a lot”, "limited a little”, “not limited at all”), while the HRS variables are binary indicators of difficulty on the items. Following Wei et al., I imputed the missing item “walking more than a mile”, using the respondent’s answer for “walking several blocks”, with the assumption that individuals who experience difficulty walking several blocks would also experience difficulty walking more than a mile (102). The 10 items are added together and rescaled for a final score ranging from 0 to 100, with higher scores indicating less functional difficulty. The SF-36 PFS is an appropriate tool to identify changes in the level of physical functioning as the items cover a wide range of functional ability from basic activities of daily living such as bathing or dressing, to vigorous physical activities such as jogging. Following the guidance in the SF-36 Manual, I calculated a score if a respondent answered at least half of the items. I imputed the missing items with the average score of the answered items (103). Appendix E provides a full item comparison of the validated SF-36 PFS and the constructed HRS equivalent.

Full Functional Recovery (post-MI): I assigned a recovery value of 1 to individuals who returned to, or exceeded their pre-MI level of functioning at the wave after the MI was reported. If the individual was still below baseline at this wave, or was deceased, I assigned a value of 0. For example, if an individual reported their first MI in wave 8, the recovery value would be determined by the difference in function between waves 7 and 9 (or alternatively, their vital status in wave 9). Importantly, the SF-36 PFS has a floor effect: those who start with zero function cannot decrease any further. This applies to 243 individuals (12.8%) in the sample. The only way for these individuals to demonstrate a lack of full recovery is to die. Thus, full recovery for this subgroup is equivalent to survival. To maximize sample size and retain those individuals in poorest health, I

included these individuals in the sample, and adjusted all recovery models for baseline (pre-MI) function.

Regression Models

Similar to the mortality analysis, I estimated a series of logistic regression models predicting mortality in four successive modelling steps (see section 5.6.1 for steps and rationale). However, the recovery models have two key differences: 1) all models are adjusted for function to account for the SF-36 PFS floor effect, and 2) sex is not adjusted for because it was determined to be insignificant in the recovery models.

5.7 Sensitivity Analyses

I ran a total of ten sensitivity analyses to examine the robustness of the results (Table 5). The main sensitivity analysis used a restricted sample of three consecutive FI measurements prior to the event of interest (MI/death), or end of follow up. The purpose of this was to check the robustness of the resilience variables to differential lengths of follow up and non-consecutive observations. Importantly, this led to second sensitivity analysis testing a modification in the estimation of RoA and DIOR-FI between the main and restricted sample, which ultimately lead to a change in how these variables were estimated (see results section 6.1.9). All subsequent sensitivity analyses used the corrected "last three" longitudinal measures, which estimated the RoA and DIOR-FI using only the last three time points per individual (while still using all longitudinal information to estimate the growth curve model).

Additional sensitivity analyses included age- and sex-stratified models, alternative cut points for resilience variables (top and bottom 15% compared to the original 25%), continuous measure models, alternative FI models (51-item and 56-item), and a household random effects model (allowing intercept and the slope of age to vary by household). I determined cut points for the age-stratified analysis using the 33rd and 67th percentiles to maximize sample size in each group. The resulting groups are 52-67 (n = 9,150), 68-79 (n = 9,002), and 80-109 (n = 9,592).[14] The final sample sizes for the sex-stratified analyses are 11,903 for males, and 15,841 for females. I

[14] Note: the youngest age is 52 because this was the age of the youngest individual who happened to be interviewed three times above the age of 50 prior to their final observation. This illustrates the variability in interview timing for the approximate 2-year wave cycle.

performed stratified analyses only in the mortality sample due to the limited sample size and power in the recovery sample. Additionally, I included two sensitivity analyses using differently constructed frailty indexes: a 51-item FI that includes disease burden variables (originally left out due to overlap with the adjusted growth curve model), and a 56-item that includes health service utilization, and is comparable to a recently publish guide by Theou et al. (98). Lastly, I estimated a cross-sectional FM for individuals in wave 10 aged 70 to 79 who reported no difficulty climbing one flight of stairs or walking several blocks (n = 3,173). The purpose of this was to match the population used by Wu et al. (43) to see if differences in population can explain differences in results. I chose wave 10 to allow enough time for deaths to occur after measurement of FM. Table 5 lists all sensitivity analyses and the corresponding rationale.

Table 5. Sensitivity Analyses

#	Sensitivity Analysis	Rationale
1	Restricted sample to three consecutive FI measurements	To check robustness to differential length of follow-up and non-consecutive observations
2	Comparing main vs. restricted samples with revised estimation of RoA and DIOR-FI	To check if observed differences between the main and restricted sample sensitivity analysis could be eliminated by revising how RoA and DIOR-FI were estimated
3	Age stratification	To check for differences by age
4	Sex stratification	To check for differences by sex
5	Alternative cut points for categorical resilience indicators : top and bottom 15%	To check robustness to the cut points used for resilience categorization
6	Continuous resilience indicators	To check how much discriminatory ability is lost by categorizing the resilience indicators
7	Alternative FI – 51-item	To check if results are robust to the inclusion of disease variables in the FI
8	Alternative FI – 56-item	To check if results are robust to the inclusion of health service utilization variables in the FI
9	Household Random Effects	To check if results are robust to the specification of household clustering effects
10	Cross-sectional FM – Wu et al. (43) sample match	To check if differences in population can explain the disparity between observed results and previous literature.

5.8 Sample Weights and Statistical Software

All estimates are unweighted as the study design does not have a single start or end date where the cross-sectional weights can be applied to all participants. Thus, estimates of variance do not account for the complex survey design in HRS. I performed all statistical analyses in R version 4.2.3, and RStudio version 2023.06.0+421 "Mountain Hydrangea." I created data visualizations using “ggplot2” (104) and “ggeffects” (105), and summary tables using “gtsummary” (106). I used package “nlme” to estimate mixed effect models (107).

5.9 Research Ethics

As this study exclusively involved secondary use of publicly available data, it was exempt from research ethics board review as outlined in the Tri-Council Policy Statement Article 2.2. (108). Publicly available datasets such as HRS have been de-identified to pose minimal risk to participant privacy, and thus there are no anticipated harms due to use of sensitive participant information in this project.

Chapter 6: Results

A Guide to Chapter 6

Results are divided into two sections, mortality analysis (section 6.1) and recovery analysis (section 6.2). Each of these sections has two main sub-components: main analysis (sections 6.1.1 to 6.1.8 for mortality analysis and sections 6.2.1 to 6.2.9 for recovery analysis) and main sensitivity analysis (sections 6.1.9 to 6.1.14 for mortality analysis and sections 6.2.10 to 6.2.15 for recovery analysis). The main sensitivity analysis used a restricted sample of individuals with only three consecutive observations. Based on the results of the main sensitivity analysis, I made a decision to keep the full, unrestricted sample, but to change the estimation of the two longitudinal variables (RoA and DIOR-FI) to make them more comparable across individuals: the estimation was changed from using all longitudinal information to estimate RoA and DIOR-FI, to using only the last three observations for each person. For this reason, for each mortality and recovery analysis, the main analysis was repeated using the corrected estimates of RoA and DIOR-FI. While presenting a full picture, this chapter presents results with cumbersome repetitions. Readers interested in an abbreviated read of the results can read only section 6.1.5 (for frailty-disease mismatch results) and sections 6.1.9 to 6.1.14 (for the corrected RoA and DIOR-FI results) for the mortality analysis and section 6.2.6 (for frailty-disease mismatch results) and sections 6.2.10 to 6.2.15 (for the corrected RoA and DIOR-FI results) for the recovery analysis.

Each main section (mortality, 6.1, and recovery, 6.2) presents the results in the same order. The sections begin with an overview of the sample used (overall and stratified by outcome), followed by the results of the growth curve models used to estimate the resilience variables. Then, an analysis of each independent variable of interest (FI, FM, RoA, DIOR-FI) which includes univariate and bivariate descriptive analyses and presentation of estimated coefficients in the series of logistic regression models. Finally, the main results conclude with the correlation and agreement between continuous and categorical independent variables, and a comparison of the discrimination (AUC) of all estimated models. Sensitivity analyses present a brief comparison to the main results with the full details in the corresponding appendices.

Individual variable results are presented in the following order: FI, FM, RoA, DIOR-FI. FI is presented first as this is the base variable we wish to complement with additional resilience

indicators. FM is presented second as this is the only other static measure that can be estimated cross-sectionally. RoA is presented third as this represents an individuals longitudinal trajectory. DIOR-FI is fourth as this represents the variation around an individual's estimated trajectory.

6.1 Mortality Analysis

6.1.1 Sample Characteristics

The mortality sample of 27,744 individuals is 57% female with a median age of 74 (IQR bounds = 64, 82) and a median FI of 0.183 (IQR bounds = 0.085, 0.360) (Table 6). In terms of the indicators of resilience, the sample has a median rate of aging of 0.005 FI per year (IQR bounds = 0.003, 0.012), a median frailty-disease mismatch of -0.026 deviation from the expected frailty (IQR bounds = -0.082, 0.068), and a median DIOR-FI of 0.046 (IQR bounds = 0.028, 0.077).

Of the 27,744 included in the sample, 11,154 (40%) died by wave 14. Survivors and decedents are significantly different, both statistically and in terms of the effect size,[15] across all variables using the Wilcoxon rank sum test for continuous variables and Pearson's Chi-squared test for categorical variables ($p < 0.001$). Compared to survivors, decedents have a higher median age (81 years compared to 68 years), a higher median FI (0.331 compared to 0.122), a higher proportion of males (46% compared to 41%), a higher median rate of aging (0.012 compared to 0.004), a higher median frailty-disease mismatch (0.012 compared to -0.036), and a higher median variability captured by the DIOR-FI (0.066 compared to 0.037).

After categorizing by the 25th and 75th percentiles, all resilience indicators remain statistically significantly different between survivors and decedents ($p < 0.001$). In terms of the RoA, decedents have a higher proportion of fast agers (52% compared to 6.6% of survivors) and a lower proportion of slow agers (8.2% compared to 36%). In terms of the FM, decedents have a higher proportion of premature frailers (39% compared to 16%) but unexpectedly they also have a slightly higher proportion of adapters (26% compared to 24%). Lastly, in terms of the DIOR-FI,

[15] Minimal important differences in effect sizes for the continuous resilience variables have yet to be determined, though the differences between survivors and decedents seem to be of reasonable magnitude. This issue is further discussed in Chapter 7.

decedents have a higher proportion of individuals with low stability (42% compared to 14%) and a lower proportion of individuals with high stability (11% compared to 35%).

Table 6. Mortality Sample Characteristics by 2018 Vital Status

		2018 Vital Status		
Characteristic	**Overall**, N = 27,744[1]	**Alive**, N = 16,590[1]	**Deceased**, N = 11,154[1]	**p-value**[2]
Age	74 (64, 82)	68 (62, 77)	81 (73, 88)	**<0.001**
Sex				**<0.001**
Female	15,841 (57%)	9,810 (59%)	6,031 (54%)	
Male	11,903 (43%)	6,780 (41%)	5,123 (46%)	
Frailty Index	0.183 (0.085, 0.360)	0.122 (0.067, 0.226)	0.331 (0.183, 0.567)	**<0.001**
FI Category				**<0.001**
Non-frail	8,340 (30%)	7,091 (43%)	1,249 (11%)	
Vulnerable	6,964 (25%)	4,852 (29%)	2,112 (19%)	
Frail	7,405 (27%)	3,591 (22%)	3,814 (34%)	
Most Frail	5,035 (18%)	1,056 (6.4%)	3,979 (36%)	
FM	-0.026 (-0.082, 0.068)	-0.036 (-0.080, 0.020)	0.012 (-0.086, 0.169)	**<0.001**
FM Category				**<0.001**
Adapter	6,936 (25%)	4,041 (24%)	2,895 (26%)	
Expected Ager	13,872 (50%)	9,949 (60%)	3,923 (35%)	
Premature Frailer	6,936 (25%)	2,600 (16%)	4,336 (39%)	
RoA	0.005 (0.003, 0.012)	0.004 (0.002, 0.006)	0.012 (0.006, 0.020)	**<0.001**
RoA Category				**<0.001**
Slow Ager	6,936 (25%)	6,017 (36%)	919 (8.2%)	
Average Ager	13,872 (50%)	9,475 (57%)	4,397 (39%)	
Fast Ager	6,936 (25%)	1,098 (6.6%)	5,838 (52%)	
DIOR-FI	0.046 (0.028, 0.077)	0.037 (0.023, 0.058)	0.066 (0.042, 0.106)	**<0.001**
DIOR-FI Category				**<0.001**
High Stability	6,936 (25%)	5,747 (35%)	1,189 (11%)	
Average Stability	13,872 (50%)	8,559 (52%)	5,313 (48%)	
Low Stability	6,936 (25%)	2,284 (14%)	4,652 (42%)	

[1]Median (IQR Bounds); n (%)

[2]Wilcoxon rank sum test; Pearson's Chi-squared test

Age, FI, and FM represent the final values prior to death or end of follow up (wave 13). RoA and DIOR-FI represent estimates from all interviews prior to death or end of follow up (wave 13).

6.1.2 Growth Curve Models

I estimated two models to create the resilience variables. The age-only model produces RoA, and the adjusted model produces FM and DIOR-FI.

Age-only Model

The coefficients and 95% confidence intervals for all model components are presented in Table 7 below. The fixed (population average) growth is described by a linear, quadratic, and cubic age term equal to 0.00673, -0.00040, and 0.00001, respectively. The standard deviation of the linear random effect of age is 0.00586, and the autocorrelation of the residuals is 0.74192 (Table 7).

Table 7. Age-Only Mixed Effects Model Results (Mortality Sample)

Fixed Effects	Estimate	Lower 95%	Upper 95%
(Intercept)	0.08838	0.08519	0.09157
Age	0.00673	0.00619	0.00727
Age^2	-0.00040	-0.00043	-0.00037
Age^3	0.00001	0.00001	0.00001
Random Effects	**Estimate**	**Lower 95%**	**Upper 95%**
Intercept (sd)	0.09322	0.08926	0.09737
Age (sd)	0.00586	0.00567	0.00606
Correlation (age and intercept)	-0.35386	-0.40572	-0.29973
Correlation Structure	**Estimate**	**Lower 95%**	**Upper 95%**
Phi (autocorrelation of residuals)	0.74192	0.73574	0.74801
Residuals	**Estimate**	**Lower 95%**	**Upper 95%**
Within-group residuals (standard error)	0.09530	0.09434	0.09627

Adjusted Model

In the adjusted model, all nine disease burden components are statistically significantly associated with FI ($p < 0.001$, Table 8). After adjusting for sex, disease burden, and wave, the linear, quadratic, and cubic age coefficients changed to 0.00387, -0.00036, and 0.00001, respectively (Table 8). This change in age coefficients towards the null is expected as adding disease burden to the model would explain some of the change in FI over time, resulting in a smaller contribution of age. The

standard deviation of the linear random effect of age was reduced to 0.00464, and the autocorrelation of the residuals was reduced to 0.67949.

Table 8. Adjusted Mixed Effects Model Results (Mortality Sample)

Fixed Effects	Estimate	Lower 95%	Upper 95%
(Intercept)	0.04069	0.03750	0.04388
Age	0.00387	0.00340	0.00435
Age^2	-0.00036	-0.00038	-0.00033
Age^3	0.00001	0.00001	0.00001
Sex: Male	-0.02347	-0.02548	-0.02147
SRH: Very Good	0.00718	0.00581	0.00856
SRH: Good	0.02588	0.02438	0.02738
SRH: Fair	0.06471	0.06302	0.06640
SRH: Poor	0.13689	0.13479	0.13899
Ever had stroke: Yes	0.09571	0.09321	0.09822
Ever had arthritis: Yes	0.03728	0.03582	0.03874
Ever had cancer: Yes	0.00864	0.00649	0.01079
Ever had high blood pressure: Yes	0.02687	0.02537	0.02837
Ever had diabetes: Yes	0.01671	0.01484	0.01858
Ever had lung disease: Yes	0.05025	0.04777	0.05273
Ever had heart problems: Yes	0.02121	0.01946	0.02297
Regularly Takes Rx Meds: Yes	0.01310	0.01177	0.01442
Wave 4	-0.00517	-0.00661	-0.00373
Wave 5	0.00020	-0.00148	0.00189
Wave 6	0.00280	0.00095	0.00464
Wave 7	0.00436	0.00244	0.00627
Wave 8	0.00734	0.00534	0.00935
Wave 9	0.00349	0.00141	0.00558
Wave 10	0.00917	0.00706	0.01128
Wave 11	0.00778	0.00558	0.00998
Wave 12	0.00923	0.00691	0.01154
Wave 13	0.00589	0.00340	0.00838
Random Effects	Estimate	Lower 95%	Upper 95%
Intercept (sd)	0.06272	0.05983	0.06576
Age (sd)	0.00464	0.00452	0.00477
Correlation (age and intercept)	-0.55432	-0.58765	-0.51912
Correlation Structure	Estimate.	Lower 95%	Upper 95%
Phi (autocorrelation of residuals)	0.67949	0.67321	0.68571
Residuals	Estimate.	Lower 95%	Upper 95%
Within-group residuals (standard error)	0.08382	0.08318	0.08447

SRH stands for self-rated health. The reference category is “Excellent”. The reference category for Wave is 3.

6.1.3 Frailty Index

Distribution of the FI

The frailty index has a right-skewed distribution with a median of 0.183, a 25th percentile of 0.085, a 75th percentile of 0.360, a 99th percentile of 0.859,[16] a skewness of 1.174, and a kurtosis of 0.491 (Figure 12).

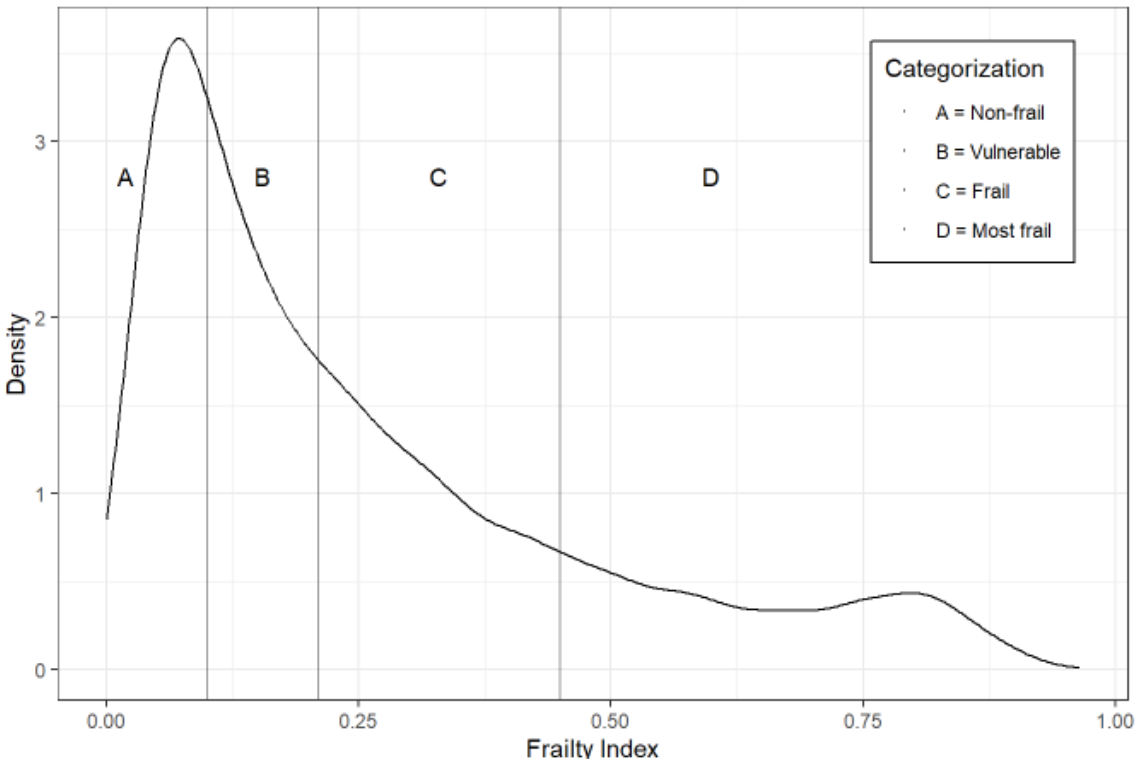


Figure 12. Distribution and Categorization of the Frailty Index (Mortality Sample)

Vertical lines illustrate categorical cut points: less than 0.1 is non-frail, less than or equal to 0.21 is vulnerable, less than or equal to 0.45 is frail, and greater than 0.45 is most frail (34). Note, I used FI categorization for descriptive purposes only.

[16] The 99th percentile of the FI is well above the empirically observed upper limit of 0.7. This corresponds to 2,009 individuals with an FI equal to or greater than 0.70 at their final observation. Of these individuals, 89% (n = 1,779) died before the next wave (within two years). This is consistent with the empirical observation that the body will fail soon after reaching this point. The high 99th percentile is likely a result of two factors: 1) the sample design capturing a large proportion of decedents, and 2) excluding the disease burden variables from the FI. This issue is addressed in the discussion in Chapter 7.

Comparison of FI Categories

All variables are statistically significantly different across frailty index categories using the Kruskal-Wallis test for continuous variables and Pearson's Chi-squared test for categorical variables ($p < 0.001$, Table 9). Age, sex, and mortality show the expected gradient across FI categories: the median age ranges from 66 years in the non-frail group to 83 years in the most frail group, the proportion of females ranges from 50% in the non-frail to 65% in the most frail group, and the proportion of deaths ranges from 15% in the non-frail group to 79% in the most frail group (Table 9).

Similarly, all resilience variables show the expected gradient across all frailty categories: the proportion of DIOR-FI low stability ranges from 3.8% in non-frail to 77% in most frail, and the proportion of DIOR-FI high stability ranges from 58% in non-frail to 1% in most frail (Table 9). The proportion of RoA fast agers ranges from 1.5% in non-frail to 82% in most frail, and the proportion of RoA slow agers ranges from 63% in non-frail to 0.7% in most frail (Table 9). And lastly, the proportion of FM premature frailers ranges from 0% in non-frail to 85% in most frail, and the proportion of FM adapters ranges from 37% in non-frail to 3.7% in most frail.

Table 9. Comparison of FI Categories (Mortality Sample)

Characteristic	FI Category				p-value[2]
	Non-frail, N = 8,340[1]	**Vulnerable**, N = 6,964[1]	**Frail**, N = 7,405[1]	**Most Frail**, N = 5,035[1]	
Age	66 (61, 74)	73 (65, 80)	77 (68, 84)	83 (75, 90)	**<0.001**
Sex					**<0.001**
Female	4,179 (50%)	3,830 (55%)	4,539 (61%)	3,293 (65%)	
Male	4,161 (50%)	3,134 (45%)	2,866 (39%)	1,742 (35%)	
2018 Vital Status					**<0.001**
Alive	7,091 (85%)	4,852 (70%)	3,591 (48%)	1,056 (21%)	
Deceased	1,249 (15%)	2,112 (30%)	3,814 (52%)	3,979 (79%)	
FM Category					**<0.001**
Adapter	3,061 (37%)	2,315 (33%)	1,374 (19%)	186 (3.7%)	
Expected Ager	5,279 (63%)	4,465 (64%)	3,574 (48%)	554 (11%)	
Premature Frailer	0 (0%)	184 (2.6%)	2,457 (33%)	4,295 (85%)	
RoA Category					**<0.001**
Slow Ager	5,257 (63%)	1,378 (20%)	265 (3.6%)	36 (0.7%)	
Average Ager	2,958 (35%)	5,076 (73%)	4,985 (67%)	853 (17%)	
Fast Ager	125 (1.5%)	510 (7.3%)	2,155 (29%)	4,146 (82%)	
DIOR-FI Category					**<0.001**
High Stability	4,852 (58%)	1,615 (23%)	419 (5.7%)	50 (1.0%)	
Average Stability	3,169 (38%)	4,687 (67%)	4,917 (66%)	1,099 (22%)	
Low Stability	319 (3.8%)	662 (9.5%)	2,069 (28%)	3,886 (77%)	

[1]Median (IQR Bounds); n (%)

[2]Kruskal-Wallis rank sum test; Pearson's Chi-squared test

The Effect of FI on Mortality

The unadjusted analysis shows that for every 0.03 increase in FI, the odds of mortality in the next two years increase by 17% (OR = 1.17, 95% CI = 1.17, 1.18) (Table 10). Compared to the non-frail reference group, the odds of mortality are 2.47 times higher for vulnerable (95% CI = 2.28, 2.67), 6.06 times higher for frail (95% CI = 5.59, 6.50), and 21.4 times higher for most frail (95% CI = 19.5, 23.4) (Table 10).

In the age and sex-adjusted analysis, the odds of mortality in the next two years increases by 14% for every 0.03 increase in FI (OR = 1.14, 95% CI = 1.13, 1.14) (Table 10). Compared to the non-frail reference group, the odds of mortality are 1.85 times higher for vulnerable (95% CI = 1.70, 2.01), 4.14 times higher for frail (95% CI = 3.82, 4.50), and 11.8 times higher for most frail (95% CI = 10.7, 13.0) (Table 10).

Table 10. Logistic Regression for the Frailty Index and Mortality

	Unadjusted			Age-Sex Adjusted		
Characteristic	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
Continuous FI						
Frailty Index	1.17	1.17, 1.18	**<0.001**	1.14	1.13, 1.14	**<0.001**
Age				1.08	1.07, 1.08	**<0.001**
Sex						
Female				—	—	
Male				1.81	1.71, 1.92	**<0.001**
Categorical FI						
FI Category						
Non-frail	—	—		—	—	
Vulnerable	2.47	2.28, 2.67	**<0.001**	1.85	1.70, 2.01	**<0.001**
Frail	6.03	5.59, 6.50	**<0.001**	4.14	3.82, 4.50	**<0.001**
Most Frail	21.4	19.5, 23.4	**<0.001**	11.8	10.7, 13.0	**<0.001**
Age				1.08	1.07, 1.08	**<0.001**
Sex						
Female				—	—	
Male				1.82	1.72, 1.94	**<0.001**

Note: Continuous FI odds ratio represents a change of 0.03, the proposed minimal important difference

Figure 13 displays the unadjusted marginal effect of the continuous FI on the predicted probability of mortality. As the frailty index increases, the predicted probability of death monotonically increases. The shaded area around the lines represents the 95% confidence intervals (though they are small and difficult to see given the large sample size).

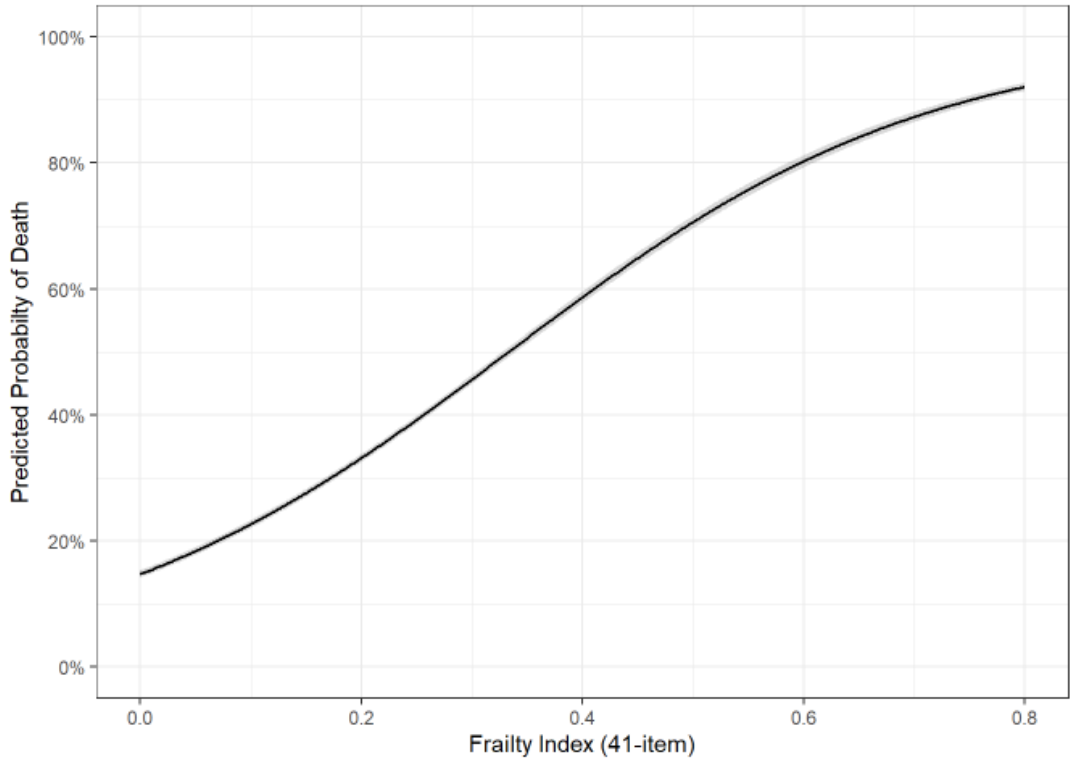


Figure 13. Marginal Effect of the Frailty Index on Mortality

Visualization of the unadjusted effect of FI on mortality (continuous unadjusted model, Table 10).

6.1.4 Frailty-Disease Mismatch

Distribution of the FM

The frailty-disease mismatch has a right-skewed distribution, with a median of -0.026, a 25th percentile of -0.082, a 75th percentile of 0.068, a skewness of 0.988, and a kurtosis of 1.762 (Figure 14).

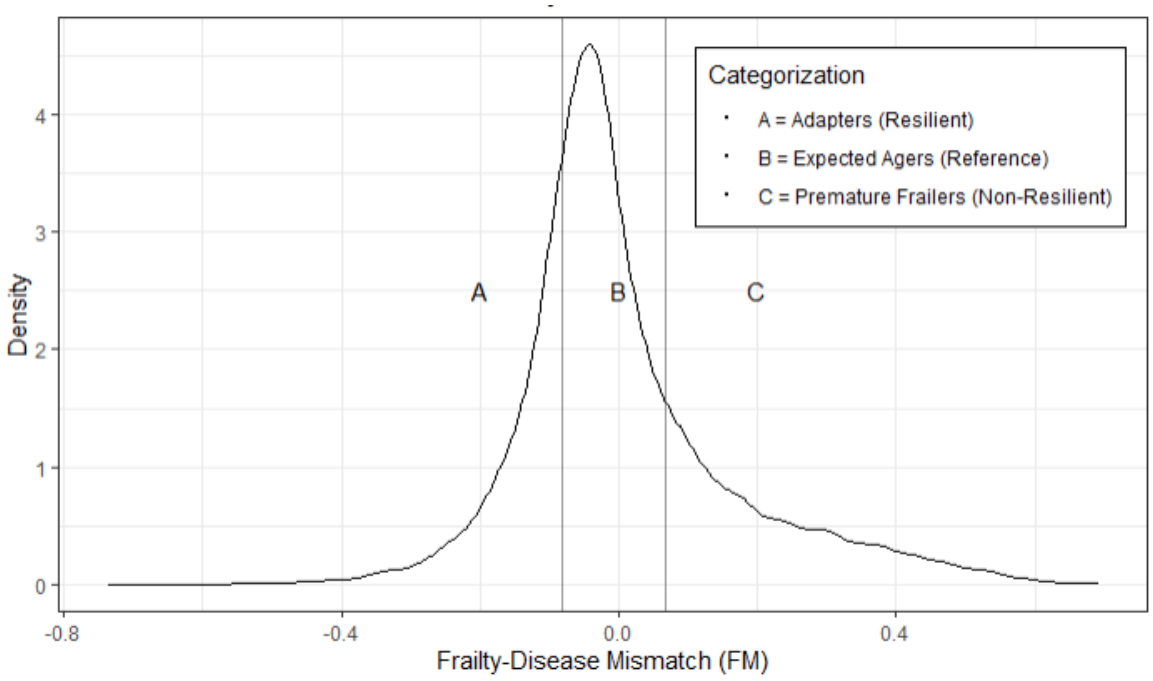


Figure 14. Distribution and Categorization of FM (Mortality Sample)

Physical resilience categories are defined by the 25th and 75th percentiles: those below the 25th percentile are adapters (resilient), those within the IQR are the expected agers (reference), and those above the 75th percentile are premature frailers (non-resilient).

Comparison of FM Categories

All variables are statistically significantly different across FM categories using the Kruskal-Wallis test for continuous variables and Pearson's Chi-squared test for categorical variables ($p < 0.001$, Table 11). Interestingly, most variables do not show the expected trend across all FM categories except frailty. The adapters have the lowest frailty, with a median FI of 0.116 (compared to 0.520 in the premature frailers). Aside from this frailty trend, adapters are the oldest but never have the highest proportion of resilient individuals or the lowest proportion of non-resilient individuals as categorized by the RoA and DIOR-FI, as expected. Compared to the adapter and premature frailer categories, the expected ager category has a higher proportion of slow agers and individuals with high stability, and a lower proportion of fast agers and low stability. Also unexpectedly, the adapters have a higher proportion of deaths compared to the expected agers (42% compared to 28%). Despite the adapters deviating from this expected trend, the comparison between the expected agers and premature frailers is consistent across all variables, with premature frailers being older (77 years compared to 69 years) and having higher proportion of deceased (63%

compared to 28%), fast agers (54% compared to 11%), and individuals with low stability (64% compared to 10%).

Of note, the difference in median FI of 0.018 from FM adapters to expected agers is below the minimum important difference threshold of 0.03.

Table 11. Comparison of FM Categories (Mortality Sample)

Characteristic	FM Category			p-value[2]
	Adapter, N = 6,936[1]	**Expected Ager**, N = 13,872[1]	**Premature Frailer**, N = 6,936[1]	
Age	80 (71, 87)	69 (62, 77)	77 (67, 85)	**<0.001**
Frailty Index	0.116 (0.067, 0.196)	0.134 (0.073, 0.232)	0.520 (0.372, 0.714)	**<0.001**
Sex				**<0.001**
Female	3,955 (57%)	7,631 (55%)	4,255 (61%)	
Male	2,981 (43%)	6,241 (45%)	2,681 (39%)	
2018 Vital Status				**<0.001**
Alive	4,041 (58%)	9,949 (72%)	2,600 (37%)	
Deceased	2,895 (42%)	3,923 (28%)	4,336 (63%)	
RoA Category				**<0.001**
Slow Ager	1,967 (28%)	4,782 (34%)	187 (2.7%)	
Average Ager	3,259 (47%)	7,599 (55%)	3,014 (43%)	
Fast Ager	1,710 (25%)	1,491 (11%)	3,735 (54%)	
DIOR-FI Category				**<0.001**
High Stability	1,439 (21%)	5,339 (38%)	158 (2.3%)	
Average Stability	4,429 (64%)	7,117 (51%)	2,326 (34%)	
Low Stability	1,068 (15%)	1,416 (10%)	4,452 (64%)	

[1]Median (IQR Bounds); n (%)
[2]Kruskal-Wallis rank sum test; Pearson's Chi-squared test

The Effect of FM on Mortality

Unadjusted logistic regression analysis shows that adapters have 1.82 times greater odds of dying compared to the expected agers (95% CI = 1.71, 1.93), and premature frailers have 4.23 times greater odds of dying compared to the expected agers (95% CI = 3.98, 4.50) (Table 12). The frailty-adjusted model shows an increase in the odds ratio for adapters (OR = 2.41, 95% CI = 2.25, 2.57), and a large decrease and change in direction in the odds ratio for premature frailers (OR = 0.39, 95% CI = 0.35, 0.43). Further adjusting for age and sex diminished the effects of both categories, with the odds ratio for adapters decreasing to 1.37 (95% CI = 1.27, 1.48), and the odds ratio for premature frailers increasing to 0.64 (95% CI = 0.58, 0.71).

Table 12. Logistic Regression Models for FM and Mortality

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
FM Category												
Expected Ager	—	—		—	—		—	—		—	—	
Adapter	1.82	1.71, 1.93	**<0.001**	2.41	2.25, 2.57	**<0.001**	1.37	1.27, 1.48	**<0.001**	1.85	1.62, 2.12	**<0.001**
Premature Frailer	4.23	3.98, 4.50	**<0.001**	0.39	0.35, 0.43	**<0.001**	0.64	0.58, 0.71	**<0.001**	1.62	1.34, 1.95	**<0.001**
Frailty Index				1.25	1.24, 1.26	**<0.001**	1.18	1.17, 1.19	**<0.001**	1.24	1.22, 1.26	**<0.001**
Age							1.06	1.06, 1.07	**<0.001**	1.06	1.06, 1.06	**<0.001**
Sex												
Female							—	—		—	—	
Male							1.85	1.75, 1.97	**<0.001**	1.90	1.79, 2.01	**<0.001**
FM Category * Frailty Index												
Adapter * Frailty Index										0.96	0.94, 0.98	**<0.001**
Premature Frailer * Frailty Index										0.92	0.90, 0.93	**<0.001**

Model 1 is unadjusted. Model 2 is adjusted for frailty. Model 3 is further adjusted for age and sex. Model 4 further includes a frailty interaction. Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference.

The interaction terms between FI and FM are statistically significant for both FM categories (adapters and premature frailers, $p < 0.001$, Table 12), and Figure 15 below visualizes this interaction. At low levels of frailty, adapters start with the highest predicted probability of death, which converges with expected agers around an FI of 0.4 (with confidence intervals overlapping around 0.3). The premature frailers have the lowest predicted probability of death at higher levels of FI, starting around 0.2.

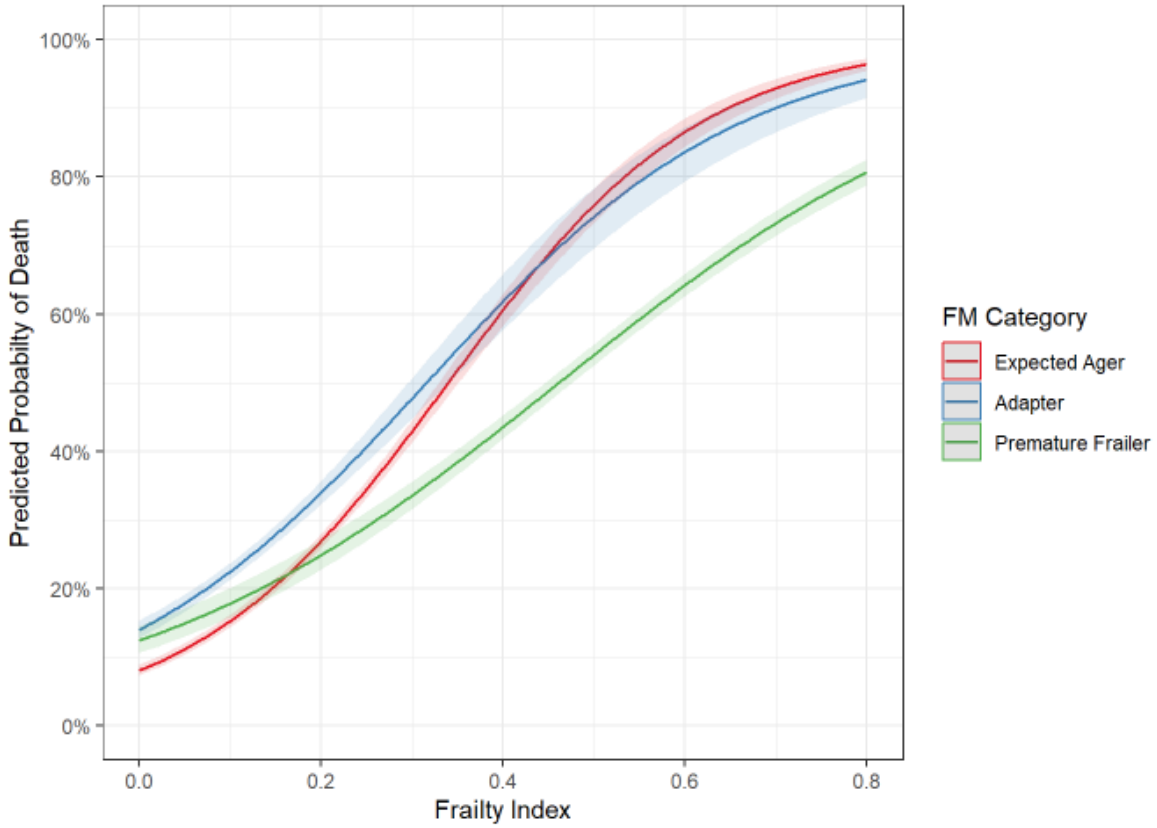


Figure 15. Interaction Effects of the Frailty Index and FM on Mortality

Visualization of the interaction between FI and FM on mortality (Model 4, Table 12).

This unexpected result prompted further investigation of differences between the FM categories. Further descriptive analysis determined that despite having similar levels of frailty, the adapters have significantly higher disease burden (Appendix F). Thus, it appears that disease burden may be driving this unexpected relationship. As an additional robustness check, I estimated a cross-sectional FM for wave 10 (n = 3,173) using a sample matching the characteristics of the Health ABC Study used by the original method authors, Wu et al. (43). The results of this sensitivity analysis are consistent with the main analysis, with the adapter group having the highest predicted probability of death in the age, sex, and frailty-adjusted model (Appendix F). As one final check I additionally separated FM into four categories by quartiles. Interestingly, only the bottom quartile deviates from the expected trend, suggesting that FM has a different association with adverse outcomes among those who are the most underestimated by the growth curve model (Appendix F).

6.1.5 Rate of Aging

Distribution of the RoA

The rate of aging has a right-skewed distribution, with a median of 0.005, a 25th percentile of 0.003, a 75th percentile of 0.012, a skewness of 1.612, and a kurtosis of 2.558 (Figure 16).

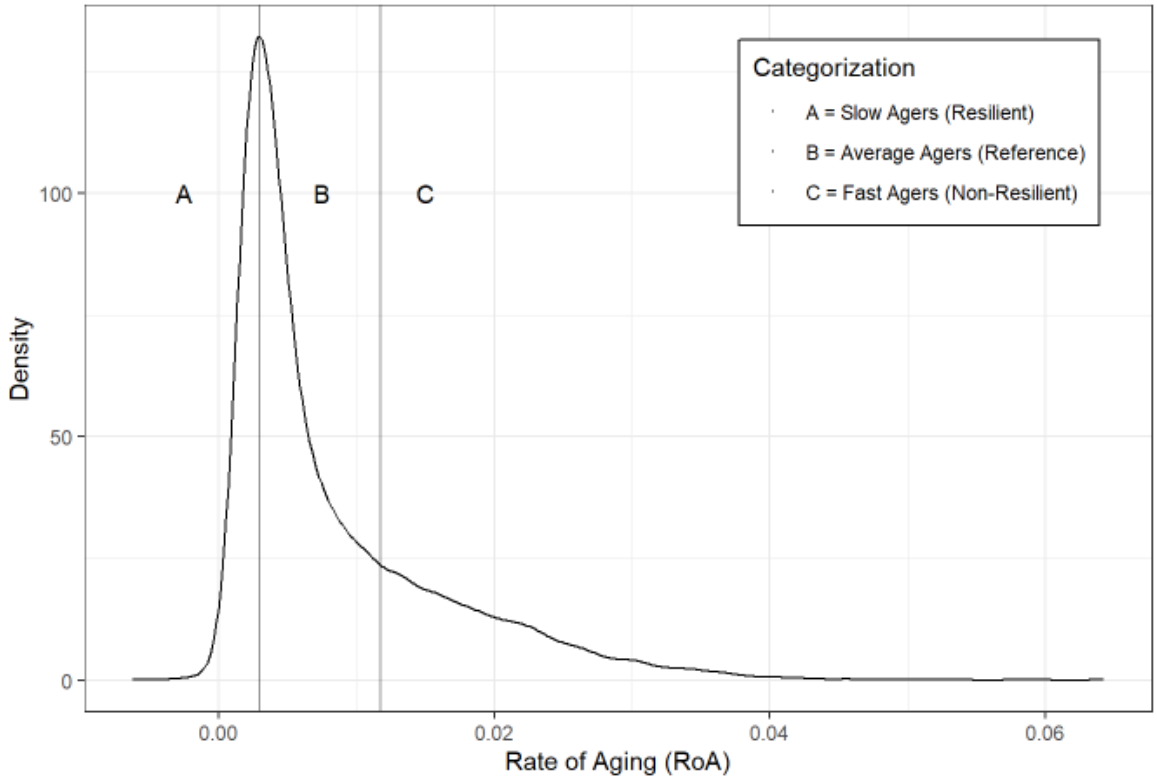


Figure 16. Distribution and Categorization of RoA (Mortality Sample)

Physical resilience categories are defined by the 25th and 75th percentiles: those below the 25th percentile are slow agers (resilient), those within the IQR are average agers (reference), and those above the 75th percentile are fast agers (non-resilient).

Comparison of RoA Categories

All variables are statistically significantly different across RoA categories using the Kruskal-Wallis test for continuous variables and Pearson's Chi-squared test for categorical variables ($p < 0.001$, Table 13). Trends across categories follow the expected patterns for all variables except FM adapters. Specifically, comparing the RoA slow agers (i.e., resilient) to the RoA fast agers (i.e., non-

resilient), the slow agers have the lowest median age (66 vs. 87), the lowest median FI at (0.067 vs. 0. 518), the lowest proportion of females (52% vs. 63%), deaths (13% vs. 84%), FM premature frailers (2.7% vs 54%), and DIOR-FI low stability (7.2% vs. 60%), and the highest proportion of DIOR-FI high stability (52% vs 3%). FM adapters deviate from this expected trend with the RoA average agers having the lowest proportion of FM adapters (23%), followed by the fast agers (25%), then the slow agers (28%).

Table 13. Comparison of RoA Categories (Mortality Sample)

	RoA Category			
Characteristic	**Slow Ager**, N = 6,936[1]	**Average Ager**, N = 13,872[1]	**Fast Ager**, N = 6,936[1]	**p-value**[2]
Age	66 (62, 71)	72 (62, 79)	87 (82, 91)	**<0.001**
Frailty Index	0.067 (0.046, 0.098)	0.183 (0.111, 0.283)	0.518 (0.338, 0.725)	**<0.001**
Sex				**<0.001**
Female	3,583 (52%)	7,902 (57%)	4,356 (63%)	
Male	3,353 (48%)	5,970 (43%)	2,580 (37%)	
2018 Vital Status				**<0.001**
Alive	6,017 (87%)	9,475 (68%)	1,098 (16%)	
Deceased	919 (13%)	4,397 (32%)	5,838 (84%)	
FM Category				**<0.001**
Adapter	1,967 (28%)	3,259 (23%)	1,710 (25%)	
Expected Ager	4,782 (69%)	7,599 (55%)	1,491 (21%)	
Premature Frailer	187 (2.7%)	3,014 (22%)	3,735 (54%)	
DIOR-FI Category				**<0.001**
High Stability	3,615 (52%)	3,116 (22%)	205 (3.0%)	
Average Stability	2,822 (41%)	8,461 (61%)	2,589 (37%)	
Low Stability	499 (7.2%)	2,295 (17%)	4,142 (60%)	

[1]Median (IQR Bounds); n (%)
[2]Kruskal-Wallis rank sum test; Pearson's Chi-squared test

The Effect of RoA on Mortality

Unadjusted logistic regression analysis shows that fast agers have 11.5 times greater odds of dying in the next two years than the average agers (95% CI = 10.6, 12.3), and the odds of dying in slow agers are 0.33 times those of the average agers (95% CI = 0.30, 0.36) (Table 14). These effects are reduced after adjusting for frailty, and then reduced further after adjusting for age and sex. After adjusting for age, sex, and frailty, fast agers have 3.14 times greater odds of dying than the average agers (95% CI = 2.83, 3.48), and the odds of slow agers are 0.55 times those of the average agers (95% CI = 0.50, 0.60).

Table 14. Logistic Regression Models for RoA and Mortality

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
RoA Category												
Average Ager	—	—		—	—		—	—		—	—	
Slow Ager	0.33	0.30, 0.36	**<0.001**	0.44	0.41, 0.48	**<0.001**	0.55	0.50, 0.60	**<0.001**	0.55	0.48, 0.62	**<0.001**
Fast Ager	11.5	10.6, 12.3	**<0.001**	6.35	5.83, 6.93	**<0.001**	3.14	2.83, 3.48	**<0.001**	6.80	5.66, 8.19	**<0.001**
Frailty Index				1.07	1.06, 1.08	**<0.001**	1.09	1.08, 1.09	**<0.001**	1.11	1.10, 1.12	**<0.001**
Age							1.05	1.04, 1.05	**<0.001**	1.05	1.04, 1.05	**<0.001**
Sex												
Female							—	—		—	—	
Male							1.84	1.74, 1.96	**<0.001**	1.87	1.76, 1.99	**<0.001**
RoA Category * Frailty Index												
Slow Ager * Frailty Index										1.04	1.01, 1.07	**0.005**
Fast Ager * Frailty Index										0.94	0.93, 0.95	**<0.001**

Model 1 is unadjusted. Model 2 is adjusted for frailty. Model 3 is further adjusted for age and sex. Model 4 further includes a frailty interaction Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference.

The interaction term between FI and RoA is statistically significant for both RoA categories (slow agers and faster agers, $p < 0.001$, Table 14). Figure 17 below visualizes this interaction: After adjusting for age and sex, the difference between RoA groups is reduced as the FI increases: the predicted probabilities for slow agers and average agers converges around an FI of 0.3, while the predicted probability for fast agers converges around an FI of 0.6.

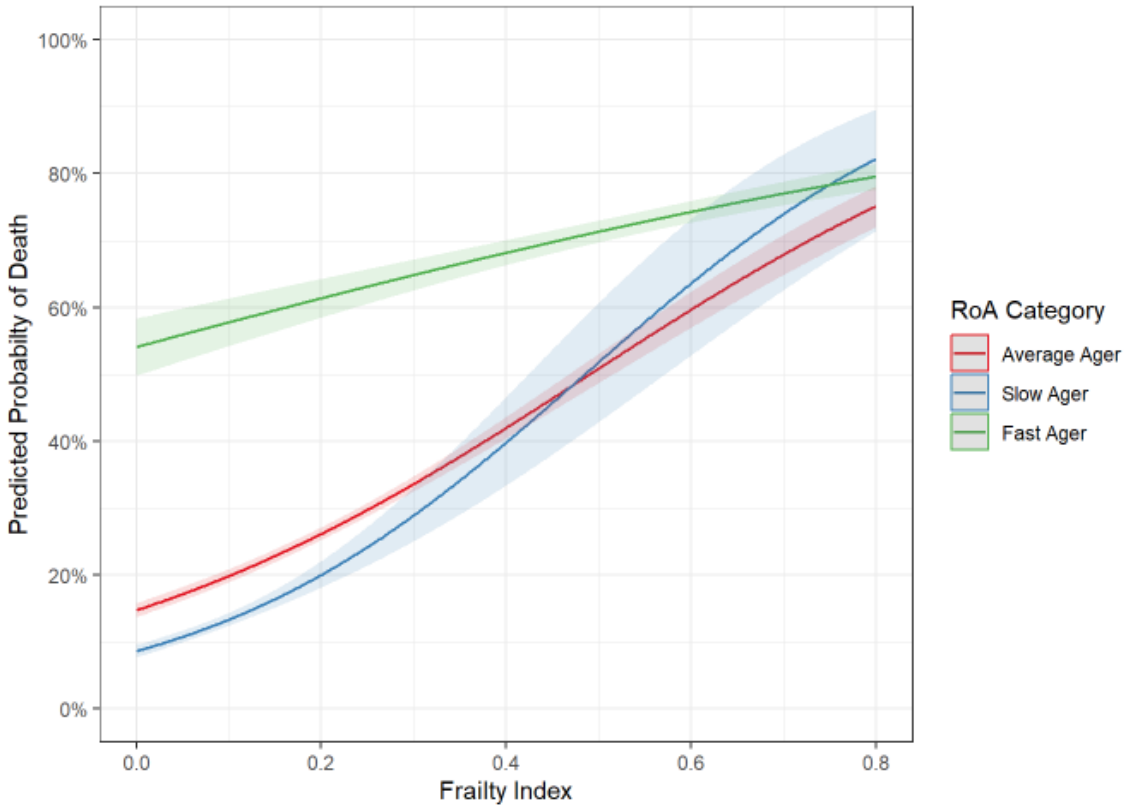


Figure 17. Interaction Effects of the Frailty Index and RoA on Mortality

Visualization of the interaction between FI and RoA on mortality (Model 4, Table 14).

6.1.6 DIOR-FI

Distribution of the DIOR-FI

The DIOR-FI has a right-skewed distribution, with a median of 0.046, a 25th percentile of 0.028, a 75th percentile of 0.077, a skewness of 1.701, and a kurtosis of 3.648 (Figure 18).

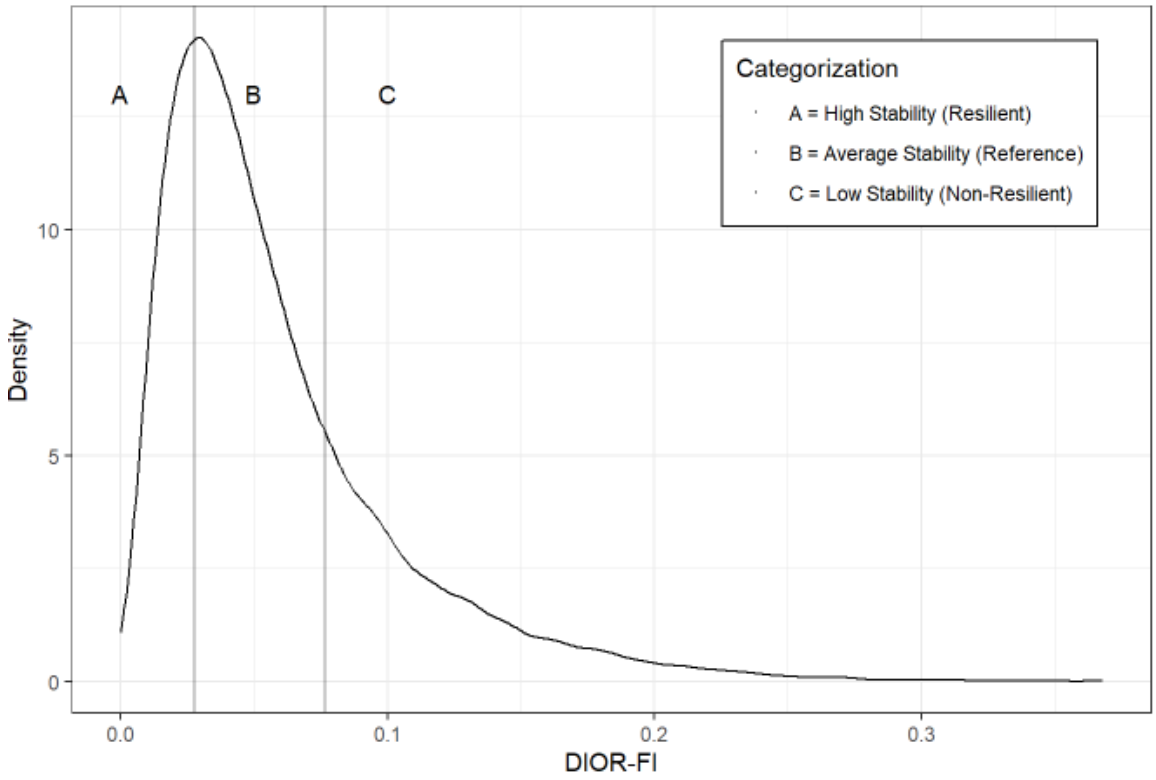


Figure 18. Distribution and Categorization of DIOR-FI (Mortality Sample)

Physical resilience categories are defined by the 25th and 75th percentile: those below the 25th percentile have high stability (resilient), those within the IQR have average stability (reference), and those above the 75th percentile have low stability (non-resilient).

