Abstract

A health risk score was created to investigate the possibility of using data provided by wearable technology to help predict overall health and mortality, with the ultimate goal of using this score to enhance the pricing of health or life insurance. Subjects were categorized into low, increased and high risk groups, and after results were adjusted for age and sex, Cox proportional hazards analysis revealed a high level of significance when predicting mortality. High risk subjects were shown to have a hazard ratio of 2.1 relative to those in the low risk group, which can be interpreted as an equivalent increase in age of 7.8 years. Our findings help to demonstrate the predictive capabilities of potential new rating factors, measured via wearables, that could feasibly be incorporated into actuarial insurance pricing models. The model also provides an initial step for insurers to begin to consider the incorporation of continuous wearable data into current risk models. With this in mind, an emphasis is placed on the limitations of the study in order to highlight the areas that must be addressed before incorporating aspects of this model within current pricing models.

Keywords: emerging risk, wearable technology, insurance, Cox regression, underwriting

1 Introduction

Much like the disruptions seen in the banking industry over the past decade, emerging technologies are revolutionising the insurance industry. Traditional insurers are under pressure to innovate existing business models to retain a competitive edge (Hilton, 2017). Data from CB Insights showed funding to startups in the newly coined *InsurTech* industry has risen from $140m in 2011 to $2.7bn in 2016, and investment in the sector is expected to continue to grow as new technologies arise (Catlin et al., 2017; Jubraj et al., 2017).

The primary driver of this change has been the increasingly larger amounts of personal data available to insurers, which offers the opportunity to predict the risk for each customer and charge them accordingly. Traditionally in life insurance underwriting data comes from a questionnaire and a medical examination performed by a registered nurse or licensed physician, depending on the coverage amount and the age of the customer. A new potential source of this information is the *Internet of Things* (IoT), comprised of the network of physical objects that can connect to the internet and communicate with one another. Examples of these include mobile phones, pacemakers, on board computers in cars and of most interest to this study, *wearable technology*. This term, often shortened to just wearables, describes all technology that is worn comfortably on the body or combined with clothing (Tehrani and Michael, 2014). Common wearables include Fitbit,Garmin fitness bands and the Oura ring. By use of the IoT, insurers have the potential to access huge amounts of real-time data, allowing them to build far more accurate risk profiles concerning the people they insure. Not only can this allow a more personal and fairer method of pricing, but through improved engagement with the customer risks can be both managed and reduced.

The overall research aim of this paper is to demonstrate the potential that data derived from wearable devices may provide to insurance companies in terms of new rating factors for their pricing models. As such we develop a conceptual risk model that utilises data measurable by wearables, and can classify a policy holders relative risk to the rest of the population. This risk model serves to highlight the potential for insurance companies to incorporate wearable device data in their own

health and life insurance related pricing models. As this is to be a preliminary model demonstrating potential and acting as a proof of concept, simplicity will be key in order to retain the model's generalisability. This is the first study that attempts to create a health risk score comprising solely of data which can be collected in a continuous manner by wearable technology.

The rest of the paper is outlined as follows. Section 2 investigates the current state of the insurance industry with respect to the use of wearable technology, and then provides a review of medical literature used in the formation of health risk scores. Section 3 describes the methods used to analyse the data. Section 4 comprises of the results of the analyses and the diagnostic tests performed to confirm the validity of the findings. Section 5 discusses the possible implications of the findings, and follows with an in-depth investigation into the limitations of the analyses performed. The paper is concluded with suggestions for possible extensions of the study.

2 Literature review

2.1 Current State of the Insurance Industry

The insurance industry is well aware of the challenges it is likely to face over the coming years, and so is investing heavily in research and data in order to evolve (Sultan, 2015). Over 63 percent of insurers expect wearables to effect the industry significantly in the next two years (Schwartz and Hamilton, 2015). A huge advantage of these pieces of technology is ability to record and analyse data continuously with minimal interaction.

Already, a number of insurers have begun to incorporate wearables into their products, trialling new innovative programmes in an attempt to get ahead of their competitors and break into markets of potential customers previously considered uninsurable. A main area of concern for insurers is the willingness to participate in these kinds of programs. Opt-in rates can be as low as 5 percent, however PwC found that if the wearable was provided for free, over two thirds of customers or employees would wear the device (Dart, 2017). With this in mind, companies such as John Hancock Financial and MLC have provided Fitbits and smartwatches to customers for free (Becher, 2016). As the main goal at this stage of the process is the collection of data to analyse, they have agreed to lower premiums as an incentive for customers to release their personal data. A more original method to collect data was used by United Health, which used a penalty rather than awards system to motivate consumers (Dart, 2017). By requiring users to reach fitness targets in order to avoid the purchase cost of the wearable, their system was three times more successful in collecting data.

Other companies have gone even further and have launched programmes making direct use of the technology available and the data they receive. One of the first insurers to use wearables in their 'Vitality' product was South African company Discovery (Abraham, 2016). Vitality provides rewards such as discounted travel or accommodation if certain activity levels are met. This is profitable as when policyholders become healthier, the expected cost of the risk pool is lowered. Similar programmes are offered by other insurers such as MLC in Australia, who provide premium discounts when healthy behaviours are displayed. These products are also being marketed towards large businesses that purchase insurance, as healthier employees have on average greater levels of productivity (Abraham, 2016). Existing wearables used for these sorts of programmes include

the Fitbit[1] and the Apple Watch[2]. These products often contain accelerometers to detect movement/sleep/heart rate; data can usually be accessed remotely via smartphone with the relevant GPS tracking data.

The usage of real time data from wearables draws parallels with insurance telematics programmes, which have increasingly gained market share in the automotive industry (Wahlström et al., 2015). GPS can measure metrics such as mileage, speeding and location, whilst an accelerometer known as the "Black Box" can record instances of hard breaking, sharp turns and sudden acceleration (Iqbal and Lim, 2006). Analysis of this data can help build a more personalised and accurate estimate of the level of risk the policy holder places on the insurer. In addition to this, drivers tend to improve driving method when monitored by telematic devices in order to lower their premiums (Azzopardi and Cortis, 2013); thus the policies can encourage safer driving behaviours, lowering the expected cost of the risk pool. If wearables can follow automotive telematics and gain a foothold in the insurance industry, they have the potential to become an integral part of many health and life insurance policies.

2.2   Health Risk Scores in Literature

The key question for research is how insurers can transform wearable technology's raw data into meaningful information that could be used to price their products. Without being able to find a quantifiable link between the measurements and the health of an individual, the data has no value (Abraham, 2016). One possible option to achieve this could be using this data to create a *health risk score*. The concept of summarising a patient's data into a single score is not new in academia. The Framingham Risk Score (Wilson et al., 1998) is used worldwide as an estimate of cardiovascular risk, and the probability of onset of type 2 diabetes is typically predicted by the Diabetes Risk Score (Lindström and Tuomilehto, 2003).

