Tampereen yliopisto

Lassi Virtanen

ESTIMATING BIOLOGICAL AGE USING QUESTIONNAIRE AND INTERVIEW DATA

Lääketieteen ja terveysteknologian tiedekunta
Kandidaatin tutkielma
Syyskuu 2022

ABSTRACT

Lassi Virtanen: Estimating biological age using questionnaire and interview data
Bachelor's thesis
Tampere University
Bioteknologian ja biolääketieteen tekniikan kandidaattiohjelma
9 / 2022

Human population is not only growing but also aging rapidly. Thus there is a growing need for biomedical research to develop new medical interventions. A key element in this development is to gain accurate information regarding a person's healthspan and lifespan. Chronological age is tightly linked to human biological aging, but it is not enough to describe the variation in the phase of aging in individuals with a same age. For example among the elderly, some need assistance for most daily activities while others can live independently. Biological age, distinct information from chronological age would serve as great additional insight to an individual's health status. Currently biological age is mostly measured using biological data obtained from e.g. a blood sample, such as DNA methylation or telomere length. While these measurements provide good data for predicting lifespan and health span, they are time-consuming and expensive leaving a need for a cheaper, faster, and more accessible measurement of biological age. This leads to this literature reviews research question. Could non-invasive, feasible and cost-effective questionnaire-based information on health be an alternative indicator of biological age?
The literature review was conducted using Scopus and PubMed as the primary sources. Search terms included "biological age" "questionnaire" "interview" "biomarker" "Frailty index" "Self-rated health" and "aging".
Several potential questionnaire-based indicators of biological age were identified, with the most prevalent ones being frailty index and self-rated health. When biomarker data-based biological age indicators have been compared to questionnaire-based indicators they seem to perform equally well in mortality prediction, and combination of the two produces slightly better results.

In some cases, questionnaire data-based biological age indicators performed even better than biological data-based estimates in health span and lifespan predictions. The holistic approach in questionnaire-data-based indicators capture aging rate at an organismal level while biomarker-based biological age is often tissue specific. Questionnaire-based biological age is affordable and easy to assess when compared to biomarkers measured using biochemical laboratory methods. This culminates in questionnaire and interview data-based information being a great tool for biological age estimation at population level. However questionnaire-based information isn't always considered as valid source for biological age estimation. This is likely because commonly known definition of biological age is defined from the perspective of cell biology

Keywords: Biological age, frailty index, self-rated health, biomarker of age, indicator of biological age, healthcare, aging

TIIVISTELMÄ

Lassi Virtanen: Estimating biological age using questionnaire and interview data
Kandidaatin tutkielma
Tampereen yliopisto
Bioteknologian ja biolääketieteen tekniikan kandidaattiohjelma
9 / 2022

Maailmaanlaajuisesti väestön määrä ei ainoastaan kasva suuremmaksi, vaan se myös vanhenee. Lääketieteellisellä tutkimuksella on sen vuoksi kasvava tarve kehittää hoitoja ja interventioita, joilla voidaan vastata vanhenemisen tuottamiin terveydellisiin ongelmiin ja haasteisiin sekä mahdollisesti ennaltaehkäistä niitä. Tätä varten tarvitaan tarkkaa tietoa siitä, paljonko ihmisellä on jäljellä elinvuosia ja mikä osa niistä on terveitä. Kronologinen eli kalenteri-ikä tarjoaa jo itsessään tärkeää, mutta lopulta vajavaista tietoa biologisen vanhenemisen vaiheesta. Ihmisten kyky toimia itsenäisesti vaihtelee hyvinkin paljon saman ikäisillä. Kahdesta samanikäisestä henkilöstä toinen voi tarvita apua ympärivuorokautisesti, kun taas toinen voi elää hyvinkin itsenäistä elämää omassa kodissaan. Biologinen ikä, kalenteri-iästä erillinen mittari tarjoaa kaivattua lisätietoa ihmisen terveydentilasta ja biologisen vanhenemisen vaiheesta. Biologisen iän mittaaminen suoritetaan nykyisin pääasiallisesti biomarkkereiden avulla, tästä esimerkkejä ovat telomeerien pituus sekä DNA:n metylaatio. Niiden tuottama data on korkealaatuista, mutta myös kallista ja hidasta niissä käytettävien laboratoriomenetelmien vuoksi. Biologisen iän mittausmetodille, joka on halpa, nopea ja helposti saavutettavissa olisi siis tarvetta. Tästä syntyikin tämän kirjallisuuskatsauksen tutkimuskysymys. Voisiko ei-invasiivinen, helppokäyttöinen ja halvempi kyselypohjainen informaatio terveydestä toimia vaihtoehtoisena biologisen iän indikaattorina?
Kirjallisuuskatsauksen kaksi pääasiallista tietokantaa olivat PubMed ja Scopus. Hakusanoihin kuuluivat “biological age” “questionnaire” “interview” “biomarker” “Frailty index” “Self-rated health” ja “aging”.
On olemassa useampia potentiaalisia kyselypohjaisia biologisen iän indikaattoreita. Kaikkein olennaisimpana pidetään haurausindeksiä (engl. frailty index) ja itsearvioitua terveyttä (engl. self-rated health). Useampi tutkimus, jossa kyselydataa verrattiin biomarkkereihin, päätyivät siihen tulokseen, että niillä on vastaava kyky arvioida biologista ikää ja kuolleisuutta. Toinen yleinen huomio on, että kyselydatan ja biomarkkereiden yhdistäminen tuottaa tarkempia tuloksia.
Kyselydatan holistinen lähestymistapa mitata ikääntymistä organismisella tasolla, biomarkkereiden molekulaarisen ja solutason mittaustavan sijaan ratkaisee tiettyjä ongelmia, kuten eri kudoksien vaihteleva vanhenemisnopeus. Tämän ansiosta kyselydata on tietyissä tilanteissa jopa tuottanut biomarkkereita tarkempia tuloksia. Lisäksi niiden helppokäyttöisyys, tulosten nopea selvittäminen sekä pienet kustannukset tekevät niistä huomattavasti biomarkkeridataa saavutettavampaa. Siksi kysely- ja haastatteludataan pohjautuvien biologisen iän arviointimetodit ovat erinomaisia työkaluja biologisen iän arvioinnissa väestötasolla.