Comparison of DIOR-FI Categories

All variables are statistically significantly different across DIOR-FI categories using the Kruskal-Wallis test for continuous variables and Pearson's Chi-squared test for categorical variables ($p < 0.001$, Table 15). Trends across categories follow the expected patterns for all variables except FM adapters. Specifically, comparing the DIOR-FI high stability (i.e., resilient) group to the low stability (i.e., non-resilient) group, the high stability group has the lowest median age (66 vs. 80), the lowest median FI at (0.073 vs. 0. 494), the lowest proportion of females (51% vs. 61%), deaths (17% vs 67%), RoA fast agers (3% vs 60%), and FM premature frailers (2.3% vs 64%), and the highest proportion of RoA slow agers (52% vs 7.2%). FM adapters deviate from this expected trend with the DIOR-FI average stability having a higher proportion than the high stability group (32% compared to 21%).

Table 15. Comparison of DIOR-FI Categories (Mortality Sample)

Characteristic	DIOR-FI Category			p-value[2]
	High Stability, N = 6,936[1]	**Average Stability**, N = 13,872[1]	**Low Stability**, N = 6,936[1]	
Age	66 (61, 74)	75 (66, 83)	80 (69, 88)	**<0.001**
Frailty Index	0.073 (0.043, 0.110)	0.183 (0.107, 0.294)	0.494 (0.305, 0.709)	**<0.001**
Sex				**<0.001**
Female	3,545 (51%)	8,060 (58%)	4,236 (61%)	
Male	3,391 (49%)	5,812 (42%)	2,700 (39%)	
2018 Vital Status				**<0.001**
Alive	5,747 (83%)	8,559 (62%)	2,284 (33%)	
Deceased	1,189 (17%)	5,313 (38%)	4,652 (67%)	
FM Category				**<0.001**
Adapter	1,439 (21%)	4,429 (32%)	1,068 (15%)	
Expected Ager	5,339 (77%)	7,117 (51%)	1,416 (20%)	
Premature Frailer	158 (2.3%)	2,326 (17%)	4,452 (64%)	
RoA Category				**<0.001**
Slow Ager	3,615 (52%)	2,822 (20%)	499 (7.2%)	
Average Ager	3,116 (45%)	8,461 (61%)	2,295 (33%)	
Fast Ager	205 (3.0%)	2,589 (19%)	4,142 (60%)	

[1]Median (IQR Bounds); n (%)

[2]Kruskal-Wallis rank sum test; Pearson's Chi-squared test

Unadjusted and Frailty-Adjusted Odds Ratios for Mortality

Unadjusted logistic regression analysis shows that the odds death in the high stability are 0.33 times those of the average stability group (95% CI = 0.31, 0.36), and the low stability group has 3.28 times greater odds of death compared to the average stability group (95% CI = 3.09, 3.49) (Table 16). The frailty-adjusted model shows an attenuation of these effects with an increase in the odds ratio for the high stability group (OR = 0.59, 95% CI = 0.55, 0.64), and a decrease in the odds ratio for the low stability group (OR = 1.12, 95% CI = 1.04, 1.21). Additional adjustment for age and sex further reduced the effect of the high stability group (OR = 0.77, 95% CI = 0.71, 0.84), and slightly increased the effect of the low stability group (OR = 1.26, 95% CI = 1.16, 1.37).

Table 16. Logistic Regression Models for DIOR-FI and Mortality

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
DIOR-FI Category												
Average Stability	—	—		—	—		—	—		—	—	
High Stability	0.33	0.31, 0.36	**<0.001**	0.59	0.55, 0.64	**<0.001**	0.77	0.71, 0.84	**<0.001**	0.45	0.39, 0.52	**<0.001**
Low Stability	3.28	3.09, 3.49	**<0.001**	1.12	1.04, 1.21	**0.004**	1.26	1.16, 1.37	**<0.001**	1.82	1.56, 2.11	**<0.001**
Frailty Index				1.15	1.14, 1.16	**<0.001**	1.12	1.12, 1.13	**<0.001**	1.13	1.12, 1.14	**<0.001**
Age							1.07	1.07, 1.08	**<0.001**	1.07	1.07, 1.08	**<0.001**
Sex												
Female							—	—		—	—	
Male							1.82	1.71, 1.93	**<0.001**	1.84	1.73, 1.95	**<0.001**
DIOR-FI Category * Frailty Index												
High Stability * Frailty Index										1.16	1.13, 1.19	**<0.001**
Low Stability * Frailty Index										0.97	0.96, 0.98	**<0.001**

Model 1 is unadjusted. Model 2 is adjusted for frailty. Model 3 is further adjusted for age and sex. Model 4 further includes a frailty interaction. Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference.

The interaction terms between FI and DIOR-FI are statistically significant for both DIOR-FI categories (high stability and low stability, $p < 0.001$, Table 16), and Figure 19 below visualizes this interaction. The high stability group has the lowest predicted probability of death below an FI of approximately 0.15, then it crosses over the other categories to have the highest predicted probability around an FI of 0.25. However, it should be noted that the 75th percentile of FI in the low stability category is 0.110, so this association at higher levels of FI may be driven by a small subset of highly stable but frail individuals.

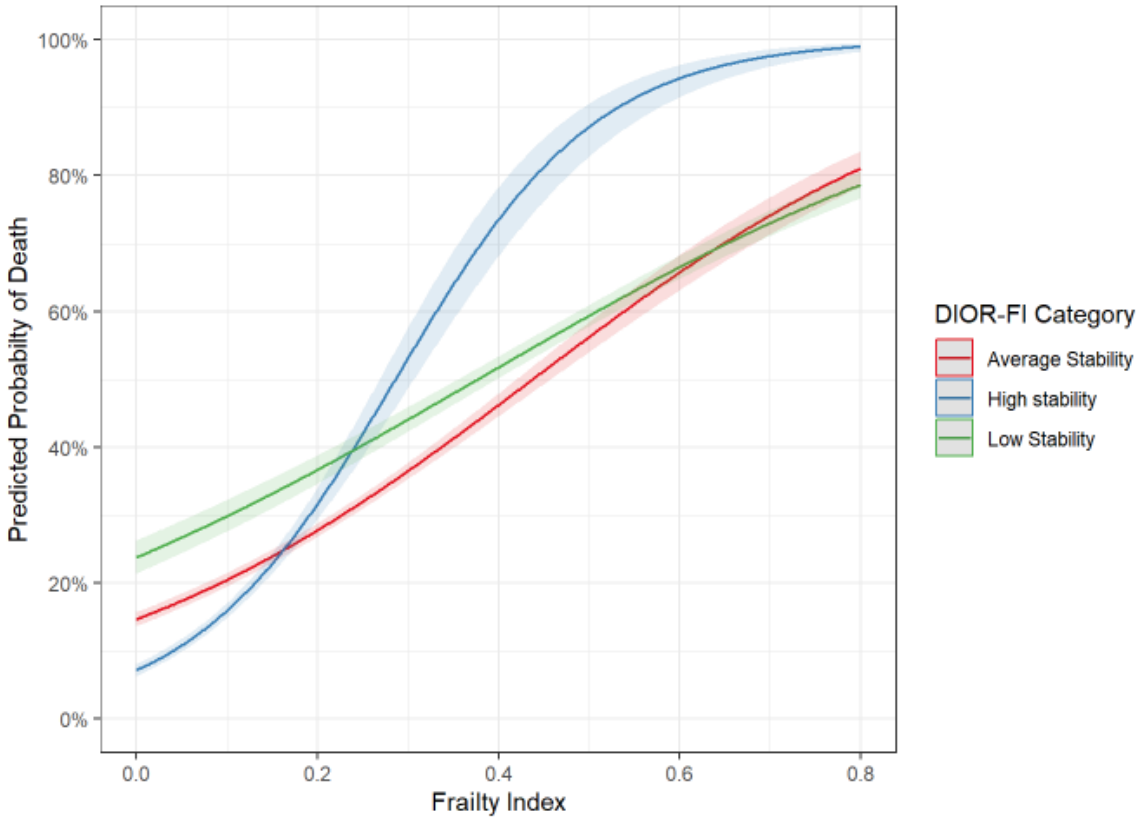


Figure 19. Interaction Effects of the Frailty Index and DIOR-FI on Mortality

Visualization of the interaction between FI and DIOR-FI on mortality (Model 4, Table 16).

6.1.7 Correlation and Agreement

Continuous variables show varying degrees of correlation using Pearson's correlation coefficient (Table 17). The resilience indicators have the highest correlations with FI, with coefficients ranging from 0.729 (DIOR-FI) to 0.785 (RoA). FM has the lowest correlation with the other resilience variables (0.332 and 0.555 for RoA and DIOR-FI, respectively).

Table 17. Correlation between FI and Resilience Indicators (Mortality Sample)

	FI	FM	RoA	DIOR-FI
FI	1.0			
FM	0.757	1.0		
RoA	0.785	0.332	1.0	
DIOR-FI	0.729	0.555	0.573	1.0

Cells represent Pearson's correlation coefficients for pairs of continuous variables.

The unweighted Cohen's Kappa for two raters shows slight to fair agreement between categorical resilience indicators, with the highest agreement between RoA and DIOR-FI (0.335), followed by RoA and FM (0.167), and lastly DIOR-FI and FM (0.150) (Table 18). A Kappa statistic of less than or equal to 0.20 is considered to represent slight agreement, while between 0.21 to 0.40 is considered to represent fair agreement (109).

Table 18. Agreement between Categorical Resilience Indicators (Mortality Sample)

	FM	RoA	DIOR-FI
FM	1.0		
RoA	0.167	1.0	
DIOR-FI	0.150	0.335	1.0

Cells represent unweighted Cohen's Kappa statistics for pairs of categorical variables.

6.1.8 Discrimination

The AUC for the unadjusted models, frailty-adjusted models, and frailty, age, and sex adjusted models are displayed in Table 19. Though interpretations vary, conventionally, AUC values below 0.6 are considered to be uninformative, above 0.6 but below 0.7 are considered to be poor to fair, while those between 0.7 and 0.8 are considered to be fair to good, and those above 0.8 are considered to be good to very good (7). The unadjusted models show that all the resilience variables have a lower AUC than the FI (0.778). However, the RoA is negligibly lower than the FI with an AUC of 0.777. A Delong's test for two correlated ROC curves confirmed that these two models did not have a significantly different AUC ($p = 0.5282$). The other two resilience variables, FM and DIOR-FI based on unadjusted models variables have a much lower AUC: 0.651 for FM and 0.695 for DIOR-FI. The largest increase when adding a single resilience indicator to the base FI model is when adding RoA, which increased the AUC from 0.778 to 0.810. Similarly, in the age, sex, and frailty adjusted models, the RoA shows the highest single variable increase over the base model (i.e., FI only). Adding RoA increases the AUC from 0.824 to 0.831. The final combined model that includes all resilience indicators and interactions shows a maximal AUC of 0.837, a modest increase from the base age-sex-adjusted FI model.

Table 19. Discrimination of Mortality Models

Model	AUC	Lower 95%	Upper 95%
Unadjusted Models			
FI Only	**0.778**	0.772	0.784
FM Only	0.651	0.645	0.657
RoA Only	0.777	0.771	0.782
DIOR Only	0.695	0.689	0.700
Frailty-Adjusted Models			
FI + FM	0.801	0.796	0.806
FI + RoA	**0.810**	0.805	0.816
FI + DIOR	0.779	0.773	0.784
Age, sex, and frailty-adjusted models			
FI Only	0.824	0.819	0.829
FI + FM	0.827	0.822	0.832
FI + FM Interaction	0.828	0.823	0.833
FI + RoA	0.831	0.826	0.836
FI + RoA Interaction	**0.832**	0.828	0.837
FI + DIOR	0.825	0.820	0.830
FI + DIOR Interaction	0.827	0.822	0.832
FI + FM + RoA + DIOR	0.834	0.830	0.839
FI + FM + RoA + DIOR + All FI Interactions	0.837	0.832	0.842

Highest AUC for each category of models is bolded, excluding the combined resilience models (bottom two rows). 95% confidence intervals are calculated using the Delong method.

6.1.9 Restricted Sample Sensitivity Analysis

In the restricted sample using only individuals with three consecutive FI measurements prior to death/survival, FM shows almost identical results, while RoA and DIOR-FI show large differences from the full models (Figure 20). In particular, the RoA is no longer the best predictor, and rather, is outperformed by the FI in unadjusted models, and FM in the adjusted models (Table 20). Appendix G provides the full results for the restricted sample.

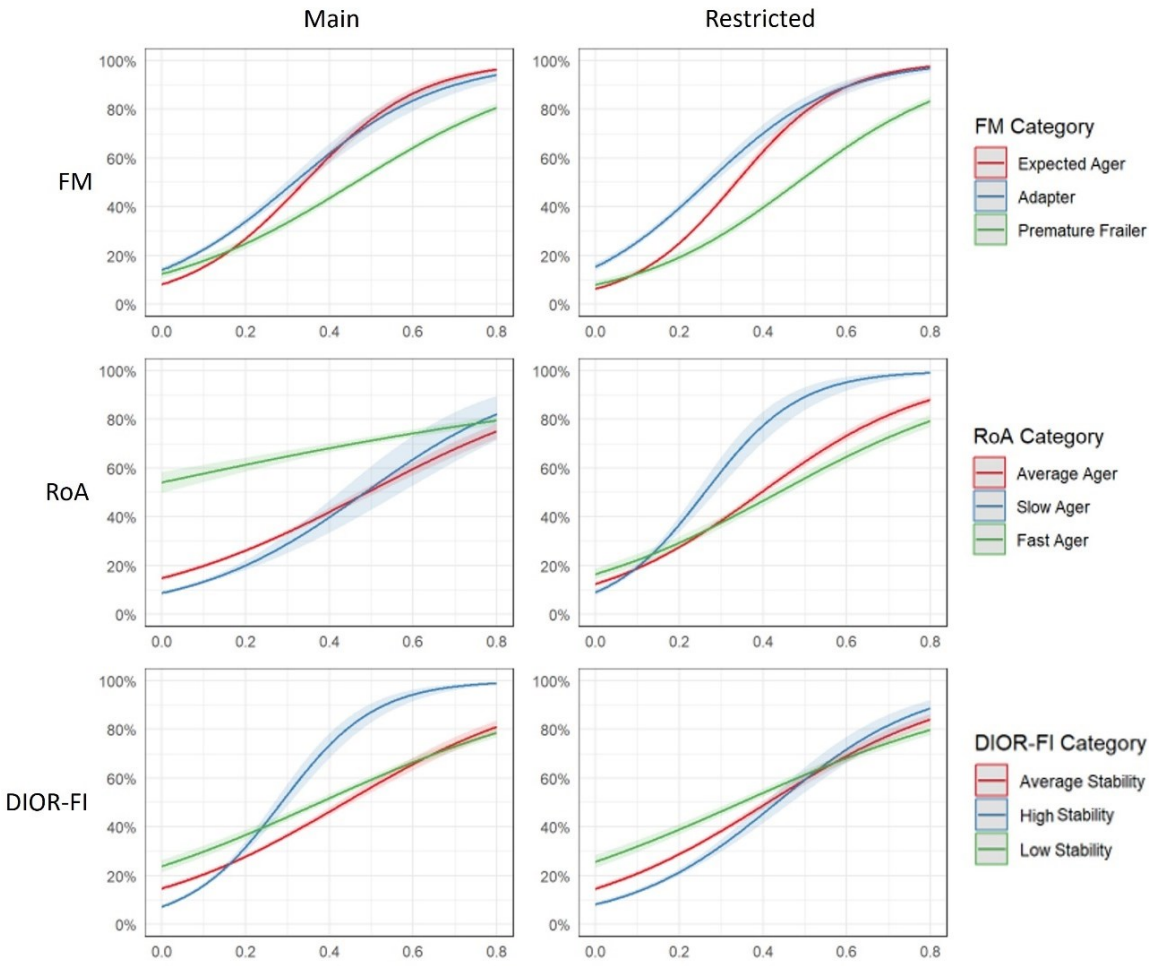


Figure 20. Interaction Effects in the Full Sample vs Restricted Sample (Mortality)

Comparison of the interaction models in the full sample (n=27,744) vs the restricted sample with only three consecutive final interviews per individual (n=23,644). The X-axis is the level of FI. The Y-axis is the predicted probability of mortality.

Table 20. Discrimination of Mortality Models in the Main and Restricted Sample

	Main Sample			Restricted Sample		
Model	AUC	Lower 95%	Upper 95%	AUC	Lower 95%	Upper 95%
Unadjusted Models						
FI Only	**0.778**	0.772	0.784	**0.782**	0.777	0.788
FM Only	0.651	0.645	0.657	0.653	0.647	0.660
RoA Only	0.777	0.771	0.782	0.737	0.731	0.743
DIOR Only	0.695	0.689	0.700	0.684	0.678	0.690
Frailty-Adjusted Models						
FI + FM	0.801	0.796	0.806	**0.812**	0.807	0.818
FI + RoA	**0.810**	0.805	0.816	0.799	0.793	0.804
FI + DIOR	0.779	0.773	0.784	0.786	0.780	0.792
Age and Frailty-Adjusted Models						
FI Only	0.824	0.819	0.829	0.828	0.822	0.833
FI + FM	0.827	0.822	0.832	**0.838**	0.832	0.843
FI + FM Interaction	0.828	0.823	0.833	0.838	0.833	0.843
FI + RoA	0.831	0.826	0.836	0.828	0.822	0.833
FI + RoA Interaction	**0.832**	0.828	0.837	0.828	0.823	0.834
FI + DIOR	0.825	0.820	0.830	0.831	0.826	0.836
FI + DIOR Interaction	0.827	0.822	0.832	0.832	0.827	0.837
FI + FM + RoA + DIOR	0.834	0.830	0.839	0.840	0.835	0.845
FI + FM + RoA + DIOR + All FI Interactions	0.837	0.832	0.842	0.841	0.836	0.846

Highest AUC for each category of models is bolded, excluding the combined resilience models (bottom two rows). 95% confidence intervals are calculated using the Delong method.

This difference prompted further investigation into the differences between the three resilience variables in each sample. In both samples, the RoA and DIOR-FI are estimated using all available data for each individual. In the main sample, this means that if a person is interviewed 10 times, they have 10 derivatives contributing to their RoA, and 10 residuals contributing to their DIOR-FI. However, if they are only observed 3 times (the minimum), they have only 3 derivatives and residuals contributing to their respective measures. In the sensitivity analysis with the restricted sample, RoA and DIOR-FI are still estimated using all past derivatives/residuals, but in this case, everyone only has 3. Since the rate of FI change and FI stability is related with age, including additional derivatives/residuals from younger ages (i.e., a longer follow up period) will decrease

the RoA/DIOR-FI. Thus, the results of RoA and DIOR-FI in the main sample are likely an artefact resulting from differential lengths of observation. The restricted sample sensitivity analysis clearly suggests that corrective measures must be taken, however, truncating the data to the last three interviews as done in the restricted sensitivity analysis results in significant loss of information, and subsequently, less accurate estimation of the growth curve model. Additionally, there are two potential contributing factors to the observed differences: the number of interviews included in the estimation of RoA and DIOR-FI, and whether the individual missed any of the final three interviews (i.e., the difference in sample size). Before simply removing the 15% of the sample with non-consecutive observations, I wanted to see if adjusting the estimation of RoA and DIOR-FI to the last three observations (regardless of whether they were consecutive), while still using all longitudinal information in the estimation of the growth curve model (rather than truncating the observations as done in the restricted sensitivity analysis above) could eliminate the differences between the two samples. In this comparison, the only difference between the two sets of results is that main analysis includes individuals that missed at least one of their last three interviews (but still have a minimum of three interviews total). The results are highly consistent in the two samples, suggesting that changing the estimation of RoA and DIOR-FI eliminates discrepancies between the samples, and that including the 15% of individuals with non-consecutive observations does not impact the results (Figure 21).

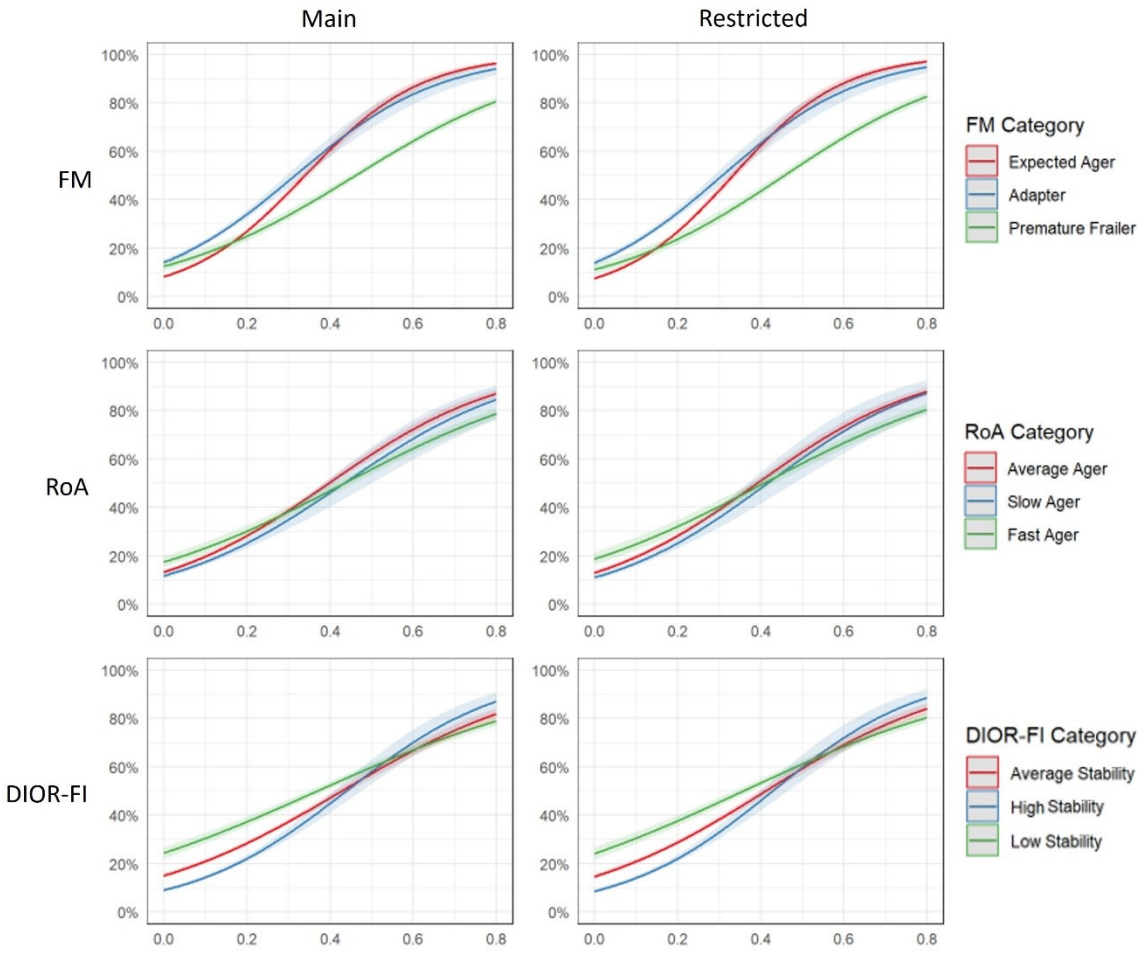


Figure 21. Alternative Estimation of RoA and DIOR-FI in Main vs. Restricted Sample

Comparison of the interaction models in the full sample (n=27,744) vs the restricted sample (n=23,644) using alternative estimation of RoA and DIOR-FI. Both samples use all longitudinal information for each individual in estimation of the growth curve model, but only the last three derivatives/residuals to estimate RoA and DIOR-FI. The X-axis is the level of FI. The Y-axis is the predicted probability of mortality.

Thus, to overcome this artefact and retain the full sample, I revised the estimation of the RoA and the DIOR-FI to use the last three derivatives/residuals only. The benefit to this approach is that 1) it retains the full sample size, 2) it uses all longitudinal information in the estimation of the growth curves, and 3) it makes individual estimates of RoA and DIOR-FI more comparable by using the same number of time points per individual (i.e., the last three). The main results are repeated below with the corrected estimation of RoA and DIOR-FI.

6.1.10 Sample Characteristics (corrected for last three observations only)

The corrected values for RoA and DIOR-FI are shaded in Table 21 below. As expected, the median RoA and DIOR FI are higher after correcting for the number of observations used to derive the estimates: RoA now has a median of 0.007 (IQR bounds = 0.003, 0.017), and DIOR-FI now has a median of 0.040 (IQR bounds = 0.021, 0.075).

Table 21. Sample Characteristics by 2018 Vital Status (corrected RoA and DIOR-FI)

		2018 Vital Status		
Characteristic	**Overall**, N = 27,744[1]	**Alive**, N = 16,590[1]	**Deceased**, N = 11,154[1]	**p-value**[2]
Age	74 (64, 82)	68 (62, 77)	81 (73, 88)	**<0.001**
Sex				**<0.001**
Female	15,841 (57%)	9,810 (59%)	6,031 (54%)	
Male	11,903 (43%)	6,780 (41%)	5,123 (46%)	
Frailty Index	0.183 (0.085, 0.360)	0.122 (0.067, 0.226)	0.331 (0.183, 0.567)	**<0.001**
FI Category				**<0.001**
Non-frail	8,340 (30%)	7,091 (43%)	1,249 (11%)	
Vulnerable	6,964 (25%)	4,852 (29%)	2,112 (19%)	
Frail	7,405 (27%)	3,591 (22%)	3,814 (34%)	
Most Frail	5,035 (18%)	1,056 (6.4%)	3,979 (36%)	
FM	-0.026 (-0.082, 0.068)	-0.036 (-0.080, 0.020)	0.012 (-0.086, 0.169)	**<0.001**
FM Category				**<0.001**
Adapter	6,936 (25%)	4,041 (24%)	2,895 (26%)	
Expected Ager	13,872 (50%)	9,949 (60%)	3,923 (35%)	
Premature Frailer	6,936 (25%)	2,600 (16%)	4,336 (39%)	
RoA	0.007 (0.003, 0.017)	0.004 (0.002, 0.009)	0.016 (0.008, 0.025)	**<0.001**
RoA Category				**<0.001**
Slow Ager	6,936 (25%)	6,077 (37%)	859 (7.7%)	
Average Ager	13,872 (50%)	8,930 (54%)	4,942 (44%)	
Fast Ager	6,936 (25%)	1,583 (9.5%)	5,353 (48%)	
DIOR-FI	0.040 (0.021, 0.075)	0.030 (0.016, 0.054)	0.062 (0.034, 0.108)	**<0.001**
DIOR-FI Category				**<0.001**
High Stability	6,936 (25%)	5,630 (34%)	1,306 (12%)	
Average Stability	13,872 (50%)	8,594 (52%)	5,278 (47%)	
Low Stability	6,936 (25%)	2,366 (14%)	4,570 (41%)	

[1]Median (IQR Bounds); n (%)

[2]Wilcoxon rank sum test; Pearson's Chi-squared test

Age, FI, and FM represent the final values prior to death or end of follow up (wave 13). RoA and DIOR-FI represent estimates using the last three interviews prior to death or end of follow up (wave 13).

6.1.11 Rate of Aging (corrected for last three observations only)

Distribution of the RoA

The corrected rate of aging has a right-skewed distribution, with a median of 0.007, a 25th percentile of 0.003, a 75th percentile of 0.017, a skewness of 1.259, and a kurtosis of 1.106 (Figure 22).

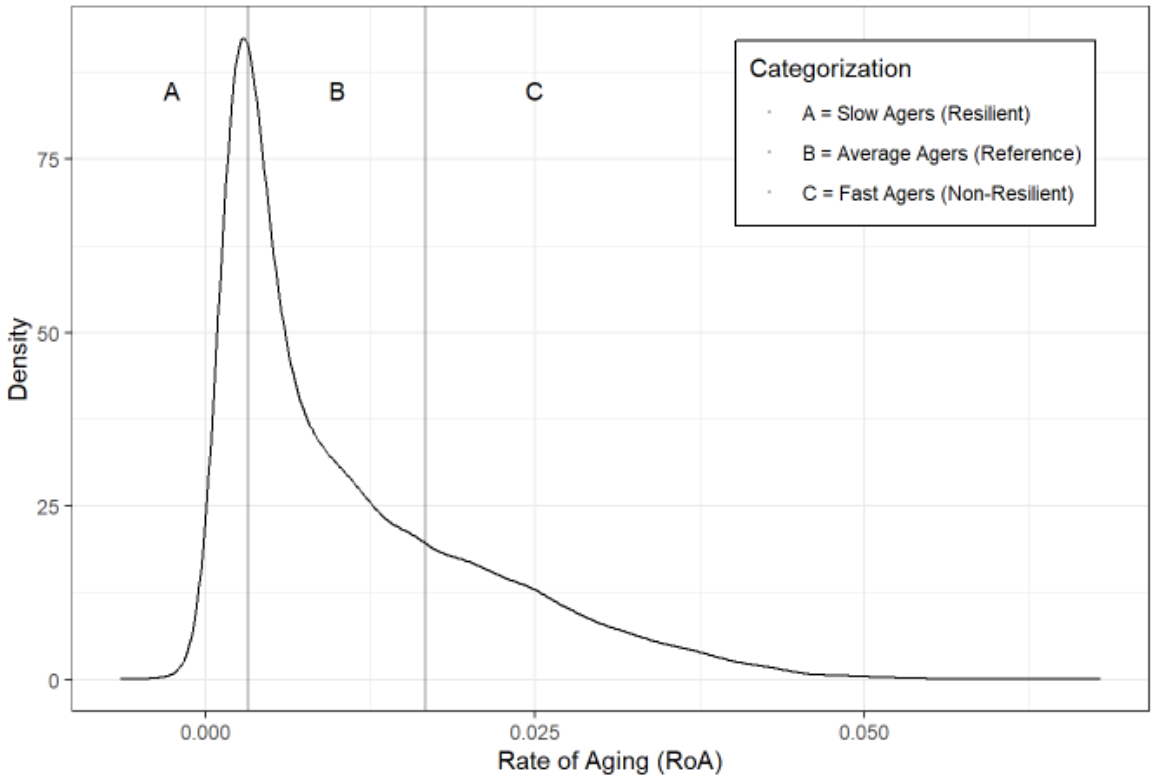


Figure 22. Distribution and Categorization of RoA (Mortality Sample – last three)

Physical resilience categories are defined by the 25th and 75th percentiles: those below the 25th percentile are slow agers (resilient), those within the IQR are average agers (reference), and those above the 75th percentile are fast agers (non-resilient). "Last three" refers to the second (and preferred) version of RoA which uses only the last three derivatives in its calculation to ensure all individuals have a comparable estimate regardless of how many interviews they completed.

Comparison of RoA Categories

All variables are statistically significantly different across RoA categories using the Kruskal-Wallis test for continuous variables and Pearson's Chi-squared test for categorical variables ($p < 0.001$, Table 22). Trends across categories follow the expected patterns for all variables except FM adapters. Specifically, comparing the RoA slow agers (i.e., resilient) to the RoA fast agers (i.e., non-resilient), the slow agers have the lowest median age (63 vs. 88), the lowest median FI at (0.073

vs. 0. 486), the lowest proportion of females (54% vs. 63%), deaths (12% vs. 77%), FM premature frailers (4.3% vs 48%), and DIOR-FI low stability (6.9% vs. 51%), and the highest proportion of DIOR-FI high stability (47% vs 6.9%). FM adapters deviate from this expected trend with the RoA slow agers having the lowest proportion (21%), followed by the average agers (25%), then the fast agers (30%).

Table 22. Comparison of RoA Categories (Mortality Sample – last three)

Characteristic	RoA Category			p-value[2]
	Slow Ager, N = 6,936[1]	**Average Ager**, N = 13,872[1]	**Fast Ager**, N = 6,936[1]	
Age	63 (61, 67)	74 (66, 79)	88 (84, 91)	**<0.001**
Frailty Index	0.073 (0.049, 0.110)	0.183 (0.104, 0.294)	0.486 (0.299, 0.713)	**<0.001**
Sex				**<0.001**
Female	3,722 (54%)	7,715 (56%)	4,404 (63%)	
Male	3,214 (46%)	6,157 (44%)	2,532 (37%)	
2018 Vital Status				**<0.001**
Alive	6,077 (88%)	8,930 (64%)	1,583 (23%)	
Deceased	859 (12%)	4,942 (36%)	5,353 (77%)	
FM Category				**<0.001**
Adapter	1,430 (21%)	3,412 (25%)	2,094 (30%)	
Expected Ager	5,211 (75%)	7,148 (52%)	1,513 (22%)	
Premature Frailer	295 (4.3%)	3,312 (24%)	3,329 (48%)	
DIOR-FI Category				**<0.001**
High Stability	3,244 (47%)	3,216 (23%)	476 (6.9%)	
Average Stability	3,211 (46%)	7,758 (56%)	2,903 (42%)	
Low Stability	481 (6.9%)	2,898 (21%)	3,557 (51%)	

[1]Median (IQR Bounds); n (%)

[2]Kruskal-Wallis rank sum test; Pearson's Chi-squared test

"Last three" refers to using the second (and preferred) versions of RoA and DIOR-FI, which use only the last three derivatives/residuals in their calculation to ensure all individuals have a comparable estimate regardless of how many interviews they completed.

The Effect of RoA on Mortality

Unadjusted logistic regression analysis shows that fast agers have 6.11 times greater odds of dying[17] than the average agers (95% CI = 5.72, 6.53), and the odds of dying in slow agers are 0.26 times those of the average agers (95% CI = 0.24, 0.28) (Table 23). These effects are reduced after

[17] This value is almost half of the previous, uncorrected OR of 11.5 shown in Table 14.

adjusting for frailty, and then reduced further after adjusting for age and sex, with the fast agers becoming insignificant (p = 0.4).

Table 23. Logistic Regression Models for RoA and Mortality (last three)

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
RoA Category												
Average Ager	—	—		—	—		—	—		—	—	
Slow Ager	0.26	0.24, 0.28	**<0.001**	0.39	0.36, 0.42	**<0.001**	0.80	0.72, 0.88	**<0.001**	0.87	0.76, 1.00	**0.047**
Fast Ager	6.11	5.72, 6.53	**<0.001**	2.95	2.74, 3.18	**<0.001**	0.96	0.86, 1.07	0.4	1.38	1.18, 1.62	**<0.001**
Frailty Index				1.10	1.10, 1.11	**<0.001**	1.14	1.13, 1.14	**<0.001**	1.15	1.14, 1.16	**<0.001**
Age							1.07	1.07, 1.08	**<0.001**	1.07	1.07, 1.08	**<0.001**
Sex												
Female							—	—		—	—	
Male							1.81	1.70, 1.92	**<0.001**	1.82	1.72, 1.93	**<0.001**
RoA Category * Frailty Index												
Slow Ager * Frailty Index										1.00	0.97, 1.02	0.8
Fast Ager * Frailty Index										0.97	0.96, 0.98	**<0.001**

Model 1 is unadjusted. Model 2 is adjusted for frailty. Model 3 is further adjusted for age and sex. Model 4 further includes a frailty interaction Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference. "Last three" refers to the second (and preferred) version of RoA which uses only the last three derivatives in its calculation to ensure all individuals have a comparable estimate regardless of how many interviews they completed.

The interaction term between FI and RoA is statistically significant for fast agers (p < 0.001, Table 23), and the main effect of fast agers becomes significant again. Figure 23 below visualizes this interaction. This shows that after adjusting for age and sex, there is a small difference between groups at low levels of FI, but the groups quickly begin to overlap as FI increases.

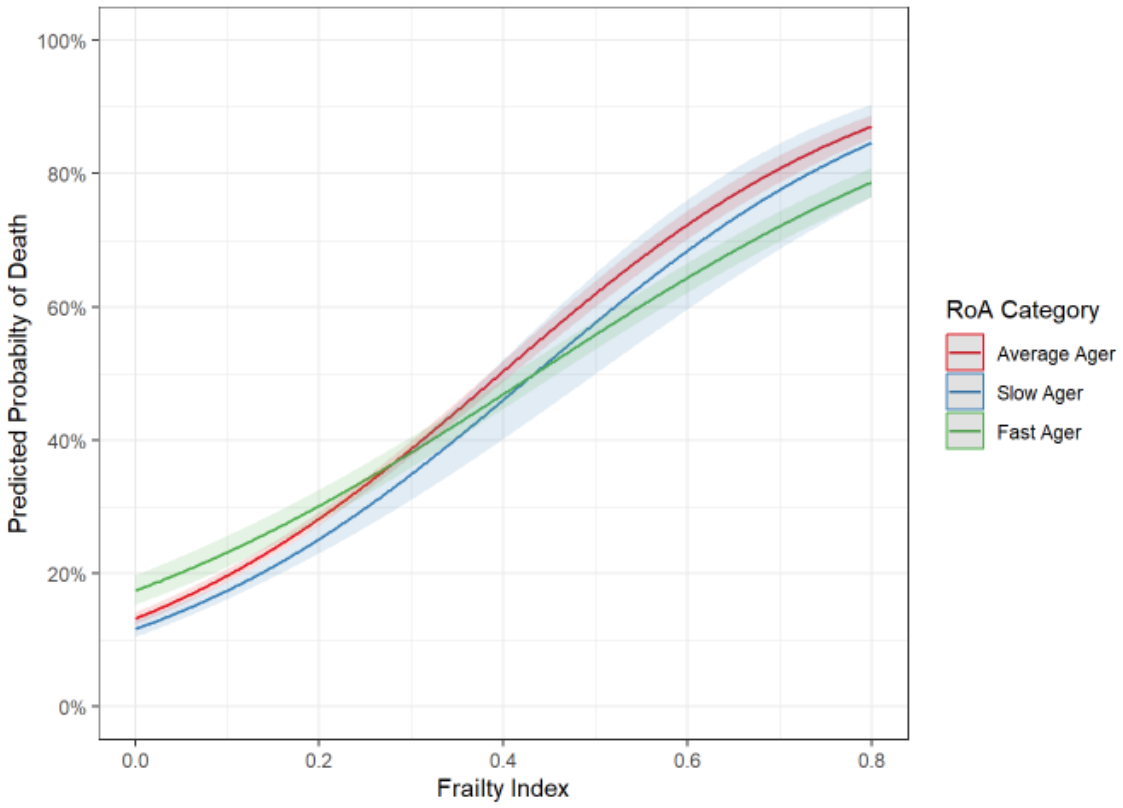


Figure 23. Interaction Effects of the Frailty Index and RoA on Mortality (last three)

Visualization of the interaction between FI and RoA on mortality (Model 4, Table 23). "Last three" refers to the second (and preferred) version of RoA which uses only the last three derivatives in its calculation to ensure all individuals have a comparable estimate regardless of how many interviews they completed.

6.1.12 DIOR-FI (corrected for last three observations only)

Distribution of the DIOR-FI

The DIOR-FI has a right-skewed distribution, with a median of 0.040, a 25th percentile of 0.021, a 75th percentile of 0.075, a skewness of 2.05, and a kurtosis of 4.977 (Figure 24).

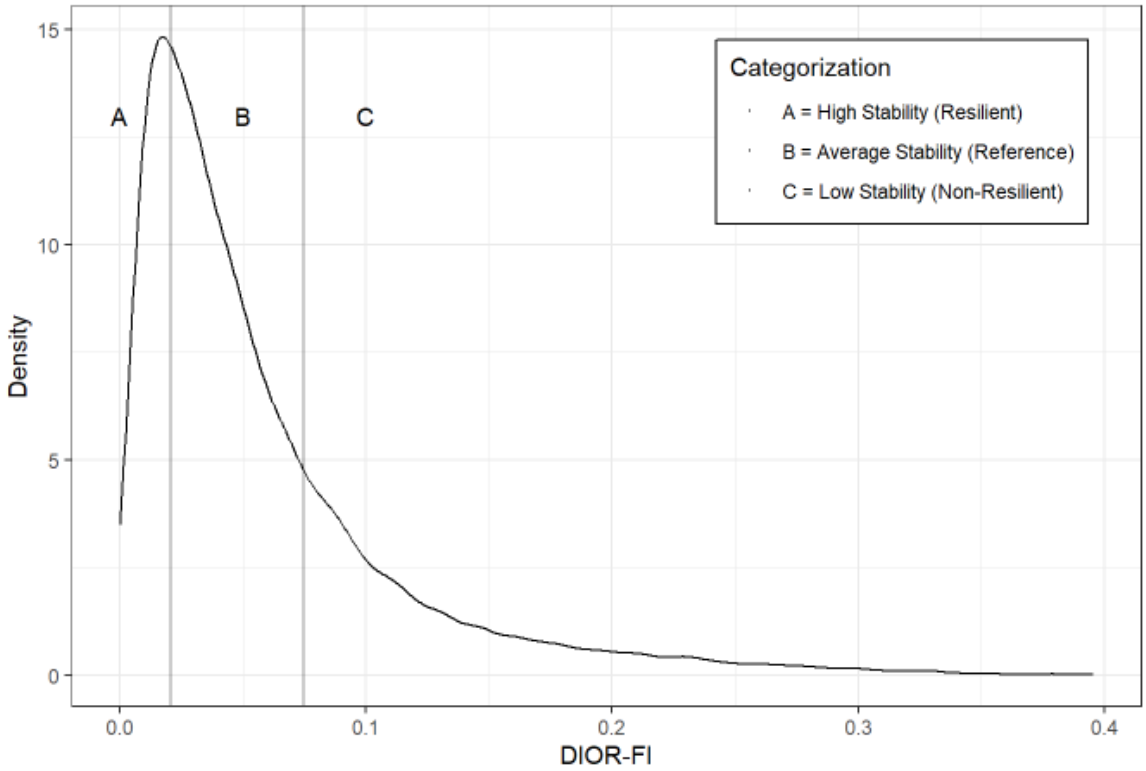


Figure 24. Distribution and Categorization of DIOR-FI (Mortality Sample - last three)

Physical resilience categories are defined by the 25th and 75th percentile: those below the 25th percentile have high stability (resilient), those within the IQR have average stability (reference), and those above the 75th percentile have low stability (non-resilient). "Last three" refers to the second (and preferred) version of DIOR-FI which uses only the last three residuals in its calculation to ensure all individuals have a comparable estimate regardless of how many interviews they completed.

Comparison of DIOR-FI Categories

All variables are statistically significantly different across DIOR-FI categories using the Kruskal-Wallis test for continuous variables and Pearson's Chi-squared test for categorical variables ($p < 0.001$, Table 24). Trends across categories follow the expected patterns for all variables except FM adapters. Specifically, comparing the DIOR-FI high stability (i.e., resilient) group to the low stability (i.e., non-resilient) group, the high stability group has the lowest median age (68 vs. 80), the lowest median FI at (0.079 vs. 0. 451), the lowest proportion of females (54% vs. 60%), deaths (19% vs 66%), RoA fast agers (6.9% vs 51%), and FM premature frailers (5% vs 60%), and the highest proportion of RoA slow agers (47% vs 6.9%). FM adapters deviate from this expected trend with the

DIOR-FI average stability having a higher proportion than the high stability group (29% compared to 24%).

Table 24. Comparison of DIOR-FI Categories (Mortality Sample - last three)

Characteristic	DIOR-FI Category			p-value[2]
	High Stability, N = 6,936[1]	**Average Stability**, N = 13,872[1]	**Low Stability**, N = 6,936[1]	
Age	68 (61, 76)	74 (65, 82)	80 (69, 88)	**<0.001**
Frailty Index	0.079 (0.049, 0.138)	0.177 (0.098, 0.293)	0.451 (0.274, 0.671)	**<0.001**
Sex				**<0.001**
Female	3,711 (54%)	7,963 (57%)	4,167 (60%)	
Male	3,225 (46%)	5,909 (43%)	2,769 (40%)	
2018 Vital Status				**<0.001**
Alive	5,630 (81%)	8,594 (62%)	2,366 (34%)	
Deceased	1,306 (19%)	5,278 (38%)	4,570 (66%)	
FM Category				**<0.001**
Adapter	1,649 (24%)	4,091 (29%)	1,196 (17%)	
Expected Ager	4,940 (71%)	7,321 (53%)	1,611 (23%)	
Premature Frailer	347 (5.0%)	2,460 (18%)	4,129 (60%)	
RoA Category				**<0.001**
Slow Ager	3,244 (47%)	3,211 (23%)	481 (6.9%)	
Average Ager	3,216 (46%)	7,758 (56%)	2,898 (42%)	
Fast Ager	476 (6.9%)	2,903 (21%)	3,557 (51%)	

[1]Median (IQR Bounds); n (%)

[2]Kruskal-Wallis rank sum test; Pearson's Chi-squared test

"Last three" refers to using the second (and preferred) versions of RoA and DIOR-FI, which use only the last three derivatives/residuals in their calculation to ensure all individuals have a comparable estimate regardless of how many interviews they completed.

Unadjusted and Frailty-Adjusted Odds Ratios for Mortality

Unadjusted logistic regression analysis shows that the odds of dying in the high stability group are 0.38 times those of the average stability group (95% CI = 0.35, 0.40), and the low stability group has 3.15 times greater odds of dying compared to the average stability group (95% CI = 2.96, 3.34) (Table 25). The frailty-adjusted model shows an attenuation of these effects with an increase in the odds ratio for the high stability group (OR = 0.58, 95% CI = 0.54, 0.63), and a decrease in the odds ratio for the low stability group (OR = 1.23, 95% CI = 1.15, 1.33). Further adjusting for age and sex further reduced the effect of the high stability group (OR = 0.64, 95% CI = 0.60, 0.71), and slightly increased the effect of the low stability group (OR = 1.29, 95% CI = 1.20, 1.40).

Table 25. Logistic Regression Models for DIOR-FI and Mortality (last three)

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
DIOR-FI Category												
Average Stability	—	—		—	—		—	—		—	—	
High Stability	0.38	0.35, 0.40	**<0.001**	0.58	0.54, 0.63	**<0.001**	0.65	0.60, 0.71	**<0.001**	0.56	0.49, 0.63	**<0.001**
Low Stability	3.15	2.96, 3.34	**<0.001**	1.23	1.15, 1.33	**<0.001**	1.29	1.20, 1.40	**<0.001**	1.83	1.59, 2.11	**<0.001**
Frailty Index				1.15	1.14, 1.16	**<0.001**	1.12	1.11, 1.13	**<0.001**	1.13	1.12, 1.14	**<0.001**
Age							1.07	1.07, 1.08	**<0.001**	1.07	1.07, 1.08	**<0.001**
Sex												
Female							—	—		—	—	
Male							1.81	1.71, 1.92	**<0.001**	1.83	1.73, 1.95	**<0.001**
DIOR-FI Category * Frailty Index												
High Stability * Frailty Index										1.04	1.02, 1.06	**<0.001**
Low Stability * Frailty Index										0.97	0.96, 0.98	**<0.001**

Model 1 is unadjusted. Model 2 is adjusted for frailty. Model 3 is further adjusted for age and sex. Model 4 further includes a frailty interaction. Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference. “Last three” refers to the second (and preferred) version of DIOR-FI which uses only the last three residuals in its calculation to ensure all individuals have a comparable estimate regardless of how many interviews they completed.

The interaction terms between FI and DIOR-FI are statistically significant for both DIOR-FI categories (high stability and low stability, $p < 0.001$, Table 25), and Figure 25 below visualizes this interaction. The high stability group has the lowest predicted probability of death below an FI of approximately 0.35, where it becomes similar to the reference category (average agers). The low stability category has the highest predicted probability of death until approximately an FI of 0.5.

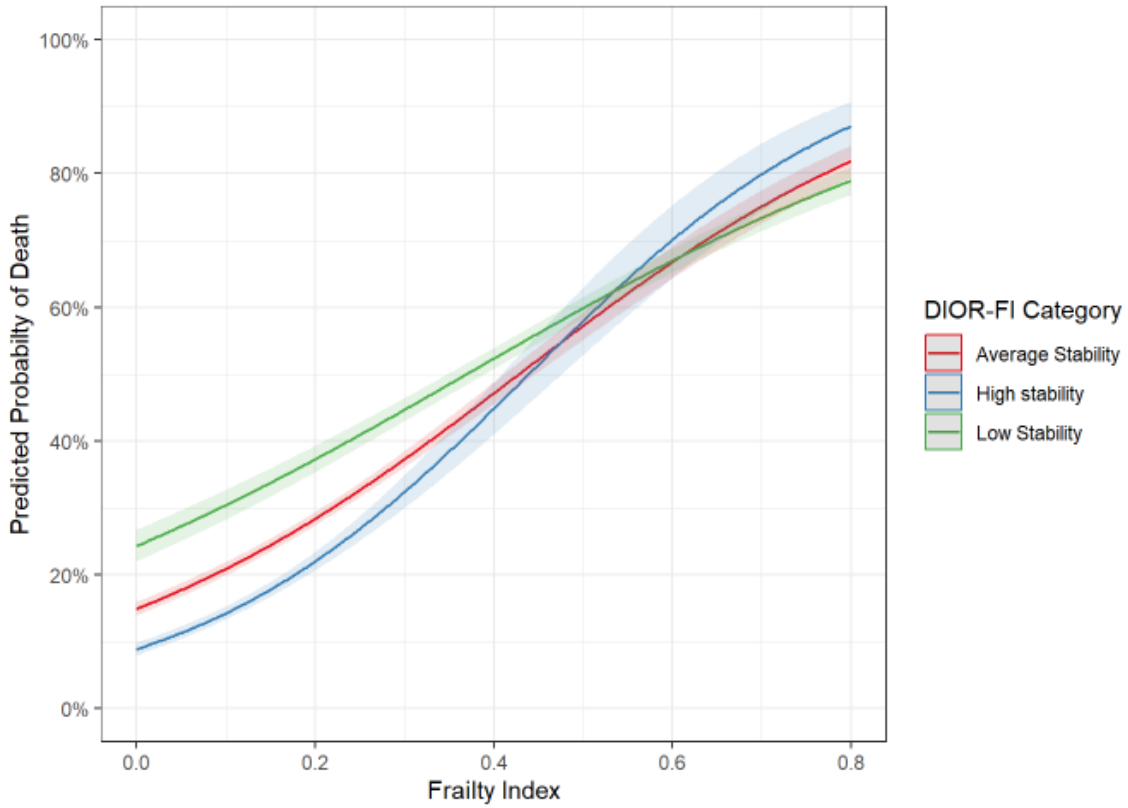


Figure 25. Interaction Effects of the Frailty Index and DIOR-FI on Mortality (last three)

Visualization of the interaction between FI and DIOR-FI on mortality (Model 4, Table 25). "Last three" refers to the second (and preferred) version of DIOR-FI which uses only the last three residuals in its calculation to ensure all individuals have a comparable estimate regardless of how many interviews they completed.

6.1.13 Correlation and Agreement (corrected for last 3 observations only)

Continuous variables show varying degrees of correlation using Pearson's correlation coefficient (Table 26). The resilience indicators have the highest correlations with FI, with coefficients ranging from 0.757 (FM) to 0.658 (DIOR-FI). FM has the lowest correlation with the other resilience variables (0.212 and 0.519 for RoA and DIOR-FI, respectively).

Table 26. Correlation between FI and Resilience Indicators (Mortality Sample – last three)

	FI	FM	RoA	$DIOR\text{-}FI_i$
FI	1.0			
FM	0.757	1.0		
RoA	0.723	0.212	1.0	
DIOR-FI	0.658	0.519	0.454	1.0

Cells represent Pearson's correlation coefficients for pairs of continuous variables. "Last three" refers to using the second (and preferred) versions of RoA and DIOR-FI, which use only the last three derivatives/residuals in their calculation to ensure all individuals have a comparable estimate regardless of how many interviews they completed.

The unweighted Cohen's Kappa for two raters shows only slight to fair agreement between categorical resilience indicators, with the highest agreement between RoA and DIOR-FI (0.240), followed by FM and DIOR-FI (0.155), and lastly RoA and FM (0.087) (Table 27). A Kappa statistic of less than or equal to 0.20 is considered to represent slight agreement, while between 0.21 to 0.40 is considered to represent fair agreement (109). A Kappa statistic of less than or equal to 0.20 is considered to represent slight agreement, while between 0.21 to 0.40 is considered to represent fair agreement (109).

Table 27. Agreement between Categorical Resilience Indicators (Mortality Sample – last three)

	FM	RoA	DIOR-FI
FM	1.0		
RoA	0.087	1.0	
DIOR-FI	0.155	0.240	1.0

Cells represent unweighted Cohen's Kappa statistics for pairs of categorical variables. "Last three" refers to using the second (and preferred) versions of RoA and DIOR-FI, which use only the last three derivatives/residuals in their calculation to ensure all individuals have a comparable estimate regardless of how many interviews they completed.

6.1.14 Discrimination (corrected for last three observations only)

Using the corrected RoA and DIOR-FI, FI is the best individual predictor (AUC = 0.778), followed by RoA (AUC = 0.753), DIOR-FI (AUC = 0.684), then FM (AUC = 0.651) (Table 28). After adjusting for frailty, RoA and FM perform similarly (0.802 and 0.801, respectively), with DIOR-FI being the worst (AUC = 0.780). After further adjusting for age and sex, all three perform similarly with very

small differences, though RoA (AUC = 0.824) performs the worst compared to the other two (AUC = 0.828). The maximum increase in AUC with all resilience indicators included is only 0.007.

