Do they know what's at risk?
Health Risk Perception Among the Obese.

Joachim Winter*, Amelie Wuppermann†

Preliminary Version

February 2011

Abstract

While causes of obesity are multifactorial, body weight depends partly on individual behavior and thereby possibly on individual health risk perceptions. This paper empirically analyzes the accuracy of obese individuals' health risk perception. Subjective health risk perceptions for various diseases elicited in the American Life Panel (ALP) are compared to individual's objective risks of the same diseases. The results show that obese individuals significantly underestimate their 5-year risks of diabetes, arthritis or rheumatism, and hypertension. Health education programs targeted at the obese might thus help these individuals lose weight by making better-informed lifestyle choices.

JEL-Code: I10, I18, D84

Keywords: Obesity, health risk, subjective expectations

This research has profited from comments by seminar participants at the University of Mainz, the Research Institute for the Economics of Aging (MEA) Mannheim, and the Helmholtz Center Munich. Financial support from the National Institute on Aging (NIA), grant no. 2 R012 AG 20717-06A1 ("Internet Interviewing and the HRS") and from the Deutsche Forschungsgemeinschaft through GRK 801 is gratefully acknowledged.

*University of Munich, Ludwigstr. 28, 80539 Munich, Germany. Email: winter@lmu.de

†University of Mainz, Jakob-Welder-Weg 4, 55128 Mainz, Germany. Email: wuppermann@uni-mainz.de

1 Introduction

Obesity – that is being too heavy for your height as measured by a Body Mass Index (BMI) of 30 or higher[1] – is a risk factor for various diseases. In particular, it increases the risks of type 2 diabetes, cardiovascular disease, several cancers, arthritis, and psychological problems (Haslam and James, 2005; Dixon, 2010). Furthermore, obesity is responsible for a large share of medical expenditures (Cawley and Meyerhoefer, 2010). Reducing excess weight would not only benefit the affected individuals' health but could also help to improve health care finances (Goldman et al., 2010).

While the causes of obesity are multifactorial, body weight depends on each individual's behavior. Individuals can choose how many calories they consume and whether they engage in activities with high or low calorie expenditure, for example they can choose whether they are physically active. Despite the health risks associated with obesity and the relation of obesity to individual behavior, more than 15 percent of adults are obese in the average OECD country and more than 30 percent of adults are obese in the US (Sassi, 2010). One possible explanation for this is that individuals who carry excess weight are not aware of the danger that this weight poses to their health and thus underestimate their personal future costs of being obese.

In this paper we analyze whether middle-aged individuals in the US with excess body weight underestimate their personal risks of diabetes, stroke, heart attack, lung disease, hypertension and arthritis or rheumatism. Health risk perceptions among obese individuals have been studied by Kan and Tsai (2004) and Gregory et al. (2008). Our analysis contributes to this literature by measuring individual risk perceptions in a different way. In particular, while Kan and Tsai (2004) and Gregory et al. (2008) use information on individual risk perception stemming from qualitative questions, we ask individuals about numerical probabilities of developing different diseases in the future. An advantage of our numerical subjective measures is that they can be compared to objective probabilities. This allows us to investigate how accurate individual risk perceptions are.

Our data on subjective disease probabilities comes from the American Life Panel (ALP), a panel study of US American adults administered by RAND. In an ongoing

[1] The BMI is measured as weight in kg divided by the square of height in meters $\frac{weight(kg)}{height^2(m^2)}$.

survey in the ALP that went into the field in the summer of 2010 we ask the survey participants to assess their chances of developing different diseases in the next 5 years on a 0-100 scale. In the design of these health expectations questions we follow Manski (2004) who also provides a general assessment of the validity of similar expectation questions in different domains, such as stock market returns and returns to schooling.

In order to gauge how well individuals are informed about their health risks we estimate objective probabilities of developing the different diseases in the next 5 years. More specifically, we estimate relationships between individual characteristics and future disease onsets in a different panel study, the Health and Retirement Study (HRS). Under the assumption that these relationships are the same in the ALP today as they were in the HRS a few years ago, we use the estimates from the HRS to predict the probabilities of disease onsets for individuals in the ALP.

Comparing individual subjective to objective disease probabilities, we find that obese individuals in the ALP underestimate certain health risks. In particular, we find evidence that obese individuals underestimate the risks of developing arthritis or rheumatism, diabetes and hypertension within the next 5 years. The risks of developing other diseases, such as a heart attack or a stroke, however, are significantly overestimated by the obese. As the objective risks of developing these diseases are relatively small on average, even among obese individuals, this overestimation corresponds to results from other studies that find that individuals tend to overestimate the probabilities of low probability events (see e.g. Lichtenstein et al., 1978).

The significant underestimation of the risks of arthritis, diabetes and hypertension among the obese indicates that there is a need for health education in this group. While health education programs targeted at the obese will probably not suffice to control the obesity epidemic, our results indicate that there is room for health education programs to help these individuals make better-informed lifestyle choices.

The paper is structured as follows. Section 2 summarizes the related literature. Section 3 describes the two data sets that we use in this analysis. The empirical methods are explained in Section 4. Section 5 presents the results. Section 6 explores the robustness of the results, and Section 7 concludes.

2 Related Literature

Our analysis is closely related to two strands of literature. The first strand focuses directly on risk perceptions among the overweight and obese. The second strand analyzes risk perception among smokers.

