Journal of Economic Behavior & Organization 142 (2017) 259–274

Contents lists available at ScienceDirect

Journal of Economic Behavior & Organization

journal homepage: www.elsevier.com/locate/jebo

Research Paper

Biomarkers and long-term labour market outcomes: The case of creatine☆

Petri Böckerman[a,b,*], Alex Bryson[b,c,d], Jutta Viinikainen[e], Christian Hakulinen[f], Laura Pulkki-Råback[g], Olli Raitakari[h,i], Jaakko Pehkonen[e]

[a] Turku School of Economics, Labour Institute for Economic Research, Helsinki, Finland
[b] IZA, Bonn, Germany
[c] UCL Institute of Education, London, United Kingdom
[d] NIESR, London, United Kingdom
[e] Jyväskylä University School of Business and Economics, Jyväskylä, Finland
[f] Department of Psychology and Logopedics, Faculty of Medicine, University of Helsinki, Helsinki, Finland
[g] Helsinki Collegium for Advanced Studies, University of Helsinki and Department of Psychology and Logopedics, Faculty of Medicine, University of Helsinki, Finland
[h] Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland
[i] Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland

ARTICLE INFO

Article history:
Received 26 August 2016
Received in revised form 27 July 2017
Accepted 2 August 2017
Available online 8 August 2017

JEL clasifications:
I12
J24
J31

Keywords:
Biomarkers
Creatine
Creatinine
Labour market
Earnings
Employment

ABSTRACT

Using the Young Finns Study (YFS) combined with the Finnish Linked Employer-Employee Data (FLEED) we show that quantities of creatine measured in 1980 prior to labour market entry affect labour market outcomes over the period 1990–2010. Those with higher levels of creatine (proxied by urine creatinine) prior to labour market entry spend more time in the labour market in the subsequent two decades and earn more. The associations between creatine and labour market outcomes are robust to controlling for other biomarkers, educational attainment and parental background. Creatine is a naturally occurring nitrogenous organic acid which supplies energy to body cells, including muscles. Our findings are consistent with high energy levels, induced by creatine, leading to productivity-enhancing traits such as a high propensity for effort, perseverance, and high-commitment.

☆ We are grateful to the Associate Editor (Matteo Picchio), two anonymous reviewers, Stephanie von Hinke Kessler Scholder, Pekka Ilmakunnas, Antti Kauhanen, Andrew Oswald, Jari Vainiomäki and participants at the EALE and Finnish Economic Association conferences for helpful comments. The data used in this study are confidential, but other researchers can obtain access to them for replication purposes at the Research Laboratory of the Department of Standards and Methods Unit of Statistics Finland. Obtaining access to the data requires approval by the administrators of the YFS data (Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku) and by Statistics Finland.

* Corresponding author at: Turku School of Economics, Labour Institute for Economic Research, Helsinki, Finland.
E-mail addresses: petri.bockerman@labour.fi (P. Böckerman), a.bryson@ucl.ac.uk (A. Bryson), jutta.viinikainen@jyu.fi (J. Viinikainen), christian.hakulinen@helsinki.fi (C. Hakulinen), laura.pulkki-raback@helsinki.fi (L. Pulkki-Råback), olli.raitakari@utu.fi (O. Raitakari), jaakko.k.pehkonen@jyu.fi (J. Pehkonen).

http://dx.doi.org/10.1016/j.jebo.2017.08.003
0167-2681/

1. Introduction

Standard micro-economic theory assumes that individuals behave according to an economic calculus in which their utility varies with effort. Effort is thus at the heart of labour economics because it can account for a substantial proportion of the variance in individuals' labour productivity and thus their earnings. In the standard model effort is a disutility creating both incentives and monitoring problems for employers eager to raise labour productivity. Individuals' propensity for effort is rarely directly observed, although proxies are available. Test scores and qualifications are often treated as markers of ability which affects returns to effort. Effort can be observed at the intensive margin as hours worked, although these are often determined by labour demand factors beyond the worker's control. In the absence of direct measures of effort, it is commonly assumed that worker heterogeneity is fixed over time such that estimates within individuals over time can identify factors influencing the labour market outcome of interest net of differences in effort and ability.[1]

Behavioural economists directly observe effort in laboratory experiments enabling them to draw inferences about which worker types, treatments and social settings are linked to productivity and other work-related outcomes (Charness and Kuhn, 2011). Other economists have begun to explore links between biomarkers and labour market behaviours in an attempt to get inside the "black box".[2] The assumption is that individuals are "wired" differently due to differences in their genetic make-up and physical processes in ways which can help predict their success in acquiring human capital and labour market behaviours. However, knowledge about links between biomarkers and labour market behaviours has been hampered by the paucity of longitudinal data containing biomarkers and labour market outcomes. Most studies have had to rely on cross-sectional data in which labour market status is measured at roughly the same time as the biomedical measurements are taken. There are some exceptions, such as birth cohort studies (e.g. Black et al., 2007), many of which focus on height or weight at an early age and subsequent labour market outcomes. Other studies take biomarker measurements when individuals have already entered the labour market and seek to predict future outcomes. The results are often disputed, as are the hypothesised mechanisms linking these biomarkers to labour market outcomes.

We contribute to the literature by examining the association between a biomarker which is linked to individuals' energy levels and, potentially, to long-term labour market outcomes. Using the Young Finns Study (YFS) linked to the Finnish Linked Employer-Employee Data (FLEED) we show that quantities of creatine proxied by urine creatinine measured in 1980 prior to labour market entry affect labour market outcomes over the period 1990–2010. Those with higher levels of creatine in 1980 spend more time in the labour market in the subsequent two decades and earn more. The associations between creatine and labour market outcomes are robust to controlling for other biomarkers, educational attainment and parental background. Creatine is a naturally occurring nitrogenous organic acid which supplies energy to body cells, including muscles and brain. Our findings are consistent with high energy levels, induced by creatine, leading to productivity-enhancing traits – such as a high propensity for effort, perseverance, or high-commitment – which results in positive labour market outcomes.

Section Two reviews the existing literature on biomarkers and labour market outcomes, identifying how we contribute to that literature. Section Three outlines the possible mechanisms linking creatine and labour market outcomes. Section Four presents the data. Section Five describes the estimation approach and presents the results while Section Six concludes.

2. Literature

Individuals differ biologically in ways that may be linked to the acquisition of human capital, labour market behaviour and thus income for a variety of reasons. First, biological differences may affect individuals' economic preferences, tastes or predispositions. There is suggestive evidence linking circulating testosterone with risk taking, an attribute that can be rewarded in the labour market (Stanton et al., 2011). For example, Coates et al. (2009) find a biomarker proxying testosterone exposure at birth was linked to the success of financial traders, a high-risk occupation.[3]

Second, individuals differ in their innate abilities which are due, in part, to their genetic make-up. These abilities include mental ability, such as intelligence, and physical aptitudes. These innate abilities may translate into higher productivity and therefore higher earnings or better employment prospects. For example, psychometric intelligence is a good predictor of educational achievement (Deary et al., 2007), occupational level attainment and performance within one's chosen occupation (Schmidt and Hunter, 2004). Some abilities make individuals particularly suited to specific occupations. For instance, manual dexterity is sought after among machinists (Fischer and Sobkow, 1979). Biomarkers may also be predictors of severe diseases which impair individuals' productivity or labour market attachment.

Third, individuals' labour market performance may be partially determined by their subjective well-being and emotional traits. For instance, traits like emotional resilience enable some individuals to persevere in tasks such as job search that others may give up on (Moorhouse and Caltabiano, 2007). There is also evidence that creatine is negatively associated with

[1] In the personnel economics literature examining the effects of incentive pay on productivity the difference between OLS and individual fixed effects results is often interpreted as the impact of effort (e.g. Pekkarinen and Riddell, 2008).

[2] The term "biomarker", or biological marker, refers to objective indicators of medical states, which can be measured accurately and reproducibly (Strimbu and Tavel, 2010). There are recent applications of biometric information in health economics (see Lee et al., 2015). Some of these studies have considered the labour market effects (Schultz-Nielsen et al., 2016; Seuring et al., 2015).

[3] Sapienza et al. (2009) find individuals high in testosterone and low in risk aversion were more likely to choose risky careers in finance. Määttänen et al. (2013) report that total testosterone is associated with higher reward dependence which reflects the importance of social rewards to an individual.

mental disorders such as depression (Allen, 2012). Conversely, depression lowers employment probabilities, though it seems to have no impact on weekly hours worked nor hourly earnings (Peng et al., 2013). Green (2010) confirms that different aspects of subjective wellbeing have different effects on labour mobility while Hardy et al. (2003) show that job-related depression-enthusiasm is better at predicting absence than job-related anxiety-contentment. This literature leaves open the important question of the extent to which subjective wellbeing simply reflects the underlying physiological well-being captured in biomarkers.[4]

Fourth, the reactions of other labour market actors to an individual's biometric characteristics may determine labour market outcomes. For instance, physical traits affect the way in which employers respond to job applicants and employees. Audit and correspondence studies indicate traits such as skin colour, facial abnormalities and gender can affect job seekers' probability of receiving a job offer, either because of employers' taste-based preferences or statistical discrimination.[5] The wage premium for beauty is consistent with employer taste-based preferences, though this has been challenged. There are certainly other potential reasons for the beauty premium (Hamermesh, 2011).[6,7] Traits such as beauty may also affect the individual's own self-perception in ways that affect their labour market behaviour. For instance, Mobius and Rosenblat (2006) show that part of the beauty wage premium they identify in a laboratory experiment arises because physically attractive workers are more confident and higher confidence increases wages.

Fifth, individuals' educational attainment and labour market performance, such as earnings, occupational choices and employment, may also be partly determined by their personality (see e.g. Almlund et al., 2011). Recent studies have shown that personality traits may be related to biometric characteristics. For example, double-blind placebo controlled studies showed that subjects with artificially raised testosterone became less trusting (Bos et al., 2010) and less generous (Zak et al., 2009) towards strangers. This implies that biomarkers might be related to labour market performance indirectly via personality.[8]

The main difficulty interpreting results from most of the earlier studies is that the biomarkers appearing as independent variables in analyses of labour market outcomes may, in fact, be simultaneously determined by labour market experiences. This makes interpretation of results tricky. The problem is accentuated in studies using cross-sectional data or where the lag between biometric measurements and labour market outcomes is short. There may also be omitted variables problems which bias the estimates of biometric effects on labour market outcomes. For instance, the positive association between birth weight and subsequent labour market outcomes (e.g. Behrman and Rosenzweig, 2004; Black et al., 2007) may be due to birth weight proxying early life social advantages that carry on through to adulthood. This may also partially explain the wage premium associated with height (LaFeve and Thomas, 2013).[9] Alternatively, these biomarkers may also reflect healthiness, leading to less sickness later in life, and thus a better employment record which is rewarded in the labour market through higher earnings.

Some studies have begun to exploit plausibly exogeneous genetic differences across individuals in an effort to identify the causal relationship between biomarkers and labour market outcomes. For example, when Norton and Han (2008) instrument for obesity with a genetic marker that predisposes individuals to obesity they find no effect of lagged obesity on subsequent employment or wages. In principle, such an approach promises to overcome endogeneity issues provided one can rule out links between the genetic marker and labour market outcomes arising through mechanisms other than obesity. In practice, as von Hinke Kessler Scholder et al. (2010) discuss, genetic markers are not a "silver bullet" in the search for causal identification. In the first place, they are often weak instruments and thus require very large samples to identify robust effects. Second, genetic components operate as a system, rather than in isolation, making it difficult to isolate the impact of a specific genetic marker. Third, specific genetic markers can have multiple pheotypic effects (pleiotropy) leading to uncertain net outcomes.