Avainsanat: Biologinen ikä, haurausindeksi, itsearvioitu terveys, iän biomarkkeri, biologisen iän indikaattori, terveydenhuolto, ikääntyminen

FOREWORD

I would like to thank several people who helped me while I was writing this bachelor's thesis.

First and foremost I need to thank my supervisor Laura Kananen, who not only helped me find the topic for my thesis, but also gave me a good direction at the very start so I didn't have to wander around. Her supervision and comments regarding the contents of my writing were instrumental in completing this thesis.

I also want to thank my fellow students Niklas, Katri and Sini who shared this journey of writing a full-sized academic text with me, providing emotional support, and who I could always approach with "how did you do this" type of questions.

Last but not least all my friends from Peräkommari, who would always listen to me rant about whatever troubles I came across and making me feel better with their kind and supportive words.

Tampere, 16.9.2022

Lassi Virtanen

SISÄLLYSLUETTELO

1. INTRODUCTION

Aging is something that is presently inevitable for all humans. We will first grow into maturity, with our bodies finishing their development usually in our early 20's. Beyond this point people will over time, on average, gradually become weaker, albeit very minutely at first, with most of us dying before we reach 100 years of age. And so long as people will age and grow old, there is innate value in understanding the aging process and its components. This understanding can be used also for development of health interventions that could ease the problems that become more common at later stages of life.

Over the years world population has been progressively getting older. According to WHO, over the next 30 years the amount of people who are 60 or older will double from 1 billion in 2020 to 2.1 billion in by 2050 (*Ageing and health*, 2021), while overall population is expected to grow from 7.6 billion to 9.8 billion (Nations); this means percentage of this age group will increase from ~13% to ~21%. The growing amount of older people means there is also a growing need for healthcare to match the needs of this population. The most effective way to match this need would be prevention of problems that typically arise at later stages of life. Thankfully there is room for medical care to assist in the biological aging process; aging is affected by genetics and environmental factors, with genetics accounting for around 10-25% (Kananen and Marttila, 2019). This means that environmental factors, the component that we can actually control to some extent play a part in the aging process.

In order to verify how effective interventions are at controlling the aging process, we need a method for measuring the rate of the aging process. Thankfully such a method exist, several in fact. Biomarker of biological is a way to measure the rate of aging in an individual. Some of the most popular are telomere length and epigenetic clock (Jylhävä, Pedersen and Hägg, 2017), but there are many others with more being constantly developed. However, obtaining biochemical information like the aforementioned biomarkers are still quite expensive and time-consuming due to requirement of specialized laboratory methodology. Additionally they require that a biological sample is taken from the person, typically a blood sample. But could non-invasive, feasible and cost-effective questionnaire-based information on health offer an alternative indicator for biological age? This is a question that I will attempt to answer through this literature review.

2. BIOLOGICAL AGE

2.1 What is biological age?

Biological age is a concept in medicinal research that encompasses the idea that people age at different rates, and thus it would add more insight on top of chronological age in medicine and healthcare(Rockwood and Theou, 2020). The existence of biological age is easy to see in the real world. For example, some 70-year-old individuals need assistance in daily activities, and may live in a nursing home, while others with the same age are able to live alone and drive a car. It is this functional disparity that alone clearly shows the lacking ability of chronological age to accurately describe health and functionality, even if it is still a major risk factor for health decline. But this shouldn't come as a surprise, after all while chronological age only accounts for the amount of time since your birth, biological age is calculated by additionally taking into accounts physiological factors, like cholesterol, capability of physical functioning and diseases.

Contrary to chronological age, biological age is dynamic. Everyone ages chronologically at the same fixed rate, but the rate of biological aging varies from person to person, and over one's lifetime (Ji, Jazwinski and Kim, 2021). Not only is there a genetic component that is responsible for 10-30% of the rate of aging, with the most recent studies leaning more towards 10% or less (Kananen and Marttila, 2019), but also environmental factors such as lifestyle changes (e.g. starting or stopping smoking) can respectively accelerate or decelerate the rate of aging (Pyrkov and Fedichev, 2019). There is also a possibility of a negative and cumulative cycle where biological aging increases the risk of contradicting diseases, and a disease could accelerate aging even further by reducing individuals' functional capabilities. All of this culminates into aging process that exists as a phenomenon at multiple levels: at molecular, cellular, tissue, organ and organismal level (Figure 1).

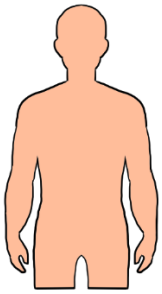
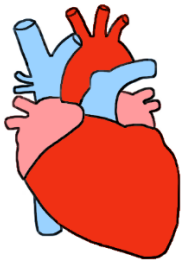
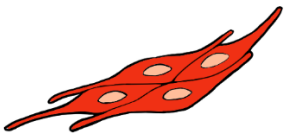
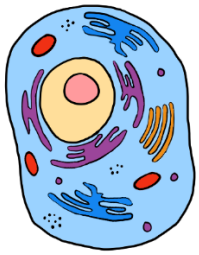
	**Organismal** - Decreased physical functioning **Indicators of biological age** - Frailty - Self-rated health
	**Organ** - Increased risk of heart, liver and kidney problems - Decreased cognitive functions **Indicators of biological age** - Brain-age
	**Tissue** - Accumulation of senescent cells - Decreased regeneration of new tissue **Indicators of biological age** - Blood cell subtype counts - Altered intercellular communication
	**Cellular** - Transcriptomic changes - Decreased mitochondrial efficiency **Indicators of biological age** - Transcriptomics - Proteomics
	**Molecular** - Accumulation of DNA damage - Telomere shortening **Indicators of biological age** - Epigenetic clock - Telomere length

Figure 1: Examples of symptoms of aging at various levels, along with both established and hypothetical indicators of biological age. Illustrations commissioned from Rilla Väre.