The different studies that analyze risk perception among the overweight and obese focus primarily on the relationship between risk knowledge and the tendency to be overweight or obese rather than on the accuracy of the risk knowledge. Kan and Tsai (2004) find that there is a relationship between individuals' perception of health risks resulting from excess weight and the tendency to carry excess weight in Taiwanese men. In particular, men with very high BMI seem to be less aware of obesity-related health risks. For women, the authors find no relationship between health risk perception and BMI. Similarly, Gregory et al. (2008) find that overweight and obese adults in the US who do not perceive their weight as a risk factor are less likely to engage in weight lowering activities. Furthermore, the authors show that significant fractions of overweight and obese adults do not agree to the statement that their weight is a health risk. This suggests a lack of health knowledge on weight as a risk factor. However, the question that measures health knowledge is targeted at overall health risks and is thus very general. It does not allow to distinguish between risks of different diseases. Moreover, because of the question's qualitative nature it is not clear how accurate individual risk perceptions are.

Analyses of risk perception among smokers have focused on the accuracy of the risk perception. Viscusi (1990) pioneered this type of analysis in economics by comparing individuals' assessments of the probability that smokers get lung cancer because they smoke with objective measures of this probability. His results indicate that smokers and non-smokers significantly overestimate the risk of lung cancer due to smoking. Viscusi and Hakes (2008) extend the earlier analysis to the accuracy of individuals' assessments of the probability that smokers die from lung cancer, heart disease, or any other illness because they smoke, and to the accuracy of the assessment of loss in life expectancy due to smoking. In line with Viscusi's earlier findings their results indicate that smokers and non-smokers overestimate the risk of dying from smoking and the years of life lost due to smoking when compared to the risks of dying from smoking and life years lost as reported in the medical literature. As smokers overestimate these risks less than non-smokers, the results are consistent with the idea that higher risk perceptions protect individuals from smoking.

While Viscusi (1990) and Viscusi and Hakes (2008) study perceptions of general risks of smoking, i.e. individuals are asked about the risks of a hypothetical smoker, Schoenbaum (1997) and Khwaja et al. (2009) analyze the accuracy of risk estimates concerning the individuals' personal risks. Schoenbaum (1997) compares individual expectations of reaching age 75 measured in the HRS with actuarial predictions based on life tables for the groups of never smokers, former smokers, current light smokers, and current heavy smokers. Heavy smokers are found to significantly overestimate the probability of reaching age 75, while the other groups have about accurate risk perceptions. In Khwaja et al. (2009), individual expectations of reaching age 75 and expectations of getting certain diseases until the age of 75 are compared with individual objective risk estimates for each event. The objective risk estimates are based on the relationship between individual characteristics and disease probabilities as estimated from the HRS. Their results indicate that smokers do not underestimate disease and mortality risks.

3 Data

The analysis in this paper is based on two data sets. The main analysis is conducted with data from the American Life Panel (ALP) in which we elicit subjective risk expectations. Data from the Health and Retirement Study (HRS) is used in the calculation of objective health risks for individuals in the ALP.

The HRS was started in 1992 as a representative study of the US non-institutionalized population born between 1931 and 1941. Blacks, Hispanics and residents of Florida were oversampled. Sampling weights are provided to correct for this oversampling. Data was collected every two years since 1992. The last currently available data is from 2008. Additional cohorts were added to the HRS after the first wave. In particular, in 1998 the "children of depression (CODA)" cohort (born 1924-1930) and the "war baby" cohort (born 1942-1947) were added to the data. We use the 1992 HRS wave and follow-up until 2008 to estimate prediction models for onset of diseases for a population aged 50-62. As a robustness check we also use a later HRS wave as a baseline year in the risk prediction models. In particular, we use the 1998 wave. The integration of additional cohorts in this wave allows to focus the analysis on a representative sample for the same age group (50-62 year-olds) as when the

1992 wave is used as baseline.

The second data source is the American Life Panel (ALP), an internet survey of about 3,200 American adults administered by RAND (see *http://rand.org/labor/roybalfd/american_life.html* for a full description). We implemented a survey in the ALP that went into the field in July 2010 and is still ongoing. The focus of our survey lies primarily on eliciting individuals' subjective expectations of developing certain diseases in the future. We elicit the expectations as numerical probabilities with the question "What do you think is the percent chance that you will develop, or re-develop if you have already been diagnosed with it, the following conditions in the next 5 years and ever in you lifetime?".[2] [3] The question is preceded by a text that explains the probability scale from 0 to 100, where 0 means that there is absolutely no chance, or 0 percent, and 100 means the event is absolutely sure to happen, or 100 percent. Similar types of questions on survival expectations have proven useful in eliciting subjective probabilities that have predictive power for actual outcomes (see Hurd, 2009, for a summary).

In addition to the subjective expectations we elicit information on risk factors for the specific diseases, including questions on co-morbidities, family history, and lifestyle. 915 of the surveyed individuals in our sample are between 50 and 62 years old at the time of the survey. While the ALP is not a representative sample of the US population, RAND provides weights to make it comparable to the Current Population Survey. These weights are used in all following analyses.

In table 1 weighted means and standard deviations of relevant health variables, risk factors and demographics are displayed for individuals aged 50 to 62 in the ALP. For means of comparison, descriptives are also displayed for the HRS 1992 sample.

While the HRS and ALP samples seem to be similar in age and sex, there are differences in BMI categories, smoking status, and education. Fewer individuals in the ALP sample are current or former smokers, but more of them are overweight or obese. These differences align with changes in the prevalence of smoking and obesity over the last decades. While the fraction of smokers decreased, the fraction

[2] Individuals who report that they have been diagnosed with chronic diseases, such as diabetes or chronic lung disease, are not asked about their chances of developing the specific conditions. For conditions that may relapse, like a stroke or a heart attack, all individuals are asked about the chances independent of their prior history of the disease.

[3] In this paper, we focus on the 5 year risks. In a follow-up analysis, we plan to also investigate the accuracy of the lifetime risk perceptions.