[4] Kahneman and Krueger (2006, p. 14) suggest that 'personality variables' such as sleep are predictors of life satisfaction and affect. If so, the effects of subjective wellbeing on labour market behaviour may be biased: they might even disappear once one introduces biometric data. Alternatively, subjective wellbeing may contain information which is not contained in biomarkers. They may even be orthogonal to one another, in which case their use together may explain more variance in labour market behaviour than the use of one or the other. However, there is no clear evidence linking specific well-being biomarkers to subjective wellbeing in large samples.

[5] For a review of recent developments in research into statistical and taste-based discrimination in markets, including labour markets, see Guryan and Charles (2013).

[6] Kanazawa and Kovar (2004) argue that beautiful people are more successful in the labour market because they are more intelligent.

[7] More broadly, individuals may suffer labour market disadvantage by virtue of being treated as different from others, or because society does not accommodate their specific needs, resulting in fewer opportunities for advancement. This is often the case with respect to physical or mental "disabilities" which, if accommodated by employers and society more generally, may no longer prove to be limitations for those with the "disabilities". Such considerations have led to legislative changes requiring employers to make physical adjustments to their workplaces to ensure equal access to potential workers (and customers) with physical limitations.

[8] Mueller and Plug (2006) distinguished three possible mechanisms which might explain the connections between personality and labour market outcomes: 1) Personality contributes to productivity; 2) Personality is linked to preferences and therefore e.g. on occupational choices or attitudes towards leisure; 3) Labour market discrimination.

[9] This is contested by Persico et al. (2004) who pinpoint differences in height in adolescence as an important contributor to the height wage premium.

3. Possible mechanisms for creatine

In this paper we consider the effects of a particular biomarker – creatine – collected prior to labour market entry on labour market outcomes over the subsequent two decades. Although we can not claim to have isolated the causal impact of creatine on labour market outcomes this sequencing helps overcome the simultaneity problem referred to above. Controlling for family background and a range of other relevant biomarkers also collected prior to labour market entry helps overcome omitted variables biases which might arise if we focused on the effects of a single biomarker which might, conceivably, be proxying the effects of other biomarkers.

Creatine is a naturally occurring nitrogenous organic acid which supplies energy to body cells, including muscles and brain. There is a vast literature on the links between creatine and pathology and metabolic capacity (see Wyss and Kaddurah-Daouk, 2000, for a comprehensive review). Two of its effects – one physical, the other mental – seem particularly pertinent to labour market performance.

The first potential mechanism is creatine's ability to improve work performance, most notably "in a variety of short-term intermittent, supramaximal exercise tests" (Wyss and Kaddurah-Daouk, 2000, p. 1183). For this reason, creatine is used by athletes as a performance-enhancing supplement. In their review of creatine supplementation in sports physiology Wyss and Kaddurah-Daouk (2000, p. 1177–1182) state it is a popular supplement for athletes engaged in high-intensity exercise including "bodybuilders, wrestlers, tennis players, cyclists, mountain bikers, rowers, ski-jumpers [and] cross-country skiers as well as among ski, rugby, handball, basketball, football and ice hockey teams" (Wyss and Kaddurah-Daouk, 2000, p. 1177). This performance-enhancing effect of creatine was not known in 1980 when the YFS cohort creatine measurements were taken. Instead we measure the amount of creatine occurring naturally in the body. It seems plausible that, by reducing (increasing) the costs (returns) to effort, the ergogenic effects of creatine may lead to improvements in labour market performance.[10] Greater effort intensity may result in more successful job search,[11] longer periods in employment (for instance, via an increased propensity to supply labour, or reduced absenteeism), and faster earnings growth (as in the case of Lazear's (2000) windshield installers who increase their productivity via effort following the firm's switch to performance pay).[12]

The second mechanism by which creatine may improve labour market performance is via its impact on cognitive ability. Creatine affects brain function and, because creatine is found in large quantities in meat, fish and animal products, researchers have examined the effects of creatine on vegans and vegetarians. One such placebo-controlled double-blind experiment gave small daily doses of creatine to a treatment group for six weeks. Compared with the control group they were able to repeat longer sequences of numbers from memory and had higher overall IQ scores (Rae et al., 2003).[13,14] The results are consistent with creatine increasing the amount of energy available to the brain for computational tasks, thus improving general mental ability. The study did not test for long-term effects but, as the authors note, short-term effects may be sufficient to assist with examination work, for instance, so short-term impacts may nevertheless lead to long-term gains in the labour market and in education. Memory improvements were also found in a similar study, though only for vegetarians, not omnivores, and creatine supplementation had no impact on verbal fluency (Benton and Donohoe, 2011). Also double-blinded placebo-controlled tests for non-vegetarians have found a positive connection between creatine supplementation and cognitive performance. Watanabe et al. (2002) report that a dietary supplement of creatine reduced mental fatigue when subjects repeatedly performed a simple mathematical calculation in a stressful time-pressured test. McMorris et al. (2006, 2007a) found that following significant sleep deprivation with mild exercise, creatine supplementation improved mood and performance in cognitive tasks. Creatine supplementation can also increase backward memory span (Hammett et al., 2010) and cognitive performance of elderly people (McMorris et al., 2007b). In addition, studies based on individuals with inborn errors in creatine production and rat experiments confirm that creatine has an important role in normal brain development and cognitive functioning (for a review see Allen, 2012).

Approximately half of an individual's daily creatine requirement is produced naturally by the human body in the kidneys, liver, pancreas and possibly in brain cells and the rest of the daily requirement is replenished by dietary intake (Allen, 2012).

[10] An additional potential channel for the positive effect of creatine on labour market outcomes in adulthood is that high levels of creatine are related to physical activity when young. Leisure sport activities can in turn be associated with improvements in non-cognitive skills and formation of social capital that are eventually positively rewarded in the labour market (e.g. Lechner, 2009).

[11] In the standard search models job search intensity is endogenous with respect to the job offer arrival rate. There are increasing returns to job search intensity, though at a diminishing rate. See, for example, Chirinko (1982).

[12] Indeed, in Lazear's (2000) and Prendergast's (1999) models employers solve the principal-agent problem by eliciting greater effort when linking pay to output, as opposed to paying a fixed wage. In such models effort is costly, so factors such as high levels of creatine which reduce the costliness of effort imply the potential for higher earnings. Effort is multi-dimensional and may show up in a variety of ways in a job context.

[13] The experiment showed significant improvements for the treatment group in two tests of what is known as "fluid intelligence", namely the ability to think logically and solve problems in novel settings independently of acquired knowledge. The two tests were the Raven's Progressive Matrices and the backward digit span text from the Wechsler Adult Intelligence Scale.

[14] However, another study found no such effects in young adults (Rawson et al., 2008). It could be that creatine has positive effects on cognitive performance particularly in stressful situations. Positive effects of creatine on cognitive behaviour have been found especially in situations associated with reduced brain creatine, such as sleep deprivation, mild exercise and stressful mathematical tasks (Allen, 2012).

Because creatine is contained in animal products including meat, dietary products and fish, it may be useful to control for food intake in isolating any relationship between creatine and subsequent labour market outcomes.[15]

We wish to isolate the independent association between creatine levels measured in 1980 and the subsequent labour market and income outcomes for individuals in the following two decades. For reasons noted above we hypothesise that higher levels of creatine prior to labour market entry will result in greater labour market attachment and higher earnings due to the labour market returns to higher effort levels and improved cognitive ability. These, in turn, may result in higher income levels later in life.

4. Data

To examine the relationship between biomarkers and labour market outcomes we link data on biomarkers and parental background drawn from the Cardiovascular Young Finns Study (YFS) to the Finnish Longitudinal Employer-Employee Data (FLEED) of Statistics Finland which records periods of employment and earnings. To account for parental background, we also use information from the Longitudinal Population Census (LPC) of Statistics Finland in 1980.

The YFS began in 1980 when 4320 participants in six age cohorts (aged 3, 6, 9, 12, 15 and 18) were randomly chosen from five Finnish university regions using the national population register (Raitakari et al., 2008). A total of 3596 persons participated in the study in 1980. Seven follow-up studies have been conducted, most recently in 2011/12. The aim of the study is to examine how childhood lifestyle, biological, and psychological factors contribute to the risk of cardiovascular diseases in adulthood. These include factors such as smoking status, alcohol use, diet, physical activity, parental background and chronic diseases. The data contain information on a variety of biomarkers such as height, skinfold measures, pulse (rest heart rate), blood pressure, cholesterol, triglyceride, insulin, copper and zinc in serum, C-reactive protein, several fatty acids and urine creatinine. Data is collected using questionnaires, physical measurements and blood tests. All anthropometric measures originate from professional health examinations conducted at local health centres.

Seven biomarkers used in this study were obtained in 1980 when the participants were between 3 and 18 years old. Height (in millimeters) was measured by Seca anthropometer. Each skinfold measure is based on the average of three measurements obtained on the non-dominant arm using a Harpenden skinfold caliper (Porkka et al., 1997). We use information on triceps and subscapular skinfolds to measure body fat using the Slaughter equation (Slaughter et al., 1988).[16] YFS data also contain information on BMI (Body mass index).[17] However, BMI blurs the distinction between fat and fat-free mass such as muscle and bone (Yusuf et al., 2005; Burkhauser and Cawley, 2008). Pulse (or the count of arterial pulse per minute) is equivalent to measuring the heart rate at rest. It was measured by listening to the heart beat directly and counting it for a minute. Blood pressure was measured from the brachial artery whilst in a sitting position after 5 min rest with a standard mercury sphygmomanometer. Readings to the nearest even number of millimeters of mercury were performed 3 times on each subject. For 3-year-olds, blood pressure was measured with an ultrasound device (Juonala et al., 2005; Kivimäki et al., 2006). The mean of three measurements of systolic blood pressure measurement is used in this study.

Serum total cholesterol and triglycerides, were determined based on fasting venous blood samples after 12 h of fasting. Non-fasting subjects were excluded from the analysis (n = 38). Serum samples were stored frozen at −20 C for no more than 6 months until analysed. All lipid determinations were performed in duplicate in the same season (fall) and as simultaneously as possible and the averages of the two measurements were used to determine the level of serum lipids. Triglycerides were determined by using a fully enzymatic method (Boehringer Mannheim) (Porkka et al., 1997). Serum insulin was measured with a modification of the immunoassay method of Herbert et al. (1965) (Juonala et al., 2005).

The YFS also contains information on the amount of creatinine in the urine. Creatinine is a major breakdown product of creatine so that urine creatinine is a good proxy for creatine.[18] Our results using creatinine provide conservative estimates for creatine. Consequently, we obtain the lower bound for the true effect of creatine on long-term labour market outcomes if creatinine is subject to classical measurement error. Creatinine is removed from the body entirely by the kidneys. Creatinine is essentially the waste created from the body's use of creatine. In healthy children with normal kidney function, urine creatinine can be interpreted as an indicator that a child is extensively supplied by creatine. The amount of creatinine was measured using standard urine drug tests, based on the Jaffe method (e.g. Delanghe and Speeckaert, 2011). In analyses triglycerides and creatinine are log-transformed because of their skewed distributions.