2.2 What is biological aging?

The previous chapter discussed the concept of biological age, and how it changes at differing rates in different individuals and over one's lifespan. It is important to understand what constitutes biological aging in order to then start measuring it. Aging is a nigh inevitable (Kananen and Marttila, 2019) and a complex mechanism with multitudes of interacting phenomena contributing to it, but it can be divided into nine major hallmarks that cover many aspects of aging on a molecular and cellular level. These hallmarks are genomic instability, telomere attrition, epigenetic alteration, loss of proteostasis, deregulated nutrient sensing, mitochondrial dysfunction cellular senescence, stem cell exhaustion and altered intercellular communication. (López-Otín *et al.*, 2013)

Genomic instability is one of the more commonly discussed reasons for aging and age-related illnesses. Human DNA undergoes several mutations throughout the day, but organisms have a finely tuned mechanism that detects these mistakes, and then either fixes them or eliminates the potentially hazardous cell trough apoptosis. Not all of these mutations are detected and thus, they accumulate over the years. And it is not just nuclear DNA that accumulates damage, but also mitochondrial DNA (mtDNA) that is especially vulnerable due to oxidative environment because it lacks histones and has less efficient repair mechanisms. Genomic instability can also be directly linked to aging, with premature aging diseases being caused by increased DNA damage accumulation.(López-Otín *et al.*, 2013)

Telomere attrition is a key element of aging and often the first of the many reasons pointed out as lifespan limiting factor. Telomeres are sequences of DNA located at the ends of chromosomes that function mainly as extra pieces to be cut off. During cell division the DNA-replication system is unable to completely recreate the ends of linear DNA, resulting the telomere sequence shortening. While there exists an enzyme called telomerase that would allow complete replication, most mammal cells do not express it. In addition to getting shorter with each cell division, telomeres are also prone to getting damaged. Telomeres are protected by a multi-protein complex known as shelterin, that serves to protect telomeres from being interfered by the DNA repair system that would otherwise fix the loose ends by fusing the ends of chromosomes together. On the flipside, being isolated from the DNA repair system also means that should telomere damage is hard to repair. This results in telomere damage having a high tendency of inducing apoptosis. Shelterin production has been linked to rapid decline in regenerative capacity and accelerated aging, even if the telomeres are of normal length.(López-Otín *et al.*, 2013)

Epigenetic alteration is the modification of DNA without directly changing the base sequence. It encompasses DNA methylation, histone modification and chromatin remodeling which all partake in regulation of transcription. DNA methylation is the basis for epigenetic clock, a popular indicator of biological age that will be described further in a later chapter. Histone modification is a major hallmark of aging in invertebrates, and controlled by *Sir2***,** but it's closest mammalian homolog

SIRT1 hasn't yet been linked to increased longevity, albeit it has been shown to improve health by increased genomic stability and enhanced metabolism. Chromatin is a fiber like structure of DNA and histones that can be remodeled to be either tightly packed or loose to inhibit or promote transcription respectively. As organism ages, the rate of remodeling is diminished resulting in heterochromatin loss and redistribution. (López-Otín *et al.*, 2013)

Proteostasis or protein homeostasis is the regulation of protein concentrations inside a cell. Cells require several proteins to maintain normal functionality, but the amount of these proteins needs to be controlled. A key part in aging are chaperone proteins, proteins that ensure proper folding in newly synthetized proteins that is required for proper functionality. Synthesis of these chaperones is impaired in aged organisms and accelerated aging diseases have been shown to impair proper proteostasis. Overexpression of chaperones in worms and flies has been shown to increase their lifespan, with an especially important family of chaperones being chaperones of heat-shock family. A regulator of heat-shock response HSF-1 has been also shown to possess increased activity after deacetylation by SIRT1.(López-Otín *et al.*, 2013)

Growth hormones naturally play a big part in maturing not only during the juvenile stages of an organism's life, but throughout its whole life. One of these, insulin-like growth factor 1 (IGF-1) is considered as a key factor in aging due to its association with insulin and they share an intracellular signaling pathway. This shared signaling is called insulin and IGF-1 signaling (IIS). IIS is a very conserved pathway, and its reduced function by genetic polymorphisms has been linked to increased longevity. It is also considered to be one of the key reasons why controlled diet increases longevity. (López-Otín *et al.*, 2013)

Mitochondria is the powerhouse of the cell. The ATP it produces is required for most of the chemical reactions that take place in the cell. However as organisms age, the mitochondria are damaged by their local environment. This damage results in increased electron leakage in the respiratory chain, and overall dysfunctionality due to mtDNA damage. This not only results in decreased ATP-production but disrupted apoptotic signaling by increased mitochondria permeabilization in response to stress. Increased production of reactive oxygen species (ROS) due to age associated mitochondrial dysfunction was for a long while believed to mean that ROS causes aging. More recent studies have however shown that ROS does not accelerate aging. Instead ROS is a stress-induced homeostatic response that increases cell proliferation and survival. Since increased age results in more cellular stress and damage, ROS levels also rise. (López-Otín *et al.*, 2013)

Cellular senescence is a product of telomere attrition, but it can also be caused by DNA damage and increased activity of *INK4/ARF* locus. A senescent cell has stopped the cell cycle and will thus no longer undergo cell division. The mechanism of senescence by itself is a beneficial because it helps to keep the organism healthy by arresting cells that are likely damaged and possibly oncogenic. A problem however arises at later ages of an organism's lifespan. In young and healthy

individuals, replacement of senescent cells is simple, resulting in tissue maintaining a healthy state. As an organism ages, its ability to produce new cells to replace the senescent ones becomes impaired, resulting in a more inefficient tissue maintenance. This leads to accumulation of senescent cells that simply aggravates damage on tissue level and contributes to symptoms of aging. (López-Otín *et al.*, 2013)

Exhaustion of stem cells is a big characteristic of aging, since it directly leads to decreased regeneration of tissue. A notable example of this is decline of hematopoiesis, or generation of blood cells. This results in a lack of immune cells at an older age causing the individual to be more susceptible to diseases. (López-Otín *et al.*, 2013)

Aging involves changes in intercellular communications at multiple levels. These changes involve endocrinal, neuroendocrinal and neuronal deregulation. Deregulation leads to increased affinity for inflammatory reactions, but also lower immunosurveillance, meaning pathogens and premalignant cells are more like to go unnoticed. This means that harmful inflammatory reactions become more common along with diseases. This however may be prevented, and even restored by genetic therapy, proper nutrition or with the use of medicine. (López-Otín *et al.*, 2013)

3. BIOMARKERS AS MEASUREMENT OF BIOLOGICAL AGE

3.1 What are biomarkers

As defined by Merriam-Webster dictionary, biomarker is “a distinctive biological or biologically derived indicator (such as a metabolite) of a process, event, or condition (such as aging, disease or oil formation)”. (*Definition of BIOMARKER*, 25.3.2022) In specifically aging research and when talking about biological age, a specific definition for biomarkers of biological age is “biological parameters of an organism that either alone or in some multivariate composite will, in the absence of diseases, better predict functional capability at some late age than will chronological age” (Baker and Sprott, 1988). These two definitions already give a great baseline for understanding what constitutes a biomarker, and what they can be used for. In this thesis terms biomarker and indicator will be used for similar purposes, but biomarker will specifically mean it was sourced from a biological sample such as one obtained from blood, urine, or soft tissue.

