Empir Econ (2012) 42:1051–1078
DOI 10.1007/s00181-011-0453-4

Food away from home expenditures and obesity among older Europeans: are there gender differences?

Andreas C. Drichoutis · Rodolfo M. Nayga Jr. · Panagiotis Lazaridis

Received: 1 November 2009 / Accepted: 22 November 2010 / Published online: 19 March 2011

Abstract Older people (i.e., at least 50 years of age) are becoming one of the most important demographic groups in the world. We explore the effect of food away from home expenditures on obesity among the older population in Europe using instrumental variable methods. Several statistical tests were conducted to assess endogeneity of selected variables, the exogeneity, relevance, and validity of instruments used. Our results generally suggest that food-away-from-home expenditure has no statistically significant effect on body mass index (BMI) of older males but is negatively related to BMI of older females.

Keywords Body Mass Index · Food expenditures · Instrumental variables · GMM

JEL Classification C3 · I1 · D12

A. C. Drichoutis (✉)
Department of Economics, University of Ioannina, University Campus,
45110 Ioannina, Greece
e-mail: adrihout@cc.uoi.gr

R. M. Nayga Jr.
Department of Agricultural Economics & Agribusiness, University of Arkansas,
Fayetteville, AR 72701, USA
e-mail: rnayga@uark.edu

P. Lazaridis
Department of Agricultural Economics & Rural Development,
Agricultural University of Athens,
Iera Odos 75, 11855 Athens, Greece
e-mail: t.lazaridis@aua.gr

1 Introduction

Aging is becoming one of the most salient social, economic, and demographic phenomena in the world. This issue is particularly acute in Europe due to decreasing birth rates and aging baby boom generation (i.e., the largest generation born in the 1950s and 1960s). Consequently, Europe now has the highest proportion of population aged 65 or over in the world (Börsch-Supan 2005). Population estimates from Eurostat indicate that in 2008 about 17.3% of the EU25 population are aged 65 or over and that this will rise to 34.6% in 2050. If those between 60 and 65 years of age are also included, the numbers rise to 22.6% for 2008 and 41.4% for 2050.

Besides the aging population concerns in Europe, obesity is also now considered a hot issue that is getting a lot of attention. In fact, obesity is now in the forefront of many individual, government, and societal concerns (Nayga 2008). Globally, obesity is increasing in dramatic rates. The World Health Organization indicated that there were 1.6 billion overweight adults and at least 400 million obese adults in the world in 2005 (WHO 2006). By 2015, these figures are expected to rise to 2.3 billion overweight and 700 million obese adults. The effects of obesity on health are well supported by the medical literature and includes osteoarthritis, sleep apnea, asthma, high blood pressure, gallbladder disease, cholesterol, type II diabetes, cardiovascular disease, stroke, renal, and genitourinary diseases (Whitmer et al. 2005; Ejerblad et al. 2006; Esposito et al. 2004; van der Steeg et al. 2008; Grundy 2004; Bray 2004). Obesity may also inflict severe emotional harm, such as social stigmatization, depression, and poor body image.

Society is rightly alarmed at the rise in obesity among the young and middle-aged since they generally threaten the health status of the working population. However, obesity is also increasing rapidly among older people with higher health consequences. Older obese patients have been found to have worse health profiles than both non-obese older patients and obese younger patients (Palkhivala 2002). Hence, obesity combined with the increased number of older people will place additional burdens on the health-care system. In fact, according to a US study using longitudinal data from 1992 to 2001, older men (women) who are overweight or obese at age 65 have a 6–13% (11–17%) more lifetime health care expenditures than those at the same age cohort with normal weight (Yang and Hall 2008).

In the debate over obesity, a lot of attention has been paid lately on food away from home (FAFH) (i.e., food consumed in restaurants and fast food establishments). The lawsuits in the US against fast food companies regarding the alleged detrimental health effects of their products have not helped the image of the FAFH industry (e.g., the Pelman vs. McDonalds' lawsuit[1]) and have resulted in some cries for policy intervention from interest groups and public health advocates. To examine whether these

[1] In 2003, two obese New York City teens, Ashley Pelman and Jazlen Bradley, were plaintiffs in the first obesity lawsuit against a fast food restaurant (McDonalds). The Pelman case featured three core claims, sounding in negligence, products liability, and unfair business practices under New York law. The US Federal court eventually dismissed the case reasoning that the teens' own choices caused their obesity. The interested reader might consult (Frank 2006) for various obesity litigation cases with emphasis on the *Pelman* case.

demands are based on solid grounds, one would need to assess the effect that FAFH might have on body weight outcomes.