Table 28. Discrimination of Mortality Models (last three)

Model	AUC	Lower 95%	Upper 95%
Unadjusted Models			
FI Only	**0.778**	0.772	0.784
FM Only	0.651	0.645	0.657
RoA Only	0.753	0.747	0.758
DIOR Only	0.684	0.678	0.689
Frailty-Adjusted Models			
FI + FM	0.801	0.796	0.806
FI + RoA	**0.802**	0.796	0.807
FI + DIOR	0.780	0.775	0.786
Age, Sex, and Frailty-Adjusted Models			
FI Only	0.824	0.819	0.829
FI + FM	0.827	0.822	0.832
FI + FM Interaction	**0.828**	0.823	0.833
FI + RoA	0.824	0.819	0.829
FI + RoA Interaction	0.825	0.820	0.829
FI + DIOR	0.827	0.822	0.832
FI + DIOR Interaction	**0.828**	0.823	0.832
FI + FM + RoA + DIOR	0.830	0.825	0.834
FI + FM + RoA + DIOR + All FI Interactions	0.831	0.826	0.836

Highest AUC for each category of models is bolded, excluding the combined resilience models (bottom two rows). 95% confidence intervals are calculated using the Delong method. “Last three” refers to using the second (and preferred) versions of RoA and DIOR-FI, which use only the last three derivatives/residuals in their calculation to ensure all individuals have a comparable estimate regardless of how many interviews they completed.

6.1.15 **Additional Sensitivity Analyses**

All sensitivity analyses presented here use the preferred “last three” estimation of RoA and DIOR-FI, which ensures all individuals have a comparable estimate regardless of how many interviews they completed.

Age-Stratified Results

The age-stratified results (interaction effects between FI and each resilience variable in the model predicting mortality, and AUC) revealed heterogeneous effects across all age groups for all three resilience indicators (Figure 26).

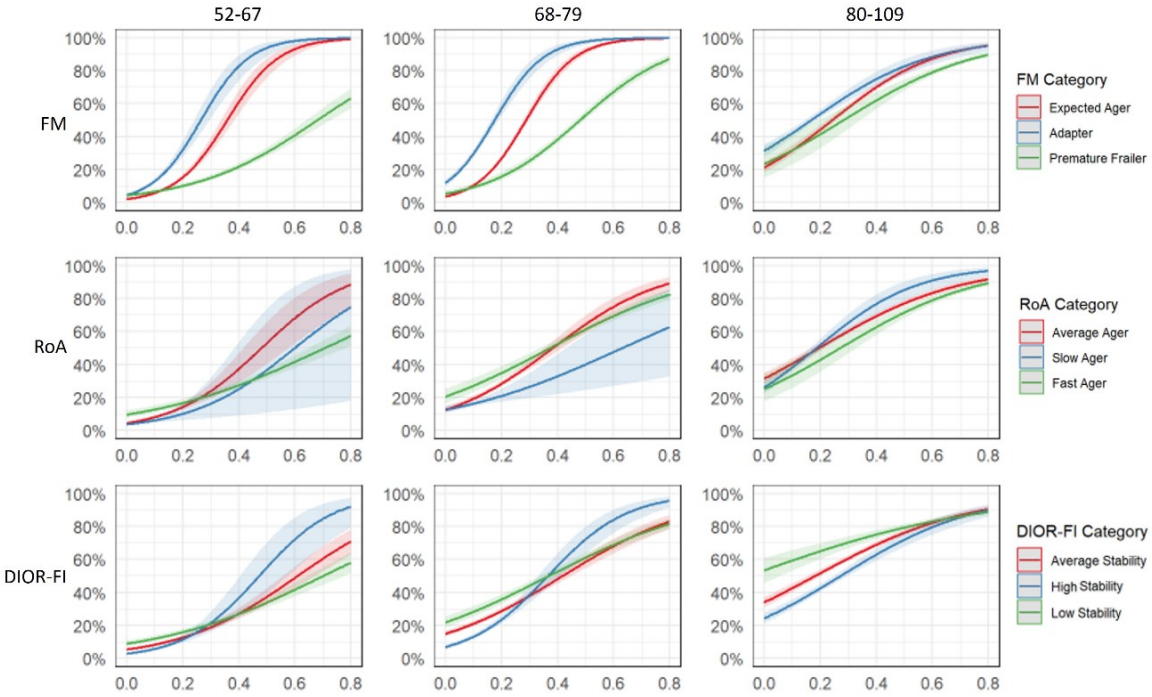


Figure 26. Age-Stratified Interaction Effects on Mortality

Pictured from top to bottom: FM, RoA, DIOR-FI. Blue lines indicate the resilient groups, red lines indicate the reference groups, and green lines indicate the non-resilient/fast aging groups. Sample size is 9,150 for ages 54-67, 9,002 for ages 68-79, and 9,592 for ages 80-109. The X-axis is the level of FI. The Y-axis is the predicted probability of mortality.

The unadjusted effects of RoA are quite similar across all age groups, consistently being the best individual predictor after FI (Table 29). In the age and sex-adjusted models the interaction differs across age groups, but the effects are quite small (middle row, Figure 26). After adjusting for age and sex, RoA becomes the worst predictor of the three resilience indicators regardless of the age group (Table 29). DIOR-FI shows more clear separation between resilience categories in the oldest age group: the convergence in predicted probabilities happens at a higher FI, with no crossover as seen in the younger groups (bottom row, Figure 26). FM shows the opposite of this with more

separation between resilience categories in the youngest and middle age groups (top row, Figure 26).

Table 29. Discrimination of Age-Stratified Models

	52-67			68-79			80-109		
	AUC	Lower 95%	Upper 95%	AUC	Lower 95%	Upper 95%	AUC	Lower 95%	Upper 95%
Unadjusted Models									
FI Only	**0.727**	0.712	0.742	**0.717**	0.706	0.728	**0.737**	0.727	0.747
FM Only	0.625	0.610	0.640	0.616	0.605	0.628	0.603	0.592	0.613
RoA Only	0.686	0.672	0.700	0.673	0.662	0.684	0.697	0.687	0.707
DIOR Only	0.642	0.628	0.656	0.650	0.640	0.661	0.659	0.648	0.669
Frailty-Adjusted Models									
FI + FM	**0.753**	0.739	0.767	**0.754**	0.744	0.765	**0.749**	0.739	0.759
FI + RoA	0.727	0.712	0.742	0.718	0.707	0.729	0.743	0.733	0.753
FI + DIOR	0.728	0.714	0.742	0.724	0.713	0.734	0.743	0.733	0.754
Age and Sex Adjusted Models									
FI Only	0.744	0.730	0.758	0.735	0.725	0.746	0.760	0.750	0.770
FI + FM	0.765	0.751	0.779	**0.774**	0.764	0.784	0.763	0.753	0.773
FI + FM Interaction	**0.772**	0.759	0.786	**0.774**	0.764	0.784	0.763	0.753	0.773
FI + RoA	0.745	0.731	0.759	0.735	0.724	0.746	0.760	0.751	0.770
FI + RoA Interaction	0.747	0.733	0.761	0.736	0.725	0.747	0.761	0.751	0.770
FI + DIOR	0.748	0.734	0.762	0.741	0.731	0.752	0.763	0.754	0.773
FI + DIOR Interaction	0.751	0.737	0.765	0.744	0.733	0.754	**0.764**	0.755	0.774
FI + FM + RoA + DIOR	0.768	0.754	0.782	0.776	0.766	0.786	0.766	0.757	0.776
FI + FM + RoA + DIOR + All FI Interactions	0.774	0.761	0.788	0.780	0.771	0.790	0.767	0.758	0.777

AUCs for each age stratified model. Highest AUC for each category of models is bolded, excluding the combined resilience models (bottom two rows). 95% confidence intervals are calculated using the Delong method.

Interestingly, the youngest and middle age group show larger increases in AUC by adding the resilience variables compared to the oldest age group and the main analysis including all ages. The maximum increase by adding all resilience variables (i.e., the final combined model, bottom row in Table 29) is 0.030 in the 52-67 age group, and 0.045 in the 68-79 group. Appendix H has full result tables for each model.

Sex-Stratified Results

In the sex-stratified results, males (n = 11,903) and females (n = 15,841) show similar overall trends with some minor differences. Females overall have a lower predicted probability of death, and more separation between resilience categories (Figure 27). Notably, the DIOR-FI groups in males converge at a lower FI and display a crossover effect where the highest stability group has the lowest predicted probability at low FI, and the highest predicted probability at high FI (bottom row, Figure 27). This effect is not seen in females.

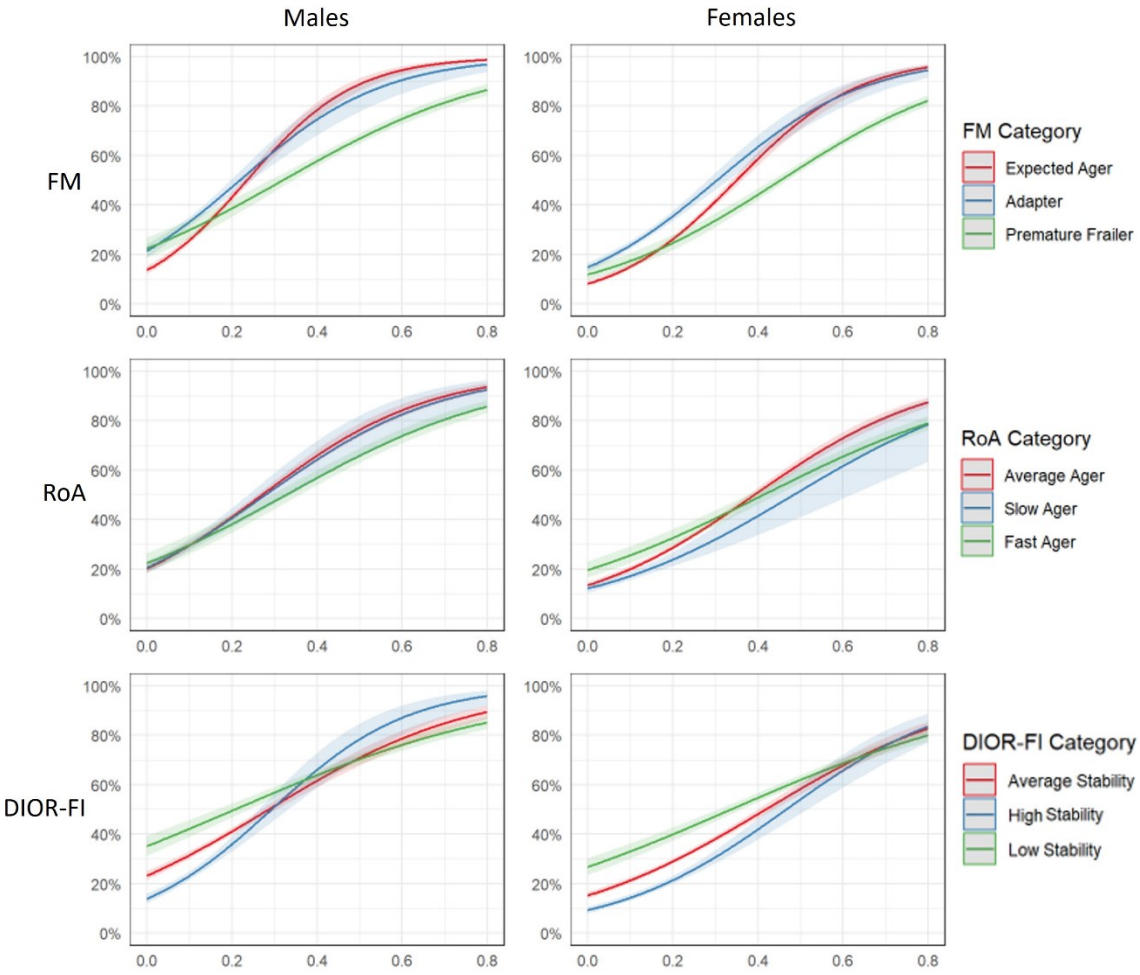


Figure 27. Sex-Stratified Interaction Effects on Mortality

Pictured from top to bottom: FM, RoA, DIOR-FI. Blue lines indicate the resilient/slow aging groups, red lines indicated the reference groups, and green lines indicate the non-resilient group. X-axis is the level of FI. Y-axis is the predicted probability of mortality. Appendix I provides the accompanying regression coefficients.

In terms of AUC, female models consistently have higher AUC than male models, though both groups show similar patterns: FI is the single best predictor, and RoA is the best resilience variable in the unadjusted, frailty-adjusted, and frailty, age, and sex-adjusted models (Table 30). Though the differences between the resilience model AUCs are negligible in the frailty, age, and sex-adjusted models. The maximal increase in the frailty, age, and sex-adjusted AUC for males was 0.012, and for females was 0.011.

Table 30. Discrimination of Sex-Stratified Models

	Females			Males		
Model	AUC	Lower 95%	Upper 95%	AUC	Lower 95%	Upper 95%
Unadjusted Models						
FI Only	**0.799**	0.792	0.806	**0.767**	0.759	0.776
FM Only	0.658	0.650	0.666	0.647	0.638	0.656
RoA Only	0.797	0.790	0.803	0.757	0.749	0.765
DIOR Only	0.706	0.699	0.714	0.686	0.678	0.695
Frailty-Adjusted Models						
FI + FM	0.823	0.817	0.830	0.789	0.781	0.798
FI + RoA	**0.833**	0.827	0.840	**0.795**	0.787	0.803
FI + DIOR	0.800	0.793	0.807	0.767	0.759	0.776
Frailty, Age, and Sex-Adjusted Models						
FI Only	0.835	0.828	0.841	0.808	0.800	0.815
FI + FM	0.838	0.832	0.845	0.810	0.803	0.818
FI + FM Interaction	0.839	0.833	0.845	0.812	0.805	0.820
FI + RoA	0.841	0.834	0.847	0.812	0.804	0.819
FI + RoA Interaction	**0.842**	0.835	0.848	**0.814**	0.807	0.822
FI + DIOR	0.836	0.829	0.842	0.809	0.801	0.816
FI + DIOR Interaction	0.837	0.831	0.844	0.811	0.804	0.819
FI + FM + RoA + DIOR	0.845	0.838	0.851	0.815	0.808	0.823
FI + FM + RoA + DIOR + All FI Interactions	0.847	0.840	0.853	0.819	0.812	0.827

Highest AUC for each category of models is bolded, excluding the combined resilience models (bottom two rows). 95% confidence intervals are calculated using the Delong method.

Alternative Cut Points Analysis

The alternative cut point sensitivity analysis using the top and bottom 15% (compared to 25% in the main analysis) shows the same trends with slightly larger effect sizes across all variables (Appendix J). In terms of discrimination, the alternative cut points resulted in slightly lower AUC values. Appendix J provides the full results for the alternative cut point analysis.

Continuous Variable Sensitivity Analysis

Continuous variable sensitivity analysis shows that the RoA predicts better than the FI when used as a continuous variable (FI AUC = 0.778, RoA AUC = 0.801). However, RoA adds nothing to the frailty, age and sex-adjusted models. Out of the resilience indicators, FM provides the best increase in AUC in the frailty, age, and sex-adjusted models. However, the differences in AUC between the different resilience models are negligible. Appendix K presents full results for the continuous sensitivity.

Alternative FI and Household Slope Sensitivity Analyses

The results of the alternative FI sensitivity analyses and the household random effect sensitivity analyses are highly consistent with the main results. Appendices L and M show the full results for the alternative FI and household random effect sensitivity analyses, respectively.

6.1.16 Summary of Mortality Analysis Results

Key takeaways from the mortality analysis results are listed below.

1. The restricted sample sensitivity analysis suggested that longitudinal variables (i.e., RoA and DIOR-FI) need to be estimated on equal time points to ensure estimates are comparable across individuals.
2. Before adjusting for age, RoA typically performs better than DIOR-FI and FM in the unadjusted and frailty-adjusted models. However, RoA loses its performance edge after adjusting for age. When used as a continuous variable, RoA predicts even better than the FI in unadjusted models. But again, the adjustment for FI and age seems to largely eliminate the effect of RoA.

3. Age-stratified results show that AUC improvements are largest in the youngest (52-67) and middle (68-79) age groups.
4. Sex-stratified results show similar patterns in males and females, with larger effect sizes and better AUC in females. DIOR-FI shows a crossover effect in males but not females. This crossover effect shows the high stability group having the lowest predicted probability of mortality at low levels of FI, but the highest predicted probability at high levels of FI.
5. FM unexpectedly shows an effect opposite to expectations with the adapters (the resilient group) having the highest predicted probability of death. Further exploratory analysis indicates that compared to the expected agers (reference group), the adapters are older and have higher disease burden despite having approximately the same level of frailty. Thus, given the distribution of risk factors between the groups, this result makes sense. With four categories, as opposed to three categories, the association holds in the expected direction if those with the very lowest FM (i.e., those most underestimated by the model, the largest outliers on the negative end) are separated.
6. There is generally low agreement between FM, RoA, and DIOR-FI categories (kappa ≤ 0.24).
7. Results are robust to different cut points, frailty indexes constructed with alternative variables, and clustering specification of the growth curve model.

6.2 Recovery Analysis

6.2.1 Sample Characteristics

The recovery sample of 1,905 individuals is 50% female with a median age of 74 (IQR bounds = 67, 82), a median FI of 0.197 (IQR bounds = 0.104, 0.345), and a median modified SF-36 PFS of 50 (IQR bounds = 20, 80) (Table 31). In terms of the resilience indicators in the study, the sample has a median rate of aging of 0.005 FI per year (IQR bounds = 0.003, 0.011), a median frailty-disease mismatch of -0.022 deviation from the expected frailty (IQR bounds = -0.083, 0.061), and a median DIOR-FI of 0.046 (IQR bounds = 0.029, 0.073).

Of the 1,905 individuals included in the sample, 506 (26.6%) fully recovered by the next wave. Those who fully recovered are significantly younger ($p < 0.001$) and less frail ($p = 0.004$) compared

to those who did not (Table 31). There is no significant difference between the groups in terms of sex (p = 0.7) or baseline modified SF-36 PFS (p = 0.8). In terms of the continuous resilience indicators, those who recovered have a significantly lower frailty-disease mismatch (p = 0.004), lower RoA (p<0.001), and a lower DIOR-FI (p < 0.001). After categorizing each of these measures by the 25th and 75th percentiles, all three measures remain statistically significantly different between those who did and did not recover (p < 0.001, p < 0.001, and p = 0.003, respectively).

Table 31. Sample Characteristics by Recovery Status

		Full Recovery		
Characteristic	**Overall**, N = 1,905[1]	**Not Recovered**, N = 1,399[1]	**Recovered**, N = 506[1]	**p-value**[2]
Age	74 (67, 82)	77 (68, 83)	70 (64, 77)	**<0.001**
Sex				0.7
Female	953 (50%)	703 (50%)	250 (49%)	
Male	952 (50%)	696 (50%)	256 (51%)	
Frailty Index	0.197 (0.104, 0.345)	0.207 (0.110, 0.348)	0.177 (0.091, 0.334)	**0.004**
Modified SF-36 PFS	50 (20, 80)	50 (20, 80)	50 (20, 80)	0.8
FI Category				**0.007**
Non-frail	443 (23%)	299 (21%)	144 (28%)	
Vulnerable	564 (30%)	417 (30%)	147 (29%)	
Frail	618 (32%)	464 (33%)	154 (30%)	
Most Frail	280 (15%)	219 (16%)	61 (12%)	
FM	-0.022 (-0.083, 0.061)	-0.026 (-0.089, 0.058)	-0.017 (-0.065, 0.067)	**0.004**
FM Category				**<0.001**
Adapter	477 (25%)	382 (27%)	95 (19%)	
Expected Ager	952 (50%)	673 (48%)	279 (55%)	
Premature Frailer	476 (25%)	344 (25%)	132 (26%)	
RoA	0.005 (0.003, 0.011)	0.006 (0.003, 0.012)	0.004 (0.002, 0.007)	**<0.001**
RoA Category				**<0.001**
Slow Ager	477 (25%)	315 (23%)	162 (32%)	
Average Ager	952 (50%)	682 (49%)	270 (53%)	
Fast Ager	476 (25%)	402 (29%)	74 (15%)	
DIOR-FI	0.046 (0.029, 0.073)	0.048 (0.030, 0.076)	0.041 (0.027, 0.064)	**<0.001**
DIOR-FI Category				**0.003**
High stability	477 (25%)	323 (23%)	154 (30%)	
Average Stability	952 (50%)	711 (51%)	241 (48%)	
Low Stability	476 (25%)	365 (26%)	111 (22%)	

[1]Median (IQR Bounds); n (%)

[2]Wilcoxon rank sum test; Pearson's Chi-squared test

Age, Modified SF-36 PFS, FI, and FM represent the final values prior to first reported myocardial infarction. RoA and DIOR-FI represent estimates from all interviews prior to first reported myocardial infarction.

6.2.2 Growth Curve Models

I estimated two models to create the resilience variables. The age-only model produces RoA, and the adjusted model produces FM and DIOR-FI.

Age-Only Model

The coefficients and 95% confidence intervals for all model components are presented in Table 32 below. The fixed (population average) growth is described by a linear, quadratic, and cubic term equal to 0.01066, -0.00065, and 0.00002, respectively. The standard deviation of the linear random effect is 0.00670, and the autocorrelation of the residuals is 0.71847 (Table 32).

Table 32. Age-Only Mixed Effects Model Results (Recovery Sample)

Fixed Effects	Estimate	Lower 95%	Upper 95%
(Intercept)	0.09308	0.07362	0.11254
Age	0.01066	0.00761	0.01372
Age^2	-0.00065	-0.00080	-0.00049
Age^3	0.00002	0.00001	0.00002
Random Effects	**Estimate**	**Lower 95%**	**Upper 95%**
Intercept (sd)	0.13556	0.12170	0.15100
Age (sd)	0.00670	0.00600	0.00748
Correlation (age and intercept)	-0.66423	-0.73711	-0.57607
Correlation Structure	**Estimate**	**Lower 95%**	**Upper 95%**
Phi (autocorrelation of residuals)	0.71847	0.68676	0.74814
Residuals	**Estimate**	**Lower 95%**	**Upper 95%**
Within-group residuals (standard error)	0.09476	0.09053	0.09918

Adjusted Model

In the adjusted model, all nine disease burden components are statistically significantly associated with FI ($p < 0.001$, Table 33). After adjusting for sex, disease burden, and wave, the linear, quadratic, and cubic age coefficients changed to 0.00420, -0.00039, and 0.00001, respectively (Table 33). This change in age coefficients towards the null is expected as adding disease burden to the model will explain some of the change in FI over time, resulting in a smaller contribution of age. The standard deviation of the linear random effect of age was reduced to 0.00468, and the autocorrelation of the residuals was reduced to 0.64793.

Table 33. Adjusted Mixed Effects Model Results (Recovery Sample)

Fixed Effects	Estimate	Lower 95%	Upper 95%
(Intercept)	0.05229	0.03489	0.06969
Age	0.00420	0.00163	0.00677
Age^2	-0.00039	-0.00052	-0.00026
Age^3	0.00001	0.00001	0.00001
Sex: Male	-0.03651	-0.04471	-0.02831
SRH: Very Good	0.00836	0.00131	0.01541
SRH: Good	0.02791	0.02054	0.03529
SRH: Fair	0.06280	0.05477	0.07083
SRH: Poor	0.13721	0.12782	0.14661
Ever had stroke: Yes	0.07611	0.06692	0.08530
Ever had arthritis: Yes	0.04478	0.03849	0.05107
Ever had cancer: Yes	0.00986	0.00071	0.01901
Ever had high blood pressure: Yes	0.02627	0.01982	0.03271
Ever had diabetes: Yes	0.01733	0.00987	0.02480
Ever had lung disease: Yes	0.04292	0.03299	0.05284
Ever had heart problems: Yes	0.02258	0.01593	0.02923
Regularly takes Rx meds: Yes	0.01483	0.00861	0.02106
Wave 4	-0.00250	-0.00729	0.00229
Wave 5	0.00175	-0.00395	0.00746
Wave 6	0.00774	0.00118	0.01429
Wave 7	0.01286	0.00560	0.02013
Wave 8	0.01863	0.01046	0.02679
Wave 9	0.01342	0.00418	0.02266
Wave 10	0.02410	0.01343	0.03477
Wave 11	0.02444	0.01219	0.03669
Wave 12	0.03096	0.01614	0.04579
Random Effects	**Estimate**	**Lower 95%**	**Upper 95%**
Intercept (sd)	0.08143	0.07055	0.09400
Age (sd)	0.00468	0.00416	0.00526
Correlation (age and intercept)	-0.68668	-0.76752	-0.58433
Correlation Structure	**Estimate**	**Lower 95%**	**Upper 95%**
Phi (autocorrelation of residuals)	0.64793	0.61750	0.67720
Residuals	**Estimate**	**Lower 95%**	**Upper 95%**
Within-group residuals (standard error)	0.08340	0.08074	0.08615

SRH stands for self-rated health. The reference category is “Excellent”. The reference category for Wave is 3.

6.2.3 Modified SF-36 Physical Function Subscale

Univariate Descriptive Statistics

The modified SF-36 PFS has an asymmetrical concave distribution with 0, 100, and 90 being the most frequently observed scores, respectively (Figure 28).

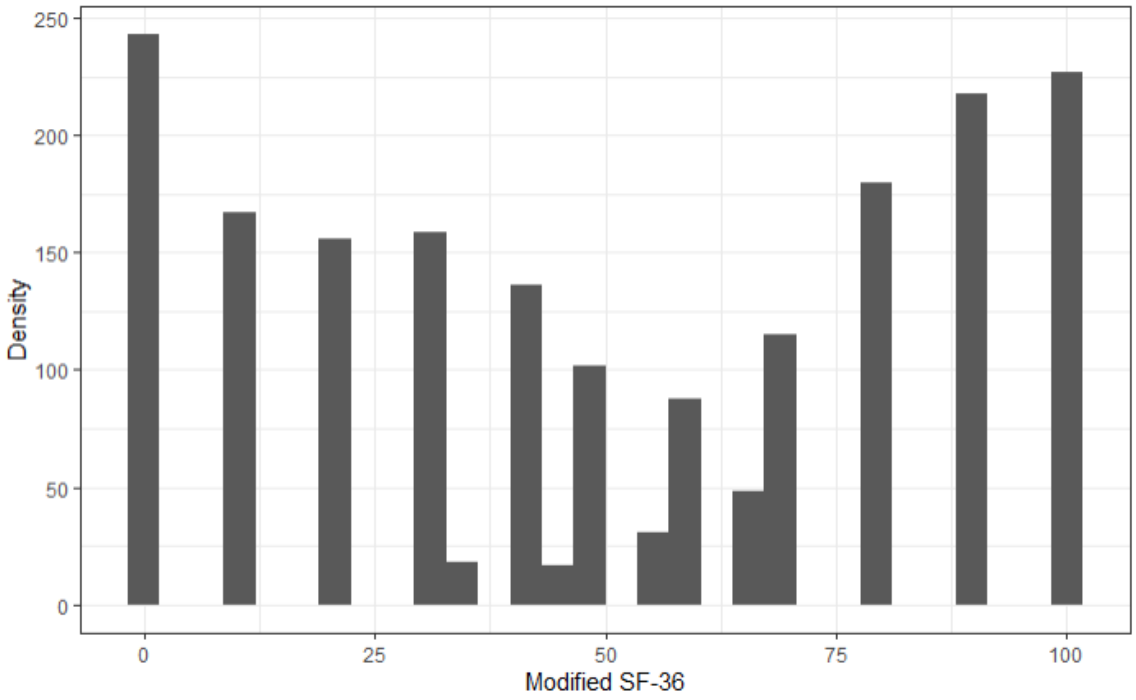


Figure 28. Distribution of the Modified SF-36 PFS (Recovery Sample)

Note: The smaller bars between increments of 10 represent imputed missing values following the guidance in the SF-36 manual (103).

The effect of pre-event physical function is not statistically significantly associated with full recovery with an odds ratio of 1.00 (95% CI = 0.97, 1.03) for a ten-unit change in the modified SF-36 PFS (Table 34).

Table 34. Univariate Logistic Regression for the Modified SF-36 PFS and Recovery

Characteristic	OR	95% CI	p-value
Modified SF-36 PFS	1.00	0.97, 1.03	0.887

Note: Odds ratio represents a SF-36 PFS change of 10, equal to one functional difficulty variable.

6.2.4 Frailty Index

Distribution of the FI

The frailty index has a right-skewed distribution with a median of 0.197, a 25th percentile of 0.104, a 75th percentile of 0.345, a 99th percentile of 0.837, a skewness of 1.244, and a kurtosis of 1.048 (Figure 29).

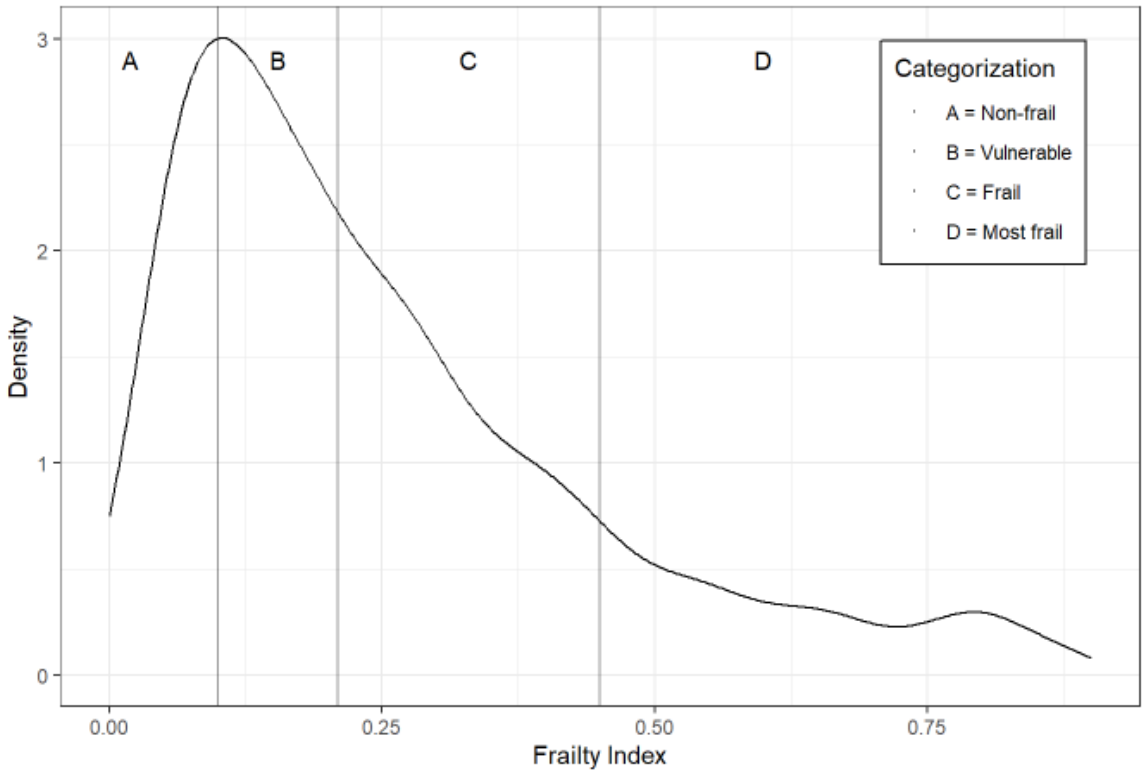


Figure 29. Distribution and Categorization of the Frailty Index (Recovery Sample)

Vertical lines illustrate categorical cut points: less than 0.1 is non-frail, less than or equal to 0.21 is vulnerable, less than or equal to 0.45 is frail, and greater than 0.45 is most frail.

Comparison of FI Categories

All variables are statistically significantly different across frailty index categories using the Kruskal-Wallis test for continuous variables and Pearson's Chi-squared test for categorical variables ($p < 0.01$, Table 35). Age, sex, physical function, and recovery show the expected gradient across FI categories: comparing the non-frail group to the most frail, the median age ranges from 70 to 80, the proportion of females ranges from 33% to 67%, and the median modified SF-36 PFS ranges from 90 to 0, and the proportion of full recovery ranges from 33% to 22%.

All resilience variables show the expected gradient across all frailty categories: the proportion of DIOR-FI low stability ranges from 6.1% in non-frail to 77% in most frail, and the proportion of DIOR-FI high stability ranges from 48% in non-frail to 1.4% in most frail (Table 35). Similarly, the proportion of RoA fast agers ranges from 2.9% in non-frail to 74% in most frail, and the proportion of RoA slow agers ranges from 47% in non-frail to 8.6% in most frail (Table 35). Lastly, the proportion of FM premature frailers ranges from 0% to 92%, and the proportion of FM adapters ranges from 49% to 0.7%. These patterns are consistent with those observed in the mortality sample.

Table 35. Comparison of FI Categories (Recovery Sample)

Characteristic	FI Category				p-value[2]
	Non-frail, N = 443[1]	**Vulnerable**, N = 564[1]	**Frail**, N = 618[1]	**Most Frail**, N = 280[1]	
Age	70 (64, 77)	73 (66, 80)	77 (68, 83)	80 (72, 88)	**<0.001**
Sex					**<0.001**
Female	144 (33%)	250 (44%)	372 (60%)	187 (67%)	
Male	299 (67%)	314 (56%)	246 (40%)	93 (33%)	
Modified SF-36 PFS	90 (89, 100)	70 (50, 80)	30 (10, 40)	0 (0, 10)	**<0.001**
Full Recovery					**0.007**
Not Recovered	299 (67%)	417 (74%)	464 (75%)	219 (78%)	
Recovered	144 (33%)	147 (26%)	154 (25%)	61 (22%)	
FM Category					**<0.001**
Adapter	219 (49%)	179 (32%)	77 (12%)	2 (0.7%)	
Expected Ager	224 (51%)	369 (65%)	338 (55%)	21 (7.5%)	
Premature Frailer	0 (0%)	16 (2.8%)	203 (33%)	257 (92%)	
RoA Category					**<0.001**
Slow Ager	206 (47%)	156 (28%)	91 (15%)	24 (8.6%)	
Average Ager	224 (51%)	359 (64%)	319 (52%)	50 (18%)	
Fast Ager	13 (2.9%)	49 (8.7%)	208 (34%)	206 (74%)	
DIOR-FI Category					**<0.001**
High Stability	212 (48%)	184 (33%)	77 (12%)	4 (1.4%)	
Average Stability	204 (46%)	327 (58%)	360 (58%)	61 (22%)	
Low Stability	27 (6.1%)	53 (9.4%)	181 (29%)	215 (77%)	

[1]Median (IQR Bounds); n (%)

[2]Kruskal-Wallis rank sum test; Pearson's Chi-squared test

The Effect of FI on Recovery

The function-adjusted analysis shows that for every 0.03 increase in FI, the odds of full recovery after incident MI decrease by 9% (OR = 0.91, 95% CI = 0.88, 0.95) (Table 36). The odds of full recovery are 0.39 times those of the non-frail reference group among the vulnerable (95% CI = 0.28, 0.55), 0.15 among the frail (95% CI = 0.09, 0.26), and 0.08 among the most frail (95% CI = 0.04, 0.15) (Table 36).

In the age-adjusted analysis, the odds of full recovery decreases by 7% for every 0.03 increase in FI (OR = 0.93, 95% CI = 0.90, 0.97) (Table 36). The odds of full recovery are 0.45 times those of the non-frail reference group among the vulnerable (95% CI = 0.32, 0.63), 0.19 among the frail (95% CI = 0.11, 0.33), and 0.11 among the most frail (95% CI = 0.06, 0.23) (Table 36). Unlike the mortality models, I did not adjust for sex as it was not statistically significant in the recovery models.

Table 36. Logistic Regression Models for the Frailty Index and Recovery

	Model 1			Model 2		
Characteristic	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
Continuous FI						
Modified SF-36 PFS	0.99	0.98, 0.99	**<0.001**	0.99	0.98, 0.99	**<0.001**
Frailty Index	0.91	0.88, 0.95	**<0.001**	0.93	0.90, 0.97	**<0.001**
Age				0.95	0.94, 0.97	**<0.001**
Categorical FI						
Modified SF-36 PFS	0.98	0.97, 0.98	**<0.001**	0.98	0.97, 0.98	**<0.001**
FI Category						
Non-frail	—	—		—	—	
Vulnerable	0.39	0.28, 0.55	**<0.001**	0.45	0.32, 0.63	**<0.001**
Frail	0.15	0.09, 0.26	**<0.001**	0.19	0.11, 0.33	**<0.001**
Most Frail	0.08	0.04, 0.15	**<0.001**	0.11	0.06, 0.23	**<0.001**
Age				0.96	0.94, 0.97	**<0.001**

Model 1 is adjusted for function. Model 2 is further adjusted for age. Note: Continuous FI odds ratio represents a change of 0.03, the proposed minimal important difference

Figure 30 displays the function-adjusted effect of the frailty index on the predicted probability of full recovery. As the frailty index increases, the predicted probability of full recovery monotonically decreases. The shaded area around the lines represents the 95% confidence intervals.

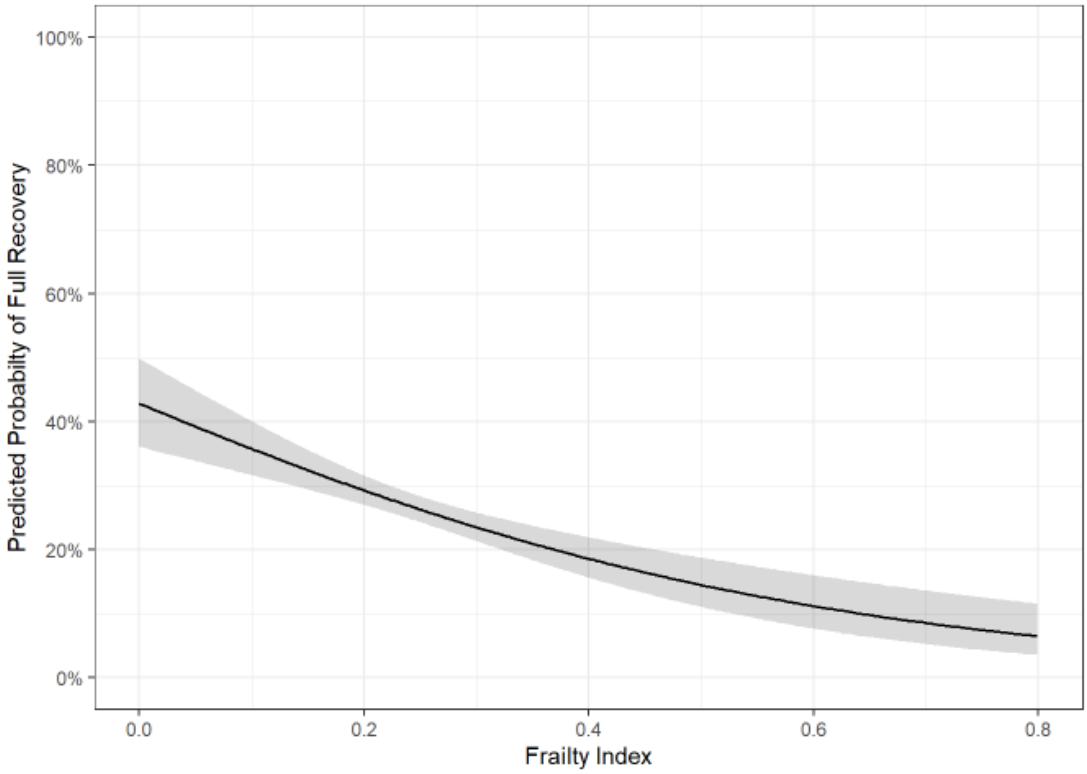


Figure 30. Function-Adjusted Effect of the Frailty Index on Full Recovery

Visualization of the function-adjusted effect of FI on full recovery (continuous Model 1, Table 36).

6.2.5 Frailty-Disease Mismatch

Distribution of the FM

The frailty-disease mismatch has a right-skewed distribution, with a median of -0.022, a 25th percentile of -0.083, a 75th percentile of 0.061, a skewness of 1.151, and a kurtosis of 2.044 (Figure 31).

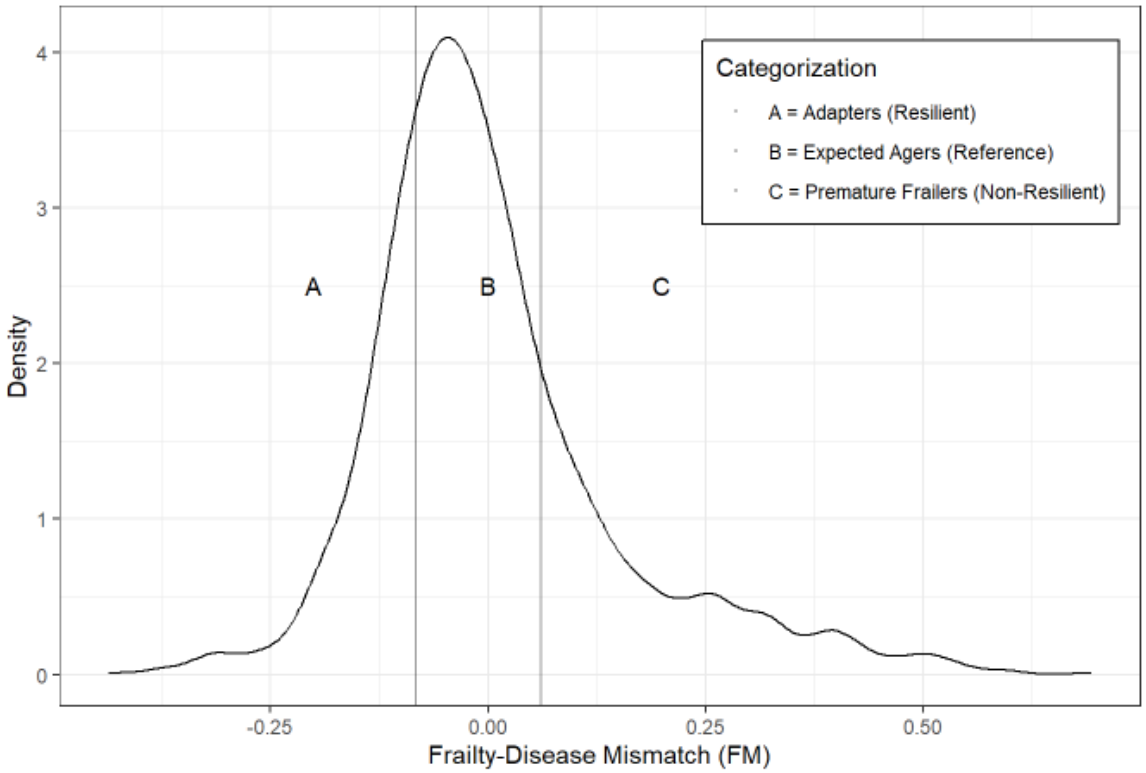


Figure 31. Distribution and Categorization of FM (Recovery Sample)

Physical resilience categories are defined by the 25th and 75th percentiles: those below the 25th percentile are adapters (resilient), those within the IQR are expected agers (reference), and those above the 75th percentile are premature frailers (non-resilient).

Comparison of FM Categories

All variables are statistically significantly different across FM categories using the Kruskal-Wallis test for continuous variables and Pearson's Chi-squared test for categorical variables ($p < 0.001$, Table 37). Mirroring the mortality analysis, most variables do not show the expected trend across FM categories except frailty and physical function. The adapters have the lowest frailty, with a median FI of 0.116 (compared to 0.520 in the premature frailers), and the highest physical function, with a median SF-36 PFS of 80 (compared to 10 in the premature frailers). Aside from frailty and physical function, the adapters are the oldest (median age of 78) and have the lowest proportion of fully recovered (20%). Additionally, FM adapters never have the highest or lowest proportion of resilient or non-resilient individuals as categorized by the RoA and DIOR-F: the expected ager category has a higher proportion of slow agers and individuals with high stability, and a lower proportion of fast agers and low stability. Despite the adapters deviating from this

expected trend, the comparison between the expected agers and premature frailers is consistent across all variables, with premature frailers being older (76 years compared to 73 years) having lower proportion of fully recovered (28% compared to 29%), and a higher proportion of fast agers (49% compared to 15%), and individuals with low stability (61% compared to 11%). These patterns are consistent with those observed in the mortality analysis.

Table 37. Comparison of FM Categories (Recovery Sample)

	FM Category			
Characteristic	**Adapter**, N = 477[1]	**Expected Ager**, N = 952[1]	**Premature Frailer**, N = 476[1]	**p-value**[2]
Age	78 (68, 85)	73 (66, 80)	76 (67, 83)	**<0.001**
Frailty Index	0.104 (0.067, 0.171)	0.174 (0.104, 0.262)	0.470 (0.356, 0.646)	**<0.001**
Modified SF-36 PFS	80 (60, 90)	60 (30, 80)	10 (0, 20)	**<0.001**
Sex				**<0.001**
Female	261 (55%)	413 (43%)	279 (59%)	
Male	216 (45%)	539 (57%)	197 (41%)	
Full Recovery				**<0.001**
Not Recovered	382 (80%)	673 (71%)	344 (72%)	
Recovered	95 (20%)	279 (29%)	132 (28%)	
RoA Category				**<0.001**
Slow Ager	133 (28%)	279 (29%)	65 (14%)	
Average Ager	246 (52%)	528 (55%)	178 (37%)	
Fast Ager	98 (21%)	145 (15%)	233 (49%)	
DIOR-FI Category				**<0.001**
High Stability	118 (25%)	341 (36%)	18 (3.8%)	
Average Stability	282 (59%)	503 (53%)	167 (35%)	
Low Stability	77 (16%)	108 (11%)	291 (61%)	

[1]Median (IQR Bounds); n (%)
[2]Kruskal-Wallis rank sum test; Pearson's Chi-squared test

The Effect of FM on Recovery

The function-adjusted logistic regression analysis shows that the odds full recovery after incident MI for the adapters are 0.57 times those of the expected agers (95% CI = 0.43, 0.75), while premature frailers do not significantly differ from the expected agers (OR = 1.05, 95% CI = 0.77, 1.41) (Table 38). Adjusting for frailty does not change the effect of adapter group, but it does increase the effect of the premature frailer group to 1.84 times higher odds of full recovery after incident MI, becoming statistically significant (95% CI = 1.31, 2.59). Further adjusting for age diminishes these effects, but they remain statistically significant ($p < 0.05$).

Table 38. Logistic Regression Models for FM and Recovery

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
Modified SF-36 PFS	1.00	1.00, 1.01	0.2	0.99	0.98, 0.99	**<0.001**	0.99	0.98, 0.99	**<0.001**	0.97	0.97, 0.98	**<0.001**
FM Category												
Expected Ager	—	—		—	—		—	—		—	—	
Adapter	0.57	0.43, 0.75	**<0.001**	0.57	0.43, 0.74	**<0.001**	0.69	0.52, 0.91	**0.010**	0.55	0.33, 0.92	**0.022**
Premature Frailer	1.05	0.77, 1.41	0.8	1.84	1.31, 2.59	**<0.001**	1.43	1.00, 2.04	**0.048**	0.18	0.07, 0.41	**<0.001**
Frailty Index				0.88	0.85, 0.92	**<0.001**	0.91	0.88, 0.95	**<0.001**	0.76	0.70, 0.82	**<0.001**
Age							0.96	0.95, 0.97	**<0.001**	0.96	0.95, 0.97	**<0.001**
FM Category * Frailty Index												
Adapter * Frailty Index										1.03	0.92, 1.14	0.6
Premature Frailer * Frailty Index										1.23	1.14, 1.33	**<0.001**

Model 1 is adjusted for function. Model 2 is further adjusted for frailty. Model 3 is further adjusted for age. Model 4 further includes a frailty interaction. Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference.

The interaction term between FI and FM is statistically significant for the premature frailer group (Table 38). Figure 32 visualizes the interaction below. The interaction shows that expected agers have the highest predicted probability of full recovery at low levels of FI, while the premature frailers have the highest predicted probability of full recovery at high levels of FI.

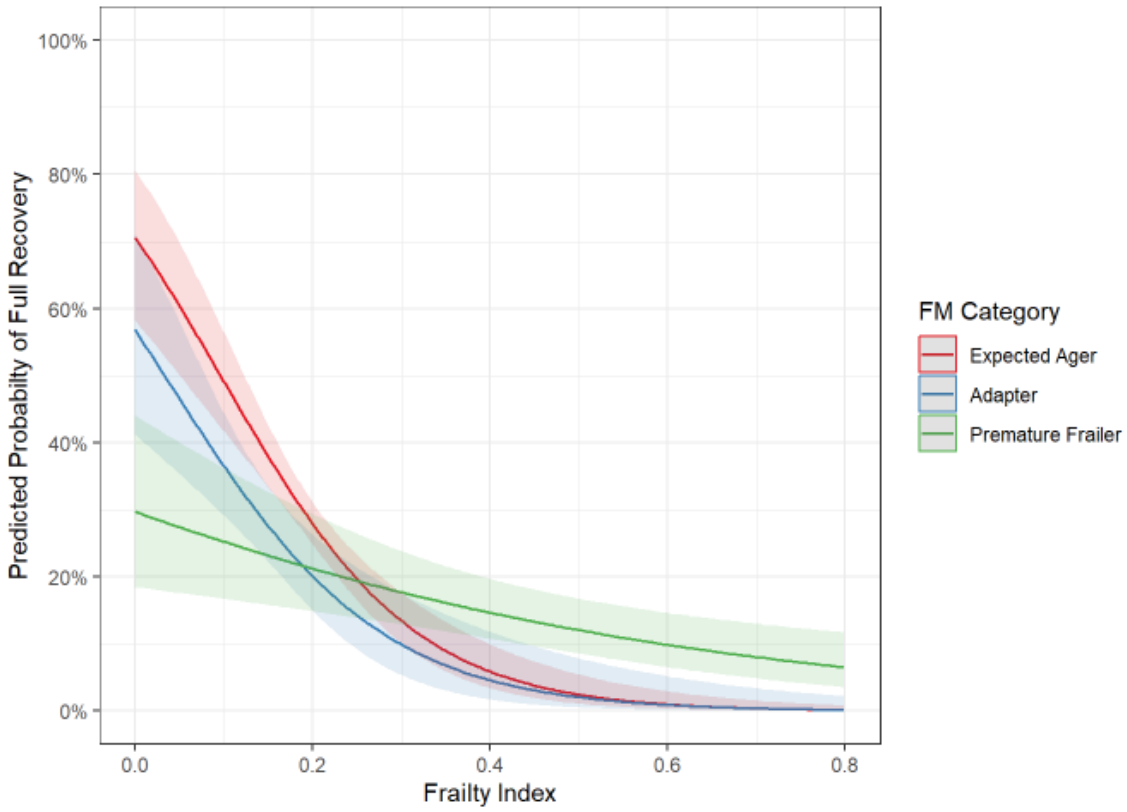


Figure 32. Interaction Effects of the Frailty Index and FM on Recovery

Visualization of the interaction between FI and FM on recovery (Model 4, Table 38).

Comparison of the disease burden variables across FM categories shows that similar to the mortality sample, the adapters have consistently higher levels of disease burden compared to the expected agers, in addition to being older (Appendix F).

6.2.6 Rate of Aging

Distribution of the RoA

The rate of aging has a right-skewed distribution, with a median of 0.005, a 25th percentile of 0.003, a 75th percentile of 0.011, a skewness of 1.672, and a kurtosis of 3.538 (Figure 33).

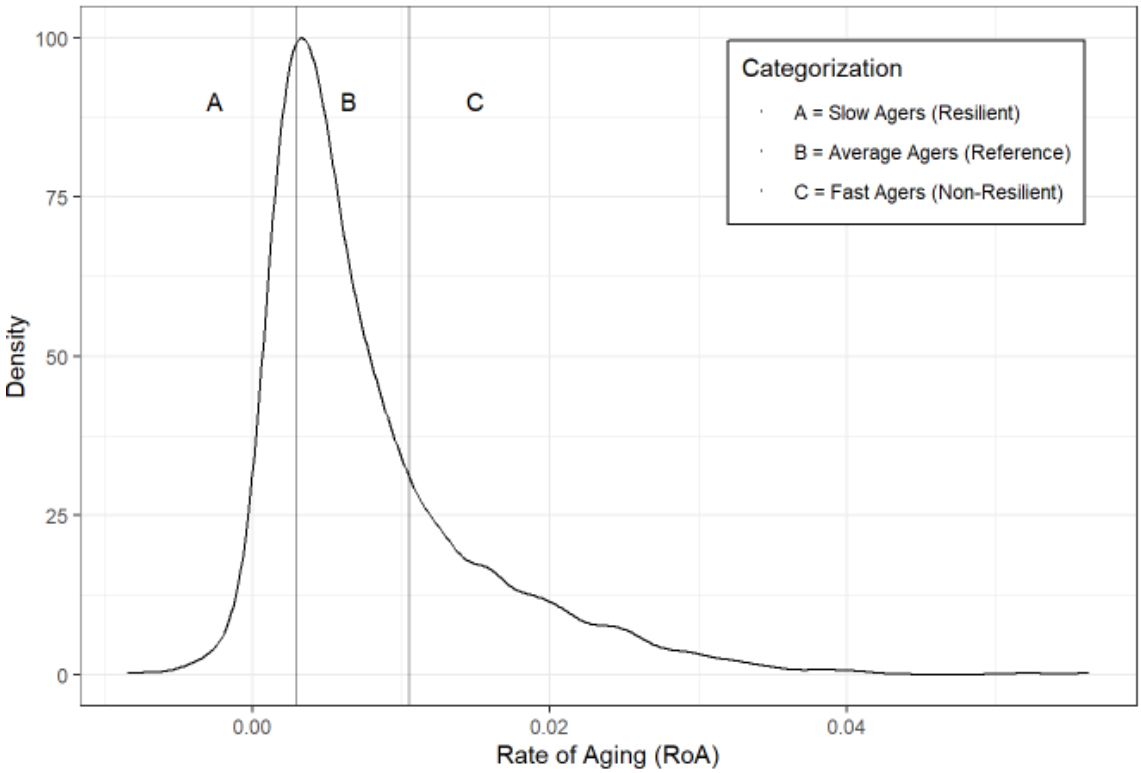


Figure 33. Distribution and Categorization of the RoA (Recovery Sample)

Physical resilience categories are defined by the 25th and 75th percentiles: those below the 25th percentile are slow agers (resilient), those within the IQR are average agers (reference), and those above the 75th percentile are fast agers (non-resilient).