The YFS data are linked to Statistic Finland's Finnish Longitudinal Employer-Employee Data (FLEED). We match the YFS and FLEED using unique personal identifiers. This is exact matching and there are no misreported ID codes. We therefore avoid problems created by errors in record linkages (e.g. Ridder and Moffitt, 2007). Thus, every person in the YFS data is identified in FLEED. We use FLEED to measure labour market outcomes. FLEED includes information on individuals' labour market status, and salaries and other income, taken directly from tax and other administrative registers that are collected

[15] Blanchflower et al. (2011) find links between income and some biomarkers weaken considerably when controlling for fruit and vegetable intake.

[16] For boys the Slaughter equation is 0.783 × (sum of triceps and subscapular skinfold) − 1.7. For girls it is 0.546 × (sum of triceps and subscapular skinfold) + 9.7.

[17] Body mass index is calculated as a person's weight in kilograms divided by height in metres squared.

[18] If our data contained both creatine and creatinine, then creatinine could be used as an instrument for creatine. From this perspective, our estimation results can be interpreted as reduced-form IV estimates.

and/or maintained by Statistics Finland. Thus, our income data do not suffer from underreporting or recall error, nor is it top-coded. Short-term, cross-sectional measures of income, such as yearly earnings and hourly wages, contain idiosyncratic components that diminish the precision of the estimates (cf. Dahl et al., 2011). Register-based earnings also have much less measurement error than short-term measures that often originate from surveys. This accuracy increases the efficiency of the estimates, which is particularly important for relatively small samples such as the YFS.

To measure formal human capital we use indicators for educational attainment, based on the highest obtained degree in 2010. This information originates from the comprehensive register of completed degrees by Statistics Finland.

We use FLEED to construct two income measures for an individual's early life experiences in the labour market. Our income dependent variable in the regression analyses is the logarithm of the average of annual taxable income over the period of 1990–2010. Taxable income is a broad income measure which includes annual wage and salary earnings and self-employment income. It also includes income transfers and social security benefits, such as unemployment and parental leave benefits, which often are proportional to past wage and salary earnings. Our earnings dependent variable is the logarithm of the average of annual wage and salary earnings over the period of 1990–2010. This is a much narrower measure of income than our first measure. The income measures are deflated using the consumer price index (base year 2000).

Data on employment status also originate from FLEED that contain information from Employment Statistics. Employment Statistics record the exact labour market status during the last week of each year. This information is based on the state-run pension registers that cover all legal employment contracts. Our measure for labour market attachment is the average share of employment years over the period of 1990–2010.

Thus, we begin to measure labour market outcomes in 1990 when the YFS cohort is aged between 13 and 28 years old. The final labour market outcomes are measured when we last survey them in 2010 when they are aged between 33 and 48 years old. Since the youngest cohorts are still at school in the early 1990s we use in all our models labour market outcomes measured after the age of 24 to capture the relevant outcomes after school-leaving over the observation window 1990–2010.[19]

To establish the baseline we estimate a set of models using OLS.[20] The models have the following structure:

$$Y_i = \beta_0 Creatinine_i + \beta_1 X_i + \varepsilon_i$$

where Y_i stands for the labour market outcomes for individual *i*. The parameter of interest is β_0 (i.e. the effect of creatinine on labour market outcomes). As is standard, our income and earnings outcomes are logged and are approximately normally distributed. The baseline models include a vector of control variables X_i that includes a set of indicators for gender, birth cohort and birth month[21] – all clearly predetermined variables. In additional specifications we account for parental background using the LPC and also other potentially important confounders.

Table 1 reports descriptive statistics for the variables that are used in the models. Means are presented together with standard deviations in parentheses.[22] Table 2 shows the correlation coefficients between biomarkers. All biomarker correlations were statistically significant with the exception of the correlation between pulse and body fat. The correlation coefficient between creatinine and height is particularly high at 0.56. This is due to the fact that the level of creatinine steadily increases with age. The correlations between height and creatinine are significantly smaller when calculated within birth cohort.[23] Age effects are accounted for fully in our models through indicators for birth cohort.

5. Results

5.1. Baseline estimates

We estimate the effect of biomarkers on three labour market outcomes – total income, earnings and employment – over the observation window 1990–2010 (Table 3, Columns 1–3). The explanatory variables are jointly statistically significant in all specifications that are reported in Table 3. We find that the quantity of creatinine measured in 1980 prior to labour market entry is positively related to labour market outcomes in the subsequent 10–20 years (depending on their birth cohort). This

[19] Working while studying is common among university students in Finland. An earlier working paper version reports the estimation results without the age restriction. The findings remain intact (cf. Table A1 in Supplementary material).

[20] All other models are also estimated using OLS if not otherwise stated.

[21] There is evidence that birth month is associated with family background (Buckles and Hungerman, 2013). Also in the YFS data the average family income (in 1980) seemed to be higher among those children who were born at the beginning of the year. We estimated the models also without the indicators for birth month. The creatinine result remained intact (Table A1 in Supplementary material).

[22] Descriptive statistics are based on the estimation sample used in Table 3. Because of missing information on some variables the estimation sample is smaller than the original sample of the YFS data. We tested the randomness of this attrition with a two-sample test of proportions and a two-sample *t*-test for equality of means (Tables A2 and A3 in Supplementary material). The results based on the available data indicated that average income was slightly higher in the estimation sample compared to the total sample. Furthermore, the youngest age cohort was slightly under-represented in the estimation sample (with the share of 14.5% in the estimation sample vs. 16.1% in the total sample). A discussion of attrition of the YFS data is provided in Raitakari et al. (2008).

[23] The correlations within cohorts from the youngest to the oldest were as follows: 0.15***; 0.03; 0.09**; 0.13***; 0.13*** and 0.02. "Growth spurt" (i.e. a rapid rise in height) may be related to the positive correlation between creatinine and height for those aged 12 and 15.

Table 1
Descriptive statistics.

	Mean (SD)	5th percentile	95th percentile
Labour market outcomes			
Average annual total income	21896.05 (15029.46)	5396.06	42151.39
Average annual earnings	19194.06 (12217.63)	1419.74	39819.05
Average employment years	0.779 (0.276)	0.111	1
Biomarkers			
Height (cm)	141.887 (25.571)	97.6	178
Body fat (kg)	16.828 (7.749)	6.913	30.666
Pulse	80.503 (14.413)	60	106
Diastolic blood pressure	73.656 (11.405)	52.67	90.67
Triglyceride	0.764 (0.290)	0.44	1.3
Insulin	9.566 (5.858)	2	20.5
Creatinine	8.784 (5.087)	2.75	18.83
C-reactive protein (N = 2067)	1.001 (3.025)	0.06	4.3
Other control variables			
Meat consumption (N = 943)	136.200 (111.025)	11.35	328.6
Fish consumption (N = 943)	21.845 (39.930)	0	100
Parental income	13262.13 (7562.50)	3493	25928
Educational years			
Educational attainment (share)	13.432 (2.835)	9	18
Lower or upper secondary education	0.556 (0.497)	–	–
Postsecondary non-tertiary or short cycle tertiary education	0.162 (0.368)	–	–
Unversity education	0.283 (0.450)	–	–

Notes: N = 3182 unless indicated otherwise. Biomarker information was obtained in 1980 when the participants were 3, 6, 9, 12, 15 and 18 years old. Income and employment measures are measured after age 24 over the period 1990–2010.

Table 2
Correlation coefficients between biomarkers.

	Height	Body fat	Pulse	Diastolic blood pressure	Logarithm of triglyceride	Insulin	Logarithm of creatinine
Height	1						
Body fat	0.233***	1					
Pulse	−0.525***	−0.003	1				
Diastolic blood pressure	0.471***	0.153***	−0.256***	1			
Logarithm of triglyceride	0.187***	0.246***	−0.055***	0.094***	1		
Insulin	0.575***	0.402***	−0.209***	0.284***	0.322***	1	
Logarithm of creatinine	0.559***	0.167***	−0.288***	0.206***	0.092***	0.300***	1

Notes: N = 3182. Significant at *** 1% level.

pattern applies to both income measures and to the average share of employment years over the observation window. The effect is statistically significant at the 1% level in all models.[24]

The quantitative magnitude of the effect of creatinine is non-negligible. We find that an increase in creatinine by one percent increases earnings by ∼0.16% (Table 3, Column 2). One standard deviation increase from the average level of creatinine

[24] The models contain seven biomarkers as explanatory variables. We have also estimated the models adjusting only for height and body fat (Table A1 in Supplementary material). This has only a small impact on the estimates and the conclusion for creatinine remains the same. This is not surprising because the other biomarkers are generally not statistically significant in the models.

Table 3
The baseline estimation results.

	(1) Total Income	(2) Earnings	(3) Average employment years
Height	0.0104***	0.0185***	0.0028***
	(0.0019)	(0.0040)	(0.0008)
Body fat	−0.0038	−0.0130**	−0.0010
	(0.0028)	(0.0062)	(0.0010)
Pulse	−0.0004	−0.0006	0.0001
	(0.0013)	(0.0024)	(0.0004)
Diastolic blood pressure	0.0023	0.0025	−0.0002
	(0.0017)	(0.0029)	(0.0005)
Logarithm of triglyceride	−0.0502	−0.1025	−0.0327**
	(0.0461)	(0.0821)	(0.0154)
Insulin	−0.0027	−0.0017	0.0007
	(0.0029)	(0.0061)	(0.0011)
Logarithm of creatinine	0.1077***	0.1616**	0.0298***
	(0.0364)	(0.0651)	(0.0105)
Gender	x	x	x
Cohort	x	x	x
Birth month	x	x	x
R^2	0.0400	0.0330	0.0273

Notes: N = 3182. Biomarker information was obtained in 1980. Income and employment measures are measured after age 24 over the period 1990–2010. Significant at *10%, ** 5%, and *** 1% level. Heteroskedasticity-robust standard errors are reported in parentheses.

is associated with a 7.36% increase in earnings. Because the average annual earnings in our estimation sample are ∼19,000 euros, this converts to an increase of ∼1400 euros.[25] The effect of creatinine on total income later in life is lower at ∼0.11% (Table 3, Column 1). This is in line with the view that total income includes elements of social insurance in a Nordic welfare state. For the reasons noted above, these estimates are likely to constitute the lower bound for the true effect of creatine on long-term labour market outcomes.

In contrast, the quantitative magnitude of the effect of creatinine on labour market attachment is very close to zero. An increase in the amount of creatinine by one percent increases average employment years only by ∼0.03 (Table 3, Column 3).[26] This implies that creatinine has an influence on labour market outcomes on the intensive margin of labour supply. The finding is consistent with the idea that high energy levels, induced by creatinine, lead to productivity-enhancing traits such as a high propensity for effort, perseverance, and high-commitment. It is also consistent with creatinine increasing cognitive ability and thus earnings potential.

We also find interesting effects for other biomarkers that are included as controls (Table 3). Height has a significant positive effect on all labour market outcomes later in life. Our results show that height measured in younger ages is also related to subsequent labour market outcomes, in contrast to the results in Persico et al. (2004).[27] Body fat has a negative effect on earnings[28] but its effect on total income is much smaller and it is also statistically weaker. Body fat is not a significant determinant of labour market attachment. The results also show that triglyceride (a serum lipid) is negatively related to labour market attachment.[29]

5.2. Robustness

The baseline estimates establish a positive correlation between creatinine and labour market outcomes later in life. To evaluate the robustness of these findings, we have estimated several additional specifications that exploit the richness of our linked data. We briefly discuss each of these results.

