UHASSELT
KNOWLEDGE IN ACTION

Maastricht University

Faculty of Sciences
School for Information Technology
Master of Statistics and Data Science

Master's thesis

Comparing AI predicted PPG age vs comorbidities, mortality and ECG age

Eddy Ceyssens
Thesis presented in fulfillment of the requirements for the degree of Master of Statistics and Data Science, specialization Data Science

SUPERVISOR :
Prof. dr. Inigo BERMEJO DELGADO

SUPERVISOR :
Myrte BARTHELS

2024
2025

Transnational University Limburg is a unique collaboration of two universities in two countries: the University of Hasselt and Maastricht University.

UHASSELT
KNOWLEDGE IN ACTION

www.uhasselt.be
Universiteit Hasselt
Campus Hasselt:
Martelarenlaan 42 | 3500 Hasselt
Campus Diepenbeek:
Agoralaan Gebouw D | 3590 Diepenbeek

Maastricht University

Faculty of Sciences
School for Information Technology
Master of Statistics and Data Science

Master's thesis

Comparing AI predicted PPG age vs comorbidities, mortality and ECG age

Eddy Ceyssens

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics and Data Science, specialization Data Science

SUPERVISOR :
Prof. dr. Inigo BERMEJO DELGADO

SUPERVISOR :
Myrte BARTHELS

Acknowledgements

Firstly, I would like to thank Professor Inigo Bermejo Delgado for his valuable guidance and insights. Many thanks also to Myrte Barthels for her support and commitment as a supervisor.

I would also like to express my sincere appreciation to Ziekenhuis Oost-Limburg for the opportunity to undertake an internship there. It was an educational and enriching experience.

Abstract

Background: Although photoplethysmography (PPG) was not originally developed for this purpose, researchers have recently started analysing PPG signals to detect age-related features. These features include changes in pulse shape, arterial stiffness and other cardiovascular characteristics that change with age. Similar studies have been conducted using ECG signals, revealing relationships between patients' ECG age measurements, comorbidities and mortality. The studies use machine learning and deep learning to estimate a person's age (AI-PPG age) from PPG data, but this is still in development.

Objectives: This study examines the reliability of AI-PPG age as a predictor of chronological age. It also aims to explore whether the deviation between estimated age using AI-PPG and actual chronological age is associated with certain comorbidities and survival time.

Methods: Clinical and demographic patient data was collected from Ziekenhuis Oost-Limburg (ZOL) between 2017 and 2025. PPG age predictions were provided by FibriCheck, the developers of the AI-PPG age model. Pearson's correlation coefficient was used to compare AI-PPG age with chronological age. Chi-squared tests, multivariate regression and ANOVA models were employed to examine the association between AI-PPG age deviation and risk factors such as body mass index (BMI), diabetes, smoking, hypertension, and hypercholesterolemia. Survival analysis employed Kaplan–Meier curves and Cox proportional hazards models to assess the association between AI-PPG age deviation and patient survival time.

Results: The final dataset comprised 181,685 PPGs from 1,414 patients. AI-PPG age showed a moderate correlation ($r = 0.57$) with chronological age. No significant associations were found between AI-PPG age deviation and risk factors such as BMI, smoking, diabetes and hypertension. There was also no significant association detected between survival time and corrected PPG age deviation.

Conclusion: Although PPG is an interesting technology with potential, this study was unable to demonstrate any associations between PPG age abnormalities and risk factors. There were not enough events for certain analyses such as mortality. Further research is needed to refine its use as a biological age estimator.

Keywords: AI-PPG age, biological age, comorbidities, survival analysis, age deviation.

Contents

1 Introduction

Individuals and society both consider health to be an important concern. A person's age is a key indicator of their health status. As people grow older, their risk of developing diseases such as high blood pressure and high cholesterol increases. Consequently, the risk of impaired health also increases with age. However, the rate at which health declines varies from person to person. Genetic factors play a role, as do lifestyle and dietary habits. Life expectancy in Western Europe has increased significantly over the past 50 years. While the average life expectancy for newborns in Europe was just under 62 years in 1950, it is expected to rise to 79 years by 2025 [1]. This increase is largely due to improvements in medicine and health care. This increase appears to be slowing in recent years. According to research, factors such as poor nutrition, obesity, and physical inactivity have a negative effect on life expectancy. Despite the availability of better treatments for high blood pressure and cholesterol, it is difficult to fully offset these health problems. There is still a necessity for enhancement in public health and medicine.

Due to both genetic differences and differences in lifestyle habits, a person's chronological age is not always the best indicator of his or her health status. One also looks at a person's biological age or vascular age. Biological age is a measure of how quickly cells, tissues and organs are aging [2]. It is influenced by factors such as DNA methylation, blood pressure and cholesterol levels. Having a lower biological age is associated with better health and a longer life, whereas having a higher biological age can indicate an increased risk of diseases such as heart disease, diabetes, and dementia [3]. Vascular age is a measure of how old a person's arteries are compared to their chronological age. It reflects the health and stiffness of your blood vessels, which can affect your risk of cardiovascular disease. Vascular age is determined by the pulse wave velocity (PWV), a metric used to assess arterial stiffness [4]. Faster PWV indicates stiffer arteries, which can be linked to aging and cardiovascular risk. Diet, exercise, smoking and blood pressure all influence vascular aging. A lower vascular age suggests healthier arteries, while a higher vascular age may indicate an increased risk of heart disease.

Efforts are constantly being made to improve existing medical technology and develop new applications to examine people's health. For cardiovascular health in particular, one well-known technology to examine the condition of the heart is the electrocardiogram (ECG). It is a medical test that measures and records the heart's electrical activity. During the test, small sensors (electrodes) are placed on the skin of the chest, arms, and legs. These sensors then record the electrical signals that the heart sends out when it contracts and relaxes. Among other things, an ECG can be used to detect heart rhythm abnormalities, heart attacks and other heart problems. The procedure is painless and non-invasive, and is often performed in hospitals or doctors' surgeries.

Recently, a substantial amount of research has been devoted to predicting age from an ECG, which has been suggested as a biomarker for biological age. In younger patients, using AI to predict age from ECGs was found to be a better way of predicting cardiovascular events than using their actual age, according to Hirota et al. But this was not the case for older patients [5]. Chang et al. developed a deep learning model (DLM) to predict the biological age via ECG. They found that ECG-age estimated via DLM provided additional information for cardiovascular disease CVD incidence. A positive deviation from chronological age was correlated correlated not only on mortality but also on other CVDs. [6]. Baek et al. developed a deep learning-based algorithm to estimate AI-ECG heart age, and evaluated its ability to predict mortality and cardiovascular outcomes. They found that biological heart age, as estimated by AI, significantly impacted mortality and major adverse cardiovascular events. AI-based ECG heart age estimation could facilitate prevention and healthcare for cardiovascular outcomes [7].

Another technology for monitoring the condition of the heart is photoplethysmography (PPG). This noninvasive technique uses light to measure changes in blood flow. PPG is widely used in medical applications such as wearable heart rate monitors (e.g. smartwatches) and saturators that measure oxygen saturation in the blood ([8][9]). A light source, typically an LED, illuminates the skin and a detector then measures the amount of transmitted or reflected light. The amount of light reflected or absorbed changes depending on blood flow in the tissues. These changes are recorded by a photodetector. When plotted and analyzed over time, PPG waveforms provide insight into the cardiac cycle. Each heartbeat is reflected by a peak in the PPG waveform. The frequency of these peaks as well as the time interval between them, provide insights into heart rate and heart rhythm regularity, respectively.

The PPG signal can also be used to construct a waveform representing blood pressure. The rising edge of the waveform primarily relates to the systolic phase, while the dicrotic notch marks the end of aortic systole and the start of diastole (see Figure 1). During the diastolic phase, the heart muscle and capillaries relax and the blood volume in the capillary system decreases. This results in a diminished light absorption by the pulsatile component. The falling edge of the PPG waveform, or the catacrotic phase, is associated with the diastolic phase of the cardiac cycle.

The features that are most widely understood are the systolic amplitude (the height of the peak), which resembles a volumetric component linked to arterial blood pressure waves, and the peak-to-peak intervals ($\text{PPI}_{systolic}$) between systolic peaks, which represent the frequency and regularity of heartbeats. The clinical relevance and usability of other waveform-related features in clinical practice is still under investigation.

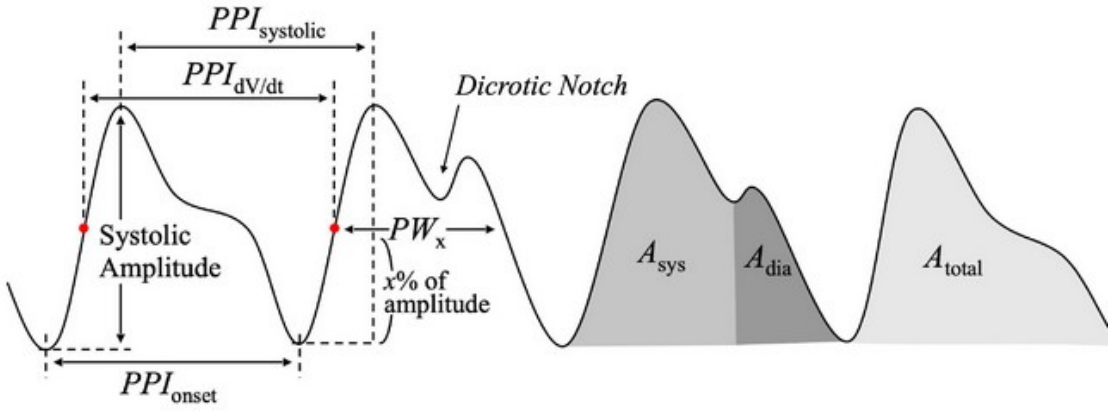

Figure 1: The blood pressure waveform. The dicrotic notch indicates the border between the systolic and the diastolic phase of the heart cycle. As people grow older, the depth of the dicrotic notch changes. Source: FibriCheck academy (https://academy.fibricheck.com/hc/en-be/categories/6240962556060-Educational-content)

There are many age-related parameters that effect the PPG signal. As blood vessels become less flexible with age, blood flow as well as the resulting waveform changes [10]. In young people, blood vessels are more flexible and expand further during the systolic phase. In older people, the blood vessels are less elastic. Therefore, flow rate and blood pressure increase in older people, changing light absorption. In [11], Fine et al. give an overview of age-related parameters and the resulting effect on PPG. For example, the slope of the systolic rising edge, the systolic time and the diastolic peak amplitude all decrease with age. Besides age, there are other factors that can affect the PPG signal. Obesity, for instance, can result in thicker skin and changes in blood flow and saturation.

These changes impact the optical properties of the skin and blood, consequently affecting the PPG signal [12] [13]. In summary, a multitude of factors must be considered. These include, in addition to those listed, respiration, body temperature, ambient light, applied pressure and movement during measurement.

FibriCheck trained a neural network to predict age based on PPG data collected through their app from healthy users. We will refer to the output of this model as AI-PPG age or PPG age throughout this thesis.

It is evident that, while a PPG signal comprises a substantial amount of information, it is also subject to a considerable degree of noise. The utilisation of PPG technology for the estimation of a person's (vascular) age remains in its infancy. The objective of this thesis is to ascertain whether or not the predicted ages resulting from FibriCheck's algorithm are indicative of the subject's actual chronological age. In addition, the objective is to investigate whether the predicted age can serve as a substitute for vascular age.

2 Research questions

Could AI-enhanced PPG age assessment be a useful tool to evaluate a patient's health? The primary objective is to verify the reliability of the model when estimating a patient's chronological age. However, other questions are also relevant, such as whether there is a relationship between PPG age and health status. Can information about health, such as vascular or biological age and risk of death, be inferred based on a PPG prediction of a person's age? The study will also include factors such as BMI, diabetes, hypertension and hypercholesterolemia. To investigate this, the following research questions were formulated:

1. Is PPG age a reliable indicator of chronological age? To answer this question, the differences between PPG predicted age and chronological age are examined in various ways.
2. What is the relationship between a person's PPG predicted age and their probability of death? The present study will examine the existence of a larger positive bias between the patients' PPG predicted age and their chronological age.
3. What is the relationship between the AI-PPG age deviation, this is the difference between the PPG predicted age and the chronological age, and cardiovascular risk factors? What impact do BMI, gender and comorbidities such as diabetes, hypertension, hypercholesterolemia and smoking have on PPG-age? Different subgroups will be compared. Patients with no comorbidity are compared to patients with one or more comorbidities.

3 The data

3.1 Data description

Patient related health data was collected from a hospital environment, specifically obtained from Hospital East-Limburg between September 2017 and April 2025. The PPG age predictions are provided by Fibricheck. They developed and trained a convolutional neural network using the InceptionTime architecture ([14]) to predict chronological age based on a large, real-world, low-frequency, multi-smartphone PPG dataset. Fibricheck used photoplethysmography (PPG) measurements from real-world users aged 18 to 75, collected via their app between 1 January 2023 and 31 March 2024. In order to focus on healthy ageing, individuals with conditions affecting PPG signals, such as diabetes, hypertension and heart disease were excluded, as were users lacking age or gender data, and those with pacemakers. PPG recordings with irregularities or heart rates outside the range of 50–100 bpm were also removed. The dataset was divided into training (70%), validation (10%), and test (20%) sets. Multiple PPGs from each user were included in the training and validation sets, but only the first available PPG was used in the test set, ensuring demographic accuracy in performance evaluation. The estimates from the resulting model will be referred to as AI-PPG age (or PPG age) throughout the thesis.

The available data for this study consists of AI-PPG age estimates from 181,685 PPG measurements from 1414 patients registered through the FibriCheck app on a smartphone. Importantly, this data was not used for the development of the AI-PPG age model. The variability in these measurements is high, both between-patient and within-patient (see Figure 2). The number of measurements per patient varies from only one measurement up to 5696. Most of the patients have between five and 95 PPG readings. There are 118 patients with only one PPG measurement and there are 90 patients who have two measurements. In total 11.1% of the PPG readings comes from only five patients. Patients measure their PPG age during a period ranging from one day to the entire duration of the study, which is more than seven years.