There are also two sets of criteria established for a biomarker of, or an indicator of biological age. One is set by American Federation for Aging Research (AFAR) and the other is a joint product

by the International Longevity Center-USA, The Ellison Medical Foundation, Kronos Longevity Research Institute, the Institute for the Study of Aging, and Canyon Ranch Health Resort. The criteria set by AFAR is as follows (Jylhävä, Pedersen and Hägg, 2017):

1. It must predict the rate of aging. In other words, it would tell exactly where a person is in their total life span. It must be a better predictor of life span than chronological age.
2. It must monitor a basic process that underlies the aging process, not the effects of disease.
3. It must be able to be tested repeatedly without harming the person. For example, a blood test or an imaging technique.
4. It must be something that works in humans and in laboratory animals, such as mice. This is so that it can be tested in lab animals before being validated in humans.

The criteria in the joint product has three parameters: (Lohman *et al.*, 2021)

1. The biomarker should predict the outcome of a wide range of age-sensitive tests in multiple physiological and behavioral domains, in an age-coherent way, and do so better than chronological age.
2. It should predict remaining longevity at an age at which 90% of the population is still alive and do so for most of the specific illnesses that afflict the species under study.
3. Its measurement should not alter life expectancy or the outcome of subsequent tests of other age-sensitive tests.

These sets of parameters are quite similar, with the notable difference being that AFAR parameters focus on how old a human is biologically and how long they would have to live if they remain free of disease, while the joint product instead also accommodates for specific diseases and thus predicts longevity more realistically. However both criteria are hard to satisfy, and a biomarker that perfectly meets all of it may not even exist. (Johnson, 2006). But while there may not be a perfect biomarker that fully satisfies these parameters, there are however biomarkers that serve as functional predictors of biological age.

3.2 Biomarkers of biological age

There are six major categories of biomarkers of biological age, epigenetic clock, telomere length, transcriptomics, proteomics, metabolomics, and multi-biomarker (Jylhävä, Pedersen and Hägg, 2017). Even before going further into their properties, many of them can be linked to the nine hallmarks of biological aging based purely on their names. This should already serve as insight as to why these biomarkers of biological age are popular.

Epigenetic clock or DNA methylation age has been a popular research topic during the last decade, with a number of studies showing its capabilities as a biomarker of biological age. Several clocks have been constructed, but two of perhaps the most robust ones are Hannum and Horvath. They

show high correlation with age (r = 0.91 and r = 0.96 respectively) along with each other (r = 0.76) (Chen *et al.*, 2016). In addition to simply measuring biological age, epigenetic clock is a great predictor of all-cause mortality regardless of ethnicity, sex, BMI, smoking and major chronic diseases (Chen *et al.*, 2016)

Epigenetic clocks can be divided into two distinct categories, chronological clocks and biological clocks (Bergsma and Rogaeva, 2020). Hannum and Horvath are examples of excellent chronological clocks which excel in predicting chronological age. Biological clocks instead are designed to be measurements for healthspan, while still associating with chronological age. While chronological clocks are still functional and fit for many purposes, the trend of new epigenetic clock development has shifted more towards the biological clocks thanks to their possibly wider applicability. The functional difference between the clocks is fairly simple. Chronological clocks reflect the DNA methylation changes that predict universally chronological age and are more representative of intrinsic aging processes that aren't directly linked to a disease. Biological DNA methylation clocks however include the DNA methylation changes that predict chronological age but also other health factors, such as diseases and smoking. This shows in Yang's clock accelerating with cancerous lesions. Some of the recent examples of good biological clocks are Zhang's mortality clock and Lu's GrimAge, with Zhang's being a very robust example that outperforms both Hannum and Horvath clocks in all-cause mortality prediction (Bergsma and Rogaeva, 2020)

Telomere length is a benchmark biomarker of biological age as it associates with healthspan and lifespan and is a hallmark of biological aging. It also showcases difference in biological age between sexes, with women having longer telomeres than men. Short telomeres have been shown to correlate with increased mortality risk, and are risk factors for cardiovascular diseases, Alzheimer's and multiple cancers (Jylhävä, Pedersen and Hägg, 2017).

Transcriptomics, the expression of genes is less studied but existing indicator of biological age. A machine learning algorithm was first trained in 7074 blood samples, after which it was used to analyze 1497 transcriptome samples of European ancestry (Peters *et al.*, 2015). The study showed varying, but overall good correlation between chronological age and transcriptomic age. This study also highlights one of the more difficult aspects in aging research. The variance in correlation was easily attributed to different tissues having their own specific gene expression levels. This variance between tissues is an aspect of what makes aging complicated, since it expresses itself on molecular, cellular, tissue, organ, and organismal level.

Studies focusing on proteomics based predictors of biological age have focused mainly on protein glycosylation, and how it's affected by aging (Jylhävä, Pedersen and Hägg, 2017). But while there

have been multiple individual studies, they haven't produced comparable results due to being single cohort based. While the usage of proteomics as predictor of biological age shows promise, further research is necessary for proper validation.

Metabolomics-based predictors are studied relatively little and have mostly been conducted with independent methods and techniques. The results however have been largely promising, with e.g. a study based on proton nuclear magnetic resonance spectroscopy analysis in human urine samples. Total of 59 metabolites were combined into a metabolic age score that was validated in two independent cohorts to associate all-cause survival, clinical outcomes such as kidney problems (Hertel *et al.*, 2016).

In addition to finding entirely new biomarkers of biological age, biological age predictors can also be constructed by combining multiple previously established biomarkers into a new and more accurate predictor. A predictor called "PhenoAge" combining CRP, serum creatinine, glycated hemoglobin, systolic blood pressure, serum albumin, total cholesterol, cytomegalovirus optical density, serum alkaline phosphatase, forced expiratory volume and serum urea nitrogen was constructed in NHANES III study from 2013 by Levine. This predictor was further validated in an independent study that linked higher biological ages with lower IQ-test scores, and worse balance, strength, and motor coordination (Belsky *et al.*, 2015).