While the risk of a specific health condition can be modeled to a reasonable level of accuracy using known causal variables, the formation of a health score using simple metrics to quantity an individual's overall health and attempt to predict all-cause mortality is a more challenging task. Due to their known strong association with mortality, certain factors such as smoking, alcohol consumption, diet and physical activity are prevalent in the majority of studies (Brandt, 2011; Dam et al., 2008; Ding et al., 2015; Gopinath et al., 2010; Hamer et al., 2011; Khaw et al., 2008; Knoops et al., 2004; Kvaavik et al., 2010; Nechuta et al., 2010). By creating health risk scores whereby each good (bad) behaviour is assigned a point, higher scores were consistently associated with an increased (decreased) risk of mortality. Some studies went further and considered the individual risk combinations and the possibility of synergistic relationships between the factors (Ding et al., 2015), with smoking and excess alcohol consumption having substantially more effect on mortality when combined. A key objective of many studies was to attempt to discover new factors to incorporate into their risk scores. Nechuta et al. (2010) found waist-hip ratio may be an even stronger predictor of mortality than BMI, and decided to include both of these factors in their health risk scores. Ding et al. (2015) incorporated metrics to measure a sedentary lifestyle, finding that both prolonged sitting and unhealthy sleep duration could be used in combination with physical activity levels in a health score.

---

[1]www.fitbit.com

[2]www.apple.com/uk/watch/

Methods used to classify diets also vary considerably between studies. The most common way this is performed is by summing up the quantity of fruit and vegetables eaten and using this as a proxy for a healthy diet, but researchers have endeavored to improve this simple method by including a range of different foods eaten (Brandt, 2011). A further step was performed by Khaw et al. (2008) who used blood plasma vitamin C concentrations as a proxy, allowing a measured value to be reported rather than a potentially biased and inaccurate self reported value.

An interesting idea can be drawn from Glei et al. (2014) and Gruenewald et al. (2006), who discuss the notion of using particular biomarkers to represent the functionality of different biological systems. Gruenewald et al. (2006) give suggestions of biomarkers to act as proxies for neurological function, immune activity, cardiovascular function and metabolic activity. From the perspective of creating a health risk score, ensuring chosen metrics can represent the functionality of all major biological systems could be a route to creating a more complete picture of overall health.

Using walking activity as a metric to predict mortality has had success in both young as well as elderly populations (Tudor-Locke et al., 2011). Walking is a particularly useful metric; while improving cardiovascular or respiratory health, it can also suggest a conscious decision to lead a healthy lifestyle when done for pleasure. Furthermore, a lack of walking can also be indicative of underlying chronic conditions. Simple measures of walking associated with mortality include distance walking per day (Hakim et al., 1998), or equivalently the average number of steps each day (Tudor-Locke et al., 2011). Ganna and Ingelsson (2015) found self reported walking pace to be one of the strongest lifestyle predictors of mortality, greater even than smoking habits. This measurement could be recorded by wearables using a combination of GPS data and accelerometers with the ability to distinguish between walking, running and other types of movement.

Due to the failure of the cardiovascular system being responsible for a large proportion of deaths, it is only natural that many studies have focused on finding ways to measure this risk. Elevated resting heart rate has been shown by many studies to be an independent predictor of both cardiovascular and all-cause mortality (Jensen et al., 2013; Zhang et al., 2015; Seccareccia et al., 2001). This is not the only possible metric however; blood pressure has been shown to be strongly associated with the occurrence of a stroke, and is also highly correlated to all-cause mortality (Georgakis et al., 2017). As technology progresses heart rate variability, which is typically measured by ECG, will likely become measurable on a continuous basis, and shows much promise in being used to predict heart failure (Lucena et al., 2016).

Ding et al. (2015) incorporated sleep duration into their health risk score, yet this is only one way to measure sleep. Wong et al. (2012) found that in addition to duration, sleep quality also had a marked effect on physical wellbeing on a sample of Chinese students. A connection between poor sleeping patterns and the risk of onset of type 2 diabetes was found by Cappuccio et al. (2010a), which as a long term illness can be very expensive as an insurer.

The list of metrics discussed here is not exhaustive, but merely an indication of how much potential exists for this academic area to be developed. Subject to availability of data, Section 3.2 attempts to utilise these possible metrics to create a wearable focused score.

3 Methodology

3.1 Study Population

This analysis is based on data from participants of the Health and Lifestyle Survey (HALS) (Cox, 1988), with the target population defined as individuals of eighteen and over in England, Wales and Scotland. The methods and rationale of this study have been reported elsewhere (Cox et al., 1987). In brief, 12,672 addresses were selected randomly from electoral registers, yielding 12,254 suitable households, each from which one person was randomly chosen. A response rate of 73 per cent generated 9,003 in-person interviews, with 82 per cent (7,414) agreeing to a further visit from a study nurse to carry out various health measurements. Comparison with the 1981 census showed that the sample was representative of the adult British population (Blaxter, 1987). The current status of the participants (alive or deceased) as of June 2009 was provided by the United Kingdom National Health Service (NHS) Central Registry.

Relevant areas of the interviews included lifestyle habits such as alcohol consumption, smoking, physical activity and sleep duration. Information about previous diagnoses and health history were also recorded at this time. Height, weight, blood pressure and resting heart rate were measured by a study nurse in the follow-up visit.

3.2 Assessment of Health Metrics

The chosen health metrics were included for a number of reasons, but the most important factor was the ability to effectively quantify the information provided by the HALS to separate subjects into healthy and unhealthy groups. The commonly used metrics (alcohol consumption, smoking and body mass index) are health metrics that have been shown in previous literature to have an effect on mortality risk (Hamer et al., 2011; Khaw et al., 2008; Kuh et al., 2009). The metrics measurable by wearables were chosen due to the existence of wearables that can currently record these metrics to a certain level of accuracy. Exercise activity can be detected and distinguished from walking or regular activity through movement sensors and heart rate increases (Comstock, 2015). Sleeping activity can be tracked to various degrees of accuracy through a multitude of devices, including most mobile phones via applications (Mann, 2017). Blood pressure measurement has only begun to enter the wearables market recently, however as the technology is developed it is likely to become more widespread and available in mainstream devices (Redlitz, 2017). Heart rate is one of the most common health metrics, available in virtually all mainstream fitness wearables including Fitbit, Jawbone and Apple Watch. Similarly, almost all devices have some form of pedometer measuring steps, walking duration, pace and distance walked when combined with GPS.

Poor health metrics were classified as shown in Table 1 using results from previous literature and official bodies (Cappuccio et al., 2010b; Department of Health, 2016; Jensen et al., 2013; Kvaavik et al., 2010; Leitzmann et al., 2007; World Health Organization, 1995; World Health Organization and International Society of Hypertension Writing Group, 2003). The effect of these metrics on subject survival time is assessed by the Cox proportional hazards model[3].

---

[3]For further information on the Cox model see Cox et al. (1972).