Comparison of RoA Categories

All variables are statistically significantly different across RoA categories using the Kruskal-Wallis test for continuous variables and Pearson's Chi-squared test for categorical variables ($p < 0.001$, Table 39). Trends across categories follow the expected patterns for all variables. Specifically, comparing the DIOR-FI high stability (i.e., resilient) group to the low stability (i.e., non-resilient) group, the high stability group has the lowest median age (68 vs. 85), the lowest median FI at (0.116 vs. 0. 407), the lowest proportion of females (44% vs. 63%), FM premature frailers (14% vs. 49%), and DIOR-FI low stability (16% vs. 49%), the highest median SF-36 PFS (78 vs. 11), and the highest proportion of fully recovered (34% vs. 16%), FM adapters (28% vs. 21%), and DIOR-FI high stability (34% vs. 10%).

Table 39. Comparison of RoA Categories (Recovery Sample)

Characteristic	RoA Category			p-value[2]
	Slow Ager, N = 477[1]	**Average Ager**, N = 952[1]	**Fast Ager**, N = 476[1]	
Age	68 (65, 72)	73 (64, 79)	85 (82, 89)	**<0.001**
Frailty Index	0.116 (0.067, 0.201)	0.177 (0.104, 0.268)	0.407 (0.287, 0.605)	**<0.001**
Modified SF-36 PFS	78 (40, 90)	60 (30, 80)	11 (0, 33)	**<0.001**
Sex				**<0.001**
Female	212 (44%)	442 (46%)	299 (63%)	
Male	265 (56%)	510 (54%)	177 (37%)	
Full Recovery				**<0.001**
Not Recovered	315 (66%)	682 (72%)	402 (84%)	
Recovered	162 (34%)	270 (28%)	74 (16%)	
FM Category				**<0.001**
Adapter	133 (28%)	246 (26%)	98 (21%)	
Expected Ager	279 (58%)	528 (55%)	145 (30%)	
Premature Frailer	65 (14%)	178 (19%)	233 (49%)	
DIOR-FI Category				**<0.001**
High Stability	163 (34%)	266 (28%)	48 (10%)	
Average Stability	236 (49%)	522 (55%)	194 (41%)	
Low Stability	78 (16%)	164 (17%)	234 (49%)	

[1]Median (IQR Bounds); n (%)

[2]Kruskal-Wallis rank sum test; Pearson's Chi-squared test

The Effect of RoA on Recovery

The function-adjusted logistic regression analysis shows that slow agers have 1.40 times greater odds of full recovery after incident MI compared to average agers (95% CI = 1.10, 1.77), and the odds of full recovery in fast agers are 0.37 times those of the average agers (95% CI = 0.27, 0.51). After adjusting for frailty, the effect of slow agers remains the same, while the effect of fast agers decreases slightly to 0.45 times the odds of the average agers (95% CI = 0.32, 0.62). Further adjusting for age results in the effects of both categories becoming insignificant (Table 40).

Table 40. Logistic Regression Models for RoA and Recovery

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
Modified SF-36 PFS	0.99	0.99, 1.00	**<0.001**	0.99	0.98, 0.99	**<0.001**	0.99	0.98, 0.99	**<0.001**	0.98	0.98, 0.99	**<0.001**
RoA Category												
Average Ager	—	—		—	—		—	—		—	—	
Slow Ager	1.40	1.10, 1.77	**0.007**	1.40	1.10, 1.78	**0.006**	1.20	0.94, 1.54	0.14	1.45	0.97, 2.17	0.069
Fast Ager	0.37	0.27, 0.51	**<0.001**	0.45	0.32, 0.62	**<0.001**	0.84	0.57, 1.25	0.4	0.50	0.23, 1.06	0.076
Frailty Index				0.95	0.92, 0.98	**0.004**	0.94	0.91, 0.97	**<0.001**	0.92	0.88, 0.97	**0.002**
Age							0.96	0.95, 0.97	**<0.001**	0.96	0.94, 0.97	**<0.001**
RoA Category * Frailty Index												
Slow Ager * Frailty Index										0.97	0.91, 1.02	0.2
Fast Ager * Frailty Index										1.04	0.99, 1.10	0.2

Model 1 is adjusted for function. Model 2 is further adjusted for frailty. Model 3 is further adjusted for age. Model 4 further includes a frailty interaction. Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference.

The interaction model shows a no statistically significant interaction between FI and RoA. For completeness, this interaction is visualized in Figure 34 below.

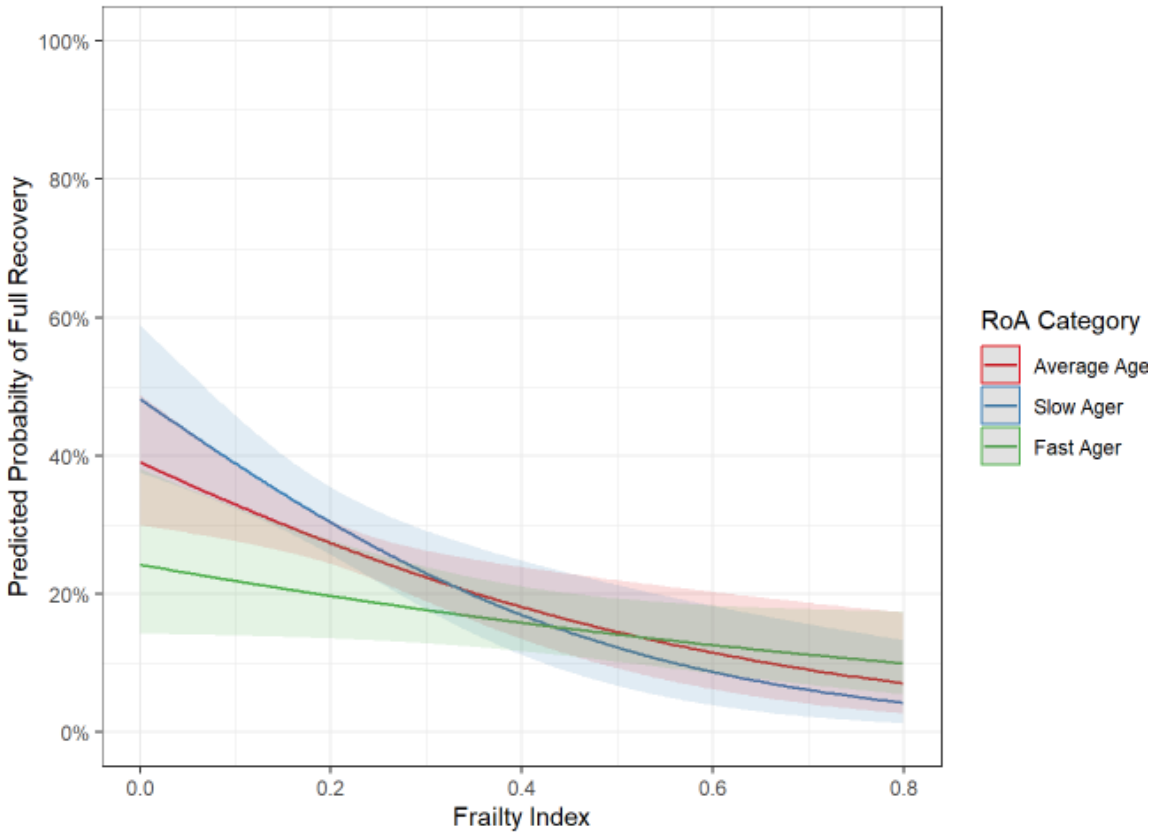


Figure 34. Interaction Effects of the Frailty Index and RoA on Recovery

Visualization of the interaction between FI and RoA on recovery (Model 4, Table 40).

6.2.7 DIOR-FI

Distribution of the DIOR-FI

The DIOR-FI has a right-skewed distribution, with a median of 0.046, a 25th percentile of 0.029, a 75th percentile of 0.073, a skewness of 1.970, and a kurtosis of 4.700 (Figure 35).

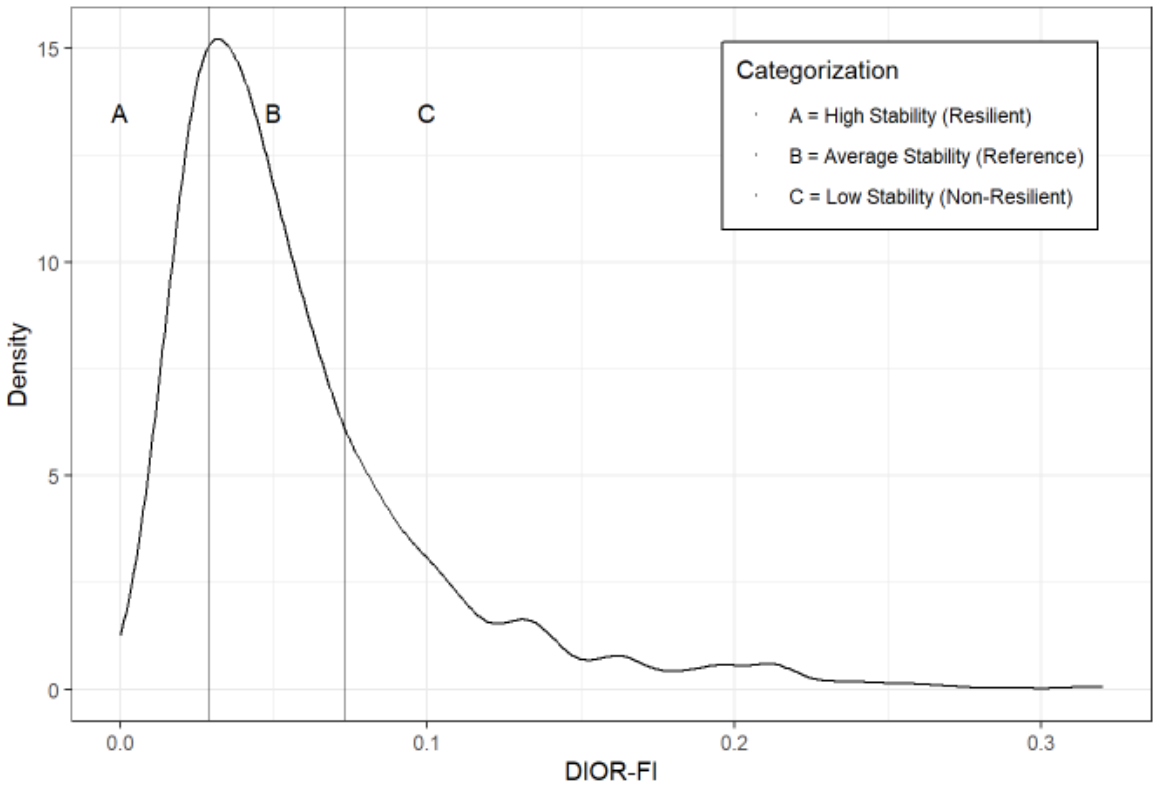


Figure 35. Distribution and Categorization of DIOR-FI (Recovery Sample)

Physical resilience categories are defined by the 25th and 75th percentiles: those below the 25th percentile have high stability (resilient), those within the IQR have average stability (reference), and those above the 75th percentile have low stability (non-resilient).

Comparison of DIOR-FI Categories

All variables are statistically significantly different across DIOR-FI categories using the Kruskal-Wallis test for continuous variables and Pearson's Chi-squared test for categorical variables ($p < 0.01$, Table 41). Trends across categories follow the expected patterns for all variables except FM adapters. Specifically, comparing the DIOR-FI high stability (i.e., resilient) group to the low stability (i.e., non-resilient) group, the high stability group has the lowest median age (72 vs. 78), the lowest median FI at (0.110 vs. 0. 415), the lowest proportion of females (41% vs. 58%), RoA fast agers (10% vs. 49%), and FM premature frailers (3.8% vs. 61%), the highest median SF-36 PFS (80 vs. 10), and the highest proportion of fully recovered (32% vs. 23%), and RoA slow agers (34% vs 16%). FM adapters deviate from this expected trend with the DIOR-FI average stability having a higher proportion than the high stability group (30% compared to 25%).

Table 41. Comparison of DIOR-FI Categories (Recovery Sample)

Characteristic	DIOR-FI Category			p-value[2]
	High Stability, N = 477[1]	**Average Stability**, N = 952[1]	**Low Stability**, N = 476[1]	
Age	72 (65, 79)	75 (67, 82)	78 (67, 85)	**<0.001**
Frailty Index	0.110 (0.067, 0.175)	0.189 (0.116, 0.287)	0.415 (0.269, 0.624)	**<0.001**
Modified SF-36 PFS	80 (60, 90)	50 (30, 80)	10 (0, 35)	**<0.001**
Sex				**<0.001**
Female	195 (41%)	484 (51%)	274 (58%)	
Male	282 (59%)	468 (49%)	202 (42%)	
Full Recovery				**0.003**
Not Recovered	323 (68%)	711 (75%)	365 (77%)	
Recovered	154 (32%)	241 (25%)	111 (23%)	
FM Category				**<0.001**
Adapter	118 (25%)	282 (30%)	77 (16%)	
Expected Ager	341 (71%)	503 (53%)	108 (23%)	
Premature Frailer	18 (3.8%)	167 (18%)	291 (61%)	
RoA Category				**<0.001**
Slow Ager	163 (34%)	236 (25%)	78 (16%)	
Average Ager	266 (56%)	522 (55%)	164 (34%)	
Fast Ager	48 (10%)	194 (20%)	234 (49%)	

[1]Median (IQR Bounds); n (%)

[2]Kruskal-Wallis rank sum test; Pearson's Chi-squared test

The Effect of DIOR-FI on Recovery

The function-adjusted logistic regression analysis shows the high stability group has 1.51 times greater odds of full recovery after incident MI compared to the average stability group (95% CI = 1.17, 1.94, Table 42). The low stability group was not statistically significantly different compared to the average stability group (p = 0.14). The effects remained largely unchanged after adjusting for both frailty and age, with only a slight reduction in the odds ratio for the high stability group (OR = 1.44, 95% CI = 1.11, 1.86), and the low stability group remaining insignificant (p = 0.8).

Table 42. Logistic Regression Models for DIOR-FI and Recovery

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
Modified SF-36 PFS	1.00	0.99, 1.00	0.054	0.98	0.98, 0.99	**<0.001**	0.99	0.98, 0.99	**<0.001**	0.98	0.98, 0.99	**<0.001**
DIOR-FI Category												
Average Stability	—	—		—	—		—	—		—	—	
High Stability	1.51	1.17, 1.94	**0.001**	1.50	1.17, 1.94	**0.002**	1.44	1.11, 1.86	**0.006**	1.79	1.13, 2.83	**0.013**
Low Stability	0.81	0.61, 1.07	0.14	1.08	0.80, 1.45	0.6	0.96	0.71, 1.30	0.8	0.62	0.33, 1.14	0.13
Frailty Index				0.92	0.88, 0.95	**<0.001**	0.94	0.91, 0.97	**<0.001**	0.91	0.86, 0.96	**<0.001**
Age							0.95	0.94, 0.97	**<0.001**	0.96	0.94, 0.97	**<0.001**
DIOR-FI Category * Frailty Index												
High Stability * Frailty Index										0.95	0.88, 1.02	0.2
Low Stability * Frailty Index										1.04	0.99, 1.10	0.14

Model 1 is adjusted for function. Model 2 is further adjusted for frailty. Model 3 is further adjusted for age. Model 4 further includes a frailty interaction. Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference

The interaction model shows no statistically significant interaction between FI and the DIOR-FI. For completeness and comparison to the other figures, this Figure 36 below visualizes this interaction.

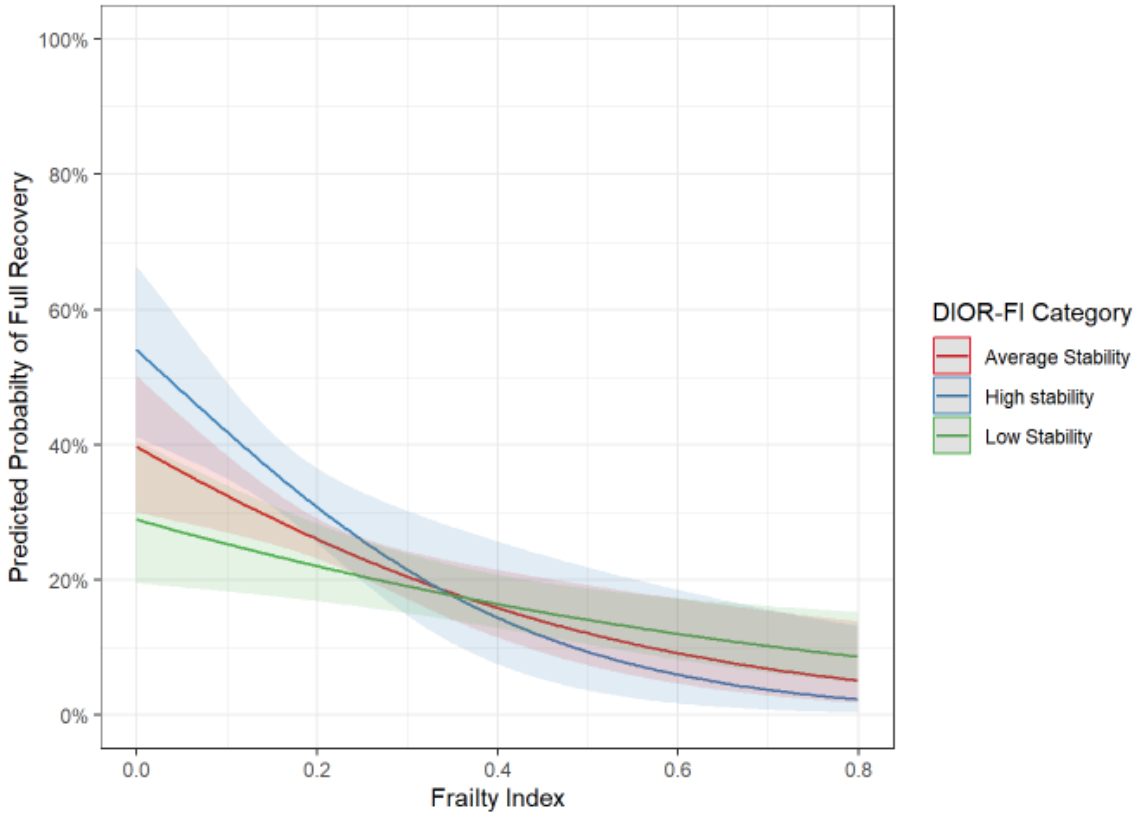


Figure 36. Interaction Effects of the Frailty Index and DIOR-FI on Recovery

Visualization of the interaction between FI and DIOR-FI on recovery (Model 4, Table 42).

6.2.8 Correlation and Agreement

Continuous variables show varying degrees of correlation using Pearson's correlation coefficient (Table 43). The highest correlations are between FI and FM (0.821), followed by FI and DIOR-FI (0.665), then FI and RoA (0.641). RoA is overall the least correlated measure in the recovery sample (Table 43).

Table 43. Correlation between FI and Resilience Indicators (Recovery Sample)

	FI	FM	RoA	DIOR-FI
FI	1.0			
FM	0.821	1.0		
RoA	0.641	0.338	1.0	
DIOR-FI	0.665	0.544	0.436	1.0

Cells indicate Pearson's correlation coefficients for pairs of continuous variables.

The unweighted Cohen's Kappa for two raters shows only slight agreement (109) between categorical indicators of resilience, with the highest agreement between RoA and DIOR-FI (0.172), followed by DIOR-FI and FM (0.166), and lastly, FM and RoA (0.151) (Table 44).

Table 44. Agreement between Categorical Resilience Indicators (Recovery Sample)

	FM	RoA	DIOR-FI
FM	1.0		
RoA	0.151	1.0	
DIOR-FI	0.166	0.172	1.0

Cells represent unweighted Cohen's Kappa statistics for pairs of categorical variables.

6.2.9 Discrimination

In contrast to the mortality analysis, the AUC values for predicting full recovery are relatively low (Table 45). Below 0.6 is generally considered to be the poor to non-informative range (7). RoA is the best individual predictor in the function-adjusted models (AUC = 0.614), outperforming the frailty index (AUC = 0.581). In the function and frailty adjusted models, RoA and FM performed similarly (AUC = 0.618). FM is the strongest predictor in the age-adjusted models (AUC = 0.672 with interaction). No other predictors are statistically significant after including the interaction term between FI and FM, leaving the best model to be the FM model with interaction, rather than the combined model.

Table 45. Discrimination of Recovery Models

Model	AUC	Lower 95%	Upper 95%
Function-Adjusted Models			
FI Only	0.581	0.553	0.610
FM Only	0.556	0.527	0.585
RoA Only	**0.614**	0.586	0.641
DIOR Only	0.558	0.530	0.587
Function and Frailty-Adjusted Models			
FI + FM	**0.618**	0.591	0.646
FI + RoA	**0.618**	0.590	0.646
FI + DIOR	0.594	0.566	0.622
Function, Frailty, Age-Adjusted Models			
FI Only	0.643	0.616	0.670
FI + FM	0.651	0.624	0.678
FI + FM Interaction	**0.672**	0.646	0.698
FI + RoA	0.645	0.618	0.672
FI + RoA Interaction	0.648	0.622	0.675
FI + DIOR	0.649	0.622	0.676
FI + DIOR Interaction	0.653	0.627	0.680
FI + FM + DIOR	0.656	0.629	0.683

Highest AUC for each category of models is bolded, excluding the combined resilience model (bottom row). 95% confidence intervals are calculated using the Delong method.

6.2.10 Restricted Sample Sensitivity Analysis

In contrast to the mortality results, the restricted sample results are very similar to those of the unrestricted sample (Figure 37, Table 46). Regardless, to remain consistent with the mortality analysis, I kept the full sample as the primary sample but corrected the estimates of RoA and DIOR-FI to incorporate only the last three derivatives/residuals. This ensures that estimates are comparable, while still using the full sample size and all longitudinal data to estimate the frailty index growth curve model. The corrected results for RoA and DIOR-FI are provided below (FI and FM remain unchanged), and all subsequent sensitivity analyses use these corrected measures.

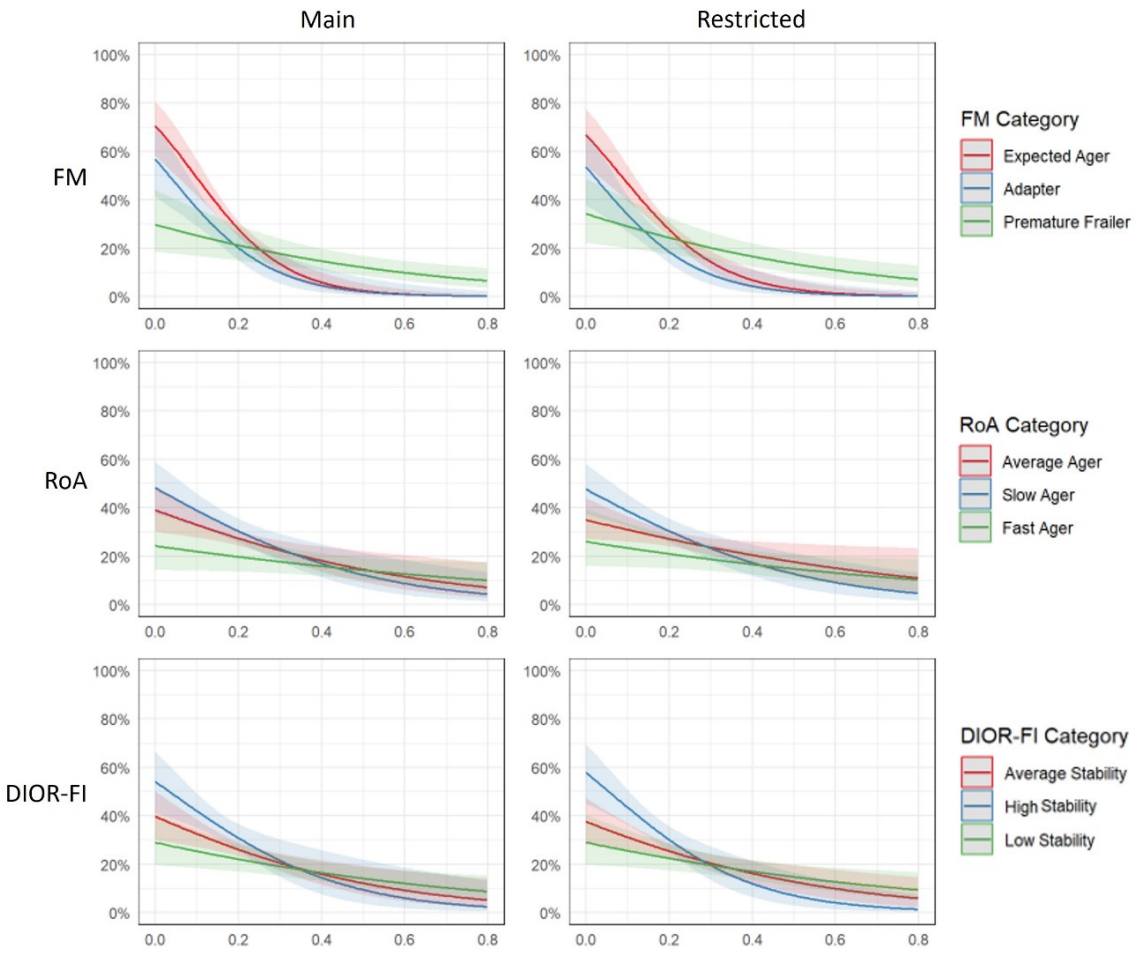


Figure 37. Recovery Main and Restricted Sample Comparison

Comparison of the interaction models in the full sample (n=1,905) vs the restricted sample with only three consecutive final interviews per individual (n=1,839). The X-axis is the level of FI. The Y-axis is the predicted probability of recovery.

Table 46. Discrimination of Recovery Models (Restricted Sample)

	Main Sample			Restricted Sample		
Model	AUC	Lower 95%	Upper 95%	AUC	Lower 95%	Upper 95%
Unadjusted Models						
FI Only	0.581	0.553	0.610	0.578	0.549	0.607
FM Only	0.556	0.527	0.585	0.563	0.533	0.592
RoA Only	**0.614**	0.586	0.641	**0.618**	0.590	0.646
DIOR Only	0.558	0.530	0.587	0.559	0.530	0.589
Frailty-Adjusted Models						
FI + FM	**0.618**	0.591	0.646	0.619	0.591	0.647
FI + RoA	**0.618**	0.590	0.646	**0.622**	0.594	0.650
FI + DIOR	0.594	0.566	0.622	0.596	0.568	0.625
Age and Frailty-adjusted models						
FI Only	0.643	0.616	0.670	0.642	0.614	0.669
FI + FM	0.651	0.624	0.678	0.654	0.626	0.681
FI + FM Interaction	**0.672**	0.646	0.698	**0.671**	0.644	0.697
FI + RoA	0.645	0.618	0.672	0.645	0.617	0.672
FI + RoA Interaction	0.648	0.622	0.675	0.651	0.623	0.678
FI + DIOR	0.649	0.622	0.676	0.651	0.624	0.678
FI + DIOR Interaction	0.653	0.627	0.680	0.659	0.632	0.686
FI + FM + DIOR	0.656	0.629	0.683	0.661	0.634	0.688

Highest AUC for each category of models is bolded, excluding the combined resilience model (bottom row). 95% confidence intervals are calculated using the Delong method.

6.2.11 Sample Characteristics (corrected for last three observations only)

The corrected values for RoA and FI are shaded in Table 47 below. As expected, the median RoA and DIOR FI are higher after correcting for the number of observations used to derive the estimates: RoA now has a median of 0.006 (IQR bounds = 0.003, 0.014), and DIOR-FI now has a median of 0.043 (IQR bounds = 0.024, 0.071).

Table 47. Sample Characteristics by Recovery Status (corrected RoA and DIOR-FI)

Characteristic	**Overall**, N = 1,905[1]	**Full Recovery**		**p-value**[2]
		Not Recovered, N = 1,399[1]	**Recovered**, N = 506[1]	
Age	74 (67, 82)	77 (68, 83)	70 (64, 77)	**<0.001**
Sex				0.7
Female	953 (50%)	703 (50%)	250 (49%)	
Male	952 (50%)	696 (50%)	256 (51%)	
Frailty Index	0.197 (0.104, 0.345)	0.207 (0.110, 0.348)	0.177 (0.091, 0.334)	**0.004**
Modified SF-36 PFS	50 (20, 80)	50 (20, 80)	50 (20, 80)	0.8
FI Category				**0.007**
Non-frail	443 (23%)	299 (21%)	144 (28%)	
Vulnerable	564 (30%)	417 (30%)	147 (29%)	
Frail	618 (32%)	464 (33%)	154 (30%)	
Most Frail	280 (15%)	219 (16%)	61 (12%)	
FM	-0.022 (-0.083, 0.061)	-0.026 (-0.089, 0.058)	-0.017 (-0.065, 0.067)	**0.004**
FM Category				**<0.001**
Adapter	477 (25%)	382 (27%)	95 (19%)	
Expected Ager	952 (50%)	673 (48%)	279 (55%)	
Premature Frailer	476 (25%)	344 (25%)	132 (26%)	
RoA	0.006 (0.003, 0.014)	0.007 (0.003, 0.015)	0.004 (0.002, 0.009)	**<0.001**
RoA Category				**<0.001**
Slow Ager	477 (25%)	304 (22%)	173 (34%)	
Average Ager	952 (50%)	691 (49%)	261 (52%)	
Fast Ager	476 (25%)	404 (29%)	72 (14%)	
DIOR-FI	0.043 (0.024, 0.071)	0.045 (0.025, 0.074)	0.037 (0.020, 0.063)	**<0.001**
DIOR-FI Category				**<0.001**
High Stability	477 (25%)	318 (23%)	159 (31%)	
Average Stability	952 (50%)	711 (51%)	241 (48%)	
Low Stability	476 (25%)	370 (26%)	106 (21%)	

[1]Median (IQR Bounds); n (%)

[2]Wilcoxon rank sum test; Pearson's Chi-squared test

Age, Modified SF-36 PFS, FI, and FM represent the final values prior to first reported myocardial infarction. RoA and DIOR-FI represent estimates using only the last three interviews prior to first reported myocardial infarction.

6.2.12 Rate of Aging (corrected for last three observations only)

Distribution of the RoA

The rate of aging has a right-skewed distribution, with a median of 0.006, a 25th percentile of 0.003, a 75th percentile of 0.014, a skewness of 1.416, and a kurtosis of 2.077 (Figure 38).

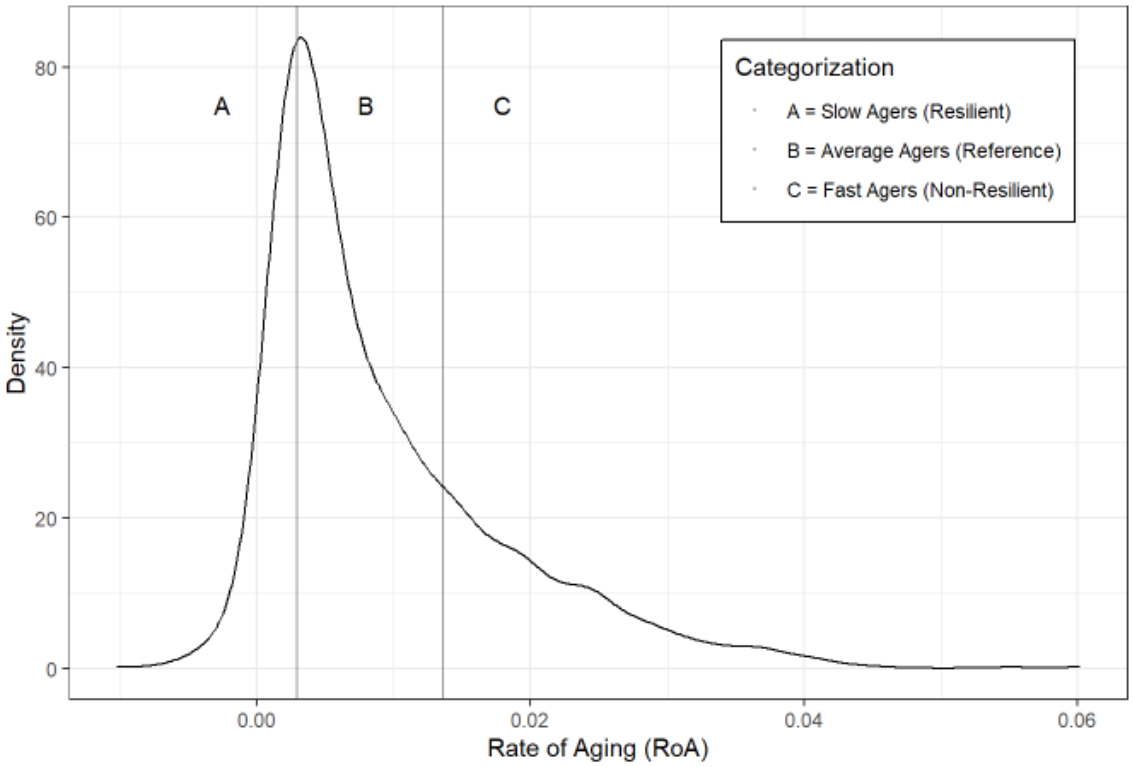


Figure 38. Distribution and Categorization of the RoA (Recovery Sample - last three)

Physical resilience categories are defined by the 25th and 75th percentiles: those below the 25th percentile are slow agers (resilient), those within the IQR are average agers (reference), and those above the 75th percentile are fast agers (non-resilient). "Last three" refers to the second (and preferred) version of RoA which uses only the last three derivatives in its calculation to ensure all individuals have a comparable estimate regardless of how many interviews they completed.

Comparison of RoA Categories

All variables are statistically significantly different across RoA categories using the Kruskal-Wallis test for continuous variables and Pearson's Chi-squared test for categorical variables ($p < 0.001$, Table 48). Trends across categories follow the expected patterns for all variables except sex and FM adapters. Specifically, comparing the DIOR-FI high stability (i.e., resilient) group to the low stability (i.e., non-resilient) group, the high stability group has the lowest median age (67 vs. 86), the lowest median FI at (0.122 vs. 0. 396), FM premature frailers (15% vs. 44%), and DIOR-FI low stability (15% vs. 46%), the highest median SF-36 PFS (70 vs. 20), and the highest proportion of fully recovered (36% vs. 15%), and DIOR-FI high stability (36% vs. 10%). Sex and FM adapters

deviate from this trend with the RoA average agers having the lowest proportion of females (45%), and all three RoA categories having an equal proportion of FM adapters (25%).

Table 48. Comparison of RoA Categories (Recovery Sample - last three)

	RoA Category			
Characteristic	**Slow Ager**, N = 477[1]	**Average Ager**, N = 952[1]	**Fast Ager**, N = 476[1]	**p-value**[2]
Age	67 (63, 69)	75 (67, 79)	86 (83, 89)	**<0.001**
Frailty Index	0.122 (0.067, 0.213)	0.177 (0.104, 0.274)	0.396 (0.264, 0.597)	**<0.001**
Modified SF-36 PFS	70 (40, 90)	60 (30, 89)	20 (0, 40)	**<0.001**
Sex				**<0.001**
Female	223 (47%)	432 (45%)	298 (63%)	
Male	254 (53%)	520 (55%)	178 (37%)	
Full Recovery				**<0.001**
Not Recovered	304 (64%)	691 (73%)	404 (85%)	
Recovered	173 (36%)	261 (27%)	72 (15%)	
FM Category				**<0.001**
Adapter	120 (25%)	239 (25%)	118 (25%)	
Expected Ager	286 (60%)	518 (54%)	148 (31%)	
Premature Frailer	71 (15%)	195 (20%)	210 (44%)	
DIOR-FI Category				**<0.001**
High Stability	171 (36%)	258 (27%)	48 (10%)	
Average Stability	235 (49%)	507 (53%)	210 (44%)	
Low Stability	71 (15%)	187 (20%)	218 (46%)	

[1]Median (IQR Bounds); n (%)

[2]Kruskal-Wallis rank sum test; Pearson's Chi-squared test

"Last three" refers to using the second (and preferred) versions of RoA and DIOR-FI, which use only the last three derivatives/residuals in their calculation to ensure all individuals have a comparable estimate regardless of how many interviews they completed.

The Effect of RoA on Recovery

The function-adjusted logistic regression analysis shows that slow agers have 1.59 times greater odds of full recovery after incident MI compared to average agers (95% CI = 1.25, 2.02), and the odds of full recovery in fast agers are 0.39 times those of the average agers (95% CI = 0.29, 0.53). After adjusting for frailty, the effect of slow agers remains the same, while the effect of fast agers decreases slightly to 0.46 times the odds of the average agers (95% CI = 0.33, 0.64). Further adjusting for age results in the effects of both categories becoming insignificant (Table 49).

Table 49. Logistic Regression Models for RoA and Recovery (last three)

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
Modified SF-36 PFS	0.94	0.91, 0.97	**<0.001**	0.87	0.82, 0.93	**<0.001**	0.88	0.82, 0.93	**<0.001**	0.86	0.80, 0.91	**<0.001**
RoA Category												
Average Ager	—	—		—	—		—	—		—	—	
Slow Ager	1.59	1.25, 2.02	**<0.001**	1.59	1.25, 2.01	**<0.001**	1.25	0.97, 1.62	0.086	1.71	1.14, 2.56	**0.009**
Fast Ager	0.39	0.29, 0.53	**<0.001**	0.46	0.33, 0.64	**<0.001**	0.83	0.55, 1.23	0.4	0.59	0.28, 1.21	0.2
Frailty Index				0.95	0.91, 0.98	**0.003**	0.94	0.91, 0.97	**<0.001**	0.94	0.89, 0.98	**0.009**
Age							0.96	0.95, 0.98	**<0.001**	0.96	0.95, 0.98	**<0.001**
RoA Category * Frailty Index												
Slow Ager * Frailty Index										0.95	0.90, 1.00	**0.043**
Fast Ager * Frailty Index										1.02	0.97, 1.08	0.4

Model 1 is adjusted for function. Model 2 is further adjusted for frailty. Model 3 is further adjusted for age. Model 4 further includes a frailty interaction. Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference. “Last three” refers to the second (and preferred) version of RoA which uses only the last three derivatives in its calculation to ensure all individuals have a comparable estimate regardless of how many interviews they completed.

The interaction model shows a statistically significant interaction between FI and RoA slow agers, with the main effect for slow agers becoming significant again (Table 49). Figure 39 below illustrates this interaction. This shows the differences between groups decreasing as FI increases.

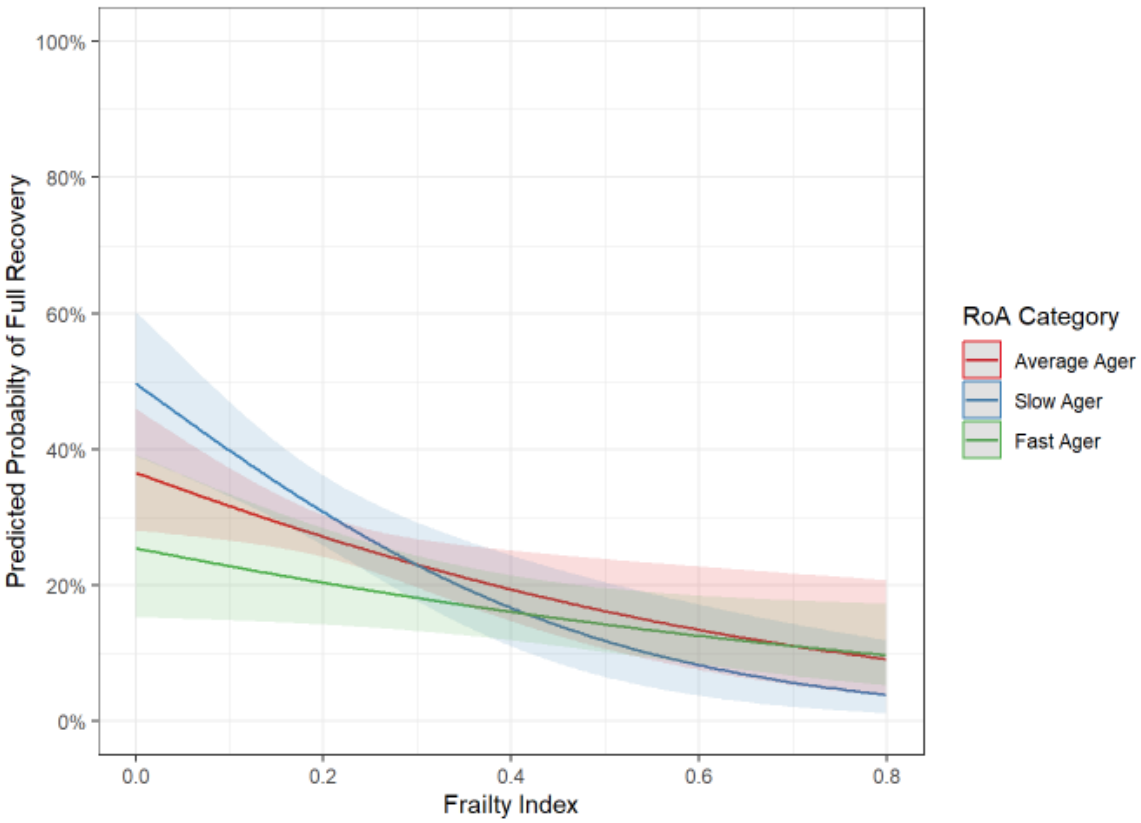


Figure 39. Interaction Effects of the Frailty Index and RoA on Recovery (last three)

Visualization of the interaction between FI and RoA on recovery (Model 4, Table 49). "Last three" refers to the second (and preferred) version of RoA which uses only the last three derivatives in its calculation to ensure all individuals have a comparable estimate regardless of how many interviews they completed.

6.2.13 DIOR-FI (corrected for last three observations only)

Distribution of the DIOR-FI

The DIOR-FI has a right-skewed distribution, with a median of 0.043, a 25th percentile of 0.024, a 75th percentile of 0.071, a skewness of 2.242, and a kurtosis of 6.484 (Figure 40).

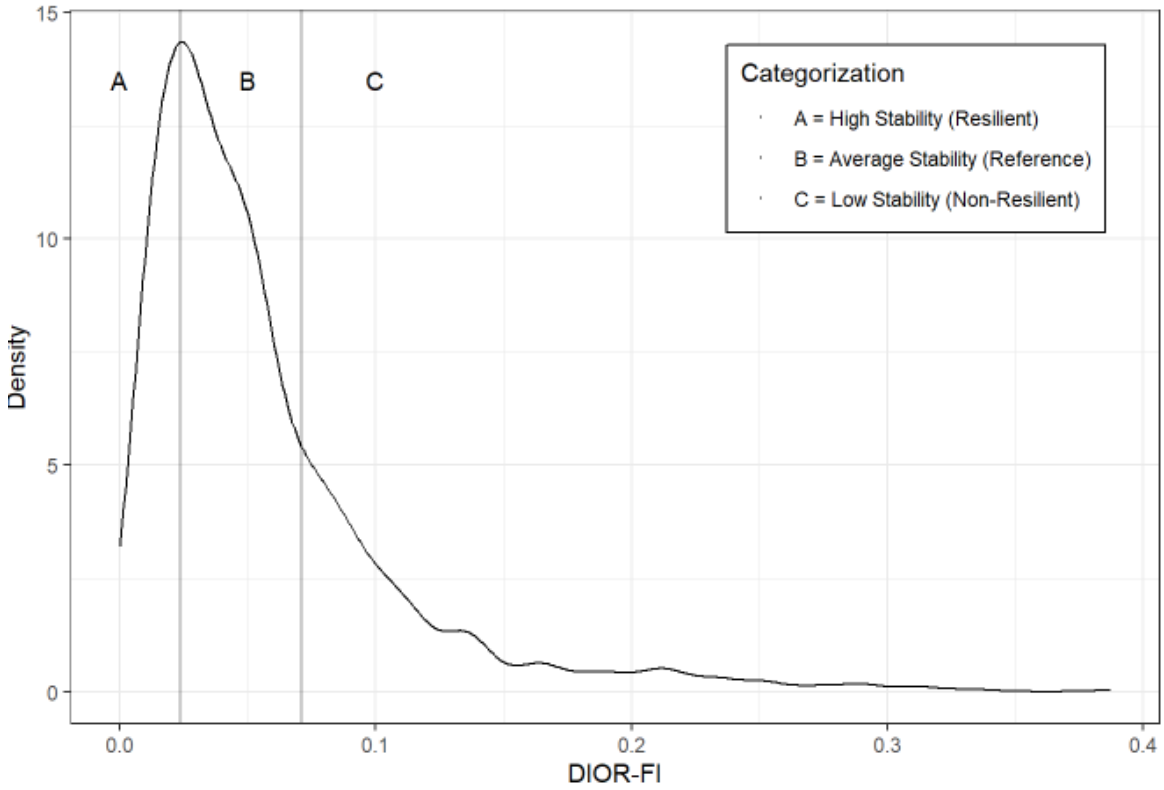


Figure 40. Distribution and Categorization of DIOR-FI (Recovery Sample - last three)

Physical resilience categories are defined by the 25th and 75th percentiles: those below the 25th percentile have high stability (resilient), those within the IQR have average stability (reference), and those above the 75th percentile have low stability (non-resilient). "Last three" refers to the second (and preferred) version of DIOR-FI which uses only the last three residuals in its calculation to ensure all individuals have a comparable estimate regardless of how many interviews they completed.

Comparison of DIOR-FI Categories

All variables are statistically significantly different across DIOR-FI categories using the Kruskal-Wallis test for continuous variables and Pearson's Chi-squared test for categorical variables ($p < 0.01$, Table 50). Trends across categories follow the expected patterns for all variables except FM adapters. Specifically, comparing the DIOR-FI high stability (i.e., resilient) group to the low stability (i.e., non-resilient) group, the high stability group has the lowest median age (71 vs. 78), the lowest median FI at (0.116 vs. 0. 396), the lowest proportion of females (41% vs. 57%), RoA fast agers (10% vs. 46%), and FM premature frailers (5.9% vs. 58%), the highest median SF-36 PFS (80 vs. 11), and the highest proportion of fully recovered (33% vs. 22%), and RoA slow agers (36% vs

15%). FM adapters deviate from this expected trend with the DIOR-FI average stability having a higher proportion than the high stability group (29% compared to 26%).

Table 50. Comparison of DIOR-FI Categories (Recovery Sample - last three)

Characteristic	DIOR-FI Category			p-value[2]
	High Stability, N = 477[1]	**Average Stability**, N = 952[1]	**Low Stability**, N = 476[1]	
Age	71 (65, 78)	75 (67, 82)	78 (68, 85)	**<0.001**
Frailty Index	0.116 (0.063, 0.183)	0.186 (0.104, 0.288)	0.396 (0.262, 0.593)	**<0.001**
Modified SF-36 PFS	80 (50, 90)	56 (30, 80)	11 (0, 40)	**<0.001**
Sex				**<0.001**
Female	196 (41%)	488 (51%)	269 (57%)	
Male	281 (59%)	464 (49%)	207 (43%)	
Full Recovery				**<0.001**
Not Recovered	318 (67%)	711 (75%)	370 (78%)	
Recovered	159 (33%)	241 (25%)	106 (22%)	
FM Category				**<0.001**
Adapter	124 (26%)	274 (29%)	79 (17%)	
Expected Ager	325 (68%)	504 (53%)	123 (26%)	
Premature Frailer	28 (5.9%)	174 (18%)	274 (58%)	
RoA Category				**<0.001**
Slow Ager	171 (36%)	235 (25%)	71 (15%)	
Average Ager	258 (54%)	507 (53%)	187 (39%)	
Fast Ager	48 (10%)	210 (22%)	218 (46%)	

[1]Median (IQR Bounds); n (%)

[2]Kruskal-Wallis rank sum test; Pearson's Chi-squared test

"Last three" refers to using the second (and preferred) versions of RoA and DIOR-FI, which use only the last three derivatives/residuals in their calculation to ensure all individuals have a comparable estimate regardless of how many interviews they completed.

The Effect of DIOR-FI on Recovery

The function-adjusted logistic regression analysis shows the high stability group has 1.58 greater odds of full recovery after incident MI compared to the average stability group (95% CI = 1.23, 2.02, Table 51). The low stability group was not statistically significantly different compared to the average stability group (p = 0.053). The effects remained largely unchanged after adjusting for both frailty and age, with only a slight reduction in the odds ratio for the high stability group (OR = 1.48, 95% CI = 1.15, 1.91), and the low stability group remaining insignificant (p = 0.3).

Table 51. Logistic Regression Models for DIOR-FI and Recovery (last three)

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
Modified SF-36 PFS	0.96	0.93, 1.00	**0.031**	0.86	0.81, 0.91	**<0.001**	0.87	0.81, 0.92	**<0.001**	0.83	0.77, 0.88	**<0.001**
DIOR-FI Category												
Average Stability	—	—		—	—		—	—		—	—	
High Stability	1.58	1.23, 2.02	**<0.001**	1.57	1.22, 2.01	**<0.001**	1.48	1.15, 1.91	**0.003**	3.03	1.93, 4.77	**<0.001**
Low Stability	0.76	0.57, 1.00	0.053	0.94	0.70, 1.26	0.7	0.86	0.64, 1.16	0.3	0.68	0.37, 1.23	0.2
Frailty Index				0.92	0.89, 0.95	**<0.001**	0.94	0.91, 0.98	**0.001**	0.93	0.89, 0.98	**0.006**
Age							0.95	0.94, 0.97	**<0.001**	0.96	0.94, 0.97	**<0.001**
DIOR-FI Category * Frailty Index												
High Stability * Frailty Index										0.87	0.80, 0.93	**<0.001**
Low Stability * Frailty Index										1.01	0.97, 1.06	0.6

Model 1 is adjusted for function. Model 2 is further adjusted for frailty. Model 3 is further adjusted for age. Model 4 further includes a frailty interaction. Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference. “Last three” refers to the second (and preferred) version of DIOR-FI which uses only the last three residuals in its calculation to ensure all individuals have a comparable estimate regardless of how many interviews they completed.

The interaction model shows a statistically significant interaction between FI and the DIOR-FI high stability group. Figure 41 below visualizes this interaction. At low levels of FI, the high stability group has the highest predicted probability of full recovery. This converges with and eventually crosses over the other two groups around an FI of 0.3, where the high stability group has the lowest predicted probability of full recovery (though the confidence intervals still slightly overlap).

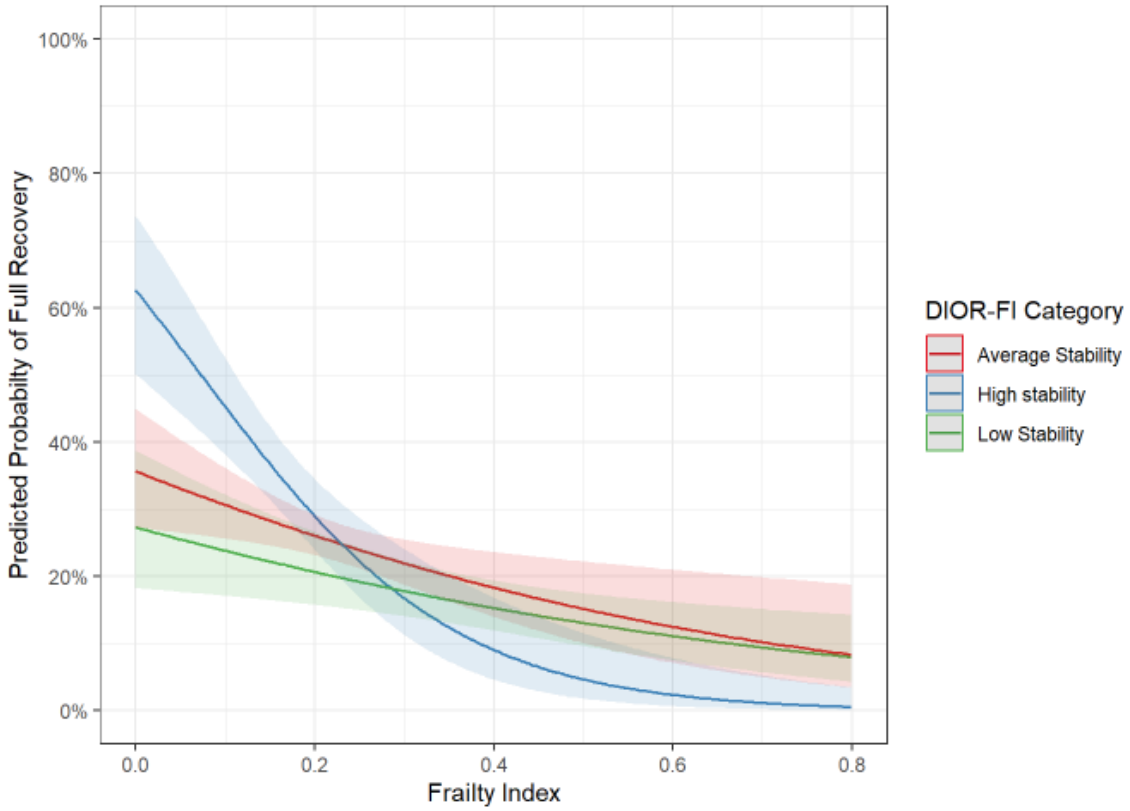


Figure 41. Interaction Effects of the Frailty Index and DIOR-FI on Recovery (last three)

Visualization of the interaction between FI and DIOR-FI on recovery (Model 4, Table 51). "Last three" refers to the second (and preferred) version of DIOR-FI which uses only the last three residuals in its calculation to ensure all individuals have a comparable estimate regardless of how many interviews they completed.

6.2.14 Correlation and Agreement (corrected for last three observations only)

Continuous variables show varying degrees of correlation using Pearson's correlation coefficient (Table 52). The highest correlations are between FI and FM (0.821), followed by FI and DIOR-FI (0.622), then FI and RoA (0.595). RoA is overall the least correlated measure in the recovery sample.

Table 52. Correlation between FI and Resilience Indicators (Recovery Sample – last three)

	FI	FM	RoA	DIOR-FI
FI	1.0			
FM	0.821	1.0		
RoA	0.595	0.253	1.0	
DIOR-FI	0.622	0.531	0.369	1.0

Cells indicate Pearson's correlation coefficients for pairs of continuous variables. "Last three" refers to using the second (and preferred) versions of RoA and DIOR-FI, which use only the last three derivatives/residuals in their calculation to ensure all individuals have a comparable estimate regardless of how many interviews they completed.