Table 1: Descriptive Statistics

	HRS 1992		ALP	
Variable	Mean	SD	Mean	SD
Self-rated health				
Excellent	0.24	0.43	0.08	0.32
Very Good	0.30	0.47	0.41	0.61
Good	0.27	0.44	0.33	0.58
Fair	0.13	0.34	0.15	0.44
Poor	0.07	0.26	0.03	0.22
BMI				
Normal weight (BMI<25)	0.35	0.48	0.27	0.55
Overweight (25≤BMI<30)	0.41	0.49	0.36	0.59
Obese 1 (30≤BMI<35)	0.16	0.37	0.22	0.51
Obese 2 (35≤BMI<40)	0.04	0.20	0.07	0.31
Obese 3 (BMI≥40)	0.02	0.14	0.08	0.33
Smoking Status				
Current Smoker	0.27	0.44	0.21	0.50
Former Smoker	0.37	0.48	0.32	0.57
Never Smoker	0.36	0.48	0.47	0.62
Demographics				
Age	55.56	3.25	55.55	4.52
White	0.86	0.345	0.84	0.45
Male	0.48	0.50	0.49	0.62
Married	0.74	0.44	0.68	0.57
Education				
Less than High School	0.23	0.42	0.06	0.28
High School or equiv	0.39	0.49	0.33	0.58
Some College	0.20	0.40	0.30	0.56
BA or equiv	0.11	0.31	0.20	0.49
More than BA	0.08	0.27	0.12	0.40
N	9,793		906	

Notes: Sampling weights used in estimation of means and standard deviations (SD).

of individuals with excess weight increased.[4] The individuals in the ALP 2010 are furthermore on average better educated than the individuals in the HRS in 1992. This could reflect the steady advance in educational accomplishments in

[4]See National Center for Health Statistics (2010), page 24ff. for trends in smoking and obesity in the US.

the US.[5] For the results of our analysis the differences in individual characteristics between the samples do not matter as long as the relationship between the individual characteristics and the probabilities of developing different diseases in the future are constant over time. We shed light on the question how strong this assumption is in the aforementioned robustness check in which we use a later HRS wave as baseline.

The first two columns of table 2 display averages and standard errors of the subjective disease probabilities in the ALP for individuals who are between 50 and 62 years old. For each disease, only individuals are included who do not report ever having been diagnosed with the respective disease. In addition to the overall averages, means of the subjective probabilities are shown for the different BMI categories. It can be observed that the overall means vary between conditions. The means range from an 8 percent risk of developing lung disease in the next 5 years to a 21 percent risk of developing arthritis or rheumatism. Furthermore, the average risks vary by categories of BMI. For example, individuals with a BMI between 30 and 35 on average rate their risk of diabetes in the next 5 years more than 2 times as high as individuals with a BMI lower than 25. For none of the conditions, however, there is a clear graded relationship between average subjective risk and BMI.

4 Estimation Methods

In order to evaluate the accuracy of individuals' subjective beliefs on different risks we construct objective risks of getting different diseases in the next 5 years for every individual in the ALP. In a second step, we analyze whether the difference between the subjective and objective risk is related to the individuals' BMI.

4.1 Calculation of Objective Risk

For the calculation of the objective risks we use data from the HRS to estimate the relationship between individual characteristics at a baseline year and the observed onset of diseases in the future. Similar to Khwaja et al. (2009) we use d, $d \in \{1, ..., D\}$ duration models to model the relationship between the characteristics and

[5] The trend in educational attainment can be seen in data from the U.S. Census available at *www.census.gov/hhes/socdemo/education/data/cps/historical/index.html.* The data shows that in 1992 25.7% of the 35 to 55 year old individuals and 14.2% of the individuals aged 55 or above in the US had finished their education with 4 years of college or more compared to 27.3% of the younger and 26.4% of the older group in 2009.

Table 2: Subjective and Objective Risk

		Subjective Risk		Objective Risk	
		Mean	SE	Mean	SE
Diabetes	Total	10.58	0.72	5.34	0.36
	BMI<25	7.14	1.04	1.22	0.12
	25≤BMI<30	10.45	1.09	3.89	0.27
	30≤BMI<35	15.15	2.04	8.79	0.71
	35≤BMI<30	10.48	1.64	12.76	1.48
	BMI≥40	14.44	3.81	17.62	3.34
Stroke	Total	12.37	0.81	1.91	0.18
	BMI<25	10.05	1.37	1.41	0.17
	25≤BMI<30	11.28	1.06	1.52	0.18
	30≤BMI<35	14.88	2.12	2.12	0.29
	35≤BMI<40	18.44	3.50	2.47	0.55
	BMI≥40	13.37	3.75	4.54	1.30
Lung Disease	Total	8.14	0.62	2.55	0.17
	BMI<25	7.71	1.21	2.39	0.25
	25≤BMI<30	8.70	1.18	2.19	0.18
	30≤BMI<35	9.00	1.19	2.70	0.32
	35≤BMI<40	6.22	1.32	3.37	0.61
	BMI≥40	5.95	1.18	3.84	0.81
Heart Attack	Total	13.07	0.79	3.13	0.25
	BMI<25	9.90	1.38	1.99	0.23
	25≤BMI<30	13.81	1.17	3.07	0.31
	30≤BMI<35	14.35	2.06	4.13	0.49
	35≤BMI<40	17.07	2.79	4.17	0.90
	BMI≥40	14.13	3.19	3.99	1.12
Hypertension	Total	11.53	0.94	12.80	0.49
	BMI<25	9.00	1.51	9.29	0.37
	25≤BMI<30	11.54	1.25	12.47	0.53
	30≤BMI<35	11.51	2.14	15.45	0.79
	35≤BMI<40	23.06	5.30	18.33	1.52
	BMI≥40	15.48	5.43	21.75	3.31
Arthritis	Total	21.14	1.37	21.06	0.72
	BMI<25	23.16	2.57	15.63	0.64
	25≤BMI<30	20.32	2.13	19.44	0.68
	30≤BMI<35	19.84	2.64	23.85	1.18
	35≤BMI<40	19.58	4.85	30.04	2.68
	BMI≥40	22.77	8.17	37.32	4.99