Table 1: Health metric classifications

Health metric	Variable	Point awarded	Percentage
Alcohol consumption	Al	intake of > 14 units of alcohol per week	18.3%
Smoking	Sm	current smoker	36.1%
BMI	Bmi	≥ 30 kg/m^2 (obese)	9.5%
Physical activity	Phy	≤ 120 min/week leisure time exercise	76.0%
Sleep duration	Sd	sleeping < 7 or > 9 hours/day	39.8%
Blood pressure	Bp	hypertensive reading	8.6%
Resting heart rate	Rhr	rate ≥ 90bpm	4.7%
Walking Duration	Wd	walking < 20 minutes/day	18.70%

4 Statistical Analysis - Results

Out of the 7,414 participants that agreed to a visit from the study nurse, 291 (3.9%) had incomplete measurements and had to be excluded from the analysis. A further 238 (3.3%) were unable to be categorised in the June 2009 survey, due to reasons such as having departed overseas, no longer being registered on the NHS, or simply being unable to be contacted. This left $n = 6,885$ suitable subjects for the following analyses out of which 2,160 (31.4%) died prior to 1 June 2009. The principle outcome in this study was survival time, measured as the time in years between collection of baseline data until death or date of censorship (1 June 2009). All statistical tests and analyses were performed using Stata 14.1 (StataCorp).

A log-rank test was performed to assess the Kaplan–Meier (survival) functions of males and females, with the null hypothesis for the test assuming the estimates for the two sexes are equal. A p-value of 0.00 indicated this was not the case. Accordingly, all Cox proportional hazards regression models were adjusted for sex and age.

4.1 Individual Health Metrics

Adjusted Cox regressions were run for all eight metrics individually, with results displayed in Table 2. The p-values in the table report the significance of the coefficients using the Wald test statistic, where the null hypothesis assumes $\beta_i = 0$. We can see that all the variables were statistically significant to a 99.5 per cent level of confidence except for alcohol consumption, which was significant to the 96 per cent level. The hazard ratios show smoking raised the probability of death to the greatest extent, whilst alcohol consumption and abnormal sleep duration had the least effect. It is important to note the non-linear relationship between the percentage of deaths of those with a poor health metric and the corresponding hazard rate. 68.1 per cent of those with high blood pressure died in the study versus only 33.7 per cent of smokers, however the hazard ratio of smokers was much higher. Adjusting for age or taking into account the survival times after measurement could have marked effects on the coefficients of our model. For reference, the hazard ratios for smoking and high blood pressure after removing the adjustment for age were 1.11 (from 1.69) and 3.37 (from 1.26) respectively. This could suggest high blood pressure may be in part caused by age, be more likely to cause death when the subject is at an older age, or be related to

Table 2: Results of individual Cox regressions

Health metric	Deaths	Coefficient	P-value	HR (95% CI)
Alcohol consumption	25.9%	0.13	0.04	1.14 (1.01-1.29)
Smoking	33.7%	0.53	0.00	1.69 (1.55-1.85)
BMI	43.4%	0.28	0.00	1.32 (1.17-1.50)
Physical activity	36.4%	0.33	0.00	1.39 (1.22-1.59)
Sleep duration	39.6%	0.15	0.00	1.16 (1.06-1.26)
Blood pressure	68.1%	0.23	0.00	1.26 (1.13-1.41)
Resting heart rate	45.3%	0.47	0.00	1.60 (1.36-1.90)
Walking Duration	40.7%	0.22	0.00	1.25 (1.13-1.38)

age in another way entirely.

4.2   Combined Health Metrics

For variables to be analyzed in a combined health metrics discussed below a point of 1 was given to each poor health metric classification and 0 otherwise, with classifications defined as in Section 3.2.

A Cox regression was then run on all the variables at once. There were no major differences between this and the individual regressions, except that the variable for alcohol consumption was no longer significant. Further investigation showed that this was mainly due to collinearity between alcohol consumption and with another explanatory variable: smoking. This would be expected considering 51 per cent of those with a poor health classification for alcohol consumption were current smokers, whilst only 26 per cent of smokers were considered to have poor drinking behavior, thus much of alcohol's effects would likely be incorporated into the smoking variable. This lack of significance of an alcohol variable has been seen in similar studies on different populations such as that by Ding et al. (2015), who note that while it may not show significance by itself, when in combination with other metrics such as smoking of physical inactivity it can have a strong association with all-cause mortality. There is also a general consensus of alcohol and mortality having a U-shaped relationship, in which both drinkers and non-drinkers have an increased risk (Khaw et al., 2008). The model showed a high level of significance overall with $\chi^2(10)$ for the log-rank test giving a p-value of 0.00. We can also see the significance of the wearable related health metrics when combined with the more commonly used metrics. This suggests that there may be some benefit to a model consisting of measurements made by wearable technology.

Health Score I   We formulate Health Score I, comprised of all significant variables in the previous analyses. The health score was created by summing the points for each individual subject, giving a possible range of 0–7 points. A lower score was indicative of a healthier lifestyle, and thus it was hypothesised that survival probability would decrease with the increase of poor health metrics. Health Score I can be summarised as

$$\text{Health Score I} = Sm + Bmi + Phy + Sd + Bp + Rhr + Wd \tag{4.1}$$

Table 3: Distribution of Health Score I

Points	Frequency	Percentage of total	Deaths	Percentage died
0	573	8.3%	54	9.4%
1	1,949	28.3%	355	18.2%
2	2,378	34.5%	797	33.5%
3	1,428	20.7%	646	45.2%
4	475	6.9%	250	52.6%
5	73	1.1%	51	69.9%
6	9	0.1%	7	77.8%

Table 4: Results of Cox regression for Health Score I

Variable	Coefficient	P-value	HR (95% CI)
Score I	0.27	0.00	1.31 (1.26-1.36)
0			1 (Reference)
1	0.02	0.88	1.02 (0.77-1.36)
2	0.51	0.00	1.66 (1.25-2.19)
3	0.74	0.00	2.10 (1.59-2.78)
4	0.86	0.00	2.37 (1.76-3.19)
5	1.31	0.00	3.69 (2.51-5.42)
6	1.10	0.01	3.00 (1.36-6.61)

with variables defined in Table 1. A Cox regression was run on Health Score I, assessing both the overall explanatory power of the score and its effectiveness across its range.

The distribution of Score I is shown in Table 3. We can see that no subjects achieved the maximum points tally of 7, and only 9 subjects had 6 points, which could affect the significance of the regressions at this value. The final column shows that as the number of points increases, the percentage of deaths for each total number of points are strictly increasing as hypothesised.

Running an adjusted Cox regression showed that Health Score I was able to predict survival time to a high level of significance, with p-value 0.00. The first row of Table 4 tells us that for an increase of 1 point, a subject has on average a 31 per cent higher chance of dying during the next year. In a stratified analysis of the score we can see that there is little evidence to suggest that the presence of one poor metric had any effect on the hazard ratio of a participant. After this, each additional poor metric shows an increase in predicted hazard ratio until a total of 6 is reached; however this is likely due to the small sample size for this score value, which is apparent on inspection of the wide range seen in the 95 per cent confidence intervals. The confidence intervals for adjacent point totals also overlap, which may suggest that an increase of just one point in Score I may not be statistically significant or indicative that too many metrics are being used.