The unweighted Cohen's Kappa for two raters shows only slight agreement (109) between categorical indicators of resilience, with the highest agreement between FM and DIOR-FI (0.158), followed by DIOR-FI and RoA (0.153), and lastly, FM and RoA (0.112) (Table 53).

Table 53. Agreement between Categorical Resilience Indicators (Recovery Sample – last three)

	FM	RoA	DIOR-FI
FM	1.0		
RoA	0.112	1.0	
DIOR-FI	0.158	0.153	1.0

Cells represent unweighted Cohen's Kappa statistics for pairs of categorical variables. "Last three" refers to using the second (and preferred) versions of RoA and DIOR-FI, which use only the last three derivatives/residuals in their calculation to ensure all individuals have a comparable estimate regardless of how many interviews they completed.

6.2.15 Discrimination (corrected for last three observations only)

RoA is the best individual predictor in the function-adjusted models (AUC = 0.625), beating the frailty index (AUC = 0.581, Table 54). In the function and frailty adjusted models, RoA is the best again (AUC = 0.628), followed by FM (AUC = 0.618). However, FM is the strongest predictor in the age-adjusted models (AUC = 0.672 with interaction). No other predictors are statistically significant after including the interaction term between FI and FM, leaving the best model to be the FM model with interaction, rather than the combined model. The largest possible increase in AUC in the fully adjusted model was 0.017.

Table 54. Discrimination of Recovery Models (last three)

Model	AUC	Lower 95%	Upper 95%
Function-Adjusted Models			
FI Only	0.581	0.553	0.610
FM Only	0.556	0.527	0.585
RoA Only	**0.625**	0.597	0.652
DIOR Only	0.573	0.544	0.601
Function and Frailty-Adjusted Models			
FI + FM	0.618	0.591	0.646
FI + RoA	**0.628**	0.601	0.655
FI + DIOR	0.604	0.576	0.632
Function, Frailty, Age-Adjusted Models			
FI Only	0.643	0.616	0.670
FI + FM	0.651	0.624	0.678
FI + FM Interaction	**0.672**	0.646	0.698
FI + RoA	0.647	0.620	0.674
FI + RoA Interaction	0.652	0.625	0.679
FI + DIOR	0.654	0.627	0.680
FI + DIOR Interaction	0.666	0.639	0.693
FI + FM + DIOR	0.660	0.633	0.686

Highest AUC for each category of models is bolded, excluding the combined resilience model (bottom row). 95% confidence intervals are calculated using the Delong method. “Last three” refers to using the second (and preferred) versions of RoA and DIOR-FI, which use only the last three derivatives/residuals in their calculation to ensure all individuals have a comparable estimate regardless of how many interviews they completed.

6.2.16 Additional Sensitivity Analysis Results

Following the structure of the mortality analyses, all sensitivity analyses presented here use the preferred “last three” estimation of RoA and DIOR-FI, which ensures all individuals have a comparable estimate regardless of how many interviews they completed.

Alternative Cut Points Analysis

The alternative cut point sensitivity analysis using the top and bottom 15% (compared to 25% in the main analysis) revealed similar same overall trends, but many variables lose significance in

the interaction models, likely due to the reduced size of the resilient and non-resilient categories (Appendix J).

Continuous Sensitivity Analysis

Continuous resilience variables (FM, RoA, DIOR-FI) show slightly higher AUC compared to categorical. RoA remains the best individual predictor, however, after adjusting for frailty, FM becomes the best (Appendix K).

Alternative FI Sensitivity Analysis

The results of the alternative FI sensitivity analyses are highly consistent with the main results (Appendix L).

6.2.17 Summary of Recovery Analysis Results

Key takeaways from the recovery analysis results are listed below.

1. The results are consistent with mortality results.
2. FM again shows the opposite effect than expected with the adapters (the resilient group) having the lowest predicted probability of full recovery. Additional exploratory analysis again indicates that compared to the expected agers (reference group), the adapters are older and have higher disease burden.
3. RoA is the best individual predictor until adjusting for age (in categorical models) or frailty (in continuous models).
4. DIOR-FI shows convergence and crossover at high levels of FI. This shows the high stability group having the highest predicted probability of full recovery at low levels of FI, but the lowest predicted probability at high levels of FI.
5. Similar to the mortality analysis, there is low agreement between the resilience indicators (kappa ≤ 0.158).
6. Stratified analyses were not performed due to the lower sample size.
7. Results are robust to alternative frailty indexes constructed with alternative variables. Alternative cut point results were consistent, but many terms lost significance due to the small group sizes.

6.3 Comparison of Mortality and Recovery Analyses

Overall, the two analyses show similar patterns: For RoA (middle row, Figure 42), the fast agers (green line) start with the highest (lowest) predicted probability of mortality (recovery), but the three groups converge at higher levels of FI. For FM (top row, Figure 42), the premature frailers (green line) show the lowest (highest) predicted probability of morality (recovery) at high levels of FI. Lastly, DIOR-FI (bottom row, Figure 42) shows the expected gradient with high stability (blue line) having the lowest (highest) predicted probability of mortality (recovery), but three groups converge as FI increases, eventually high stability becoming the highest (lowest) at high levels of FI.

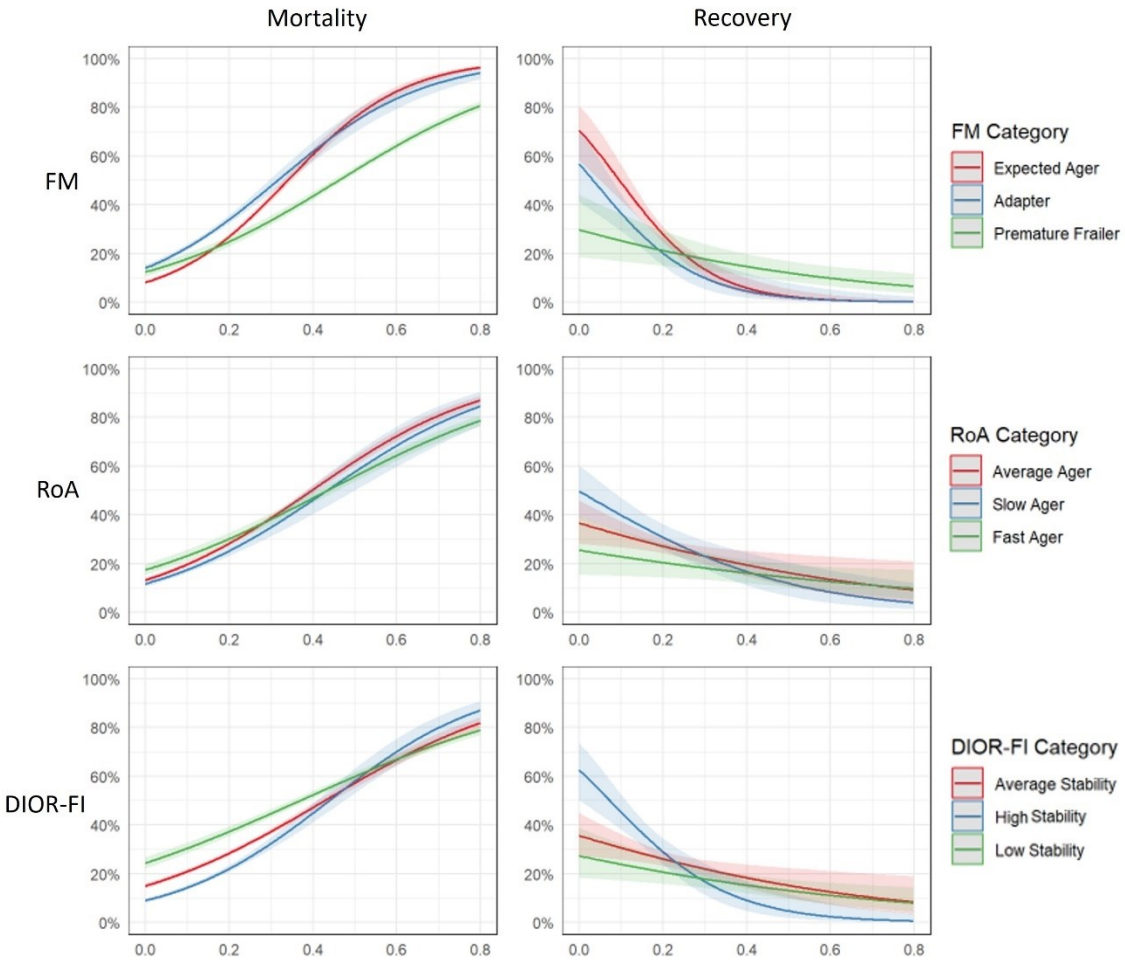


Figure 42. A Comparison of the Mortality and Recovery Analysis Results

The main results for the mortality and recovery analyses are presented-side by-side using the preferred "last three" estimation of RoA and DIOR-FI. X-axis is the level of FI. Y-axis is the predicted probability of mortality (left) or recovery (right).

In terms of AUC, the mortality and recovery models show similar results with the main difference being that the recovery models start with a lower AUC and, thus, are improved more by adding the resilience variables (Table 55).

Table 55. Comparison of Mortality and Recovery Model AUC

	Mortality			Recovery		
Model	AUC	Lower 95%	Upper 95%	AUC	Lower 95%	Upper 95%
Unadjusted/Function-Adjusted Models						
FI Only	**0.778**	0.772	0.784	0.581	0.553	0.610
ROA Only	0.753	0.747	0.758	**0.625**	0.597	0.652
FM Only	0.651	0.645	0.657	0.556	0.527	0.585
DIOR Only	0.684	0.678	0.689	0.573	0.544	0.601
Function and Frailty-Adjusted Models						
FI + ROA	**0.802**	0.796	0.807	**0.628**	0.601	0.655
FI + FM	0.801	0.796	0.806	0.618	0.591	0.646
FI + DIOR	0.780	0.775	0.786	0.604	0.576	0.632
Function, Frailty, Age and Sex-Adjusted Models						
FI Only	0.824	0.819	0.829	0.643	0.616	0.670
FI + ROA	0.824	0.819	0.829	0.647	0.620	0.674
FI + ROA Interaction	0.825	0.820	0.829	0.652	0.625	0.679
FI + FM	0.827	0.822	0.832	0.651	0.624	0.678
FI + FM Interaction	**0.828**	0.823	0.833	**0.672**	0.646	0.698
FI + DIOR	0.827	0.822	0.832	0.654	0.627	0.680
FI + DIOR Interaction	**0.828**	0.823	0.832	0.666	0.639	0.693
Final combined model	0.831	0.826	0.836	0.660	0.633	0.686

Highest AUC for each category of models is bolded, excluding the combined resilience model (bottom row). 95% confidence intervals are calculated using the Delong method. Key modelling differences between mortality and recovery are that all recovery models are adjusted for function (while mortality models are not), and sex is only adjusted for in the mortality models (insignificant in recovery models).

Chapter 7: Discussion

This thesis started with a simple question: how can we improve risk estimation in aging populations? More specifically, how can we improve risk estimation based on frailty, the most prominent concept to date used for risk assessment? To address this question, I explored the relatively recently introduced concept of physical resilience. This exploration led to the identification of diverse concepts and empirical approaches. Currently, the first major clinical studies, such as SPRING (52) and PRIME-KNEE (48), are underway supported by new conceptual frameworks. However, these frameworks focus on clinical settings in the short term and do not incorporate alternative concepts of resilience, such as resilience as adaptation. To complement these ongoing clinical studies by taking a population approach, I proposed the Integrated Stress Response Framework for Frailty and Physical Resilience (ISRF-FPR). This framework helps to understand how different concepts relate to different measurement approaches in longitudinal population data by explicitly acknowledging the longitudinal cycle of stress and the influence of frailty and resilience on this cycle. As the SPRING (52) and PRIME-KNEE (48) conceptual frameworks provide a basis for clinical investigation of physical resilience, the ISRF-FPR facilitates concurrent examinations of frailty and physical resilience using longitudinal population data.

Importantly, the ISRF-FPR is flexible and can be operationalized in many ways. The empirical demonstration in this thesis is by no means definitive, as it represents only one possibility for operationalization in one context (HRS). However, the empirical results do provide support for these methodological approaches. Overall, the key findings from the empirical work are: 1) All resilience variables (FM, RoA, and DIOR-FI) show significant unadjusted associations with mortality and recovery, supporting the idea that these can be used to indicate resilience; 2) FM requires careful evaluation and interpretation but also shows the greatest promise for adding predictive ability beyond what age, sex, and frailty can offer; 3) Continuous RoA, as currently operationalized, is the best individual predictor, surpassing even the frailty index. However, its contribution to predictive power is negligible when both frailty and age are accounted for (increase in AUC < 0.01); and 4) DIOR-FI interactions with frailty highlight a potential crossover effect, where the high stability group is the healthiest at low levels of FI (i.e., lowest predicted probability of death and highest predicted probability of full recovery), but the unhealthiest at high levels of FI (i.e., highest predicted probability of death and lowest predicted probability of

full recovery). This effect was more prominent in younger age groups and males. Taken together, these analyses demonstrate key insights and lessons learned for future analyses of frailty and physical resilience. Below I discuss the results for each of the three resilience variables and the implications of the key analytic decisions made in my analyses.

7.1 The Unexpected Results and Predictive Potential of the Frailty Disease Mismatch (FM)

Interestingly, the results for the frailty-disease mismatch do not support the previous findings by the original method authors. In Wu et al. the measure showed the expected gradient across FM categories, with the adapters, i.e., the resilient group, having the highest (lowest) indicators of good (poor) health, and the premature frailers having the lowest (highest) indicators of good (poor) health, including mortality (43). In this thesis, the adapters were older and had higher disease burden compared to the average agers (reference group), despite having similar levels of frailty. This distribution of risk factors is consistent with the concept behind the frailty-disease mismatch: individuals who are older with higher disease burden would be expected to have higher levels of frailty. If they had lower frailty than expected, they would be considered adapters. This thesis revealed a group who appeared to fit this description of an adapter but had the highest predicted probability of mortality and the lowest predicted probability of recovery. This association was robust to all sensitivity analyses, including additional post-hoc sensitivity analyses to specifically test the FM.

One key difference between this thesis and Wu et al. is the study population. Wu et al. used a sample of 2457 initially well-functioning older adults aged 70 to 79 from the Health ABC Study. In contrast, this thesis used adults aged 52 to 109 spanning all levels of function. The Health ABC Study participants were selected to have no difficulty walking a quarter mile or climbing ten steps at baseline (110). To investigate if this was the source of the discrepancy, I ran a sensitivity analysis to mirror this population within HRS by estimating FM using individuals in wave 10 aged 70 to 79 who reported no difficulty climbing one flight of stairs or walking several blocks. Using this sample to match Wu et al. still resulted in the adapters having the highest predicted probability of death. For comparison, my linear cross-sectional FM model had a coefficient of determination (R^2) of 0.27. This value is larger than the R^2 of 0.17 reported by Wu et al. for their linear model of the SAVE frailty scale.

In light of these unexpected results, I also considered modelling issues: model assumptions such as normality of the residuals and heteroscedasticity were violated. However, the purpose of this thesis was to quantify the residuals after using theoretically informed variables. Wu et al. also provided a figure of their residuals, which visually do not appear to be normally distributed but are more symmetrical than the residuals in this thesis. Alternatively, it is possible that small differences in disease burden operationalization could make a large impact. Given that this study was longitudinal and aimed to include as many waves as possible with comparable FI variables, fewer clinical diseases were used. In comparison, Wu et al. additionally included osteoporosis, kidney disease, depression, and Parkinson's disease, as well as the number of medications taken.

However, perhaps the more likely cause of this discrepancy is the use of a different frailty measure. Wu et al. used the SAVE frailty scale, a discrete 10-point scale based on the frailty phenotype (66). In contrast, this thesis used the frailty index. These two measures have different distributions and likely identify different (but related) groups of individuals as robust or frail, as previously discussed regarding the frailty phenotype and the frailty index (see section 2.1.2). Thus, it is possible that using the frailty phenotype for estimation of the FM would lead to a different distribution of risk factors between resilience categories, and subsequently result in the expected gradient of health indicators (or in the case of this thesis, result in adapters having the lowest predicted probability of death and the highest predicted probability of full recovery). A key consideration here is that the frailty phenotype reflects signs and symptoms, while the frailty index reflects multiple domains including clinical disease. I took this into consideration when creating the frailty index for this thesis: for the main analysis, disease burden variables were not included in the frailty index to prevent any strange regression results from regressing components of the FI on the FI. However, after testing multiple frailty indices, this appears to not make a difference in the behaviour of the FM.

As a final check, I further categorized FM into four categories using quartiles. Removing the bottom quartile (i.e., the most "resilient"), produced the expected association between the remaining three categories: as FM increases, so does the odds of death. Thus, it appears that those individuals who were most overestimated by the growth curve model demonstrated a different risk profile. These individuals had highly discrepant age and disease burden for their level of frailty. This still appears to fit the definition of resilience as adaptation, but it seems in the most extreme cases, the effects of age and disease burden trump the effects of frailty. In light of this,

attention to detail is necessary when implementing residual-based approaches to physical resilience. The resulting categories (or continuous variable) should be carefully assessed for a non-linear relationship with the outcome variables of interest. Future analyses should investigate these potential issues by comparing the frailty phenotype and the frailty index in the same population, comparing different operationalizations of disease burden, or perhaps using a different operationalization of stress entirely, such as a simple chronological age-biological age mismatch, as suggested by Whitson et al. (9).

Despite this abnormality, FM still consistently performed equal to or better than the other two measures in the frailty, age, and sex-adjusted models in terms of improvement in discriminatory ability. In fact, FM shows the most promise of all measures for improving AUC when estimated in age-stratified models (this is further discussed in the discriminatory ability section below).

7.2 Rate of Aging (RoA) – the Best Individual Predictor but not in Adjusted Models

RoA showed great initial promise until the restricted sample sensitivity analyses revealed that this large effect was an artifact resulting from individuals having different numbers of observations. This prompted me to change the way that both longitudinal measures (RoA and DIOR-FI) were estimated: rather than using all past estimates (i.e., derivatives for RoA and residuals for DIOR-FI) in creating these variables, I used only the last three. This ensured estimates were proximal to outcome/event of interest and comparable across all individuals. After doing this, RoA still performed quite well in the unadjusted models, being the best individual predictor among the univariate models in both the recovery and mortality analyses when the continuous variable was used, and even remained the best in the recovery analysis when the categorical variable was used.

Additionally, RoA increased the AUC by a small but non-negligible amount when added to FI. However, the benefit of adding RoA was reduced to a negligible amount (<0.01) in the frailty, age, and sex-adjusted models. Estimation of variance inflation factors for the continuous logistic regression model predictors flagged RoA and age as having problematic multicollinearity, suggesting the two measures are too highly correlated to produce reliable estimates. This supports the idea that RoA captures the aging process and thus the effect is largely diminished after accounting for age and frailty.

This collinearity could possibly be a result of the random effect structure of the mixed effect model: only the linear effect was allowed to vary by individual. All individuals were forced to have the same curvature, defined by the fixed polynomial age terms. It is possible that this did not allow for enough variability between individuals. However, I was unable to get a model with a polynomial term in the random effects to converge. If this had been possible, it would have been interesting to see how the model with random effects affected the performance of RoA. Of course, it would be possible to impose a simpler linear specification for age. In this case, the random effects would be easily estimated for every individual, allowing individuals' rates of aging to vary completely. However, this would need to be estimated in a sample with the same number of data points to avoid any bias introduced by differing lengths of follow up. This would result in an undesirable trade off: losing many repeated measures or losing many individuals. Additionally, imposing a linear specification to a non-linear phenomenon may not accurately reflect the aging process.

Though many studies have investigated frailty trajectories, only a few have examined the individual rate of change in FI and its potential in terms of mortality prediction. Bai et al. (111) examined frailty trajectories by age at death and found that most recent FI was a stronger predictor than the rate of change[18] when estimated individually, and that the rate of change was no longer significant when estimated together. Bai et al. suggested that the reason previous research found that the change in FI is independently predictive of mortality is because previous studies were only comparing to baseline frailty, not most recent. In contrast to these results, Stolz et al. (112) found that the rate of change in FI is more important than the current FI when predicting short-term mortality in a cohort of older adults aged 75 and older. The results of this thesis are more in line with Bai et al., as RoA added little beyond the most recent FI and age. Bai et al. and Stolz et al. have key methodological differences from this thesis which may explain the different results. Stolz et al. examined one and two-year survival over a 4.5-year follow-up, with data that included repeated measures every nine months. In comparison to Bai et al. and this thesis, Stolz et al.'s finding may reflect the length of time and the frequency of observations. More repeated measures over a shorter time period may be more predictive, particularly in the oldest individuals where a terminal decline may happen over a relatively short period. Additionally, it is

[18] Though these studies do not label their measure as the "rate of aging", the "rate of change" (in FI) similarly refers to the rate of deficit accumulation over time.

worth noting that all three studies used a different modelling approach: Stolz et al. used joint longitudinal and time-to-event models, Bai et al. used cox regression models, and this thesis used a two-stage approach with a longitudinal (mixed effect) model and logistic regression. This comparison highlights the importance of temporal resolution when estimating the effects of RoA. Future investigations of RoA should use data with more frequent observations than every two years.

7.3 Using Frailty Index Instability as a Dynamical Indicator of Resilience

As discussed in Sections 2.2.1 and 2.3.4, the idea behind DIORs is that variability can indicate proximity to failure and, thus, is thought to be an indicator of the resilience of the system under study. This concept has most often been operationalized at the subsystem-level in the existing literature, whereas, in this thesis, this concept was operationalized at the whole person level. To the best of my knowledge, this study is the first to apply this concept directly to the whole person level using an overall marker of health that covers multiple body systems and domains. In addition, this approach covers a much longer observation period than previous implementations.

The descriptive results of this study support the work of Stolz et al. (79), who found that FI instability increases with age and frailty and is higher among women and those who died. This study builds upon Stolz et al. to relate the instability of the frailty index to the physical resilience literature and evaluates its discriminatory ability to predict mortality and functional recovery. Unfortunately, the increase in AUC beyond what frailty and age provided was minimal. In contrast, Gijzel et al. (64), found that including multiple different DIORs increased the AUC for prediction of three-month recovery from 0.70 to 0.79, a larger increase than seen by any measure in this thesis. However, the improvement in Gijzel et al. included adding the mean and variability of multiple physical and mental responses, such as heart rate, physical activity, life satisfaction, anxiety, and discomfort (64). Additionally, looking at three-month recovery is quite different than a coarse measure of full recovery over approximately two years. Despite the lacklustre increase in discriminatory ability of the DIOR-FI developed in this thesis, the statistically significant association with recovery and mortality support the potential for repeated measures of the FI to be used in estimation of a DIOR, particularly when contrasted to the numerous other studies highlighting negative results for different variables. For example, a previous study found that variability in step count, pain, and fatigue were not associated with decline and recovery (74) (see

Section 2.3.5 for more details). Still, further research is needed to determine which variables and timescales provide the most effective dynamical indicators of resilience. Future studies may benefit from using high temporal resolution data and including multiple DIORs rather than the single DIOR-FI. In fact, Gijzel et al. advocates for linking the psychological and physiological subsystems in the estimation of resilience (47) and included the variation of such measures to obtain the AUC increase of 0.09 (64). This presents a potential opportunity for future investigations.

Lastly, in some subgroups, there was a pronounced cross-over effect of DIOR-FI at high levels of frailty. This suggests that for some groups the level of frailty matters for the interpretation of DIOR-FI resilience. For males and younger age groups (<80), the highly stable group has the highest predicted probability of death at high levels of FI. However, this effect was much less prominent in females and the oldest age group (80+). This finding should be further investigated in future studies.

7.4 Discriminatory Ability – How Much of an Increase is Meaningful?

Overall, the increase in discriminatory ability across models was quite low, with a few notable exceptions. The base FI mortality model had an AUC of 0.778, consistent with previous literature (3). Adding ROA or FM to the FI model added approximately 0.02 to the AUC. The age and sex adjusted FI model had an AUC of 0.824, and adding all resilience variables only increased the AUC by 0.007, despite statistical significance of the model predictors.

In contrast to the mortality results, the base FI recovery model had an AUC of 0.581. This was quite low, and generally considered to be in the non-informative range (7). Adding ROA or FM to the FI model added approximately 0.04 or 0.03 to the AUC, respectively. The age and sex adjusted FI model had an AUC of 0.643, and only FM remained statistically significant in the combined model with all resilience variables, with an increase of 0.017. This was larger than the mortality analysis, but still quite small.

The larger increase in AUCs for the recovery models may simply be due to a lower baseline AUC: it is easier to add predictive power when the starting point is lower. Additionally, the low AUC in the recovery analysis is likely a result of the coarse nature of the outcome variable. Though I could not find any studies specifically reporting a frailty index model AUC for functional recovery after

incident myocardial infarction, I did find one study reporting the AUC of a logistic regression model predicting death or poor functional recovery after transcatheter and surgical aortic valve replacement (113):[19] the frailty index showed an AUC of 0.74, which is relatively high in comparison to 0.581 estimated in this thesis. A key difference is that this aortic valve replacement study examined a short 6-month follow up period. This again points to the potential limitation of using data with longer time between repeated measurements.

Interestingly, the age-stratified mortality analysis showed higher increases in AUC for the frailty, age, and sex-adjusted models. The maximum increase in AUC was 0.030 for the 52-67 age group, and 0.045 for the 68-79 age group. However, this was much lower for the oldest age group, with a maximum increase of 0.007. Importantly, the FM was responsible for most of this increase in the younger age groups, and it is likely that this categorization did not have optimal cut points given the discovered non-linear association. In fact, rerunning the 68-79 age group model with the continuous variables led to the frailty, age, and sex-adjusted FM model having an AUC of 0.799, an increase of 0.064 over the age, sex, and FI model (AUC = 0.735). In terms of differences in AUC, age-stratification appears to make a difference for FM, but not so much for the other RoA and DIORs. It appears that the residuals provide more information after reducing the variability introduced by age (except in the oldest age group), potentially reflecting different effects of the disease burden variables on frailty in different age groups.

What do we make of these small differences in AUC? While there is no established standard for a minimally important difference in AUC, a few studies have addressed the question of whether small differences in AUC are meaningful. Martens et al. (114) provided a simulation study that illustrates that minimal changes in AUC (e.g., 0.01 to 0.03) correspond to minimal changes in predicted risks. The exception to this is when the AUC of the baseline model is very high (greater than 0.90), in which case there is a more substantial improvement in predicted risks. Baker et al. (115) suggested that the best way to determine whether a small change in AUC is worth including additional predictors is through a decision analysis. This requires a relevant risk threshold at which an individual would be indifferent regarding a clinical decision. In their case, they used the risk of later non-elective operative delivery at which the patient is indifferent in choosing between

[19] This paper made no mention of resilience or physical resilience and, thus, was not captured in the literature review.

proceeding with usual care or an early elective cesarian section. The authors then determined the test trade-off: the minimum number of individuals who need to take a test for an additional marker to gain one additional correct prediction. For a range of risk thresholds, the authors found that 68 to 124 women need to be tested for every additional correct prediction, which they deemed as acceptable given that the data collection is non-invasive.

This strategy is an excellent way to determine the utility of additional prediction markers, unfortunately, such risk thresholds are based on clinical decisions, and thus are not directly applicable to the current population-based analyses. However, decision analysis is highly applicable to physical resilience studies evaluating recovery after clinical stressors such as elective surgery. For interpretation of the results of this thesis, if data was routinely collected and available for clinical settings, then determination of such measures would be of relatively low cost (i.e., no additional testing would be needed), and even small increases in predictive ability may be deemed worthwhile. Future studies of frailty and physical resilience should aim to incorporate decision analysis if applicable.

I believe it is safe to assume that the increase in discrimination in the main mortality model is negligible (<0.01). The increases in the recovery model are also small, but perhaps more than negligible (0.017). On the population scale, small differences can result in a sizable number of additional correct assignments. By far the most promising results are seen when using the FM in the age stratified models. In these cases, especially when using the continuous variable, adding FM to the age, sex, and frailty model produced a reasonable increase in AUC (up to 0.064).

7.5 Using the Frailty Index as a Tool to Investigate Physical Resilience

The decision to base the analysis on the frailty index in this thesis has both conceptual and empirical implications. In addition to obvious advantages (e.g., continuous variable, strong association with mortality, and is considered by some to indicate biological age), the frailty index is a particularly relevant measure for the investigation of physical resilience. This is because there is an intrinsic link between the theory of deficit accumulation and recovery. Mitnitski et al. (30) describe a stochastic model in which deficit accumulation is the net result of environmental stress and damage control/recovery. Put simply, if the rate of damage from environmental stress exceeds the rate of recovery, then deficits will accumulate. Under this perspective, individuals

with better damage mitigation/recovery capacity (i.e., resilient individuals) would accumulate less deficits over time. Interestingly, a recent article by Farrell et al. (116) expands on this idea by using longitudinal changes in the frailty index to extract damage and repair rates from state transitions of binary deficits. They use these damage and repair rates to represent robustness and resilience, respectively. This approach appears to suggest that the processes of robustness and resilience over time determine the level of frailty, an idea that is compatible with the ISRF-FPR. This recent work, along with this thesis, illustrates the flexibility and potential of the frailty index to be used as a tool to explore and operationalize multiple different vulnerability-related concepts.

An important consideration is whether the performance of the resilience indicators in this study is tied to the performance of the frailty index. Two possibilities exist: a "worse" frailty measure could also mean a worse resilience indicator, or in contrast, a "worse" frailty measure could allow the resilience indicators to compensate for the performance of the "worse" frailty measure. When comparing different frailty indexes, the 51-item FI shows the best prediction in both the recovery and mortality samples. However, the differences between the mortality models using different FIs was quite small (<0.01), so it is difficult to discern patterns of change in the AUC of resilience indicators for the different FIs. Though in the recovery sample where the 51-item FI had a higher AUC of 0.03, the difference in the AUC of the resilience indicators was negligible. These differences may suggest that the predictive ability of the frailty index does not influence the predictive ability of the resilience indicators. This is not definitive, as this result may not hold if the difference in predictive ability of the different FIs was larger.

Despite the many reasons to advocate for the frailty index as the best option to investigate physical resilience, research implementing and comparing alternatives would offer valuable insight, such as determining whether the characteristics of the FM depend on the frailty measure. Future studies should investigate these potential issues by comparing multiple operationalizations based on different frailty measures as well as some operationalizations that are not based on frailty.

A final notable result regarding the frailty index is that the contents of the frailty index did not have a major impact on results: all conclusions remain the same whether including chronic disease or health service utilization variables in the index (Appendix L). The only difference when using different FIs was a small change in regression coefficients. Interestingly, all three FIs, based on

different sets of variables, had a 99th percentile above 0.7, the empirically observed upper limit. This violation showed a dose-response with the number of items included, with the 41-item having the highest 99th percentile, and the 56-item having the lowest. As briefly mentioned in a footnote in the results (footnote 15), this is likely due to the nature of sample selection. The mortality analysis included a large proportion of decedents, and the recovery analysis selected individuals who experienced a heart attack. In both cases, the sample was less healthy than the overall HRS sample at any one point in time. Additionally, in the mortality analysis, 89% of individuals above 0.7 were individuals who died before the next interview.

7.6 Strengths and Limitations

Aside from the large sample size in the mortality analyses, the main strengths of this study are the conceptual framework and the breadth of empirical analyses. This thesis connected and integrated numerous often discussed, but rarely synthesized, ideas and distills them into an integrated framework to aid our understanding of vulnerability in aging populations. Importantly, this allowed for multiple concepts of physical resilience and application to longitudinal population data. The ISRF-FPR guided the empirical analysis, ensuring the empirical work was grounded in theory. This is an important contribution of this thesis as previous literature pointed out the lack of an underlying theory or framework as a frequent weakness in the empirical frailty literature (13). The proposed framework is flexible and, together with the empirical work, offers guidance for future investigations. In terms of the breadth of analysis, this study compared the FI with three resilience indicators (i.e., RoA, FM, DIOR-FI) and evaluated two outcomes (i.e., mortality and recovery). To the best of my knowledge, this is the first study to provide such a comprehensive empirical comparison of frailty and physical resilience. Furthermore, I included several sensitivity analyses to evaluate the robustness of these results. Together, these empirical results combined with the ISRF-FPR represent a comprehensive body of work that sets the stage for future research by generating new insights and providing guidance for the concurrent investigation of frailty and resilience in longitudinal population data.

The main limitations of this study are the lack of statistical power in the recovery analyses and the use of biennial data for operationalizing recovery and estimating variability (DIOR-FI).

First, the limited statistical power in the recovery analyses resulted in imprecise estimation of coefficients and did not allow for subgroup analyses (i.e., age- and sex-stratified analyses). This is likely the result of the smaller sample size (compared to the mortality analysis) and the nature of the predictor and response variables. Having unbalanced categorical predictors and response variables combined with a smaller sample size led to inadequate representation for a specific group. Partitioning of the sample for individual groups requires larger sample sizes to detect effects. For example, having a three-category predictor and a two-category outcome effectively splits the sample into six groups. If the categories are unbalanced, then the group representing the intersection of the smallest predictor and outcome categories may have a very low sample size. Additionally, the coarseness of the response variable may make prediction difficult. For example, it may be more difficult to predict the coarse binary outcome of full recovery than the amount of function regained over a shorter period. We may thus be limited to small effect sizes given the suboptimal operationalization of recovery.

Second, biennial data collection in HRS limits the ascertainment of recovery. Given the temporal resolution of the data, this study cannot determine how much functional decline and subsequent recovery had taken place. Rather, the only certainty was how the individual's function had changed at subsequent data collections. Thus, given the limitations in the data, the most accurate (though restrictive) benchmark of recovery was return to baseline (i.e., full recovery). This was a coarse operationalization that did not allow for differentiation among those who did not return to baseline. There is likely informative heterogeneity among those who did not fully recover. For example, this study put someone who recovered 75% of lost function and someone who recovered 10% into the same category despite the two likely having different levels of physical resilience. Additionally, reports of myocardial infarction were taken from the surviving respondent or captured in an exit interview with a spouse or relative if the respondent died. HRS obtains exit interviews for approximately 80% of deaths, so it is likely that some of the least resilient individuals were not captured (i.e., those who had an MI and died, but no exit interview was provided). This likely resulted in a healthier sample, potentially underestimating the true association between the independent variables and recovery. Additionally, HRS lacks information on the severity of the MI. To try to reduce variability in predisposing factors and severity, I limited my sample to first-time MIs. Despite this attempt, there is likely still significant unaccounted variation in severity in the sample. Individuals with a more severe MI would be less likely to fully

recover, and it is not possible to disentangle whether the resilience indicators are associated with recovery or whether they are associated with disease severity.

Third, biennial data collection in HRS may not provide adequate temporal resolution to accurately capture variability to estimate DIOR-FI. There could be significant variation in the two-year period between observations which is not captured. Given the positive association between frailty and instability observed in this thesis and previous work by Stolz et al. (79), it is reasonable to expect that it is more likely for unobserved fluctuations to occur in frail individuals in between waves. This could lead to systematic measurement error where frail individuals' variability may be underestimated, potentially resulting in the underestimation of the association of the DIOR-FI with recovery and mortality. Additionally, estimates of variability would be improved not only by reducing the time between measurements, but also increasing the number of repeated measures used. Only three time points were used in this study (after correcting for the different number of observations), giving a coarse approximation of the true variability. Future studies should aim to overcome these limitations by using data with more frequent observations and additional measures of stressor severity, and a more sensitive measure of recovery.

In addition to these main limitations, there are four further considerations worth noting. These include: the possibility of selection bias introduced by requiring individuals to have three repeated measures, variability in the age range used to estimate RoA and DIOR-FI, the choice of analytical model and approach used (i.e., logistic regression in a two-stage approach), and the use of categorical variables over continuous for the primary analyses. I elaborate on each of these points below.

First, by selecting individuals with at least three FI observations, it is possible the sample systematically excluded individuals who were too sick to complete at least three interviews (or were too sick to answer at least 80% of FI variable questions). To assess the potential impact of using this inclusion criterion, I compared the characteristics of those excluded from the sample to those included to see if they differ. Compared to those included in the sample, those excluded had a lower median age (56 vs. 74), a lower median frailty index (0.122 vs. 0.183), a lower proportion of females (52% vs. 57%), and a similar proportion of deaths (41% vs. 40%). These differences are unexpectedly in the opposite direction, with those excluded being younger and less frail, despite having a similar proportion of deaths. This result is likely due to the rolling

enrollment in HRS: newly recruited younger and healthier individuals who have not been present for three waves were also excluded in addition to those who were too unhealthy to continue.

Second, differences in the age range in the last three waves used for both DIOR-FI and RoA could lead to some imprecision in estimates. For example, though HRS data collections are on a two-year cycle, not all interviews are spaced by exactly two years. This is illustrated by the example of the youngest individual included in the sample being age 52 – this individual was interviewed three times between the age of 50 and 52. If interviews had taken place exactly two years apart, this individual should have been age 54 in the third interview. This is an example where the first interview was very late in the cycle, and the third interview was very early in the cycle, resulting in a timespan of less than three years for their last three interviews. However, I would not expect there to be any systematic association between proximity of interviews and other variables, so overall I would expect this non-systematic measurement error to dilute the results towards the null rather than systematically bias them in any particular direction.

Additionally, after finding the artefact with the number of observations, I chose to limit the estimation of RoA and DIOR to the last three derivatives/residuals, while still using the full sample. The choice of using the full sample rather than the restricted sensitivity sample means that some of the individuals in the full sample (approximately 15% - the amount dropped to create the restricted sample) did not have three consecutive final interviews. This means that the three values used to estimate their RoA and DIOR-FI did not come from the last three interviews, but rather cover a timespan of at least four interviews. For example, if an individual died in wave 10, but missed wave 8, they are still included in the sample. However, their RoA and DIOR-FI are estimated based on their derivatives/residuals from waves 6, 7 and 9 (approximately a 6-year window) rather than 7, 8 and 9 (approximately a 4-year window) if they had been observed every wave. The benefit to this approach is that we do not lose individuals who missed an interview. The drawback, however, is that the age range for estimating RoA and DIOR-FI will be systematically longer than those with three consecutive measures (the remaining 85%). I would expect longer estimation periods to underestimate RoA and DIOR-FI (compared to shorter periods), as both rate of change and variability increase with age, and having a longer estimation period means estimates at younger ages are being included in the creation of RoA and DIOR-FI. To check the impact of non-consecutive observations I compared the restricted sample (using all longitudinal data in the growth curve model) to the full sample using only the last three

observations to estimate RoA and DIOR-FI (see Figure 21, section 6.1.9). The results were highly consistent in the two samples, suggesting that including the 15% of individuals with non-consecutive observations does not impact the results. Thus, I opted to retain them in the analysis. In summary, while non-consecutive observations and variation in the age range used to estimate RoA and DIOR-FI is a concern that needs to be evaluated, it does not appear to have influenced the results presented.

Third, the choice of logistic regression over survival models has important implications. I chose logistic regression to compare the most recently estimated resilience profiles among those who died vs. those who did not, and those who full recovered after incident MI vs. those who did not. Importantly, this approach does not allow for the investigation of within-person changes in resilience over time. In contrast, survival models offer the advantages of retaining the information from individuals lost to follow-up and accommodating time-varying covariates. However, incorporating longitudinal estimates of stability in survival models is less straight forward. Thus, to compare the three measures operationalized in this study I opted for the simpler logistic models as a proof of concept. Future studies should explore the opportunity to expand upon the current work by implementing survival models or other modelling approaches that allow estimation of the effects of within-person change in indicators of physical resilience.

Additionally, the choice of a two-stage analytic approach (i.e., estimating separate longitudinal and logistic models) has its drawbacks compared to joint modelling approaches (i.e., estimating both models simultaneously). I chose the two-stage analytic approach so that I could estimate and assess the predictive ability of the frailty index and the three resilience indicators together. While joint models exist for analyzing longitudinal and time to event data, or longitudinal and variability models (e.g., the location-scale model used by Stolz et al. (79)), to my knowledge, there is no joint modelling framework that can accomplish the goal of this study: estimating the effects of FI, RoA, FM, and DIOR-FI on mortality/recovery using a single model. Joint modelling approaches have the advantage of improving accuracy of statistical inference by retaining uncertainty in estimates and accounting for the correlation structure between survival and longitudinal outcomes. However, the empirical analysis in this thesis is meant to be exploratory, not definitive. What the empirical results in this thesis show is a proof of concept. The three resilience indicators are clearly indicating what they were hypothesized to. The exact coefficients estimated would of course vary depending on the relationship between variables in the chosen

data. The empirical analyses presented in this thesis will serve as a foundation for future research on frailty and physical resilience.

Fourth, this thesis used somewhat arbitrary categorizations of the resilience indicators. Cut points in this thesis were determined simply based on the non-normal distribution of the variables and were not optimized for sensitivity or specificity. This limitation is not specific to this study; it is a recurring theme in the physical resilience literature. Many studies use plus or minus one standard deviation to define three resilience categories. Given that the resilience variables in this study are not normally distributed, I used the non-parametric equivalent of the 25th and 75th percentiles to try to remain consistent with this previous work. Results from the alternative cut points sensitivity analysis (top and bottom 15%) remained consistent with the main analysis, suggesting that the observed associations are robust to this somewhat arbitrary decision (Appendix J). The benefits of categorization include a meaningful comparison between groups and accommodation of non-linear effects (as seen with FM). However, categorization suffers from the loss of potentially informative heterogeneity, illustrated by the higher AUC values in the continuous sensitivity analyses (Appendix K). An interesting avenue for future research would be to characterize minimal important differences in these continuous measures. This would allow for the interpretation of meaningful results without the need for categorization.

7.7 Implications

This thesis offers a comprehensive conceptual and empirical comparison of frailty and physical resilience, supported by a flexible framework suitable for the use of longitudinal population data. The empirical work is an exploratory investigation of different methods and provides one operationalization of the ISRF-FPR based on the cumulative deficit model. Many other opportunities for operationalization exist and should be explored in future analyses to further elucidate the empirical relationship between the concepts of frailty and physical resilience. Future studies can take the lessons learned from this thesis to improve methodology for more effective operationalization of resilience indicators. Promising avenues for future research are listed below.

1. Repeat the empirical analyses presented in this thesis using data with shorter time between observations. Using such data should improve estimates of RoA and DIOR-FI, particularly among older individuals.

2. Investigate whether stratifying by age and sex to examine resilience indicators would have an effect on AUCs (rather than always age- and sex-stratify examination of resilience indicators, particularly for FM and DIOR-FI).
3. Investigate FM with different variables (e.g., frailty phenotype) and the FI in different datasets to provide insights about the unexpected results in this thesis.
4. Compare the use of the frailty index, as a whole-person level measure of reserve, in this thesis, to the use of multiple specific measures of reserve (e.g., measures of function in different subsystems). Explore the possibility to measure reserve at the subsystems level in population surveys.
5. Explore the effect of alternative modelling strategies, such as time-to-event models instead of logistic regression and joint models, instead of two-stage approaches. Modelling strategies that can accommodate within-person change in resilience will be beneficial.

7.8 Conclusion

Exploring the world of resilience offers exciting opportunities to better understand the dynamics of health in aging populations and leads to many interesting methodological and philosophical questions. This thesis builds the foundation upon which we can further explore the relationship between frailty and physical resilience and various measurement approaches using population-level data. Results of this thesis demonstrate that resilience indicators have potential to improve predictive ability over what age and frailty can offer and further investigation is warranted. In particular, continuous FM shows the most promise in age-stratified mortality models, with an improvement in AUC of 0.064 when added to the age, sex, and frailty-adjusted model for the 68-79 age group. Overcoming methodological challenges, performance of RoA and DIOR-FI may also improve. With further refinement of the methods proposed in this thesis, the combination of population data (for estimating FM) and routinely collected health data (for estimating RoA and DIOR-FI) offer promising opportunities to improve risk estimation in aging populations. Ultimately, this avenue of research can help identify vulnerable groups and individuals, which can in turn support health policy, beyond clinical settings. Furthermore, if successful, there is potential to integrate these longitudinal and population-based approaches to improve individual risk assessment in clinical settings.

REFERENCES

1. Morley JE, Haren MT, Rolland Y, Kim MJ. Frailty. Med Clin North Am. 2006;90(5):837–47.

2. Theou O, Squires E, Mallery K, Lee JS, Fay S, Goldstein J, et al. What do we know about frailty in the acute care setting? A scoping review. BMC Geriatr. 2018;18(1):139.

3. Theou O, Brothers TD, Mitnitski A, Rockwood K. Operationalization of frailty using eight commonly used scales and comparison of their ability to predict all-cause mortality. J Am Geriatr Soc. 2013;61(9):1537–51.

4. op het Veld LPM, Beurskens AJHM, de Vet HCW, van Kuijk SMJ, Hajema K, Kempen GIJM, et al. The ability of four frailty screening instruments to predict mortality, hospitalization and dependency in (instrumental) activities of daily living. Eur J Ageing. 2019;16(3):387–94.

5. Gonzalez-Colaço Harmand M, Meillon C, Bergua V, Tabue Teguo M, Dartigues JF, Avila-Funes JA, et al. Comparing the predictive value of three definitions of frailty: Results from the Three-City study. Arch Gerontol Geriatr. 2017;72:153–63.

6. Si H, Jin Y, Qiao X, Tian X, Liu X, Wang C. Predictive performance of 7 frailty instruments for short-term disability, falls and hospitalization among Chinese community-dwelling older adults: A prospective cohort study. Int J Nurs Stud. 2021;117:103875–103875.

7. de Hond AA, Steyerberg EW, van Calster B. Interpreting area under the receiver operating characteristic curve. Lancet Digit. 2022;4(12):e853–5.

8. Andrew M, Searle SD, McElhaney JE, McNeil SA, Clarke B, Rockwood K, et al. COVID-19, frailty and long-term care: Implications for policy and practice. J Infect Dev Ctries. 2020;14(05):428–32.

9. Whitson HE, Duan-Porter W, Schmader KE, Morey MC, Cohen HJ, Colón-Emeric CS. Physical resilience in older adults: Systematic review and development of an emerging construct. J Gerontol A Biol Sci Med Sci. 2016;71(4):489–95.

10. Chhetri JK, Xue QL, Ma L, Chan P, Varadhan R. Intrinsic capacity as a determinant of physical resilience in older adults. J Nutr Health Aging. 2021;25(8):1006–11.

11. Whitson HE, Cohen HJ, Schmader K, Morey MC, Kuchel G, Colon-Emeric C. Physical resilience: Not simply the opposite of frailty. J Am Geriatr Soc. 2018;66(8):1459–61.

12. Varadhan R, Walston JD, Bandeen-Roche K. Can a link be found between physical resilience and frailty in older adults by studying dynamical systems? J Am Geriatr Soc. 2018;66(8):1455–8.

13. Kuchel GA. Frailty and resilience as outcome measures in clinical trials and geriatric care: Are we getting any closer? J Am Geriatr Soc. 2018;66(8):1451–4.

14. Rikkert M, Melis R. Rerouting geriatric medicine by complementing static frailty measures with dynamic resilience indicators of recovery potential. Front Physiol. 2019;10:1–9.

15. Vaupel JW, Manton KG, Stallard E. The impact of heterogeneity in individual frailty on the dynamics of mortality. Demography. 1979;16(3):439–54.

16. Rockwood K, Howlett SE. Age-related deficit accumulation and the diseases of ageing. Mech Ageing Dev. 2019;180:107–16.

17. Afilalo J. Conceptual models of frailty: The sarcopenia phenotype. Can J Cardiol. 2016;32(9):1051–5.

18. Rockwood K. Conceptual models of frailty: Accumulation of deficits. Can J Cardiol. 2016;32(9):1046–50.

19. Marzetti E, Calvani R, Tosato M, Cesari M, Di Bari M, Cherubini A, et al. Sarcopenia: An overview. Aging Clin Exp Res. 2017;29:11–7.

20. Searle SD, Mitnitski A, Gahbauer EA, Gill TM, Rockwood K. A standard procedure for creating a frailty index. BMC Geriatr. 2008;8(1):1–10.

21. Rockwood K MD, FRCPC, FRCP, Mitnitski A PhD. Frailty defined by deficit accumulation and geriatric medicine defined by frailty. Clin Geriatr Med. 2011;27(1):17–26.

22. Cesari M, Gambassi G, van Kan GA, Vellas B. The frailty phenotype and the frailty index: Different instruments for different purposes. Age Ageing. 2014;43(1):10–2.

23. Morley JE, Vellas B, Van Kan GA, Anker SD, Bauer JM, Bernabei R, et al. Frailty consensus: A call to action. J Am Med Dir Assoc. 2013;14(6):392–7.

24. Dent E, Kowal P, Hoogendijk EO. Frailty measurement in research and clinical practice: A review. Eur J Intern Med. 2016;31:3–10.

25. Fried LP, Tangen CM, Walston J, Newman AB, Hirsch C, Gottdiener J, et al. Frailty in older adults: Evidence for a phenotype. J Gerontol A Biol Sci Med Sci. 2001;56(3):M146–57.

26. Rockwood K, Mitnitski A. Limits to deficit accumulation in elderly people. Mech Ageing Dev. 2006;127(5):494–6.

27. Karasik D, Demissie S, Cupples LA, Kiel DP. Disentangling the genetic determinants of human aging: Biological age as an alternative to the use of survival measures. J Gerontol A Biol Sci Med Sci. 2005;60(5):574–87.

28. Belsky DW, Caspi A, Houts R, Cohen HJ, Corcoran DL, Danese A, et al. Quantification of biological aging in young adults. Proc Natl Acad Sci U S A. 2015;112(30):E4104–10.

29. Wagner KH, Cameron-Smith D, Wessner B, Franzke B. Biomarkers of aging: From function to molecular biology. Nutrients. 2016;8(6):1–12.

30. Mitnitski A, Song X, Rockwood K. Assessing biological aging: the origin of deficit accumulation. Biogerontology. 2013;14(6):709–17.

31. Mitnitski A, Howlett SE, Rockwood K. Heterogeneity of human aging and its assessment. J Gerontol A Biol Sci Med Sci. 2017;72(7):877–84.

32. Kim S, Myers L, Wyckoff J, Cherry KE, Jazwinski SM. The frailty index outperforms DNA methylation age and its derivatives as an indicator of biological age. Geroscience. 2017;39(1):83–92.

33. Li X, Ploner A, Wang Y, Magnusson PK, Reynolds C, Finkel D, et al. Longitudinal trajectories, correlations and mortality associations of nine biological ages across 20-years follow-up. Elife. 2020;9:1–20.

34. Blodgett J, Theou O, Kirkland S, Andreou P, Rockwood K. Frailty in NHANES: Comparing the frailty index and phenotype. Arch Gerontol Geriatr. 2015;60(3):464–70.

35. Kulminski AM, Ukraintseva SV, Kulminskaya IV, Arbeev KG, Land K, Yashin AI. Cumulative deficits better characterize susceptibility to death in elderly people than phenotypic frailty: Lessons from the Cardiovascular Health Study. J Am Geriatr Soc. 2008;56(5):898–903.

36. Watanabe D, Yoshida T, Yamada Y, Watanabe Y, Yamada M, Fujita H, et al. Combined use of two frailty tools in predicting mortality in older adults. Sci Rep. 2022;12(1):1–9.

37. Gijzel SM, van de Leemput IA, Scheffer M, Roppolo M, Olde Rikkert MG, Melis RJ. Dynamical resilience indicators in time series of self-rated health correspond to frailty levels in older adults. J Gerontol A Biol Sci Med Sci. 2017;72(7):991–6.

38. Resnick B, Galik E, Dorsey S, Scheve A, Gutkin S. Reliability and validity testing of the physical resilience measure. Gerontologist. 2011;51(5):643–52.

39. Ukraintseva S, Yashin AI, Arbeev KG. Resilience versus robustness in aging. J Gerontol A Biol Sci Med Sci. 2016;71(11):1533–4.

40. Whitson HE, Duan-Porter W, Schmader K, Morey M, Cohen HJ, Colón-Emeric C. Response to Ukraintseva et al. letter: Resilience versus robustness in aging. J Gerontol A Biol Sci Med Sci. 2016;71(11):1535–6.

41. Hadley EC, Kuchel GA, Newman AB. Report: NIA workshop on measures of physiologic resiliencies in human aging. J Gerontol A Biol Sci Med Sci. 2017;72(7):980–90.

42. Scheffer M, Bolhuis JE, Borsboom D, Buchman TG, Gijzel SMW, Goulson D, et al. Quantifying resilience of humans and other animals. Proc Natl Acad Sci U S A. 2018;115(47):11883–90.

43. Wu C, Li YX, Marron MM, Odden MC, Newman AB, Sanders JL. Quantifying and classifying physical resilience among older adults: The Health, Aging, and Body Composition Study. J Gerontol A Biol Sci Med Sci. 2019;75(10):1960–6.

44. Colón-Emeric C, Pieper CF, Schmader KE, Sloane R, Bloom A, McClain M, et al. Two approaches to classifying and quantifying physical resilience in longitudinal data. J Gerontol A Biol Sci Med Sci. 2020;75(4):731–8.

45. Witham MD, Sayer AA. Biological resilience in older people – a step beyond frailty? Eur Geriatr Med. 2015 Apr 1;6(2):101–2.

46. Rockwood K, Mitnitski A. Resilience and frailty: Further steps, best taken together. Eur Geriatr Med. 2015;5(6):405–7.

47. Gijzel SM, Whitson HE, van de Leemput IA, Scheffer M, van Asselt D, Rector JL, et al. Resilience in clinical care: Getting a grip on the recovery potential of older adults. J Am Geriatr Soc. 2019;67(12):2650–7.

48. Whitson HE, Crabtree D, Pieper CF, Ha C, Au S, Berger M, et al. A template for physical resilience research in older adults: Methods of the PRIME-KNEE study. J Am Geriatr Soc. 2021;69(11):3232–41.

49. Hamaker M, Gijzel S, Rostoft S, van den Bos F. Intrinsic capacity and resilience: Taking frailty to the next level. J Geriatr Oncol. 2023;14(2):101421.

50. Cesari M, Azzolino D, LeBrasseur N, Whitson H, Rooks D, Sourdet S, et al. Resilience: Biological basis and clinical significance — a perspective report from the International Conference on Frailty and Sarcopenia Research (ICFSR) Task Force. J Frailty Aging. 2022;11(4):342–7.