[25] The average level of creatinine in 1980 for all cohorts was 8.734 and the standard deviation within cohorts was 3.995. Therefore, a one standard deviation increase from the average level of creatinine is equivalent to a 46% change. We found in Table 3 (Column 2) that an increase in creatinine by one percent increases earnings by ∼0.16%. This implies that a 46%, or one standard deviation, increase in creatinine is associated with ∼7.36% increase in annual earnings.

[26] Because the average share of employment years using information during the last week of each year is a relatively crude measure of labour market attachment, we have also estimated models using average employment months that are available for the shorter time period 1997–2010 (Table A1 in Supplementary material). The coefficient of creatinine is 0.2871***.

[27] Persico et al. (2004) use the National Longitudinal Survey of Youth (NLSY) from 1979 and focus on white men. Their baseline specifications explain wages with height measured at the ages of 7, 11, 16 and 33. Persico et al. (2004, p. 1033) find that, among all recorded heights, only height at age 16 has an economically large and statistically significant effect on adult wages. Böckerman and Vainiomäki (2013) provide earlier evidence on the height premium for Finland.

[28] Wada and Tekin (2010) find similar results for the United States.

[29] Because there is a statistically significant positive correlation (0.25***) between body fat and triglyceride, we have also estimated the employment model without including body fat among the controls (Table A1 in Supplementary material). In this specification triglyceride has a negative coefficient that is statistically significant at the 5% level.

5.2.1. Extensions

Because we use cross-sectional information on biomarkers, it is important to check the influence of potential outliers on the estimates. For this reason, we have excluded income, earnings and creatinine observations that are outside the 1st and 99th percentiles (Table A1 in Supplementary material). This restriction of the estimation sample does not change the effect of creatinine on total income and earnings. Thus, the results are not driven by outliers. This finding is not surprising because we use a logarithmic transformation for both creatinine and our long-term average measures of labour market outcomes to suppress the influence of potential outliers in the baseline specifications.

We have also evaluated whether the estimates are sensitive to the way that creatinine is entered in the model. The amount of creatinine and height are significantly positively correlated at 0.56 (cf. Table 2). For this reason, we have estimated the models excluding height from the set of biomarkers (Table A1 in Supplementary material). The baseline models also used the log-transformation of creatinine and triglyceride. Thus, we have estimated the models without using the logarithmic transformation (Table A1 in Supplementary material). Furthermore, we have estimated the models excluding all other biomarkers except creatinine from the set of covariates (Table A4 in Supplementary material). Our creatinine results are robust to these alternative specifications.

To assess the existence of possible non-linear effects of creatinine on labour market outcomes, we added a quadratic creatinine term to the specifications but this was not statistically significant (Table A4 in Supplementary material). Therefore, a parsimonious linear specification seems to be adequate to describe the relationship between creatinine and labour market outcomes at least for the values of creatinine found in our data.

The linked data that we use contain information on the date of obtaining the highest degree. However, this information is not available for all persons (N = 2783). For this reason, we have estimated a set of additional models in which the labour market outcomes are calculated after obtaining the highest degree over the observation window 1990–2010 (Table A4 in Supplementary material). The baseline results remain largely intact. However, the effect of creatinine on earnings is smaller and statistically significant at the 11% level, while the effect on income also falls, though remains statistically significant. It is possible that individuals with lower levels of creatinine on average work fewer hours while in school. For example, they may need more time to achieve the same school grades than similar pupils with higher level of creatinine.

We have also estimated models in which the dependent variables are measured as annual earnings/income at different ages (i.e. the year in which the individual turned 25, 30, 35, 40). These results compare income generated at the same age by individuals with different creatine levels. These results show that there is a statistically significant positive relationship at age 40 (Table A4 in Supplementary material). All other yearly point estimates are positive but imprecisely estimated. This finding is consistent with the estimates based on our long-term measures, because earnings after age 35 approximate quite closely long-term life-time measures of earnings (Dahl et al., 2011). However, we prefer to use *long-term* labour market outcomes measuring the average values over the observation window to suppress the idiosyncratic variation from year to year in the outcomes that diminishes the precision of the estimates.

Business cycle conditions may have some impact on the estimated effects. Therefore, we estimated two additional models (Table A5 in Supplementary material). First, we have estimated the models using labour market outcomes for the shorter period 2000–2010, because there was a great depression in Finland during the early 1990s. This observation window additionally ensures that all sample persons are old enough to have some relevant labour market experience. The point estimate of creatinine is larger for the period 2000–2010 most likely reflecting better labour market opportunities after the downturn. Second, Finland was badly hit by the global financial crisis that started in September 2008. For this reason, we have also estimated the baseline models of Table 3 excluding the labour market outcomes from the years 2009–2010. The results remain intact in these specifications.

We run various subsample estimates to establish how heterogeneous the creatinine effects might be. The analyses of subsamples reduce the sample size that is used in the estimates substantially. Therefore, we have to treat these results cautiously.

Given occupational differences and the potential for biomarker effects to differ by gender we ran estimates separately by sex (Table A5 in Supplementary material). Using earnings and total income as the outcome variables we find that the positive effect of creatinine is statistically significant only for men. The gender difference is significant at the 5% level. The positive effect of creatinine on labour market attachment remains statistically significant for men, but is positive and non-significant for women. Weaker effects for women are plausible, because men are more strongly attached to the labour market. Women's labour supply decisions are much more affected by family and fertility choices, making it somewhat harder to capture the independent effect of creatinine in the case of women.[30] It is also possible that effort induced by creatine is more highly rewarded in typical male occupations.

We have also conducted analyses stratified by age groups estimating separate models for two age cohorts i.e. “young” (those who were aged 3–9 in 1980) and “old” (those who were aged 12–18 in 1980).[31] We find that the effect of creatinine

[30] Reflecting this, work values differ significantly by gender (e.g. Clark 1997). Men rank promotion prospects and pay more highly than do women.

[31] A fixed unit difference in urinary creatinine may not mean the same thing in a toddler, a teen going through puberty, and a post-pubertal young adult. One reason for this might be the large changes in metabolism over the developmental span.

Table 4
Estimation results with educational attainment as an additional control variable.

	(1) Total income	(2) Earnings	(3) Average employment years
Height	0.0065***	0.0113***	0.0020***
	(0.0019)	(0.0039)	(0.0008)
Body fat	−0.0029	−0.0113	−0.0007
	(0.0027)	(0.0060)	(0.0010)
Pulse	−0.0010	−0.0018	−0.0001
	(0.0013)	(0.0023)	(0.0004)
Diastolic blood pressure	0.0027	0.0032	−0.0002
	(0.0016)	(0.0028)	(0.0005)
Logarithm of triglyceride	−0.0412	−0.0855	−0.0301**
	(0.0450)	(0.0796)	(0.0152)
Insulin	−0.0012	0.0011	0.0009
	(0.0028)	(0.0060)	(0.0011)
Logarithm of creatinine	0.1022***	0.1514**	0.0289***
	(0.0357)	(0.0633)	(0.0103)
Educational attainment (reference group: lower or upper secondary education)	–	–	–
Postsecondary non-tertiary or short cycle tertiary education	0.3359***	0.6314***	0.1119***
	(0.0280)	(0.0500)	(0.0117)
University education	0.4391***	0.8090***	0.0831***
	(0.0318)	(0.0555)	(0.0107)
Gender	x	x	x
Cohort	x	x	x
Birth month	x	x	x
R^2	0.0968	0.0909	0.0553

Notes: N = 3182. Significant at *10%, ** 5%, and *** 1% level. Biomarker information was obtained in 1980. Income and employment measures are measured after age 24 over the period 1990–2010. Education is measured in 2010. Heteroskedasticity-robust standard errors are reported in parentheses.

on labour market outcomes is notably stronger for the older age cohort (Table A5 in Supplementary material).[32] A plausible explanation for this pattern is that we can measure the labour market outcomes for the older age cohort over a longer time horizon than for the younger age cohort, thus reducing idiosyncratic measurement error and supporting precision of the estimates.

We have also estimated models in which we use the years of completed education as the outcome variable (Table A6 in Supplementary material). The years of completed education are based on the official estimates by Statistics Finland to obtain a specific degree. These results indicate a positive link between creatinine and the years of education for men. The relationship is not statistically significant for women. We also run an ordered logit model with educational categorical outcomes. Again, the results indicate a positive link between creatinine and educational attainment for men (Table A7 in Supplementary material).

5.2.2. *Potential confounders*

We have estimated several additional specifications that extend the set of covariates. As a first check, we control for educational attainment measured using comprehensive register-based information in 2010 with indicators for those who have postsecondary education and those with at least some university education. These results reveal that the positive effect of creatinine on labour market outcomes over the period 1990–2010 remains intact while controlling for formal education (Table 4). The point estimates of creatinine change only slightly for all three outcomes.[33]

There is a potential omitted variable bias related to family background. It is possible that biomarkers reflect parental background while having no independent effect on subsequent labour market outcomes, raising concerns about the interpretation of the baseline estimates in Table 3.

To account for parental background, we extend our longitudinal research design and link the YFS to register-based information from the LPC in 1980 using unique personal identifiers for parents and their children. Parental background is predetermined for offspring. To account for family background we use the family's total annual income in 1980. Using the YFS we also added a health measure that captures parents' chronic conditions to the set of covariates. We incorporate three groups of chronic conditions, namely cardiovascular diseases, lung disorders and other diseases because most of the chronic conditions are rare in the population.

[32] We have also estimated the models separately for the young (cohorts 1–2) who had not entered the schooling system in the 1980s and the old (cohorts 3–6) who already had some school experience. The effect of creatinine is statistically significant for the cohorts 3–6 (Table A8 in Supplementary material).
[33] This is also true if we use years of education instead of educational attainment. The coefficients for years of education are 0.0848, 0.1583 and 0.0175 respectively for the total income, earnings and employment equations. The logarithm of creatine coefficients in these models are 0.0983, 01440 and 0.2079 respectively, all statistically significant at a 99% confidence level.

Table 5
Estimation results with parental income as an additional control variable.

	(1) Total income	(2) Earnings	(3) Average employment years
Height	0.0107***	0.0165***	0.0026***
	(0.0020)	(0.0040)	(0.0008)
Body fat	−0.0028	−0.0108*	−0.0006
	(0.0029)	(0.0064)	(0.0010)
Pulse	0.0001	−0.0005	0.0001
	(0.0014)	(0.0025)	(0.0004)
Diastolic blood pressure	0.0026	0.0020	−0.0003
	(0.0017)	(0.0029)	(0.0005)
Logarithm of triglyceride	−0.0419	−0.0824	−0.0276*
	(0.0453)	(0.0834)	(0.0160)
Insulin	−0.0046	−0.0042	0.0001
	(0.0030)	(0.0063)	(0.0012)
Logarithm of creatinine	0.0964**	0.1586**	0.0267**
	(0.0384)	(0.0683)	(0.0109)
Parental income	0.1163***	0.2186***	0.0380***
	(0.0244)	(0.0446)	(0.0085)
Gender	x	x	x
Cohort	x	x	x
Birth month	x	x	x
R^2	0.0503	0.0428	0.0356

Notes: N = 2979. Significant at *10%, ** 5%, and *** 1% level. Biomarker information was obtained in 1980. Logarithm of parental income is measured in 1980 using LPC. Income and employment measures are measured after age 24 over the period 1990–2010. Heteroskedasticity-robust standard errors are reported in parentheses.