Notes: Individuals in the ALP aged 50-62 who have not been diagnosed with the respective disease before. Weights used in calculating means and standard errors (SE). SE of the objective risks are bootstrapped using 500 bootstrap replications. Subjective expectation on developing arthritis/rheumatism is a combination of subjective expectation on arthritis/rheumatism and on rheumatoid arthritis (see appendix for details).

the time to onset of each disease, d. We assume that the survivor functions follow Weibull distributions and allow for Gamma distributed unobserved heterogeneity. The survivor function for disease d is

$$S(t_i^d; \mathbf{X_i}, \boldsymbol{\theta}^d, \mu^d | \eta_i^d) = \Big(exp(-\lambda_i^d (t_i^d)^{\mu^d})\Big)^{\eta_i^d} \tag{1}$$

where λ_i^d is parameterized as $exp(-\mu^d \mathbf{X_i}' \boldsymbol{\theta}^d)$ so that the survival distribution is a Weibull in accelerated failure time (AFT) metric. η_i^d stands for the unobserved heterogeneity that is Gamma($\frac{1}{\sigma^d}$,σ^d) distributed. t_i^d represents the time until individual i is diagnosed with disease d. If individual i does not report a diagnosis of the disease within the time that she is observed in the HRS, i is treated as a right-censored observation. In this case, t_i^d contains the time to censoring, i.e. the time in which i is observed.[6] μ^d is the shape parameter of the Weibull distribution, $\mathbf{X_i}$ represents the vector of individual characteristics and includes information on smoking status, BMI categories, self-rated health, age, sex, educational degree, marital status, and race. $\boldsymbol{\theta}^d$ is the corresponding vector of coefficients.

We estimate μ^d, $\boldsymbol{\theta^d}$, and σ^d by maximum likelihood. Based on these parameter estimates, we calculate each individual i's probability of not getting disease d in the next t years as

$$\widehat{O_i^d}(t; \mathbf{X_i}, \hat{\boldsymbol{\theta}}^d, \hat{\mu}^d) = E_\eta \left(\widehat{O_i^d}(t; \mathbf{X_i}, \hat{\boldsymbol{\theta}}^d, \hat{\mu}^d | \hat{\eta}^d)\right) = \left[1 + \hat{\sigma}^d exp(-\hat{\mu}^d \mathbf{X_i}' \hat{\boldsymbol{\theta}}^d) t^{\hat{\mu}^d}\right]^{-\frac{1}{\hat{\sigma}^d}} \tag{2}$$

Naturally, the probability of getting the disease within the next t years is then

$$\widetilde{O_i^d} = 1 - \widehat{O_i^d}(t; \mathbf{X_i}, \hat{\boldsymbol{\theta}}^d, \hat{\mu}^d). \tag{3}$$

As the time horizon of the subjective risk is 5 years, we set $t = 5$.

We have to make one main assumption to use the estimated relationship for the calculation of the objective risks in the ALP. As above-mentioned, we assume that the relationships between the characteristics and disease onsets stay constant over time and thus are the same in the ALP today as they were in the HRS. The validity of this assumption is evaluated by using different HRS waves as baseline years in robustness analyses.

[6]The time to disease onset, t_i^d, is measured in years. A report of a new diagnosis in wave A is coded as the time in years between wave 1 and wave A minus 1. As there are two years between consecutive waves, t_i^d is coded as 5, for example, in the case of a new report of a disease d by individual i in wave 4. For right-censored observations t_i^d is measured in years between 1992 and the year of the observation i's last interview.

4.2 Comparison Subjective and Objective Risk

In the next step, we compare the subjective and the objective disease risks. Following Khwaja et al. (2009), we regress the difference between the subjective probability and the objective probability on the different BMI categories with normal- or underweight as reference

$$S_i^d - \widetilde{O_i^d} = \gamma_0^d + \gamma_1^d OV_i + \gamma_2^d OB1_i + \gamma_3^d OB2_i + \gamma_4^d OB3_i + \epsilon_i^d \quad (4)$$

where S_i^d indicates the subjective risk of developing disease d within the next 5 years, OV is a dummy for having a BMI between 25 and 30, and $OB1$, $OB2$, and $OB3$ stand for the 3 categories of obesity, $30 \leq$ BMI < 35, $35 \leq$ BMI $<$ 40, and BMI$\geq$ 40, respectively. If γ_0^d is significantly positive, individuals who are normal- or underweight significantly overestimate their risk of developing disease d. Significantly positive estimates for $\gamma_1^d, \gamma_2^d, \gamma_3^d$ or γ_4^d indicate that in the respective group the difference between subjective and objective risk is on average larger than among the normal- or underweight individuals. Individuals in group k on average underestimate their risk of disease d if $\gamma_0^d + \gamma_k^d$ is significantly smaller than 0.

In addition to these simple comparisons of means between the groups, we include information on income and education in equation (4) to analyze whether differences in these socio-economic characteristics can explain differences in the accuracy of risk perception between different BMI groups.

5 Results

Table 3 reports coefficients and standard errors after estimation of the model specified in equation (1) for the different diseases. As the models are specified in accelerated failure time metric, a positive coefficient indicates that the time to disease onset increases with the associated variable, i.e. the risk decreases with the variable. Conversely, a negative coefficient indicates that the risk increases with an increase in the associated variable. The risks of all diseases except diabetes and hypertension significantly increase with age. The disease risks are higher for worse assessments of self-rated health, and they increase with smoking and BMI categories. Lower educated individuals are at higher risk for diabetes, lung disease and hypertension, and men are at higher risk for diabetes, stroke and a heart attack, while women have a higher risk of arthritis or rheumatism.