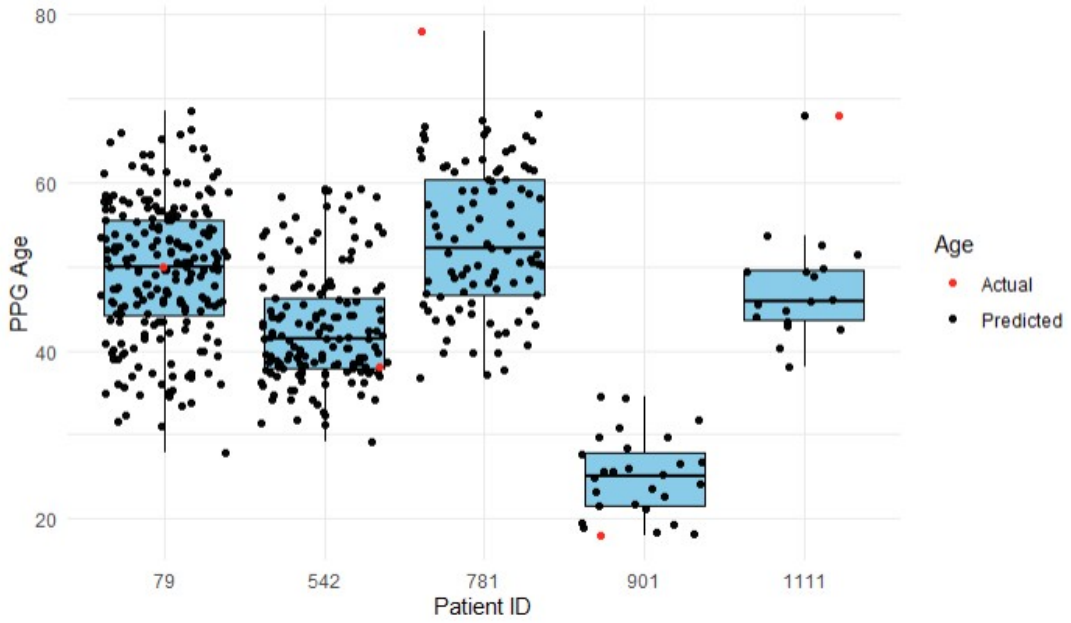

Figure 2: Predicted PPG ages from five patients with multiple PPG measurements. The red dots are their actual ages. They are 50, 38, 78, 18 and 68.

Furthermore, the age of the patient, the time of the measurement, heart rate at the time of the measurement and additional information on heart rate irregularities were recorded, together with the gender and date of birth. 1414 patients, including 615 women and 799 men were included in the study, with ages ranging from 18 years to 94 years at the time of their first PPG measurement. The largest group of patients have ages between 50 and 75 (Table 1). Patients' personal information like date of birth and gender was available in a separate dataset.

Age	**Patients**		**PPG**	
$18-35$	99	(7%)	8437	(4.6%)
$36-45$	112	(8%)	8009	(4.4%)
$46-55$	221	(16%)	26558	(14.6%)
$56-65$	425	(30%)	54104	(29.8%)
$66-75$	433	(31%)	66956	(36.9%)
$76-95$	124	(9%)	17621	(9.7%)
Total	1414		181685	

Table 1: Number of patients and PPG measurements by age group.

In addition to this information, there are five datasets available related to the patient's health. The first one includes information on diabetes, smoking behavior, hypercholesterolemia and hypertension which is extracted from patient's risk profiles (Tables 2 - 5). This dataset contains information on diabetes of 1117 patients, 107 of whom are diabetic. For 1070 patients, information regarding hypertension is available, 458 of which have high blood pressure. For 1072 patients, information with respect to hypercholesterolemia was available and 543 patients have excessive cholesterol. There are 59 smokers out of a total of 1072 patients whose smoking behavior was registered. There are also 290 patients (20.5%) with no known comorbidities.

Diabetes	**Yes**	**No**	**Total**
Female	25	434	459
Male	82	576	658
	107	1010	1117

Table 2: Diabetes by gender.

Smoking	**Yes**	**No**	**Total**
Female	22	417	439
Male	37	596	633
	59	1013	1072

Table 3: Smoking by gender.

Hypertension	**Yes**	**No**	**Total**
Female	171	266	437
Male	287	346	633
	458	612	1070

Table 4: Hypertension by gender.

Hyperchol.	**Yes**	**No**	**Total**
Female	188	243	431
Male	355	274	629
	543	517	1060

Table 5: Hypercholestorolemia by gender.

A second dataset is available containing information regarding the smoking behaviour derived from patient questionnaires. In this dataset, smokers are divided into four categories: never smoked, stopped smoking, passive smoking and active smoking. Table 6 displays the numbers in the different groups by gender. There are 106 patients (7.5%) for whom there is no data.

	Never	Stopped	Passive	Active	NA
Female	347	138	2	80	48
Male	358	297	1	85	58
Total	705	435	3	165	106

Table 6: Numbers of (non)-smoking patients by gender. NA indicates there is no registration.

In the third dataset, information is available on the patients' medication intake (Table 7). These are medications related to the previously mentioned comorbidities: diabetes, hypertension and hypercholesterolemia. There are 281 patients taking diabetes medication, of whom 104 are female and 177 are male. A total of 259 women and 490 men are taking medication for hypercholesterolemia. Medication for hypertension was prescribed for 503 women and 693 men. Of 121 patients (8.6%) no data on medication used were reported.

	Diabetes	Cholesterolemia	Hypertension
Female	104	259	503
Male	177	490	693
Total	281	749	1196

Table 7: Medication intake by gender.

Finally, a dataset containing BMI measurements, taken at regular check up, and a dataset containing mortality information of the patients involved are available. De mortality data was gathered from the Belgium national register. BMI values of 1368 patients can be found. In addition to BMI, patient height and weight were also noted. Between November 2019 and April 2025, 23 patients died: 12 women and 11 men. Their ages at death ranged from 50 years to 87 years.

3.2 Data pre-processing

The datasets with the PPG predictions and patient personal data (gender and date of birth) are complete. However, there is a striking difference in the number of PPG measurements between patients as well as a high variability between within patient measurements. To ensure equal weighting for each patient, we follow two approaches. In the first approach, only the first PPG measurement of each patient is taken into account. That way, every patient is included but almost all PPG measurements are lost. Due to the high variability of predicted ages within patients, this could potentiality lead to unreliable results (Figure 2). This approach will be referred to as First measurement (FIRST). In the second approach, the PPG measurements were grouped by patient and by year. Then, for each patient, the year in which he or she took the most measurements was selected. The PPG predicted age was calculated as the average of all the patient's PPG predicted ages in that year. Finally, patients who had taken at least three measurements and and measurements at least 16.6% of days during their usage time (between the first and last measurement) were selected. These values (3 and 0.166) are equal to the first quartile of the number of measurements and time usage. This approach will be referred to as Active users (ACTIVE).

There are two datasets for smoking behavior. In the comorbidities dataset, patients are categorized as smokers or non-smokers. The separate dataset on smoking behavior divides individuals into four categories: never smoked, quit smoking, passive smoking and active smoking. These four categories are reduced to two: once a person has been registered as 'quit', 'passive smoker' or 'smoker', they are categorized as a 'smoker'. All others are 'non-smokers'. Then the information is combined. If someone is categorized as a 'smoker' in one dataset and a 'non-smoker' in the other, they are regis-

tered as a 'smoker' (Table 8).

Information about other comorbidities (such as diabetes, hypercholesterolemia and hypertension) can be found in the comorbidity dataset and medication records. If a patient is taking medication for a specific comorbidity, it is assumed that he has that condition. Once again, the information is aggregated: if a patient is registered as suffering from a particular comorbidity in at least one of the datasets, then he is registered as suffering from that comorbidity in the final dataset.

	Yes	**No**	**NA**
Diabetes	290 (20.5%)	848 (60.0%)	276 (19.5%)
Smoking	607 (42.9%)	754 (53.3%)	53 (3.7%)
Hypercholesterolemia	842 (59.5%)	848 (23.7%)	237 (16.8%)
Hypertension	1203 (85.1%)	120 (8.5%)	91 (6.4%)

Table 8: Overview of the comorbidities after aggregating all of the available information.

Looking at the BMI data, patients' heights can fluctuate considerably. Therefore, the average height per patient was used to recalculate the BMI. When using only the first PPG measurement for each patient, the BMI measurement closest in time to that measurement was taken. For the Active users data, the overall average BMI of the patient is used.

3.3 Missing data

The data related to the PPG measurements and the patient (sex, date of birth, and mortality) are complete. After merging the data on comorbidities, smoking behavior and medication, missingness was observed in the Active users data: For about 3% of patients there is no information about their smoking behavior, in 6% of patients there is no information about hypertension, in 16% of patients there is no information about hypercholesterolemia, in 19% of patients there is no information about diabetes and in 3% of the patients BMI information is missing. In Figure 3 one can see that when information is missing in a patient, in many cases this spans multiple comorbidities. After consulting with a clinician, no explicit mechanism driving the missingness based on the missing variables themselves could be identified. We will assume missingness at random (MAR) for further analysis.

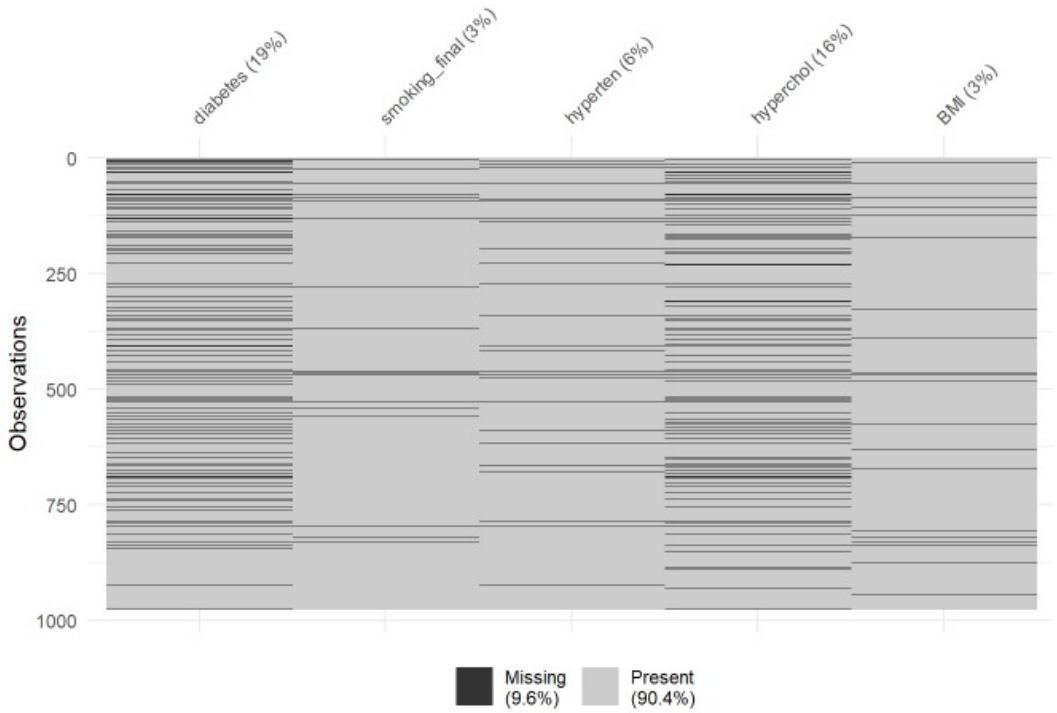

Figure 3: Missing data (9.6%) on comorbidities in the Active users dataset. In many cases it is about multiple comorbidities.

Missingness was also observed in the First measurements data: in 19.5% of the patients there is no information about diabetes, in 16.8% of the patients there is no information about hypercholesterolemia, in 6.4% of the patients, there is no information about hypertension, in 3.8% of the patients there is no information on smoking and in 3.2% of the patients BMI information is missing. Figure 4 gives a graphical presentation of the missingness.

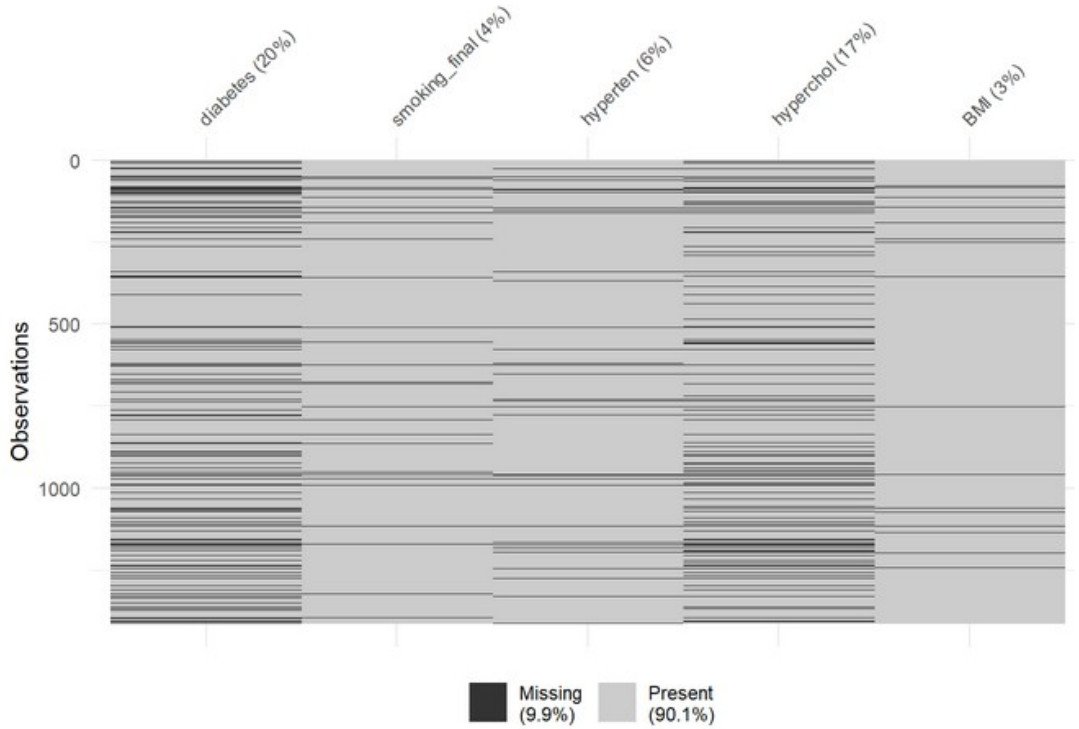

Figure 4: Missing data (9.9%) on comorbidities in the First measurements data set. In many cases it is about multiple comorbidities.

4 Methodology

4.1 Validation of the AI-PPG age

The first research question examines whether PPG age prediction is indicative of chronological age. To this aim, two data sets are employed: one comprising only the first measurement of each patient, and another in which patients are selected based on how frequently they measured their PPG age (Active users data). The quality of the prediction is measured using the following statistics: the Pearson correlation between age and predicted age, mean difference (deviation) between age and predicted age, mean absolute error (MAE), and standard deviation of absolute errors. Finally, linear regression was performed to compare predicted and chronological age.