Table 5: Distribution of Health Score II

Points	Frequency	Percentage of total	Deaths	Percentage died
0	871	12.7%	93	10.7%
1	2,820	41.0%	647	22.9%
2	2,345	34.1%	927	39.5%
3	740	10.8%	415	56.1%
4	104	1.5%	73	70.2%
5	5	0.1%	5	100.0%

Table 6: Results of Cox regression for Health Score II

Variable	Coefficient	P-value	HR (95% CI)
Score II	0.23	0.00	1.25 (1.19-1.31)
0			1 (Reference)
1	0.19	0.08	1.21 (0.97-1.51)
2	0.46	0.00	1.58 (1.27-1.96)
3	0.64	0.00	1.88 (1.50-2.38)
4	0.92	0.00	2.50 (1.83-3.41)
5	0.90	0.05	2.45 (0.99-6.05)

Health Score II   Here, we create an alternative health score which consisted of only the health metrics deemed viable to be measured by wearables, as seen in Section 3.2. Health Score II was defined by

$$\text{Health Score II} = Phy + Sd + Bp + Rhr + Wd \tag{4.2}$$

and was analysed in a similar manner to Health Score I.

Again the maximum points total, in this case 5, had a very low frequency of subjects and so was unlikely to be able to provide a strong level of significance of increased mortality above those with 4 points. As before, the percentage of deaths for each total number of points is increasing, with the difference between each level more distinct.

The Cox regression for this score illustrated a strong ability to predict relative survival time, producing a p-value of 0.00 as seen in Table 6. Score II shows an expected 25 per cent increase of death in the next year for each increase of 1 point. The presence of one single poor metric was again not statistically significant to the 95 per cent level, however with a p-value of 0.08 and a hazard ratio of 1.25, there appears to be an indication of one poor metric having some effect on survival time. When stratifying the score, again we see the hazard ratio increases as we increase the number of points until we reach a total of 5. Once more, this is likely due to the sample size of 5 subjects (0.1%) above anything else. Overlapping confidence intervals are seen again with the individual points totals, suggesting that there is still room to improve the model.

Table 7: Classification of Health Score III

Points in Score II	Score III variable	Risk classification
0	0	
1	0	Low risk
2	1	
3	1	Increased risk
4	2	
5	2	High risk

Table 8: Results of Cox regression for Health Score III

Variable	Coefficient	P-value	HR (95% CI)
Score III	0.35	0.00	1.42 (1.31-1.54)
Age	0.10	0.00	1.10 (1.10-1.11)
Sex	0.49	0.00	1.63 (1.49-1.77)
0			1 (Reference)
1	0.34	0.00	1.40 (1.27-1.53)
2	0.74	0.00	2.10 (1.66-2.65)

Health Score III   We further refine Health Score II in order to achieve non-overlapping confidence intervals, resulting in the creation of Health Score III, a final model which categorised subjects as low, increased and high risk. This score categorised subjects into 3 groups as shown in Table 7. The aim of Health Score III was to summarise participants into distinct groups without any overlapping of 95 per cent confidence intervals.

Adjusted Cox regressions were run on Score III with results displayed in Table 8. Similarly to Score II the model is statistically significant, with a greater hazard ratio as would be expected due to the presence of more poor metrics between each neighbouring value.

In the stratified analysis, we can also see a score of 1 is no longer insignificant, a major flaw present in the previous models. Additionally, we have now managed to successfully remove the overlapping of 95 per cent confidence intervals with adjacent scores. We can interpret this as that there is over a 95 per cent probability that a arbitrary subject will be classified in the survival category that describes their survival function best.

We can now write an equation for the hazard function $\lambda(t)$ of a subject, with respect to vector X_i consisting of Health Score III, age and sex as

$$\lambda(t|X_i) = \lambda_0(t)\exp[0.35(\text{score III}_i) + 0.10(age_i) + 0.49(sex_i)], \qquad (4.3)$$

where $\lambda_0(t)$ denotes the baseline hazard function.

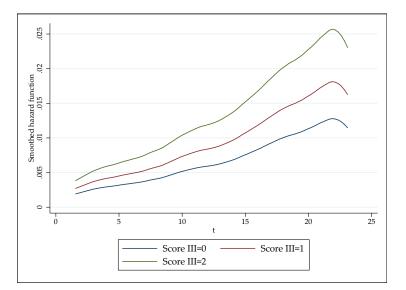The age and sex variables included are as a result of the model adjustment. Our Cox model

Figure 1: Estimated hazard functions

suggests a hazard ratio of $\exp[0.10] = 1.11$ for an increase in age of one year, so the probability of death for a arbitrary subject increase by approximately 11 per cent each year. A value of 0 for the variable sex denoted a female, and 1 a male, thus the model predicts that all else being equal, the hazard rate of a male subject will be 63 per cent higher than that of a female.

A graphical estimate of the baseline hazard function (low risk) was calculated to complement the model, along with estimates for the other categories. The baseline hazard term $\lambda_0(t)$ cannot be represented as a function in (4.3), however Stata can use standard kernel-smoothing methodology (Gray, 1990) to approximate a smooth curve to the Nelson–Aalen[4] estimate, thus making it differentiable. The resulting estimated baseline hazard rate can be viewed in Figure 1. It is represented by the blue line in the plot, for which Health Score III is 0. Hazard rates for scores of 1 and 2 are also plotted for reference. As expected, the hazard rate increases exponentially over time, noticeable through the slight convexity of the functions. The sharp drop at the end is caused by the censoring of subjects after differing lengths of time under observation, resulting from different measurement dates, and has no bearing on the true hazard function.

This graph also illustrates the consequence of the proportional-hazards assumption. It is clear that the smoothed hazard functions are proportional, and would be parallel if scaled logarithmically [5].

4.3   Diagnostics

If we are to discuss the Cox model with Health Score III, and it's ability to be used in the life insurance market, we must first run several diagnostic tests. Although the Cox model is semi-parametric, it still must be checked for mis-specification, goodness of fit, outliers and influential points.

---

[4]See Nelson (1972) and Aalen (1978) for further information.

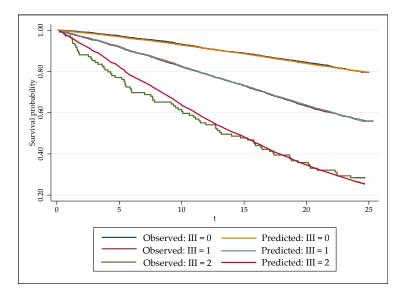[5]With a slight allowance due to the kernel-smoothing process.