51. Colon-Emeric C, Schmader K, Cohen HJ, Morey M, Whitson H. Ageing and physical resilience after health stressors. Stress Health. 2023;39(S1):48–54.

52. Walston J, Varadhan R, Xue Q, Buta B, Sieber F, Oni J, et al. A study of physical resilience and aging (SPRING): Conceptual framework, rationale, and study design. J Am Geriatr Soc. 2023;71(8):2393–405.

53. Chhetri JK, Ma L, Chan P. Physical resilience: A novel approach for healthy aging. J Frailty Sarcopenia Falls. 2022;7(1):29.

54. Frolova E, Arosio B, Lim WS. Intrinsic capacity and resilience vs. frailty: On the way to healthy aging. Front Med. 2023;10:1155648.

55. World Health Organization. World report on ageing and health [Internet]. World Health Organization; 2015. 246 p. Available from: https://iris.who.int/handle/10665/186463.

56. Cesari M, Araujo de Carvalho I, Amuthavalli Thiyagarajan J, Cooper C, Martin FC, Reginster JY, et al. Evidence for the domains supporting the construct of intrinsic capacity. J Gerontol A Biol Sci Med Sci. 2018;73(12):1653–60.

57. Belloni G, Cesari M. Frailty and intrinsic capacity: Two distinct but related constructs. Front Med. 2019;6:1–5.

58. Calle A, Onder G, Morandi A, Bellelli G, Ortolani E, Pérez L, et al. Frailty related factors as predictors of functional recovery in geriatric rehabilitation: The sarcopenia and function in aging rehabilitation (SAFARI) multi-centric study. J Nutr Health Aging. 2018;22(9):1099–106.

59. Colón-Emeric C, Whitson HE, Pieper CF, Sloane R, Orwig D, Huffman KM, et al. Resiliency groups following hip fracture in older adults. J Am Geriatr Soc. 2019;67(12):2519–27.

60. Hladek MD, Zhu J, Crews DC, McAdams-DeMarco MA, Buta B, Varadhan R, et al. Physical resilience phenotype trajectories in incident hemodialysis: Characterization and mortality risk assessment. Kidney Int Rep. 2022;7(9):2006–15.

61. Presley CJ, Arrato NA, Shields PG, Carbone DP, Wong ML, Benedict J, et al. Functional trajectories & resilience among adults with advanced lung cancer. JTO Clin Res Rep. 2022;3(6):1–11.

62. Pedone C, Costanzo L, Finamore P, Bandinelli S, Ferrucci L, Antonelli Incalzi R. Defining resilience in older people: does a subjective definition of stressor work? J Gerontol A Biol Sci Med Sci. 2021;76(8):1480–5.

63. Duan-Porter W, Cohen HJ, Demark-Wahnefried W, Sloane R, Pendergast JF, Snyder DC, et al. Physical resilience of older cancer survivors: an emerging concept. J Geriatr Oncol. 2016;7(6):471–8.

64. Koivunen K, Löppönen A, Palmberg L, Rantalainen T, Rantanen T, Karavirta L. Autonomic nervous system and postural control regulation during orthostatic test as putative markers of physical resilience among community-dwelling older adults. Exp Gerontol. 2023;182:112292.

65. Sanders JL, Arnold AM, Hirsch CH, Thielke SM, Kim D, Mukamal KJ, et al. Effects of disease burden and functional adaptation on morbidity and mortality on older adults. J Am Geriatr Soc. 2016;64(6):1242–9.

66. Sanders JL, Boudreau RM, Fried LP, Walston JD, Harris TB, Newman AB. Measurement of organ structure and function enhances understanding of the physiological basis of frailty: the Cardiovascular Health Study. J Am Geriatr Soc. 2011;59(9):1581–8.

67. Wu C, Lin TZ, Sanders J. A simplified approach for classifying physical resilience among community-dwelling older adults: The Health, Aging, and Body Composition Study. J Frailty Aging. 2022;11:281–5.

68. Zhang H, Hao M, Li Y, Hu Z, Liu Z, Jiang S, et al. Assessment of Physical Resilience Using Residual Methods and Its Association With Adverse Outcomes in Older Adults. Innov Aging. 2023;7(9):igad118.

69. Milman S, Lerman B, Ayers E, Zhang Z, Sathyan S, Levine M, et al. Frailty Resilience Score: A Novel Measure of Frailty Resilience Associated with Protection from Frailty and Survival. J Gerontol A Biol Sci Med Sci. 2023;78(10):1771–7.

70. Sotos-Prieto M, Ortolá R, López-García E, Rodríguez-Artalejo F, García-Esquinas E. Adherence to the Mediterranean diet and physical resilience in older adults: The Seniors-ENRICA Cohort. J Gerontol A Biol Sci Med Sci. 2021;76(3):505–12.

71. García-Esquinas E, Ortolá R, Prina M, Stefler D, Rodríguez-Artalejo F, Pastor-Barriuso R. Trajectories of accumulation of health deficits in older adults: are there variations according to health domains? J Am Med Dir Assoc. 2019;20(6):710–7.

72. Gijzel SM, van de Leemput IA, Scheffer M, van Bon GE, Weerdesteyn V, Eijsvogels TM, et al. Dynamical indicators of resilience in postural balance time series are related to successful aging in high-functioning older adults. J Gerontol A Biol Sci Med Sci. 2019;74(7):1119–26.

73. Gijzel SM, Rector J, van Meulen FB, van Der Loeff RS, van de Leemput IA, Scheffer M, et al. Measurement of dynamical resilience indicators improves the prediction of recovery following hospitalization in older adults. J Am Med Dir Assoc. 2020;21(4):525–30.

74. Kolk D, Melis RJ, MacNeil-Vroomen JL, Buurman BM, Reichardt LA, Aarden J, et al. Physical resilience in daily functioning among acutely ill hospitalized older adults: The Hospital-ADL Study. J Am Med Dir Assoc. 2022;23(5):903.e1-903.e12.

75. Rector JL, Gijzel SM, van de Leemput IA, van Meulen FB, Rikkert MGO, Melis RJ. Dynamical indicators of resilience from physiological time series in geriatric inpatients: Lessons learned. Exp Gerontol. 2021;149:1–7.

76. Lucas A, Rutledge J, Sloane R, Hall K, Green C, Pieper C, et al. Physical activity is a potential measure of physical resilience in older adults receiving hemodialysis. Front Nephrol. 2023;2:1032468.

77. Zhu Y, Chen X, Geng S, Li Q, Yuan H, Zhou X, et al. Association between ambulatory blood pressure variability and frailty among older hypertensive patients. J Clin Hypertens (Greenwich). 2020;22(9):1703–12.

78. Rouch L, De Souto Barreto P, Hanon O, Vidal JS, Amar J, Andrieu S, et al. Visit-to-visit blood pressure variability and incident frailty in older adults. J Gerontol A Biol Sci Med Sci. 2021;76(8):1369–75.

79. Stolz E, Mayerl H, Freidl W. Fluctuations in frailty among older adults. Age Ageing. 2019;48(4):547–52.

80. Lin J, Kelley-Moore JA. From noise to signal: the age and social patterning of intra-individual variability in late-life health. J Gerontol B Psychol Sci Soc Sci. 2017;72(1):168–79.

81. Park J, Lee J, Lee H, Kim S, Kim CO, Park CG. Physical resilience as a moderator of the relationship between frailty and disability in older adults with osteoarthritis. J Adv Nurs. 2022;78(7):2085–94.

82. Gobbens RJ, Boersma P, Uchmanowicz I, Santiago LM. The Tilburg Frailty Indicator (TFI): New evidence for its validity. Clin Interv Aging. 2020;15:265–74.

83. Hu FW, Lin CH, Lai PH, Lin CY. Predictive validity of the Physical Resilience Instrument for Older Adults (PRIFOR). J Nutr Health Aging. 2021;25(9):1042–5.

84. Rockwood K, Song X, MacKnight C, Bergman H, Hogan DB, McDowell I, et al. A global clinical measure of fitness and frailty in elderly people. CMAJ. 2005;173(5):489–95.

85. Rockwood K, Theou O. Using the clinical frailty scale in allocating scarce health care resources. Can Geriatr J. 2020;23(3):210–5.

86. Arbeev KG, Ukraintseva SV, Bagley O, Zhbannikov IY, Cohen AA, Kulminski AM, et al. “Physiological Dysregulation” as a promising measure of robustness and resilience in studies of aging and a new indicator of preclinical disease. J Gerontol A Biol Sci Med Sci. 2019;74(4):462–8.

87. Galvin A, Ukraintseva S, Arbeev K, Feitosa M, Christensen K. Physical robustness and resilience among long-lived female siblings: A comparison with sporadic long-livers. Aging. 2020;12(14):15157–68.

88. Galvin A, Feitosa M, Arbeev K, Kuipers AL, Wojczynski M, Ukrainsteva S, et al. Physical resilience after a diagnosis of cardiovascular disease among offspring of long-lived siblings. Eur J Ageing. 2021;1–9.

89. Ukraintseva S, Arbeev K, Duan M, Akushevich I, Kulminski A, Stallard E, et al. Decline in biological resilience as key manifestation of aging: Potential mechanisms and role in health and longevity. Mech Ageing Dev. 2021;194:111418.

90. Sonnega A, Faul JD, Ofstedal MB, Langa KM, Phillips JW, Weir DR. Cohort profile: the health and retirement study (HRS). Int J Epidemiol. 2014;43(2):576–85.

91. Weir DR. Validating mortality ascertainment in the health and retirement study. Ann Arbor, Michigan: Survey Research Center, Institute for Social Research, University of Michigan; 2016 p. 1–18.

92. Heeringa SG, Connor JH. Technical description of the Health and Retirement Survey sample design. Ann Arbor, Michigan: Sampling Section, Institute for Social Research, University of Michigan; 1995 p. 1–59.

93. Lee S, Nishimura R, Burton P, McCammon R. HRS 2016 Sampling Weights [Internet]. 2021. Available from: https://hrs.isr.umich.edu/sites/default/files/biblio/HRS2016SamplingWeights.pdf.

94. Curran PJ, Obeidat K, Losardo D. Twelve frequently asked questions about growth curve modeling. J Cogn Dev. 2010;11(2):121–36.

95. McNeish D, Matta T. Differentiating between mixed-effects and latent-curve approaches to growth modeling. Behav Res Methods. 2018;50:1398–414.

96. Peugh JL. A practical guide to multilevel modeling. J Sch Psychol. 2010;48(1):85–112.

97. Blodgett J. The association between sedentary behaviour, moderate–vigorous physical activity and frailty. [Halifax, Nova Scotia]: Dalhousie University; 2014.

98. Theou O, Haviva C, Wallace L, Searle SD, Rockwood K. How to construct a frailty index from an existing dataset in 10 steps. Age Ageing. 2023;52(12):afad221.

99. Eendebak R, Theou O, van der Valk A, Godin J, Andrew M, McNeil S, et al. Defining minimal important differences and establishing categories for the frailty index. Innov Aging. 2018 Nov 1;2(suppl_1):715–6.

100. Theou O, van der Valk AM, Godin J, Andrew MK, McElhaney JE, McNeil SA, et al. Exploring clinically meaningful changes for the frailty index in a longitudinal cohort of hospitalized older patients. J Gerontol A Biol Sci Med Sci. 2020;75(10):1928–34.

101. Hoover M, Rotermann M, Sanmartin C, Bernier J. Validation of an index to estimate the prevalence of frailty among community-dwelling seniors. Health Rep. 2013;24(9):10–7.

102. Wei MY, Kabeto MU, Langa KM, Mukamal KJ. Multimorbidity and physical and cognitive function: Performance of a new multimorbidity-weighted index. J Gerontol A Biol Sci Med Sci. 2018;73(2):225–32.

103. Ware J, Snow K, Kosinski M, Gandek B. SF-36 health survey manual and interpretation guide. 1993.

104. Wickham H. ggplot2: Elegant graphics for data analysis [Internet]. New York: Springer-Verlag; 2016. Available from: https://ggplot2.tidyverse.org/.

105. Lüdecke D. ggeffects: Tidy data frames of marginal effects from regression models. J Open Source Softw. 2018;3(26):772.

106. Sjoberg DD, Whiting K, Curry M, Lavery JA, Larmarange J. Reproducible summary tables with the gtsummary package. R J. 2021;13(1):570–80.

107. Pinheiro J, Bates D, DebRoy S, Sarkar D. Nonlinear mixed-effects models. R package version. 2012;3:1–89.

108. Canadian Institutes of Health Research, Natural Sciences and Engineering, Research Council of Canada, Social Sciences and Humanities Research. Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans. 2018.

109. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;159–74.

110. National Institute on Aging. Introducing the Health ABC Study: The dynamics of health, aging, and body composition [Internet]. [cited 2023 Oct 1]. Available from: https://healthabc.nia.nih.gov/.

111. Bai G, Szwajda A, Wang Y, Li X, Bower H, Karlsson IK, et al. Frailty trajectories in three longitudinal studies of aging: Is the level or the rate of change more predictive of mortality? Age Ageing. 2021;50(6):2174–82.

112. Stolz E, Mayerl H, Hoogendijk EO. Frailty in the oldest old: Is the current level or the rate of change more predictive of mortality? Age Ageing. 2022;51(2):afac020.

113. Shi S, Afilalo J, Lipsitz LA, Popma JJ, Khabbaz KR, Laham RJ, et al. Frailty phenotype and deficit accumulation frailty index in predicting recovery after transcatheter and surgical aortic valve replacement. J Gerontol A Biol Sci Med Sci. 2019;74(8):1249–56.

114. Martens FK, Tonk EC, Kers JG, Janssens ACJ. Small improvement in the area under the receiver operating characteristic curve indicated small changes in predicted risks. J Clin Epidemiol. 2016;79:159–64.

115. Baker SG, Schuit E, Steyerberg EW, Pencina MJ, Vickers A, Moons KG, et al. How to interpret a small increase in AUC with an additional risk prediction marker: Decision analysis comes through. Stat Med. 2014;33(22):3946–59.

116. Farrell S, Kane AE, Bisset E, Howlett SE, Rutenberg AD. Measurements of damage and repair of binary health attributes in aging mice and humans reveal that robustness and resilience decrease with age, operate over broad timescales, and are affected differently by interventions. Elife. 2022;11:e77632.

APPENDICES

Appendix A. PubMed Search

Search terms for Figure 1:

((((((((("physical resilience"[Title/Abstract]) OR ("physical resiliency"[Title/Abstract])) OR ("physical resiliencies"[Title/Abstract])) OR ("physiologic resilience"[Title/Abstract])) OR ("physiologic resiliency"[Title/Abstract])) OR ("physiologic resiliencies"[Title/Abstract])) OR ("physiological resilience"[Title/Abstract])) OR ("physiological resiliency"[Title/Abstract])) OR ("physiological resiliencies"[Title/Abstract])) AND ((older[Title/Abstract]) OR (gerontol*[Title/Abstract]) OR (geriatr*[Title/Abstract]) OR aging[Title/Abstract] OR ageing[Title/Abstract]))

The search was last run on January 3rd, 2024. Search terms were optimized to balance between sensitivity and specificity of literature pertaining to physical resilience in older adults. These results are an underestimate of the true number of publications on the topic. For example, a number of recent review articles cited in this thesis only use the term resilience in the title/abstract but discuss overall resilience (including physical and psychological) in the body of the paper (e.g., Hamaker 2023, Abadir 2023, Cesari 2022). Removing the physical/physiological term results in an excess of irrelevant articles given the broad nature of the term resilience.

Appendix B. Empirical Literature Review Table

Table 56. Summary of the Empirical Literature Measuring Physical Resilience

Author and Year	Physical Resilience Measure	Stressor	Accounted for severity?	Outcome Measure	Time Scale	Population vs Clinical Sample (N)
Measures of Post-Stressor Change						
Calle et al. 2018 (58)	Relative and absolute functional gain (AFG, RFG)	Acute - orthopaedic surgery (hip fracture and hip or knee replacement) and stroke (ischemic or hemorrhagic	Stratified by orthopedics or stroke, but no further differentiation of stressor.	Barthel Index (Functional independence)	Baseline plus 2 measurements with 30 days in between	Clinical – SAFARI Study (n=450), aged 65+, Spain and Italy
Colon-Emeric et al. 2019 (59)	Recovery Phenotype – Latent Class Analysis (LCA) using multiple outcomes to determine high, medium, and low resilience	Acute - Hip fracture	Yes - Stressor characteristics included anesthesia type, duration of surgery, partial vs total arthroplasty, and post-operative complications	Hip fracture: LCA of 10 Self-reported physical functioning and activity measures	Baseline, 2, 6 and 10 months	Clinical – Baltimore Hip Studies cohorts (n=541), Aged 60+, USA
Colon-Emeric et al. 2020 (44)	Recovery Phenotype – Latent Class Analysis (LCA) using multiple outcomes OR Principal Component Analysis (PCA) for multiple classification variables Expected Recovery Differential	Acute - Hip fracture, pneumonia	Yes - Restricted to known probable viral infection (excluded bacterial to reduce stressor variability). Also excluded individuals with known or suspected coinfection at any other site.	Hip fracture: LCA of 10 Self-reported physical functioning and activity measures Pneumonia: a variety of factors such as length of hospital stay and ICU admission	Hip fracture: Baseline, 2, 6 and 10 months Pneumonia: 28 day follow up	Clinical – Baltimore Hip Studies cohorts (n=541) And Community Acquired Pneumonia and Sepsis Outcome

Author and Year	Physical Resilience Measure	Stressor	Accounted for severity?	Outcome Measure	Time Scale	Population vs Clinical Sample (N)
	(expected vs observed based on population)		Mentioned stressor characteristics (mentioned above for hip fracture) and laboratory findings			Diagnostics (n=185), Aged 60+, USA
Hladek et al. 2022 (60)	Phenotypic trajectory of 4 variables (not recovery-improving, stable, declining)	Chronic - incident hemodialysis	Yes – accounted for disease severity with disease-specific markers and dialysis details such as access type	SF-36: Physical function, mental health, vitality, and general health	Baseline, 3, 6 12 months	Clinical - CHOICE Cohort (n=394), Aged 55+, USA
Presley et al. 2022 (61)	Maintenance or improvement in disability scores	Chronic - newly diagnosed lung cancer, just starting treatment	Yes – Restricted recruitment to stage 4, limiting variation in disease severity. Treatment type considered.	Disability measured using EQ-5D-5L	Monthly for 8 months	Clinical – Beating Lung cancer in Ohio cohort (n=207), aged 34-91, USA
Pedone et al. 2020 (62)	Categorization based on observed decline (resilient, non-resilient, decliners)	Non-specific "major health event" (could be acute or initiation of chronic/ recurrent)	No	Short physical performance battery (SBBP)	Baseline plus follow-up at three years after baseline, with or without a major health event in between.	Population - InCHIANTI Study (n=726 for mortality outcome, n=567 for functional status outcome), Aged 65+, Italy
Duan-porter et al. 2016 (63)	Regaining at least 50% of lost function after decline	Unspecified decline in a sample of cancer survivors	Accounted for years since diagnosis and treatment type. Included prostate, breast, and colorectal – type and stage not accounted for in analyses.	SF-36 physical function subscale	Quarterly over two years	Clinical – RENEW RCT (n=594) Overweight cancer survivors, 65+, at least 5 years post diagnosis, USA

Author and Year	Physical Resilience Measure	Stressor	Accounted for severity?	Outcome Measure	Time Scale	Population vs Clinical Sample (N)
Stimulus Response Measures						
Koivunen et al. 2023 (64)	Stimulus response	Orthostatic challenge - change from prone to supine	N/A (stimulus response measures are standardized by definition)	Several hemodynamic and postural sway indices	N/A	Population - Community dwelling Finnish older adults aged 75, 80, and 85 at baseline (n=689)
Whitson et al. 2021 (PRIME-KNEE) (48)	Stimulus response	Physical, immunological, and cognitive tests		PBMC ex vivo response to LPS, influenza vaccine, cerebrovascular reactivity, dual task effect of gait speed	N/A	Clinical - (n=250), aged 60+, USA
Walston et al. 2023 (SPRING) (52)	Stimulus response	Physical and endocrine tests		Orthostatic blood pressure, ACTH stimulation, oral glucose tolerance test	N/A	Clinical – (pilot n=32, 22, and 23 for bonemarrow, dialysis, and knee substudies, respectively), aged 55+, USA
Residual-based Measures						
Wu et al. 2019 (43)	Frailty-disease Mismatch: mismatch between frailty and disease burden: adapters, expected agers, premature frailers	Existing chronic disease and disease burden	Used self-rated health and number of medications as indictors of disease burden (stand-in for severity)	Residuals of linear regression of SAVE Frailty Scale	Estimated using data from second annual clinic visit	Population – Health ABC Study (n=2,457), Aged 70-79, USA
Wu et al. 2022 (67)	Simplified frailty-disease mismatch: mismatch between frailty and disease burden: adapters,	Existing chronic disease	No	Follow up simplified approach not using regression, bur rather using group	Estimated using data from second annual clinic visit	Population – Health ABC Study (n=2,457), Aged 70-79, USA

Author and Year	Physical Resilience Measure	Stressor	Accounted for severity?	Outcome Measure	Time Scale	Population vs Clinical Sample (N)
	expected agers, premature frailers			means and cut points		
Zhang et al. 2023 (68)	Residuals of Short physical performance battery (SBBP) model	Existing chronic diseases, age, sex, race, self-rated health	Not explicitly stated, but similar to Wu et al., self-rated health may indicate burden.	Residuals of linear regression on SBBP	Estimated using baseline data	Population – National Health and Aging Trends Study (n=6508), Aged 65+, USA
Milman et al. 2023 (69)	Frailty Resilience Scale: mismatch between frailty and polygenic risk score	Polygenic risk score	No	Residual of linear mixed model of 41-item frailty index. Residual based on predictions using fixed effects only	Average residual across earliest three visits	Population - LonGenity Cohort (n=467), mean age of 74.4, USA
Sotos-Prieto et al. 2021 (70)	Resilience defined as accumulating fewer deficits than expected (based on cohort average) despite exposure to chronic stressors	No, but limited to those above median number of deficits, suggesting exposure to chronic stressors	No	52-item "deficit accumulation index" – essentially FI	3.2 year follow up (accumulated more or less than expected based on the cohort average)	Population – Seniors-ENRICA cohort (n=1301), Aged 60+, Spain
Dynamical Indicators of Resilience						
Gijzel et al. 2017 (37)	DIOR – variance, cross-correlation, and temporal autocorrelation	No	N/A	Self-rated physical, mental, and social health	daily for 100 days	Clinical (institutionalized) – residential care facilities (n=22), Aged 70+, Italy
Gijzel et al. 2019 (72)	DIOR – variance, cross-correlation, and temporal autocorrelation	No	N/A	Postural balance	30 second continuous feed	Population – (n=212), aged 80-94, Netherlands

Author and Year	Physical Resilience Measure	Stressor	Accounted for severity?	Outcome Measure	Time Scale	Population vs Clinical Sample (N)
Gijzel et al. 2020 (73)	DIOR – variance, cross-correlation, and temporal autocorrelation	No	N/A	Heart rate and physical activity, well being	Continuous monitoring of heart rate and physical activity during hospitalization (mean stay of 7.4 days), as well as momentary well being 4x per day	Clinical – patients admitted to geriatric ward for acute illness (n=121), aged 65+, Netherlands
Kolk et al. 2021 (74)	DIOR: variance	No - but estimated in a sample of acutely hospitalized individuals to see relationship with recovery	N/A	Step count, self-rated levels of pain, fatigue, fear of falling. Coefficient of variation for DIOR rather than SD. (KATZ 15-tem ADL as recovery outcome)	3 months of observations-continuous for steps, daily for self-rated measures. Minimum of three days of observations	Clinical – Hospital-ADL study (n=207), Aged 70+, Netherlands
Rector et al. 2021 (75)	DIOR: Critical slowing down	No	N/A	Heart rate and physical activity (measured via accelerometer)	11 hours of recording	Clinical – Wellbeing and Resilience Study (n=121), geriatric inpatients aged 65+, Netherlands
Lucas et al. 2023 (76)	DIOR: physiological complexity	No	N/A	Step counts and step count variability	4-hour blocks of accelerometer data	Clinical - ambulatory older adults receiving hemodialysis (n=37), mean age of 70.6, USA
Self-reported Measures						

Author and Year	Physical Resilience Measure	Stressor	Accounted for severity?	Outcome Measure	Time Scale	Population vs Clinical Sample (N)
Park et al. 2022 (81)	The Physical Resilience Instrument	No	N/A	Questionnaire	Cross-sectional	Clinical – individuals with osteoarthritic symptoms (n=235), Aged 65-92, Korea
Hu et al. 2021 (83)	The Physical Resilience Instrument for Older Adults (PRIFOR)	No	N/A	Questionnaire	Cross-sectional ascertainment, but followed over time to validate outcomes (baseline then one month after discharge)	Clinical – patients admitted to medical ward of tertiary medical centres (n=192), Aged 65+, Taiwan
Static Surrogates, Proxies, and Aggregate Measures						
Walston et al. 2023 (SPRING) (52)	Static surrogates – phenotypic frailty, SF-36, nutrition, Karnofsky performance measures, specific measures for outcomes	N/A	N/A	N/A	Static	Clinical – (pilot n=32, 22, and 23 for bonemarrow, dialysis, and knee substudies, respectively), aged 55+, USA
Arbeev et al. 2019 (86)	Proxy measures: onset of "unhealthy life" and survival following onset/avoid diseases at age 65+, and survival to extreme ages	N/A	N/A	Unhealthy life: first occurrence of a major complex disease including cancer, CVD, and type II diabetes.	2-year resolution for Framingham and 1 year for CHS.	Population – Framingham Cohort (n=5079), aged 30-62 at recruitment, USA Cardiovascular Health Study (n=5795), Aged 65+, USA

Author and Year	Physical Resilience Measure	Stressor	Accounted for severity?	Outcome Measure	Time Scale	Population vs Clinical Sample (N)
Galvin et al. 2020 (87)	Proxy measures: onset of "unhealthy life" and survival following onset/avoid diseases at age 65+, and survival to extreme ages	N/A	N/A	Unhealthy life: first occurrence of a major complex disease including cancer, CVD, and type II diabetes.	Followed since 1968	Population - Danish national population registers. Female siblings (n=1156) and controls (n=1156), Aged 68+, Denmark
Galvin et al. 2021 (88)	Survival after onset of CVD	N/A	N/A	Mortality	Followed since 1968	Population - Same as above – national registers (n=1206) (offspring and controls 1:2 ratio)

Appendix C. Creation of the HRS Frailty Index

Table 57 includes 59 items that meet criteria for inclusion in the Frailty Index. These criteria are 1) associated with age, 2) do not saturate too early, 3) are not too common (>80%) or too rare (<1%), 4) have minimal missing data (<5%), 5) are not too highly correlated with other variables (<0.95), and 6) are comparable across all waves considered (3-13).

FI-41: Main FI that excludes the clinical disease domain and self-rated health (so the FI is compatible with the adjusted FI model), as well as health care utilization (left out over concerns regarding access in the United States).

FI-51: Sensitivity analysis that only excludes health care utilization variables.

FI-56: Sensitivity analysis that includes health care utilization variables, but drops assistance with IADLs. This was created to be comparable with a recently published guide by Theou et al. (98).

Assumptions regarding missing values in FI:

The equipment use questions, bede and walkre, were skipped and assigned a special missing code if the ADL question was skipped due to prior answers indicating no ADL difficulties. I coded "skip" as no difficulty (0), because the only reason for skips is the assumption of no difficulty.

For ADLs, IADLs, and functional limitations and mobility, "don't do" was assumed to correspond to difficulty (1). This only applied to a small fraction of answers. The only variables this was not assumed for were jogging and using a map (neither of which were included).

Help variables are set to no help (0) if the difficulty variable was 0, as this question is not asked if the respondent does not report difficulty.

Variable Notes:

There are two options for chronic condition variables in HRS: one that corresponds to the individual ever having the condition, and one that corresponds to the raw response of reporting the condition in the last two years. The ever variables created by RAND were deemed a better choice than the variables corresponding to the last two years due to uncertainty in the answers due to disputes at following waves. The ever variables have been corrected for disputes, but also assume any refusal or don't know is a 0, does not have condition.

The ever variables for cardiovascular diseases are different – not available in early waves, but the last two years are available. Angina excluded due to cross wave differences (change in question wording in wave 10).

Similarly, any medications with uncertainty in when it was and was not asked across waves were left out (e.g., diabetes in wave 13, cardiovascular conditions for all waves).

Table 57. Variables Meeting FI Inclusion Criteria Waves 3-14

Domain	Variable	Variable Name	FI-41	FI-51	FI-56
Self-rated health, hearing, and vision	Self-rated health	shlt		✓	✓
	Self-rated eyesight	sight	✓	✓	✓
	Self-rated near eyesight	nsight	✓	✓	✓
	Self-rated distal eyesight	dsight	✓	✓	✓
	Self-rated hearing	hearing	✓	✓	✓
Activities of daily living (ADLs)	Difficulty bathing or showering	bath	✓	✓	✓
	Difficulty getting in or out of bed	bed	✓	✓	✓
	Difficulty dressing	dress	✓	✓	✓
	Difficulty eating	eat	✓	✓	✓
	Difficulty walking across room	walkr	✓	✓	✓
	Difficulty using toilet	toilt	✓	✓	✓
Instrumental activities of daily living (IADLS)	Difficulty using a phone	phone	✓	✓	✓
	Difficulty managing money	money	✓	✓	✓
	Difficulty shopping for groceries	shop	✓	✓	✓
Assistance with ADLs and IADLs	Help walking	walkrh	✓	✓	✓
	Help bathing or showering	bathh	✓	✓	✓
	Help dressing	dressh	✓	✓	✓
	Help eating	eath	✓	✓	✓
	Help getting in/out of bed	bedh	✓	✓	✓
	Help using toilet	toilth	✓	✓	✓
	Help managing money	moneyh	✓	✓	
	Help shopping	shoph	✓	✓	
	Help using phone	phoneh	✓	✓	

Domain	Variable	Variable Name	FI-41	FI-51	FI-56
Other functional/ mobility limitations	Difficulty getting up from chair	chair	✓	✓	✓
	Difficulty climbing one flight of stairs	clim1	✓	✓	✓
	Difficulty climbing several slights of stairs		✓	✓	✓
	Difficulty picking up a dime	dime	✓	✓	✓
	Difficulty Reach/extend arms up	arms	✓	✓	✓
	Difficulty Lift/carry 10lbs	lift	✓	✓	✓
	Difficulty push/pull large object	push	✓	✓	✓
	Difficulty Stoop/Kneel/Crouch	stoop	✓	✓	✓
	Difficulty walking several blocks	walks	✓	✓	✓
	Difficulty walking one block	walk1	✓	✓	✓
Clinical Disease	Ever had arthritis or rheumatism (ever)	arthre		✓	✓
	Ever had diabetes or high blood sugar (ever)	diabe		✓	✓
	Ever had high blood pressure/hypertension (ever)	hibpe		✓	✓
	Ever had cancer or malignant tumor of any kind except skin cancer (ever)	cancre		✓	✓
	Ever had chronic lung disease except asthma, chronic bronchitis, or emphysema	lunge		✓	✓
	Ever had heart problems	hearte		✓	✓
	Heart attack (last two years)	hrtatt		✓	✓
	Congestive heart failure (last two years)	conhrtf		✓	✓
	Ever had stroke or TIA	stroke		✓	✓
Signs and Symptoms	Urinary incontinence	urinai	✓	✓	✓
	Urinary incontinence frequency	urinaif	✓	✓	✓
	Underweight	bmicat (dichotomized for underweight)	✓	✓	✓
Equipment Use	Wears hearing aid	hearaid	✓	✓	✓

Domain	Variable	Variable Name	FI-41	FI-51	FI-56
	Equipment to walk	walkre	✓	✓	✓
	Equipment to get into bed	bede	✓	✓	✓
Medication	Takes meds for high blood pressure	rxhibp	✓	✓	✓
	Takes meds for stroke	rxstrok	✓	✓	✓
	Takes meds for lung condition	rxlung	✓	✓	✓
Health Care Utilization	Nursing home stay, last two years	nrshom			✓
	Home health care, last two years	homcar			✓
	Outpatient surgery, last two years	output			✓
	Special health facility, last two years	spcfac			✓
	Hospital stay, last two years	hosp			✓
	Lives in nursing home (at time of interview)	nhmliv			✓
	Heart surgery since last wave	hrtsrg			✓
	Joint replacement since last wave	jointr			✓

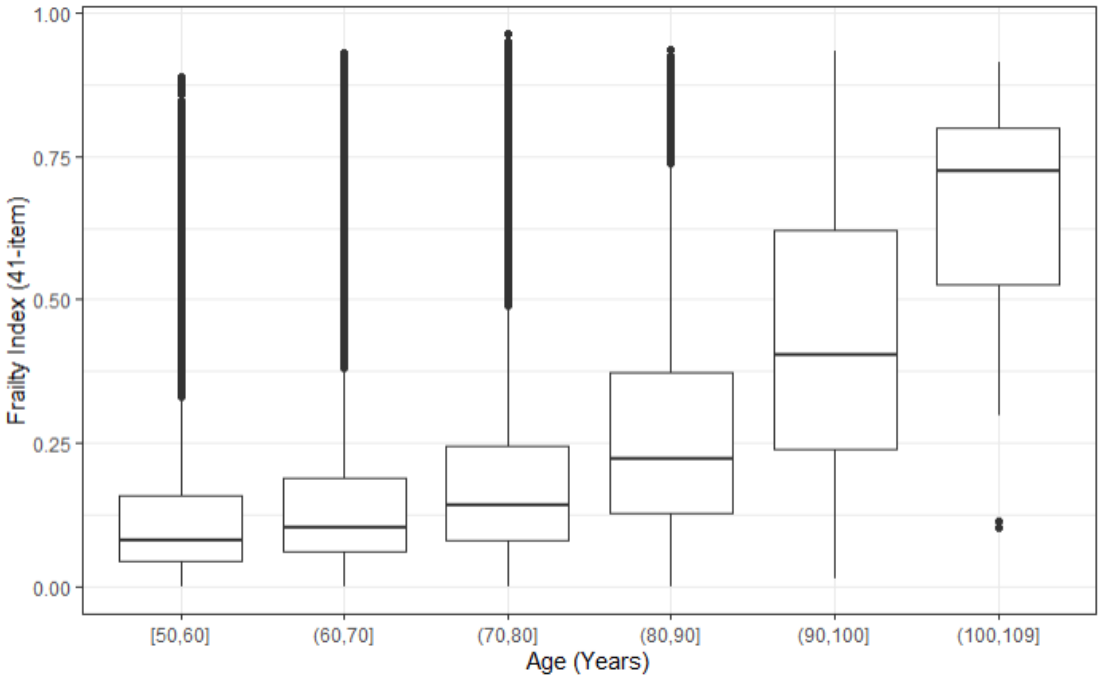
Appendix D. Growth Curve Modelling

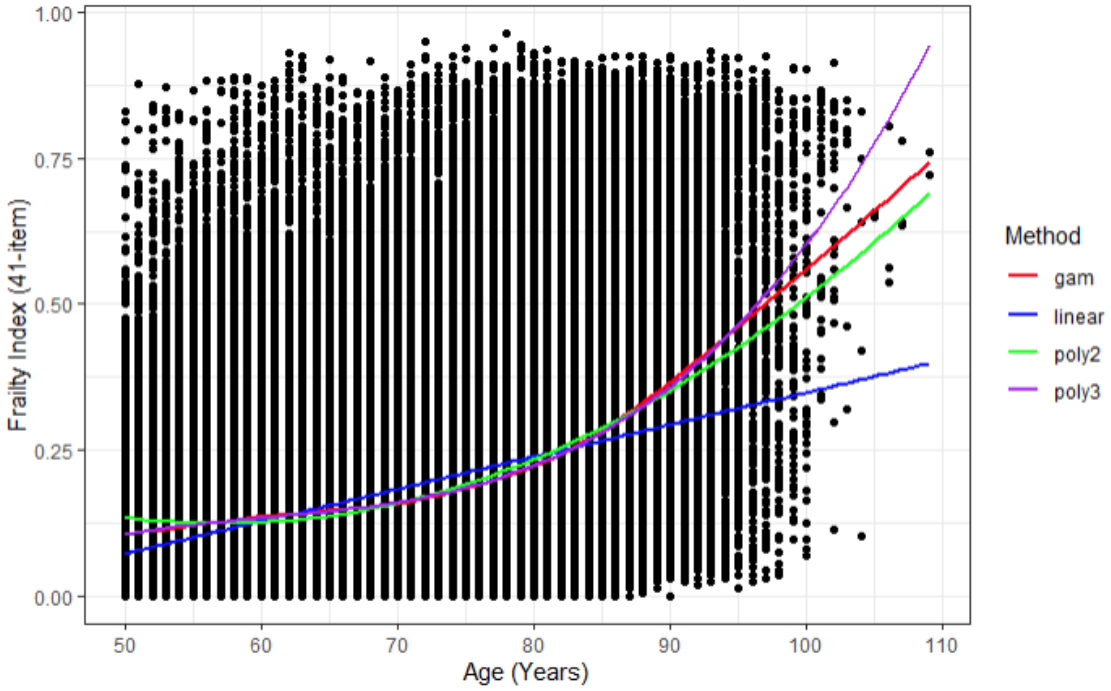
Cubic Specification of age

A likelihood ratio test indicated that quadratic is a better fit than linear (used ML to compare nested models, $p < 0.0001$).

Similarly, cubic fits better than quadratic, as expected by the plots below ($p < 0.0001$). This applies to both unadjusted and adjusted models.

Plots of age and FI in the pooled sample suggesting cubic fit:





Assessment of Clustering Effects

I based my assessment of clustering effects on the design effect:

$$Design\ Effect = 1 + (nc - 1)ICC$$

Where nc is equal to the average cluster size (i.e., average number of individuals per household) (96). The intraclass correlation coefficient (ICC) is a measure of dependence that represents the expected correlation between two observations within the same cluster (in this instance, between two individuals in the same household). The ICC is equal to the between-cluster variance divided by the total variance:

$$ICC = \frac{\tau_{00}}{\tau_{00}+\sigma^2}$$

Where τ_{00} is the between cluster variance (i.e., household), and σ^2 is the within cluster variance (i.e., individual) (96).

Mortality clustering:

Individual level ICC of 0.62175

N = 27,744, with a mean of 6.87 observations per person. Minimum of 3, maximum of 11. 190553 observations total.

Household ICC of 0.54138

N = 18,595, with a mean of 10.25 repeated observations per household. Average of 1.49 individuals per household

27,744/18,595 = 1.49

Design Effect=1+(nc-1)ICC

= 1+(1.49-1)0.54138 = 1.265.

➔ Below 2, don't need to include given addition complexity added.

Recovery clustering:

Individual level ICC of 0.64108

N = 1905, with a mean of 5.29 observations per person. Minimum of 3, maximum of 10. 10,085 observations total.

Household ICC of 0.62671

N = 1,839, with a mean of 5.48 observations per household. Minimum of 3, maximum of 20. Average of 1.04 individuals per household.

1,839/1,905 = 1.036

Design Effect=1+(nc-1)ICC

= 1+(1.036-1)0.62671 = 1.023

➔ Negligible

Residual Correlation Structure:

I tried the following:

- uncorrelated (default)

- AR1 – similar fit to CAR1, but given that some individuals are not observed every wave, CAR1 is more appropriate
- CAR1
- ARMA – convergence issues
- Compound symmetry
- Unstructured – would not converge

To determine what residual correlation fit best, I estimated models with identical fixed and random effects, but different residual correlation using ML. The model with the lowest AIC was considered the better fitting model. Continuous autoregressive order 1 fit the best. I confirmed this in both the age-only mixed effect model, and the adjusted model. Final models were re-estimated using REML.

Random Effect Structure:

Using the same approach described above, a general positive-definite matrix, with no additional structure using a log-Cholesky parameterization, fit better than the general positive-definite matrix, with no additional structure (with no log-Cholesky parameterization).

Appendix E. Comparison of SF-36 Physical Function Subscale and HRS Equivalent

Table 58. Comparison of SF-36 Physical Function Subscale and HRS Equivalent

SF-36 Item #	SF-36 Question	HRS Equivalent
	"The following items are about activities you might do during a typical day. Does **your health now limit you** in these activities? If so, how much?"	"Because of a health problem do you have any difficulty with.."
3	**Vigorous activities**, such as running, lifting heavy objects, participating in strenuous sports	Running or jogging about a mile
4	**Moderate activities**, such as moving a table, pushing a vacuum cleaner, bowling, or playing golf	Pulling or pushing large objects like a living room chair
5	Lifting or carrying groceries	Lifting or carrying weights over 10 pounds, like a heavy bag of groceries
6	Climbing **several** flights of stairs	Climbing several flights of stairs without resting
7	Climbing **one** flight of stairs	Climbing one flight of stairs without resting
8	Bending, kneeling, or stooping	Stooping, kneeling, or crouching
9	Walking **more than a mile**	Imputed assuming those with difficulty walking several blocks would also have difficulty walking more than a mile.
10	Walking **several blocks**	Walking several blocks
11	Walking **one block**	Walking one block
12	Bathing or dressing yourself	Bathing or showering And Dressing, including putting on shoes and socks

SF-36 responses are three categories: "Yes, limited a lot", "Yes, limited a little", and "No, not limited at all". The HRS equivalent responses are "yes", "no." Binary indicators of any difficulty coded 0, 1, where a higher score indicates more functional ability (i.e., coded opposite as FI items, where 0 = difficulty while 1 = no difficulty). Following the guidance in the SF-36 Manual (103), I calculated a score as long as a respondent answered half of the items. I used the average score of the answered items to impute the missing items (see page 6:16 or 79 of the manual). The final imputed score sum is then divided by 10 to yield an approximation of the SF-36 PFS, which ranges from 0-100, where a higher score indicates higher functional ability.

The SF-36 Questionnaire can be viewed here: 36-Item Short Form Survey Instrument (SF-36) | RAND.

Appendix F. Additional FM Exploration

FM Category Comparison of Disease Burden (Mortality and Recovery)

Table 59. Disease Burden Comparison (Mortality Sample)

Characteristic	FM Category			p-value[2]
	Adapter, N = 6,936[1]	**Expected Ager**, N = 13,872[1]	**Premature Frailer**, N = 6,936[1]	
Age	80 (71, 87)	69 (62, 77)	77 (67, 85)	**<0.001**
Frailty Index	0.12 (0.07, 0.20)	0.13 (0.07, 0.23)	0.52 (0.37, 0.71)	**<0.001**
Ever had Arthritis	5,027 (72%)	7,952 (57%)	5,386 (78%)	**<0.001**
Ever had Diabetes	2,000 (29%)	3,420 (25%)	2,434 (35%)	**<0.001**
Ever had High Blood Pressure	4,941 (71%)	8,395 (61%)	5,166 (74%)	**<0.001**
Ever had Cancer	1,729 (25%)	2,380 (17%)	1,533 (22%)	**<0.001**
Ever had Lung Disease	1,154 (17%)	1,553 (11%)	1,399 (20%)	**<0.001**
Ever had Heart Problems	3,026 (44%)	3,578 (26%)	3,038 (44%)	**<0.001**
Ever had a Stroke	1,517 (22%)	1,034 (7.5%)	1,721 (25%)	**<0.001**
Regularly Takes Prescription Drugs	6,377 (92%)	11,266 (81%)	6,547 (94%)	**<0.001**
Self-rated Health				**<0.001**
Excellent	388 (5.6%)	1,134 (8.2%)	121 (1.7%)	
Very Good	1,611 (23%)	4,063 (29%)	517 (7.5%)	
Good	2,261 (33%)	4,632 (33%)	1,531 (22%)	
Fair	1,813 (26%)	2,863 (21%)	2,459 (35%)	
Poor	863 (12%)	1,180 (8.5%)	2,308 (33%)	

[1]Median (IQR); n (%)

[2]Kruskal-Wallis rank sum test; Pearson's Chi-squared test

Table 60. Disease Burden Comparison (Recovery Sample)

Characteristic	FM Category			p-value[2]
	Adapter, N = 477[1]	**Expected Ager**, N = 952[1]	**Premature Frailer**, N = 476[1]	
Age	78 (68, 85)	73 (66, 80)	76 (67, 83)	**<0.001**
Frailty Index	0.10 (0.07, 0.17)	0.17 (0.10, 0.26)	0.47 (0.36, 0.65)	**<0.001**
Ever had Arthritis	360 (75%)	577 (61%)	376 (79%)	**<0.001**
Ever had Diabetes	143 (30%)	267 (28%)	180 (38%)	**<0.001**
Ever had High Blood Pressure	343 (72%)	628 (66%)	366 (77%)	**<0.001**
Ever had Cancer	104 (22%)	159 (17%)	83 (17%)	0.055
Ever had Lung Disease	72 (15%)	120 (13%)	87 (18%)	**0.016**
Ever had Heart Problems	237 (50%)	346 (36%)	229 (48%)	**<0.001**
Ever had a Stroke	116 (24%)	104 (11%)	118 (25%)	**<0.001**
Regularly Takes Prescription Drugs	443 (93%)	822 (86%)	460 (97%)	**<0.001**
Self-rated Health				**<0.001**
Excellent	26 (5.5%)	53 (5.6%)	10 (2.1%)	
Very Good	117 (25%)	206 (22%)	40 (8.4%)	
Good	164 (34%)	333 (35%)	91 (19%)	
Fair	105 (22%)	264 (28%)	173 (36%)	
Poor	65 (14%)	96 (10%)	162 (34%)	

[1]Median (IQR); n (%)

[2]Kruskal-Wallis rank sum test; Pearson's Chi-squared test

Table 61. Comparison of FM Categories (Mortality - Four Category FM)

Characteristic	FM Category				p-value[2]
	Most resilient, N = 6,936[1]	**Resilient**, N = 6,936[1]	**Less Resilient**, N = 6,936[1]	**Least Resilient**, N = 6,936[1]	
Age	80 (71, 87)	68 (62, 77)	69 (62, 78)	77 (67, 85)	**<0.001**
Sex					**<0.001**
Female	3,955 (57%)	3,926 (57%)	3,705 (53%)	4,255 (61%)	
Male	2,981 (43%)	3,010 (43%)	3,231 (47%)	2,681 (39%)	
2018 Vital Status					**<0.001**
Alive	4,041 (58%)	5,257 (76%)	4,692 (68%)	2,600 (37%)	
Deceased	2,895 (42%)	1,679 (24%)	2,244 (32%)	4,336 (63%)	
Frailty Index	0.116 (0.067, 0.196)	0.091 (0.055, 0.165)	0.183 (0.110, 0.280)	0.520 (0.372, 0.714)	**<0.001**
RoA Category					**<0.001**
Slowest Ager	1,430 (21%)	3,119 (45%)	2,092 (30%)	295 (4.3%)	
Slow Ager	1,304 (19%)	1,953 (28%)	2,369 (34%)	1,310 (19%)	
Fast Ager	2,108 (30%)	1,256 (18%)	1,570 (23%)	2,002 (29%)	
Fastest Ager	2,094 (30%)	608 (8.8%)	905 (13%)	3,329 (48%)	
DIOR-FI Category					**<0.001**
Highest Stability	1,649 (24%)	3,035 (44%)	1,905 (27%)	347 (5.0%)	
High Stability	2,051 (30%)	2,090 (30%)	2,039 (29%)	756 (11%)	
Low Stability	2,040 (29%)	1,251 (18%)	1,941 (28%)	1,704 (25%)	
Lowest Stability	1,196 (17%)	560 (8.1%)	1,051 (15%)	4,129 (60%)	

[1]Median (IQR Bounds); n (%)

[2]Kruskal-Wallis rank sum test; Pearson's Chi-squared test

Table 62. Logistic Regression Models for FM and Mortality (Four Category FM)

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
FM Category												
Most resilient	—	—		—	—		—	—		—	—	
Resilient	0.45	0.41, 0.48	**<0.001**	0.46	0.43, 0.50	**<0.001**	0.74	0.68, 0.81	**<0.001**	0.50	0.43, 0.59	**<0.001**
Less Resilient	0.67	0.62, 0.72	**<0.001**	0.37	0.34, 0.40	**<0.001**	0.72	0.66, 0.79	**<0.001**	0.52	0.44, 0.61	**<0.001**
Least Resilient	2.33	2.17, 2.49	**<0.001**	0.16	0.14, 0.17	**<0.001**	0.46	0.41, 0.53	**<0.001**	0.86	0.70, 1.05	0.14
Frailty Index				1.26	1.25, 1.27	**<0.001**	1.18	1.17, 1.19	**<0.001**	1.19	1.17, 1.21	**<0.001**
Age							1.06	1.06, 1.07	**<0.001**	1.06	1.06, 1.06	**<0.001**
Sex												
Female							—	—		—	—	
Male							1.86	1.75, 1.97	**<0.001**	1.90	1.79, 2.02	**<0.001**
FM Category * Frailty Index												
Resilient * Frailty Index										1.08	1.05, 1.10	**<0.001**
Less Resilient * Frailty Index										1.04	1.01, 1.06	**0.003**
Least Resilient * Frailty Index										0.95	0.94, 0.97	**<0.001**

Model 1 is unadjusted. Model 2 is adjusted for frailty. Model 3 is further adjusted for age and sex. Model 4 further includes a frailty interaction. Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference.

Table 63. Disease Burden Comparison (Mortality Sample - Four Category FM)

Characteristic	FM Category				p-value[2]
	Most resilient, N = 6,936[1]	**Resilient**, N = 6,936[1]	**Less Resilient**, N = 6,936[1]	**Least Resilient**, N = 6,936[1]	
Age	80 (71, 87)	68 (62, 77)	69 (62, 78)	77 (67, 85)	**<0.001**
Frailty Index	0.12 (0.07, 0.20)	0.09 (0.05, 0.16)	0.18 (0.11, 0.28)	0.52 (0.37, 0.71)	**<0.001**
Ever had Arthritis	5,027 (72%)	3,832 (55%)	4,120 (59%)	5,386 (78%)	**<0.001**
Ever had Diabetes	2,000 (29%)	1,717 (25%)	1,703 (25%)	2,434 (35%)	**<0.001**
Ever had High Blood Pressure	4,941 (71%)	4,261 (61%)	4,134 (60%)	5,166 (74%)	**<0.001**
Ever had Cancer	1,729 (25%)	1,191 (17%)	1,189 (17%)	1,533 (22%)	**<0.001**
Ever had Lung Disease	1,154 (17%)	647 (9.3%)	906 (13%)	1,399 (20%)	**<0.001**
Ever had Heart Problems	3,026 (44%)	1,681 (24%)	1,897 (27%)	3,038 (44%)	**<0.001**
Ever had a Stroke	1,517 (22%)	441 (6.4%)	593 (8.5%)	1,721 (25%)	**<0.001**
Regularly Takes Prescription Drugs	6,377 (92%)	5,713 (82%)	5,553 (80%)	6,547 (94%)	**<0.001**
Self-rated Health					**<0.001**
Excellent	388 (5.6%)	663 (9.6%)	471 (6.8%)	121 (1.7%)	
Very Good	1,611 (23%)	2,231 (32%)	1,832 (26%)	517 (7.5%)	
Good	2,261 (33%)	2,349 (34%)	2,283 (33%)	1,531 (22%)	
Fair	1,813 (26%)	1,252 (18%)	1,611 (23%)	2,459 (35%)	
Poor	863 (12%)	441 (6.4%)	739 (11%)	2,308 (33%)	

[1]Median (IQR); n (%)

[2]Kruskal-Wallis rank sum test; Pearson's Chi-squared test

Cross-sectional FM Test (Mortality)

This model is a simple linear version of the main model using cross sectional data from Wave 10. Individuals are included if they are between ages 70 to 79 and report no difficulty climbing one flight of stairs or walking several blocks (n = 3,173). Adjusted R^2 is 0.2722.

Table 64. Logistic Regression Models for FM and Mortality (Cross-sectional, ABC Match)

	Model 1			Model 2			Model 3			Model 4		
Characteristic	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
FM Category												
Expected Ager	—	—		—	—		—	—		—	—	
Adapter	1.46	1.19, 1.80	**<0.001**	1.98	1.58, 2.48	**<0.001**	2.08	1.65, 2.61	**<0.001**	1.32	0.75, 2.32	0.3
Premature Frailer	1.41	1.14, 1.73	**0.001**	0.67	0.51, 0.89	**0.006**	0.65	0.49, 0.87	**0.004**	0.53	0.27, 1.05	0.073
Frailty Index				1.29	1.21, 1.38	**<0.001**	1.31	1.23, 1.40	**<0.001**	1.24	1.10, 1.39	**<0.001**
Age							1.11	1.07, 1.14	**<0.001**	1.11	1.07, 1.15	**<0.001**
Sex												
Female							—	—		—	—	
Male							2.17	1.81, 2.61	**<0.001**	2.16	1.80, 2.60	**<0.001**
FM Category * Frailty Index												
Adapter * Frailty Index										1.21	0.98, 1.49	0.084
Premature Frailer * Frailty Index										1.06	0.92, 1.23	0.4

Model 1 is unadjusted. Model 2 is adjusted for frailty. Model 3 is further adjusted for age and sex. Model 4 further includes a frailty interaction. Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference.