4. QUESTIONNAIRE AND INTERVIEW DATA-BASED INDICATORS OF BIOLOGICAL AGE

4.1 Types

Unlike biomarkers of biological age, which can easily be associated with hallmarks of biological aging, questionnaire and interview (Q&I) data-based indicators of biological age are more indicative of the symptoms of aging. A person can't tell by themselves about the state of their telomeres or levels of stem cells, but they are well aware of their symptoms and signs of aging, health-related habits, physical capability, and lifestyle choices. These all either directly affect the physiological status or reflect it. This relationship gives baseline validity for why data obtained purely through questionnaires and interviews without any physical examinations or blood tests could predict a person's biological age. Questionnaires and interviews don't require any invasive procedures passing the requirement #3 for both criteria for indicators of biological age as presented in chapter 3.1. Questionnaires and interviews on health are well available, have been intensively used and studied as well as they are cheap, easy, and relatively fast to conduct.

Frailty index (FI) is one of the most studied Questionnaire and Interview (Q&I) data-based indicator of biological age (Ji et al., 2021; Lohman et al., 2021; Lucicesare et al., 2010; Rockwood & Mitnitski, 2007; Searle et al., 2008). FI is a measurement of accumulation of deficits, which results in frailty (Rockwood and Mitnitski, 2007). As for what constitutes frailty, while it is a fairly complex attribute that manifests on multiple levels through various signs it can also be defined succinctly with “One practicable definition of frailty is age greater than 65 years and dependence on others to perform activities of daily living” (Ji, Jazwinski and Kim, 2021). Usage of FI can start by either using a pre-existing one, or construction of a new one. There are a multitude of different items that can be used in the formation of a new FI making it fairly simple. But some rules should be followed so that the resulting index can produce consistent results. The items must be deficits associated with health status, their prevalence must generally increase with age but shouldn’t saturate too early, the list of items must cover a range of systems and should contain at least 30-40 items to be sufficiently accurate (Searle *et al.*, 2008). However the size of a valid list of items used to construct a FI can vary from 70 to as low as 20 items (Lohman *et al.*, 2021). Despite, or perhaps because of its simplicity frailty indexes abiding by the previous ruleset have a tendency of producing consistent, comparable, and reproducible results. They also show common trends, such as deficit accumulation of 0.03/year and maximal limit of 2/3 deficit (Searle *et al.*, 2008), along with tendency of 0.12 being the cut-off for being well and without disease with index of 0.43 or higher meaning severe frailty and complete dependence on others (Lucicesare *et al.*, 2010). In addition to being a good measurement of functionality, FI has been repeatedly shown to be a good predictor of mortality (Rockwood and Mitnitski, 2007; Searle *et al.*, 2008; Lucicesare *et al.*, 2010; Ji, Jazwinski and Kim, 2021; Lohman *et al.*, 2021), even outperforming those based on biological samples (Li *et al.*, 2020).

Self-rated health (SRH) is another popular Q&I data-based indicator of health, and is frequently used in health and social research (Kananen *et al.*, 2021), and might be considered as indicator of biological age if it meets the criteria for a valid biological age predictor. Whereas FI is based on multiple questions having yes or no answers, SRH is typically a question with multiple answer choices. Other types of SRH-questionnaires also exist. As critique for singular questions, earlier researchers have proposed that a single question based SRH can be a lot vaguer than instruments with many questions (e.g. as in FI), thus possibly relying more on the participants view on what health is. They thought that these cause difficulties in establishing a causal relationship between SRH and mortality, indicating SRH just being more exhaustive in capturing the state of health (Idler and Benyamini, 1997). The biological basis for SRH is still quite poorly understood, but since the 1990s, its robust capability to predict mortality has been repeatedly proven in various societies and populations (Kananen *et al.*, 2021).

Self-rated health deficit index (SRHDI) is a combination of SRH and FI, where a SRH questionnaire with multiple questions are converted into a FI (Lucicesare *et al.*, 2010). This was created in an

attempt to analyze the relationship between FI and SRH since they are both indicators of health and mortality. They used four questions, and three of them had an answer score scale of 0-3 and one had a scale of 0-5, resulting in a total sum of scores on a scale 0-14. This SRH result was then transformed into a SRHDI by dividing the sum of scores by the number, 14. This resulted in an index in a scale of 0 to 1. Interestingly enough, while this index wasn't created according to the rules of FI, having notably fewer items, it resulted in data abiding by general FI key values and correlated moderately with multiple measures of health, including FI. (Lucicesare *et al.*, 2010)

While not considered traditionally as an indicator of biological age, social capital can be used to predict mortality, a trait shared with frailty and SRH (Nieminen *et al.*, 2015). In this context social capital is a resource born trough social interactions and relationships on both individual and societal level. Higher social capital has been reported several times to associate with better health and lower mortality, albeit most of the research hasn't been exhaustive enough to properly accommodate for mediating factors between social capital and mortality. The results are thus promising but not yet conclusive enough, leaving a need for further research (Nieminen *et al.*, 2015). A comparatively more thorough study by Nieminen et al. however did produce results supporting earlier research. In the study they broke down social capital into 3 dimensions, social support, social participation and networks, and trust and reciprocity with a combined total of 29 items on a questionnaire. The questionnaire was conducted along with an interview and health examination to adjust for demographic, behavioral and biological risk factors they made following observations. Lower level of social support had significant mortality correlation among men, but not women. Low social participation had a strong association with mortality among both genders, and trust wasn't found to have notable correlation with mortality.

