Data, Informatics and Technology: An Inspiration for Improved Healthcare
A. Hasman et al. (Eds.)
IOS Press, 2018

doi:10.3233/978-1-61499-880-8-125

Physical Fitness Forecasting and Risk Estimation in Slovenian Schoolchildren

Anton GRADIŠEK[a,1] , Miha MLAKAR[a], Matej CIGALE[a], Luka LAJOVIC[a], Mitja LUŠTREK[a], Maroje SORIĆ[b,c], Gregor STARC[b], Bojan LESKOŠEK[b] and Gregor JURAK[b]
[a] *Jožef Stefan Institute, Ljubljana, Slovenia*
[b] *Faculty of Sports, University of Ljubljana, Ljubljana, Slovenia*
[c] *Faculty of Kinesiology, University of Zagreb, Croatia*

Abstract. Physical fitness is important in view of reducing risks for a number of non-communicable diseases, both for individuals and policy-makers. In this paper, we present a prototype tool that combines forecasting of individual fitness parameters of schoolchildren to early adulthood with estimation of relative risk for all-cause early mortality in adulthood based on the forecasted fitness. This tool is a first step in the development of a platform that will show age, gender, and geographical distributions of risk and suggest potential interventions.

Keywords. SLOfit, exercise, machine learning, CrowdHEALTH

1. Introduction

The prevalence of overweight and obesity is rising globally [1], with the current obesity epidemic especially alarming among children and adolescents [2]. Lack of physical activity (PA), alongside poor nutrition, is proposed as one of the major contributors to childhood obesity [3] and one of the major public health problems in the world [4]. A wealth of evidence demonstrates that regular PA reduces all-cause mortality and the incidence of cardiovascular diseases, type-2 diabetes, and cancer, and enhances bone strength and psychological health [5]. Therefore, both halting adiposity and rising physical activity have been included as global non-communicable disease targets in the forthcoming period by the WHO [4].

Physical fitness is closely linked to PA level of an individual and is often used as a proxy measure for it. Abundant evidence links both overweight and low fitness with higher incidence of several non-communicable diseases and mortality in adult population [4], with the strongest protective effect being found for cardiorespiratory fitness (CRF) and muscular fitness (MF) [6-8]. However, these outcomes are very rare in children. Hence, it is difficult to relate either weight status or low fitness to hard health outcomes in this age group. Instead, the associations of fitness and obesity with proxy measures of health are usually described. Examples include clustered metabolic or cardiovascular risk [9] (i.e. a set of biochemical measures of carbohydrate and lipid metabolism). Yet, a stronger message to policy-makers, as well as the public, would be

---

[1] Corresponding Author, Anton Gradišek; E-mail: anton.gradisek@ijs.si

conveyed if the risk of future death based on childhood health-related fitness could be presented.

In Slovenia, every April, almost the entire Slovenian population aged 6 to 18 (220,000 students) is measured using 8 motor tests and 3 anthropometric measurements (SLOfit, www.slofit.org) [10]. The study contains data on physical fitness of Slovenian primary and high school students and enables annual monitoring of physical and motor status of children in all Slovenian schools from 1987 onwards. On the national level, the SLOfit data serves as the scientific backbone for most of the policies related to enhancement of physical activity of children and youth, and the policies related to school physical education. To date, the SLOfit database includes over 7 million sets of measurements for over 1 million children and is one of the largest cross-sectional and cohort databases of physical and motor development in the world. Slovenian educational policy, informed by the SLOfit data, managed to develop one of the most efficient system of physical education and extracurricular sports programs in the world, which results in a very favourable level of physical fitness and physical activity of children in Slovenia in comparison to the rest of the world [11].

Here, we present the prototype of a tool that allows us to estimate the total relative risk for premature death related to weight status, CRF, and MF in adolescence. The tool is based on the SLOfit dataset and shows the relative risks for all-cause mortality based on the BMI, MF, and CRF at the age of 18. The forecasting of parameters from a given age to the age of 18 is done using methods of artificial intelligence. This tool, developed within the framework of the EU H2020 project CrowdHEALTH (http://crowdhealth.eu/), will serve to suggest health-related interventions for policy-makers.