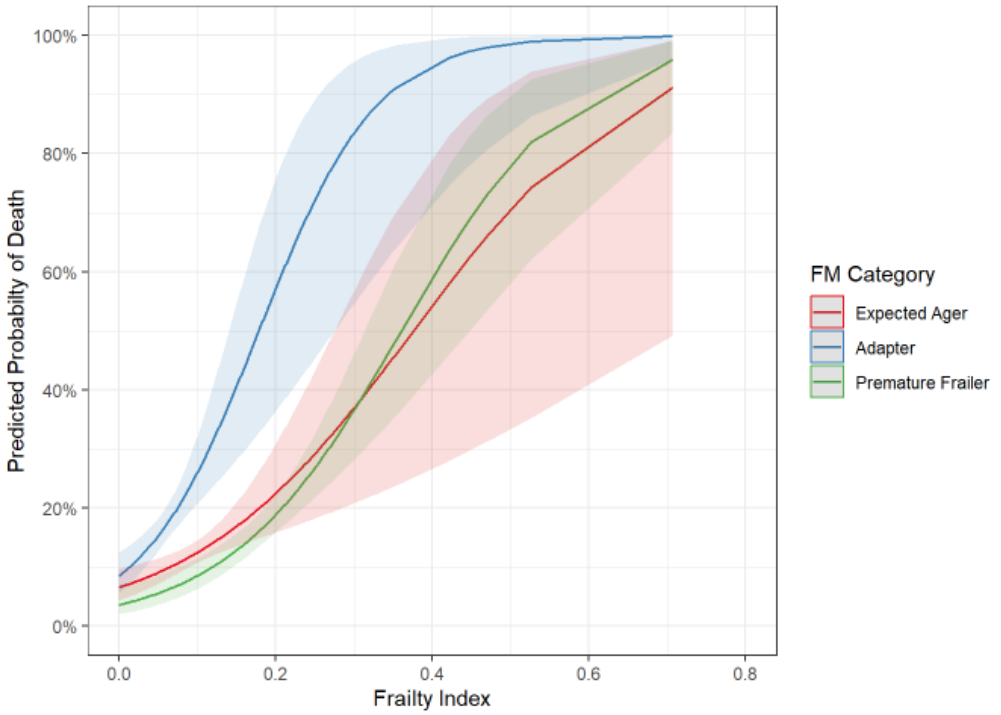


Figure 43. Interaction Effects of FI and FM on Mortality (Cross-sectional, ABC Match)

Visualization of the interaction between FI and FM on mortality (Model 4, Table 64).

Table 65. Disease Burden Comparison (Cross-sectional, ABC Match)

Characteristic	FM Category			p-value[2]
	Adapter, N = 794[1]	**Expected Ager**, N = 1,586[1]	**Premature Frailer**, N = 793[1]	
Age	74.08 (72.00, 76.83)	73.92 (71.77, 76.50)	74.17 (72.08, 76.92)	0.085
Frailty Index	0.05 (0.04, 0.07)	0.09 (0.07, 0.12)	0.16 (0.13, 0.20)	**<0.001**
Ever had Arthritis	519 (65%)	843 (53%)	495 (62%)	**<0.001**
Ever had Diabetes	182 (23%)	280 (18%)	181 (23%)	**0.001**
Ever had High Blood Pressure	501 (63%)	942 (59%)	492 (62%)	0.2
Ever had Cancer	156 (20%)	263 (17%)	150 (19%)	0.13
Ever had Lung Disease	70 (8.8%)	91 (5.7%)	60 (7.6%)	**0.016**
Ever had Heart Problems	221 (28%)	340 (21%)	209 (26%)	**<0.001**
Ever had a Stroke	75 (9.4%)	74 (4.7%)	53 (6.7%)	**<0.001**
Regularly Takes Prescription Drugs	705 (89%)	1,362 (86%)	703 (89%)	0.055
Self-rated Health				**<0.001**
Excellent	73 (9.2%)	226 (14%)	86 (11%)	
Very Good	276 (35%)	704 (44%)	300 (38%)	
Good	290 (37%)	534 (34%)	270 (34%)	
Fair	131 (16%)	115 (7.3%)	117 (15%)	
Poor	24 (3.0%)	7 (0.4%)	20 (2.5%)	

[1]Median (IQR); n (%)

[2]Kruskal-Wallis rank sum test; Pearson's Chi-squared test

Appendix G. Restricted Sample Results

Mortality

Frailty-disease Mismatch (FM)

Table 66. Logistic Regression Models for FM and Mortality (Restricted Sample)

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
FM Category												
Expected Ager	—	—		—	—		—	—		—	—	
Adapter	2.18	2.04, 2.32	**<0.001**	2.87	2.68, 3.09	**<0.001**	1.94	1.79, 2.10	**<0.001**	2.72	2.35, 3.16	**<0.001**
Premature Frailer	4.13	3.87, 4.42	**<0.001**	0.37	0.33, 0.41	**<0.001**	0.49	0.43, 0.54	**<0.001**	1.30	1.05, 1.59	**0.015**
Frailty Index				1.26	1.25, 1.27	**<0.001**	1.21	1.20, 1.22	**<0.001**	1.27	1.26, 1.29	**<0.001**
Age							1.06	1.05, 1.06	**<0.001**	1.05	1.05, 1.06	**<0.001**
Sex												
Female							—	—		—	—	
Male							1.99	1.87, 2.13	**<0.001**	2.04	1.91, 2.18	**<0.001**
FM Category * Frailty Index												
Adapter * Frailty Index										0.95	0.93, 0.97	**<0.001**
Premature Frailer * Frailty Index										0.91	0.90, 0.93	**<0.001**

Model 1 is unadjusted. Model 2 is adjusted for frailty. Model 3 is further adjusted for age and sex. Model 4 further includes a frailty interaction. Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference.

Rate of Aging (RoA)

Table 67. Logistic Regression Models for RoA and Mortality (Restricted Sample)

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
RoA Category												
Average Ager	—	—		—	—		—	—		—	—	
Slow Ager	0.35	0.32, 0.38	**<0.001**	0.62	0.57, 0.67	**<0.001**	0.99	0.90, 1.09	0.8	0.70	0.59, 0.83	**<0.001**
Fast Ager	6.23	5.80, 6.70	**<0.001**	3.09	2.86, 3.35	**<0.001**	0.93	0.83, 1.04	0.2	1.40	1.18, 1.65	**<0.001**
Frailty Index				1.12	1.11, 1.13	**<0.001**	1.15	1.14, 1.15	**<0.001**	1.16	1.15, 1.17	**<0.001**
Age							1.08	1.07, 1.08	**<0.001**	1.08	1.07, 1.08	**<0.001**
Sex												
Female							—	—		—	—	
Male							1.88	1.76, 2.01	**<0.001**	1.90	1.78, 2.02	**<0.001**
RoA Category * Frailty Index												
Slow Ager * Frailty Index										1.12	1.08, 1.17	**<0.001**
Fast Ager * Frailty Index										0.96	0.95, 0.98	**<0.001**

Model 1 is unadjusted. Model 2 is adjusted for frailty. Model 3 is further adjusted for age and sex. Model 4 further includes a frailty interaction Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference.

Table 68. Logistic Regression Models for DIOR-FI and Mortality (Restricted Sample)

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
DIOR-FI Category												
Average Stability	—	—		—	—		—	—		—	—	
High Stability	0.38	0.35, 0.41	**<0.001**	0.58	0.53, 0.63	**<0.001**	0.62	0.57, 0.67	**<0.001**	0.52	0.45, 0.59	**<0.001**
Low Stability	3.18	2.98, 3.40	**<0.001**	1.27	1.18, 1.38	**<0.001**	1.31	1.20, 1.42	**<0.001**	2.02	1.73, 2.36	**<0.001**
Frailty Index				1.15	1.15, 1.16	**<0.001**	1.13	1.12, 1.13	**<0.001**	1.14	1.13, 1.15	**<0.001**
Age							1.07	1.07, 1.08	**<0.001**	1.07	1.07, 1.08	**<0.001**
Sex												
Female							—	—		—	—	
Male							1.89	1.77, 2.01	**<0.001**	1.91	1.79, 2.04	**<0.001**
DIOR-FI Category * Frailty Index												
High Stability * Frailty Index										1.04	1.02, 1.06	**<0.001**
Low Stability * Frailty Index										0.96	0.95, 0.98	**<0.001**

Model 1 is unadjusted. Model 2 is adjusted for frailty. Model 3 is further adjusted for age and sex. Model 4 further includes a frailty interaction. Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference.

Recovery

Frailty-disease Mismatch (FM)

Table 69. Logistic Regression Models for FM and Recovery (Restricted Sample)

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
Modified SF-36 PFS	1.03	0.99, 1.07	0.12	0.89	0.84, 0.95	**<0.001**	0.89	0.84, 0.95	**<0.001**	0.79	0.73, 0.85	**<0.001**
FM Category												
Expected Ager	—	—		—	—		—	—		—	—	
Adapter	0.55	0.42, 0.73	**<0.001**	0.55	0.41, 0.72	**<0.001**	0.65	0.48, 0.86	**0.003**	0.57	0.34, 0.98	**0.042**
Premature Frailer	1.08	0.80, 1.47	0.6	1.81	1.29, 2.54	**<0.001**	1.50	1.06, 2.13	**0.023**	0.26	0.11, 0.58	**0.001**
Frailty Index				0.89	0.85, 0.92	**<0.001**	0.91	0.88, 0.95	**<0.001**	0.78	0.72, 0.84	**<0.001**
Age							0.96	0.95, 0.97	**<0.001**	0.96	0.95, 0.97	**<0.001**
FM Category * Frailty Index												
Adapter * Frailty Index										1.00	0.90, 1.11	>0.9
Premature Frailer * Frailty Index										1.19	1.11, 1.29	**<0.001**

Model 1 is adjusted for function. Model 2 is further adjusted for frailty. Model 3 is further adjusted for age. Model 4 further includes a frailty interaction. Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference.

Rate of Aging (RoA)

Table 70. Logistic Regression Models for RoA and Recovery (Restricted Sample)

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
Modified SF-36 PFS	0.95	0.92, 0.98	**0.003**	0.88	0.83, 0.94	**<0.001**	0.88	0.83, 0.94	**<0.001**	0.87	0.81, 0.93	**<0.001**
RoA Category												
Average Ager	—	—		—	—		—	—		—	—	
Slow Ager	1.39	1.09, 1.77	**0.007**	1.39	1.09, 1.77	**0.007**	1.21	0.94, 1.55	0.13	1.69	1.13, 2.51	**0.010**
Fast Ager	0.37	0.27, 0.51	**<0.001**	0.43	0.31, 0.59	**<0.001**	0.79	0.53, 1.19	0.3	0.65	0.32, 1.31	0.2
Frailty Index				0.95	0.91, 0.98	**0.003**	0.94	0.91, 0.98	**0.001**	0.95	0.90, 0.99	**0.018**
Age							0.96	0.95, 0.98	**<0.001**	0.96	0.94, 0.97	**<0.001**
RoA Category * Frailty Index												
Slow Ager * Frailty Index										0.95	0.90, 1.00	**0.034**
Fast Ager * Frailty Index										1.01	0.96, 1.07	0.6

Model 1 is adjusted for function. Model 2 is further adjusted for frailty. Model 3 is further adjusted for age. Model 4 further includes a frailty interaction. Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference.

Dynamical Indicator of Resilience (DIOR-FI)

Table 71. Logistic Regression Models for DIOR-FI and Recovery (Restricted Sample)

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
Modified SF-36 PFS	0.97	0.94, 1.01	0.12	0.87	0.82, 0.92	**<0.001**	0.87	0.82, 0.93	**<0.001**	0.84	0.78, 0.90	**<0.001**
DIOR-FI Category												
Average Stability	—	—		—	—		—	—		—	—	
High Stability	1.58	1.22, 2.03	**<0.001**	1.56	1.21, 2.01	**<0.001**	1.50	1.16, 1.94	**0.002**	2.28	1.45, 3.60	**<0.001**
Low Stability	0.89	0.67, 1.17	0.4	1.09	0.82, 1.45	0.6	1.02	0.76, 1.38	0.9	0.68	0.37, 1.22	0.2
Frailty Index				0.92	0.89, 0.95	**<0.001**	0.94	0.91, 0.98	**<0.001**	0.92	0.87, 0.97	**0.001**
Age							0.95	0.94, 0.97	**<0.001**	0.96	0.94, 0.97	**<0.001**
DIOR-FI Category * Frailty Index												
High Stability * Frailty Index										0.92	0.85, 0.98	**0.018**
Low Stability * Frailty Index										1.03	0.98, 1.09	0.2

Model 1 is adjusted for function. Model 2 is further adjusted for frailty. Model 3 is further adjusted for age. Model 4 further includes a frailty interaction. Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference.

Appendix H. Age-Stratified Results

Note: provided for mortality only

Table 72. Age-Only Mixed Effects Model Results (52-67)

Fixed Effects	Lower 95%	Estimate	Upper 95%
(Intercept)	0.10162	0.10592	0.11022
Age	0.00289	0.00457	0.00626
Age^2	-0.00026	-0.00003	0.00019
Age^3	0.00000	0.00000	0.00001
Random Effects	**Lower 95%**	**Estimate**	**Upper 95%**
Intercept (sd)	0.09746	0.10058	0.10379
Age (sd)	0.00699	0.00739	0.00782
Correlation (age and intercept)	0.00379	0.09234	0.17946
Correlation Structure	**Lower 95%**	**Estimate**	**Upper 95%**
Phi (autocorrelation of residuals)	0.49869	0.53239	0.56579
Residuals	**Lower 95%**	**Estimate**	**Upper 95%**
Within-group residuals (standard error)	0.06625	0.06771	0.06920

Table 73. Adjusted Mixed Effects Model Results (52-67)

Fixed Effects	Lower 95%	Estimate	Upper 95%
(Intercept)	0.04669	0.05372	0.06075
Age	0.00002	0.00165	0.00327
Age^2	-0.00024	-0.00003	0.00018
Age^3	0.00000	0.00000	0.00001
Sex: Male	-0.02312	-0.01967	-0.01621
SRH: Very Good	0.00243	0.00503	0.00763
SRH: Good	0.02109	0.02396	0.02682
SRH: Fair	0.06223	0.06547	0.06871
SRH: Poor	0.13213	0.13619	0.14025
Ever had stroke: Yes	0.08321	0.08886	0.09451
Ever had arthritis: Yes	0.03839	0.04094	0.04349
Ever had cancer: Yes	0.00657	0.01099	0.01540
Ever had high blood pressure: Yes	0.02692	0.02963	0.03235
Ever had diabetes: Yes	0.00726	0.01049	0.01371
Ever had lung disease: Yes	0.05113	0.05582	0.06052
Ever had heart problems: Yes	0.02310	0.02680	0.03049
Regularly Takes Rx Meds: Yes	0.01201	0.01428	0.01654
Wave 4	-0.01827	-0.01277	-0.00726
Wave 5	-0.01090	-0.00499	0.00091
Wave 6	-0.01144	-0.00518	0.00107
Wave 7	-0.01152	-0.00551	0.00050
Wave 8	-0.00959	-0.00343	0.00272
Wave 9	-0.01797	-0.01165	-0.00533
Wave 10	-0.01390	-0.00759	-0.00128
Wave 11	-0.01656	-0.00997	-0.00338
Wave 12	-0.01912	-0.01215	-0.00518
Wave 13	-0.02670	-0.01912	-0.01154
Random Effects	Lower 95%	Estimate	Upper 95%
Intercept (sd)	0.06289	0.06691	0.07118
Age (sd)	0.00548	0.00600	0.00658
Correlation (age and intercept)	-0.22322	-0.09763	0.03115
Correlation Structure	Lower 95%	Estimate	Upper 95%
Phi (autocorrelation of residuals)	0.41181	0.44319	0.47503
Residuals	Lower 95%	Estimate	Upper 95%
Within-group residuals (standard error)	0.06066	0.06186	0.06308

SRH stands for self-rated health. The reference category is “Excellent”. The reference category for Wave is 3.

Table 74. Age-Only Mixed Effects Model Results (68-79)

Fixed Effects	Lower 95%	Estimate	Upper 95%
(Intercept)	0.07161	0.07729	0.08297
Age	0.00636	0.00761	0.00885
Age^2	-0.00053	-0.00044	-0.00035
Age^3	0.00001	0.00002	0.00002
Random Effects	**Lower 95%**	**Estimate**	**Upper 95%**
Intercept (sd)	0.08433	0.08807	0.09197
Age (sd)	0.00662	0.00684	0.00706
Correlation (age and intercept)	-0.38516	-0.34138	-0.29607
Correlation Structure	**Lower 95%**	**Estimate**	**Upper 95%**
Phi (autocorrelation of residuals)	0.62835	0.63928	0.65005
Residuals	**Lower 95%**	**Estimate**	**Upper 95%**
Within-group residuals (standard error)	0.07386	0.07475	0.07566

Table 75. Adjusted Mixed Effects Model Results (68-79)

Fixed Effects	Lower 95%	Estimate	Upper 95%
(Intercept)	0.04876	0.05422	0.05967
Age	0.00151	0.00269	0.00387
Age^2	-0.00037	-0.00028	-0.00020
Age^3	0.00001	0.00001	0.00001
Sex: Male	-0.02750	-0.02437	-0.02125
SRH: Very Good	0.00516	0.00711	0.00907
SRH: Good	0.02336	0.02552	0.02767
SRH: Fair	0.05912	0.06160	0.06407
SRH: Poor	0.12940	0.13254	0.13568
Ever had stroke: Yes	0.08479	0.08879	0.09279
Ever had arthritis: Yes	0.02898	0.03109	0.03320
Ever had cancer: Yes	0.00265	0.00581	0.00897
Ever had high blood pressure: Yes	0.01861	0.02077	0.02294
Ever had diabetes: Yes	0.00932	0.01203	0.01474
Ever had lung disease: Yes	0.03409	0.03766	0.04124
Ever had heart problems: Yes	0.01372	0.01634	0.01897
Regularly Takes Rx Meds: Yes	0.00945	0.01135	0.01324
Wave 4	-0.00815	-0.00610	-0.00404
Wave 5	-0.00280	-0.00034	0.00211
Wave 6	-0.00452	-0.00170	0.00112
Wave 7	-0.00443	-0.00130	0.00182
Wave 8	-0.00411	-0.00058	0.00296
Wave 9	-0.01058	-0.00659	-0.00260
Wave 10	-0.00941	-0.00483	-0.00026
Wave 11	-0.01871	-0.01360	-0.00848
Wave 12	-0.02467	-0.01892	-0.01318
Wave 13	-0.04054	-0.03394	-0.02735
Random Effects	**Lower 95%**	**Estimate**	**Upper 95%**
Intercept (sd)	0.05649	0.06033	0.06443
Age (sd)	0.00528	0.00548	0.00568
Correlation (age and intercept)	-0.52906	-0.48397	-0.43615
Correlation Structure	**Lower 95%**	**Estimate**	**Upper 95%**
Phi (autocorrelation of residuals)	0.55327	0.57127	0.58908
Residuals	**Lower 95%**	**Estimate**	**Upper 95%**
Within-group residuals (standard error)	0.06668	0.06761	0.06855

SRH stands for self-rated health. The reference category is “Excellent”. The reference category for Wave is 3.

Table 76. Age-Only Mixed Effects Model Results (80+)

Fixed Effects	Lower 95%	Estimate	Upper 95%
(Intercept)	0.08070	0.10442	0.12814
Age	-0.00051	0.00227	0.00506
Age^2	-0.00041	-0.00031	-0.00021
Age^3	0.00001	0.00001	0.00001
Random Effects	**Lower 95%**	**Estimate**	**Upper 95%**
Intercept (sd)	0.10715	0.11813	0.13024
Age (sd)	0.00761	0.00791	0.00822
Correlation (age and intercept)	-0.94799	-0.93481	-0.91842
Correlation Structure	**Lower 95%**	**Estimate**	**Upper 95%**
Phi (autocorrelation of residuals)	0.74669	0.75553	0.76416
Residuals	**Lower 95%**	**Estimate**	**Upper 95%**
Within-group residuals (standard error)	0.10766	0.10931	0.11098

Table 77. Adjusted Mixed Effects Model Results (80+)

Fixed Effects	Lower 95%	Estimate	Upper 95%
(Intercept)	0.02674	0.04855	0.07035
Age	0.00092	0.00351	0.00610
Age^2	-0.00050	-0.00040	-0.00031
Age^3	0.00001	0.00001	0.00001
Sex: Male	-0.03113	-0.02774	-0.02436
SRH: Very Good	0.00805	0.01056	0.01308
SRH: Good	0.02704	0.02972	0.03240
SRH: Fair	0.06574	0.06873	0.07171
SRH: Poor	0.13655	0.14019	0.14383
Ever had stroke: Yes	0.08735	0.09124	0.09512
Ever had arthritis: Yes	0.03379	0.03639	0.03899
Ever had cancer: Yes	0.00258	0.00613	0.00968
Ever had high blood pressure: Yes	0.02362	0.02626	0.02890
Ever had diabetes: Yes	0.01273	0.01632	0.01991
Ever had lung disease: Yes	0.03971	0.04424	0.04877
Ever had heart problems: Yes	0.01401	0.01685	0.01969
Regularly Takes Rx Meds: Yes	0.01180	0.01434	0.01688
Wave 4	-0.00660	-0.00434	-0.00207
Wave 5	-0.00354	-0.00077	0.00201
Wave 6	0.00080	0.00398	0.00716
Wave 7	0.00159	0.00510	0.00862
Wave 8	0.00417	0.00806	0.01194
Wave 9	0.00007	0.00437	0.00868
Wave 10	0.00305	0.00795	0.01285
Wave 11	0.00120	0.00661	0.01201
Wave 12	0.00189	0.00795	0.01400
Wave 13	-0.00404	0.00294	0.00992
Random Effects	**Lower 95%**	**Estimate**	**Upper 95%**
Intercept (sd)	0.09219	0.10080	0.11022
Age (sd)	0.00626	0.00650	0.00675
Correlation (age and intercept)	-0.98914	-0.97932	-0.96077
Correlation Structure	**Lower 95%**	**Estimate**	**Upper 95%**
Phi (autocorrelation of residuals)	0.69124	0.70066	0.70991
Residuals	**Lower 95%**	**Estimate**	**Upper 95%**
Within-group residuals (standard error)	0.09497	0.09615	0.09734

SRH stands for self-rated health. The reference category is “Excellent”. The reference category for Wave is 3.

Frailty-disease Mismatch (FM)

Table 78. Logistic Regression Models for FM and Mortality (Ages 52-67)

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
FM Category												
Expected Ager	—	—		—	—		—	—		—	—	
Adapter	1.73	1.50, 2.00	**<0.001**	2.13	1.83, 2.48	**<0.001**	2.01	1.72, 2.34	**<0.001**	2.24	1.67, 2.99	**<0.001**
Premature Frailer	3.04	2.66, 3.48	**<0.001**	0.48	0.39, 0.59	**<0.001**	0.50	0.41, 0.62	**<0.001**	2.15	1.55, 2.98	**<0.001**
Frailty Index				1.22	1.20, 1.24	**<0.001**	1.22	1.20, 1.24	**<0.001**	1.38	1.34, 1.43	**<0.001**
Age							1.07	1.05, 1.09	**<0.001**	1.06	1.04, 1.08	**<0.001**
Sex												
Female							—	—		—	—	
Male							1.78	1.57, 2.01	**<0.001**	1.84	1.63, 2.09	**<0.001**
FM Category * Frailty Index												
Adapter * Frailty Index										1.02	0.96, 1.09	0.5
Premature Frailer * Frailty Index										0.83	0.80, 0.86	**<0.001**

Model 1 is unadjusted. Model 2 is adjusted for frailty. Model 3 is further adjusted for age and sex. Model 4 further includes a frailty interaction. Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference.

Table 79. Logistic Regression Models for FM and Mortality (Ages 68-79)

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
FM Category												
Expected Ager	—	—		—	—		—	—		—	—	
Adapter	1.46	1.31, 1.63	**<0.001**	2.43	2.16, 2.73	**<0.001**	2.81	2.47, 3.20	**<0.001**	3.65	2.82, 4.73	**<0.001**
Premature Frailer	3.03	2.73, 3.37	**<0.001**	0.35	0.30, 0.41	**<0.001**	0.30	0.25, 0.36	**<0.001**	1.54	1.12, 2.12	**0.008**
Frailty Index				1.25	1.24, 1.27	**<0.001**	1.29	1.27, 1.31	**<0.001**	1.41	1.38, 1.45	**<0.001**
Age							0.96	0.95, 0.98	**<0.001**	0.95	0.94, 0.97	**<0.001**
Sex												
Female							—	—		—	—	
Male							2.06	1.87, 2.28	**<0.001**	2.16	1.96, 2.39	**<0.001**
FM Category * Frailty Index												
Adapter * Frailty Index										1.00	0.95, 1.05	>0.9
Premature Frailer * Frailty Index										0.85	0.82, 0.87	**<0.001**

Model 1 is unadjusted. Model 2 is adjusted for frailty. Model 3 is further adjusted for age and sex. Model 4 further includes a frailty interaction. Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference.

Table 80. Logistic Regression Models for FM and Mortality (Ages 80-109)

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
FM Category												
Expected Ager	—	—		—	—		—	—		—	—	
Adapter	1.0	0.90, 1.10	>0.9	1.85	1.66, 2.08	**<0.001**	1.44	1.26, 1.64	**<0.001**	1.68	1.34, 2.11	**<0.001**
Premature Frailer	3.63	3.21, 4.12	**<0.001**	0.51	0.42, 0.61	**<0.001**	0.62	0.51, 0.75	**<0.001**	1.13	0.67, 1.92	0.6
Frailty Index				1.18	1.17, 1.19	**<0.001**	1.16	1.15, 1.18	**<0.001**	1.18	1.16, 1.19	**<0.001**
Age							1.06	1.04, 1.07	**<0.001**	1.06	1.04, 1.07	**<0.001**
Sex												
Female							—	—		—	—	
Male							1.81	1.64, 1.99	**<0.001**	1.81	1.64, 2.00	**<0.001**
FM Category * Frailty Index												
Adapter * Frailty Index										0.98	0.95, 1.01	0.2
Premature Frailer * Frailty Index										0.96	0.94, 0.99	**0.010**

Model 1 is unadjusted. Model 2 is adjusted for frailty. Model 3 is further adjusted for age and sex. Model 4 further includes a frailty interaction. Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference.

Table 81. Logistic Regression Models for RoA and Mortality (Ages 52-67)

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
RoA Category												
Average Ager	—	—		—	—		—	—		—	—	
Slow Ager	0.52	0.43, 0.63	**<0.001**	0.66	0.54, 0.80	**<0.001**	0.64	0.52, 0.77	**<0.001**	0.82	0.59, 1.12	0.2
Fast Ager	3.64	3.22, 4.13	**<0.001**	1.40	1.15, 1.69	**<0.001**	1.34	1.11, 1.63	**0.003**	2.29	1.68, 3.13	**<0.001**
Frailty Index				1.11	1.09, 1.13	**<0.001**	1.12	1.10, 1.13	**<0.001**	1.21	1.16, 1.26	**<0.001**
Age							1.09	1.07, 1.11	**<0.001**	1.09	1.07, 1.11	**<0.001**
Sex												
Female							—	—		—	—	
Male							1.72	1.52, 1.95	**<0.001**	1.76	1.55, 1.99	**<0.001**
RoA Category * Frailty Index												
Slow Ager * Frailty Index										0.97	0.86, 1.08	0.6
Fast Ager * Frailty Index										0.91	0.87, 0.95	**<0.001**

Model 1 is unadjusted. Model 2 is adjusted for frailty. Model 3 is further adjusted for age and sex. Model 4 further includes a frailty interaction Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference.

Table 82. Logistic Regression Models for RoA and Mortality (Ages 68-79)

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
RoA Category												
Average Ager	—	—		—	—		—	—		—	—	
Slow Ager	0.52	0.46, 0.58	**<0.001**	0.72	0.63, 0.82	**<0.001**	0.74	0.63, 0.86	**<0.001**	1.00	0.77, 1.29	>0.9
Fast Ager	3.55	3.19, 3.94	**<0.001**	1.29	1.11, 1.49	**<0.001**	1.26	1.07, 1.49	**0.006**	1.81	1.37, 2.39	**<0.001**
Frailty Index				1.12	1.11, 1.13	**<0.001**	1.13	1.11, 1.14	**<0.001**	1.16	1.14, 1.19	**<0.001**
Age							1.00	0.99, 1.02	0.6	1.01	0.99, 1.03	0.2
Sex												
Female							—	—		—	—	
Male							1.94	1.76, 2.13	**<0.001**	1.95	1.78, 2.15	**<0.001**
RoA Category * Frailty Index												
Slow Ager * Frailty Index										0.94	0.89, 1.00	**0.038**
Fast Ager * Frailty Index										0.96	0.93, 0.98	**<0.001**

Model 1 is unadjusted. Model 2 is adjusted for frailty. Model 3 is further adjusted for age and sex. Model 4 further includes a frailty interaction Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference.

Table 83. Logistic Regression Models for RoA and Mortality (Ages 80-109)

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
RoA Category												
Average Ager	—	—		—	—		—	—		—	—	
Slow Ager	0.33	0.29, 0.36	**<0.001**	0.60	0.53, 0.67	**<0.001**	0.99	0.85, 1.15	>0.9	0.77	0.60, 1.00	**0.047**
Fast Ager	3.39	2.97, 3.90	**<0.001**	1.45	1.23, 1.71	**<0.001**	0.72	0.58, 0.88	**0.002**	0.73	0.47, 1.13	0.2
Frailty Index				1.10	1.09, 1.11	**<0.001**	1.13	1.12, 1.14	**<0.001**	1.13	1.11, 1.14	**<0.001**
Age							1.09	1.08, 1.11	**<0.001**	1.09	1.08, 1.11	**<0.001**
Sex												
Female							—	—		—	—	
Male							1.75	1.59, 1.93	**<0.001**	1.76	1.59, 1.94	**<0.001**
RoA Category * Frailty Index												
Slow Ager * Frailty Index										1.05	1.01, 1.09	**0.014**
Fast Ager * Frailty Index										1.00	0.98, 1.02	>0.9

Model 1 is unadjusted. Model 2 is adjusted for frailty. Model 3 is further adjusted for age and sex. Model 4 further includes a frailty interaction Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference.

Table 84. Logistic Regression Models for DIOR-FI and Mortality (Ages 52-67)

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
DIOR-FI Category												
Average Stability	—	—		—	—		—	—		—	—	
High Stability	0.47	0.39, 0.56	**<0.001**	0.65	0.54, 0.78	**<0.001**	0.65	0.54, 0.78	**<0.001**	0.51	0.39, 0.67	**<0.001**
Low Stability	2.28	2.01, 2.58	**<0.001**	1.17	1.01, 1.35	**0.037**	1.18	1.02, 1.36	**0.029**	1.72	1.35, 2.18	**<0.001**
Frailty Index				1.13	1.12, 1.14	**<0.001**	1.13	1.12, 1.15	**<0.001**	1.15	1.13, 1.17	**<0.001**
Age							1.09	1.07, 1.11	**<0.001**	1.09	1.07, 1.11	**<0.001**
Sex												
Female							—	—		—	—	
Male							1.73	1.53, 1.96	**<0.001**	1.75	1.55, 1.98	**<0.001**
DIOR-FI Category * Frailty Index												
High Stability * Frailty Index										1.09	1.03, 1.14	**0.001**
Low Stability * Frailty Index										0.96	0.94, 0.98	**<0.001**

Model 1 is unadjusted. Model 2 is adjusted for frailty. Model 3 is further adjusted for age and sex. Model 4 further includes a frailty interaction. Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference.

Table 85. Logistic Regression Models for DIOR-FI and Mortality (Ages 68-79)

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
DIOR-FI Category												
Average Stability	—	—		—	—		—	—		—	—	
High Stability	0.47	0.42, 0.53	**<0.001**	0.63	0.55, 0.71	**<0.001**	0.62	0.55, 0.71	**<0.001**	0.42	0.34, 0.52	**<0.001**
Low Stability	2.52	2.27, 2.79	**<0.001**	1.27	1.13, 1.43	**<0.001**	1.24	1.10, 1.41	**<0.001**	1.59	1.28, 1.98	**<0.001**
Frailty Index				1.12	1.11, 1.13	**<0.001**	1.13	1.12, 1.14	**<0.001**	1.13	1.12, 1.15	**<0.001**
Age							1.03	1.02, 1.04	**<0.001**	1.03	1.01, 1.04	**<0.001**
Sex												
Female							—	—		—	—	
Male							1.93	1.76, 2.13	**<0.001**	1.96	1.78, 2.16	**<0.001**
DIOR-FI Category * Frailty Index												
High Stability * Frailty Index										1.09	1.05, 1.13	**<0.001**
Low Stability * Frailty Index										0.98	0.96, 1.00	**0.034**

Model 1 is unadjusted. Model 2 is adjusted for frailty. Model 3 is further adjusted for age and sex. Model 4 further includes a frailty interaction. Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference.

Table 86. Logistic Regression Models for DIOR-FI and Mortality (Ages 80-109)

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
DIOR-FI Category												
Average Stability	—	—		—	—		—	—		—	—	
High Stability	0.45	0.41, 0.50	**<0.001**	0.64	0.58, 0.71	**<0.001**	0.68	0.61, 0.76	**<0.001**	0.62	0.51, 0.76	**<0.001**
Low Stability	2.89	2.55, 3.29	**<0.001**	1.26	1.09, 1.46	**0.002**	1.25	1.08, 1.45	**0.003**	2.23	1.62, 3.09	**<0.001**
Frailty Index				1.11	1.10, 1.12	**<0.001**	1.11	1.10, 1.12	**<0.001**	1.12	1.10, 1.13	**<0.001**
Age							1.08	1.07, 1.09	**<0.001**	1.08	1.07, 1.09	**<0.001**
Sex												
Female							—	—		—	—	
Male							1.75	1.58, 1.93	**<0.001**	1.77	1.60, 1.95	**<0.001**
DIOR-FI Category * Frailty Index												
High Stability * Frailty Index										1.02	0.99, 1.04	0.2
Low Stability * Frailty Index										0.96	0.95, 0.98	**<0.001**

Model 1 is unadjusted. Model 2 is adjusted for frailty. Model 3 is further adjusted for age and sex. Model 4 further includes a frailty interaction. Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference.

Appendix I. Sex-Stratified Results

Note: provided for mortality only

Table 87. Age-Only Mixed Effects Model Results (Males)

Fixed Effects	Lower 95%	Estimate	Upper 95%
(Intercept)	0.06710	0.07217	0.07724
Age	0.00615	0.00703	0.00791
Age^2	-0.00047	-0.00043	-0.00038
Age^3	0.00001	0.00001	0.00001
Random Effects	**Lower 95%**	**Estimate**	**Upper 95%**
Intercept (sd)	0.08106	0.08690	0.09317
Age (sd)	0.00607	0.00635	0.00664
Correlation (age and intercept)	-0.54281	-0.48401	-0.42049
Correlation Structure	**Lower 95%**	**Estimate**	**Upper 95%**
Phi (autocorrelation of residuals)	0.73166	0.74191	0.75191
Residuals	**Lower 95%**	**Estimate**	**Upper 95%**
Within-group residuals (standard error)	0.09079	0.09231	0.09385

Table 88. Adjusted Mixed Effects Model Results (Males)

Fixed Effects	Lower 95%	Estimate	Upper 95%
(Intercept)	0.02068	0.02552	0.03036
Age	0.00313	0.00390	0.00467
Age^2	-0.00040	-0.00036	-0.00032
Age^3	0.00001	0.00001	0.00001
SRH: Very Good	0.00384	0.00584	0.00784
SRH: Good	0.02075	0.02291	0.02508
SRH: Fair	0.05393	0.05638	0.05882
SRH: Poor	0.13013	0.13324	0.13634
Ever had stroke: Yes	0.08812	0.09178	0.09544
Ever had arthritis: Yes	0.03285	0.03500	0.03715
Ever had cancer: Yes	0.00724	0.01031	0.01338
Ever had high blood pressure: Yes	0.02178	0.02401	0.02623
Ever had diabetes: Yes	0.01237	0.01505	0.01774
Ever had lung disease: Yes	0.04521	0.04891	0.05261
Ever had heart problems: Yes	0.01268	0.01517	0.01766
Regularly Takes Rx Meds: Yes	0.00999	0.01190	0.01381
Wave 4	-0.00717	-0.00502	-0.00288
Wave 5	-0.00226	0.00026	0.00278
Wave 6	0.00042	0.00317	0.00593
Wave 7	0.00190	0.00476	0.00762
Wave 8	0.00416	0.00716	0.01016
Wave 9	-0.00047	0.00265	0.00576
Wave 10	0.00572	0.00887	0.01202
Wave 11	0.00466	0.00794	0.01121
Wave 12	0.00764	0.01109	0.01454
Wave 13	0.00417	0.00788	0.01159
Random Effects	**Lower 95%**	**Estimate**	**Upper 95%**
Intercept (sd)	0.05643	0.06063	0.06514
Age (sd)	0.00474	0.00493	0.00512
Correlation (age and intercept)	-0.68321	-0.64558	-0.60452
Correlation Structure	**Lower 95%**	**Estimate**	**Upper 95%**
Phi (autocorrelation of residuals)	0.66923	0.67965	0.68989
Residuals	**Lower 95%**	**Estimate**	**Upper 95%**
Within-group residuals (standard error)	0.08007	0.08108	0.08210

SRH stands for self-rated health. The reference category is “Excellent”. The reference category for Wave is 3.

Table 89. Age-Only Mixed Effects Model Results (Females)

Fixed Effects	Lower 95%	Estimate	Upper 95%
(Intercept)	0.09528	0.09939	0.10350
Age	0.00609	0.00678	0.00747
Age^2	-0.00044	-0.00040	-0.00036
Age^3	0.00001	0.00001	0.00001
Random Effects	**Lower 95%**	**Estimate**	**Upper 95%**
Intercept (sd)	0.09037	0.09659	0.10324
Age (sd)	0.00527	0.00557	0.00588
Correlation (age and intercept)	-0.37918	-0.28436	-0.18364
Correlation Structure	**Lower 95%**	**Estimate**	**Upper 95%**
Phi (autocorrelation of residuals)	0.73497	0.74218	0.74927
Residuals	**Lower 95%**	**Estimate**	**Upper 95%**
Within-group residuals (standard error)	0.09621	0.09740	0.09859

Table 90. Adjusted Mixed Effects Model Results (Females)

Fixed Effects	Lower 95%	Estimate	Upper 95%
(Intercept)	0.02982	0.03386	0.03790
Age	0.00306	0.00366	0.00426
Age^2	-0.00037	-0.00034	-0.00031
Age^3	0.00001	0.00001	0.00001
SRH: Very Good	0.00649	0.00839	0.01028
SRH: Good	0.02629	0.02836	0.03043
SRH: Fair	0.06889	0.07121	0.07354
SRH: Poor	0.13751	0.14035	0.14320
Ever had stroke: Yes	0.09564	0.09905	0.10245
Ever had arthritis: Yes	0.03677	0.03874	0.04071
Ever had cancer: Yes	0.00505	0.00805	0.01105
Ever had high blood pressure: Yes	0.02617	0.02819	0.03022
Ever had diabetes: Yes	0.01543	0.01801	0.02059
Ever had lung disease: Yes	0.04766	0.05098	0.05430
Ever had heart problems: Yes	0.02432	0.02677	0.02922
Regularly Takes Rx Meds: Yes	0.01349	0.01533	0.01717
Wave 4	-0.00719	-0.00526	-0.00333
Wave 5	-0.00200	0.00027	0.00253
Wave 6	0.00019	0.00266	0.00512
Wave 7	0.00159	0.00415	0.00672
Wave 8	0.00491	0.00760	0.01028
Wave 9	0.00144	0.00423	0.00701
Wave 10	0.00671	0.00954	0.01236
Wave 11	0.00494	0.00788	0.01082
Wave 12	0.00507	0.00817	0.01126
Wave 13	0.00143	0.00476	0.00808
Random Effects	**Lower 95%**	**Estimate**	**Upper 95%**
Intercept (sd)	0.05960	0.06332	0.06727
Age (sd)	0.00428	0.00444	0.00460
Correlation (age and intercept)	-0.54119	-0.49344	-0.44251
Correlation Structure	**Lower 95%**	**Estimate**	**Upper 95%**
Phi (autocorrelation of residuals)	0.67052	0.67829	0.68596
Residuals	**Lower 95%**	**Estimate**	**Upper 95%**
Within-group residuals (standard error)	0.08480	0.08561	0.08643

SRH stands for self-rated health. The reference category is “Excellent”. The reference category for Wave is 3.

Frailty-disease Mismatch (FM)

Table 91. Logistic Regression Models for FM and Mortality (Males)

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
FM Category												
Expected Ager	—	—		—	—		—	—		—	—	
Adapter	1.83	1.68, 2.01	**<0.001**	2.24	2.03, 2.47	**<0.001**	1.22	1.09, 1.36	**<0.001**	1.73	1.41, 2.12	**<0.001**
Premature Frailer	4.08	3.71, 4.47	**<0.001**	0.39	0.34, 0.45	**<0.001**	0.65	0.56, 0.76	**<0.001**	1.83	1.40, 2.38	**<0.001**
Frailty Index				1.26	1.25, 1.28	**<0.001**	1.17	1.16, 1.19	**<0.001**	1.27	1.24, 1.29	**<0.001**
Age							1.07	1.06, 1.08	**<0.001**	1.06	1.06, 1.07	**<0.001**
FM Category * Frailty Index												
Adapter * Frailty Index										0.94	0.91, 0.98	**0.003**
Premature Frailer * Frailty Index										0.89	0.87, 0.91	**<0.001**

Model 1 is unadjusted. Model 2 is adjusted for frailty. Model 3 is further adjusted for age and sex. Model 4 further includes a frailty interaction. Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference.

Table 92. Logistic Regression Models for FM and Mortality (Females)

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
FM Category												
Expected Ager	—	—		—	—		—	—		—	—	
Adapter	1.83	1.69, 1.98	**<0.001**	2.65	2.41, 2.90	**<0.001**	1.47	1.33, 1.64	**<0.001**	1.98	1.63, 2.40	**<0.001**
Premature Frailer	4.52	4.17, 4.90	**<0.001**	0.36	0.31, 0.41	**<0.001**	0.62	0.54, 0.71	**<0.001**	1.53	1.16, 2.02	**0.002**
Frailty Index				1.27	1.26, 1.28	**<0.001**	1.18	1.17, 1.19	**<0.001**	1.23	1.21, 1.25	**<0.001**
Age							1.06	1.06, 1.07	**<0.001**	1.06	1.05, 1.06	**<0.001**
FM Category * Frailty Index												
Adapter * Frailty Index										0.96	0.94, 0.99	**0.007**
Premature Frailer * Frailty Index										0.93	0.91, 0.95	**<0.001**

Model 1 is unadjusted. Model 2 is adjusted for frailty. Model 3 is further adjusted for age and sex. Model 4 further includes a frailty interaction. Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference.

Rate of Aging (RoA)

Table 93. Logistic Regression Models for RoA and Mortality (Males)

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
RoA Category												
Average Ager	—	—		—	—		—	—		—	—	
Slow Ager	0.40	0.35, 0.44	**<0.001**	0.50	0.45, 0.56	**<0.001**	0.64	0.57, 0.72	**<0.001**	0.66	0.55, 0.79	**<0.001**
Fast Ager	10.4	9.30, 11.7	**<0.001**	5.70	4.99, 6.52	**<0.001**	2.59	2.22, 3.03	**<0.001**	6.12	4.68, 8.03	**<0.001**
Frailty Index				1.07	1.06, 1.08	**<0.001**	1.09	1.08, 1.10	**<0.001**	1.13	1.11, 1.14	**<0.001**
Age							1.06	1.05, 1.06	**<0.001**	1.05	1.05, 1.06	**<0.001**
RoA Category * Frailty Index												
Slow Ager * Frailty Index										1.03	0.99, 1.07	0.11
Fast Ager * Frailty Index										0.92	0.91, 0.94	**<0.001**

Model 1 is unadjusted. Model 2 is adjusted for frailty. Model 3 is further adjusted for age and sex. Model 4 further includes a frailty interaction Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference.

Table 94. Logistic Regression Models for RoA and Mortality (Females)

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
RoA Category												
Average Ager	—	—		—	—		—	—		—	—	
Slow Ager	0.26	0.23, 0.29	**<0.001**	0.38	0.34, 0.43	**<0.001**	0.48	0.42, 0.54	**<0.001**	0.50	0.41, 0.61	**<0.001**
Fast Ager	13.0	11.8, 14.4	**<0.001**	6.83	6.10, 7.65	**<0.001**	3.55	3.10, 4.08	**<0.001**	8.73	6.79, 11.3	**<0.001**
Frailty Index				1.08	1.07, 1.09	**<0.001**	1.08	1.08, 1.09	**<0.001**	1.11	1.10, 1.12	**<0.001**
Age							1.04	1.04, 1.05	**<0.001**	1.04	1.04, 1.05	**<0.001**
RoA Category * Frailty Index												
Slow Ager * Frailty Index										1.03	0.99, 1.07	0.14
Fast Ager * Frailty Index										0.93	0.92, 0.95	**<0.001**

Model 1 is unadjusted. Model 2 is adjusted for frailty. Model 3 is further adjusted for age and sex. Model 4 further includes a frailty interaction Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference.

Dynamical Indicator of Resilience (DIOR-FI)

Table 95. Logistic Regression Models for DIOR-FI and Mortality (Males)

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
DIOR-FI Category												
Average Stability	—	—		—	—		—	—		—	—	
High Stability	0.33	0.29, 0.36	**<0.001**	0.53	0.48, 0.60	**<0.001**	0.74	0.65, 0.83	**<0.001**	0.46	0.37, 0.56	**<0.001**
Low Stability	2.90	2.65, 3.18	**<0.001**	1.02	0.91, 1.15	0.7	1.22	1.08, 1.37	**0.002**	1.93	1.56, 2.38	**<0.001**
Frailty Index				1.15	1.14, 1.17	**<0.001**	1.12	1.11, 1.13	**<0.001**	1.14	1.12, 1.16	**<0.001**
Age							1.08	1.07, 1.08	**<0.001**	1.07	1.07, 1.08	**<0.001**
DIOR-FI Category * Frailty Index												
High Stability * Frailty Index										1.18	1.12, 1.24	**<0.001**
Low Stability * Frailty Index										0.96	0.94, 0.97	**<0.001**

Model 1 is unadjusted. Model 2 is adjusted for frailty. Model 3 is further adjusted for age and sex. Model 4 further includes a frailty interaction. Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference.

Table 96. Logistic Regression Models for DIOR-FI and Mortality (Females)

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
DIOR-FI Category												
Average Stability	—	—		—	—		—	—		—	—	
High Stability	0.31	0.28, 0.35	**<0.001**	0.62	0.56, 0.69	**<0.001**	0.81	0.72, 0.90	**<0.001**	0.48	0.39, 0.59	**<0.001**
Low Stability	3.61	3.34, 3.92	**<0.001**	1.15	1.04, 1.28	**0.006**	1.30	1.17, 1.44	**<0.001**	1.96	1.59, 2.42	**<0.001**
Frailty Index				1.16	1.15, 1.17	**<0.001**	1.12	1.11, 1.13	**<0.001**	1.13	1.12, 1.14	**<0.001**
Age							1.07	1.07, 1.08	**<0.001**	1.07	1.07, 1.08	**<0.001**
DIOR-FI Category * Frailty Index												
High Stability * Frailty Index										1.13	1.09, 1.17	**<0.001**
Low Stability * Frailty Index										0.97	0.96, 0.99	**<0.001**

Model 1 is unadjusted. Model 2 is adjusted for frailty. Model 3 is further adjusted for age and sex. Model 4 further includes a frailty interaction. Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference.

Appendix J. Alternative Cut Point Sensitivity Analysis (Mortality)

Mortality

Frailty-disease Mismatch (FM)

Table 97. Logistic Regression Models for FM and Mortality (Alternative Cut Point)

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
FM Category												
Expected Ager	—	—		—	—		—	—		—	—	
Adapter	2.00	1.87, 2.14	**<0.001**	3.04	2.82, 3.28	**<0.001**	1.50	1.38, 1.64	**<0.001**	1.88	1.62, 2.18	**<0.001**
Premature Frailer	5.08	4.73, 5.47	**<0.001**	0.28	0.25, 0.32	**<0.001**	0.48	0.42, 0.54	**<0.001**	1.01	0.74, 1.38	>0.9
Frailty Index				1.25	1.24, 1.25	**<0.001**	1.18	1.18, 1.19	**<0.001**	1.20	1.19, 1.21	**<0.001**
Age							1.06	1.06, 1.07	**<0.001**	1.06	1.06, 1.07	**<0.001**
Sex												
Female							—	—		—	—	
Male							1.87	1.76, 1.99	**<0.001**	1.88	1.77, 2.00	**<0.001**
FM Category * Frailty Index												
Adapter * Frailty Index										0.96	0.94, 0.98	**<0.001**
Premature Frailer * Frailty Index										0.95	0.94, 0.97	**<0.001**

Model 1 is unadjusted. Model 2 is adjusted for frailty. Model 3 is further adjusted for age and sex. Model 4 further includes a frailty interaction. Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference.

Rate of Aging (RoA)

Table 98. Logistic Regression Models for RoA and Mortality (Alternative Cut Point)

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
RoA Category												
Average Ager	—	—		—	—		—	—		—	—	
Slow Ager	0.21	0.19, 0.23	**<0.001**	0.38	0.34, 0.42	**<0.001**	0.76	0.68, 0.85	**<0.001**	0.77	0.66, 0.89	**<0.001**
Fast Ager	8.28	7.60, 9.03	**<0.001**	2.95	2.67, 3.25	**<0.001**	0.87	0.77, 0.97	**0.017**	1.38	1.11, 1.72	**0.004**
Frailty Index				1.13	1.12, 1.13	**<0.001**	1.14	1.13, 1.14	**<0.001**	1.15	1.14, 1.15	**<0.001**
Age							1.08	1.07, 1.08	**<0.001**	1.08	1.07, 1.08	**<0.001**
Sex												
Female							—	—		—	—	
Male							1.81	1.70, 1.92	**<0.001**	1.82	1.71, 1.93	**<0.001**
RoA Category * Frailty Index												
Slow Ager * Frailty Index										1.01	0.98, 1.04	0.7
Fast Ager * Frailty Index										0.97	0.96, 0.98	**<0.001**

Model 1 is unadjusted. Model 2 is adjusted for frailty. Model 3 is further adjusted for age and sex. Model 4 further includes a frailty interaction Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference.

Table 99. Logistic Regression Models for DIOR-FI and Mortality (Alternative Cut Point)

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
DIOR-FI Category												
Average Stability	—	—		—	—		—	—		—	—	
High Stability	0.33	0.30, 0.36	**<0.001**	0.56	0.51, 0.62	**<0.001**	0.64	0.58, 0.71	**<0.001**	0.52	0.44, 0.60	**<0.001**
Low Stability	4.11	3.82, 4.42	**<0.001**	1.16	1.06, 1.27	**0.001**	1.23	1.12, 1.36	**<0.001**	2.17	1.79, 2.63	**<0.001**
Frailty Index				1.16	1.15, 1.16	**<0.001**	1.13	1.12, 1.13	**<0.001**	1.13	1.13, 1.14	**<0.001**
Age							1.08	1.07, 1.08	**<0.001**	1.07	1.07, 1.08	**<0.001**
Sex												
Female							—	—		—	—	
Male							1.81	1.70, 1.92	**<0.001**	1.83	1.72, 1.94	**<0.001**
DIOR-FI Category * Frailty Index												
High Stability * Frailty Index										1.05	1.03, 1.08	**<0.001**
Low Stability * Frailty Index										0.96	0.95, 0.97	**<0.001**

Model 1 is unadjusted. Model 2 is adjusted for frailty. Model 3 is further adjusted for age and sex. Model 4 further includes a frailty interaction. Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference.

Interaction Figures

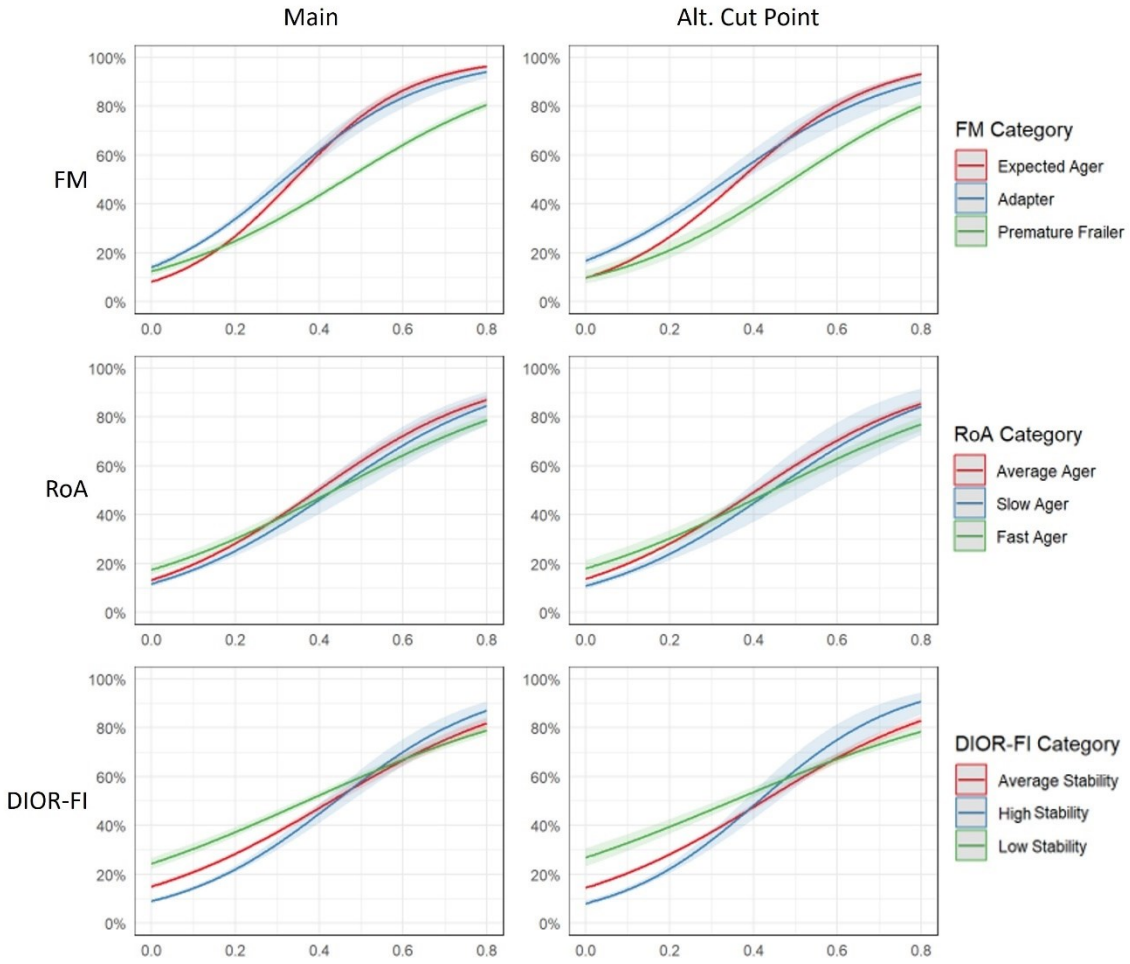


Figure 44. Mortality Main and Alternative Cut Point Comparison

Visualized interactions correspond to the frailty, age, and sex adjusted models. X-axis is the level of FI. Y-axis is the predicted probability of mortality.