4.2 Validity

Does Q&I-data based indicators of biological age match the criteria set by AFAR detailed in chapter 3.1? As already discussed earlier, and further examined in this chapter, Q&I-data based indicators do indeed predict the rate of aging, passing the first criterion. The second criterion is further discussed in chapter 4.2, but both SRH and FI function independently of a certain disease. The third criterion is an important factor, as questionnaires and interviews are the least invasive and harmful methods of collecting data from an individual. With the fourth criterion not applying due to requirement of sentience and ability to communicate, we can deem that Q&I-data could be used to construct valid indicators of biological age. The validity is also confirmed by statistics. A statistically significant correlation is required before a claim can be made about the efficacy of an indicator of biological age. These numbers can however feel meaningless without something where to be compared. As such there have also been direct comparisons made between biomarkers and Q&I data-based indicators of biological age. One recent study developed several methods for identifying

biological aging acceleration (Pyrkov and Fedichev, 2019). They found that biomarker-based models did not produce significantly better results compared to Q&I data-based models, and the model combining both source of information produced the best results (Pyrkov and Fedichev, 2019). In their study, Pyrkov and Fedichey developed several methods to measure an individual's biological age that could then be compared to mean biological age of their peers and thus measure their rate of aging, dubbed “biological aging acceleration” (BAA). They ended up developing eight different models of biological age prediction. The results of these models are summarized in table 1. 1) They started out with models with biomarker data only and obtained LIN-bioage. Model was built simply using a linear regression of NHANES-data to find blood features associated with chronological age. 2) Then they also aimed to improve the biological age estimation by using a deep neural network to select the best blood features for DNN-bioage. This approach was able to consider also non-linear associations between blood features and chronological age. DNN-bioage was much more accurate than LIN-bioage in prediction of person’s chronological age but was not more strongly associated with mortality. 3) Next Pyrkov & Fedichev aimed to predict mortality using the existence of at least one age-related disease as a binary covariate in the prediction model. This model was named MORBID-bioage. 4) Paired with MORBID-bioage, using already existing mortality statistics they created another age predictor, MORTAL-bioage. Both MORBID- and MORTAL-bioage performed well in mortality prediction, outperforming LIN-bioage and DNN-bioage. 5) They later aimed to improve MORTAL-bioage by incorporating sex and chronological age as covariates, resulting in MORTAL-bioage with explicit age. This addition resulted in slight increase in statistical power. 6) Next, they developed HAZARDS-survey, a model based purely on questionnaire data, including chronological age and sex as covariates. The HAZARDS-survey had slightly weaker statistical power compared to that of MORTAL-bioage with explicit age. This was true despite comparing a purely questionnaire data-based model (HAZARDS-survey) to that with blood-based biomarkers (MORTAL-bioage with explicit age). 7) Based on the findings so far, blood and questionnaire data-based HAZARDS-survey and MORTAL-bioage with explicit age models were combined into HAZARDS-blood-survey, resulting in a significantly improved biological age prediction. 8) Lastly, they rebuild the age prediction model using a deep neural network and created Deep HAZARDS-blood-survey. Mortality prediction capacity of this model was the best when compared to models 1-7. In summary the study suggests that questionnaire data can be used as a source for biological age indicator, is comparable to blood biomarker-based biological age indicators, and addition of Q&I data to biomarker-based biological age models improves performance of the predictor.

Table 1. *The results of Pyrkov's and Fedichey's models. Current/never smoker refers to difference between smokers and non-smokers, health status refers to difference between people with a chronic disease and those without. Modified from* (Pyrkov and Fedichev, 2019)

	All-cause mortality		**Current/never smoker**		**Health status**	
Model	HR (95% CI)	*p*-value	Δ Bioage (years)	p-value	Δ Bioage (years)	p-value
LIN-bioage	1.04	9.0E-72	2.4	6.4E-07	2.5	1.2E-36
DNN-bioage	1.05	1.1E-27	1.8	1.4E-05	0.3	2.4E-19
MORBID-bioage	1.05	2.7E-72	2.1	1.8E-09	3.0	4.2E-55
MORTAL-bioage	1.08	4.5E-138	2.0	3.7E-10	0.8	2.4E-15
MORTAL-bioage with explicit age	1.06	1.2E-163	2.6	1.1E-13	1.8	6.4E-28
HAZARDS-survey	1.07	7.2E-77	2.6	7.0E-15	1.8	8.8E-32
HAZARDS-blood-survey	1.07	3.9E-194	4.1	1.0E-18	2.4	3.8E-37
Deep HAZARDS-blood-survey	1.10	5.5E-219	4.2	6.7E-19	2.4	2.9E-36

The association between SRH and the levels of several biomarkers in body fluids have also been investigated. A study by Kananen et al. (2021) analyzed a total of 150 different biomarkers, . Of these 57 were associated with SRH. After adjusting for diseases and physical functioning, 26 biomarkers were still associated with SRH. While there have been multiple studies earlier that link biomarker levels with SRH, they have not been quite as exhaustive and have only a few indicators with small sample sizes. The correlation between SRH and biomarker levels shows that there are biological factors explaining the efficacy of SRH as a health indicator, but the causal pathways underlying these associations are still quite poorly understood (Kananen *et al.*, 2021). Kananen et al. analyzed for example in Health 2000, including health information from over 6000 Finns, that even after adjusting for age, gender, diseases, and also for biomarkers that significantly associate with SRH, people with poor SRH were twice as likely to die than people who rated their health as good.

SRHDI, the transformation of SRH into FI developed by Lucicesare et al. also lends credibility to standard FI. Higher SRHDI was associated with higher mortality; 0.01 increase in SRHDI resulted in a 2% increased risk of death, a change comparable to increased mortality associated with growing a year older. But while 2% sounds small, it should be noted that thanks to the way it was constructed, your SRHDI changes at intervals of 0.071, so if you would rate your health as very poor instead of poor, it would directly result in a 14% increased risk of death over the next 50 months (Lucicesare *et al.*, 2010). While SRHDI is constructed using SRH as the base, as shown

by Figure 2, SRHDI behaves like standard FI (albeit having slightly higher indexes).