4.2 AI-PPG Age as Predictor of Overall Mortality

4.2.1 AI-PPG Age correction

Regression towards the mean is a well-known statistical phenomenon, which is also often observed in age estimation models: younger people are estimated to be older, and vice versa [15]. This systematic error occurs because extreme values (very young or very old individuals) tend to move closer to the average when estimated using statistical models. Notably, the risk factors diabetes, hypertension, hypercholesterolemia and BMI are also correlated with age. The likelihood of experiencing one or more of these conditions increases with age. Thus, there are multiple causes of the correlation between age and deviation. Therefore, the deviation must be corrected; otherwise, older patients may be assigned to certain groups for the wrong reasons. For this purpose, a modified version of Beheshti's [16] method is used to correct the predicted PPG age.

At first, there is a regression of the predicted age PA on the chronological age CA:

$$PA = \alpha + \beta \times CA + \varepsilon$$

The deviation DEV is equal to

$$DEV = PA - CA$$

The corrected deviation DEV_c is defined as

$$DEV_c = PA - (\alpha + \beta \times CA)$$

and the corrected PPG age PA_c becomes

$$PA_c = CA + DEV_c = CA + PA - (\alpha + \beta \times CA)$$

But even after this correction, age-related bias may remain. Therefore, Zhang et al. [17] suggest another correction by age group i, a standardization by age. For each age group, the mean PPG age deviation $MPA_c(i)$ and the standard deviation $SDPA_c(i)$ of the deviations are calculated. Then, each PPG age deviation is reduced with the corresponding mean ($MDEV_c(i)$) and this result is divided by the standard deviation ($SDDEV_c(i)$).

$$MDEV_c(i) = \frac{1}{n_i} \sum_{j=1}^{n_i} DEV_c(j)$$

The bias corrected deviation:

$$DEV_{bc} = \frac{DEV_c - MDEV_c(i)}{SDDEV_c(i)}$$

and the bias corrected PPG age:

$$PA_{bc} = CA + DEV_{bc}$$

4.2.2 Kaplan-Meier analysis

How predictive is the AI-PPG age of survival time, in addition to chronological age? Survival time is defined as the difference in years between the first predicted PPG age and the patient's date of death or the patient's last registration (comorbidity, BMI, medication, etc.). When mortality is not registered, the outcome is right-censored, meaning the event did not occur and the actual survival time is unknown.

Kaplan-Meier estimates were used to estimate the survival function for different PPG age deviation groups. There are three groups:

- "Underestimated" age group: the bias corrected deviation is smaller than the negative standard deviation
- "Overestimated" age group: the bias corrected deviation is larger than de standard deviation
- "Correctly estimated" age group: the magnitude of the bias corrected deviation is smaller than de standard deviation

The logrank test is used to compare the different groups. The null hypothesis states that all groups have the same survival functions [18].

4.2.3 Cox Proportional hazards

The Cox proportional hazards model is used to analyse the association of various factors with survival time. This semi parametric model can handle censored data [19]. Covariates can be either categorical or continuous. Two models will be tested.

(1) A model with gender, age and deviation:

$$h_i(t) = h_0(t) \cdot \exp\left(\beta_1.\mathrm{GEN}_i + \beta_2.\mathrm{DEVbc}_i + \beta_3.\mathrm{Age}_i\right)$$

with:

- $h_i(t)$ the hazard function for patient i at time t
- $h_0(t)$ the baseline hazard function
- β_1 the coefficient for the gender indicator (GEN_i)
- β_2 the coefficient for the bias corrected deviation (DEVbc_i)
- β_3 the coefficient for the age (Age_i)

(2) And a model with gender, comorbidities, age and deviation:

$$h_i(t) = h_0(t) \cdot \exp\left(\beta_1.\mathrm{GEN}_i + \beta_2.\mathrm{DIA}_i + \beta_3.\mathrm{SMO}_i + \beta_4.\mathrm{HCL}_i + \beta_5.\mathrm{HTN}_i + \beta_6.\mathrm{DEVbc}_i + \beta_7.\mathrm{Age}_i + \beta_8.\mathrm{BMI}_i\right)$$

with:

- $h_i(t)$ the hazard function for patient i at time t
- $h_0(t)$ the baseline hazard function
- β_1 the coefficient for the gender indicator (GEN_i)

- β_2 the coefficient for the diabetes indicator (DIA_i)
- β_3 the coefficient for the smoking indicator (SMO_i)
- β_4 the coefficient for the hypercholesterolemia indicator (HCL_i)
- β_5 the coefficient for the hypertension indicator (HTN_i)
- β_6 the coefficient for the bias corrected deviation (DEVbc_i)
- β_7 the coefficient for the age (Age_i)
- β_8 the coefficient for the BMI (BMI_i)

The Cox proportional hazard model is used under the assumption that the effect of the covariates is independent of time, which means that the hazard ratio between groups is constant. Schoenfeld Residuals are calculated to test these assumptions [20]. The functional form of age and corrected age deviation will be tested using Martingale residuals [21]. For categorized continuous variables, the log-cumulative hazard is plotted against the log-time.

4.3 Correlation Between Risk Factors and AI-PPG Age Deviation

To examine the correlation between deviation and risk factors, the patients were divided into groups. They were classified according to comorbidities and the number of comorbidities they suffered from. The following statistics were calculated for the whole group and for each subgroup: mean age; mean corrected deviation; mean absolute corrected deviation; standard deviation from absolute deviation; correlation between age and corrected PPG age; and correlation between age and corrected deviation.

Another way to examine relationships is multivariate regression to predict the bias corrected PPG age from the comorbidities, gender and age. The final regression model is:

$$\text{PA}_{bc} = \beta_0 + \beta_1.\text{Age} + \beta_2.\text{BMI} + \beta_3.\text{GEN} + \beta_4.\text{DIA} + \beta_5.\text{SMO} + \beta_6.\text{HTN} + \beta_7.\text{HCL} + \varepsilon$$

ANOVA will also be used to examine whether there are significant differences in the mean corrected deviation between the aforementioned groups. The Tukey test is applied to determine which specific groups within a dataset are significantly different from each other. Levene's test is applied to test the homogeneity of variance. Chi-squared tests will be conducted to determine whether there is a relationship between certain comorbidities and the corrected deviation. For this, the patient's bias corrected PPG deviations DEV_{bc} are divided into three categories based on the standard deviation.

5 Results

5.1 Data pre-processing

The First measurements data only uses the first PPG measurement of every patient and consists of 1,414 patients and 1,414 PPG measurements. The Active users group consists of a selection of 977 patients (428 women and 549 men), combining 110,169 PPG age prediction measurements. Information about the distribution of the comorbidities for both data sets can be found in the appendix (H and I).

5.2 Validation of the AI-PPG age

Table 9 shows the resulting relationship between the PPG predicted age PA and the chronological age. These are similar for the two datasets. There is a moderate positive correlation between age and PPG age. However, the correlation is stronger for the Active users data. The correlation between age and deviation (DEV) is strongly negative: the algorithm overestimates the age of young patients and underestimates the age of older patients. This can also be seen in Figure 5. The transition point occurs at around the age of 41. The mean deviation is negative: the ages of most of the patients are underestimated by the model. This can be attributed to the fact that the majority of patients are over 50 years old.

The mean absolute error (MAE) and associated standard deviation are similar for both data sets. The R^2 values in the linear regression of PPG age against chronological age are low, which is a consequence of the high variability in the data. When regressing chronological age against PPG age, the slope for the Active users data is almost equal to 1; however, for the other data set, this is not the case. This may indicate bias in the model.

When the correlation between age and PPG age of males and females is calculated separately, that of females in both data sets is considerably higher. For the Active users data: 0.66 versus 0.49; and for the First measurements data: 0.54 versus 0.41.

Statistic	**Active users**	**First measurement**
Cor(Age, PA)	0.57	0.47
C.I.	$[0.55; 0.61]$	$[0.42; 0.51]$
Cor(Age, DEV)	-0.78	-0.66
Mean DEV	-11.5	-10.2
MAE (SD)	13.1 (9.31)	13.0 (10.0)
R^2	0.32	0.22
Intercept (SD)	16.0 (2.07)	30.7 (1.47)
Slope (SD)	0.91 (0.042)	0.58 (0.029)

Table 9: Statistics for the validation of the PPG predicted age. The correlation between Age and PPG Age, the 95% confidence interval of this correlation, the correlation between Age and Deviation (DEV), the mean Deviation, the MAE, R^2 from a regression of PPG Age against Age, Slope and Intercept of a regression of Age against PPG Age.

Both plots in Figure 5 show that in younger individuals, the predicted age is systematically larger than the actual age. In persons older than 42, the opposite happens, they are systematically estimated younger. The deviation DEV increases the older the patients are.

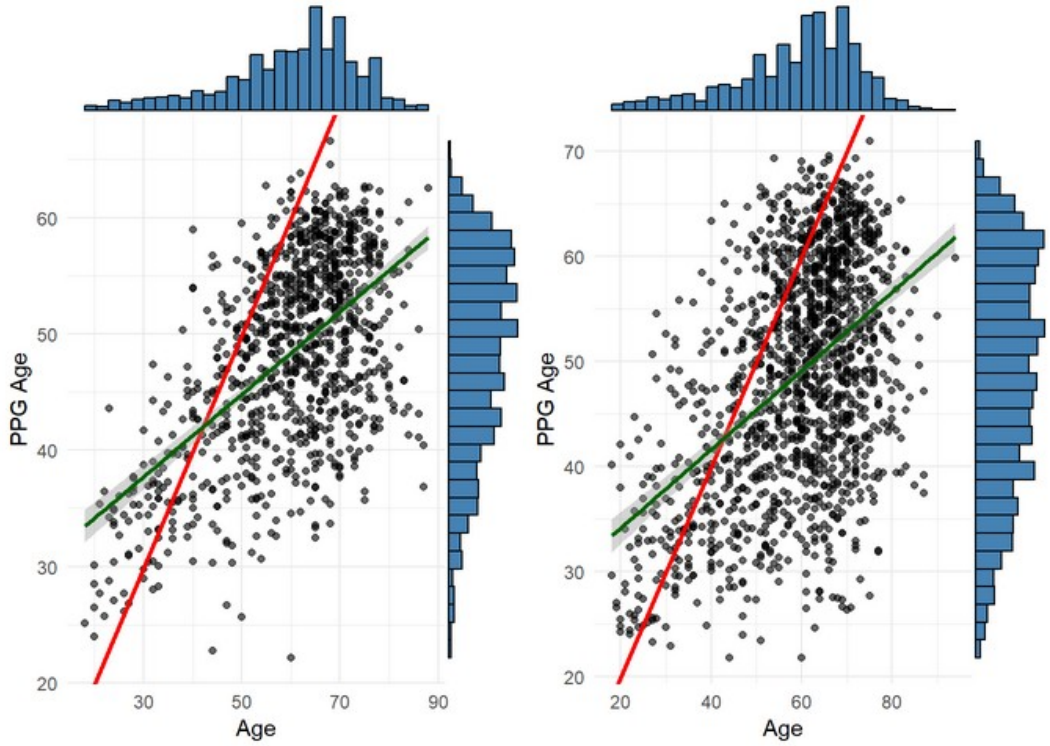

Figure 5: Scatterplots of chronological age vs PPG predicted age. The plot on the left is based on the Active users data, the one on the right on the First measurements data. The red line is the identity line, the green line is the regression line.

Figures 6 and 7 show the residuals versus the fitted values from a linear regression of chronological age against the PPG predicted age. Clearly, there is no linear relationship between chronological age and PPG predicted age.

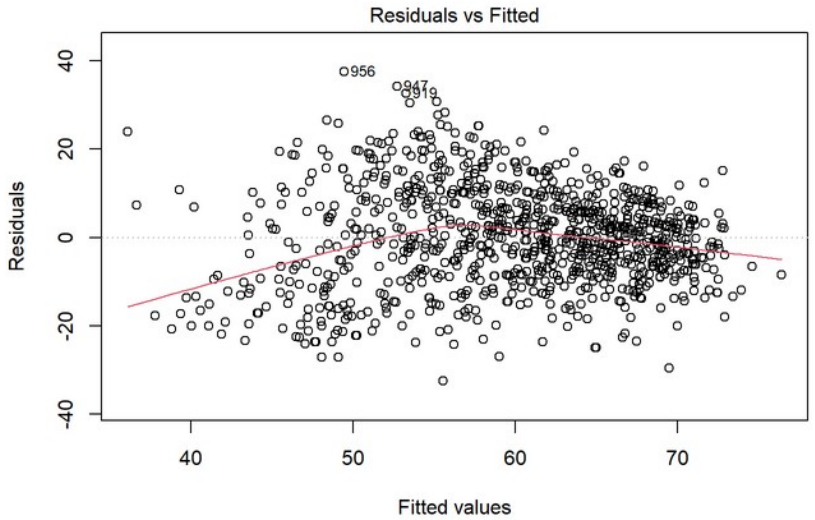

Figure 6: Scatterplot of the residuals against the fitted values after a linear regression of chronological age against PPG predicted age. (Active users data).

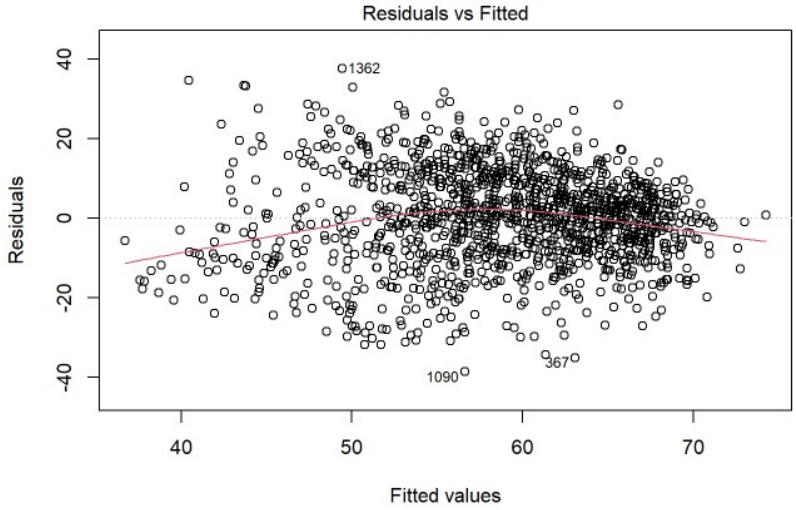

Figure 7: Scatterplot of the residuals against the fitted values after a linear regression of chronological age against PPG predicted age. (First measurements data).