11

Figure 2: Comparison of Kaplan–Meier (observed) and Cox (predicted) survival functions

The simple yet powerful *link test* was run as a general specification test for the model (Cleves et al., 2010), with no evidence of mis-specification uncovered. When Schoenfeld (1982) residuals were analysed graphically to check the proportional hazards assumption, no violation of the assumption was apparent. The assumption was further supported using a log-log plot for Health Score III, with curves appearing roughly parallel as expected.

Model agnostic observed Kaplan–Meier curves (Kaplan and Meier, 1958) were plotted alongside the predicted survival functions for each score produced by our Cox model to observe how they compared to the data. This is shown in Figure 2, with the Cox model appearing to be an excellent fit for estimating survival probability of subjects who scored 0 or 1 in Score III. There is a slight deviation between predicted and observed for those who scored 3. We would hope that this is due to the smaller sample size of this category, and perhaps due to the presence of a few outliers, rather than a deviation from proportionality. Ideally, if the sample size were to approach infinity, the observed Kaplan–Meier curves would become indistinguishable from those predicted by the Cox regression.

Cox–Snell residuals (Cox and Snell, 1968) examining the overall fit of the model showed little divergence between observed and predicted values. The predictive power of the model was evaluated with *Harrell's* C *concordance statistic* (Harrell et al., 1982), which indicated the model correctly predicting the order of survival times in 86 per cent of instances.

Martingale residuals were calculated, with no evidence in the residuals to suggest that there were covariates with incorrect functional form. Deviance residuals[6] and DFBETAs were used to determine the the influence that outliers exerted on the Score III with no individuals subjects considered highly influential (Belsley et al., 2005).

---

[6]For further information on deviance residuals see Therneau et al. (1990).

5 Discussion

The model created in this study demonstrates that insurance rating factors that can feasibly be measured by wearable devices have predictive power in relation to all-cause mortality.

5.1 Benefits of Model

A key benefit of Health Score III is its simplicity of being categorised into low, increased and high risk. The simplicity should help facilitate the possible inclusion of a similar model within current developed pricing models. Furthermore, using simple logarithm rules and the data in Table 8, we can interpret the scores in terms of years added on to the age of the insured. For each additional point added in Health Score III, an increase of 3.7 years above the the subject's true age would be equivalent. Using the hazard rates, we can also say that being considered high risk is equivalent to a low risk subject being 7.8 years older. The difference in mortality risk due to being male can be quantified as in increase in age of 5.1 years, which is slightly higher than the average difference in life expectancy of males and females of 3.7 years (Office for National Statistics, 2016). The effect on life expectancy would likely be greater at younger ages as well, due to the higher expected survival times in that population. Being able to transform the health scores into an equivalent increase in age could drastically simplify the process needed to integrate them into existing models.

Data was available in the HALS for other factors which could have been used to adjust the analyses, such as household income or socioeconomic status, however in the interests of simplicity they were not included. By only adjusting for age and sex, the predictive ability of the model could be retained, and the goal was not to address causality (Ding et al., 2015). Simplicity reasons were also used to justify the lack of weighting variables, however this imprecision would only be expected to reduce the significance of the score, and strong significance was still present.

Our model acts as a simple proof of concept to help demonstrate, in a rudimentary sense, that insurance companies may be able to utilise data from wearables as part of their premium rating process. Thus we extend our discussion to the benefits of using wearable derived data, in insurance pricing, from a general point of view.

5.2 General Benefits of Using Wearables Data in Insurance Pricing Models

The recent proliferation of wearable devices, and the resulting explosion in personal self-quantified health data, has opened up the potential for new rating factors to be included as part of current life and health insurance pricing models.

There are numerous benefits regarding the use of a wearables model. Firstly, by recording data on individuals health behaviour (e.g. biometric self-quantification data collected via wearable device technology), the information asymmetry between the policyholder and the insurer is reduced, thus enabling an enhanced granular risk differentiation based on the true risk levels of the drivers to be achieved. This potentially reduces the problems of adverse-selection, allowing the insurer to price individuals at a more personalised and accurate level, which should result in a more stable cohort of policyholders which are fairly priced. It should be noted however that there is a concern that being under obligation to provide personal data may penalise uninsurables (Yates, 2017), however it could

be argued that those who do not attempt to remain healthy penalise low risk policy holders due to the information asymmetry between insurer and consumer, and the inevitable adverse selection that comes with it (Gatzert and Wesker, 2014). Whilst more individualised pricing may open insurance cover to previously uninsurable risks (e.g. diabetics who very carefully manage their diet and exercise regimes) it is important to also consider that some restrictions on risk classification and hence an acceptable level of adverse-selection can increase loss coverage and so make insurance work better for society as a whole (Thomas, 2008).

A further significant is the potential reduction in underwriting costs borne by the insurer and consequently the policy holder. In younger and healthier age groups, costs from frequent medical examinations can actually exceed expected value of claims over the same period (Pitacco, 2014). If used in combination with other non-wearable metrics that require measurement, the select period required between examinations could be increased due to the reduced risk. The presence of more information throughout the life of the policy holder, due to the ideally continuous nature of the model, will reduce the variability of costs from their expected level (Pitacco, 2014). While this could not be achieved in this paper's model, the use of a continuous data set would not require much modification, as at this point the model coefficients are assumed to remain constant over time and only the covariates can change.

One area insurance companies have always struggled is in customer engagement, with customers considering insurance policies more of an obligation, or a "grudge purchase", rather than a product. The ability to self-monitor could be an incentive to increase good behaviours in individuals due to the increased engagement with their own health data through mobile devices (Abraham, 2016). Policies can also be tailored to individuals specific needs. These factors will be important to retain interest, as up to 50 percent of customers become disinterested and stop using their wearables within one year of purchase (Gore, 2015). Thus it could be argued that wearables may play a future role in enhancing the customer relationship, possibly even to the extent that insurance companys begin to play a greater role in the pre-claim period, by incentivising healthy behaviours. This has many implications including the potential for insurers to help with earlier identification of conditions such as diabetes and heart disease, chronic disease management and improving the obesity epidemic through financial incentives and encouragement. Clearly there is potential for enhancing the policyholder relationship as well as also ultimately bringing about interventions that ultimately lead to a reduction in claims. Wearables may also bring about a more fluid and continuous relationship between the insurer and policyholder. Historically there was little to no interaction between the two parties between point of sale and claim or renewal. Data provided on a more continuous basis with potential ensuing rewards (e.g. monthly premium discounts based on activity levels) provides the opportunity for greater "touch points" in the relationship and thus may lead to lower churn rates.

Whilst there is great potential for insurance companies to incorporate wearables into their insurance products, there are many hurdles yet to overcome. Of paramount importance are the issues of fraud detection and the questionable accuracy of many devices/metrics. At present, some wearable data is open to fraudulent reporting as individuals may be able to record data that is not indicative of their own behaviour. As can be easily imagined, this is particularly problematic in relation to metrics such as number of steps taken. Device accuracy also represents another problem. Certain metrics are currently measured consistently and accurately, via wearables, whereas other metrics show a large discrepancy. For example, a 2017 Stanford study found that energy expenditure

readings were very inaccurate, whereas in contrast, heart rate metrics were found to be within 5% of the true value for most devices (Shcherbina et al., 2017).