Diet is a parental input at the ages when creatinine was measured in 1980. For our purposes, household total income is a crucial additional covariate because it is closely related to the composition of food intake (e.g. Roos et al., 1996) that is a determinant of the amount of creatine in the body. Laitinen et al. (1995) show that the main differences in terms of socio-economic status in Finland are found in the fat, vitamin D, vitamin C and fatty acid content of the diet. In contrast, differences in energy intake and in mineral density of the diet are minor.

The correlation between the logarithm of family's total income (1980) and creatinine is statistically significantly positive but small at 0.09***. Thus, only a small portion of the variance in creatinine can be related to family income.

The positive effect of creatinine on labour market outcomes later in life remains intact in all specifications that account for parental background (Table 5 and Table A8 in Supplementary material).[34] The estimates that control for family total annual income in 1980 are reported in Table 5. As expected, there is a positive intergenerational correlation in income measures that is stronger for earnings compared to total income. We have also estimated these models allowing for nonlinear effects of parental income in 1980 by adding a quadratic parental income term to the specifications. The creatinine results remain intact (Table A8 in Supplementary material).

The extensive controls for family background narrow the scope for non-causal explanations of the relationship. Therefore, the relationship between creatinine and labour market outcomes does not seem to arise because of omitted variable bias related to childhood advantages in parental investments.

We have also used individuals' own chronic conditions (measured in 1980) as additional controls (Table A8 in Supplementary material). These results should be treated with some caution, because individuals' own chronic conditions are not likely to be predetermined and they may thus capture some of the total effect of biomarkers on labour market outcomes later in life. Their addition had only minor effect on the results. Thus, the effect of creatinine on labour market outcomes remained intact.

Because fruit and vegetable consumption reflect parental investments, we have added them as covariates as well (Table A11 in Supplementary material). This reduces the sample size substantially, because this information is available only for relatively small subset of the original sample. To account for the composition of diet, we have also added the intake of meat and fish measured in 1980 as controls because meat and fish are particularly important sources of creatine. The positive effect of creatinine on labour market outcomes remains intact in these models. Encouragingly, creatinine is the only biomarker that is significant in all models along with height after controlling for the consumption of meat and fish (Table 6). The estimates are less precise because of the much smaller sample size (N = 943). Thus, we conclude that creatinine is not simply picking up diet effects.

[34] The YFS data also contain comprehensive retrospective self-reported information on parents' education, their labour market status, and income level in eight categories that are all measured in 1979. The results remain intact using these variables to account for family background (Table A8 in Supplementary material). We have also used birth weight as an additional control (Table A8 in Supplementary material). It has only a small impact on the estimate of creatinine on labour market outcomes.

Table 6
Estimation results with consumption of meat and fish as additional control variables.

	(1) Total Income	(2) Earnings	(3) Average employment years
Height	0.0054*	0.0113**	0.0022*
	(0.0032)	(0.0049)	(0.0013)
Body fat	0.0009	−0.0126	0.0006
	(0.0043)	(0.0137)	(0.0019)
Pulse	0.0002	−0.0004	0.0007
	(0.0017)	(0.0032)	(0.0007)
Diastolic blood pressure	−0.0012	−0.0018	−0.0010
	(0.0023)	(0.0036)	(0.0009)
Logarithm of triglyceride	0.0307	0.1582	−0.0119
	(0.0910)	(0.1512)	(0.0267)
Insulin	−0.0023	−0.0176	−0.0015
	(0.0048)	(0.0110)	(0.0019)
Logarithm of creatinine	0.0699*	0.1361*	0.0362**
	(0.0412)	(0.0772)	(0.0158)
Meat consumption	0.0004**	0.0008**	0.0001
	(0.0002)	(0.0003)	(0.0001)
Fish consumption	0.0000	0.0003	−0.0000
	(0.0007)	(0.0011)	(0.0002)
Gender	x	x	x
Cohort	x	x	x
Birth month	x	x	x
R^2	0.1131	0.0854	0.0765

Notes: N = 943. Biomarker information was obtained in 1980. Consumption of meat and fish was measured in 1980. Income and employment measures are measured after age 24 over the period 1990–2010. Significant at *10%, ** 5%, and *** 1% level. Heteroskedasticity-robust standard errors are reported in parentheses.

Table 7
Estimation results with C-reactive protein as an additional control variable.

	(1) Total Income	(2) Earnings	(3) Average employment years
Height	0.0069***	0.0102***	0.0023***
	(0.0018)	(0.0038)	(0.0008)
Body fat	0.0019	−0.0066	0.0009
	(0.0024)	(0.0073)	(0.0011)
Pulse	0.0004	0.0016	0.0007
	(0.0010)	(0.0021)	(0.0005)
Diastolic blood pressure	0.000	−0.0007	−0.0009*
	(0.0013)	(0.0023)	(0.0005)
Logarithm of triglyceride	−0.0527	−0.0117	−0.0113
	(0.0455)	(0.0803)	(0.0164)
Insulin	−0.0049*	−0.0064	−0.0008
	(0.0026)	(0.0068)	(0.0012)
Logarithm of creatinine	0.0719**	0.1585**	0.0339***
	(0.0293)	(0.0682)	(0.0116)
Logarithm of C-reactive protein	−0.0265*	−0.0372*	−0.0052
	(0.0140)	(0.0221)	(0.0041)
Gender	x	x	x
Cohort	x	x	x
Birth month	x	x	x
R^2	0.1128	0.0617	0.0625

Notes: N = 2067. Biomarker information was obtained in 1980. C-reactive protein is log-transformed because of its skewed distribution. Income and employment measures are measured after age 24 over the period 1990–2010. Significant at *10%, ** 5%, and *** 1% level. Heteroskedasticity-robust standard errors are reported in parentheses.

We have also extended the set of biomarkers that are used as controls, because the YFS data also contain some relevant biomarkers that were not used in the baseline models. These biomarkers are not available for the whole data that were gathered in 1980. Therefore, it is not reasonable to limit the sample size in the baseline models by including these additional biomarkers.

The most important finding that stems from the models with additional biomarkers is that the log-transformation of the amount of C-reactive protein (measured in 1980) is negatively related to earnings and total income later in life (Table 7). This pattern is reasonable, because C-reactive protein is a general marker for inflammation and infection in the body, suggesting

these effects may be health-related. The point estimate for earnings is notably larger than for total income. Therefore, market income is more sensitive to biomarkers. Interestingly, C-reactive protein is not significantly correlated with labour market attachment. We also tested the use of systolic blood pressure measured in 1980 as an explanatory variable (Table A11 in Supplementary material). It was not significant in any of the models. Importantly, the creatinine results were robust to the addition of these extra biomarkers. Even the quantitative magnitude of the effect of creatinine remains almost unchanged in Table 7.

5.2.3. Mechanisms

We argued earlier that creatine can potentially affect labour market outcomes via physical effort and mental endeavour. Both of these are possible due to the way creatine releases energy in the body. If creatine does improve cognitive function it can then have an effect on labour market outcomes either directly via cognitive ability or indirectly through the acquisition of formal educational attainment.

The fact that the baseline creatine coefficients do not change very much when we introduce educational attainment in Table 4 suggests that the positive effect of creatinine manifests itself mostly directly via cognitive ability in jobs in the labour market, or through effort, rather than through educational attainment per se. However, as noted earlier, creatinine was positively and significantly associated with educational attainment among men but not women. We therefore reran the separate income and labour market estimates for men (Table A9 in Supplementary material) and women (Table A10 in Supplementary material) incorporating educational attainment to see if the introduction of educational attainment affected the size or significance of the creatine coefficients, at least in the case of men. It did not. Among men, although the creatine coefficients were a little smaller with the introduction of educational attainment they remained large and statistically significant for income, earnings and employment. Among women, the introduction of educational attainment resulted in a very small increase in the size of the creatine coefficients, but they remained statistically non-significant for income, earnings and employment.

Even using a longitudinal research design, it is challenging to establish the exact mechanisms at play. To this end, we have used the opportunities that the linked data and the later waves of the YFS offer to provide potential directions for future studies.

First, we have estimated models in which we have added a full set of occupational indicators from FLEED at the 2-digit level as additional controls (Table A11 in Supplementary material). The aim is to detect whether the effect of creatinine also prevails within occupational groups, because it is possible that it emerges solely as a consequence of occupational sorting. We find that a non-negligible part of the total effect is plausibly attributable to occupational sorting, but there is also a substantial positive effect within occupations using total income as the outcome variable.

Second, we have estimated models in which we use occupation as an outcome variable (Table A12 in Supplementary material). The YFS data contain self-reported information on physical strenuousness of work from the year 1989. These results do not show statistically significant relationship between creatinine and physical strenuousness of work. However, information on physical strenuousness of work is only available for a subsample of the original data and self-reported information may contain systematic measurement error.

Third, we have evaluated the impact of creatinine along the earnings distribution by using a quantile regression (Table A13 in Supplementary material). This approach allows us to move beyond mean impacts and consider whether biomarker effects are constant across the earnings distribution, or whether there are larger impacts in certain parts of the distribution. We find evidence that the effect of creatinine is significantly larger at the lower part of the earnings distribution. This is consistent with the notion that creatine improves performance in lower-paid manual work.

Fourth, in the latest follow-up of the YFS (2011–12), cognitive function was evaluated with the commercially available Cambridge Neuropsychological Test Automated Battery (CANTAB®). This information is not available for all persons in the YFS (N = 1644). There is a small, but statistically significant negative correlation (-0.1036^{***}) between creatinine (1980) and overall cognitive performance later in life. This finding discounts the mechanism based on mental endeavour. Thus, there is some indirect evidence that the correlation between creatine and labour market outcomes might originate from improved physical performance at work.

6. Conclusions

Biomarker analysis is a growing and vital field of study in empirical social science. Our contribution to this emerging literature is to provide the first population-based longitudinal study about the relationship between creatine, a common amino acid, and long-term labour market outcomes later in life. To obtain information on childhood urinary creatinine we use the Young Finns Study, a prospective epidemiological study of a population sample of Finnish youth born in the 1960s and 1970s.

Linking multiple data sources we show that quantities of creatine measured in 1980 prior to labour market entry affect labour market outcomes over the period 1990–2010. In this way we avoid concerns regarding the simultaneity of creatine measurement and labour market effects.

Across several specifications including statistical adjustments for a range of potential confounders, prospective associations between urinary creatinine and adult earnings remain positive and statistically significant. We find that those with higher levels of creatine, which is proxied by urine creatinine measurements taken prior to labour market entry, spend

more time in the labour market in the subsequent two decades and earn more. They also have higher total income in adulthood, much of it derived from labour market activity at the intensive margin. Both the income and earnings effects are quantitatively substantial, the labour market attachment outcomes less so.

The effect of creatine on labour market outcomes does not arise through the acquisition of formal education. Nor does it appear to be proxying the effects of other biomarkers, nor parental background. Instead the findings are consistent with high energy levels, induced by creatine, perhaps leading to productivity-enhancing traits such as a high propensity for effort, perseverance, physical strength and high-commitment. Alternatively, creatine may also enhance cognitive ability by supplying additional energy to the brain, an ability that individuals are able to translate into better earnings and greater labour market attachment.

