

# Deviation of Physiological from Chronological Age Is Associated with Health

Lin PERETZ<sup>a</sup> and Nadav RAPPOPORT<sup>a,1</sup>

<sup>a</sup> *Software and Information Systems Engineering, Ben-Gurion University of the Negev, ISRAEL*

**Abstract.** Biological age may be of higher importance than chronological age, yet biological age is not trivial to estimate. This study presents a regression model to predict age using routine clinical tests like laboratory tests using the UK Biobank (UKBB) data. We run different machine learning regression models for this predictions task and compare their performance according to RMSE. The models were trained using data from 472,189 subjects aged 37–82 years old and 61 different laboratory tests results. Our chosen model was an XGboost model, which achieved an RMSE of 6.67 years. Subjects whose the model predicted to be younger than their actual age were found to be healthier as they had fewer diagnoses, fewer operations, and had a lower prevalence of specific diseases than age-matched controls. On the other hand, subjects predicted to be older than their chronological age had no significant differences in the number of diagnoses, number of operations, and specific diseases than age-matched controls.

**Keywords.** Biological age, Machine Learning, Laboratory tests, Chronological age, Electronic Health Records, BioBank

## 1. Introduction

As the world's aging population grows at an unprecedented rate, there is a clear need to learn more about the biological aging process and the determinants of healthy aging. To achieve this goal, researchers seek biological markers and other factors that can track biophysiological aging and, ideally, provide insight into the underlying mechanisms [1,2,3,4].

Biological age (BA), also called physiological age, measures how well or poorly a person's body functions. BA is correlated with calendar age (CA), also called chronological age, which is an objective measure of elapsed time since birth. The BA of a person can be higher or lower than their CA, since aging is not only a matter of time but is, in fact, a complex process with multiple causes. Studies have shown that young individuals of the same chronological age varied in their BA. These individuals showed cognitive decline and brain aging, self-reported worse health, and looked older [5].

There are several ways for determining BA, but none are definitive or truly accurate. Previous studies used a variety of ways to estimate the BA of a person. For example, studies used human physical activity as recorded by a wearable device [1], by cognitive variation independently of chronological age [6]. Other studies used molecular

---

<sup>1</sup> Corresponding Author, Nadav Rappoport, Department of Software and Information Systems Engineering, Ben-Gurion University of the Negev, ISRAEL, 1 Ben-Gurion Boulevard, Beer-Sheva, ISRAEL; E-mail: nadavrap@bgu.ac.il.

biomarkers like telomeres' length and others [1,2,3,4]. These approaches for estimating BA are limited as it requires the active collection of data specifically for this purpose and is costly.

This study presents a predictive model of chronological age applied to the UK Biobank (UKBB) data. The type of data used is often routinely collected laboratory test results. The advantage of using routinely collected laboratory results for estimating BA is that such a resource is very readily available, easy, and cheap to execute and does not need to be actively collected for our purpose.

Laboratory test results from the UKBB were used as features to train different machine learning regression models, and we chose the model with the lowest RMSE on a 25% held-out test set. This model has good performance in predicting the subjects' age, but the model makes significant errors in some cases. The cases where there is a large deviation between the predicted age and the actual age are the focus of this study. These cases were divided into two groups. Group P-Y (predicted younger) for which the model predicted to be younger than their actual age, and Group P-O (predicted older) for which the model predicted to be older than their actual age. On these deviation groups, we conducted further investigation determining their overall health vis-à-vis the rest of the population and vis-à-vis their age-matched groups.

In another study, Zichen et al. used vital signs and lab tests from an EMR to predict CA [7]. They have found specific diagnoses enriched in subjects with a high discrepancy between CA and BA.

Our study focused on adults population and demonstrated the use of BioBank data to assess the overall health, using different measures, of the population with a high discrepancy between CA and BA.

## 2. Methods

### 2.1 Data

The UK Biobank (UKBB) is an open-access research database and a large prospective study with about 500,000 adults (aged 40-69 at recruitment) participants recruited from 2006 to 2010. This database covers thousands of clinical and environmental variables such as demographic features (e.g., sex, age), laboratory test values, previous and current illnesses, lifestyle, imaging data, hospital discharges, and more [8]. We used data from 472,189 subjects consisting of 61 types of laboratory tests and the subject's gender and chronological age (CA). Missing values laboratory tests values were imputed using the median.

### 2.2 A predictive model for aging

To find the best model to predict the CA, we trained different machine learning regression models and compared their performance according to the root mean squared error (RMSE). The models were trained on a 75% random sample, and performances were evaluated using the 25% held out test set.