6 Limitations

Despite the numerous diagnostic tests to used to validate the model, the findings in this paper must be interpreted in the light of the study's limitations.

6.1 Measurement Limitations

Despite the large sample size in the HALS certain categories investigated were small, such as the high risk category for Health Score III, and to an even greater extent the higher points totals for Health Scores I and II. An increased number of subjects in these categories would allow for a more accurate calculation of model coefficients, hopefully narrowing their 95 per cent confidence intervals to a more acceptable level.

Its is also worth considering the possibility that the selection of subjects analysed themselves were biased. As there was a response rate of 73 per cent in the survey, nonparticipation bias might have affected the prevalence of associations, which would impacts their generalisability; however due to the multi-variable nature of the model, the health scores would not be affected to the same degree. In addition to this, Galea and Tracy (2007) find that lower participation rates are unlikely to have a substantial effect on exposure event associations. This suggests that associations, relative to prevalence, are less reliant on sample representativeness.

A particularly significant shortfall in the HALS was the length of the follow up period spanning only 25 years. This is simply due to the date of the initial data collection, and so the dataset will improve in time, however it was speculated that the model may only be relevant for certain segments of the population. For example, a 30 year old with high blood pressure who exercises infrequently is still unlikely to die in the next 25 years, whereas an 80 year old is likely to die over the same period regardless of the underlying health metrics they possess. Thus the effect of possessing these metrics is hidden from our dataset. In order to investigate this, the sample was stratified into age groups spanning ten years starting from 30 years old. Log-rank tests were performed for Score III within each of the age groups. The results showed that the model is only successful in predicting survival times between ages 40–80. This means that there may not be optimal data for 44.7 per cent of the participants in our dataset. In fact, these 44.7 per cent of participants only accounted for 13.1 per cent of deaths. While 100 per cent of those above 80 years old died in the follow up period, there were deaths in only 4.0 per cent of those under 40 years old. With 96.0 per cent of subjects under forty being censored in June 2009, it would be very difficult to find significant results for that population due to a very small proportion providing an exact survival time.

A final area identified in the HALS was the possibility for measurement error to take place. While several metrics were taken by a study nurse, others such as physical activity, walking and sleeping duration were not subject to the same level of scrutiny. When variables are self reported they are almost always subject to misclassification bias (Maudsley and Williams, 1996). Physical activity and walking were estimated by calculating the average amount achieved over the previous fortnight, however for many participants these two weeks may not have been representative of their

lifestyle as a whole. Something as simple as bad weather in a region could have impacted reported walking levels for its subjects. The survey question on average sleep duration was little more than a best guess by participants, and an estimate over the long term would likely be difficult for most to answer accurately. There is also the possibility that this bias could be non-random: on average subjects tend to report favourable behaviours due to social desirability bias. Fortunately, the nature of this study means that this bias will more often be towards the null (Ford et al., 2011).

6.2 Methodological Limitations

While the HALS had its limitations, the design of the model itself had its own shortcomings. Poor indicators for different metrics may have the same underlying cause, and thus the presence of a second poor metric may exaggerate the mortality risk due to the additive nature of the health score. This is particularly relevant as our model considers the metrics to be indicators of risk rather than causes. Confounding could also be in effect, which is defined as when an included variable is correlated to both the dependent and an independent variable. Walking duration is a prime example of this, as it was shown with physical activity in Section 4.2 to be individually significant in the same model, and are also likely related to one another. On the other hand, studies have shown that associations of particular pairs of metrics with mortality can be much higher than the sum of their individual associations as measured by hazard ratios (Ding et al., 2015), and so the inclusion of a score multiplier could be a useful addition in these cases. A further modification that could be made to improve the modeling of multiple metrics would be the inclusion of weighting factors to represent their effects on hazard ratios. In the Cox regression performed in Section 4.2 resting heart rate had a much higher hazard ratio (1.45) relative to high blood pressure (1.17). Allowing a greater weighting to resting heart rate in this scenario might increase the predictability of our health scores. As mentioned before, simplicity was a key aspect of the model and the capturing of these kinds of interactions were not a primary concern when there was already a strong relationship with mortality present.

A common step taken in epidemiology studies is to exclude any subjects with previous diagnoses or chronic diseases such as cancer, heart disease or stroke. This conditions could effect not just survival time, but whether a subject presents poor health metrics or not. However as this study was not investigating causation, only indicators of poor health, it was opted to leave these subjects in the main analyses. Poor metrics associated with these conditions was something we hoped to capture, as from an insurance perspective our score must be based on the likelihood of a claim for death or illness being. The presence of previous conditions affecting results in this way is known as *reverse causality*. In the interest of thoroughness a separate Cox regression was run excluding all those who possessed these conditions at measurement or died within the first two years of follow-up as per the methods of Ding et al. (2015) with results displayed in Table 9. The exclusion of 144 subjects did not effect the significance of the Score and comparison with the main analysis results in Table 8 showed little difference apart from a slight reduction of hazard ratio for a value of 2. This is more than likely a result of the sample size, with this additional regression excluding 13 of 78 deaths in this category.

In the interests of simplicity, only all-cause mortality was considered as the primary outcome, yet much information could be gained by recording cause-specific mortality or onset of particular conditions as well. Without consideration of cause of death, we may be misrepresenting the signifi-

Table 9: Results of Cox regression adjusted for reverse causality

Variable	Coefficient	P-value	HR (95% CI)
Score III	0.32	0.00	1.38 (1.27-1.51)
0			1 (Reference)
1	0.32	0.00	1.38 (1.26-1.52)
2	0.66	0.00	1.93 (1.49-2.49)

cance of our health score as deaths may not always be for health reasons. For example, the leading cause of death for 20–34 year olds was suicide, with 24 and 12 per cent of male and female deaths in this age group respectively (Office for National Statistics, 2017). These deaths, among others, would have no causal relationship with our health score.

Finally, a key attraction of using wearables to price insurance is their ability to take measurements consistently through time, allowing the insured's risk profile to be updated in real time without the need to visit a doctor. Due to the nature of the survey data used in this study the covariates in our model are assumed to remain constant in time. Future waves of follow up data could be incorporated, increasing the applicability of the model to the real world at the sacrifice of simplicity. This would help to account for behavioural or physical changes resulting in misclassification (Kvaavik et al., 2010). The stability of certain behaviours of metrics differs over time, with several studies describing the stability of physical activity over time as low or moderate (Parsons et al., 2006; Telama et al., 2005). In this we must assume that some degree of stability exists, as evidenced by the significance in the model's coefficients.