While a number of recent papers have explicitly focused on availability of fast food restaurants as potential contributors to obesity (e.g., Rashad et al. 2006; Dunn 2007, 2010; Anderson and Matsa 2009; Currie et al. 2010; Chou et al. 2008), few studies have examined the relationship between FAFH expenditures and obesity. You and Davis (2010) assessed the influence of household food expenditures, parental time allocation, and other parental factors on children's obesity-related health outcomes. They developed a theoretical model and modeled the parents–child interaction as a two-stage Stackelberg game. They then tested the model using a unique dataset drawn from 311 households in Houston, Texas. You and Davis (2010) found that only household food at home (FAH) expenditures positively affects child's BMI (FAFH expenditures did not have a statistically significant effect). Kyureghian et al. (2007) used data from National Eating Trends in the US and modeled the relationship between number of foods consumed and obesity by service type and meal occasion. Their results suggest that FAFH consumption is positively related to BMI and that foods consumed from quick service restaurants affect BMI more than foods from full service restaurants. They also found that lunch away from home has a sizeable positive influence on BMI. Their study confirms the public held view that fast food can increase obesity levels.

In this study, we examine the effect of FAFH on body weight outcomes among a very specific segment of people, namely older Europeans. We focus our analysis on older Europeans due to the increasing importance of the older segment of the European population. We employ instrumental variable (IV) methods, using an instrument list that has been thoroughly scrutinized for exogeneity, relevance, and validity. We address the endogeneity issue using the IV method since it is possible that unobserved characteristics (e.g., preference for good health) associated with obesity may also affect food expenditures. Moreover, it is also possible that obesity could influence food expenditures. For example, obese individuals could have higher food expenditures because they have higher BMI than others.

The following sections present a short literature review, the data, empirical specification, estimation, results, and concluding remarks.

2 Literature review

Many researchers have provided economic and non-economic explanations on what determines the caloric imbalance that causes obesity and several explanations have been proposed. Some authors, for example, argue that technological innovations that reduce the amount of effort required to accomplish various tasks and the taste-nutrition trade-off have influenced food consumption and obesity (Wansink and Huckabee 2005).

Other studies suggest that obesity rates are related to agricultural innovation that has lowered food prices and to technological changes in home and market productions (Lakdawalla and Philipson 2002). That is, technological change has contributed to the rise in obesity either by lowering the cost of consuming calories and/or by raising the cost of expending them (Philipson and Posner 2003). In addition, the prevalence of fast-food restaurants, higher prices of alcohol and cigarettes (Chou et al. 2004), lower time costs of food preparation resulting from technological changes in mass food preparation (Cutler et al. 2003) and even the rise of sprawl patterns in land development (Plantinga and Bernell 2007; Ewing et al. 2003) have been linked to rising obesity. In a cross-country comparison among OECD countries, Loureiro and Nayga (2005) showed that the percentage of female labor force participation positively affects obesity rates. Likewise, Anderson et al. (2004) found that the intensity of a mother's work over a child's lifetime has a positive effect on a child's likelihood of being overweight.

Lakdawalla and Philipson (2006) discuss the theoretical approaches that economists have taken to study obesity. They distinguish between neoclassical and behavioral theories of weight gain and discuss how these interrelate. The neoclassical theories rest upon the capital investment model of weight (e.g., Lakdawalla and Philipson 2002; Philipson and Posner 2003) and the rational addiction model of weight (Cawley 1999). The capital investment model stresses that agricultural technological advancements have induced lower food prices while the cost of calorie expenditure has become higher at the same time as home and market productions became more sedentary. The rational addiction model complements the capital investment model by making food addictive (i.e., increasing food consumption today makes it necessary to increase food consumption in the future).

While the neoclassical theories of weight gain assume that individuals are rational and forward looking with respect to their weight, Cutler et al. (2003) adopt a behavioral economics view by arguing that individuals have self-control problems and therefore discount the future quasi-hyperbolically instead of exponentially, as a rational-neoclassical consumer would do. Lakdawalla and Philipson (2006) lean toward the neoclassical view since they note that while behavioral theories add a useful explanatory element, they do not adequately explain time trends and differences across countries.

Cawley (1999), on the other hand, uses the rational addiction model (Becker and Murphy 1988) and considers obesity as the result of an addiction to calories. He found support for the hypothesis that caloric consumption is addictive. Dockner and Feichtinger (1993) also apply the rational addiction model to eating decisions. They assume that food consumption is addictive and show that consumption decisions, and the consequent weight path, can exhibit cycles with gradual increases followed by gradual decreases.