2. Forecasting algorithms

The task of the forecasting algorithm is to predict a particular SLOfit parameter (height, weight, 60 s sit-ups, 600 m run) at the age of 18, based on the data form previous years and knowing the general population trends.

The simplest baseline model uses the percentile method: for example, if an individual is in the N^{th} percentile at the age of 13, we assume he would be in the same percentile at the age of 18 (of course, this model clearly has limitations, such as not taking into account the fact that the puberty can start at different ages). More advanced approaches use machine learning. To improve the prediction accuracy, we generated additional features, including average, maximum, minimum year growth, standard deviations, peak height velocity (the year with the largest increase in growth), as well as the percentile data. Next, we built a model for each year up to which we have available data. For example, the model for the age of 13 takes the measurements from 6 to 13 and forecasts the value at the age of 18. Each model is built on a single type of data (e.g. height).

Several machine learning algorithms were tested on a dataset of about 2000 children. To evaluate them, the average absolute error was calculated. The best results were obtained using a linear regression model [12]. An example forecast is shown in Fig. 1. In future work, the entire SLOfit dataset will be used to build models, as well as combinations of SLOfit parameters instead of single-type data.

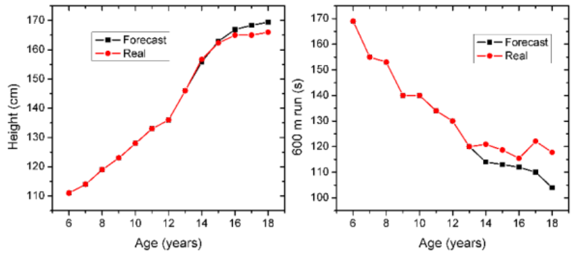

Figure 1. Real height and 600 m data together with the forecast data, based on the data from ages 6 to 13.

3. Risk assessment

Among 11 measures included in the SLOfit system, BMI, 600 m run, and 60 s sit-ups were chosen as indicators of weight status, CRF and MF, respectively. To estimate the risk of all-cause mortality in adulthood, we first predicted the values of these traits at age 18, and then related this to the relative risk of premature death these values convey at young adulthood. To ascribe risks to the specific trait, we searched the PubMed database for studies on young to middle-aged subjects (20-50 yrs.) relating obesity and/or fitness to all-cause mortality. Search terms were entered separately for each of the 3 traits included in the study (e.g. CRF OR aerobic endurance AND all-cause mortality OR premature death). The available studies were first ranked according to the type: 1. meta-analysis; 2. prospective cohort study; 3. retrospective cohort study; 4. cross-sectional study. Next, we ordered the studies based on: 1) method of obesity/fitness assessment, 2) methodological quality, 3) N of participants. Out of several meta-analyses extracted for weight status, we chose the highest-ranking one according to the criteria mentioned above [13]. Conversely, the absence of appropriate meta-analysis for CRF and MF forced us to rely on the highest ranked among the prospective cohort studies available [8,14]. The relative risk reported in these studies relates to prospective risk of early death in individuals that were healthy at baseline. Deaths in the first few years are excluded to minimise the risk of reverse causality.

Table 1. Relative increase in risk (in %) for all-cause mortality in men, related to suboptimal values of BMI, CRF, and MF (based on [12], [13], and [8], respectively)

BMI (kg/m^2)	15-18.5	18.5-20	20-22.5	22.5-25	25-27.5	27.5-30	30-35	35-40	40-60
Risk increase (%)	82	44	2	ref.	7	27	66	166	335
CRF (600m run)	Q1 (best)	Q2	Q3	Q4	Q5 (worst)				
Risk increase (%)	ref.	28	59	78	85				
MF (60s sit-ups)	Q1 (best)	Q2	Q3	Q4 (worst)					
Risk increase (%)	ref.	47	47	172					

Based on how each of the traits was presented in these studies, we divided participants into BMI categories, quantiles of 600 m run and quartiles of 60 s sit-ups. The associated relative risks of all-cause mortality for specific traits are shown in Table 1. Relative risk shows the percentage increase in risk for a specific outcome compared to a reference category. In our case, the reference individual is a lean, very fit person (specifically, an individual that belongs to top 20 % of CRF and top 25 % of MF). For women, the approach is the same, only with different risk increase percentages.