5.3 AI-PPG Age as Predictor of Overall Mortality

5.3.1 Age correction

Active users' data: As a result of the Behesti's correction, there was an increase on the predicted PPG age, an increase on the correlation between the PPG age and the chronological age, a decrease of the MAE, an increase of R^2 and the correlation between the age and the deviation went to 0 (Table 10).

	Before	**Correction**	**Bias correction**
Mean PPG age	48.3	59.9	59.9
Mean Deviation	−11.55	0	0
Corr(Age, PPG Age)	0.57	0.89	1.00
MAE (SD)	13.13 (9.31)	5.73 (4.00)	0.80 (0.53)
$\mathbf{R^2}$	0.3224	0.79	0.99
Corr(Age, Deviation)	−0.782	0	0

Table 10: Active users data: statistics of the selected patients before and after the age corrections.

After Beheshti's correction, there still was age-level bias of age deviation by chronological age (Figure 8). Older and younger patients tend to have a negative age deviation while the group in the middle tends to have a positive deviation. A second correction was performed to standardize all deviations by chronological age.

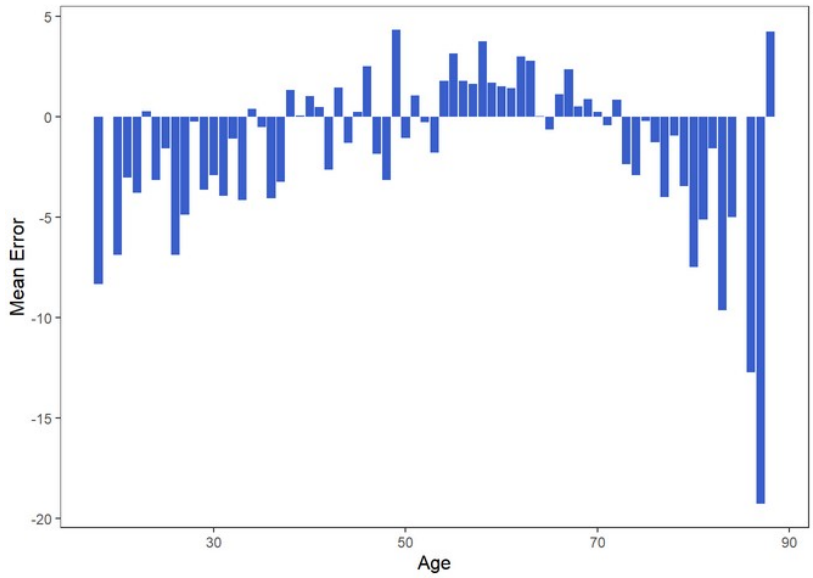

Figure 8: Age-level bias of age deviation by chronological age for the Active users data after the first correction. On the vertical axis is the mean deviation by chronological age.

First measurements data: The same corrections were made on these data, showing identical effects. The results are reported in Table 11. The age-level bias after the first correction can be seen in Appendix A.

	Before	**Correction**	**Bias correction**
Mean PPG age	48.7	59.0	59.0
Mean Deviation	−10.22	0	0
Corr(Age, PPG Age)	0.47	0.82	0.997
MAE (SD)	13.03 (10.01)	8.03 (5.34)	0.82 (0.53)
$\mathbf{R^2}$	0.2172	0.6645	0.995
Corr(Age, Deviation)	−0.66	0	−0.001

Table 11: First measurements data: statistics of the selected patients before and after the age corrections.

5.3.2 Kaplan-Meier analysis

In the Active users data, 19 out of 977 patients selected for the study died during the study period. The Kaplan-Meier method compares the survival probabilities of categorical variables. The patients were divided into three groups based on their corrected deviation: those with a strong negative deviation (Underestimated); those with a deviation around zero (Low); and those with a strong positive deviation (Overestimated). This distinction was made using the standard deviation of the corrected deviation (SD = 0.966), with the boundaries between the groups set at −0.966 and +0.966. Of the patients in the data set based on the first PPG measurement (First measurements data), 23 died. Patients were again divided into three groups based on standard deviation (SD = 0.975). According to the log-rank test, there is no significant difference between the survival curves of the different groups for both data sets ($p = 0.5$ for Active users and $p = 0.4$ for First measurements). The survival curves cross each other which can indicate a problem with the proportional hazards assumption between groups (Figure 9 and Appendix B).

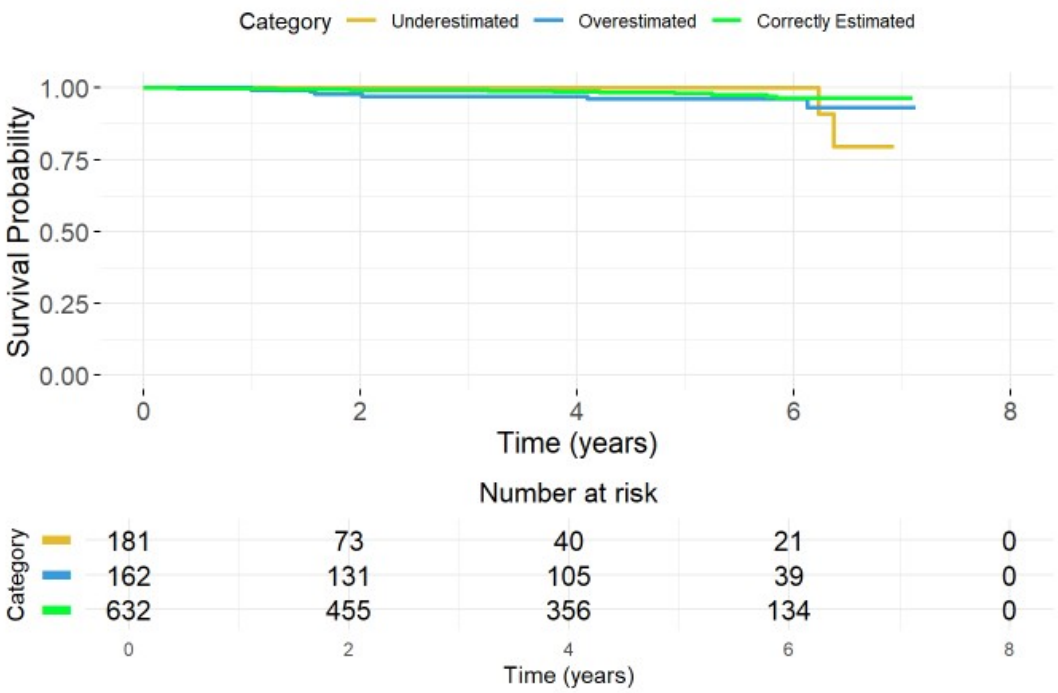

Figure 9: Survival functions of the categorized Deviation groups for the Active users data. On the bottom are the number of patients at risk.

5.3.3 Cox proportional hazards model

The cox model including gender, age and bias corrected deviation tries to model the effect of PPG age deviation and chronological age on the risk of mortality. The results are in Tables 12 and 13. The analyses show that for each unit increase in PPG age deviation, controlling for chronological age and gender, the risk of mortality decreases. None of the results are significant. Both data sets yield similar results. The model assumptions are checked based on Schoenfeld residuals test: for both datasets all the p-values are larger than 0.05, indicating that there are no violations in the model. (Plots are in Appendix J and K.) Remarkable though are the p-values for the age deviation: the are only slightly above 0.05 indicating that the corrected age deviation might depend on the survival time.

Covariate	**HR**	**95% CI**	$P(>\|z\|)$
Gender (M)	0.6969	[0.249; 1.947]	0.49
DEV$_{bc}$	0.8232	[0.467; 1.452]	0.50
Age	1.0840	[1.018; 1.154]	0.01

Table 12: Hazard Ratio (HR) and 95% confidence interval (CI) for the Cox proportional hazard model with the Active users data.

Covariate	**HR**	**95% CI**	$P(>\|z\|)$
Gender (M)	0.7030	[0.283; 1.744]	0.45
DEV$_{bc}$	0.9152	[0.567; 1.477]	0.72
Age	1.0944	[1.036; 1.156]	0.00

Table 13: Hazard Ratio (HR) and 95% confidence interval (CI) for the Cox proportional hazard model with the First measurements data.

The second model adds the risk factors as additional variables. The results are shown in Table 14 (Active users data). The risk of dying increases when one suffers from diabetes, smoking, or hypercholesterolemia. The risk also increases with age but it decreases with BMI and corrected age deviation. Only the effects for age and BMI are significant. The result obtained for the hypertension parameter can be explained by the fact that all patients who died, suffered from the condition. A Schoenfeld residual test was performed. It indicated that the effect of diabetes is not constant over time ($p = 0.012$). After correcting for multiple testing, the models were within the proportional hazards assumption. Schoenfeld residue plots are in the Appendix J and K.

Covariate	**HR**	**95% CI**	$P(>\|z\|)$
Gender (M)	0.836	[0.284; 2.457]	0.744
Diabetes	1.105	[0.354; 3.445]	0.864
Smoking	1.541	[0.510; 4.662]	0.444
Hypercholesterolemia	1.122	[0.236; 5.340]	0.885
Hypertension	$5.46 \cdot 10^7$	[0; inf	0.998
DEV$_{bc}$	0.906	[0.532; 1.542]	0.716
Age	1.070	[1.001; 1.143]	0.0475
BMI	0.854	[0.739; 0.988]	0.033

Table 14: Hazard Ratio's (HR) and 95% confidence intervals (CI) for the Cox proportional hazard model where the comorbidities are included as covariates, data set is the Active users data.

The second model was also applied to the First measurements data (Table 15). Here, only age has a significant effect. Another difference with the Active users data is the effect of hypercholesterolemia: although not significant, it appears to be beneficial. The Schoenfeld test indicated that the effect of at least one covariate changes over time (global p-value = 0.047). According to N there is a positive correlation between survival time and predicted PPG age.

Covariate	**HR**	**95% CI**	$P(>\|z\|)$
Gender (M)	0.725	[0.284; 1.847]	0.450
Diabetes	1.408	[0.531; 3.730]	0.492
Smoking	1.601	[0.601; 4.267]	0.346
Hypercholesterolemia	0.494	[0.168; 1.452]	0.200
Hypertension	$5.48 \cdot 10^7$	[0; inf	0.998
DEV$_{bc}$	0.891	[0.544; 1.460]	0.646
Age	1.096	[1.037; 1.158]	0.001
BMI	0.898	[0.797; 1.012]	0.078

Table 15: Hazard Ratio's (HR) and 95% confidence intervals (CI) for the Cox proportional hazard model where the comorbidities are included as covariates, data set is the First measurements data.

To evaluate the functional form of the covariates for the construction of a Cox proportional hazards model, the martingale residuals are plotted against chronological age and age deviation, as shown in Figure 10. The assumption of linearity between continuous covariates and the log hazard does not appear to cause any problems.

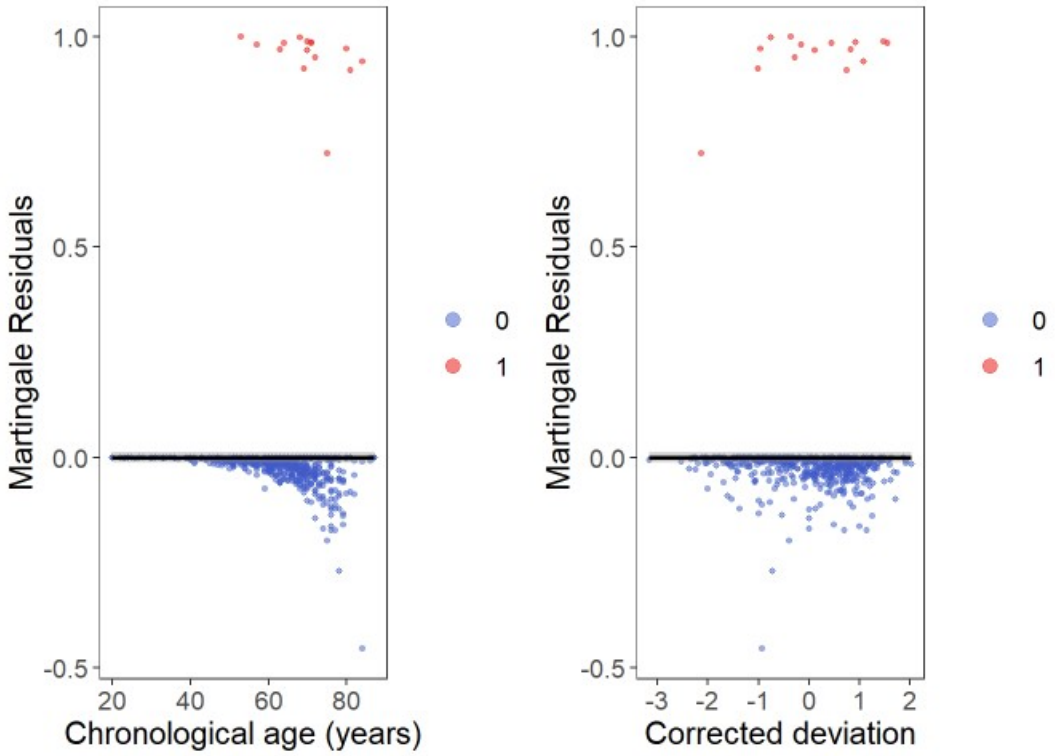

Figure 10: Evaluation of functional form for chronological age and age deviation using martingale residuals for the first Cox regression model (Active users data). Censored = 0, uncensored = 1.

Another way to check the proportional hazards (PH) assumption of the Cox regression model is to look at the complementary log-log plot (Figure 11). The intersection of the curves belonging to the various deviation categories may suggest that the proportional hazards assumption is invalid. This is consistent with the intersecting survival curves shown in Figure 9.