Table 3: Duration Model for Disease Onsets in HRS

	Diabetes		Stroke		Lung Disease		Heart Attack		Hypertension		Arthritis	
Age	-0.008	(0.006)	-0.051***	(0.009)	-0.022***	(0.007)	-0.033***	(0.011)	-0.004	(0.004)	-0.015***	(0.005)
Male	-0.185***	(0.043)	-0.131**	(0.061)	0.048	(0.051)	-0.589***	(0.078)	0.039	(0.026)	0.307***	(0.035)
SRH – very good	-0.103*	(0.060)	-0.165*	(0.090)	-0.320***	(0.087)	-0.421***	(0.110)	-0.025	(0.033)	-0.257***	(0.043)
SRH – good	-0.215***	(0.062)	-0.459***	(0.089)	-0.554***	(0.087)	-0.713***	(0.111)	-0.067*	(0.035)	-0.407***	(0.049)
SRH – fair	-0.359***	(0.074)	-0.700***	(0.111)	-0.829***	(0.097)	-1.155***	(0.132)	-0.109**	(0.047)	-0.418***	(0.067)
SRH – poor	-0.491***	(0.097)	-0.943***	(0.130)	-0.973***	(0.114)	-1.290***	(0.171)	-0.113*	(0.064)	-0.433***	(0.094)
White	0.180***	(0.051)	0.075	(0.075)	-0.356***	(0.068)	-0.282***	(0.093)	0.183***	(0.032)	-0.026	(0.043)
Married	0.047	(0.047)	0.087	(0.066)	0.054	(0.057)	-0.107	(0.086)	-0.038	(0.030)	-0.104**	(0.041)
Less than HS	-0.181***	(0.053)	0.038	(0.076)	-0.110*	(0.061)	-0.097	(0.091)	-0.049	(0.032)	-0.001	(0.046)
Some College	-0.025	(0.057)	-0.008	(0.080)	-0.048	(0.070)	-0.036	(0.094)	0.057	(0.035)	0.030	(0.045)
BA or eq.	-0.041	(0.076)	0.024	(0.105)	0.199*	(0.111)	-0.095	(0.133)	0.047	(0.045)	0.079	(0.057)
More than BA	0.095	(0.087)	0.183	(0.134)	-0.025	(0.117)	0.177	(0.158)	0.099*	(0.051)	-0.003	(0.065)
Current Smoker	-0.145***	(0.052)	-0.418***	(0.076)	-1.304***	(0.079)	-0.792***	(0.097)	-0.024	(0.032)	-0.090**	(0.044)
Former Smoker	-0.001	(0.048)	-0.149	(0.074)	-0.562***	(0.077)	-0.219**	(0.086)	0.009	(0.030)	-0.113***	(0.039)
$25 \leq BMI < 30$	-0.736***	(0.060)	-0.019	(0.068)	-0.005	(0.059)	-0.318***	(0.087)	-0.210***	(0.029)	-0.185***	(0.039)
$30 \leq BMI < 35$	-1.241***	(0.072)	-0.170*	(0.089)	-0.108	(0.073)	-0.519***	(0.107)	-0.351***	(0.038)	-0.371***	(0.050)
$35 \leq BMI < 40$	-1.488***	(0.094)	-0.171	(0.138)	-0.190*	(0.115)	-0.510***	(0.183)	-0.471***	(0.060)	-0.556***	(0.094)
$BMI \geq 40$	-1.705***	(0.153)	-0.493**	(0.194)	-0.222	(0.149)	-0.451**	(0.229)	-0.546***	(0.111)	-0.698***	(0.156)
Constant	4.922***	(0.363)	7.294***	(0.538)	7.003***	(0.449)	7.765***	(0.641)	3.258***	(0.206)	4.080***	(0.272)
μ	1.617	(0.059)	1.764	(0.105)	1.318	(0.041)	1.418	(0.069)	1.474	(0.025)	1.260	(0.032)
σ	1.121	(0.357)	4.907	(1.768)	0.000002	(0.000003)	4.438	(1.399)	0.0000001	(0.00000002)	0.450	(0.165)
N	8,850		9,552		9,876		9,223		6,825		6,872	

* $p<0.10$, ** $p<0.05$, *** $p<0.01$

Notes: Robust standard errors in parentheses. Weibull AFT parameterization with Gamma distributed heterogeneity. μ shape parameter of Weibull, σ parameter of Gamma distribution. Small values of $\widehat{\sigma}$ indicate negligible heterogeneity. In this case, models without frailty would deliver almost identical results. Estimates based on HRS 1992 cohort aged 50-62 and follow up until 2008. Each estimation only includes individuals who do not report having been diagnosed with specific disease before 1992. Results are weighted using sampling weights provided with the data.

Based on the estimates in table 3, objective disease risks for the individuals in the ALP are calculated according to equations (2) and (3). The third column of table 2 reports the average objective probabilities for the sample of 50-62 year-olds in the ALP. Bootstrapped standard errors of the average probabilities that account for the additional variation introduced through the estimation of the relationship between individual characteristics and disease onset are presented in the last column. Averages of objective probabilities are also displayed for the different categories of BMI. The average objective risks show a clear graded relationship with BMI. All disease risks are on average higher for higher BMI categories.

Table 4 reports results of the estimation of equation (4) for the different diseases. The first column of results for each disease reports estimates when only indicators for the different BMI categories are included as explanatory variables. In the reference category, i.e. among individuals who have a BMI smaller than 25, the subjective probabilities of all disease onsets but hypertension are on average significantly larger than the respective objective probabilities. Normal- and underweight individuals, for example, overestimate the risk of being diagnosed with diabetes within the next 5 years by 5.9 percentage points on average. For the other diseases the overestimation ranges from 5.3 percentage points for lung disease to 8.6 percentage points for stroke.

