Biomarkers as Precursors of Disability

Apostolos Davillas
Institute for Social and Economic Research
University of Essex

Stephen Pudney
School of Health and Related Research
University of Sheffield

INSTITUTE FOR SOCIAL
& ECONOMIC RESEARCH

E·S·R·C
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL

University of Essex

Non-Technical Summary

Recent advances in the large social-science surveys involve the collection of physical measurements and markers derived from biological samples, in addition to the conventional self-reported health assessments. This information imposes higher collection costs compared to conventional self-reported questionnaires. Although the contribution of biomarkers in medical epidemiology has been established, their potential contribution to economics and social science needs more research. In this paper, we focused on disability, which results from biological processes but is defined in terms of its implications for social functioning and well-being. To our knowledge, this is the first comprehensive study to examine the predictive power of a wide range of biomarkers – over and above the contribution of the conventional self-assessed health (SAH) – for future disability risks.

Applying nationally-representative, longitudinal UK data (*Understanding Society*), we estimated predictive models for disability 2-4 years ahead of the baseline, using a wide range of biomarkers in addition to SAH and covariates. For estimation purposes, we followed individuals who reported no disability at baseline. To explore the robustness of our results, we employed alternative disability measures, ranging from the number of functional difficulties reported to receipt of disability benefits.

In subsequent analysis, we further contribute to the literature by developing and applying –for the first time in the relevant literature– a random-effects two-dimensional latent variable (LV) approach. This LV model has a number of advantages. It explicitly takes into account the longitudinal nature of our data on disability, and it aims to address the problem of measurement error bias by allowing for measurement noise in both SAH and the biomarkers. This model may help to identify any distinct dimension of health which influence future disability and is captured by the biomarkers but missed by SAH.

We found a quantitatively and statistically significant predictive role for a large set of nurse-collected and blood-based biomarkers, over and above the strong predictive power of SAH. For example, using a biomarkers summary measure (allostatic load), we found that moving from the best (excellent) to worst (fair/poor) category of SAH increased the risk of disability 4 years later by 5-18%, depending on the disability measure, while a relevant 3-standard deviation increase in allostatic load increased disability risk by 2-7%. The results from our LV models showed that although SAH has performed well, it was far from perfect as a leading indicator for disability risks; it tended to under-emphasize grip strength, lung and liver function and stress-related steroid hormone levels, while over-emphasizing adiposity, hypertension, and cholesterol levels. These results highlight the presence of distortions in SAH in the sense of dimensions of health that are given too much or too little weight by SAH in predicting disability.

Biomarkers as precursors of disability

Apostolos Davillas

Institute for Social and Economic Research, University of Essex

Stephen Pudney

School of Health and Related Research, University of Sheffield

This version: March 27, 2018

Abstract: Some social surveys now collect physical measurements and markers derived from biological samples, in addition to self-reported health assessments. This information is expensive to collect; its value in medical epidemiology has been clearly established, but its potential contribution to social science research is less certain. We focused on disability, which results from biological processes but is defined in terms of its implications for social functioning and wellbeing. Using data from waves 2 and 3 of the UK *Understanding Society* panel survey as our baseline, we estimated predictive models for disability 2-4 years ahead, using a wide range of biomarkers in addition to self-assessed health (SAH) and other socio-economic covariates. We found a quantitatively and statistically significant predictive role for a large set of nurse-collected and blood-based biomarkers, over and above the strong predictive power of self-assessed health. We also applied a latent variable model accounting for the longitudinal nature of observed disability outcomes and measurement error in in SAH and biomarkers. Although SAH performed well as a summary measure, it has shortcomings as a leading indicator of disability, since we found it to be biased in the sense of over- or under-sensitivity to certain biological pathways.

Keywords: Biomarkers; Disability; Prediction; Self-assessed health; Understanding Society

JEL codes: C2, C8, I10

Contact: Steve Pudney, ScHARR, University of Sheffield, 30 Regent Street, S1 4DA, UK; tel. +44(0)114 222 9187; email: steve.pudney@sheffield.ac.uk

1 Introduction

An important recent development in research based on large-scale social surveys is the integration of physical health measurements and markers derived from biological samples, in addition to traditional self-reported health assessments. Biomarkers are objectively measured and evaluated as indicators of normal biological or pathogenic processes (Colburn et al., 2001), and they have potential advantages over self-assessments as early indicators of conditions that are below clinical diagnostic thresholds, or are pre-symptomatic and below individuals' threshold of perception (Colburn et al., 2001). Cardiovascular, metabolic, inflammatory, neuroendocrine and other biomarkers have been shown to be predictors of mortality when used alone or alongside self-reported health assessments (Idler and Benyamini, 1997; Gruenewald et al., 2006; Ridker, 2007; Jylhä, 2009). They have also been used to reveal the socioeconomic gradient in health risks (Seeman et al., 2004; Carrieri and Jones, 2017).

Despite their advantages, biomarkers impose significant additional costs of collection in the survey context and their potential contribution to economic and social research is not entirely clear. The wider social impacts of ill-health – on quality of life, personal and social functioning, and social costs of disease – depend critically on the duration and severity of disability prior to death, and there has been little research on the role of biomarkers in relation to disability. Disability is associated with loss of employment and serious consequences for the families affected (Jones, 2016; Pudney et al., 2011) and typically implies long-lasting impairments that may prevent independent living and generate large social costs. This is particularly so in the UK where disability prevalence is well above the European Union average (Jones, 2016) and has been rising rapidly (from 11.9 to 13.3 million over 2013/14-2015/16 (DWP, 2017). There is evidence of an increasing birth-cohort trend in functional difficulties for older individuals of low socio-economic status (Morciano et al., 2015) and developed countries like the UK may face severe problems in supporting the projected future

growth in the disabled population and providing public support to people with care needs (Commission on Funding of Care and Support, 2011). A crucial question for researchers and policymakers is whether the demand for care services will be curbed by gains in disability-free life expectancy alongside the projected continuing gains in longevity. An answer to this question requires a better understanding of the processes leading to disability, allowing the development of strategies and screening programmes to address disability more efficiently. The availability of biomarker information in population-representative data may contribute to that better understanding.