We can not discount the possibility that some or all of the labour market effect we attribute to creatine may, in fact, arise from the genetic determination of creatine. If parent and child creatine levels are highly positively correlated, we may simply be picking up the effects of a heritable biomarker. Thus, creatine may improve parental performance in the labour market, and the resources available for the child. Importantly, there is some earlier medical evidence according to which creatine deficiency could be heritable, to some extent. Calculations based on the small subsample (N = ∼300) of the Young Finns Study show the correlation between parent and child creatinine levels is statistically significant and in the range of 0.15–0.20 for those who are aged 9. There are some genetically determined disorders related to creatine deficiency such as X-linked creatine deficiency (Braissant et al., 2011). However, the prevalence of these disorders is low in the population and so it is unlikely that they can explain the overall correlation between creatinine and labour market outcomes that we report in the paper.

The strengths of the study are its relatively large sample, prospective design (20 years separates biomarker measurement from earnings data), and excellent earnings data. The results are suggestive of a causal linkage between creatine levels early in life and positive labour market outcomes in adulthood. However, we are unable to identify the exact causal mechanisms at play. It would be valuable if future research identified these mechanisms and detected whether the correlation reflects an accumulating developmental process caused by genes and/or environments. Whether urinary creatinine causes differences in labour market outcomes or is a marker for some other process remains an open question. Deeper knowledge of mechanisms is needed to draw policy inferences.

If one were to interpret these effects as causal, then the impact of sub-optimal creatine in the body goes beyond the simple health effects previously identified. We found a positive link (linear within our range) to labour market outcomes. This raises questions about what might be an optimal level of creatine for labour market performance. The optimal level may be heterogeneous across individuals and it is unclear from our study whether there could be adverse effects above certain levels. Because creatine supplements are readily available nowadays, their use is already widespread and demands further research.

From a broader perspective, our findings are related to the literature that analyses links between childhood factors (birth weight, height, obesity) and later life outcomes. Prenatal interventions that focus on mothers' nutritional and health standards may have positive effects in utero that improve health outcomes later in life (cf. Barker, 1997). Using information on biomarkers directly rather than as instruments may also have implications for the analysis of genetic versus environmental variations in later life outcomes. In particular, biomarkers may provide information on pre-disease pathways and health risks, and thus help to identify the role of environment for the incidence of diseases.

Funding

The Young Finns Study has been financially supported by the Academy of Finland: grants 286284, 134309 (Eye), 126925, 121584, 124282, 129378 (Salve), 117787 (Gendi), and 41071 (Skidi); the Social Insurance Institution of Finland; Competitive State Research Financing of the Expert Responsibility area of Kuopio, Tampere and Turku University Hospitals (grant X51001); Juho Vainio Foundation; Paavo Nurmi Foundation; Finnish Foundation for Cardiovascular Research; Finnish Cultural Foundation; Tampere Tuberculosis Foundation; Emil Aaltonen Foundation; Yrjö Jahnsson Foundation; Signe and Ane Gyllenberg Foundation; and Diabetes Research Foundation of Finnish Diabetes Association. The Palkansaaja Foundation supported the use of linked data. Böckerman thanks the Strategic Research Council funding for the project Work, Inequality and Public Policy (293120). Jutta Viinikainen appreciates financial support from the Yrjö Jahnsson Foundation (grant 6664) and OP Group Research Foundation. Jaakko Pehkonen acknowledges financial support from the Yrjö Jahnsson Foundation (grant 6646).

Conflict of interest

The authors declare that they have no conflict of interest.

Ethical approval

All participants of the YFS provided written informed consent, and the study was approved by local institutional review boards (ethics committees of the participating universities). Parents or guardians provided written informed consent on behalf of the under aged children enrolled in the study.

The FLEED data have been constructed for research purposes by Statistics Finland, under the ethical guidelines of the institution which comply to the national standards.

The use of linked YFS/FLEED data have been approved by Statistics Finland. The population-based data are analysed in an anonymous form using the remote access of Statistics Finland.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jebo.2017.08.003.

References

Allen, P.J., 2012. Creatine metabolism and psychiatric disorders: does creatine supplementation have a therapeutic value? Neurosci. Behav. Rev. 36, 1442–1462.
Böckerman, P., Vainiomäki, J., 2013. Stature and life-time labor market outcomes: accounting for unobserved differences. Labour Econ. 24, 86–96.
Barker, D.J.P., 1997. Maternal nutrition: fetal nutrition and diseases in later Life. Nutrition 13, 807–813.
Behrman, J.R., Rosenzweig, M.R., 2004. Returns to birthweight. Rev. Econ. Stat. 86, 586–601.
Benton, D., Donohoe, R., 2011. The influence of creatine supplementation on the cognitive functioning of vegetarians and omnivores. Br. J. Nutr. 105, 1100–1105.
Black, S.E., Devereux, P.J., Salvanes, K.G., 2007. From the cradle to the labor market? The effect of birth weight on adult outcomes. Q. J. Econ. 122, 409–439.
Blanchflower, D.G., Christakis, N.A., Oswald, A.J., 2011. An Introduction to the Structure of Biomarker Equations. mimeo.
Bos, P.A., Terburg, D., van Honk, J., 2010. Testosterone decreases trust in socially naïve humans. Proc. Natl. Acad. Sci. U. S. A. (PNAS) 107, 9991–9995.
Braissant, O., Henry, H., Béard, E., Uldry, J., 2011. Creatine deficiency syndromes and the importance of creatine synthesis in the brain. Amino Acids 40, 1315–1324.
Buckles, K., Hungerman, D.M., 2013. Season of birth and later outcomes: old questions, new answers. Rev. Econ. Stat. 95, 711–724.
Burkhauser, R.V., Cawley, J., 2008. Beyond BMI: the value of more accurate measures of fatness and obesity in social science research. J. Health Econ. 27, 519–529.
Charness, G., Kuhn, P., 2011. Lab labor: what can labor economists learn from the lab? In: Ashenfelter, O., Card, D. (Eds.), Handbook of Labor Economics, vol. 4A, pp. 229–330, Amsterdam : North Holland.
Chirinko, R.S., 1982. An empirical investigation of returns to job search. Am. Econ. Rev. 72, 498–501.
Clark, A.E., 1997. Job satisfaction and gender: why are women so happy at work? Labour Econ. 4, 341–372.
Coates, J.M., Gurnell, M., Rustichini, A., 2009. Second-to-fourth digit ratio predicts success among high-frequency financial traders. Proc. Natl. Acad. Sci. U. S. A. (PNAS) 106, 623–628.
Dahl, M., DeLeire, T., Schwabish, J.A., 2011. Estimates of year-to-year volatility in earnings and in household incomes from administrative survey, and matched data. J. Hum. Resour. 21, 750–774.
Deary, I.J., Strand, S., Smith, P., Fernandes, C., 2007. Intelligence and educational achievement. Intelligence 35, 13–21.
Delanghe, J.R., Speeckaert, M.M., 2011. Creatinine determination according to Jaffe–what does it stand for? Clin. Kidney J. 4, 83–86.
Fischer, D.G., Sobkow, J.A., 1979. Workers' estimation of ability requirements of their jobs. Percept. Mot. Skills 48, 519–531.
Green, F., 2010. Well-being: job satisfaction and labour mobility. Labour Econ. 17, 897–903.
Guryan, J., Charles, K.K., 2013. Taste-based or statistical discrimination: the economics of discrimination returns to its roots. Econ. J. 123, F417–F432.
Hamermesh, D.S., 2011. Beauty Pays: Why Attractive People Are More Successful. Princeton University Press, Princeton New Jersey.
Hammett, S.T., Wall, M.B., Edwards, T.C., Smith, A.T., 2010. Dietary supplementation of creatine monohydrate reduces the human fMRI BOLD signal. Neurosci. Lett. 479, 201–205.
Hardy, G.E., Woods, D., Wall, T.D., 2003. The impact of psychological distress on absence from work. J. Appl. Psychol. 88, 306–314.
Herbert, V., Lau, K.S., Gottlieb, C.W., et al., 1965. Coated charcoal immunoassay of insulin. J. Clin. Endocrinol. Metab. 25, 1375–1384.
Juonala, M., Järvisalo, M.J., Mäki-Torkko, N., et al., 2005. Risk factors identified in childhood and decreased carotid artery elasticity in adulthood. Circulation 112, 1486–1493.
Kanazawa, S., Kovar, J.L., 2004. Why beautiful people are more intelligent. Intelligence 32, 227–243.
Kivimäki, M., Laelor, D., Smith, G.D., et al., 2006. Early socioeconomic position and blood pressure in childhood and adulthood. Hypertension 47, 39–44.
LaFeve, D., Thomas, D., 2013. Height and Cognition at Work: Labor Market Performance in a Low Income Setting. mimeo.
Laitinen, S., Räsänen, L., Viikari, J., Åkerblom, H.K., 1995. Diet of Finnish children in relation to the family's socio-economic status. Scand. J. Public Health 23, 88–94.
Lazear, E.P., 2000. Performance pay and productivity. Am. Econ. Rev. 90, 1346–1361.
Lechner, M., 2009. Long-run labour market and health effects of individual sports activities. J. Health Econ. 28, 839–854.
Lee, J., McGovern, M.E., Bloom, D.E., et al., 2015. Education, gender, and state-level disparities in the health of older Indians: evidence from biomarker data. Econ. Hum. Biol. 19, 145–156.
Määttänen, I., Jokela, M., Hintsa, T., et al., 2013. Testosterone and temperament traits in men: longitudinal analysis. Psychoneuroendocrinology 38, 2243–2248.
McMorris, T., Harris, R.C., Swain, J., et al., 2006. Effect of creatine supplementation and sleep deprivation, with mild exercise, on cognitive and psychomotor performance, mood state, and plasma concentrations of catecholamines and cortisol. Psychopharmacology (Berl) 185, 93–103.
McMorris, T., Harris, R.C., Howard, A., et al., 2007a. Creatine supplementation sleep deprivation, cortisol, melatonin and behaviour. Physiol. Behav. 90, 21–28.
McMorris, T., Mielcarz, G., Harris, R.C., et al., 2007b. Creatine supplementation and cognitive performance in elderly individuals. Neuropsychol. Dev. Cogn. B Aging Neuropsychol. Cogn. 14, 517–528.
Mobius, M.M., Rosenblat, T.S., 2006. Why beauty matters. Am. Econ. Rev. 96, 222–235.
Moorhouse, A., Caltabiano, M.L., 2007. Resilience and unemployment: exploring risk and protective influences for the outcome variables of depression and assertive job searching. J. Employ. Couns. 44, 115–125.
Mueller, G., Plug, E., 2006. Estimating the effect of personality on male and female earnings. Ind. Labour Relat. Rev. 60, 3–22.
Norton, E.C., Han, E., 2008. Genetic information: obesity and labor market outcomes. Health Econ. 17, 1089–1104.
Pekkarinen, T., Riddell, C., 2008. Performance pay and earnings: evidence from personnel records. Ind. Labor Relat. Rev. 61, 297–319.
Peng, L., Meyerhoefer, C.D., Zuvekas, S.H., 2013. The Effects of Depression on Labour Market Outcomes, NBER Working Paper No. 19451.
Persico, N., Postlewaite, A., Silverman, D., 2004. The effect of adolescent experience on labor market outcomes: the case of height. J. Polit. Econ. 112, 1019–1053.
Porkka, K.V., Raitakari, O.T., Leino, A., et al., 1997. Trends in serum lipid levels during 1980–1992 in children and young adults. Am. J. Epidemiol. 146, 64–77.
Prendergast, C., 1999. The provision of incentives in firms. J. Econ. Lit. 37, 7–63.