The results for the overweight and obese individuals are mixed for the different diseases. Similar to the normal- and underweight, they seem to correctly asses their risk of hypertension and overestimate the risks of a stroke, a heart attack, and the onset of chronic lung disease within the next 5 years. The risk of diabetes is also overestimated by the overweight and mildly obese ($30 \leq \text{BMI} < 35$). The two highest BMI categories, however, seem to assess their risk accurately on average. In these two categories, the sum of the group's coefficient and the constant is not significantly different from 0.

As the objective risks of most diseases for the normal- and underweight and of chronic lung disease, heart attack and stroke among the overweight and obese are relatively small on average, the general pattern of overestimation is in line with other results in the literature according to which individuals tend to overestimate the probabilities of small probability events (Lichtenstein et al., 1978). As diabetes risk increases significantly for the higher BMI groups, this might also explain why obese individuals have on average more accurate assessments of their risks of diabetes than the normal- and underweight.

Table 4: Differences Subjective and Objective Risk

	Diabetes		Stroke		Lung Disease		Heart Attack		Hypertension		Arthritis	
Overweight (25≤BMI<30)	0.006	0.006	0.011	0.01	0.012	0.011	0.028	0.029*	-0.006	-0.009	-0.067**	-0.066*
	(0.015)	(0.014)	(0.017)	(0.016)	(0.017)	(0.016)	(0.018)	(0.017)	(0.019)	(0.018)	(0.033)	(0.034)
Obese 1 (30≤BMI<35)	0.004	0.006	0.041	0.034	0.01	0.007	0.023	0.02	-0.036	-0.032	-0.115***	-0.109***
	(0.023)	(0.021)	(0.025)	(0.025)	(0.017)	(0.017)	(0.025)	(0.023)	(0.027)	(0.022)	(0.037)	(0.038)
Obese 2 (35≤BMI<40)	-0.082***	-0.083***	0.073**	0.07**	-0.025	-0.029	0.05	0.053*	0.05	0.061	-0.18***	-0.191***
	(0.022)	(0.021)	(0.037)	(0.035)	(0.017)	(0.018)	(0.032)	(0.031)	(0.055)	(0.048)	(0.054)	(0.06)
Obese 3 (BMI≥40)	-0.091**	-0.09**	0.002	-0.007	-0.032*	-0.036*	0.022	0.019	-0.06	-0.051	-0.221***	-0.212***
	(0.041)	(0.035)	(0.041)	(0.039)	(0.017)	(0.019)	(0.036)	(0.03)	(0.056)	(0.051)	(0.081)	(0.072)
Constant	0.059***	0.072***	0.086***	0.059***	0.053***	0.024*	0.079***	0.049**	-0.003	0.026	0.075***	0.019
	(0.011)	(0.021)	(0.014)	(0.014)	(0.013)	(0.013)	(0.014)	(0.013)	(0.015)	(0.025)	(0.026)	(0.03)
Education	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Income	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
N	779	779	861	860	828	827	861	860	536	535	585	584
R^2	0.0333	0.0897	0.0148	0.0537	0.0097	0.056	0.0072	0.0562	0.0187	0.17	0.0511	0.088

* p<0.10, ** p<0.05, *** p<0.01

Notes: Standard errors in parentheses. Each estimation is based on individuals in the ALP aged 50-62 who do not report having been diagnosed with the respective condition. Results weighted using ALP sampling weights. Subjective expectation on the onset of arthritis is combination of subjective expectation on onset of arthritis/rheumatism and subjective expectation on onset of rheumatoid arthritis (see appendix for details). Number of observations drops by 1 with inclusion of income for all conditions but diabetes, as there is one individual that has diabetes but no other disease and missing income information.

Interestingly the results in table 4 suggest that obese individuals underestimate one health risk, namely the risk of developing arthritis or rheumatism in the next 5 years. On average, individuals with a BMI between 35 and 40 underestimate the risk by 10 percentage points and individuals in the highest BMI group underestimate the risk by about 15 percentage points.

The coefficients in the second column for each disease in table 4 result from estimation of equation (4) with information on an individual's educational attainment and on household income included in addition to indicators of BMI categories. The results are essentially unchanged. Neither differences in educational attainment nor differences in income can thus explain the underestimation of arthritis risk among the obese.

6 Robustness Analyses

In this section, we present several robustness analyses. First, we use estimates based on a later HRS wave to calculate the objective risks in the ALP. Second, we employ probit models instead of duration models to estimate the relationship between individual characteristics and disease onset in the HRS. Third, we report estimates using the relative difference between subjective and objective risk instead of the absolute difference as dependent variable.

The upper panel of table 5 displays coefficients and standard errors after estimation of equation (4) where the objective risk measure is based on HRS 1998 and follow-up instead of on the first HRS wave.[7] In order to ensure comparability to the main analysis, only individuals aged 50-62 from the HRS 1998 are used to estimate model (1). Using a later cohort in the HRS as baseline helps us to investigate whether and how changes in the relationship between individual characteristics and disease onset over time affect our results. Two important differences compared to the main results emerge. First, obese individuals do not only underestimate the risk of arthritis, but also the risk of diabetes. Second, the risk of being diagnosed with hypertension in the next 5 years is significantly underestimated among all BMI categories on average. The underestimation of the risk of hypertension is especially pronounced among the overweight and mildly obese.

[7]The latest available wave that allows 5 years follow-up is wave 6, elicited in 2002. When information from this wave and follow-up is used, results are qualitatively and quantitatively unchanged from the ones displayed in the upper panel of table 5.