4. Discussion

We present the prototype of the tool that allows both forecasting of fitness from childhood to young adulthood, and estimation of relative risk for all-cause mortality at middle age. This approach allowed us to link fitness in adolescence with health risks in adulthood. In this initial version, the forecasting component only uses single type of data to build the models. Next steps will include a combination of parameters, as well as risk estimates for individual chronic diseases, such as cardiovascular disease or diabetes. As for the risks, at the moment we treat them (mostly) independent of one another. Although estimates for both CRF and MF were adjusted for each other and BMI, we are not aware of a meta-analysis that adjusts BMI for fitness.

The tool will be coupled with an intervention planner. A policy-maker, such as a school principal or an employee at the ministry of education, will be able to see the main risks a particular region faces, and what interventions have proven successful to alleviate these risks in the past. Examples of such interventions include additional physical education classes or improved infrastructure for sports and recreation.

References

[1] NCD Risk Factor Collaboration, Trends in adult body-mass index in 200 countries from 1975 to 2014: a pooled analysis of 1698 population-based measurement studies with 19.2 million participants, *The Lancet* **387** (2016), 1377-1396.

[2] NCD Risk Factor Collaboration, Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128.9 million children, adolescents, and adults, *The Lancet* **390** (2017), 2627-2642.

[3] T. Lobstein, L. Baur, R. Uauy, Obesity in children and young people: a crisis in public health, *Obes. Rev.*, **5** (2004), 4–104.

[4] WHO, Global action plan for the prevention and control of noncommunicable diseases 2013–2020. World Health Organization, Geneva, Switzerland, 2013.

[5] D.S. Warburton et at., A systematic review of the evidence for Canada's Physical Activity Guidelines for Adults, *Int. J. Behav. Nutr. Phys. Act*. **7** (2010), 1-220.

[6] M. Mišigoj-Duraković, M. Sorić, D. Matika, I. Jukić, Z. Duraković, Which is more important for reducing the odds of metabolic syndrome in men: Cardiorespiratory or muscular fitness? *Obesity* **24** (2016), 238-244.

[7] S. Kodama, K. Saito, S. Tanaka, et al., Healthy men and women: a meta-analysis all- cause mortality and cardiovascular events in healthy men and women: a meta- analysis, *JAMA* **301** (2009), 2024-2035.

[8] P. T. Katzmarzyk, C. L. Craig, Musculoskeletal fitness and risk of mortality, *Med. Sci. Sports Exerc*. **34** (2002), 740–744.

[9] U. Ekelund, S.A. Anderssen, K. Froberg, L. B. Sardinha, L.B. Andersen, S. Brage, Independent associations of physical activity and cardiorespiratory fitness with metabolic risk factors in children: the European youth heart study, *Diabetologia* **50** (2007), 1832–1840.

[10] J. Strel, Sports Educational Chart, Ministry of Education and Sport, Ljubljana, 1997.

[11] V. Sember, et al., Results from the Republic of Slovenia's 2016 Report Card on Physical Activity for Children and Youth, *J. Phys. Act. Health*, **13** (2016), S256-S264.

[12] T.M. Mitchell, Machine Learning, The Mc-Graw-Hill Companies, Inc, 1997.

[13] Global BMI Mortality Collaboration, Body-mass index and all-cause mortality: individual-participant-data meta-analysis of 239 prospective studies in four continents, *The Lancet* **388** (2016), 776–786.

[14] C.E. Barlow, L.F. DeFina, et al., Cardiorespiratory fitness and long-term survival in "low-risk" adults, *J. Am. Heart Assoc.* **1** (2012), e001354.