Despite the importance of disability trends for social policy planning, relatively little is known about the association between biomarkers and future disability. The World Health Organization (WHO) proposed a framework that portrays progression from diseases to functional disabilities (WHO, 1980), and Fried et al. (1991) hypothesized the existence of preclinical disability as an intermediate stage in which health impairments have an impact on general functioning. Few studies have explored the predictive role of biomarkers in relation to this disability process, and most are limited by being based on small samples or unrepresentative data (Brex et al., 2002; Reuben et al., 1999; Baylis et al., 2013; Seeman et al., 1994; Kallaur et al., 2017); or focused exclusively on older individuals (Reuben et al., 1999; Seeman et al., 1994; Baylis et al., 2013); or concerned with disability outcomes from a specific disease or condition (Brex et al., 2002; Kallaur et al., 2017).

We examined the predictive power of a wide range of biomarkers for future disability and specifically asked whether biomarkers offer incremental value in predicting disability outcomes beyond the contribution of the conventional self-assessed health (SAH) measure. SAH may be associated with disability outcomes in parallel with biomarkers by reflecting the impact on the individual of diagnosis of health conditions defined in terms of elevated biomarkers (Idler et al., 2004; Jylhä, 2009), or through bodily sensations that are sensitive to the biochemical processes measured by biomarkers (Jylhä, 2009). Besides confirming the value of biomarkers as leading indicators of disability, we also investigated the success of SAH

as an overall summary of biomedical states relevant to disability and examined whether SAH is biased in the sense of over- or under-sensitivity to specific biological pathways.

The paper makes several new contributions to the research literature. To the best of our knowledge, this is the first study that provides a comprehensive analysis of this kind. Using baseline data from early waves of the UK Household Longitudinal Study (UKHLS, also known as *Understanding Society*), we estimate predictive models for disability two to four years ahead, exploiting a large set of nurse-collected and blood-based biomarkers. These biomarkers measure adiposity, grip strength, blood pressure, lung, kidney and liver functions, inflammation, steroid hormone levels, blood sugar and anaemia, giving an unusually broad picture of individuals health states. The use of several alternative disability measures demonstrates the robustness of our results.

In addition to simple prediction models, we also develop a latent variable (LV) approach which is new to the literature. This LV model has a number of advantages. First, unlike simpler prediction models, it takes into account the longitudinal nature of our data on disability, allowing for correlation between disability variables two to four years after baseline. Second, it addresses measurement error bias by allowing for measurement noise in both SAH (Crossley and Kennedy, 2002) and the biomarkers (Zang et al., 2015). Measurement error in biomarkers normally causes attenuation of the estimated impact of the biological pathways on disability, and the LV model is expected to give more accurate estimates of these impacts. This may be of particular importance for developing policy strategies and interventions to reduce the personal and social burden of disability. Third, our LV approach is set up to identify any distinct dimension of health which influences future disability and is captured by the biomarkers but missed by SAH. The pattern of factor loadings tells us what underlying aspects of health tend to be under- or over-represented by the SAH measure and can therefore guide the interpretation of research findings related to SAH.

2 Data

The UKHLS is a large, nationally representative panel survey, running continuously from the initial wave in 2009-10, with each panel member interviewed annually. Its predecessor, the British Household Panel Survey (BHPS) was incorporated into the UKHLS from wave 2. A set of physical health measures and non-fasted blood samples were collected by nurses, five months on average after the wave 2 interview for UKHLS respondents and similarly at wave 3 for the BHPS sample (McFall et al., 2014). Respondents were eligible for nurse visits if, at the relevant wave, they were aged 16 or over, lived in England, Wales or Scotland and were not pregnant. Blood sample collections were further restricted to those who had no clotting or bleeding disorders and no history of fits. Participants gave informed written consent for their blood to be taken and stored for future scientific analysis. The UKHLS has been approved by the University of Essex Ethics Committee and the nurse data collection by the National Research Ethics Service (10/H0604/2).

We define wave 2 as the baseline for the main UKHLS panel and wave 3 as the baseline for the BHPS sub-panel and we refer to the timing of this baseline observation as $t = 0$; these baseline observations were spread through calendar years 2010-2013. We used baseline data on personal and household characteristics and bio-medical measures as predictors of disability observed subsequently at waves 4-6 for the UKHLS sample ($t = 2, \ldots 4$) or waves 5-6 for the BHPS sample ($t = 2, 3$), where t denotes years since the baseline main interview. We did not use data from $t = 1$ since the time gap between collection of biomarkers and interview at $t = 1$ was less than 6 months for 75% of the $t = 1$ sample.

There were 15,632 and 5,053 UKHLS and BHPS respondents who participated in the wave 2 or wave 3 nurse visits. For the UKHLS group, 13,404, 12,719 and 11,434 were followed up at waves 4, 5 and 6 respectively and had non-missing information on disability; 4,513 and 4,113 of the BHPS subsample were followed up at waves 5 and 6. We further conditioned the analysis on the absence of reported disability at baseline by excluding those

4

reporting the relevant type of disability at baseline. Exclusion of these cases reduced the potential samples by 25%, 13% and 8% depending on the disability concept used. A detailed summary of available sample sizes is given in the Supplement Table S1.

2.1 Biomarkers

Measures of adiposity, grip strength, heart rate, blood pressure and lung function were collected during visits by trained nurses. We used the waist-to-height ratio (WHR) to measure adiposity. Grip strength was measured (in kg) using a hand dynamometer (McFall et al., 2014) and we took the highest reading from three repeated measurements for the dominant hand. Higher levels of grip strength are indicative of better physical functioning.

Three repeated measurements of resting heart rate (HR), systolic and diastolic blood pressure (SBP, DBP) were taken at intervals of one minute (McFall et al., 2014). We skipped the first reading, believed to impose upward biases, and computed HR, SBP and DBP as the average of the second and third readings. HR, which is sometimes regarded as a measure of fitness rather than health, was used as a continuous variable, and we also used a binary hypertension indicator recording SBP > 140, DBP > 90 and/or current use of anti-hypertensive medications (Johnston et al., 2009).

Lung function, assessed using spirometry equipment, was measured by the total amount of air forcibly blown out after a full inspiration (forced vital capacity; FVC), higher values indicating better lung function (Gray et al., 2013). Forced expiratory volume in one second (FEV1) is often used as an alternative to FVC. However, different equipment and measurement protocols were used in Scotland than in England and Wales, and comparison of matched samples showed FEV1 to be seriously affected by this, while FVC measures appeared comparable. Consequently we retained the Scottish sample and used FVC as our lung function measure.

We used blood-based biomarkers specific to inflammation, steroid hormone, cholesterol,

blood sugar, kidney function, liver function and anaemia.