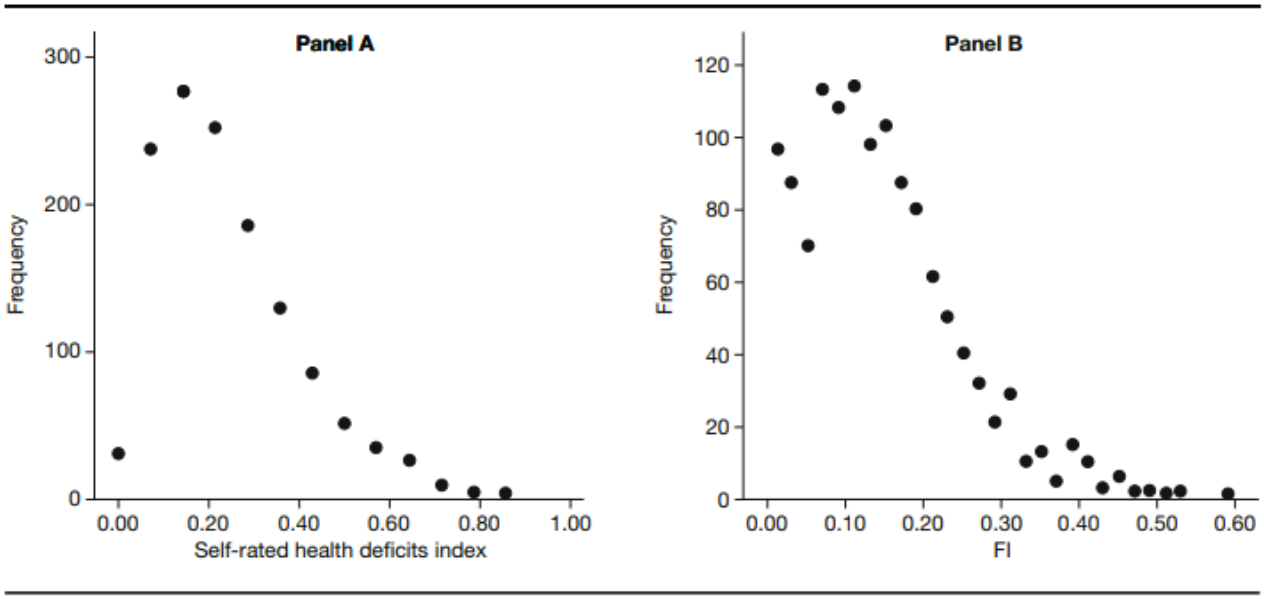


Figure 2. Distribution of SRHDI and FI in the study, showing their comparable shapes. (Lucicesare at al., 2010)

4.3 Mechanisms

Why can data from questionnaires and interview produce seemingly valid data for estimating a person's biological age? There are multiple possible explanations for this. One possible phenomenon behind this is interoception. In short, interoception is a sense of one's body, but it can be defined more in-depth as an "umbrella term that encompasses the afferent (body-to-brain) signaling through distinct neural and humoral (including immune and endocrine) channels; the neural encoding, representation and integration of this information concerning internal bodily state; the influence of such information on other perceptions, cognitions, and behaviors; and the psychological expression of these representations as consciously accessible physical sensations and feelings" (Quadt, Critchley and Garfinkel, 2018). As it has been described earlier, aging encompasses multiple levels ranging from molecular to organismal level. Aging biomarkers typically are only measures at molecular or organ level, which may be the reason why composite markers bring most accurate results due to their wider coverage of aging (Jylhävä, Pedersen and Hägg, 2017). This is why interoception has great potential as an explanatory factor behind Q&I data-based indicators, since it encapsulates and combines information from multiple organs across the whole body, and thus it measures aging at an organismal level. Interoception can also be linked to biomarkers thanks to its tight relation to both homeostatic system and immune response. Additionally interoception also shares a direct molecular link to emotions, through the use of shared substrates (Quadt, Critchley and

Garfinkel, 2018). The link between mental, social, and physical health is well researched and documented and provides a hypothetical explanation at the molecular level why social capital predicts mortality.

Frailty as a concept is pretty self-explanatory as for why FI can work as an indicator of biological age. Aging is the accumulation of molecular damage, which combines into damage at cellular, tissue and organismal level. This damage results in deficits of operational capability. Then by identifying and tallying these deficits they should reasonably present how much an individual has aged biologically.

Same applies for SRH. People are aware of their own health status quite intimately, since even if they might not know their exact cholesterol levels or amount of DNA damage, they experience the tangible effects in their daily life. Additionally, there are some more commonly known biomarkers such as cholesterol of which levels, individuals often know, and this information can be taken negatively/positively into account when rating their own health. People may also take into account healthy and unhealthy lifestyle and health choices when answering SRH questionnaires (Kananen *et al.*, 2021).

5. SUMMARY

Q&I data-based indicators of biological age are studied with an increasing popularity every year, thanks to more and more promising results creating a promise that time spent in study won't be time wasted. Population growing older also creates increasing need to measure aging rates. Studies focusing on biological age indicators add understanding of the very complex aging process. In interventions verifying an individual's health status at baseline is essential and evaluation of the efficacy of medical interventions will benefit from accurate aging rate measures. The number of different biomarkers of biological age is quite vast, each often focusing on different molecular and cellular events, such as telomere length, transcriptomics, or DNA methylation. Q&I data-based indicators of biological age measure aging at an organismal level. While there are multiple potential Q&I data-based indicators of biological age, the most commonly known ones are SRH and FI. When we compare the amount of research done, the amount of information they provide and their statistical significance in predicting lifespan and healthspan, SRH and FI are the two most promising indicators. Especially FI has a lot of variability, thanks to clearly defined rules that still give plenty of leeway in constructing a new one. This leaves a clear route on developing it further. By further refining the items on the questionnaire, I believe, even more accurate indexes can be produced.

The reason why Q&I data-based indicators of biological age produce data that is consistent with biomarkers of biological age is still not well known. There are multiple hypotheses such as interoception as a phenomenon acting as the bridge between molecular and cellular events, and both sub-conscious and conscious understanding an individual has of their own health. People are also aware of the tangible effects of aging that affect their daily life, resulting in an intuitive reasoning where the underlying cellular or molecular changes themselves aren't measured directly, but their effects are observed.

The question that served as the basis for this thesis was "could non-invasive, feasible, and cost-effective questionnaire-based information on health offer an alternative indicator for biological age?". The answer that I arrived to is yes. Q&I data-based indicators of biological age can predict remaining lifespan better than chronological age, and work independently of a certain disease. They are also repeatedly testable causing no harm to the individual. They cannot be tested with animals, but that by itself is a criterion which's inclusion could be reconsidered since it was originally formed from the perspective of purely biomarker-based indicators of biological age. While more research and improvement can be done, the same goes for biomarkers of biological health. Q&I data already produces results of similar performance compared to biomarker data, while the latter requires magnitudes more time and money to obtain. It is also much more accessible for people who might have trouble with travelling from their home to a healthcare center in order to have a blood sample taken, since the questionnaire can be mailed to their home or be filled online. Capturing health at organismal level also helps to bypass the variance in the rate of aging between tissues; tissue-specificity can skew the results when the aim is to evaluate the phase of aging at an organismal level.