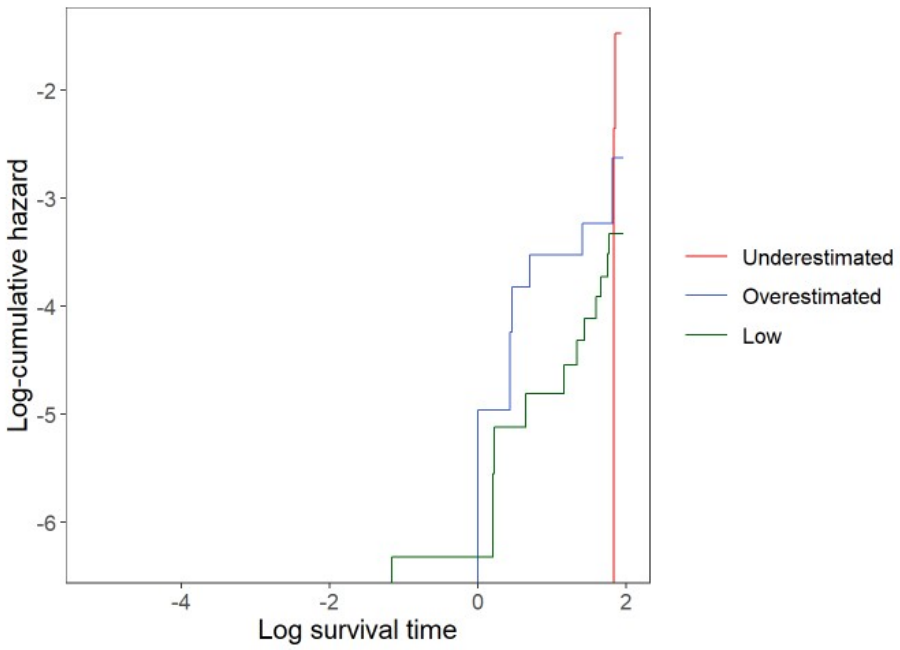

Figure 11: The log-cumulative hazards functions are plotted to assess the proportional hazards assumption. The assumption is supported if the survival functions of the deviation groups are more or less parallel which is not the case (Active users data).

The martingale residual plot and the log-cumulative hazards functions for the First measurements dataset are in the appendix P and Q. They lead to conclusions similar to those of Active users.

5.4 Correlation Between Risk Factors and AI-PPG Age Deviation

The results of the assessment of the effect of the number of comorbidities (0 - 4) on the corrected deviations for the Active users data are in Table 16. The mean age of patients increases with the number of comorbidities. The mean corrected deviation is always close to 0 (never significantly different from 0) and there is no clear increase over the number of comorbidities. The mean absolute error (MAE) is more or less constant over the different groups as is the standards deviation of the corrected deviations (SD(AE)). The correlation between age and corrected predicted age, R^2 and the slope of the regression line are in all cases almost equal to 1 due to the correction of the ages. The same statistics were calculated for the First measurements data. The results were similar and are in the Appendix (Table C).

The correlation between the individual comorbidities and the corrected deviations was also examined. The results for the Active users data are in table 17. For most comorbidities the number of patients involved is larger than in the previous analysis but the outcomes are comparable. Again the same statistics were calculated for the First measurements data (Appendix Table D). Neither the number of comorbidities nor a specific comorbidity leads to a different mean deviation or a different absolute deviation.

	All	0	1	2	3	4
n	977	26	130	262	225	114
mean Age	59.9	43.0	55.2	62.1	65.2	66.6
mean DEV$_{bc}$	0	0.04	-0.08	-0.12	0.06	0.15
MAE	0.80	0.74	0.78	0.79	0.82	0.84
sd(AE)	0.54	0.50	0.55	0.57	0.53	0.49
corr.(1)	1.00	1.00	1.00	1.00	0.99	0.99
corr.(2)	0	0.24	0.04	0.08	-0.06	-0.04
Intercept	0.30	0.71	0.51	0.97	0.30	0.41
slope	0.99	0.98	0.99	0.99	1.00	0.99
R^2	1.00	1.00	1.00	0.99	0.99	0.99

Table 16: Active users data. When patients have 0 up to 4 comorbidities. The first column gives results when all off the patients were considered. n is the number patients involved, corr.(1) is the correlation between the age and the corrected predicted age; corr.(2) is the correlation between the age and the corrected deviation. Intercept, Slope and R^2 result from a regression of age against predicted age.

	All	No	DIA	SMO	HTN	HCL
n	977	26	212	422	845	596
mean Age	59.9	43.0	65.2	62.3	61.4	64.5
mean DEV$_{bc}$	0	0.04	0.03	0.09	-0.02	0.04
MAE	0.80	0.74	0.84	0.79	0.81	0.82
sd(AE)	0.54	0.50	0.50	0.52	0.54	0.55
corr.(1)	1.00	1.00	1.00	1.00	1.00	1.00
corr.(2)	0	0.24	-0.04	-0.02	0.01	0.00
Intercept	0.30	0.71	0.28	0.20	0.44	0.61
slope	0.99	0.98	1.00	1.00	0.99	0.99
R^2	1.00	1.00	0.99	0.99	0.99	0.99

Table 17: Active users data. When patients have at least one comorbidity. The first column gives results when all off the patients were considered. n is the number patients involved, corr.(1) is the correlation between the age and the corrected predicted age; corr.(2) is the correlation between the age and the corrected deviation. Intercept, Slope and R^2 result from a regression of age against predicted age.

In the multivariate regression analysis of the Active users data, only smoking has a significant effect on corrected PPG age (Table 18). The algorithm removed 236 observations due to missing data. When running a multivariate analysis on the First measurements data, however, smoking and hypercholesterolemia do have a significant effect on the corrected PPG age. They increase the corrected PPG age of the patient. Patient gender is also significant in this dataset (Table 19). Here, 347 observations were removed due to missingness in the data.

Covariate	β_i	SE	Pr($>\|t\|$)
Intercept	0.12	0.291	0.684
Age	1.00	0.003	$< 2 \cdot 10^{E-16}$
BMI	−0.01	0.008	0.510
Gender (M)	−0.21	0.073	0.004
Diabetes	0.05	0.084	0.517
Smoking	0.16	0.073	0.030
Hyperchol.	0.15	0.086	0.077
Hypertension	−0.10	0.127	0.430

Table 18: Parameter estimates of a multivariate model with comorbidities, age and gender. Data set is the Active users data. None of the comorbidities appear to have a significant effect on the PPG age.

Covariate	β_i	SE	Pr($>\|t\|$)
Intercept	0.18	0.235	0.454
Age	1.00	0.003	$< 2 \cdot 10^{-16}$
BMI	0.00	0.006	0.966
Gender (M)	−0.32	0.062	$3. \cdot 10^{-7}$
Diabetes	−0.02	0.072	0.762
Smoking	0.15	0.062	0.016
Hyperchol.	0.22	0.073	0.002
Hypertension	−0.08	0.103	0.430

Table 19: Parameter estimates of a multivariate model with comorbidities, age and gender. Data set is the First measurements data. Patients who smoke or suffer from hypercholesterolemia, have an increased PPG age.

ANOVA tests were performed to asses significant differences in the mean corrected deviation between multiple groups. They were performed for patient gender and comorbidities, as well as for the number of comorbidities. For the Active users data this resulted in significant F-values for smoking ($\Pr(> F) = 0.011$) and hypercholesterolemia ($\Pr(> F) = 0.041$). The differences in mean deviation were 0.161 for smoking and 0.156 for hypercholesterolemia. The 95% confidence intervals are $[0.036; 0.285]$ and $[0.006; 0.307]$ respectively. Patients who smoke or suffer from hypercholesterolemia, have on average a higher positive deviation between age and predicted PPG age. Levene's tests were conducted to check the homogeneity of variance. No problems were detected.

ANOVA on the First measurements data gave similar results. There were significant F-values for smoking ($\Pr(> F) = 0.013$) and hypercholesterolemia ($\Pr(> F) = 0.018$). The differences in mean deviation were 0.133 for smoking and 0.154 for hypercholesterolemia. The 95% confidence intervals are $[0.028; 0.239]$ and $[0.027; 0.281]$ respectively. Similarly, patients who smoke or suffer from hypercholesterolemia, have on average a higher positive deviation between age and predicted PPG age. Levene's tests were conducted to check the homogeneity of variance: no problems were detected.

Chi-Squared tests of independence were conducted to determine whether there is a statistically significant relationship between a categorical version of the deviation and whether or not one has a specific comorbidity. For BMI, a one way anova was performed. After Bonferroni correction for multiple testing, there were no significant results for both the Active users data (Table 20) as the First measurements data (Appendix E). The PPG deviation category a patient ends up in is independent of the comorbidities investigated.

	Underestimated (< -0.966)	**Low**	**Overestimated** (> 0.966)	Test statistic	p-value
Gender (M)	66.3%	54.7%	56.2%	$\chi^2 = 10.09$	0.006
(n)	120	346	82		
Diabetes	28.3%	24.7%	34.4%	$\chi^2 = 5.02$	0.081
(n)	43	125	44		
Smoking	39.5%	45.5%	47.8%	$\chi^2 = 2.66$	0.264
(n)	70	276	76		
Hyperchol.	72.1%	71.0%	81.6%	$\chi^2 = 6.29$	0.043
(n)	111	374	111		
Hypertension	94.3%	90.9%	93.5%	$\chi^2 = 2.71$	0.258
(n)	165	537	143		
Mean BMI	27.8	27.5	27.1	$F = 0.828$	0.44
(SD)	5.04	4.59	4.74		

Table 20: Presence of risk factors for PPG age deviation groups in the Active users data. P-values are not corrected; n refers to the total number of patients in a group that suffers from the comorbidity.

6 Discussion & Conclusion

6.1 Interpretation of the results

This study set out to assess the accuracy of an AI-driven PPG age prediction model and to examine the significance of AI-PPG age deviation—the difference between a person's chronological age and the age estimated by the model. If AI-PPG age proves to be related to an individual's health status independently of their actual age, it could serve as a valid biomarker of biological age. This, in turn, would make AI-PPG age a potential indicator of cardiovascular health.

To explore this, the study conducted multiple analyses. First, the consistency and reliability of the neural network's predictions were evaluated. Second, Kaplan-Meier estimates were used to estimate the survival function for different PPG age deviation groups. The impact of AI-PPG age deviation on survival time was examined through a Cox proportional hazards model, both with and without additional explanatory variables. Finally, the relationship between AI-PPG age deviation groups and various health risk factors—such as smoking habits, body mass index (BMI), and existing comorbidities—was analyzed to identify patient profiles more likely to have their age over- or underestimated by the model.

6.1.1 External Validation of AI-PPG Age

The analyses show that there is a moderate correlation between chronological age and predicted PPG age (ranging from $r = 0.47$ to $r = 0.57$) but the relation between chronological age and predicted age is not linear. PPG predictions generally increase with patient age, but there is substantial variability. The mean PPG deviation is negative, indicating that patients are usually estimated to be younger than they are. The mean absolute error (MAE) of 13 years is high compared to other AI models ([8] [22]). Older individuals are underestimated in age, whereas younger individuals are overestimated. This phenomenon has also been observed in other models ([22] [23] [24]) and in other research domains ([17] [25] [26]). This bias complicates the use of prediction and should be considered when using PPG predictions in medical contexts.
When the correlation (between age and PPG age) and the MAE are calculated separately for men and women, the MAE is greater for men (12 versus 14), while the correlation is clearly greater for women (see appendix Tables F and G). The neural network is more accurate in women.

6.1.2 AI-PPG Age as Predictor of Overall Mortality

Following the first correction, the mean PPG age becomes roughly equal to the patients' mean age. The correlation between age and PPG age increases. The second correction centers the deviation by age, resulting in a strong linear relationship between age and the bias-corrected age. The bias-corrected PPG age is now the sum of chronological age and a standardised correction. The advantage is that the relationship between age and deviation disappears. One disadvantage is that if there is only one patient of a certain age, standardisation is not possible and the patient is excluded from the dataset.

Dividing patients into groups based on their PPG age deviation yielded no results. There were no significant differences in mortality between the groups. Cox regression analysis showed that the probability of death decreases with a larger positive deviation from PPG age, and increases with a more negative deviation. Once more, this outcome was not significant. The small number of patients and limited number of deaths make it difficult to obtain meaningful results.

6.1.3 Correlation Between Risk Comorbidities and AI-PPG Age Deviation

There were no significant differences in the mean deviation and MAE values across the different patient groups. In the multivariate regression analysis, there was also no evidence that specific comorbidities give rise to a different PPG age. ANOVA testing revealed that only smoking and hypercholesterolemia were significant comorbidities. Patients with any of these comorbidities have, on average, a larger PPG deviation. Also after categorising the PPG deviation, no clear differences between patients emerged. In general, it can be concluded that no evidence was found to suggest that PPG age or deviation correlated with specific comorbidities.

6.2 Limitations of Methods Used

This study has several limitations. Firstly, the data set is not an independent, random sample. The different age groups are also not equally represented in the data, which mostly consists of individuals over the age of 50. The training data for the model comprised people without comorbidities, whereas the data for this study originated from a hospital setting. Therefore, the model may not have been able to process the data adequately.

The combination of a small sample size, examination of five comorbidities, and missing data may have caused the study to lack statistical power. As no significant results were found, the impact of the missing data on the results was not examined.

The PPG measurements were not carried out in a controlled environment. The patients performed them themselves, so there is a lot of variability in the number of measurements taken and the times at which they were taken. Furthermore, several important factors were not controlled: interfering ambient light, patient movement during measurements and smartphone type.

Another observation is that the per-patient information available was not collected at the same time. This data is a combination of information recorded at different time points. Furthermore, the information is not always complete, accurate or up to date. This introduces noise into the data, making it more difficult to identify correlations. Assuming that all the measurements are labeled with the disease if the patient has a registration of the disease introduces some forecasting, as some measurements might have been taken before disease onset. While this approach may misrepresent temporal relationships, it was adopted due to the scale of the data and the limited resources available for manual annotation.

The time shifts in the anonymisation process may introduce bias into the results of certain analyses. This particularly applies to analyses where time plays a role, such as survival analysis. Bias may also arise because the neural network training data does not accurately reflect the population using the app. There was no time to figure out why the PPG predicted ages decrease over time.

6.3 Ethical Thinking, Societal Relevance, and Stakeholder Awareness

This project uses data collected from electronic patient files at the hospital Ziekenhuis Oost-Limburg. As this was a retrospective study, it was not necessary to obtain explicit informed consent from the patients. In order to guarantee and to protect data privacy only essential information is shared. Names or any other information that could directly identify patients is not accessible. The data can be considered pseudo-anonymous since a conversion key is used in conjunction with the patient identification numbers. Each patient record is also moved to a slightly different time, according to a specific random time range that has been assigned to each individual. This anonymisation process is applied to all patient records. A Good Clinical Practice (GCP) certificate was obtained by the

student before any dataset could be accessed.[1]

AI-PPG age prediction models can help to improve healthcare, particularly in the prevention and treatment of cardiovascular disease. By analysing AI-PPG age deviations, doctors could detect early warning signs of health problems in patients who appear healthy. Since PPG measurements are non-invasive and straightforward to perform, these tools could be adopted widely. Furthermore, AI-PPG age prediction could enable more personalised treatments, reducing unnecessary interventions and improving efficiency. However, successful implementation requires scientific validation, bias reduction and standardised clinical parameters to ensure consistent outcomes across hospitals.