### 2.3 *Identifying individuals with a significant discrepancy between predicted and chronological age*

To identify subjects with significantly older or younger predicted age than their CA, we retrain our model on all the data without splitting it to train and test. Subjects for whom the model predicted a difference of more than two standard deviations from their CA, were considered as cases. Cases were divided into two groups: (i) cases with a prediction younger than their CA (P-Y); (ii) cases with a prediction older than their CA (P-O). We created two age-matching control cohorts. P-Y's control group contains subjects with CA of 67-82, matching CA range of group P-Y. P-O's control group contains subjects with CA in the range of 40-47, which is the CA of group P-O.

### 2.4 *Outcome's comparisons*

We used two sets of outcomes of interest, Overall health measures and disease-specific. We used several measures for estimating the overall health of each group. We used inpatient data available for a subset of the UKBB population. The data include admissions, diagnostic, and operation codes (UKBB data fields 41272, 41211, and 41270, respectively). Diagnoses were considered in two ways. In the first, we counted the number of ICD-10 diagnoses codes for each subject. In the second, we counted the number of groups of diagnoses codes based on Clinical Classifications Software (CCS) [9]. Subjects for whom clinical data were not extracted from electronic health record systems were excluded. After this step, we were left with 413,828 subjects, 177 in group P-O and 343 in group P-Y.

We compared the number of diagnoses, admissions, and operations of the two cases groups versus their age-matched groups. We tested the differences between the groups using the non-parametric Kolmogorov–Smirnov (KS) test to examine the hypothesis that there are statistically significant differences between the P-Y/P-O groups and their age-matched groups.

To perform the disease-specific analysis, we extracted diagnoses for a set of common diseases. Disease extraction was conducted in two different ways. First, we used the health-related outcomes category from the UKBB and extracted diseases with specific data filed. Second, we extract additional diseases using the ICD10 codes. We performed a chi-squared test to test the hypothesis that there are statistically significant differences between P-Y/P-O groups versus their age-matched groups and controlled for multiple testing using the Benjamini Hochberg method.

## 3. Results

We trained different machine learning regression models and computed their performance using RMSE. The Linear Regression model had the lowest performance (RMSE=7.214) compared to more complex models such as Random Forest (RMSE=6.803), XGboost (RMSE 6.271), and Deep Learning model multilayer perceptron (MLP) (RMSE=6.389). We chose to use the XGboost model as our age prediction model due to its lowest RMSE.

Hyper-parameters tuning was performed using a grid search cross-validation approach. The chosen parameters were learning rate of 0.03, maximum depth of a tree equal to 12, minimum sum of instance weight equal to 15, number of estimators equal to

500, and subsample equal to 0.75. The RMSE of this model on the test set was 6.271 years.

### 3.1 Outcome's comparisons

The two groups of interest were compared to the matching controls using inpatient data. We found that the P-Y group had significantly fewer diagnoses, group of diagnoses, number of operations, and number of admissions. On the other hand, the P-O group showed no significant differences compared to their age-matched group (Table 1).

Next, we compared differences in the prevalence of specific diseases between the groups, and we applied Benjamini Hochberg's correction of p-values for multiple tests [10]. First, we compared group P-Y with their age-matched and found that they had statistically significant lower prevalences of all 16 tested diseases (Table 2). Second, we compare group P-O with their age-matched and didn't see significant differences between the groups (Table 2).

**Table 1.** Outcomes comparison between the groups. Mean (SD) and the p-value of the KS test results. The group of diagnosis codes is based on CCS classification (see Methods).

	Mean (SD)				P-value	
	P-O	40-47	P-Y	67-82	P-O vs. 40-47	P-Y vs. 67-82
#Diagnoses	6.98 (7.6)	6.94 (8.6)	6.59 (5.8)	11.74 (12.2)	0.27	<b>1.26e-12</b>
#Grouped Diagnoses	5.69 (5.5)	6.27 (6.7)	5.66 (4.6)	9.28 (8.4)	0.42	<b>3.65e-12</b>
#Operations	6.51 (6.5)	6.39 (6.6)	6.80 (5.4)	9.76 (8.6)	0.99	<b>2.74e-07</b>
#Admissions	6.98 (7.6)	6.94 (8.6)	6.59 (5.8)	11.74 (12.2)	0.27	<b>1.26e-12</b>

**Table 2.** Prevalence of different diseases in 4 groups of subjects. Prevalence is given in percentages. The difference between the affected group and its matched controls is adjusted for multiple false discovery rate using the Benjamini-Hochberg method.