6. REFERENCES

Ageing and health (2021). Available at: https://www.who.int/news-room/fact-sheets/detail/ageing-and-health (Accessed: 29 June 2022).

Baker, G.T. and Sprott, R.L. (1988) 'Biomarkers of aging', *Special Issue Biomarkers of Aging*, 23(4), pp. 223–239. Available at: https://doi.org/10.1016/0531-5565(88)90025-3.

Belsky, D.W. *et al.* (2015) 'Quantification of biological aging in young adults', *Proceedings of the National Academy of Sciences of the United States of America*, 112(30), pp. E4104-4110. Available at: https://doi.org/10.1073/pnas.1506264112.

Bergsma, T. and Rogaeva, E. (2020) 'DNA Methylation Clocks and Their Predictive Capacity for Aging Phenotypes and Healthspan', *Neuroscience Insights*, 15, p. 2633105520942221. Available at: https://doi.org/10.1177/2633105520942221.

Chen, B.H. *et al.* (2016) 'DNA methylation-based measures of biological age: meta-analysis predicting time to death', *Aging (Albany NY)*, 8(9), pp. 1844–1859. Available at: https://doi.org/10.18632/aging.101020.

Definition of BIOMARKER (no date). Available at: https://www.merriam-webster.com/dictionary/biomarker (Accessed: 25 March 2022).

Hertel, J. *et al.* (2016) 'Measuring Biological Age via Metabonomics: The Metabolic Age Score', *Journal of Proteome Research*, 15(2), pp. 400–410. Available at: https://doi.org/10.1021/acs.jproteome.5b00561.

Idler, E.L. and Benyamini, Y. (1997) 'Self-rated health and mortality: a review of twenty-seven community studies', *Journal of Health and Social Behavior*, 38(1), pp. 21–37.

Ji, L., Jazwinski, S.M. and Kim, S. (2021) 'Frailty and Biological Age', *Annals of Geriatric Medicine and Research*, 25(3), pp. 141–149. Available at: https://doi.org/10.4235/agmr.21.0080.

Johnson, T.E. (2006) 'Recent results: Biomarkers of aging', *Annual Reviews in Biogerontology*, 41(12), pp. 1243–1246. Available at: https://doi.org/10.1016/j.exger.2006.09.006.

Jylhävä, J., Pedersen, N.L. and Hägg, S. (2017) 'Biological Age Predictors', *EBioMedicine*, 21, pp. 29–36. Available at: https://doi.org/10.1016/j.ebiom.2017.03.046.

Kananen, L. *et al.* (2021) 'Self-rated health in individuals with and without disease is associated with multiple biomarkers representing multiple biological domains', *Scientific Reports*, 11(1), p. 6139. Available at: https://doi.org/10.1038/s41598-021-85668-7.

Kananen, L. and Marttila, S. (2019) *Vanhenemisen biologiset mekanismit ja miten niihin voi vaikuttaa*. Available at: https://www.duodecimlehti.fi/duo14951 (Accessed: 26 February 2022).

Li, X. *et al.* (2020) 'Longitudinal trajectories, correlations and mortality associations of nine biological ages across 20-years follow-up', *eLife*. Edited by D.M. Harper, E. Franco, and A. Moskalev, 9, p. e51507. Available at: https://doi.org/10.7554/eLife.51507.

Lohman, T. *et al.* (2021) 'Predictors of Biological Age: The Implications for Wellness and Aging Research', *Gerontology and Geriatric Medicine*, 7, p. 23337214211046420. Available at: https://doi.org/10.1177/23337214211046419.

López-Otín, C. *et al.* (2013) 'The Hallmarks of Aging', *Cell*, 153(6), pp. 1194–1217. Available at: https://doi.org/10.1016/j.cell.2013.05.039.

Lucicesare, A. *et al.* (2010) 'An index of self-rated health deficits in relation to frailty and adverse outcomes in older adults', *Aging Clinical and Experimental Research*, 22(3), pp. 255–260. Available at: https://doi.org/10.1007/BF03324805.

Nations, U. (no date) *World population projected to reach 9.8 billion in 2050, and 11.2 billion in 2100*, *United Nations*. United Nations. Available at: https://www.un.org/en/desa/world-population-projected-reach-98-billion-2050-and-112-billion-2100 (Accessed: 12 July 2022).

Nieminen, T. *et al.* (2015) 'Social capital and all-cause mortality among Finnish men and women aged 30–79', *European Journal of Public Health*, 25(6), pp. 972–978. Available at: https://doi.org/10.1093/eurpub/ckv058.

Peters, M.J. *et al.* (2015) 'The transcriptional landscape of age in human peripheral blood', *Nature Communications*, 6, p. 8570. Available at: https://doi.org/10.1038/ncomms9570.

Pyrkov, T.V. and Fedichev, P.O. (2019) ‘Biological Age is a Universal Marker of Aging, Stress, and Frailty’, in *Biomarkers of Human Aging*. 1st ed. 2019. Cham: Springer International Publishing (Healthy Ageing and Longevity, 10), pp. 23–36.

Quadt, L., Critchley, H.D. and Garfinkel, S.N. (2018) ‘The neurobiology of interoception in health and disease’, *Annals of the New York Academy of Sciences*, 1428(1), pp. 112–128. Available at: https://doi.org/10.1111/nyas.13915.

Rockwood, K. and Mitnitski, A. (2007) ‘Frailty in Relation to the Accumulation of Deficits’, *The Journals of Gerontology: Series A*, 62(7), pp. 722–727. Available at: https://doi.org/10.1093/gerona/62.7.722.

Rockwood, K. and Theou, O. (2020) ‘Using the Clinical Frailty Scale in Allocating Scarce Health Care Resources’, *Canadian Geriatrics Journal*, 23(3), pp. 210–215. Available at: https://doi.org/10.5770/cgj.23.463.

Searle, S.D. *et al.* (2008) ‘A standard procedure for creating a frailty index’, *BMC Geriatrics*, 8, p. 24. Available at: https://doi.org/10.1186/1471-2318-8-24.