C-reactive protein (CRP) rises as part of the immune response to infection and is associated with general chronic or systemic inflammation. We excluded those with a CRP over 10 mg/L, because those values may reflect current transient infections rather than chronic processes (Pearson et al., 2003).

Dihydroepiandrosterone suphate (DHEAS) is the most common steroid hormone in the body, considered as one of the primary mechanisms through which psychosocial stressors may affect individual health. Low levels of DHEAS are associated with cardiovascular (CVD) risks and all-cause mortality (Ohlsson et al., 2010).

High-density lipoprotein cholesterol (HDL) is known as "good" cholesterol, low levels being associated with increased CVD risks (Wannamethee et al., 2000).

Glycated haemoglobin (HbA1c) measures blood sugar, and is regarded as a validated diagnostic test for diabetes (WHO, 2011).

The estimated glomerular filtration rate (EGFR), calculated from the serum creatinine concentration, measures kidney function; higher EGFR levels indicate better kidney function (Levey et al., 2009).

We used albumin, the main protein made by the liver, as a liver function test; low albumin levels suggest impaired liver function (Howard and Sparks, 2016).

Haemoglobin (Hgb), is an iron-containing protein responsible for carrying oxygen throughout the body, and was used to proxy anaemia status; lower levels of Hgb are suggestive of anaemia (Balarajan et al., 2012).

In addition to specific markers, we also used two composite summary measures. One was an index of multi-system risk that measures the wear and tear on the body, approximating the allostatic load (Seeman et al., 2008). Our index combined the selected biomarkers for inflammation, blood pressure, HR, HbA1c, HDL cholesterol, albumin, DHEAS and the WHR

(Seeman et al., 2008; Howard and Sparks, 2016). HDL, Albumin and DHEAS were converted to negative values to reflect ill health, and then each biomarker was transformed into z-scores and summed to calculate the overall index. The second index was a cumulative risk score for CVD, created by adding the relevant z-scores for WHR, blood pressure, HbA1c and CRP (Walsemann et al., 2016). A summary of all biomarkers by reported future disability state is given in the Supplement Table S3.

2.2 Self-assessed health

SAH is considered a summary measure capturing the way that numerous aspects of health, both subjective and objective, are combined within the perceptual framework of the individual respondent (Jylhä, 2009). The SAH question asked respondents to rate their health on a five-point scale from "excellent" to "poor". It was collected in the self-completion instrument at baseline, approximately five months prior to biomarker collection. We group the lowest two SAH categories because of their small sample size, giving a four-point scale ranging from 1 = "excellent" to 4 = "fair" or "poor".

2.3 Disability measures

Our measures of disability were collected at UKHLS waves 4-6, so disability outcomes were observed for prediction horizons of $t = 2 \ldots 4$ years for the main UKHLS sample or $t = 2, 3$ years for BHPS respondents. Respondents were asked about any long-standing physical or mental impairment that they might have and then the consequent functional difficulties, from a list of twelve provided (see Supplement Table S1 for the full list). We used the report of any functional difficulty as a dichotomous variable and the number of functional difficulties (coded as an ordinal variable: 0,1,2 or 3+) as an indicator of severity. Specific difficulty with mobility is also examined as a separate dichotomous indicator because of its relatively high prevalence and significance for functioning and independence in later life (Guralnik et al.,

1993).

Our fourth disability measure came from the income module of the UKHLS questionnaire, constructed as a binary indicator of whether the respondent received income from private disability insurance or any of the UK disability benefit programmes. For all programmes, receipt requires a decision to apply for the benefit, the ability to craft a good-quality application and a positive assessment of need by the programme administrators. Consequently, while the benefit receipt indicator involves a rigorous external assessment of severity, that assessment is confounded to some degree by the incentive and capacity to apply for the benefit, which has a strong socioeconomic gradient (Hancock et al., 2016).

2.4   Covariates

The explanatory covariates that we used in our models have been found to be associated with disability (Hernández-Quevedo et al., 2008; Morciano et al., 2015), and also directly with biomarkers (Carrieri and Jones, 2017). The covariates were collected at baseline and are described and summarised in the Supplement Table S4. Gender and polynomials in age were used to capture demographic influences. Three indicators of socio-economic status were included: educational attainment, home ownership and household income. We excluded disability benefits from income to avoid spurious correlation arising from the fact that disability creates eligibility for those benefits (see Morciano et al. (2015)). We also controlled for marital status, household composition, national and urban dummies. To assess the impact on future disability of lung function over and above smoking status, we estimated models with and without the inclusion of smoking variables.

3   Methods

For each of the binary disability indicators, we applied a probit prediction model, with estimation carried out for the set of individuals who had no observed history of disability at the time of biomarker collection ($D_{i0} = 0$):

$$Pr\left(D_{i4} = 1 | X_{i0}, S_{i0}, B_{i0}\right) = \Phi\left(\alpha_0 + \alpha_1 B_{i0} + \alpha_2 S_{i0} + \alpha_3 X_{i0} + u_{it}\right) \tag{1}$$

The variable D_{i4} is the disability measure observed 4 years after baseline and $\Phi(.)$ is the $N(0,1)$ distribution function. B_{i0} is a specific biomarker or index of biomarkers observed at the nurse visit in period 0; S_{i0} is a set of dummy variables representing SAH in period 0; X_{i0} is a set of covariates describing the individual and her/his history up to the nurse visit; u_{it} is a random residual; and $\alpha_0 \ldots \alpha_3$ are coefficients. For the count indicator of disability, we extended model (1) to a 4-level ordered probit model. Note that the use of a 4-year horizon excludes BHPS members who received the nurse visits at wave 3; however, they were included in the LV model presented later in this section.

We did not use all twelve specific biomarkers simultaneously as predictors in model (1), for three reasons. First, there was a significant loss of usable data when all markers are required to be observed. Second, the full set of health measures (SAH and biomarkers) displayed a substantial degree of collinearity, so there would have been a further loss of statistical precision. A third reason for considering each biomarker separately is that, in practice, it is unlikely that any screening programme would simultaneously check blood pressure, adiposity, blood sugar, cholesterol, haemoglobin, hormone levels, liver, kidney and lung function; so there is a practical interest in the predictive power of each specific marker on its own. However, we also separately used the two composite indexes for allostatic load and CVD risk to consider the potential performance of more comprehensive tests.