Table 5: Robustness Analyses I

	Diabetes	Stroke	Lung Disease	Heart Attack	Hypertension	Arthritis
	Objective Risk based on duration model with HRS 1998 and follow-up					
Overweight (25≤BMI<30)	-0.005	0.009	0.014	0.034*	-0.049**	-0.042
	(0.015)	(0.017)	(0.017)	(0.018)	(0.02)	(0.033)
Obese 1 (30≤BMI<35)	-0.022	0.046*	0.01	0.033	-0.075***	-0.139***
	(0.023)	(0.025)	(0.018)	(0.024)	(0.028)	(0.037)
Obese 2 (35≤BMI<40)	-0.134***	0.059	-0.025	0.056*	0.008	-0.193***
	(0.023)	(0.037)	(0.018)	(0.033)	(0.053)	(0.054)
Obese 3 (BMI≥40)	-0.128***	0.007	-0.055***	0.02	-0.062	-0.236***
	(0.041)	(0.041)	(0.018)	(0.036)	(0.053)	(0.083)
Constant	0.055***	0.083***	0.048***	0.076***	-0.041***	0.053**
	(0.01)	(0.014)	(0.013)	(0.014)	(0.015)	(0.026)
N	779	861	828	861	536	585
R^2	0.0643	0.0136	0.018	0.0101	0.0326	0.0668
	Objective Risk based on probit model with HRS 1992 and follow-up					
Overweight (25≤BMI<30)	-0.002	0.014	0.018	0.03	-0.035*	-0.067**
	(0.015)	(0.017)	(0.017)	(0.018)	(0.019)	(0.033)
Obese 1 (30≤BMI<35)	-0.028	0.04	0.007	0.015	-0.078***	-0.114***
	(0.023)	(0.025)	(0.017)	(0.025)	(0.027)	(0.037)
Obese 2 (35≤BMI<40)	-0.128***	0.078**	-0.027	0.042	0.02	-0.221***
	(0.027)	(0.037)	(0.018)	(0.032)	(0.055)	(0.057)
Obese 3 (BMI≥40)	-0.177***	0.007	-0.014	0.029	-0.186***	-0.332***
	(0.043)	(0.041)	(0.017)	(0.036)	(0.06)	(0.084)
Constant	0.06***	0.083***	0.046***	0.073***	-0.011	0.037
	(0.011)	(0.013)	(0.013)	(0.014)	(0.015)	(0.026)
N	779	861	828	861	536	585
R^2	0.083	0.0151	0.0089	0.0067	0.0789	0.0899

* p<0.10, ** p<0.05, *** p<0.01

Notes: Standard errors in parentheses. Dependent variables are differences between individual subjective and objective risks. Each estimation is based on individuals in the ALP aged 50-62 without a diagnosis of the specific condition prior to the ALP interview. Results weighted to take into account sample selection. Subjective expectation on the onset of arthritis is combination of subjective expectation on arthritis/rheumatism onset and subjective expectation of rheumatoid arthritis onset (see appendix for details).

As the measures of disease onset in the HRS rely on self-reports instead of for example on results from a medical examination, these results could reflect a general trend in the reduction of the number of undiagnosed cases of diabetes and hypertension over time.[8] Furthermore, an increased awareness of the health problems related to obesity might have made diagnoses of diabetes and hypertension particularly more likely among the higher BMI group. This could explain why the calculated objective risks of hypertension and diabetes are higher for the higher BMI categories when they are based on later HRS waves.

Table 6: Robustness Analyses II

	Diabetes	Stroke	Lung Disease	Heart Attack	Hypertension	Arthritis
Overweight (25≤BMI<30)	-3.265***	0.401	0.479	-2.112	-0.050	-0.459**
	(0.884)	(1.752)	(1.184)	(1.445)	(0.184)	(0.216)
Obese 1 (30≤BMI<35)	-4.260***	-0.621	0.493	-3.834***	-0.190	-0.684***
	(0.875)	(1.912)	(1.338)	(1.452)	(0.218)	(0.211)
Obese 2 (35≤BMI<40)	-5.199***	1.799	-2.216**	-2.000	0.314	-0.869***
	(0.851)	(3.327)	(1.058)	(2.029)	(0.332)	(0.239)
Obese 3 (BMI≥40)	-5.182***	-3.525	-2.733***	-2.569	-0.274	-0.929***
	(0.872)	(2.994)	(1.008)	(2.112)	(0.279)	(0.273)
Constant	5.058***	7.864***	3.564***	7.231***	-0.036	0.509***
	(0.840)	(1.505)	(0.922)	(1.318)	(0.159)	(0.180)
N	779	861	828	861	536	585
R^2	0.0779	0.006	0.01	0.0126	0.0084	0.043

* p<0.10, ** p<0.05, *** p<0.01

Notes: Standard errors in parentheses. Dependent variables are differences between individual subjective and objective risk relative to the objective risk of each disease. Each estimation is based on individuals in ALP aged 50-62 without respective condition at time of interview. Results weighted to take into account sample selection. Subjective expectation on the onset of arthritis is combination of subjective expectation on arthritis/rheumatism onset and subjective expectation of rheumatoid arthritis onset (see appendix for details).

Results using objective risk estimates based on probit models instead of duration models are presented in the lower panel of table 5. Instead of the time to disease onset as dependent variable we use an indicator variable that takes on the value one if an individual is diagnosed with the specific condition sometime before the 4th wave of the HRS and is zero otherwise. The results confirm conclusions drawn in the main analysis. Obese individuals underestimate the risk of arthritis or rheumatism.

[8]See Smith (2007) for the development of undiagnosed diabetes over time and Cutler et al. (2008) for the development of undiagnosed hypertension.

Moreover, there is evidence that the obese are overly optimistic about their risks of diabetes and hypertension, as they also underestimate the risks of the latter two diseases.