7   Conclusion

In conclusion, Health Score III acts as a proof of concept, demonstrating the potential for the inclusion of rating factors, based on wearables data, to be included in health and life insurance pricing models. The model also potentially acts as a starting point for wearable derived data inclusion in a more fully formed pricing model, especially those that wish to utilise rating factors such as resting heart rate, blood pressure, sleep duration and walking duration. The suitability of the existing metrics would require further evaluation with weighting, substitution and erasures taking place. With this in mind, there are several areas which could provide the basis for future research.

As this model only considered all-cause mortality as an event of interest, it is not directly applicable to pricing health insurance in its current form. An investigation into cause-specific mortality however would be the first movement in this direction, and inclusion of the onset of disease or other conditions would be a logical next step. It is worth noting that a larger data set would be required to provide enough occurrences of each condition to produce statistically significant results. At a certain point it would also be necessary to run the model on a continuous data set in order to better simulate the real world data it was developed for.

References

Aalen, O. (1978). Nonparametric inference for a family of counting processes. *The Annals of Statistics*, pages 701–726.

Abraham, M. (2016). Wearable technology: A health-and-care actuary's perspective. *Institute and Faculty of Actuaries*.

Azzopardi, M. and Cortis, D. (2013). Implementing automotive telematics for insurance covers of fleets. *Journal of technology management & innovation*, 8(4):59–67.

Becher, S. (2016). Wearables–a new chance for private insurance companies from the underwriting view. *Zeitschrift für die gesamte Versicherungswissenschaft*, 105(5):563–565.

Belsley, D. A., Kuh, E., and Welsch, R. E. (2005). *Regression diagnostics: Identifying influential data and sources of collinearity*, volume 571. John Wiley & Sons.

Blaxter, M. (1987). Evidence on inequality in health from a national survey. *The Lancet*, 330(8549):30–33.

Brandt, P. A. v. (2011). The impact of a mediterranean diet and healthy lifestyle on premature mortality in men and women. *The American journal of clinical nutrition*, pages ajcn–008250.

Cappuccio, F. P., D'Elia, L., Strazzullo, P., and Miller, M. A. (2010a). Quantity and quality of sleep and incidence of type 2 diabetes. *Diabetes care*, 33(2):414–420.

Cappuccio, F. P., D'Elia, L., Strazzullo, P., and Miller, M. A. (2010b). Sleep duration and all-cause mortality: a systematic review and meta-analysis of prospective studies. *Sleep*, 33(5):585–592.

Catlin, T., Lorenz, J. T., Mnstermann, B., Olesen, B., and Ricciardi, V. (2017). *Insurtechthe threat that inspires*. Mckinsey & Company, Financial Services.

Cleves, M., Gould, W., Gutierrez, R. G., and Marchenko, Y. V. (2010). *An introduction to survival analysis using Stata*. Stata Press.

Comstock, J. (2015). Fitbit adds auto-detection of biking, running, elliptical, and more.

Cox, B. D. (1988). *Health and Lifestyle Survey, 1984-1985. [data collection]*. UK Data Service. SN: 2218.

Cox, B. D., Blaxter, M., Buckle, A. L. J., Fenner, N. P., Golding, J. F., Gore, M., Huppert, F. A., Nickson, J., Roth, M., Stark, J., et al. (1987). *The health and lifestyle survey. Preliminary report of a nationwide survey of the physical and mental health, attitudes and lifestyle of a random sample of 9,003 British adults*. Health Promotion Research Trust.

Cox, D. R. et al. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society*, 34:187–220.

Cox, D. R. and Snell, E. J. (1968). A general definition of residuals. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 248–275.

Dam, R. M. v., Li, T., Spiegelman, D., Franco, O. H., and Hu, F. B. (2008). Combined impact of lifestyle factors on mortality: prospective cohort study in us women. *Bmj*, 337:a1440.

Dart, A. (2017). *The Case for Connected Wearables in Insurance.* Asia Insurance Review.

Department of Health (2016). *UK Chief Medical Officers Low Risk Drinking Guidelines.*

Ding, D., Rogers, K., van der Ploeg, H., Stamatakis, E., and Bauman, A. E. (2015). Traditional and emerging lifestyle risk behaviors and all-cause mortality in middle-aged and older adults: evidence from a large population-based australian cohort. *PLoS medicine*, 12(12):e1001917.

Ford, E. S., Zhao, G., Tsai, J., and Li, C. (2011). Low-risk lifestyle behaviors and all-cause mortality: findings from the national health and nutrition examination survey iii mortality study. *American journal of public health*, 101(10):1922–1929.

Galea, S. and Tracy, M. (2007). Participation rates in epidemiologic studies. *Annals of epidemiology*, 17(9):643–653.

Ganna, A. and Ingelsson, E. (2015). 5 year mortality predictors in 498 103 uk biobank participants: a prospective population-based study. *The Lancet*, 386(9993):533–540.

Gatzert, N. and Wesker, H. (2014). Mortality risk and its effect on shortfall and risk management in life insurance. *Journal of Risk and Insurance*, 81(1):57–90.

Georgakis, M. K., Protogerou, A. D., Kalogirou, E. I., Kontogeorgi, E., Pagonari, I., Sarigianni, F., Papageorgiou, S. G., Kapaki, E., Papageorgiou, C., Tousoulis, D., et al. (2017). Blood pressure and all-cause mortality by level of cognitive function in the elderly: Results from a population-based study in rural greece. *The Journal of Clinical Hypertension*, 19(2):161–169.

Glei, D. A., Goldman, N., Rodríguez, G., and Weinstein, M. (2014). Beyond self-reports: Changes in biomarkers as predictors of mortality. *Population and development review*, 40(2):331–360.

Gopinath, B., Flood, V. M., Burlutsky, G., and Mitchell, P. (2010). Combined influence of health behaviors on total and cause-specific mortality. *Archives of internal medicine*, 170(17):1605–1607.

Gore, R. (2015). *Insurance, Innovation and IoT: Insurers have their say on the Internet of Things.* FC Business Intelligence.

Gray, R. J. (1990). Some diagnostic methods for cox regression models through hazard smoothing. *Biometrics*, 46(1):93–102.

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.

Hakim, A. A., Petrovitch, H., Burchfiel, C. M., Ross, G. W., Rodriguez, B. L., White, L. R., Yano, K., Curb, J. D., and Abbott, R. D. (1998). Effects of walking on mortality among nonsmoking retired men. *New England Journal of Medicine*, 338(2):94–99.

Hamer, M., Bates, C. J., and Mishra, G. D. (2011). Multiple health behaviors and mortality risk in older adults. *Journal of the American Geriatrics Society*, 59(2):370.

Harrell, F. E., Califf, R. M., Pryor, D. B., Lee, K. L., and Rosati, R. A. (1982). Evaluating the yield of medical tests. *Journal of the American Medical Association*, 247(18):2543–2546.

Hilton, A. (2017). Insurtech is set to take the insurance industry by storm. *Raconteur*.

Iqbal, M. U. and Lim, S. (2006). A privacy preserving gps-based pay-as-you-drive insurance scheme. In *Symposium on GPS/GNSS (IGNSS2006)*, pages 17–21.