Model AUC

Table 100. Discrimination of Mortality Models (Alternative Cut Point)

Model	AUC	Lower 95%	Upper 95%
Unadjusted Models			
FI Only	**0.778**	0.772	0.784
FM Only	0.631	0.625	0.636
RoA Only	0.691	0.686	0.696
DIOR Only	0.646	0.641	0.651
Frailty-Adjusted Models			
FI + FM	**0.799**	0.794	0.805
FI + RoA	0.791	0.786	0.797
FI + DIOR	0.779	0.773	0.784
Age, sex, and frailty-adjusted models			
FI Only	0.824	0.819	0.829
FI + FM	0.828	0.823	0.833
FI + FM Interaction	**0.828**	0.823	0.833
FI + RoA	0.824	0.819	0.829
FI + RoA Interaction	0.824	0.820	0.829
FI + DIOR	0.826	0.821	0.831
FI + DIOR Interaction	0.826	0.822	0.831
FI + FM + RoA + DIOR	0.829	0.824	0.834
FI + FM + RoA + DIOR + All FI Interactions	0.829	0.824	0.834

Highest AUC for each category of models is bolded, excluding the combined resilience models (bottom two rows). 95% confidence intervals are calculated using the Delong method.

Recovery

Frailty-disease Mismatch (FM)

Table 101. Logistic Regression Models for FM and Recovery (Alternative Cut Point)

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
Modified SF-36 PFS	1.02	0.99, 1.06	0.2	0.85	0.80, 0.90	**<0.001**	0.86	0.81, 0.91	**<0.001**	0.81	0.75, 0.88	**<0.001**
FM Category												
Expected Ager	—	—		—	—		—	—		—	—	
Adapter	0.47	0.33, 0.66	**<0.001**	0.45	0.32, 0.63	**<0.001**	0.58	0.41, 0.83	**0.003**	0.40	0.21, 0.77	**0.006**
Premature Frailer	1.09	0.78, 1.53	0.6	3.33	2.14, 5.23	**<0.001**	2.40	1.51, 3.84	**<0.001**	0.59	0.18, 1.91	0.4
Frailty Index				0.85	0.81, 0.89	**<0.001**	0.88	0.84, 0.92	**<0.001**	0.83	0.78, 0.88	**<0.001**
Age							0.96	0.95, 0.97	**<0.001**	0.96	0.95, 0.97	**<0.001**
FM Category * Frailty Index												
Adapter * Frailty Index										1.08	0.95, 1.22	0.2
Premature Frailer * Frailty Index										1.11	1.03, 1.20	**0.010**

Model 1 is adjusted for function. Model 2 is further adjusted for frailty. Model 3 is further adjusted for age. Model 4 further includes a frailty interaction. Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference.

Rate of Aging (RoA)

Table 102. Logistic Regression Models for RoA and Recovery (Alternative Cut Point)

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
Modified SF-36 PFS	0.96	0.93, 0.99	**0.008**	0.89	0.84, 0.94	**<0.001**	0.88	0.83, 0.93	**<0.001**	0.87	0.82, 0.93	**<0.001**
RoA Category												
Average Ager	—	—		—	—		—	—		—	—	
Slow Ager	1.56	1.18, 2.04	**0.001**	1.55	1.18, 2.03	**0.002**	1.14	0.85, 1.51	0.4	1.41	0.91, 2.18	0.12
Fast Ager	0.34	0.23, 0.49	**<0.001**	0.43	0.28, 0.64	**<0.001**	0.87	0.54, 1.38	0.6	0.60	0.22, 1.57	0.3
Frailty Index				0.95	0.91, 0.98	**0.002**	0.94	0.90, 0.97	**<0.001**	0.94	0.90, 0.98	**0.003**
Age							0.96	0.94, 0.97	**<0.001**	0.96	0.94, 0.97	**<0.001**
RoA Category * Frailty Index												
Slow Ager * Frailty Index										0.97	0.91, 1.02	0.2
Fast Ager * Frailty Index										1.02	0.96, 1.09	0.5

Model 1 is adjusted for function. Model 2 is further adjusted for frailty. Model 3 is further adjusted for age. Model 4 further includes a frailty interaction. Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference.

Dynamical Indicator of Resilience (DIOR-FI)

Table 103. Logistic Regression Models for DIOR-FI and Recovery (Alternative Cut Point)

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
Modified SF-36 PFS	0.97	0.94, 1.01	0.12	0.86	0.81, 0.92	**<0.001**	0.87	0.82, 0.93	**<0.001**	0.83	0.77, 0.89	**<0.001**
DIOR-FI Category												
Average Stability	—	—		—	—		—	—		—	—	
High Stability	1.48	1.11, 1.95	**0.007**	1.45	1.09, 1.93	**0.010**	1.34	1.00, 1.79	**0.045**	2.71	1.63, 4.55	**<0.001**
Low Stability	0.72	0.52, 0.99	**0.045**	0.99	0.69, 1.39	>0.9	0.92	0.64, 1.30	0.6	0.40	0.17, 0.88	**0.027**
Frailty Index				0.92	0.88, 0.95	**<0.001**	0.94	0.90, 0.97	**<0.001**	0.91	0.87, 0.95	**<0.001**
Age							0.95	0.94, 0.97	**<0.001**	0.96	0.94, 0.97	**<0.001**
DIOR-FI Category * Frailty Index												
High Stability * Frailty Index										0.85	0.77, 0.94	**0.001**
Low Stability * Frailty Index										1.06	1.00, 1.12	**0.034**

Model 1 is adjusted for function. Model 2 is further adjusted for frailty. Model 3 is further adjusted for age. Model 4 further includes a frailty interaction. Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference.

Interaction Figures

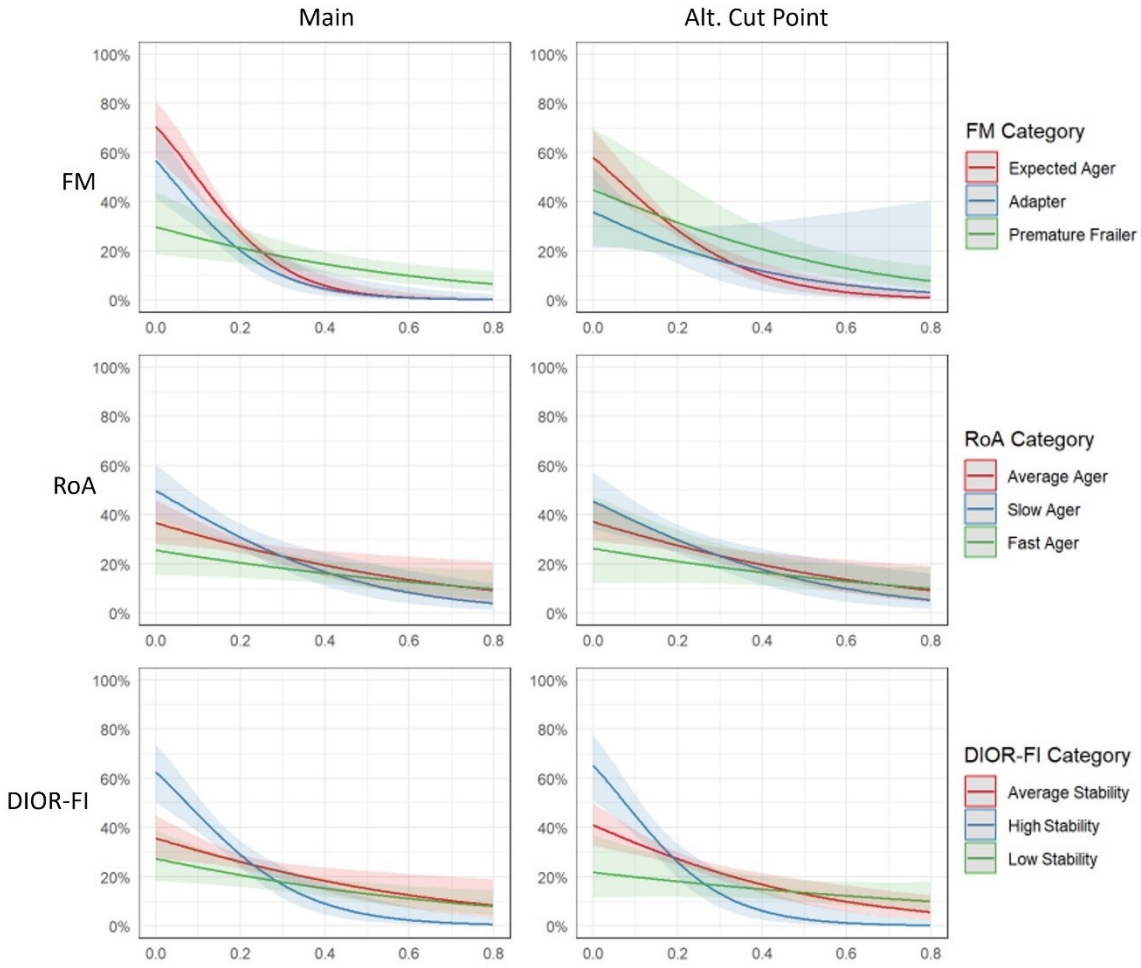


Figure 45. Recovery Main and Alternative Cut Point Comparison

Visualized interactions correspond to the frailty, age, and sex adjusted models. X-axis is the level of FI. Y-axis is the predicted probability of recovery.

Model AUC

Table 104. Discrimination of Recovery Models (Alternative Cut Point)

Model	AUC	Lower 95%	Upper 95%
Unadjusted Models			
FI Only	0.581	0.553	0.610
FM Only	0.560	0.531	0.589
RoA Only	**0.597**	0.569	0.626
DIOR Only	0.555	0.526	0.584
Frailty-Adjusted Models			
FI + FM	**0.627**	0.600	0.654
FI + RoA	0.607	0.579	0.635
FI + DIOR	0.592	0.564	0.621
Age, Sex, and Frailty-Adjusted models			
FI Only	0.643	0.616	0.670
FI + FM	0.656	0.629	0.683
FI + FM Interaction	**0.662**	0.636	0.689
FI + RoA	0.644	0.617	0.671
FI + RoA Interaction	0.645	0.618	0.672
FI + DIOR	0.647	0.620	0.674
FI + DIOR Interaction	0.659	0.632	0.686
FI + FM + DIOR	0.659	0.633	0.686

Highest AUC for each category of models is bolded, excluding the combined resilience model (bottom row). 95% confidence intervals are calculated using the Delong method.

Appendix K. Continuous Sensitivity Analysis

Mortality

FM

Table 105. Logistic Regression Models for FM and Mortality (Continuous)

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
FM	1.03	1.03, 1.03	**<0.001**	0.94	0.94, 0.94	**<0.001**	0.97	0.96, 0.97	**<0.001**	0.97	0.96, 0.97	**<0.001**
Frailty Index				1.32	1.31, 1.33	**<0.001**	1.23	1.21, 1.24	**<0.001**	1.23	1.22, 1.24	**<0.001**
Age							1.05	1.05, 1.06	**<0.001**	1.05	1.05, 1.06	**<0.001**
Sex												
Female							—	—		—	—	
Male							1.94	1.82, 2.06	**<0.001**	1.94	1.83, 2.06	**<0.001**
FM * Frailty Index										1.00	1.00, 1.00	**0.032**

Model 1 is unadjusted. Model 2 is adjusted for frailty. Model 3 is further adjusted for age and sex. Model 4 further includes a frailty interaction. Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference. FM odds ratio represents a change of 0.01.

RoA

Table 106. Logistic Regression Models for RoA and Mortality (Continuous)

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
RoA	3.65	3.52, 3.78	**<0.001**	2.48	2.38, 2.59	**<0.001**	0.89	0.80, 0.98	**0.018**	1.12	0.99, 1.26	0.066
Frailty Index				1.08	1.08, 1.09	**<0.001**	1.15	1.14, 1.15	**<0.001**	1.17	1.16, 1.18	**<0.001**
Age							1.08	1.08, 1.09	**<0.001**	1.08	1.07, 1.09	**<0.001**
Sex												
Female							—	—		—	—	
Male							1.81	1.70, 1.92	**<0.001**	1.82	1.71, 1.93	**<0.001**
RoA * Frailty Index										0.98	0.98, 0.99	**<0.001**

Model 1 is unadjusted. Model 2 is adjusted for frailty. Model 3 is further adjusted for age and sex. Model 4 further includes a frailty interaction Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference. RoA odds ratio represents a change of 0.01.

Note: there was problematic multicollinearity (indicated by variance inflation factor greater than 5) in models 3 and 4 between age and RoA, so these coefficients should not be interpreted. Rather, these models are used for comparing AUC only.

DIOR-FI

Table 107. Logistic Regression Models for DIOR-FI and Mortality (Continuous)

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
DIOR-FI	1.17	1.16, 1.17	**<0.001**	1.04	1.03, 1.05	**<0.001**	1.04	1.03, 1.05	**<0.001**	1.11	1.09, 1.12	**<0.001**
Frailty Index (41-item)				1.15	1.14, 1.16	**<0.001**	1.12	1.11, 1.12	**<0.001**	1.15	1.14, 1.16	**<0.001**
Age							1.08	1.07, 1.08	**<0.001**	1.07	1.07, 1.08	**<0.001**
Sex												
Female							—	—		—	—	
Male							1.79	1.69, 1.90	**<0.001**	1.83	1.72, 1.94	**<0.001**
DIOR-FI * Frailty Index (41-item)										1.00	1.0, 1.00	**<0.001**

Model 1 is unadjusted. Model 2 is adjusted for frailty. Model 3 is further adjusted for age and sex. Model 4 further includes a frailty interaction. Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference. DIOR-FI odds ratio represents a change of 0.01.

AUC

Table 108. Discrimination of Mortality Models (Continuous)

Model	AUC	Lower 95%	Upper 95%
Unadjusted Models			
FI Only	0.778	0.772	0.784
RoA Only	**0.801**	0.795	0.806
FM Only	0.597	0.590	0.604
DIOR Only	0.716	0.710	0.722
Frailty-Adjusted Models			
FI + RoA	**0.812**	0.807	0.817
FI + FM	0.811	0.806	0.816
FI + DIOR	0.783	0.778	0.789
Age, Sex, and Frailty-Adjusted models			
FI Only	0.824	0.819	0.829
FI + RoA	0.824	0.819	0.829
FI + RoA Interaction	0.824	0.819	0.829
FI + FM	**0.830**	0.826	0.835
FI + FM Interaction	**0.830**	0.826	0.835
FI + DIOR	0.826	0.821	0.831
FI + DIOR Interaction	0.828	0.823	0.833
FI + FM + RoA + DIOR	0.832	0.828	0.837
FI + FM + RoA + DIOR + All FI Interactions	0.833	0.828	0.838

Highest AUC for each category of models is bolded, excluding the combined resilience models (bottom two rows). 95% confidence intervals are calculated using the Delong method.

Recovery

FM

Table 109. Logistic Regression Models for FM and Recovery (Continuous)

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
Modified SF-36 PFS	1.00	1.00, 1.01	0.10	0.98	0.98, 0.99	**<0.001**	0.98	0.98, 0.99	**<0.001**	0.98	0.97, 0.99	**<0.001**
FM	1.01	1.00, 1.02	**0.008**	1.06	1.05, 1.08	**<0.001**	1.05	1.03, 1.06	**<0.001**	1.04	1.02, 1.06	**<0.001**
Frailty Index				0.80	0.76, 0.84	**<0.001**	0.84	0.80, 0.88	**<0.001**	0.82	0.77, 0.87	**<0.001**
Age							0.97	0.96, 0.98	**<0.001**	0.97	0.96, 0.98	**<0.001**
FM * Frailty Index										1.00	1.00, 1.00	0.2

Model 1 is adjusted for function. Model 2 is further adjusted for frailty. Model 3 is further adjusted for age and sex. Model 4 further includes a frailty interaction. Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference.

RoA

Table 110. Logistic Regression Models for RoA and Recovery (Continuous)

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
Modified SF-36 PFS	0.99	0.99, 1.00	**<0.001**	0.99	0.98, 0.99	**<0.001**	0.99	0.98, 0.99	**<0.001**	0.98	0.98, 0.99	**<0.001**
RoA	0.50	0.43, 0.59	**<0.001**	0.53	0.45, 0.63	**<0.001**	0.79	0.60, 1.04	0.10	0.56	0.37, 0.84	**0.006**
Frailty Index				0.96	0.93, 1.00	**0.046**	0.95	0.91, 0.98	**0.004**	0.92	0.87, 0.96	**<0.001**
Age							0.97	0.95, 0.98	**<0.001**	0.97	0.95, 0.99	**0.003**
RoA * Frailty Index										1.02	1.00, 1.04	**0.021**

Model 1 is adjusted for function. Model 2 is further adjusted for frailty. Model 3 is further adjusted for age and sex. Model 4 further includes a frailty interaction Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference.

DIOR-FI

Table 111. Logistic Regression Models for DIOR-FI and Recovery (Continuous)

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
Modified SF-36 PFS	1.00	0.99, 1.00	**0.043**	0.99	0.98, 0.99	**<0.001**	0.99	0.98, 0.99	**<0.001**	0.98	0.98, 0.99	**<0.001**
DIOR-FI	0.95	0.93, 0.98	**<0.001**	0.98	0.95, 1.00	0.094	0.97	0.94, 1.00	0.055	0.90	0.85, 0.95	**<0.001**
Frailty Index				0.93	0.89, 0.96	**<0.001**	0.95	0.91, 0.98	**0.006**	0.90	0.85, 0.94	**<0.001**
Age							0.95	0.94, 0.96	**<0.001**	0.95	0.94, 0.97	**<0.001**
DIOR-FI * Frailty Index										1.01	1.00, 1.01	**<0.001**

Model 1 is adjusted for function. Model 2 is further adjusted for frailty. Model 3 is further adjusted for age and sex. Model 4 further includes a frailty interaction. Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference.

AUC

Table 112. Discrimination of Recovery Models (Continuous)

Model	AUC	Lower 95%	Upper 95%
Unadjusted Models			
FI Only	0.581	0.553	0.610
RoA Only	**0.632**	0.605	0.659
FM Only	0.548	0.519	0.576
DIOR Only	0.559	0.531	0.587
Frailty-Adjusted Models			
FI + RoA	0.632	0.605	0.659
FI + FM	**0.645**	0.619	0.672
FI + DIOR	0.587	0.559	0.615
Age, Sex, and Frailty-Adjusted models			
FI Only	0.643	0.616	0.670
FI + RoA	0.643	0.616	0.670
FI + RoA Interaction	0.646	0.619	0.673
FI + FM	0.663	0.637	0.690
FI + FM Interaction	**0.665**	0.638	0.692
FI + DIOR	0.644	0.618	0.671
FI + DIOR Interaction	0.653	0.627	0.680
FI + FM + RoA + DIOR	0.666	0.640	0.693

Highest AUC for each category of models is bolded, excluding the combined resilience model (bottom row). 95% confidence intervals are calculated using the Delong method.

Appendix L. Alternative Frailty Index Sensitivity Analyses

51-item FI Results (Mortality)

FM

Table 113. Logistic Regression Models for FM and Mortality (51-item FI)

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
FM Category												
Expected Ager	—	—		—	—		—	—		—	—	
Adapter	1.81	1.71, 1.93	**<0.001**	2.08	1.94, 2.22	**<0.001**	1.19	1.11, 1.28	**<0.001**	1.50	1.28, 1.75	**<0.001**
Premature Frailer	4.27	4.02, 4.54	**<0.001**	0.49	0.45, 0.54	**<0.001**	0.75	0.68, 0.82	**<0.001**	1.60	1.30, 1.96	**<0.001**
Frailty Index				1.26	1.25, 1.27	**<0.001**	1.19	1.18, 1.20	**<0.001**	1.23	1.21, 1.24	**<0.001**
Age							1.07	1.06, 1.07	**<0.001**	1.07	1.06, 1.07	**<0.001**
Sex												
Female							—	—		—	—	
Male							1.80	1.70, 1.91	**<0.001**	1.82	1.72, 1.94	**<0.001**
FM Category * Frailty Index												
Adapter * Frailty Index										0.97	0.95, 0.99	**0.005**
Premature Frailer * Frailty Index										0.94	0.92, 0.95	**<0.001**

Model 1 is unadjusted. Model 2 is adjusted for frailty. Model 3 is further adjusted for age and sex. Model 4 further includes a frailty interaction. Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference.

RoA

Table 114. Logistic Regression Models for RoA and Mortality (51-item FI)

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
RoA Category												
Average Ager	—	—		—	—		—	—		—	—	
Slow Ager	0.21	0.20, 0.23	**<0.001**	0.37	0.34, 0.40	**<0.001**	0.78	0.70, 0.86	**<0.001**	0.77	0.65, 0.92	**0.003**
Fast Ager	5.96	5.58, 6.36	**<0.001**	3.00	2.78, 3.23	**<0.001**	0.99	0.89, 1.10	0.8	1.48	1.24, 1.76	**<0.001**
Frailty Index				1.12	1.11, 1.12	**<0.001**	1.15	1.15, 1.16	**<0.001**	1.17	1.16, 1.18	**<0.001**
Age							1.07	1.07, 1.08	**<0.001**	1.07	1.07, 1.08	**<0.001**
Sex												
Female							—	—		—	—	
Male							1.79	1.68, 1.90	**<0.001**	1.80	1.69, 1.91	**<0.001**
RoA Category * Frailty Index												
Slow Ager * Frailty Index										1.02	0.99, 1.05	0.2
Fast Ager * Frailty Index										0.96	0.95, 0.98	**<0.001**

Model 1 is unadjusted. Model 2 is adjusted for frailty. Model 3 is further adjusted for age and sex. Model 4 further includes a frailty interaction Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference.

DIOR-FI

Table 115. Logistic Regression Models for DIOR-FI and Mortality (51-item FI)

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
DIOR-FI Category												
Average Stability	—	—		—	—		—	—		—	—	
High Stability	0.38	0.35, 0.41	**<0.001**	0.62	0.58, 0.67	**<0.001**	0.68	0.63, 0.74	**<0.001**	0.57	0.49, 0.66	**<0.001**
Low Stability	3.20	3.02, 3.40	**<0.001**	1.25	1.16, 1.35	**<0.001**	1.29	1.20, 1.40	**<0.001**	1.82	1.55, 2.14	**<0.001**
Frailty Index				1.17	1.17, 1.18	**<0.001**	1.14	1.13, 1.15	**<0.001**	1.15	1.14, 1.16	**<0.001**
Age							1.07	1.07, 1.08	**<0.001**	1.07	1.07, 1.08	**<0.001**
Sex												
Female							—	—		—	—	
Male							1.79	1.69, 1.90	**<0.001**	1.80	1.70, 1.91	**<0.001**
DIOR-FI Category * Frailty Index												
High Stability * Frailty Index										1.03	1.02, 1.05	**<0.001**
Low Stability * Frailty Index										0.97	0.96, 0.98	**<0.001**

Model 1 is unadjusted. Model 2 is adjusted for frailty. Model 3 is further adjusted for age and sex. Model 4 further includes a frailty interaction. Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference.

51-item FI Results (Recovery)

FM

Table 116. Logistic Regression Models for FM and Recovery (51-item FI)

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
Modified SF-36 PFS	1.03	0.99, 1.07	0.2	0.86	0.81, 0.91	**<0.001**	0.85	0.80, 0.91	**<0.001**	0.74	0.68, 0.80	**<0.001**
FM Category												
Expected Ager	—	—		—	—		—	—		—	—	
Adapter	0.56	0.43, 0.74	**<0.001**	0.63	0.47, 0.82	**<0.001**	0.75	0.57, 1.00	0.051	0.80	0.44, 1.46	0.5
Premature Frailer	1.05	0.77, 1.42	0.8	1.74	1.25, 2.41	**<0.001**	1.42	1.01, 1.99	**0.043**	0.13	0.05, 0.32	**<0.001**
Frailty Index				0.86	0.82, 0.89	**<0.001**	0.88	0.84, 0.92	**<0.001**	0.75	0.69, 0.80	**<0.001**
Age							0.96	0.95, 0.97	**<0.001**	0.96	0.95, 0.97	**<0.001**
FM Category * Frailty Index												
Adapter * Frailty Index										0.99	0.90, 1.09	0.9
Premature Frailer * Frailty Index										1.23	1.15, 1.33	**<0.001**

Model 1 is adjusted for function. Model 2 is further adjusted for frailty. Model 3 is further adjusted for age. Model 4 further includes a frailty interaction. Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference.

RoA

Table 117. Logistic Regression Models for RoA and Recovery (51-item FI)

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
Modified SF-36 PFS	0.94	0.91, 0.97	**<0.001**	0.83	0.78, 0.88	**<0.001**	0.83	0.78, 0.88	**<0.001**	0.81	0.76, 0.86	**<0.001**
RoA Category												
Average Ager	—	—		—	—		—	—		—	—	
Slow Ager	1.59	1.25, 2.02	**<0.001**	1.56	1.22, 1.98	**<0.001**	1.24	0.96, 1.60	0.10	1.74	1.11, 2.74	**0.016**
Fast Ager	0.38	0.27, 0.51	**<0.001**	0.47	0.34, 0.65	**<0.001**	0.84	0.56, 1.26	0.4	0.47	0.20, 1.06	0.075
Frailty Index				0.91	0.87, 0.94	**<0.001**	0.90	0.86, 0.94	**<0.001**	0.89	0.84, 0.94	**<0.001**
Age							0.96	0.95, 0.98	**<0.001**	0.96	0.95, 0.98	**<0.001**
RoA Category * Frailty Index												
Slow Ager * Frailty Index										0.95	0.89, 1.00	0.061
Fast Ager * Frailty Index										1.04	0.98, 1.11	0.2

Model 1 is adjusted for function. Model 2 is further adjusted for frailty. Model 3 is further adjusted for age. Model 4 further includes a frailty interaction. Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference.

DIOR-FI

Table 118. Logistic Regression Models for DIOR-FI and Recovery (51-item FI)

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
Modified SF-36 PFS	0.97	0.93, 1.00	**0.044**	0.82	0.77, 0.87	**<0.001**	0.82	0.77, 0.88	**<0.001**	0.79	0.74, 0.85	**<0.001**
DIOR-FI Category												
Average Stability	—	—		—	—		—	—		—	—	
High Stability	1.56	1.22, 2.00	**<0.001**	1.52	1.18, 1.95	**0.001**	1.44	1.12, 1.86	**0.005**	3.09	1.88, 5.12	**<0.001**
Low Stability	0.79	0.59, 1.04	0.091	1.04	0.78, 1.39	0.8	0.97	0.72, 1.30	0.8	0.66	0.33, 1.31	0.2
Frailty Index				0.88	0.84, 0.91	**<0.001**	0.90	0.86, 0.94	**<0.001**	0.89	0.85, 0.94	**<0.001**
Age							0.96	0.94, 0.97	**<0.001**	0.96	0.95, 0.97	**<0.001**
DIOR-FI Category * Frailty Index												
High Stability * Frailty Index										0.87	0.81, 0.94	**<0.001**
Low Stability * Frailty Index										1.02	0.97, 1.08	0.4

Model 1 is adjusted for function. Model 2 is further adjusted for frailty. Model 3 is further adjusted for age. Model 4 further includes a frailty interaction. Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference.

56-item FI Results (Mortality)

FM

Table 119. Logistic Regression Models for FM and Mortality (56-item FI)

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
FM Category												
Expected Ager	—	—		—	—		—	—		—	—	
Adapter	1.67	1.57, 1.78	**<0.001**	2.01	1.88, 2.15	**<0.001**	1.17	1.09, 1.26	**<0.001**	1.27	1.08, 1.50	**0.003**
Premature Frailer	3.94	3.71, 4.19	**<0.001**	0.47	0.43, 0.52	**<0.001**	0.71	0.64, 0.78	**<0.001**	1.28	1.04, 1.58	**0.018**
Frailty Index				1.29	1.28, 1.30	**<0.001**	1.21	1.20, 1.22	**<0.001**	1.23	1.22, 1.25	**<0.001**
Age							1.07	1.06, 1.07	**<0.001**	1.07	1.06, 1.07	**<0.001**
Sex												
Female							—	—		—	—	
Male							1.80	1.70, 1.91	**<0.001**	1.81	1.71, 1.92	**<0.001**
FM Category * Frailty Index												
Adapter * Frailty Index										0.99	0.97, 1.01	0.5
Premature Frailer * Frailty Index										0.95	0.93, 0.96	**<0.001**

Model 1 is unadjusted. Model 2 is adjusted for frailty. Model 3 is further adjusted for age and sex. Model 4 further includes a frailty interaction. Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference.

RoA

Table 120. Logistic Regression Models for RoA and Mortality (56-item FI)

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
RoA Category												
Average Ager	—	—		—	—		—	—		—	—	
Slow Ager	0.20	0.18, 0.22	**<0.001**	0.35	0.32, 0.38	**<0.001**	0.74	0.67, 0.83	**<0.001**	0.76	0.64, 0.91	**0.003**
Fast Ager	5.84	5.47, 6.24	**<0.001**	2.99	2.78, 3.23	**<0.001**	1.00	0.90, 1.12	>0.9	1.48	1.24, 1.77	**<0.001**
Frailty Index				1.12	1.12, 1.13	**<0.001**	1.17	1.16, 1.17	**<0.001**	1.18	1.17, 1.19	**<0.001**
Age							1.07	1.07, 1.08	**<0.001**	1.07	1.07, 1.08	**<0.001**
Sex												
Female							—	—		—	—	
Male							1.78	1.68, 1.89	**<0.001**	1.79	1.69, 1.90	**<0.001**
RoA Category * Frailty Index												
Slow Ager * Frailty Index										1.01	0.98, 1.05	0.4
Fast Ager * Frailty Index										0.96	0.95, 0.98	**<0.001**

Model 1 is unadjusted. Model 2 is adjusted for frailty. Model 3 is further adjusted for age and sex. Model 4 further includes a frailty interaction Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference.

DIOR-FI

Table 121. Logistic Regression Models for DIOR-FI and Mortality (56-item FI)

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
DIOR-FI Category												
Average Stability	—	—		—	—		—	—		—	—	
High Stability	0.42	0.39, 0.45	**<0.001**	0.70	0.65, 0.75	**<0.001**	0.76	0.71, 0.82	**<0.001**	0.61	0.53, 0.71	**<0.001**
Low Stability	3.03	2.86, 3.22	**<0.001**	1.20	1.12, 1.29	**<0.001**	1.22	1.13, 1.32	**<0.001**	1.57	1.33, 1.85	**<0.001**
Frailty Index				1.20	1.19, 1.20	**<0.001**	1.16	1.15, 1.17	**<0.001**	1.16	1.15, 1.17	**<0.001**
Age							1.07	1.07, 1.08	**<0.001**	1.07	1.07, 1.08	**<0.001**
Sex												
Female							—	—		—	—	
Male							1.78	1.68, 1.89	**<0.001**	1.79	1.69, 1.90	**<0.001**
DIOR-FI Category * Frailty Index												
High Stability * Frailty Index										1.04	1.02, 1.06	**<0.001**
Low Stability * Frailty Index										0.98	0.97, 0.99	**0.003**

Model 1 is unadjusted. Model 2 is adjusted for frailty. Model 3 is further adjusted for age and sex. Model 4 further includes a frailty interaction. Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference.

56-item FI Results (Recovery)

FM

Table 122. Logistic Regression Models for FM and Recovery (56-item FI)

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
Modified SF-36 PFS	1.03	0.99, 1.07	0.2	0.86	0.81, 0.91	**<0.001**	0.85	0.80, 0.90	**<0.001**	0.76	0.71, 0.82	**<0.001**
FM Category												
Expected Ager	—	—		—	—		—	—		—	—	
Adapter	0.62	0.47, 0.81	**<0.001**	0.67	0.51, 0.88	**0.004**	0.81	0.61, 1.07	0.13	0.91	0.49, 1.67	0.8
Premature Frailer	1.07	0.80, 1.45	0.6	1.74	1.26, 2.40	**<0.001**	1.44	1.03, 2.01	**0.033**	0.19	0.08, 0.45	**<0.001**
Frailty Index				0.85	0.81, 0.89	**<0.001**	0.87	0.83, 0.91	**<0.001**	0.76	0.71, 0.82	**<0.001**
Age							0.96	0.95, 0.97	**<0.001**	0.96	0.95, 0.97	**<0.001**
FM Category * Frailty Index												
Adapter * Frailty Index										0.98	0.88, 1.08	0.7
Premature Frailer * Frailty Index										1.20	1.12, 1.30	**<0.001**

Model 1 is adjusted for function. Model 2 is further adjusted for frailty. Model 3 is further adjusted for age. Model 4 further includes a frailty interaction. Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference.

RoA

Table 123. Logistic Regression Models for RoA and Recovery (56-item FI)

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
Modified SF-36 PFS	0.94	0.91, 0.97	**<0.001**	0.83	0.78, 0.88	**<0.001**	0.83	0.79, 0.89	**<0.001**	0.82	0.77, 0.87	**<0.001**
RoA Category												
Average Ager	—	—		—	—		—	—		—	—	
Slow Ager	1.55	1.22, 1.97	**<0.001**	1.51	1.18, 1.91	**<0.001**	1.17	0.90, 1.52	0.2	1.68	1.06, 2.68	**0.027**
Fast Ager	0.37	0.27, 0.51	**<0.001**	0.46	0.33, 0.63	**<0.001**	0.83	0.55, 1.23	0.3	0.50	0.21, 1.14	0.10
Frailty Index				0.90	0.87, 0.94	**<0.001**	0.89	0.86, 0.93	**<0.001**	0.89	0.84, 0.94	**<0.001**
Age							0.96	0.95, 0.98	**<0.001**	0.96	0.95, 0.98	**<0.001**
RoA Category * Frailty Index												
Slow Ager * Frailty Index										0.94	0.88, 1.00	0.051
Fast Ager * Frailty Index										1.04	0.98, 1.11	0.2

Model 1 is adjusted for function. Model 2 is further adjusted for frailty. Model 3 is further adjusted for age. Model 4 further includes a frailty interaction. Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference.

DIOR-FI

Table 124. Logistic Regression Models for DIOR-FI and Recovery (56-item FI)

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
Modified SF-36 PFS	0.96	0.93, 1.00	**0.025**	0.83	0.78, 0.88	**<0.001**	0.83	0.78, 0.88	**<0.001**	0.81	0.76, 0.86	**<0.001**
DIOR-FI Category												
Average Stability	—	—		—	—		—	—		—	—	
High Stability	1.30	1.01, 1.66	**0.039**	1.29	1.01, 1.65	**0.044**	1.20	0.93, 1.55	0.2	2.15	1.31, 3.53	**0.003**
Low Stability	0.60	0.45, 0.80	**<0.001**	0.78	0.58, 1.04	0.095	0.72	0.53, 0.97	**0.033**	0.51	0.25, 1.03	0.063
Frailty Index				0.88	0.84, 0.92	**<0.001**	0.90	0.86, 0.94	**<0.001**	0.90	0.85, 0.95	**<0.001**
Age							0.95	0.94, 0.97	**<0.001**	0.96	0.94, 0.97	**<0.001**
DIOR-FI Category * Frailty Index												
High Stability * Frailty Index										0.91	0.84, 0.97	**0.007**
Low Stability * Frailty Index										1.02	0.97, 1.09	0.4

Model 1 is adjusted for function. Model 2 is further adjusted for frailty. Model 3 is further adjusted for age. Model 4 further includes a frailty interaction. Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference.

Comparison of 41, 51, and 56 (Mortality)

Table 125. Comparison of Model Discrimination Using Different Frailty Indexes (Mortality)

	41-Item FI			51-Item FI			56-Item FI		
	AUC	Lower 95%	Upper 95%	AUC	Lower 95%	Upper 95%	AUC	Lower 95%	Upper 95%
Unadjusted Models									
FI Only	**0.778**	0.772	0.784	**0.784**	0.779	0.789	**0.784**	0.779	0.789
FM Only	0.651	0.645	0.657	0.652	0.646	0.658	0.643	0.636	0.649
RoA Only	0.753	0.747	0.758	0.759	0.754	0.764	0.761	0.756	0.766
DIOR Only	0.684	0.678	0.689	0.685	0.679	0.690	0.673	0.667	0.679
Frailty-Adjusted Models									
FI + FM	0.801	0.796	0.806	0.799	0.794	0.804	0.799	0.793	0.804
FI + RoA	**0.802**	0.796	0.807	**0.806**	0.801	0.812	**0.807**	0.802	0.813
FI + DIOR	0.780	0.775	0.786	0.786	0.781	0.792	0.785	0.780	0.791
Age and Sex Adjusted Models									
FI Only	0.824	0.819	0.829	0.828	0.823	0.832	0.828	0.823	0.833
FI + FM	0.827	0.822	0.832	0.829	0.824	0.834	**0.829**	0.824	0.834
FI + FM Interaction	**0.828**	0.823	0.833	0.829	0.825	0.834	**0.829**	0.825	0.834
FI + RoA	0.824	0.819	0.829	0.828	0.823	0.832	0.828	0.823	0.833
FI + RoA Interaction	0.825	0.820	0.829	0.828	0.823	0.833	0.828	0.823	0.833
FI + DIOR	0.827	0.822	0.832	**0.830**	0.825	0.835	**0.829**	0.824	0.834
FI + DIOR Interaction	**0.828**	0.823	0.832	**0.830**	0.825	0.835	**0.829**	0.824	0.834
FI + FM + RoA + DIOR	0.830	0.825	0.834	0.831	0.826	0.836	0.830	0.826	0.835
FI + FM + RoA + DIOR + All FI Interactions	0.831	0.826	0.836	0.832	0.827	0.836	0.831	0.826	0.836

Highest AUC for each category of models is bolded, excluding the combined resilience models (bottom two rows). 95% confidence intervals are calculated using the Delong method.

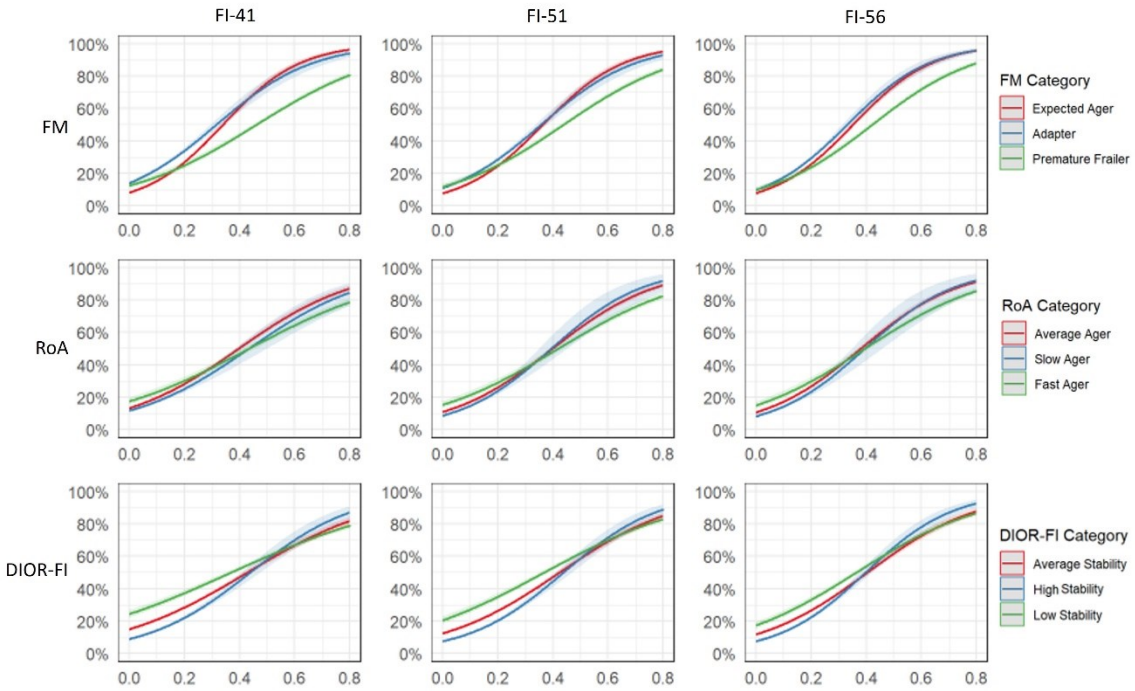


Figure 46. Comparison of Interactions Using Different Frailty Indexes (Mortality)

Visualized interactions correspond to the frailty, age, and sex adjusted models. X-axis is the level of FI. Y-axis is the predicted probability of mortality.

Comparison of 41, 51, and 56 (Recovery)

Table 126. Comparison of Model AUC using Different Frailty Indexes (Recovery)

	41-Item FI			51-Item FI			56-Item FI		
	AUC	Lower 95%	Upper 95%	AUC	Lower 95%	Upper 95%	AUC	Lower 95%	Upper 95%
Unadjusted Models									
FI Only	0.581	0.553	0.610	0.611	0.583	0.639	0.608	0.580	0.636
FM Only	0.556	0.527	0.585	0.558	0.529	0.587	0.549	0.520	0.579
RoA Only	**0.625**	0.597	0.652	**0.628**	0.600	0.655	**0.625**	0.598	0.653
DIOR Only	0.573	0.544	0.601	0.568	0.539	0.597	0.569	0.540	0.597
Frailty-Adjusted Models									
FI + FM	0.618	0.591	0.646	0.634	0.606	0.661	0.628	0.601	0.655
FI + RoA	**0.628**	0.601	0.655	**0.643**	0.616	0.670	**0.640**	0.613	0.667
FI + DIOR	0.604	0.576	0.632	0.625	0.597	0.653	0.616	0.589	0.644
Frailty and Age-Adjusted Models									
FI Only	0.643	0.616	0.670	0.655	0.628	0.682	0.654	0.628	0.681
FI + FM	0.651	0.624	0.678	0.662	0.636	0.689	0.660	0.633	0.687
FI + FM Interaction	**0.672**	0.646	0.698	**0.687**	0.661	0.713	**0.679**	0.653	0.705
FI + RoA	0.647	0.620	0.674	0.659	0.632	0.685	0.656	0.629	0.683
FI + RoA Interaction	0.652	0.625	0.679	0.665	0.638	0.691	0.663	0.636	0.690
FI + DIOR	0.654	0.627	0.680	0.663	0.636	0.689	0.659	0.632	0.685
FI + DIOR Interaction	0.666	0.639	0.693	0.675	0.649	0.702	0.667	0.641	0.694
FI + FM + DIOR	0.660	0.633	0.686	0.669	0.642	0.695	0.664	0.638	0.690

Highest AUC for each category of models is bolded, excluding the combined resilience model (bottom row). 95% confidence intervals are calculated using the Delong method.

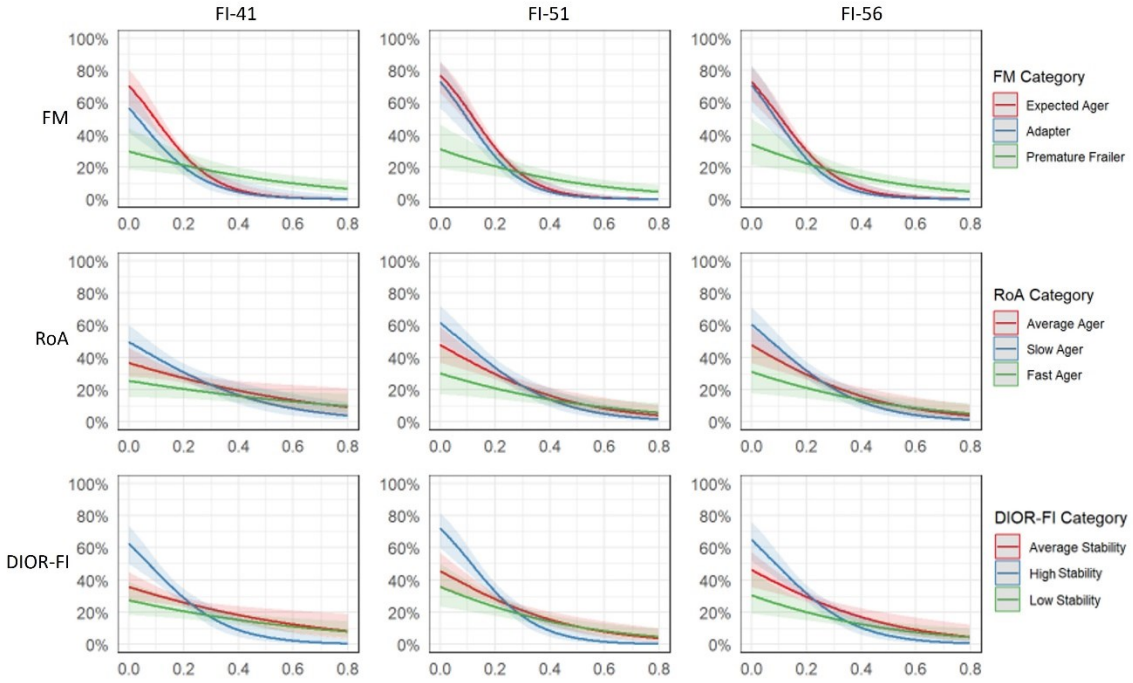


Figure 47. Comparison of Interactions Using Different Frailty Indexes (Recovery)

Visualized interactions correspond to the function, frailty, and age-adjusted models. X-axis is the level of FI. Y-axis is the predicted probability of full recovery.

Appendix M. Household Random Effects Sensitivity Analysis

Run for mortality only (not necessary in recovery sample due to negligible design effect, see section 6.1.2).

FM

Table 127. Logistic Regression Models for FM and Mortality (Household Clustering)

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
FM Category												
Expected Ager	—	—		—	—		—	—		—	—	
Adapter	1.81	1.70, 1.92	**<0.001**	2.40	2.25, 2.57	**<0.001**	1.36	1.26, 1.47	**<0.001**	1.85	1.61, 2.12	**<0.001**
Premature Frailer	4.21	3.96, 4.48	**<0.001**	0.38	0.35, 0.42	**<0.001**	0.63	0.57, 0.70	**<0.001**	1.58	1.30, 1.91	**<0.001**
Frailty Index				1.25	1.24, 1.26	**<0.001**	1.18	1.17, 1.19	**<0.001**	1.24	1.22, 1.26	**<0.001**
Age							1.06	1.06, 1.07	**<0.001**	1.06	1.06, 1.06	**<0.001**
Sex												
Female							—	—		—	—	
Male							1.85	1.74, 1.97	**<0.001**	1.89	1.78, 2.01	**<0.001**
FM Category * Frailty Index												
Adapter * Frailty Index										0.96	0.94, 0.98	**<0.001**
Premature Frailer * Frailty Index										0.92	0.90, 0.93	**<0.001**

Model 1 is unadjusted. Model 2 is adjusted for frailty. Model 3 is further adjusted for age and sex. Model 4 further includes a frailty interaction. Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference.

RoA

Table 128. Logistic Regression Models for RoA and Mortality (Household Clustering)

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
RoA Category												
Average Ager	—	—		—	—		—	—		—	—	
Slow Ager	0.27	0.25, 0.29	**<0.001**	0.41	0.37, 0.44	**<0.001**	0.84	0.76, 0.92	**<0.001**	0.88	0.76, 1.01	0.060
Fast Ager	6.22	5.83, 6.65	**<0.001**	2.99	2.77, 3.22	**<0.001**	0.96	0.87, 1.07	0.5	1.40	1.20, 1.64	**<0.001**
Frailty Index				1.10	1.10, 1.11	**<0.001**	1.14	1.13, 1.14	**<0.001**	1.15	1.14, 1.16	**<0.001**
Age							1.07	1.07, 1.08	**<0.001**	1.07	1.07, 1.08	**<0.001**
Sex												
Female							—	—		—	—	
Male							1.81	1.71, 1.92	**<0.001**	1.82	1.72, 1.94	**<0.001**
RoA Category * Frailty Index												
Slow Ager * Frailty Index										1.01	0.98, 1.03	0.5
Fast Ager * Frailty Index										0.97	0.95, 0.98	**<0.001**

Model 1 is unadjusted. Model 2 is adjusted for frailty. Model 3 is further adjusted for age and sex. Model 4 further includes a frailty interaction Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference.

DIOR-FI

Table 129. Logistic Regression Models for DIOR-FI and Mortality (Household Clustering)

Characteristic	Model 1			Model 2			Model 3			Model 4		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
DIOR-FI Category												
Average Stability	—	—		—	—		—	—		—	—	
High Stability	0.38	0.35, 0.40	**<0.001**	0.58	0.54, 0.62	**<0.001**	0.65	0.60, 0.71	**<0.001**	0.55	0.49, 0.63	**<0.001**
Low Stability	3.15	2.96, 3.34	**<0.001**	1.23	1.15, 1.33	**<0.001**	1.29	1.20, 1.40	**<0.001**	1.83	1.59, 2.11	**<0.001**
Frailty Index				1.15	1.14, 1.16	**<0.001**	1.12	1.11, 1.13	**<0.001**	1.13	1.12, 1.14	**<0.001**
Age							1.07	1.07, 1.08	**<0.001**	1.07	1.07, 1.08	**<0.001**
Sex												
Female							—	—		—	—	
Male							1.81	1.71, 1.92	**<0.001**	1.83	1.73, 1.95	**<0.001**
DIOR-FI Category * Frailty Index												
High Stability * Frailty Index										1.04	1.02, 1.06	**<0.001**
Low Stability * Frailty Index										0.97	0.96, 0.98	**<0.001**

Model 1 is unadjusted. Model 2 is adjusted for frailty. Model 3 is further adjusted for age and sex. Model 4 further includes a frailty interaction. Note: FI odds ratio represents a change of 0.03, the proposed minimal important difference.

AUC

Table 130. Discrimination of Mortality Models (Household Clustering)

Model	AUC	Lower 95%	Upper 95%
Unadjusted Models			
FI Only	**0.778**	0.772	0.784
FM Only	0.650	0.644	0.657
RoA Only	0.751	0.746	0.757
DIOR Only	0.684	0.678	0.690
Frailty-Adjusted Models			
FI + FM	**0.801**	0.796	0.806
FI + RoA	**0.801**	0.796	0.806
FI + DIOR	0.781	0.775	0.786
Age, sex, and frailty-adjusted models			
FI Only	0.824	0.819	0.829
FI + FM	0.827	0.823	0.832
FI + FM Interaction	**0.828**	0.824	0.833
FI + RoA	0.824	0.819	0.829
FI + RoA Interaction	0.824	0.819	0.829
FI + DIOR	0.827	0.822	0.832
FI + DIOR Interaction	**0.828**	0.823	0.832
FI + FM + RoA + DIOR	0.830	0.825	0.834
FI + FM + RoA + DIOR + All FI Interactions	0.831	0.826	0.836

Highest AUC for each category of models is bolded, excluding the combined resilience models (bottom two rows). 95% confidence intervals are calculated using the Delong method.

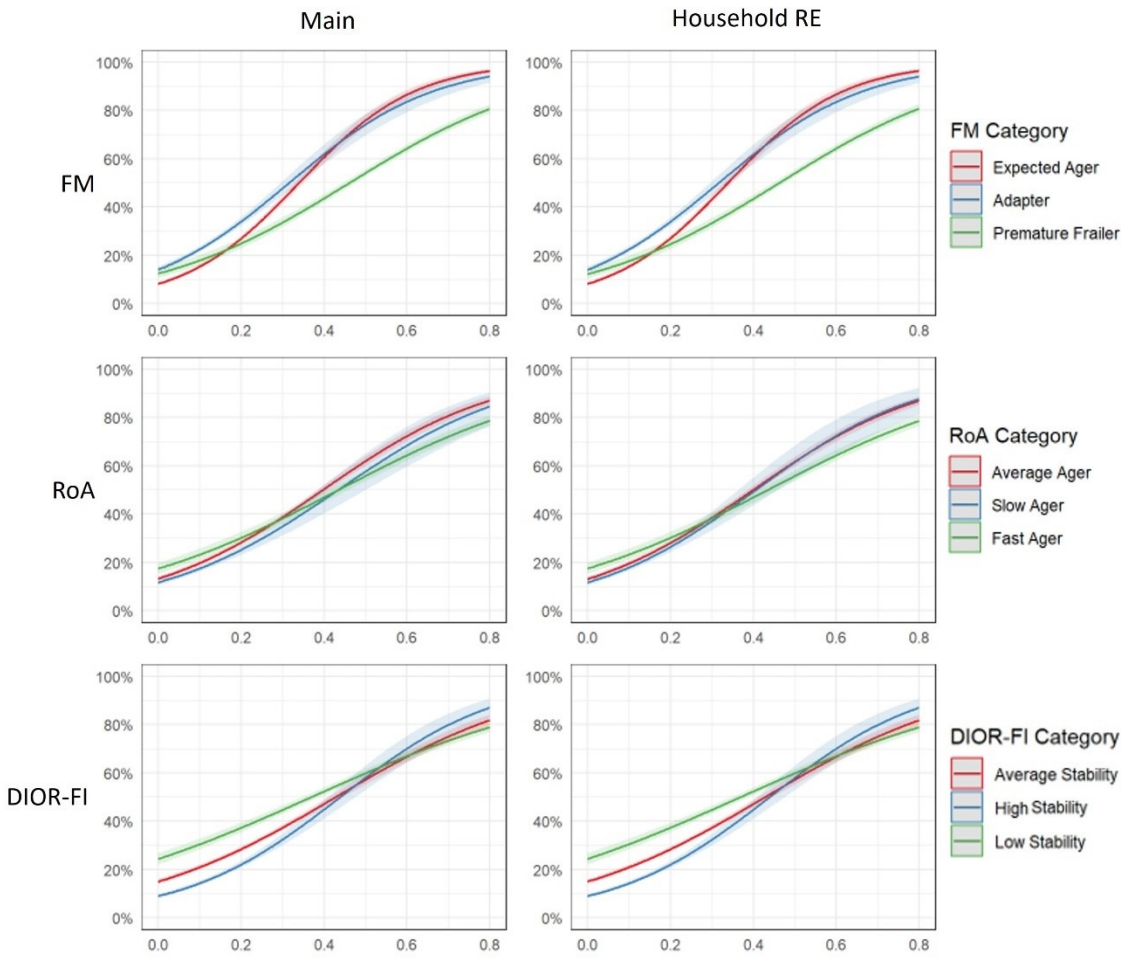


Figure 48. Comparison of Interactions Effects on Mortality (Household Clustering)

Visualized interactions correspond to the frailty, age, and sex adjusted models. X-axis is the level of FI. Y-axis is the predicted probability of mortality.