Rae, C., Digney, A.L., McEwan, S.R., Bates, T.C., 2003. Oral creatine monohydrate supplementation improves brain performance: a double-blind, placebo-controlled, cross-over trial. Proc. R. Soc. B 270, 2147–2150.
Raitakari, O.T., Juonala, M., Rönnemaa, T., et al., 2008. Cohort profile: the cardiovascular risk in Young Finns Study. Int. J. Epidemiol. 37, 1220–1226.
Rawson, E.S., Lieberman, H.R., Walsh, T.M., et al., 2008. Creatine supplementation does not improve cognitive function in young adults. Physiol. Behav. 95, 130–134.
Ridder, G., Moffitt, R., 2007. The econometrics of data combination. In: Heckman, J.J., Leamer, E.E. (Eds.), Handbook of Econometrics, vol. 6B, pp. 5469–5547, Amsterdam: North Holland.
Roos, E., Prättälä, R., Lahelma, E., et al., 1996. Modern and healthy?: Socioeconomic differences in the quality of diet. Eur. J. Clin. Nutr. 50, 753–760.
Sapienza, P., Zingales, L., Maestripieri, D., 2009. Gender differences in financial risk aversion and career choices are affected by testosterone. Proc. Natl. Acad. Sci. U. S. A. (PNAS) 106, 15268–15273.
Schmidt, F.L., Hunter, J., 2004. General mental ability in the world of work: occupational attainment and job performance. J. Pers. Soc. Psychol. 86, 162–173.
Schultz-Nielsen, M.L., Tekin, E., Greve, J., 2016. Labor market effects of intrauterine exposure to nutritional deficiency: evidence from administrative data on Muslim immigrants in Denmark. Econ. Hum. Biol. 21, 196–209.
Seuring, T., Goryakin, Y., Suhrcke, M., 2015. The impact of diabetes on employment in Mexico. Econ. Hum. Biol. 18, 85–100.
Slaughter, M.H., Lohman, T.G., Boileau, R.A., et al., 1988. Skinfold equations for estimation of body fatness in children and youth. Hum. Biol. 60, 709–723.
Stanton, S.J., O'Dhaniel, A., McLaurin, R.E., et al., 2011. Low-and high-testosterone individuals exhibit decreased aversion to economic risk. Psychol. Sci. 22, 447–453.
Strimbu, K., Tavel, J.A., 2010. What are biomarkers? Current Opin. HIV AIDS 5, 463–466.
von Hinke Kessler Scholder, S., Smith, G.D., Lawlor, D.A., et al., 2010. Genetic Markers as Instrumental Variables: an Application to Child Fat Mass and Academic Achievement, CEMMAP Working Paper No. CWP03/W10.
Wada, R., Tekin, E., 2010. Body composition and wages. Econ. Hum. Biol. 8, 242–254.
Watanabe, A., Nobumasa, K., Tadafumi, K., 2002. Effects of creatine on mental fatigue and cerebral henoglobin oxygenation. Neurosci. Res. 42, 279–285.
Wyss, M., Kaddurah-Daouk, R., 2000. Creatine and creatinine metabolism. Psychol. Rev. 80, 1107–1213.
Yusuf, S., Hawken, S., Ounpuu, S., et al., 2005. Obesity and the risk of myocardial infarction in 27000 participants from 52 countries: a case-control study. Lancet 366, 1640–1649.
Zak, P.J., Kurzban, R., Ahmadi, S., et al., 2009. Testosterone administration decreases generosity in the ultimatum game. PLoS One 4 (12), e8330.

SUPPLEMENTARY ONLINE APPENDIX

Table A1. Additional estimation results.

	(1) Total Income	(2) Earnings	(3) Average employment years
Panel A: Labour market outcomes over the period 1990-2010 (not using the age restriction to those who are aged over 24) (Footnote 19)			
Logarithm of creatinine	0.1014*** (0.0291)	0.1544*** (0.0480)	0.0315*** (0.0092)
R^2	0.1584	0.0869	0.2030
N	3201	3201	3201
Panel B: Not controlling for birth month (Footnote 21)			
Logarithm of creatinine	0.1100*** (0.0363)	0.1603** (0.0639)	0.0284*** (0.0104)
R^2	0.0376	0.0313	0.0235
N	3182	3182	3182
Panel C: Biomarkers: height, body fat, creatinine (Footnote 24)			
Logarithm of creatinine	0.1070*** (0.0357)	0.1617** (0.0647)	0.0308*** (0.0105)
R^2	0.0386	0.0322	0.0258
N	3182	3182	3182
Panel D: Employment months 1997-2000 as outcome (Footnote 26)			
Logarithm of creatinine	..	..	0.2871*** (0.1031)
R^2	..	..	0.0354
N	..	..	3138
Panel E: Body fat excluded from the controls (Footnote 29)			
Logarithm of triglyceride	-0.0572 (0.0455)	-0.1260 (0.0807)	-0.0344** (0.0153)
R^2	0.0394	0.0310	0.0270
N	3182	3182	3182
Panel F: Excluding income, earnings and logarithm of the creatinine observations outside P1 and P99			
Logarithm of creatinine	0.0411* (0.0231)	0.1602*** (0.0538)	0.0242** (0.0111)
R^2	0.0543	0.0377	0.0256
N	3056	3049	3116
Panel G: Height excluded from the controls			
Logarithm of the creatinine	0.1211*** (0.0367)	0.1853*** (0.0663)	0.0334*** (0.0106)
R^2	0.0336	0.0269	0.0231
N	3182	3182	3182
Panel H: Entering creatinine and triglyceride without log transformation			
Creatinine	0.0115*** (0.0031)	0.0169*** (0.0061)	0.0036*** (0.0011)
R^2	0.0393	0.0324	0.0270
N	3182	3182	3182

Notes: Biomarker information was obtained in 1980. Income and employment measures are measured after age 24 over the period 1990-2010. The exception is the model in Panel A that does not implement the age restriction. Significant at *10%, ** 5%, and *** 1% level. Heteroskedasticity-robust standard errors are reported in parentheses.

Table A2. Attrition (two sample test of proportions).

	Mean (Total sample)	Mean (Estimation sample)	Difference	z-statistics
Female	0.4912	0.5000	-0.0088	-0.7228
High education	0.2748	0.2825	-0.0077	-0.7064
Cohort 1977	0.1605	0.1449	0.0156	1.7768*
Cohort 1974	0.1613	0.1625	-0.0012	-0.1301
Cohort 1971	0.1798	0.1838	-0.0041	-0.4350
Cohort 1968	0.1812	0.1873	-0.0061	-0.6508
Cohort 1965	0.1680	0.1713	-0.0033	-0.3563
Cohort 1962	0.1493	0.1502	-0.0009	-0.1073
Birth month: Jan	0.1012	0.1006	0.0006	0.0867
Birth month: Feb	0.0819	0.0833	-0.0014	-0.2041
Birth month: Mar	0.0951	0.0930	0.0020	0.2851
Birth month: Apr	0.0959	0.0933	0.0026	0.3578
Birth month: May	0.0934	0.0921	0.0013	0.1830
Birth month: June	0.0746	0.0757	-0.0011	-0.1704
Birth month: July	0.0822	0.0814	0.0008	0.1192
Birth month: Aug	0.0788	0.0795	-0.0007	-0.1023
Birth month: Sep	0.0752	0.0776	-0.0024	-0.3742
Birth month: Oct	0.0780	0.0786	-0.0006	-0.0869
Birth month: Nov	0.0677	0.0676	0.0001	0.0142
Birth month: Dec	0.0760	0.0773	-0.0013	-0.1957
N	3577	3182		

Notes: Significant at *10% level.

Table A3. Attrition (two-sample t-test).

	Mean (Total sample)	Mean (Estimation sample)	Difference	t-statistics
Log of total income[a]	9.6316	9.7926	-0.1611	-5.6083***
Log of earnings[a]	9.2839	9.4531	-0.1692	-4.0178***
Years of employment[b]	0.7775	0.7785	-0.0010	-0.1527
Height[c]	141.294	141.887	-0.5930	-0.9433
Body fat[d]	16.7331	16.8277	-0.0946	-0.4999
Pulse[e]	80.7865	80.5031	0.2834	0.7964
Diastolic blood pressure[f]	73.6999	73.6561	0.0437	0.1554
Log of triglyceride[g]	-0.3208	-0.3276	0.0067	0.8260
Insulin[h]	9.5771	9.5657	0.0114	0.0787
Log of creatinine[i]	1.9928	2.0096	-0.0167	-1.1583
Log of C-reactive protein[j]	-1.2270	-1.2465	0.0195	0.4861

Notes: Significant at * 10%, ** 5%, *** 1 % level.

[a] Total sample N = 3577; Estimation sample N = 3182
[b] Total sample N = 3520; Estimation sample N = 3182
[c] Total sample N = 3554; Estimation sample N = 3182
[d] Total sample N = 3554; Estimation sample N = 3182
[e] Total sample N = 3527; Estimation sample N = 3182
[f] Total sample N = 3429; Estimation sample N = 3182
[g] Total sample N = 3535; Estimation sample N = 3182
[h] Total sample N = 3516; Estimation sample N = 3182
[I] Total sample N = 3443; Estimation sample N = 3182
[j] Total sample N = 2245; Estimation sample N = 2067

Table A4. Additional estimation results.

	(1) Total Income	(2) Earnings	(3) Average employment years
Panel A: All biomarkers except the log of creatinine excluded			
Logarithm of creatinine	0.1195*** (0.0361)	0.1848*** (0.0661)	0.0343*** (0.0106)
R^2	0.0316	0.0240	0.0211
N	3182	3182	3182
Panel B: Quadratic term of creatinine included			
Logarithm of creatinine	0.1798 (0.2300)	0.2862 (0.3224)	0.0157 (0.0480)
Logarithm of creatinine squared	-0.0190 (0.0542)	-0.0328 (0.0760)	0.0037 (0.0117)
R^2	0.0401	0.0331	0.0273
N	3182	3182	3182
Panel C: Labour market outcomes after obtaining the highest degree			
Logarithm of creatinine	0.0633** (0.0308)	0.0748 (0.0459)	0.0259** (0.0102)
R^2	0.1023	0.0640	0.0611
N	2783	2783	2783
Panel D: Labour market outcomes at age 25			
Logarithm of creatinine	0.0907 (0.0810)	0.2946** (0.1384)	0.0294 (0.0196)
R^2	0.0205	0.0479	0.0427
N	2686	2686	2686
Panel E: Labour market outcomes at age 30			
Logarithm of creatinine	0.0337 (0.0598)	0.1797 (0.1251)	0.0070 (0.0161)
R^2	0.0156	0.0316	0.0149
N	3134	3134	3134
Panel F: Labour market outcomes at age 35			
Logarithm of creatinine	0.0440 (0.0689)	0.0639 (0.1289)	0.0260* (0.0158)
R^2	0.0173	0.0248	0.0121
N	2667	2667	2667
Panel G: Labour market outcomes at age 40			
Logarithm of creatinine	0.1855** (0.0906)	0.4650*** (0.1565)	0.0518*** (0.0184)
R^2	0.0260	0.0343	0.0258
N	1577	1577	1577

Notes: Biomarker information was obtained in 1980. Income and employment measures are measured after age 24 over the period 1990-2010 in Panels A-C. Significant at *10%, ** 5%, and *** 1% level. Heteroskedasticity-robust standard errors are reported in parentheses.