Jensen, M. T., Suadicani, P., Hein, H. O., and Gyntelberg, F. (2013). Elevated resting heart rate, physical fitness and all-cause mortality: a 16-year follow-up in the copenhagen male study. *Heart*, 99(12):882–887.

Jubraj, R., Watson, S., and Tottman, S. (2017). *The Rise of Insurtech*. Accenture.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282):457–481.

Khaw, K., Wareham, N., Bingham, S., Welch, A., Luben, R., and Day, N. (2008). Combined impact of health behaviours and mortality in men and women: the epic-norfolk prospective population study. *PLoS medicine*, 5(1):e12.

Knoops, K. T., de Groot, L. C., Kromhout, D., Perrin, A.-E., Moreiras-Varela, O., Menotti, A., and Van Staveren, W. A. (2004). Mediterranean diet, lifestyle factors, and 10-year mortality in elderly european men and women: the hale project. *Jama*, 292(12):1433–1439.

Kuh, D., Hardy, R., Hotopf, M., Lawlor, D. A., Maughan, B., Westendorp, R., Cooper, R., Black, S., and Mishra, G. (2009). A review of lifetime risk factors for mortality. *British Actuarial Journal*, 15(S1):17–64.

Kvaavik, E., Batty, G. D., Ursin, G., Huxley, R., and Gale, C. R. (2010). Influence of individual and combined health behaviors on total and cause-specific mortality in men and women: the united kingdom health and lifestyle survey. *Archives of internal medicine*, 170(8):711–718.

Leitzmann, M. F., Park, Y., Blair, A., Ballard-Barbash, R., Mouw, T., Hollenbeck, A. R., and Schatzkin, A. (2007). Physical activity recommendations and decreased risk of mortality. *Archives of internal medicine*, 167(22):2453–2460.

Lindström, J. and Tuomilehto, J. (2003). The diabetes risk score. *Diabetes Care*, 26(3):725–731.

Lucena, F., Barros, A. K., and Ohnishi, N. (2016). The performance of short-term heart rate variability in the detection of congestive heart failure. *BioMed research international*, 2016.

Mann, J. (2017). The ultimate guide to sleep tracking.

Maudsley, G. and Williams, E. (1996). Inaccuracy in death certification–where are we now. *Journal of Public Health*, 18(1):59–66.

Nechuta, S. J., Shu, X.-O., Li, H.-L., Yang, G., Xiang, Y.-B., Cai, H., Chow, W.-H., Ji, B., Zhang, X., Wen, W., et al. (2010). Combined impact of lifestyle-related factors on total and cause-specific mortality among chinese women: prospective cohort study. *PLoS medicine*, 7(9):e1000339.

Nelson, W. (1972). Theory and applications of hazard plotting for censored failure data. *Technometrics*, 14(4):945–966.

Office for National Statistics (2016). *National life tables, UK: 20132015. [statistical bulletin]*.

Office for National Statistics (2017). *Avoidable mortality in England and Wales: 2015. [statistical bulletin]*.

Parsons, T. J., Power, C., and Manor, O. (2006). Longitudinal physical activity and diet patterns in the 1958 british birth cohort. *Medicine and science in sports and exercise*, 38(3):547–554.

Pitacco, E. (2014). *Health Insurance: Basic Actuarial Models.* Springer International Publishing Switzerland.

Redlitz, H. (2017). Wrist-size wearables will help you keep your blood pressure in check.

Schoenfeld, D. (1982). Partial residuals for the proportional hazards regression model. *Biometrika*, 69(1):239–241.

Schwartz, J. L. and Hamilton, M. A. (2015). *Accessory overload: Wearable technologys impact on the insurance industry.* Voice Magazine.

Seccareccia, F., Pannozzo, F., Dima, F., Minoprio, A., Menditto, A., Lo Noce, C., and Giampaoli, S. (2001). Heart rate as a predictor of mortality: the matiss project. *American Journal of Public Health*, 91(8):1258–1263.

Shcherbina, A., Mattsson, C. M., Waggott, D., Salisbury, H., Christle, J. W., Hastie, T., Wheeler, M. T., and Ashley, E. A. (2017). Accuracy in wrist-worn, sensor-based measurements of heart rate and energy expenditure in a diverse cohort. *Journal of personalized medicine*, 7(2):3.

Sultan, N. (2015). Reflective thoughts on the potential and challenges of wearable technology for healthcare provision and medical education. *International Journal of Information Management*, 35(5):521–526.

Tehrani, K. and Michael, A. (2014). *Wearable Technology and Wearable Devices–Everything You Need to Know.* Wearable Devices.

Telama, R., Yang, X., Viikari, J., Välimäki, I., Wanne, O., and Raitakari, O. (2005). Physical activity from childhood to adulthood: a 21-year tracking study. *American journal of preventive medicine*, 28(3):267–273.

Therneau, T. M., Grambsch, P. M., and Fleming, T. R. (1990). Martingale-based residuals for survival models. *Biometrika*, 77(1):147–160.

Thomas, R. G. (2008). Loss coverage as a public policy objective for risk classification schemes. *Journal of Risk and Insurance*, 75(4):997–1018.

Tudor-Locke, C., Craig, C. L., Aoyagi, Y., Bell, R. C., Croteau, K. A., De Bourdeaudhuij, I., Ewald, B., Gardner, A. W., Hatano, Y., Lutes, L. D., et al. (2011). How many steps/day are enough? for older adults and special populations. *International Journal of Behavioral Nutrition and Physical Activity*, 8(1):80.

Wahlström, J., Skog, I., and Händel, P. (2015). Driving behavior analysis for smartphone-based insurance telematics. In *2nd Workshop on Physical Analytics, WPA 2015, 22 May 2015*, pages 19–24. ACM Digital Library.

Wilson, P. W. F., DAgostino, R. B., Levy, D., Belanger, A. M., Silbershatz, H., and Kannel, W. B. (1998). Prediction of coronary heart disease using risk factor categories. *Circulation*, 97(18):1837–1847.

Wong, M. L., Lau, E. Y. Y., Wan, J. H. Y., Cheung, S. F., Hui, C. H., and Mok, D. S. Y. (2012). The interplay between sleep and mood in predicting academic functioning, physical health and psychological health: a longitudinal study. *Journal of psychosomatic research*, 74(4):271–277.

World Health Organization (1995). *Physical status: The use of and interpretation of anthropometry, Report of a WHO Expert Committee.* Geneva: World Health Organization.

World Health Organization and International Society of Hypertension Writing Group (2003). 2003 world health organization (who)/international society of hypertension (ish) statement on management of hypertension. *Journal of hypertension*, 21(11):1983–1992.

Yates, H. (2017). Personal data may penalise uninsurables. *Raconteur*.

Zhang, D., Shen, X., and Qi, X. (2015). Resting heart rate and all-cause and cardiovascular mortality in the general population: a meta-analysis. *Canadian Medical Association Journal*, pages cmaj–150535.