Biomarkers and SAH may be noisy indicators of the relevant health concepts, rather than

direct observations of those concepts. Measurement noise may bias estimates of predictive models like (1), usually causing attenuation of the estimated impact of biomarkers and SAH. The LV approach offers a way of dealing with measurement error, by exploiting the multiplicity of biomarkers that reflect to varying degrees the underlying health state. Our aim here is to develop a form of the LV model that gives a clear indication of the predictive value of the biomarker information beyond that contained in the SAH measure.

We used a two-factor structural LV model in which a latent variable h_{i0} reflects a dimension of general health at baseline, measured to varying degrees by SAH and the set of twelve biomarkers (not the indexes for allostatic load and CVD risk). To capture the incremental contribution of the biomarkers, we specified a second latent health variable, b_{i0} representing any dimension of baseline health that the biomarkers succeed in measuring, but which is not captured by SAH. The outcome variables, D_{i2}, D_{i3}, D_{i4}, represent disability 2, 3 and 4 years after baseline; the correlation between them is captured by an unobserved random effect, u_i, which may have different impacts in different periods. The resulting model has the structure shown in Figure 1 and set out algebraically in the Supplement.

We used the estimated LV model to construct two predictive probabilities, π_1 and π_2 based on different predictor sets, one comprising the baseline covariates and SAH, the other expanded to include also the twelve biomarkers. Omitting time subscripts, the predictive probabilities are:

$$\pi_1(X, S) = P(D = 1|X, S)$$

$$\pi_2(X, S, B_1 \ldots B_{12}) = P(D = 1|X, S, B_1 \ldots B_{12})$$

These were computed using Bayes' rule as $\pi_1 = \left[E\, P(D = 1, S|h, b, u, X) \right] / \left[E\, P(S|h, b, u, X) \right]$ and similarly for π_2, where the expectations with respect to the latent variables h, b, u were approximated using Monte Carlo simulations with 100,000 replications.
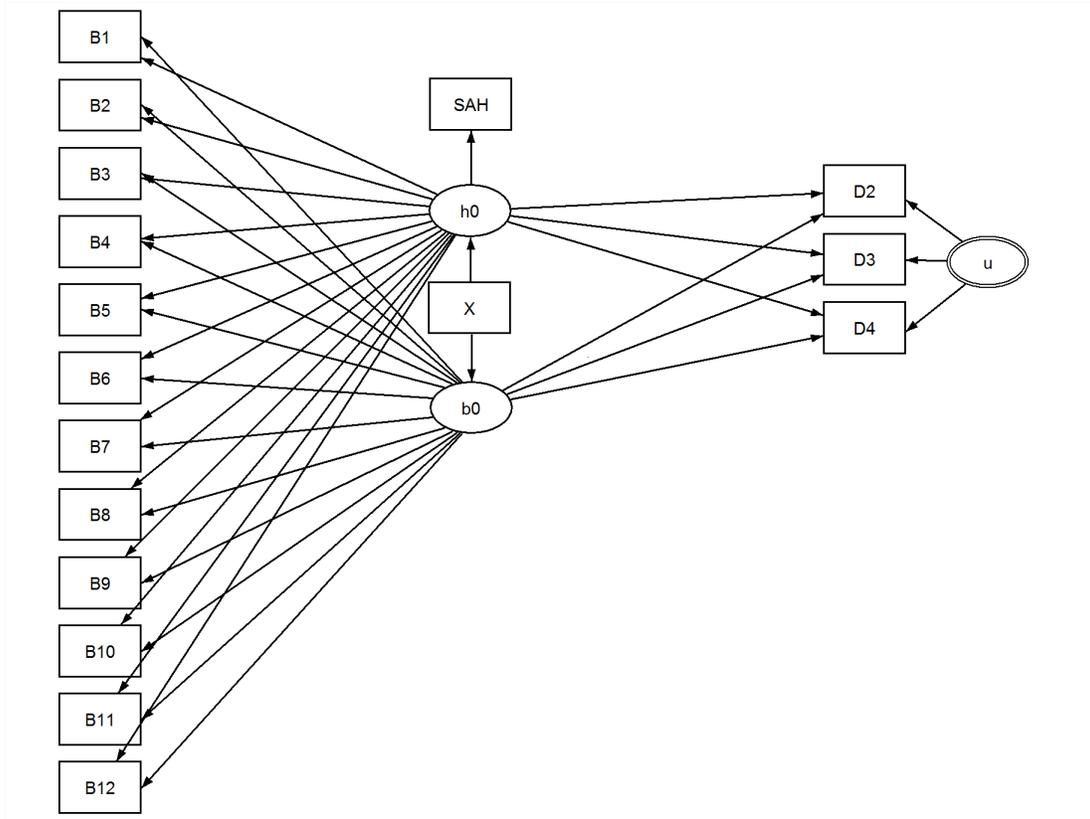
Figure 1: Path diagram for 2-factor LV model

4 Results

The results for model (1) at horizon $t = 4$ are presented in Table 1, which shows the percentage impact on the predicted number of people classified as disabled, of a 1-standard deviation increase in the relevant biomarker. Columns three (one or more functional difficulties), six (mobility) and seven (benefits) of the table were derived from binary probit models; each cell in those columns was based on a separate model for the relevant combination of biomarker and disability measure. The cells in each row of columns four and five came from the same ordered probit model for the number of reported disabilities, with the impact of each biomarker evaluated respectively at the 2+ and 3+ thresholds.

First note that, when SAH was excluded from the prediction models, almost all biomarkers had substantial and statistically significant (at least at the 5% level) predictive power.

The exceptions were for hypertension, DHEAS, EGFR and Hgb, where estimates were insignificant at the 5% level for some disability measures. When dummy variables for SAH status were introduced into the model, the magnitudes of the biomarker coefficients fell, mostly by 20-40%, but generally remained statistically significant. Consequently, the SAH measure succeeded in capturing some of the predictive power of the more objective measures, but far from all.

There was some variation across biomarkers but, overall, the nurse-collected and blood-based biomarkers were comparable in terms of their predictive power. As expected, biomarkers for which higher values represent worse health states had a positive percentage impact on future disability and *vice versa*. Note that lung function, which emerged as a strong influence on disability, was not acting as a proxy for smoking, since inclusion of smoking variables left the estimated effect practically unchanged. To save space, the results for shorter prediction horizons (t=2, 3) are reported in the Supplement (Tables S5-S7); they show systematic predictive power for most of the biomarkers, with the effects mostly rising for longer prediction horizons.