The main stakeholders in the study are Ziekenhuis Oost-Limburg hospital and FibriCheck. The results of the AI-PPG age prediction analysis could be used to improve the algorithm and its implementation in a hospital setting, which could potentially benefit patient care.

6.4 Ideas for Future Research

The high variability of PPG measurements within patients was remarkable. It might be useful to address this by developing models that take certain comorbidities into account. Alternatively, models could be developed that adapt to the user. One option would be to create models that require the user to take an initial measurement for calibration.

Another approach would be to work towards more uniformity in the data by ensuring that a sufficient number of PPG measurements are taken within a given time frame per patient, and that any comorbidities are recorded alongside the PPG measurement (potentially via an automated process). For example, the app could request the height and weight of a person at the time of measurement.

Examining the cause of the correlation between survival time and deviations from the predicted PPG age is also useful.

6.5 Conclusion

This study investigated the reliability and usefulness of AI-PPG age estimates in a medical context. However, our results show that AI-PPG age predictions are highly variable and unreliable. This, coupled with the limited amount of available data, leads us to conclude that we cannot currently make any statements about the usefulness of AI-PPG age predictions. While this technology certainly has potential, the accuracy of the model needs to improve.

[1] https://globalhealthtrainingcentre.tghn.org/ich-good-clinical-practice/

References

[1] Steel, Nicholas et al. Changing life expectancy in European countries 1990–2021: a subanalysis of causes and risk factors from the Global Burden of Disease Study 2021, *The Lancet Public Health*, Volume 10, Issue 3, e172 - e188 (2025)

[2] Mahdi Moqri, Chiara Herzog, Jesse R. Poganik, Jamie Justice, Daniel W. Belsky, Albert Higgins-Chen, Alexey Moskalev, Georg Fuellen, Alan A. Cohen, Ivan Bautmans, Martin Widschwendter, Jingzhong Ding, Alexander Fleming, Joan Mannick, Jing-Dong Jackie Han, Alex Zhavoronkov, Nir Barzilai, Matt Kaeberlein, Steven Cummings, Brian K. Kennedy, Luigi Ferrucci, Steve Horvath, Eric Verdin, Andrea B. Maier, Michael P. Snyder, Vittorio Sebastiano, Vadim N. Gladyshev, *Biomarkers of aging for the identification and evaluation of longevity interventions, Cell*, Volume 186, Issue 18, 2023, Pages 3758-3775

[3] Liu WS, You J, Ge YJ, Wu BS, Zhang Y, Chen SD, Zhang YR, Huang SY, Ma LZ, Feng JF, Cheng W, Yu JT. *Association of biological age with health outcomes and its modifiable factors. Aging Cell.* 2023 Dec;22(12):e13995. doi: 10.1111/acel.13995. Epub 2023 Sep 18. PMID: 37723992; PMCID: PMC10726867.

[4] Marshall, A.G., Neikirk, K., Afolabi, J. et al. Update on the Use of Pulse Wave Velocity to Measure Age-Related Vascular Changes. *Curr Hypertens Rep* 26, 131–140 (2024).

[5] Hirota N, Suzuki S, Motogi J, Nakai H, Matsuzawa W, Takayanagi T, Umemoto T, Hyodo A, Satoh K, Arita T, Yagi N, Otsuka T, Yamashita T. Cardiovascular events and artificial intelligence-predicted age using 12-lead electrocardiograms. Int J Cardiol Heart Vasc. 2023 Jan 6;44:101172. doi: 10.1016/j.ijcha.2023.101172. PMID: 36654885; PMCID: PMC9841236.

[6] Chang Chiao-Hsiang , Lin Chin-Sheng , Luo Yu-Sheng , Lee Yung-Tsai , Lin Chin, Electrocardiogram-Based Heart Age Estimation by a Deep Learning Model Provides More Information on the Incidence of Cardiovascular Disorders, *Frontiers in Cardiovascular Medicine* Volume 9 - 2022

[7] Baek Yong-Soo , Lee Dong-Ho , Jo Yoonsu , Lee Sang-Chul , Choi Wonik , Kim Dae-Hyeok, Artificial intelligence-estimated biological heart age using a 12-lead electrocardiogram predicts mortality and cardiovascular outcomes, *Frontiers in Cardiovascular Medicine* Volume 10 - 2023

[8] Barthels, Myrte & De Cooman, Thomas & Gruwez, Henri & Duncker, David & Vandervoort, Pieter. (2024). Predicting Age from a Real World Smartphone Acquired PPG Signal Using Deep Learning. 10.22489/CinC.2024.293.

[9] Pankaj, Maan, P., Kumar, M. et al. Integrated wearable PPG: a multi-vital sign monitoring based on group sparse mode decomposition framework in remote health care using PPG signal. Phys Eng Sci Med (2025).

[10] Karimpour P, May JM, Kyriacou PA. Photoplethysmography for the Assessment of Arterial Stiffness. Sensors (Basel). 2023 Dec 17;23(24):9882. doi: 10.3390/s23249882. PMID: 38139728; PMCID: PMC10747425.

[11] Fine J, Branan KL, Rodriguez AJ, Boonya-Ananta T, Ajmal, Ramella-Roman JC, McShane MJ, Coté GL. Sources of Inaccuracy in Photoplethysmography for Continuous Cardiovascular Monitoring. Biosensors (Basel). 2021 Apr 16;11(4):126. doi: 10.3390/bios11040126. PMID: 33923469; PMCID: PMC8073123.

[12] Jacques, S.L. Optical properties of biological tissues: A review. *Phys. Med. Biol.* 2013, 58, R37.

[13] Elgendi, M. On the analysis of fingertip photoplethysmogram signals. *Curr. Cardiol. Rev.* 2012, 8, 14–25

[14] Ismail Fawaz, H., Lucas, B., Forestier, G. et al. InceptionTime: Finding AlexNet for time series classification. *Data Min Knowl Disc* 34, 1936–1962 (2020).

[15] Aykroyd, R.G.; Lucy, D.; Pollard, A.M; Solheim, T. Regression analysis in adult age estimation *American Journal of Physical Anthropology* 1997, Volume 104(2), 259-265

[16] Iman Beheshti, Scott Nugent, Olivier Potvin, and Simon Duchesne. Bias-adjustment in neuroimagingbased brain age frameworks: A robust scheme. *NeuroImage: Clinical*, 24:102063, 2019.

[17] Zhang B, Zhang S, Feng J, Zhang S. Age-level bias correction in brain age prediction. Neuroimage Clin. 2023;37:103319. doi: 10.1016/j.nicl.2023.103319. Epub 2023 Jan 7. PMID: 36634514; PMCID: PMC9860514.

[18] J Martin Bland and Douglas G Altman. The logrank test. *Bmj*, 328(7447):1073, 2004.

[19] David R Cox. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187–202, 1972.

[20] Patricia M Grambsch and Terry M Therneau. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, 81(3):515–526, 1994.

[21] David Collett. *Modelling survival data in medical research.* Chapman and Hall/CRC, 2023.

[22] Shin H, ad B. Choi NG. Photoplethysmogram based vascular aging assessment using the deep convolutional neural network. Sci Rep Jul. 2022;12(1):11377.

[23] Zachi I Attia, Paul A Friedman, Peter A Noseworthy, Francisco Lopez-Jimenez, Dorothy J Ladewig, Gaurav Satam, Patricia A Pellikka, Thomas M Munger, Samuel J Asirvatham, Christopher G Scott, et al. Age and sex estimation using artificial intelligence from standard 12-lead ecgs. *Circulation: Arrhythmia and Electrophysiology*, 12(9):e007284, 2019.

[24] Andy Schumann, Christian Gaser, Rassoul Sabeghi, P Christian Schulze, Sven Festag, Cord Spreckelsen, and Karl-J¨urgen B¨ar. Using machine learning to estimate the calendar age based on autonomic cardiovascular function. *Frontiers in Aging Neuroscience*, 14:899249, 2023.

[25] Shammi More, Georgios Antonopoulos, Felix Hoffstaedter, Julian Caspers, Simon B Eickhoff, Kaustubh R Patil, Alzheimer's Disease Neuroimaging Initiative, et al. Brain-age prediction: A systematic comparison of machine learning workflows. *NeuroImage*, 270:119947, 2023.

[26] Hualou Liang, Fengqing Zhang, and Xin Niu. Investigating systematic bias in brain age estimation with application to post-traumatic stress disorders. Technical report, Wiley Online Library, 2019.

7 Appendix

A Table: Age-level bias (First measurements data)

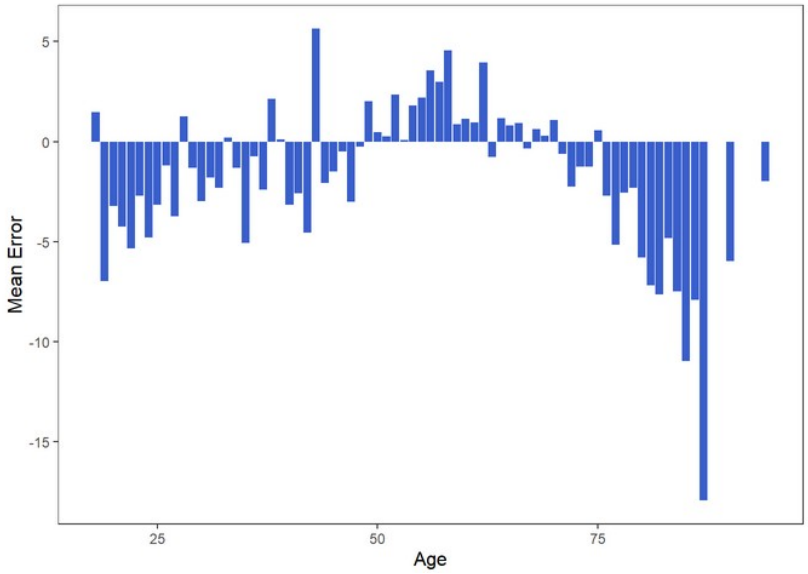

Figure 12: Age-level bias of age deviation by chronological age for the First measurements data after the first correction. On the vertical axis is the mean deviation by chronological age.

B Survival functions of the categorized Deviation groups (First measurements data).

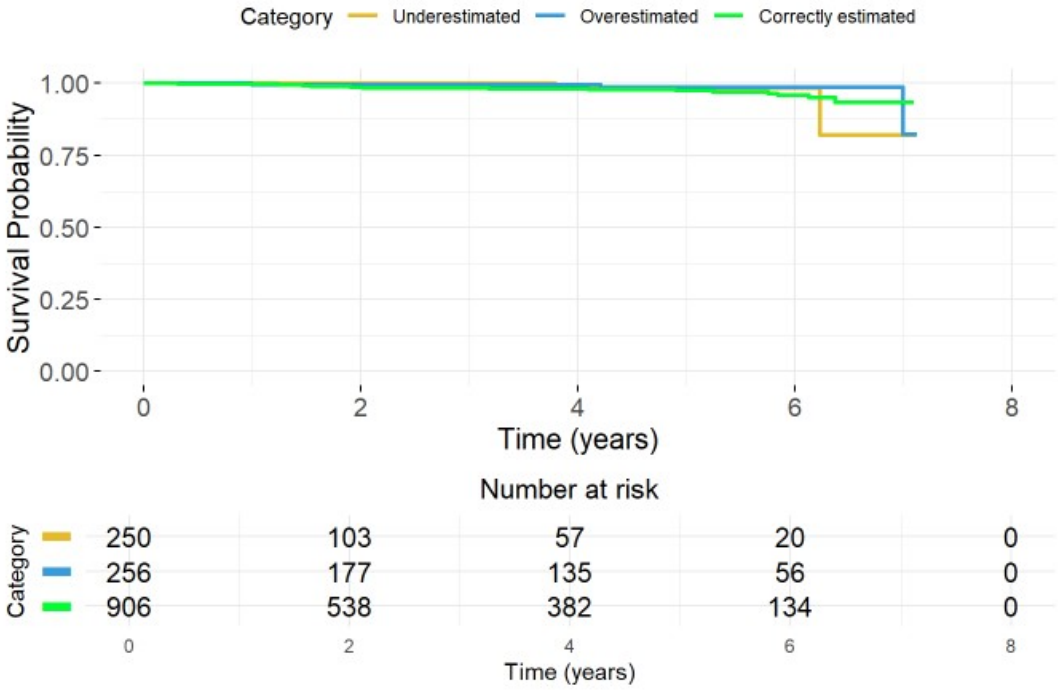

Figure 13: Survival functions of the categorized Deviation groups for the Active users data. On the bottom are the number of patients at risk.

C Table: Focus on the number of comorbidities (First measurements data)

	All	**0**	**1**	**2**	**3**	**4**
n	1414	53	193	348	318	158
mean age	59.0	42.8	53.9	61.9	64.0	65.9
Mean dev.	0	-0.09	-0.12	-0.03	0.07	0.11
MAE	0.81	0.70	0.84	0.82	0.84	0.85
sd(AE)	0.54	0.49	0.56	0.56	0.54	0.53
corr.	1.00	1.00	1.00	1.00	0.99	0.99
Intercept	0.31	0.13	0.29	0.33	0.38	-0.15
slope	0.99	1.00	1.00	1.00	0.99	1.00
R2	0.99	1.00	0.99	0.99	0.99	0.98

Table 21: When patients have 0 up to 4 comorbidities. First column gives the results when all off the patients were considered.

D Table: focus on a specific comorbidity (First measurements data)

	All	**No**	**DIA**	**SMO**	**HTN**	**HCL**
n	1414	53	284	599	1167	818
mean age	59.0	42.8	65.0	61.5	60.9	63.8
Mean dev.	0	-0.09	0.02	0.08	0.01	0.05
MAE	0.81	0.70	0.88	0.81	0.83	0.83
sd(AE)	0.54	0.49	0.54	0.54	0.55	0.54
corr.	1.00	1.00	0.99	1.00	0.99	0.99
Intercept	0.31	0.13	0.42	0.19	0.31	0.30
slope	0.99	1.00	0.99	1.00	0.99	0.99
R2	0.99	1.00	0.99	0.96	0.99	0.99

Table 22: When patients have at least one comorbidity (First measurements). First column are the results when all of the patients are taken into account. Second column are the results from de group of patients that don't suffer from any of the comorbidities.

E Table: Presence of risk factors for PPG age deviation groups in the Fitst measurements data.