Table 6 reports results when the deviation of the subjective risk from the objective risk is measured relative to the objective risk. Instead of the difference in levels, we use the difference relative to the objective risk, $\frac{S_i^d - \widetilde{O_i^d}}{\widetilde{O_i^d}}$, as dependent variable in model (4). The results are again similar to the results of the main analysis. Individuals who are normal- or underweight overestimate the risks of most diseases significantly. As the percent rather than percentage point metric puts an emphasis on deviations from smaller objective values, it comes at no surprise that there are overestimations between 350 and almost 800% for the diseases that show relatively low risks among normal- and underweight individuals. The risk of hypertension is not overestimated and the risk of arthritis is overestimated by about 50% among the normal- and underweight. Importantly, despite putting relatively less emphasis on deviations from higher objective risks, we find evidence that individuals who are obese underestimate their risk of arthritis or rheumatism. Individuals with a BMI between 35 and 40 underestimate the risk by about 36% and individuals with a BMI over 40 underestimate their risk by 42%.

Overall, the robustness analyses confirm our main findings that obese individuals underestimate the 5-year risk of arthritis or rheumatism. Furthermore, this section provides evidence suggesting that also the risks of hypertension and diabetes are underestimated by individuals with excess body weight.

7 Conclusion

In this paper, we add to the literature on risk perception among the obese by comparing subjective probabilities of developing different diseases within the following 5 years to objective probabilities of the same events. Our results indicate that individuals who are normal- or underweight overestimate most disease risks. Similarly, the overweight tend to overestimate their risks. Obese individuals, however, underestimate the risks of arthritis or rheumatism, diabetes and hypertension, while they overestimate the risks of a heart attack and a stroke.

These results are important for at least two reasons. First, diabetes and hypertension are silent conditions that often do not show symptoms. Therefore they risk to remain undiagnosed. When undiagnosed the conditions cannot be treated adequately and might thus be more harmful. If individuals who underestimate the risks of these conditions are also less likely to get checked for the conditions, the conditions might remain undiagnosed for a longer period. They would then cause greater harm and possibly higher medical expenses than if they were treated earlier.

Second, the results can provide guidance for designing health education programs targeted at the obese. While the medium term risks of cardiovascular disease, and in particular heart attack and stroke, is on average not underestimated by the obese, there is evidence that the risks of diabetes, hypertension and arthritis are not well understood. Increasing the awareness of these risks among the obese might allow them to make better-informed lifestyle choices and help them to opt for a healthier lifestyle.

References

Cawley, J. and Meyerhoefer, C. (2010). The Medical Care Costs of Obesity: An Instrumental Variables Approach. NBER Working Paper 16467.

Cutler, J., Sorlie, P., Wolz, M., Thom, T., Fields, L., and Roccella, E. (2008). Trends in Hypertension Prevalence, Awareness, Treatment, and Control Rates in United States Adults Between 1988-1994 and 1999-2004. *Hypertension*, 52:818–827.

Dixon, J. (2010). The effect of obesity on health outcomes. *Molecular and Cellular Endocrinology*, 316(2):104–108.

Goldman, D., Michaud, P., Lakdawalla, D., Zheng, Y., Gailey, A., and Vaynman, I. (2010). The Fiscal Consequences of Trends in Population Health. *National Tax Journal*, 63(2):307–330.

Gregory, C., Blanck, H., Gillespie, C., Maynard, L., and Serdula, M. (2008). Perceived health risk of excess body weight among overweight and obese men and women: differences by sex. *Preventive Medicine*, 47(1):46–52.

Haslam, D. and James, W. (2005). Obesity. *The Lancet*, 366(9492):1197–1209.

Hurd, M. (2009). Subjective probabilities in household surveys. *Annual Review of Economics*, 1:543–562.

Kan, K. and Tsai, W. (2004). Obesity and risk knowledge. *Journal of Health Economics*, 23(5):907–934.

Khwaja, A., Silverman, D., Sloan, F., and Wang, Y. (2009). Are mature smokers misinformed? *Journal of Health Economics*, 28(2):385–397.

Lichtenstein, S., Slovic, P., Fischhoff, B., Layman, M., and Combs, B. (1978). Judged frequency of lethal events. *Journal of Experimental Psychology: Human Learning and Memory*, 4(6):551–578.

Manski, C. (2004). Measuring expectations. *Econometrica*, 72(5):1329–1376.

National Center for Health Statistics (2010). *Health, United States, 2009: With Special Feature on Medial Technology*. Hyattsville, MD.

Sassi, F. (2010). Obesity and the Economics of Prevention. OECD.

Schoenbaum, M. (1997). Do smokers understand the mortality effects of smoking? Evidence from the Health and Retirement Survey. *American Journal of Public Health*, 87(5):755–759.

Smith, J. (2007). Nature and causes of trends in male diabetes prevalence, undiagnosed diabetes, and the socioeconomic status health gradient. *Proceedings of the National Academy of Sciences*, 104(33):13225–13231.

Viscusi, W. (1990). Do smokers underestimate risks? *Journal of Political Economy*, 98(6):1253–1269.

Viscusi, W. and Hakes, J. (2008). Risk beliefs and smoking behavior. *Economic Inquiry*, 46(1):49–59.

Appendix

Subjective Expectations on Arthritis/Rheumatism

Subjective expectations on developing arthritis or rheumatism in the future are elicited by two separate questions in the ALP. The first question asks individuals about their chances of developing arthritis or rheumatism except rheumatoid arthritis, $Pr(arthr)$, and the second question asks about the chances of developing rheumatoid arthritis, $Pr(RA)$. In the HRS, however, there is only information on arthritis and rheumatism including rheumatoid arthritis. We therefore combine the answers to the two subjective expectations question in the ALP.

As we are interested in whether overweight and obese individuals underestimate their risks of developing different diseases, we aggregate the two probabilities in a way that results in the largest possible subjective probability of developing arthritis. In particular, we set

$$Pr(arthritis) = min\{100; Pr(arthr) + Pr(RA)\} \tag{5}$$