Disease	Group P-Y (n=343)	Age-matched (n=64,550)	P-value	Group P-O (n=177)	Age-matched (n=61,610)	P-value
Cancer	24.78%	26.85%	0.504	9.60%	8.66%	1
Asthma	9.03%	13.00%	0.064	19.20%	15.53%	1
COPD	2.04%	6.41%	0.003	3.38%	1.37%	0.784
Dementia	0.29%	1.55%	0.15	0.00%	0.06%	1
ESRD	0.00%	0.26%	0.707	0.56%	0.18%	1
Motor Neuron	0.00%	0.13%	1	0.00%	0.02%	1
Heart Attack	4.08%	7.88%	0.027	2.25%	1.36%	1
Parkinson	0.29%	0.94%	0.482	0.00%	0.08%	1
Stroke	1.45%	4.79%	0.016	1.12%	1.05%	1
Hypertension	32.65%	50.53%	8.76E-10	14.68%	14.36%	1
Anemia	1.45%	4.16%	0.036	2.25%	2.94%	1
Atherosclerosis	0.29%	0.81%	0.504	0.00%	0.09%	1
Chronic liver disease	0.00%	0.45%	0.504	0.00%	0.40%	1
Diabetes	1.45%	9.77%	1.89E-06	5.64%	3.40%	1
Heart disease	11.37%	27.01%	8.76E-10	4.51%	5.73%	1
Kidney disease	1.45%	7.17%	2.60E-04	2.25%	1.36%	1

#### 4. Discussion

A machine learning regression model was shown to be valuable for predicting age based on laboratory test results. Our model can predict subjects' CA with an RMSE of 6.27 years. The cohort of subjects with estimated BA significantly lower CA has overall better health than the age-matched group. The overall health was approximated using the number of diagnoses, operations, and admissions. In addition, this population was found to have lower prevalences for all 16 tested diseases. On the other hand, we found that the cohort for which the BA was higher than their CA had no significant differences.

This study has a few limitations. First, the cohort is composed of only adults subjects, and its application to a broader range of ages is yet to be proven. In addition, the models are based on tens of different laboratory tests, which limits its broad use. Moreover, in this study, we showed that there is an association between the deviation of BA from CA and health. But this does not imply causation, nor any suggestion for intervention that can improve health.

#### 5. Conclusion

We demonstrated the usefulness of a machine learning model for age prediction based on laboratory test results. We showed that subjects with a significantly lower predicted age than actual age were 'healthier'.

#### References

- [1] Syed Ashiqur R, Adjero DA. "Deep Learning Using Convolutional LSTM Estimates Biological Age from Physical Activity." *Scientific Reports* 9, no. 1 (December 2019): 11425.
- [2] Janet L, McGill NI, Lindsey LA, Green DK, Cooke HJ. "In Vivo Loss of Telomeric Repeats with Age in Humans." *Mutation Research/DNAging* 256, no. 1 (January 1991): 45–48.
- [3] Bae CY, Kang YG, Kim S, Cho C, Kang HC, Yu BY, et al. "Development of models for predicting biological age (BA) with physical, biochemical, and hormonal parameters". *Archives of Gerontology and Geriatrics*, 47, 2, September–October 2008.
- [4] Belsky DW, Caspi A, Houts R, Cohen HJ, Corcoran DL, Danese A, et al. "Quantification of Biological Aging in Young Adults." *Proceedings of the National Academy of Sciences* 112, no. 30 (July 28, 2015): E4104–10.
- [5] Åke W, MacDonald SWS, de Frias CM, Nilsson LG, Dixon RA. "How Do Health and Biological Age Influence Chronological Age and Sex Differences in Cognitive Aging: Moderating, Mediating, or Both?" *Psychology and Aging* 21, no. 2 (2006): 318–32.
- [6] Zichen W, Li L, Glicksberg BS, Israel A, Dudley JT, Ma'ayan A. "Predicting Age by Mining Electronic Medical Records with Deep Learning Characterizes Differences between Chronological and Physiological Age." *Journal of Biomedical Informatics* 76 (December 2017): 59–68.
- [7] Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. "UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age." *PLOS Medicine* 12, no. 3 (March 31, 2015): e1001779.
- [8] Tianqi C, Guestrin C. "XGBoost: A Scalable Tree Boosting System." *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, August 13, 2016, 785–94
- [9] Salsabili M, Kiogou S, Adam TJ. "The Evaluation of Clinical Classifications Software Using the National Inpatient Sample Database," n.d., 10.