Both composite biomarker measures gave strong effects, with allostatic load having the strongest impact. The results for the benefit receipt measure of disability were an interesting exception to this: when SAH was included in the model, the magnitude of the biomarker effect halved and retained significance only at the 10% level. There may be two behavioural factors involved in that result. One is justification bias – receipt of benefit may lead some respondents to report a worse state of health in SAH to justify their receipt of disability benefit. Alternatively, some people may be reluctant to accept or admit that their health is poor, leading them both to under-represent their current health difficulties in SAH and avoid claiming their potential entitlement to disability benefit. Both of these behaviours would be likely to strengthen the empirical SAH effect relative to the estimated effect of allostatic load or CVD risk indexes (which are more highly correlated with SAH than individual biomarkers, and therefore more affected by bias in SAH).

Table 1: Four year ahead prediction models: % impact on mean disability prevalence of a standard deviation increase

Marker	SAH	No. of functional difficulties reported			Mobility difficulty	Benefit receipt
		1 or more[§]	2 or more[†]	3 or more[†]		
Nurse-collected measures						
Waist/height ratio	excluded	20.2***	25.1***	28.0***	38.1***	33.7***
	included	11.1***	13.9***	15.5***	27.0***	20.9***
Grip Strength	excluded	−13.5***	−20.0***	−25.1***	−18.0***	−32.6***
	included	−10.3***	−15.9***	−20.1***	−10.2	−22.3***
Hypertension[1]	excluded	14.3*	17.0*	18.8*	26.0***	24.6*
	included	3.7	4.3	4.7	11.8	11.4
Heart rate	excluded	13.3***	16.3***	18.1***	15.9***	16.1***
	included	9.9***	12.1***	13.4***	12.2***	13.8***
FVC[2]	excluded	−26.3***	−26.4***	−28.3***	−31.1***	−39.3***
	included	−15.3***	−15.1***	−16.2***	−16.6**	−22.0**
FVC[3]	excluded	−25.6***	−26.3***	−28.2***	−29.3***	−38.1***
	included	−14.7***	−15.1***	−16.2***	−15.5**	−21.2**
Blood-based biomarkers						
HDL	excluded	−20.8***	−22.4***	−24.8***	−21.1***	−8.3
	included	−15.5***	−15.9***	−17.5***	−15.0***	−1.2
CRP	excluded	15.4***	15.3***	16.4***	19.6***	20.2***
	included	11.4***	10.7***	11.4***	13.6***	14.9***
HbA1c	excluded	14.8***	16.3***	17.9***	14.0***	10.0*
	included	10.4***	11.2***	12.3***	7.2*	3.5
DHEAS	excluded	−12.1**	−12.5**	−13.8**	−13.2*	−10.0
	included	−9.1*	−9.0	−9.9	−7.0	−3.0
EGFR	excluded	−9.2*	−10.6*	−11.7*	−15.0**	−9.2
	included	−7.6	−8.4	−9.2	−11.5*	−4.7
Hgb	excluded	−5.2	−6.4	−7.1	−10.7*	−16.7**
	included	−3.0	−3.6	−4.0	−6.9	−12.7*
Albumin	excluded	−10.8***	−13.9***	−15.3***	−19.4***	−16.2**
	included	−7.6**	−10.0**	−11.0***	−14.6***	−11.2*
Systemic risk indexes						
Allostatic load	excluded	26.2***	28.8***	30.0***	39.1***	25.4***
	included	18.1***	19.5***	20.0***	27.2***	13.0*
CVD risk	excluded	20.4***	22.4***	23.3***	33.6***	24.1***
	included	13.4***	14.4***	15.0***	23.3***	12.4*
Sample proportion with disability		0.127	0.057	0.026	0.072	0.042

[§] derived from binary probit model. [†] derived from ordered probit model. [1] Impact of switch from non-hypertensive to hypertensive; [2] without smoking covariate; [3] with smoking covariate. Statistical significance: * = 10%, ** = 5%, *** = 1%.

The SAH impact and most of the biomarker impacts were statistically significant at the 5% level, but statistical significance does not necessarily imply that biomarkers make a contribution to prediction large enough to be of practical significance. Table 2 compares the magnitude of the SAH and biomarker impacts calculated from models where both SAH and the relevant biomarker were included among the covariates. Note that the estimated SAH impact varies slightly across rows, since each model used a different biomarker and there were

also slight differences in sample sizes due to missing data. The SAH impact was calculated as the mean predicted impact of switching the individual from the best ("excellent") to worst ("poor/very poor") category of SAH; with the exception of the binary hypertension marker, the biomarker effect was calculated as the mean impact of switching from a value approximately 1.5 standard deviations better to 1.5 standard deviations worse than the mean (between the 5th and 95th percentiles). Impacts are calculated using the covariates X for each sampled individual and then averaged.

Table 2 shows that the biomarkers with statistically significant impacts made predictive contributions of about 20-25% of the SAH effect. For disability measures based on the number of disabilities reported and mobility difficulties, allostatic load made the largest contribution to prediction, both absolutely and as a proportion of the SAH impact: the impact of a three standard deviation change in allostatic load was approximately one third that of the hypothetical SAH shift. For the mobility-based disability criterion, allostatic load and WHR had the largest impacts in absolute terms (roughly 40% of the SAH effect). In accordance with the results in Table 1, allostatic load contributed little to the prediction of benefit receipt, while the markers for grip strength, adiposity and lung function all gave substantially greater impacts (around one third of the SAH effect).

Despite some differences between disability definitions in the pattern of results, the overall conclusion seems robust – biomarkers made a contribution to prediction of disability four years ahead that is significant both statistically and in terms of absolute magnitude. But that contribution is moderate in comparison with the information contained in SAH.