	Underestimated (< -0.966)	**Low**	**Overestimated** (> 0.966)	Test statistic	p-value
Gender (M)	70.0%	54.9%	49.2%	$\chi^2 = 25.06$	$3.6 \cdot 10^{-6}$
(n)	175	497	126		
Diabetes	29.1%	22.8%	30.8%	$\chi^2 = 7.18$	0.028
(n)	61	164	64		
Smoking	43.0%	44.1%	48.2%	$\chi^2 = 1.64$	0.441
(n)	104	383	120		
Hyperchol.	70.1%	70.9%	71.5%	$\chi^2 = 1.67$	0.43
(n)	150	526	165		
Hypertension	92.8%	89.9%	92.7%	$\chi^2 = 3.05$	0.22
(n)	219	754	228		
Mean BMI	27.7	27.3	27.8	$F = 1.45$	0.243
(SD)	4.61	4.87	4.76		

Table 23: Presence of risk factors for PPG age deviation groups in the First measurements data. P-values are not corrected; n refers to the total number of patients in a group that suffers from the comorbidity.

F Table: Comparing female and male patients (First measurements data)

Gender	n	Mean age	Mean deviation	MAE	Correlation
Female	615	57.2	-8.1	12.0	0.54
Male	799	60.3	-11.9	13.8	0.41

Table 24: Comparing genders for mean deviation, MAE and correlation between chronological age and PPG predicted age.

G Table: Comparing female and male patients (Active users data)

Gender	n	Mean age	Mean deviation	MAE	Correlation
Female	428	57.4	-9.4	12.0	0.656
Male	549	61.8	-13.2	14.0	0.489

Table 25: Comparing genders for mean deviation, MAE and correlation between chronological age and PPG predicted age.(Active users data)

H Table: Distribution of comorbidities after data selection (Active users data).

Data obtained after combination of data tables and data selection.

Diabetes	Yes	No	Total
Female	37	254	291
Male	139	321	460
	176	575	751

Table 26: Diabetes by gender.

Smoking	Yes	No	Total
Female	153	262	415
Male	269	261	530
	422	523	945

Table 27: Smoking by gender.

Hypertension	Yes	No	Total
Female	364	33	397
Male	483	41	524
	847	74	921

Table 28: Hypertension by gender.

Hyperchol.	Yes	No	Total
Female	217	116	333
Male	380	105	485
	597	221	818

Table 29: Hypercholestorolemia by gender.

I Table: Distribution of comorbidities after data selection (First measurements data).

Data obtained after combination of data tables and data selection.

Diabetes	Yes	No	Total
Female	104	367	471
Male	186	481	667
	290	484	1138

Table 30: Diabetes by gender.

Smoking	Yes	No	Total
Female	221	369	590
Male	386	385	771
	607	754	1361

Table 31: Smoking by gender.

Hypertension	Yes	No	Total
Female	506	55	561
Male	697	65	762
	1203	120	1323

Table 32: Hypertension by gender.

Hyperchol.	Yes	No	Total
Female	304	176	480
Male	538	159	697
	842	335	1177

Table 33: Hypercholestorolemia by gender.

J Plot: Schoenfeld residual plots for the first Cox regression model (Active users data).

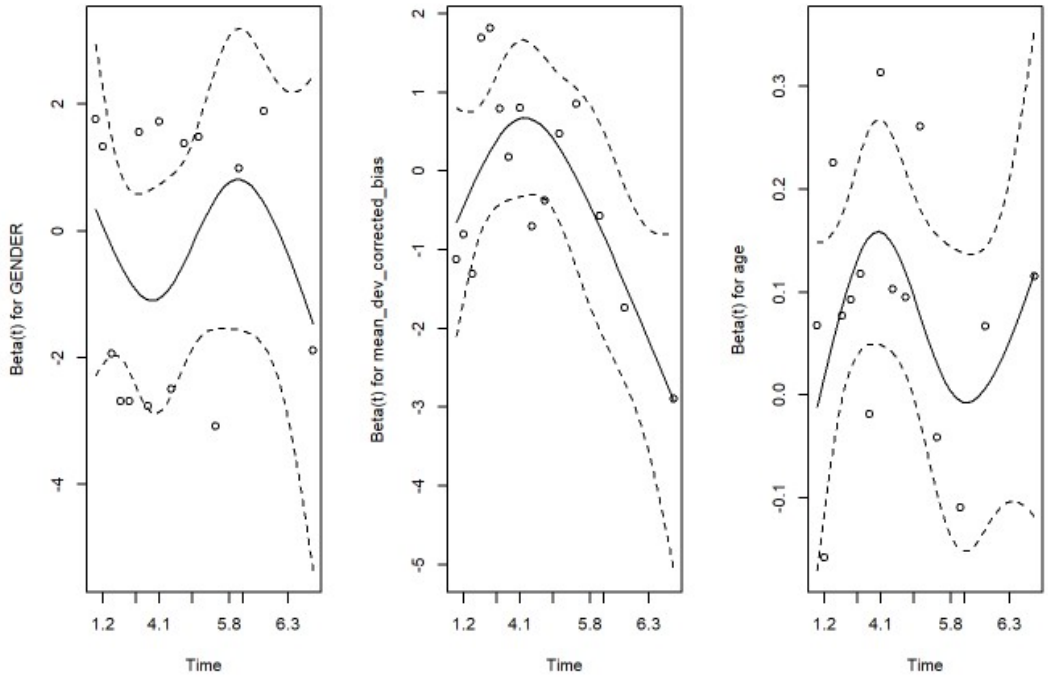

Figure 14: Schoenfeld residuals for gender, corrected age deviation and age.

K Plot: Schoenfeld residual plots for the full Cox regression model (Active users data).

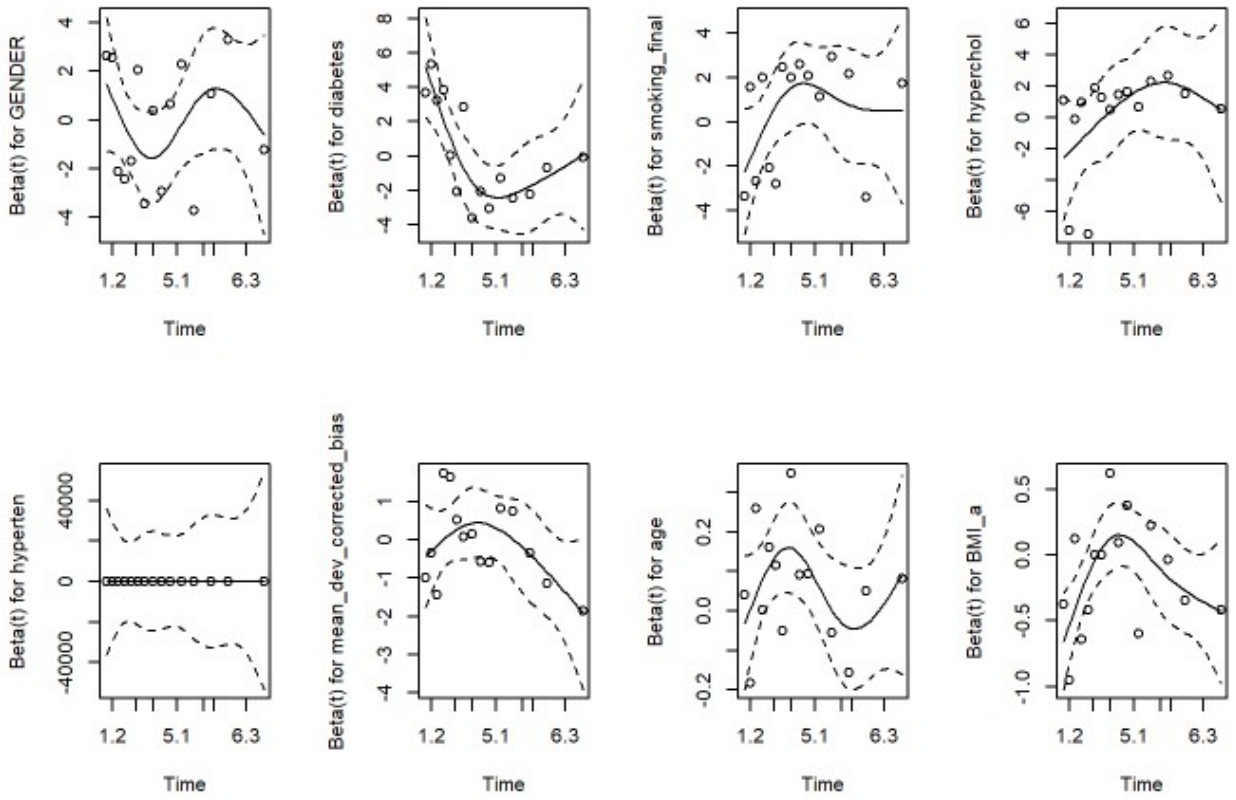

Figure 15: Schoenfeld residuals for gender, diabetes, smoking, hypercholesterolemia, hypertension, corrected age deviation, age and BMI.

L Plot: Schoenfeld residual plots for the first Cox regression model (First measurements data).

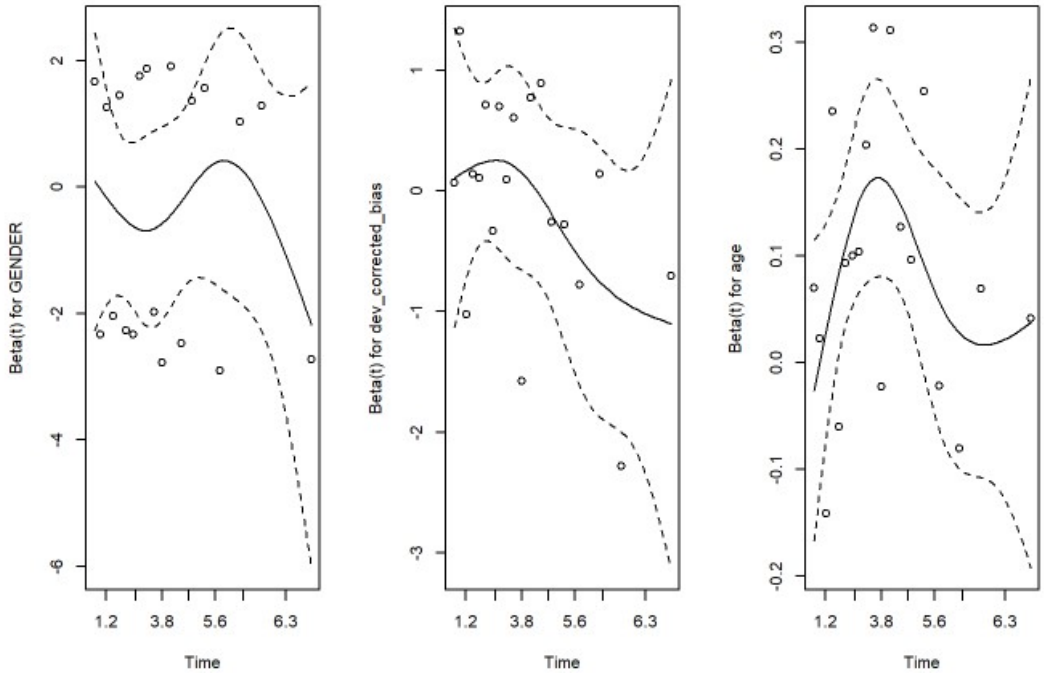

Figure 16: Schoenfeld residuals for gender, corrected age deviation and age.

M Plot: Schoenfeld residual plots for the full Cox regression model (First measurements data).

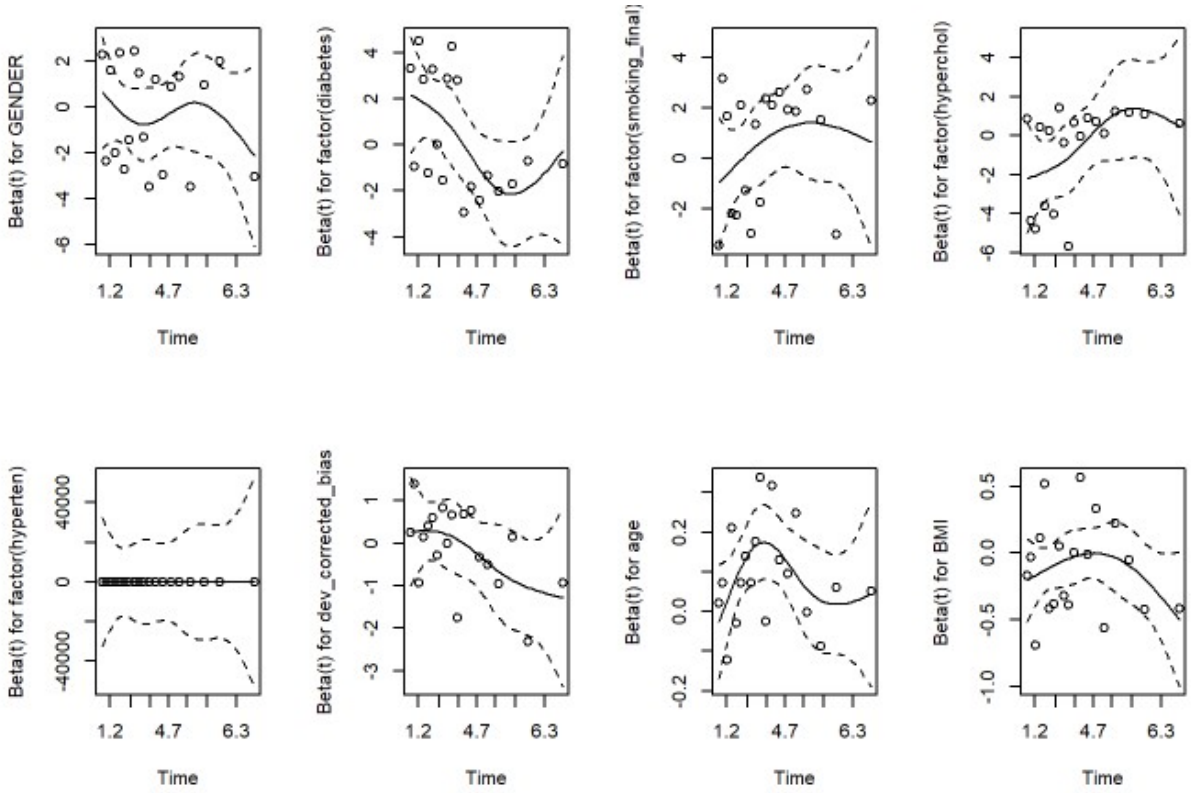

Figure 17: Schoenfeld residuals for gender, diabetes, smoking, hypercholesterolemiia, hypertension, corrected age deviation, age and BMI.

N Plot: Predicted PPG age against survival time (First measurements data).