Table A5. Additional estimation results.

	(1) Total Income	(2) Earnings	(3) Average employment years
Panel A: Labour market outcomes over the period 2000-2010 (if at least 25 years old)			
Logarithm of creatinine	0.1091** (0.0435)	0.2034** (0.0832)	0.0308*** (0.0111)
R^2	0.0380	0.0258	0.0301
N	3156	3156	3156
Panel B: Years 2009 and 2010 excluded from the calculation of labour market outcomes			
Logarithm of creatinine	0.0980*** (0.0358)	0.1597** (0.0656)	0.0286*** (0.0106)
R^2	0.0447	0.0359	0.0298
N	3180	3180	3180
Panel C: Female sample			
Logarithm of creatinine	0.0491 (0.0433)	0.0154 (0.0830)	0.0136 (0.0148)
R^2	0.0281	0.0344	0.0343
N	1591	1591	1591
Panel D: Male sample			
Logarithm of creatinine	0.1555*** (0.0586)	0.2954*** (0.1003)	0.0440*** (0.0149)
R^2	0.0406	0.0409	0.0296
N	1591	1591	1591
Panel E: Young sample (aged 3-9 in 1980)			
Logarithm of creatinine	0.0834 (0.0570)	0.0610 (0.1085)	0.0136 (0.0163)
R^2	0.0367	0.0287	0.0197
N	1563	1563	1563
Panel F: Old sample (aged 12-18 in 1980)			
Logarithm of creatinine	0.1309*** (0.0473)	0.2452*** (0.0766)	0.0422*** (0.0134)
R^2	0.0644	0.0555	0.0465
N	1619	1619	1619

Notes: Biomarker information was obtained in 1980. Income and employment measures are measured after age 24 over the period 1990-2010. Significant at *10%, ** 5%, and *** 1% level. Heteroskedasticity-robust standard errors are reported in parentheses.

Table A6. Years of education in 2010 as the dependent variable.

Panel A: Total sample	
Logarithm of creatinine	0.1111 (0.1058)
R^2	0.0622
N	3182
Panel B: Female sample	
Logarithm of creatinine	-0.1644 (0.1478)
R^2	0.0323
N	1591
Panel C: Male sample	
Logarithm of creatinine	0.3721** (0.1506)
R^2	0.0496
N	1591

Notes: Biomarker information was obtained in 1980. Significant at *10%, ** 5%, and *** 1% level. Heteroskedasticity-robust standard errors are reported in parentheses.

Table A7. Education (3 categories) in 2010 as the dependent variable (ordered logit).

Panel A: Total sample	
Logarithm of creatinine	0.0638 (0.0757)
Pseudo R^2	0.0305
N	3182
Panel B: Female sample	
Logarithm of creatinine	-0.0823 (0.1000)
Pseudo R^2	0.0175
N	1591
Panel C: Male sample	
Logarithm of creatinine	0.2360** (0.1175)
Pseudo R^2	0.0274
N	1591

Notes: The dependent variable had three categories: 1) 12 years or education or less; 2) 13-15 years of education; 3) 16 years of education or more. Models were estimated using ordered logit. Biomarker information was obtained in 1980. Significant at *10%, ** 5%, and *** 1% level. Heteroskedasticity-robust standard errors are reported in parentheses.

Table A8. Additional estimation results.

	(1) Total Income	(2) Earnings	(3) Average employment years
Panel A: Young (cohorts 1-2) (Footnote 32)			
Logarithm of creatinine	0.1140 (0.0808)	0.0271 (0.1482)	0.0074 (0.0207)
R^2	0.0426	0.0399	0.0252
N	978	978	978
Panel B: Old (cohorts 3-6) (Footnote 32)			
Logarithm of creatinine	0.1050*** (0.0387)	0.2161*** (0.0661)	0.0390*** (0.0119)
R^2	0.0543	0.0441	0.0403
N	2204	2204	2204
Panel C: Parents' chronic conditions as controls			
Logarithm of creatinine	0.1210*** (0.0406)	0.1805** (0.0735)	0.0350*** (0.0115)
R^2	0.0469	0.0382	0.0317
N	2670	2670	2670
Panel D: Self reported information on parents' education, labour market status and income level (Footnote 34)			
Logarithm of creatinine	0.1093*** (0.0418)	0.1460** (0.0725)	0.0265** (0.0113)
R^2	0.0520	0.0518	0.0479
N	2678	2678	2678
Panel E: Birth weight as a control (Footnote 34)			
Logarithm of creatinine	0.1141*** (0.0429)	0.1589** (0.0754)	0.0324*** (0.0115)
R^2	0.0416	0.0368	0.0317
N	2651	2651	2651
Panel F: Non-linear effects of parental income			
Logarithm of creatinine	0.0945** (0.0387)	0.1560** (0.0687)	0.0265** (0.0109)
R^2	0.0508	0.0430	0.0356
N	2979	2979	2979
Panel G: Person's own chronic conditions as controls			
Logarithm of creatinine	0.1051*** (0.0359)	0.1524** (0.0641)	0.0297*** (0.0105)
R^2	0.0417	0.0359	0.0278
N	3165	3165	3165

Notes: Biomarker information was obtained in 1980. Income and employment measures are measured after age 24 over the period 1990-2010. Significant at *10%, ** 5%, and *** 1% level. Heteroskedasticity-robust standard errors are reported in parentheses.

Table A9. Estimation results with educational attainment as an additional control variable (males only).

	(1) Total income	(2) Earnings	(3) Average employment years
Height	0.0118*** (0.0039)	0.0220*** (0.0062)	0.0036*** (0.0011)
Body fat	-0.0073* (0.0040)	-0.0157* (0.0084)	-0.0022* (0.0013)
Pulse	-0.0021 (0.0023)	-0.0044 (0.0037)	-0.0002 (0.0006)
Diastolic blood pressure	0.0040 (0.0028)	0.0054 (0.0047)	0.0001 (0.0008)
Logarithm of triglyceride	-0.0384 (0.0611)	-0.0185 (0.1149)	-0.0180 (0.0212)
Insulin	-0.0063 (0.0051)	-0.0123 (0.0104)	-0.0002 (0.0017)
Logarithm of creatinine	0.1298** (0.0562)	0.2535*** (0.0972)	0.0389*** (0.0147)
Educational attainment (reference group: lower or upper secondary education)	..	..	..
Postsecondary non-tertiary or short cycle tertiary education	0.3847*** (0.0491)	0.6281*** (0.0763)	0.1026*** (0.0177)
University education	0.5088*** (0.0471)	0.8280*** (0.0738)	0.0771*** (0.0149)
Cohort	x	x	x
Birth month	x	x	x
R^2	0.0991	0.0918	0.0500

Notes: N=1591. Significant at *10%, ** 5%, and *** 1% level. Biomarker information was obtained in 1980. Income and employment measures are measured after age 24 over the period 1990-2010. Education is measured in 2010. Heteroskedasticity-robust standard errors are reported in parentheses.

Table A10. Estimation results with educational attainment as an additional control variable (females only).

	(1) Total income	(2) Earnings	(3) Average employment years
Height	0.0019 (0.0026)	0.0068 (0.0053)	0.0020* (0.0012)
Body fat	0.0036 (0.0032)	-0.0089 (0.0078)	0.0011 (0.0015)
Pulse	-0.0002 (0.0012)	0.0012 (0.0029)	0.0001 (0.0006)
Diastolic blood pressure	0.0015 (0.0017)	0.0017 (0.0039)	-0.0004 (0.0007)
Logarithm of triglyceride	-0.0265 (0.0685)	-0.1090 (0.1087)	-0.0343 (0.0217)
Insulin	0.0026 (0.0032)	0.0105 (0.0067)	0.0014 (0.0014)
Logarithm of creatinine	0.0590 (0.0433)	0.0365 (0.0809)	0.0176 (0.0146)
Educational attainment (reference group: lower or upper secondary education)	..	..	..
Postsecondary non-tertiary or short cycle tertiary education	0.2836*** (0.0338)	0.6022*** (0.0678)	0.1164*** (0.0158)
University education	0.3704*** (0.0427)	0.7863*** (0.0828)	0.0875*** (.0153)
Cohort	x	x	x
Birth month	x	x	x
R^2	0.0849	0.0994	0.0693

Notes: N=1591. Significant at *10%, ** 5%, and *** 1% level. Biomarker information was obtained in 1980. Income and employment measures are measured after age 24 over the period 1990-2010. Education is measured in 2010. Heteroskedasticity-robust standard errors are reported in parentheses.

Table A11. Additional estimation results.

	(1) Total Income	(2) Earnings	(3) Average employment years
Panel A: Fruit and vegetable consumption as control			
Logarithm of creatinine	0.1631*** (0.0577)	0.2670*** (0.0834)	0.0566*** (0.0145)
R^2	0.0444	0.0478	0.0483
N	1554	1554	1554
Panel B: Systolic blood pressure as control			
Logarithm of creatinine	0.1082*** (0.0365)	0.1620** (0.0651)	0.0299*** (0.0105)
Systolic blood pressure	0.0038** (0.0018)	0.0038 (0.0029)	0.0007 (0.0005)
R^2	0.0417	0.0335	0.0279
N	3182	3182	3182
Panel C: Full set of 2-digit occupational indicators as controls			
Logarithm of creatinine	0.0690** (0.0307)	0.0761 (0.0541)	0.0108 (0.0077)
R^2	0.3324	0.3468	0.4508
N	3182	3182	3182

Notes: Biomarker information was obtained in 1980. Income and employment measures are measured after age 24 over the period 1990-2010. Significant at *10%, ** 5%, and *** 1% level. Heteroskedasticity-robust standard errors are reported in parentheses.

Table A12. Occupation choice as the outcome variable (1 = physically demanding job; 0 = otherwise).

Logarithm of creatinine	-0.0049 (0.0178)
R^2	0.1135
N	1566

Notes: Biomarker information was obtained in 1980. Self-reported information on physical strenuousness of work originates from the year 1989. Significant at *10%, ** 5%, and *** 1% level. Heteroskedasticity-robust standard errors are reported in parentheses.

Table A13. Quantile regression results (dependent variable: log of earnings after age 24 over the period 1990-2010).

Quantile: 0.25	
Logarithm of creatinine	0.1042* (0.0539)
Pseudo R^2	0.0239
N	3182
Quantile: 0.50	
Logarithm of creatinine	0.0441** (0.0216)
Pseudo R^2	0.0431
N	3182
Quantile: 0.75	
Logarithm of creatinine	0.0471** (0.0190)
Pseudo R^2	0.0601
N	3182

Notes: Biomarker information was obtained in 1980. Earnings are measured after age 24 over the period 1990-2010. Significant at *10%, ** 5%, and *** 1% level. Heteroskedasticity-robust standard errors are reported in parentheses.