Tables 3 and 4 summarise results from the LV model. Table 3 gives the estimated factor loadings relating observable biomarkers to the two latent dimensions of health. The loadings were normalised to give the proportional impacts of latent health on observed indicators in standard deviation units. The loadings of the primary latent health factor h_{i0} were mostly as expected, with disability risk raised significantly by WHR, hypertension, CRP and HbA1c;

Table 2: Estimated mean partial impacts of deterioration in SAH ("excellent" to "poor/very poor") and biomarkers (3-standard deviation change centred on mean) on disability prevalence 4 years ahead

Health indicator	No. of functional difficulties reported			Mobility difficulty	Benefit receipt
	1 or more[§]	2 or more[†]	3 or more[†]		
SAH	0.208	0.118	0.061	0.129	0.076
WHR	0.042	0.024	0.012	0.057	0.026
SAH	0.213	0.121	0.063	0.135	0.078
Grip strength	0.039	0.027	0.014	0.022	0.028
SAH	0.225	0.127	0.064	0.152	0.088
Hypertension[*]	0.004	0.003	0.001	0.009	0.005
SAH	0.221	0.125	0.063	0.152	0.086
HR	0.038	0.022	0.010	0.026	0.017
SAH	0.202	0.114	0.056	0.124	0.086
FVC	0.056	0.026	0.012	0.036	0.027
SAH	0.219	0.125	0.064	0.160	0.080
HDL	0.059	0.028	0.014	0.032	0.002
SAH	0.215	0.120	0.060	0.158	0.076
CRP	0.036	0.016	0.010	0.024	0.015
SAH	0.211	0.120	0.061	0.158	0.076
HbA1c	0.042	0.019	0.009	0.015	0.004
SAH	0.224	0.128	0.065	0.167	0.080
DHEAS	0.034	0.015	0.008	0.015	0.004
SAH	0.228	0.129	0.066	0.167	0.080
EGFR	0.028	0.015	0.007	0.025	0.006
SAH	0.216	0.123	0.063	0.160	0.075
Hgb	0.013	0.006	0.003	0.015	0.017
SAH	0.228	0.129	0.066	0.165	0.078
Albumin	0.030	0.018	0.008	0.031	0.014
SAH	0.184	0.104	0.050	0.139	0.072
Allostatic load	0.069	0.033	0.015	0.056	0.016
SAH	0.196	0.110	0.052	0.143	0.073
CVD risk	0.051	0.025	0.011	0.049	0.015

[§] derived from binary probit model. [†] derived from ordered probit model. [*] Impact of switch from non-hypertensive to hypertensive. Statistical significance: all SAH effects significant at 1% level; for statistical significance of biomarkers, see Table 1

and lowered by grip strength, lung capacity, HDL cholesterol, DHEAS, EGFR and Albumin. Small perverse effects (not all statistically significant at the 5% level) were found for HR and Hgb.

For any given marker, if the loading on h has the correct sign and the loading on b has the same sign, then this implies that the health concept implicitly captured by SAH understates the role of the biological pathway which the marker measures. Under this interpretation, our finding was that SAH strongly understates the importance of grip strength, lung function,

DHEAS and liver function and weakly understates the effect of CRP. If the loading on h has the correct sign and the loading on b has the opposite sign, then SAH can be interpreted as over-emphasising the pathway measured by the marker. This was the case for WHR, hypertension and HDL cholesterol and, more weakly, for HbA1c and EGFR.

In two cases, HR and Hgb, the loading on h was small, with an *a priori* wrong sign, which was strongly reversed in the loading for b. This finding can be interpreted to mean that, in these dimensions, SAH is a potentially misleading indicator of future disability.

Table 3: Estimated factor loadings (marginal impact on each health measure of a standard deviation increase in latent factor h or b)

| | Disability measure: | | | | | | | |
| | Any difficulty | | ≥ 2 difficulties | | Mobility | | Benefit receipt | |
Marker	h	b	h	b	h	b	h	b
SAH[§]	0.237	-	0.288	-	0.309	-	0.367	-
Nurse-measured markers								
WHR	0.498	-0.171	0.512	-0.189	0.504	-0.207	0.540	-0.193
Grip	-0.006[†]	-0.829	-0.043[†]	-0.819	-0.122	-0.794	-0.092	-0.808
Hypertension[§]	0.924	-0.219	0.953	-0.253	0.944	-0.298	0.966	-0.248
HR	-0.095	0.134	-0.090	0.132	-0.080	0.130	-0.082	0.131
FVC	-0.324	-0.676	-0.358	-0.654	-0.426	-0.618	-0.407	-0.632
Blood-based biomarkers								
HDL	-0.100	0.377	-0.111	0.380	-0.083	0.375	-0.128	0.382
CRP	0.185	0.060	0.205	0.044	0.215	0.033	0.235	0.031
HbA1c	0.424	-0.061	0.426	-0.087	0.425	-0.100	0.443	-0.088
DHEAS	-0.527	-0.309	-0.799	-0.284	-0.576	-0.248	-0.558	-0.272
EGFR	-0.796	0.082	-0.799	0.099	-0.794	0.136	-0.801	0.103
Hgb	0.131	-0.639	0.105	-0.634	0.051[†]	-0.619	0.073	-0.622
Albumin	-0.326	-0.273	-0.346	-0.253	-0.375	-0.225	-0.377	-0.235

[†] loading not significantly different from zero at 5% (2-tailed test); all other loadings significant at 1%.

[§] for discrete indicators, marginal impact on the continuous latent variable underlying the observed indicator.

Our finding from the simple predictive model (1) was that biomarkers provide significant and substantial predictive power which is nevertheless moderate in relation to SAH. The same result is also evident in results from the LV model. Table 4 gives summary statistics for the predictive probabilities of future disability conditional only on SAH and covariates (π_1), and conditional on SAH, covariates and twelve biomarkers (π_2). These predictive probabilities were calculated for all disability definitions and prediction horizons.

For three of the predictors, the mean predicted probability of disability rose as we raised

the prediction horizon from $t = 2$ to $t = 4$, by approximately 20% (1 or more functional difficulties), 28% (mobility problem) or 62% (benefit receipt). This rise in disability risk over time is a natural reflection of the age-related character of disability prevalence. However, there was no such rise in prevalence measured as the proportion of individuals reporting two or more functional difficulties. For all four disability measures, the predictive probabilities also became substantially more variable (their standard deviations increased) as the prediction horizon lengthened. For (almost) all of the predictions, there was a rise in the variability of the predictor when we expanded the baseline predictor set by adding the biomarkers. This reflects the fact that adding biomarkers gives a more detailed and diverse picture of each individual's disability risk. The correlations between π_1 and π_2 were generally high, reflecting the good – but not perfect – performance of SAH as a general health proxy.