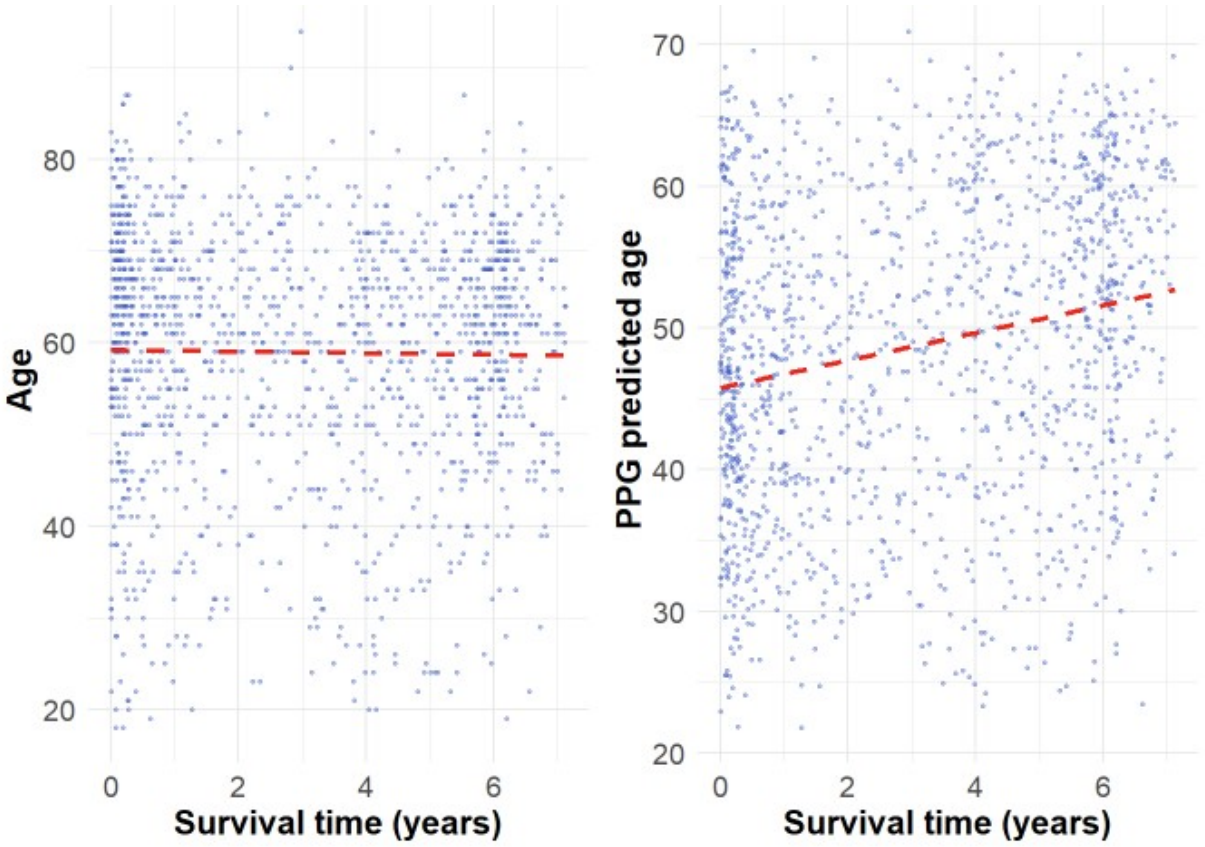

Figure 18: Left: the age of the patients at their first PPG measurement against their survival time. Right: the original predicted PPG age at the first measurement for every patient in the study against survival time. There appears to be a positive correlation between the predictions and the survival times. The average age at the first measurement for the patients in the study remains constant over time.

O Plot: Age and predicted PPG age against the date of the first PPG measurement. (First measurements data).

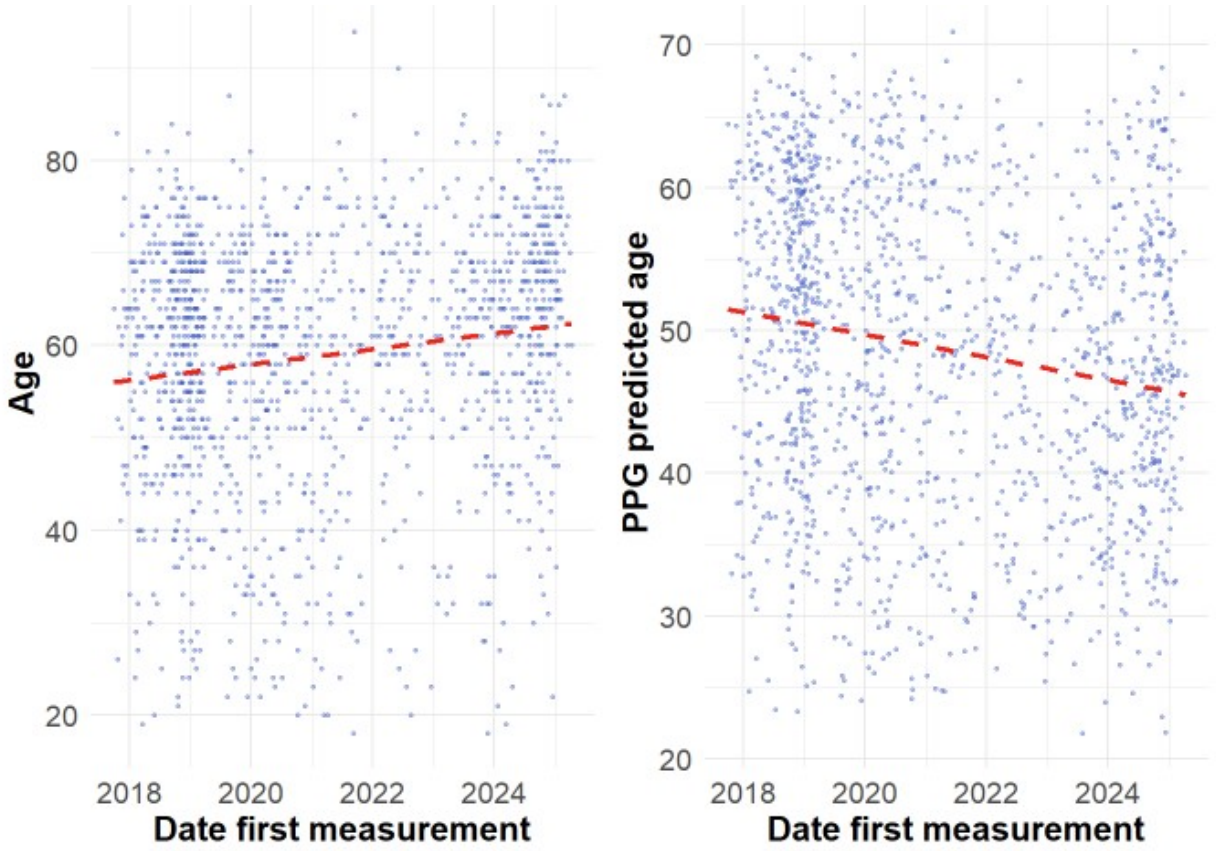

Figure 19: On average the ages of the patients at first measurement increase while the predicted PPG ages at the first measurement decrease.

P Plot: Martingale residuals plot (First measurements data).

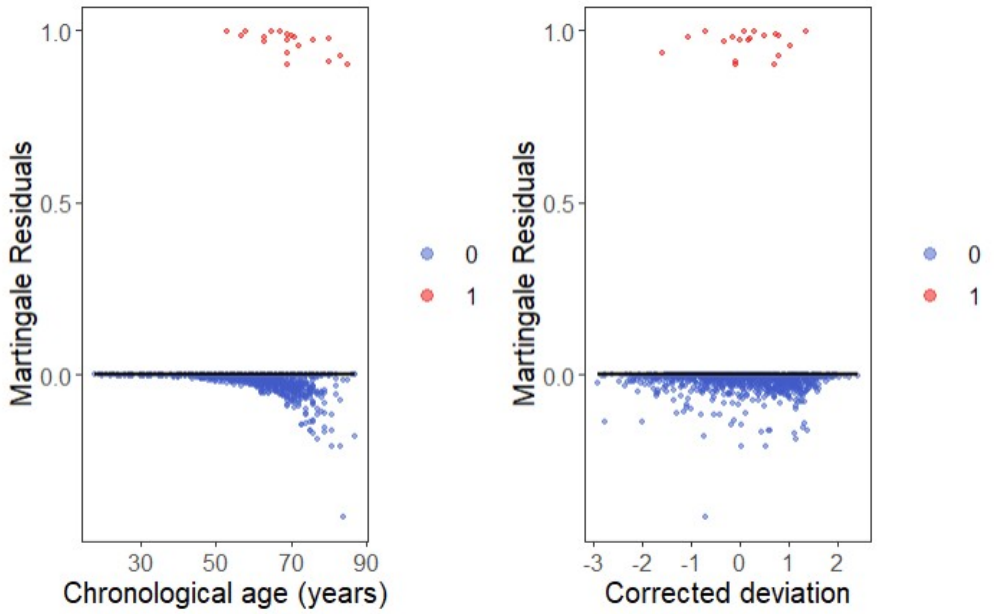

Figure 20: Evaluation of functional form for chronological age and age deviation using martingale residuals for the Cox regression model with age, age deviation and gender as variables. Censored = 0, uncensored = 1.

Q Plot: The log-cumulative hazards functions (First measurements data).

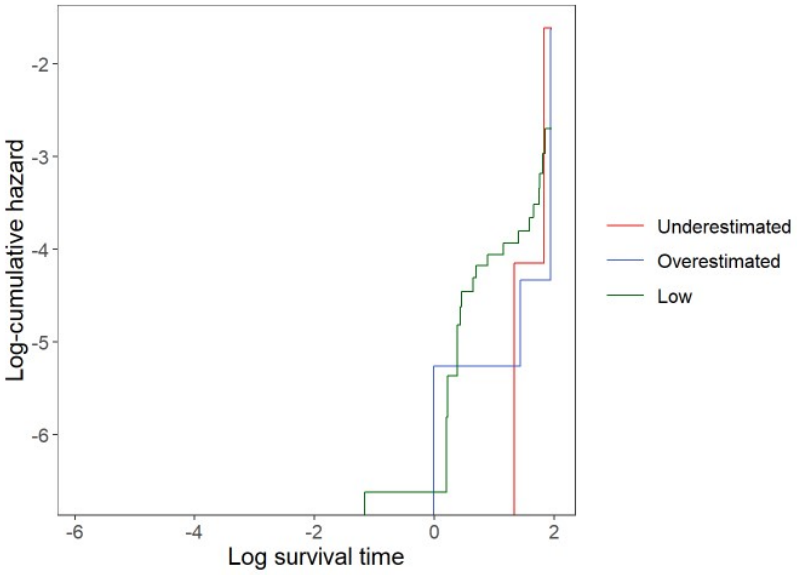

Figure 21: The log-cumulative hazards functions are plotted to assess the proportional hazards assumption. The assumption is supported if the survival functions of the deviation groups are more or less parallel which is not the case.

R R Code

```
# number of measurements and mean PPG age per patient and per year
# ndays is the number of distinct days the patient has taken PPG measurements.
# day1 = day of the first measurement
# day2 = day of the last measurement
# age = age of the patient at the end of the year
data <- measurements %>%   group_by(patient_id, GESLACHT, year) %>%
  summarise(count_Y = n(),mean_PPG = mean(predicted_age),age = max(age),
            day1 = min(date), day2 = max(date),ndays = n_distinct(date))

# Active (or motivation) is the ratio of the number day the patient has taken measurements
# over the number of days between first and last measurement.
data_motivation$active <-
    m_motivation$ndays/(as.numeric(m_motivation$day2 - m_motivation$day1) + 1)

# Standardization of predicted age (ACTIVE data)
# deviation = PPG-age - age
data$mean_dev = data$mean_PPG - data$age

# linear regression of averaged PPG age by chronological age
m_model <- lm(mean_PPG ~ age, data = data)
m_intercept <- coef(m_model)[1]
m_slope <- coef(m_model)[2]

# PAc = PA + error = PA + CA - (int + (CA*slope)) = PA + CA - int - CA*slope
data$PPGage_corrected <- data$mean_PPG + data$age - data$age*m_slope - m_intercept

data$mean_dev_corrected <- data$PPGage_corrected - data$age

# Age-level correction
age_level1m <- data %>% group_by(age) %>%
    summarize(Mean_error = mean(mean_dev_corrected), SD_error = sd(mean_dev_corrected),
              n = n(), Mean_age1 = mean(PPGage_corrected), SD_age = sd(PPGage_corrected))

#standardized mean deviation
data$mean_dev_corrected_bias <- (data$mean_dev_corrected - data$Mean_error)/data$SD_error
data$PPGage_corrected_bias <- data$mean_dev_corrected_bias + data$age

# multivariate regression with averaged BMI
mv_model_57 <- lm(PPGage_corrected_bias ~ age + BMI_a + GESLACHT + diabetes +
                    smoking + hyperchol + hyperten, data = data)


# Fit Cox regression model
cox_fit01 <- coxph(Surv(time, OVERLEDEN) ~ GESLACHT  + mean_dev_corrected_bias + age,
                 data = na.omit(patients))
# Fit full Cox regression model
cox_fit01b <- coxph(Surv(time, OVERLEDEN) ~ GESLACHT + diabetes + smoking + hyperchol +
        hyperten  + mean_dev_corrected_bias + age + BMI_a, data = na.omit(patients))
```

```
# Schoenfeld residuals test
cox_check <-cox.zph(cox_fit01)
# martingale residuals
pat <- na.omit(patients)
pat$martingale <- residuals(cox_fit01, type = "martingale")

# Fit Cox regression model
cox_fit03 <- coxph(Surv(time, OVERLEDEN) ~ GESLACHT  + dev_cat + age,
                   data = na.omit(patients))

# Create survival object
surv_obj <- Surv(time = patients$time, event = patients$OVERLEDEN)

# Fit Kaplan-Meier model
km_fit <- survfit(surv_obj ~ dev_cat, data = patients)

surv_data <- data.frame(
  time = km_fit$time,
  surv = km_fit$surv,
  strata = rep(names(km_fit$strata), unique(km_fit$strata))
)

ggplot(surv_data, aes(x = log(time), y = log(-log(surv)),color = strata)) +
  geom_step() +
  labs(
    x = "Log survival time",
    y = "Log-cumulative hazard"
  ) +
  scale_color_manual(values = c("red2", "royalblue3","darkgreen"), labels = c("Underestimated"
  theme_apa(x.font.size = 15, y.font.size = 15) +
  theme(axis.text = element_text(size = 12),
    axis.title = element_text())
```