Table 4: Summary statistics of predictive probabilities for alternative disability measures at horizons $t = 2, 3, 4$, derived from estimated LV models

Prediction	$\pi_1(X, S)$		$\pi_2(X, S, B_1 \ldots B_J)$		
horizon	Mean	S.D.	Mean	S.D.	$corr(\pi_1, \pi_2)$
One or more functional difficulties					
$t = 2$	0.113	0.075	0.110	0.081	0.911
$t = 3$	0.124	0.087	0.120	0.093	0.908
$t = 4$	0.135	0.086	0.132	0.092	0.914
Two or more functional difficulties					
$t = 2$	0.068	0.063	0.084	0.070	0.894
$t = 3$	0.067	0.066	0.082	0.074	0.891
$t = 4$	0.066	0.068	0.080	0.076	0.893
Mobility problem					
$t = 2$	0.061	0.058	0.058	0.063	0.883
$t = 3$	0.069	0.064	0.066	0.070	0.874
$t = 4$	0.078	0.069	0.074	0.075	0.879
Receipt of disability benefit					
$t = 2$	0.027	0.032	0.025	0.034	0.856
$t = 3$	0.031	0.039	0.028	0.042	0.852
$t = 4$	0.044	0.047	0.040	0.050	0.866

5   Discussion and conlusions

We investigated the predictive power of objective nurse-collected biomeasures and blood-based biomarkers for functional disability, following individuals who reported no disability at baseline for up to four years after collection of the health measures. We used a wide range of biomarkers, and alternative measures of disability, covering the existence and number of functional disabilities, mobility difficulties and receipt of disability benefits.

For almost all of the biomarkers, we found 4-year-ahead predictive effects that were substantial in magnitude and statistically significant. When SAH was introduced as a predictor alongside the biomarkers, the magnitude of the biomarker effects fell, in most cases by 20-40%, but remained important in magnitude and highly statistically significant for most of the biomarkers examined. Although there were some differences across biomarkers and disability measures, we found that measures of adiposity, grip strength, heart rate, lung functioning, cholesterol levels, inflammation, blood sugar and anaemia had strong predictive power for future disability risk, over and above SAH.

In addition to simple predictive models, we also developed a new latent variable approach capable of incorporating large numbers of biomarkers and longitudinal observation of disability outcomes, while allowing for measurement error in SAH and biomarkers. This approach allowed us to identify distortions in SAH as a measure of health, by detecting an additional predictive factor that is revealed by the biomarkers but not by SAH. The corresponding factor loadings indicate dimensions of biological function that are given too much or too little weight by SAH in predicting disability. We found that SAH is excessively sensitive to the biological pathways reflected in adiposity, hypertension and cholesterol, and insufficiently sensitive to strength, lung function, hormonal balance and liver function.

Nevertheless, SAH emerged as a good general health proxy in the sense that, when SAH and biomarkers were both used as predictors, the estimated biomarker impacts on future disability, although substantial absolutely, were moderate in comparison with the effects

18

of SAH. For example, using a composite summary measure to proxy allostatic load (our strongest biomarker predictor), we found that moving from the best (excellent) to worst (fair/poor) category of SAH increased the risk of disability 4 years later by 5-18 percentage points on average, depending on the disability concept used, while an increase in allostatic load from the fifth to the lowest 95th percentile (roughly a 3-standard deviation rise) increased disability risk by 2-7 points.

5.1 Limitations

Key strengths of our analysis come from its use of UKHLS data which allowed us to use a large, nationally representative sample covering all adult ages. The bio-social character of UKHLS provided a wide range of nurse-collected and blood-based biomarkers, in addition to SAH, disability indicators and extensive measures of household characteristics and socio-economic status. This adds breadth and depth to the small body of evidence that already exists on biomarkers as predictors of future disability. Existing studies are more limited in terms of the range of biomarkers used and also the study population, which is mainly restricted to older people, nonrepresentative samples or specific patient groups (Brex et al., 2002; Reuben et al., 1999; Baylis et al., 2013; Seeman et al., 1994). As far as we are aware, ours is the first study that makes an explicit evaluation of subjective SAH against objective biomarker information in relation to disability.

Despite these advantages, there are limitations. First, the available data follow individuals for a relatively short time horizon. We have found evidence of a rise in the estimated effect of biomarkers as the length of the prediction horizon increases, suggesting that our results may understate the full long-term predictive role of biomarkers. Second, functional disability is a slippery, hard-to-measure concept and the measures used in our analysis are necessarily limited. Our use of a range of alternative disability measures alleviates these concerns to some degree, but a complete solution requires further research. Finally, although we

19

used an unusually extensive set of biomarkers, the multidimensional nature of the biomedical processes underlying disability means that there may remain significant aspects of physical health that are not covered by our analysis.

References

Balarajan, Y., Ramakrishnan, U., Özaltin, E., Shankar, A. H., and Subramanian, S. (2012). Anaemia in low-income and middle-income countries. *The Lancet*, 378(9809):2123–2135.

Baylis, D., Bartlett, D., Syddall, H., Ntani, G., Gale, C., Cooper, C., Lord, J., and Sayer, A. (2013). Immune-endocrine biomarkers as predictors of frailty and mortality: a 10-year longitudinal study in community-dwelling older people. *Age*, 35(3):963–971.

Brex, P. A., Ciccarelli, O., O'Riordan, J. I., Sailer, M., Thompson, A. J., and Miller, D. H. (2002). A longitudinal study of abnormalities on MRI and disability from multiple sclerosis. *New England Journal of Medicine*, 346(3):158–164.

Carrieri, V. and Jones, A. M. (2017). The income–health relationship beyond the mean: New evidence from biomarkers. *Health Economics*, 26(7):937–956.

Colburn, W., DeGruttola, V. G., DeMets, D. L., Downing, G. J., Hoth, D. F., Oates, J. A., Peck, C. C., Schooley, R. T., Spilker, B. A., Woodcock, J., et al. (2001). Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. Biomarkers Definitions Working Group. *Clinical Pharmacol & Therapeutics*, 69:89–95.

Commission on Funding of Care and Support (2011). *Fairer care funding: The report of the Commission on Funding of Care and Support*, volume 1. The Stationery Office.

Crossley, T. F. and Kennedy, S. (2002). The reliability of self-assessed health status. *Journal of health economics*, 21(4):643–658.

DWP (2017). *Family Resources Survey: financial year 2015/16.* Department for Work and Pensions.

Fried, L. P., Herdman, S. J., Kuhn, K. E., Rubin, G., and Turano, K. (1991). Preclinical disability: hypotheses about the bottom of the iceberg. *Journal of Aging and Health*, 3(2):285–300.

Gray, L. A., Leyland, A. H., Benzeval, M., and Watt, G. C. (2013). Explaining the social patterning of lung function in adulthood at different ages: the roles of childhood precursors, health behaviours and environmental factors. *Journal of Epidemiology and Community Health*, 67(11):905–911.

Gruenewald, T. L., Seeman, T. E., Ryff, C. D., Karlamangla, A. S., and Singer, B. H. (2006). Combinations of biomarkers predictive of later life mortality. *Proceedings of the National Academy of Sciences*, 103(38):14158–14163.

Guralnik, J. M., LaCroix, A. Z., Abbott, R. D., Berkman, L. F., Satterfield, S., Evans, D. A., and Wallace, R. B. (1993). Maintaining mobility in late life. I. Demographic characteristics and chronic conditions. *American Journal of Epidemiology*, 137(8):845–857.

Hancock, R. M., Morciano, M., and Pudney, S. E. (2016). *Disability and Poverty in Later Life.* Joseph Rowntree Foundation.

Hernández-Quevedo, C., Jones, A. M., and Rice, N. (2008). Persistence in health limitations: a European comparative analysis. *Journal of Health Economics*, 27(6):1472–1488.

Howard, J. T. and Sparks, P. J. (2016). Does allostatic load calculation method matter? Evaluation of different methods and individual biomarkers functioning by race/ethnicity and educational level. *American Journal of Human Biology*, 28(5):627–635.

Idler, E., Leventhal, H., McLaughlin, J., and Leventhal, E. (2004). In sickness but not in health: Self-ratings, identity, and mortality. *Journal of health and social behavior*, 45(3):336–356.

Idler, E. L. and Benyamini, Y. (1997). Self-rated health and mortality: a review of twenty-seven community studies. *Journal of health and social behavior*, pages 21–37.

Johnston, D. W., Propper, C., and Shields, M. A. (2009). Comparing subjective and objective measures of health: Evidence from hypertension for the income/health gradient. *Journal of Health Economics*, 28:540–552.

Jones, M. (2016). Disability and labor market outcomes. *IZA World of Labor*.

Jylhä, M. (2009). What is self-rated health and why does it predict mortality? towards a unified conceptual model. *Social Science and Medicine*, 69(3):307–316.

Kallaur, A. P., Reiche, E. M. V., Oliveira, S. R., Pereira, W. L. d. C. J., Alfieri, D. F., Flauzino, T., de Meleck Proença, C., Lozovoy, M. A. B., Kaimen-Maciel, D. R., Maes, M., et al. (2017). Genetic, immune-inflammatory, and oxidative stress biomarkers as predictors for disability and disease progression in multiple sclerosis. *Molecular neurobiology*, 54(1):31–44.

Levey, A. S., Stevens, L. A., Schmid, C. H., Zhang, Y. L., Castro, A. F., Feldman, H. I., Kusek, J. W., Eggers, P., Van Lente, F., Greene, T., et al. (2009). A new equation to estimate glomerular filtration rate. *Annals of internal medicine*, 150(9):604–612.

McFall, S. L., Petersen, J., Kaminska, O., and Lynn, P. (2014). Understanding Society waves 2 and 3 Nurse Health Assessment, 2010-2012. Guide to Nurse Health Assessment. ISER, University of Essex.

Morciano, M., Hancock, R. M., and Pudney, S. E. (2015). Birth-cohort trends in older-age functional disability and their relationship with socio-economic status: evidence from a pooling of repeated cross-sectional population-based studies for the UK. *Social Science and Medicine*, 136:1–9.

Ohlsson, C., Labrie, F., Barrett-Connor, E., Karlsson, M. K., Ljunggren, O., Vandenput, L., Mellstrom, D., and Tivesten, A. (2010). Low serum levels of dehydroepiandrosterone

sulfate predict all-cause and cardiovascular mortality in elderly Swedish men. *The Journal of Clinical Endocrinology and Metabolism*, 95(9):4406–4414.

Pearson, T. A., Mensah, G. A., Alexander, R. W., Anderson, J. L., Cannon, R. O., and Criqui, M. (2003). Markers of inflammation and cardiovascular disease. application to clinical and public health practice: A statement for healthcare professionals from the Centers for Disease Control and Prevention and the American Heart Association. *Circulation*, 107:499–511.

Pudney, S. E., Skew, A., and Taylor, M. P. (2011). The economic impacts of leaving employment for health-related reasons. Technical report, ISER, University of Essex.

Reuben, D. B., Ix, J. H., Greendale, G. A., and Seeman, T. E. (1999). The predictive value of combined hypoalbuminemia and hypocholesterolemia in high functioning community-dwelling older persons: Macarthur studies of successful aging. *Journal of the American Geriatrics Society*, 47(4):402–406.

Ridker, P. M. (2007). Inflammatory biomarkers and risks of myocardial infarction, stroke, diabetes, and total mortality: implications for longevity. *Nutrition Reviews*, 65:S253–S259.

Seeman, T., Merkin, S. S., Crimmins, E., Koretz, B., Charette, S., and Karlamangla, A. (2008). Education, income and ethnic differences in cumulative biological risk profiles in a national sample of US adults: NHANES III (1988-1994). *Social Science and Medicine*, 66(1):72–87.

Seeman, T. E., Charpentier, P. A., Berkman, L. F., Tinetti, M. E., Guralnik, J. M., Albert, M., Blazer, D., and Rowe, J. W. (1994). Predicting changes in physical performance in a high-functioning elderly cohort: Macarthur studies of successful aging. *Journal of gerontology*, 49(3):M97–M108.

Seeman, T. E., Crimmins, E., Huang, M.-H., Singer, B., Bucur, A., Gruenewald, T., Berkman, L. F., and Reuben, D. B. (2004). Cumulative biological risk and socio-economic dif-

ferences in mortality: Macarthur studies of successful aging. *Social Science and Medicine*, 58(10):1985–1997.

Walsemann, K. M., Goosby, B. J., and Farr, D. (2016). Life course SES and cardiovascular risk: Heterogeneity across race/ethnicity and gender. *Social Science and Medicine*, 152:147–155.

Wannamethee, S. G., Shaper, A. G., and Ebrahim, S. (2000). HDL-cholesterol, total cholesterol, and the risk of stroke in middle-aged British men. *Stroke*, 31(8):1882–1888.

WHO (1980). *International classification of impairments, disabilities, and handicaps: a manual of classification relating to the consequences of disease.* Geneva: World Health Organization.

WHO (2011). Use of glycated haemoglobin (HbA1c) in the diagnosis of diabetes mellitus. Technical report, World Health Organization, Geneva.

Zang, Y., Liu, S., and Yuan, Y. (2015). Optimal marker-adaptive designs for targeted therapy based on imperfectly measured biomarkers. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 64(4):635–650.