Using a different model, Levy (2002) set out a dynamic model where eating is neither addictive nor habit forming. He found that when physiological, psychological, environmental, and socio-cultural reasons for divergence from a physiologically optimal weight do not exist, the steady state for an expected lifetime-utility maximizer is a state of overweightness. He also showed that the rationally optimal stationary level of overweightness increases with an individual's rate of time-preference but declines with his/her rate of caloric expenditures. The positive association between the rate of time preference and obesity has also been confirmed in other studies (Smith et al. 2005; Zhang and Rashad 2008).

Levy (2003) developed a similar dynamic model where he incorporated taste, price, and risk differences between junk food and healthy food into an expected lifetime-utility-maximizing framework under the assumptions that junk food is cheaper and tastier than healthy food. Goldfarb et al. (2006), on the other hand, develop a static model that includes the benefits of food consumption ("satisfaction from food") and the negative effects that arise from weight above or below ideal weight. Their model, similar to Levy (2002), generates "optimal overweightness."

Suranovic and Goldfarb (2006) adopt a bounded rationality approach to an individual's food consumption and dieting decisions. They assume that food consumption has three possible effects on individual utility: a positive benefit from food consumption, a negative utility effect resulting from weight gain, and a negative effect caused by dieting. Their results show that an individual will occasionally choose to diet, but that diet will reduce weight only temporarily. Hence, recurrence of weight gain provides a rationale for cyclical dieting.

None of the studies discussed above assessed the effect of food expenditures on body weight outcomes. Exceptions to this strand of the literature would be Huffman and Rizov (2010) which depart from a theoretical framework and link food expenditures to obesity but do not specifically consider FAFH consumption. You and Davis (2010) consider separately FAH and FAFH expenditures but focus on child's BMI outcomes and parental role. Our aim is to fill this void in the literature, focusing on an increasingly important segment of the population: older Europeans. The cross-sectional nature of our data, which will be discussed next, allows us to examine the relationship between two types of food expenditures, FAH and FAFH, and body weight outcomes. However, we are not able to focus on the dynamic patterns of these inter-relationships due to lack of access to longitudinal data.

3 Data description

We use a European micro-dataset, the Survey of Health, Ageing, and Retirement in Europe (SHARE), which contains data on health, socio-economic status, and social and family networks of individuals aged 50 or over. Eleven countries have participated in the 2004 SHARE baseline study, representing the various regions in Europe; namely Scandinavia (Denmark and Sweden), Central Europe (Austria, France, Germany, Switzerland, Belgium, and the Netherlands), and Southern Europe (Spain, Italy and Greece).

The SHARE data used in our analysis are drawn from Release 2.0.1 of the Survey of Health, Ageing, and Retirement in Europe. SHARE was designed following the US Health and Retirement Study (HRS) and the English Longitudinal Study of Ageing (ELSA). SHARE is the first European dataset that includes information on physical and mental health as well as income and assets information of the older Europeans.

The data consist of information from 28,517 individuals in the 11 countries mentioned above. Table 1 provides sample characteristics by country, gender, and age groups. The age group of less than 50 years of age represents younger spouses or partners of age eligible respondents. Table 1 also displays the household response rates and the individual response rates. For methodological details see Börsch-Supan (2005) and Börsch-Supan and Jürges (2005).

Table 1 Sample characteristics: SHARE data

Country	Total	Gender		Age				Response rates[a]	
		Male	Female	Under 50	50–64	65–74	Over 75	Household response rate (%)	Individual response rate (%)
Austria	1,893	782	1,111	44	949	544	356	55.60	87.50
Germany	3,008	1,380	1,628	65	1,569	887	487	39.20	90.50
Sweden	3,053	1,414	1,639	56	1,589	816	592	63.20	93.00
Netherlands	2,979	1,368	1,611	102	1,697	716	464	81.00	93.30
Spain	2,396	994	1,402	42	1,080	701	573	63.40	86.20
Italy	2,559	1,132	1,427	51	1,342	785	381	63.10	91.80
France	3,193	1,386	1,807	141	1,627	768	657	54.50	79.70
Denmark	1,707	771	936	92	916	369	330	61.60	87.80
Greece	2,898	1,244	1,654	218	1,450	714	516	53.00	73.70
Switzerland	1,004	462	542	42	505	252	205	46.90	84.60
Belgium	3,827	1,739	2,088	128	1,947	992	760	38.80	86.90
Total	28,517	12,672	15,845	981	14,671	7,544	5,321	61.60	85.30

[a] Weighted average for main sample (see Börsch-Supan and Jürges 2005, for methodological details)

The SHARE dataset contains a variety of physical and mental health measures like Body Mass Index (BMI), chronic diseases, and hand grip strength.[2] Several other variables reflect individual's lifestyle like smoking, exercise, and excessive drinking of alcohol as well as demographic factors like age, gender, household size, urbanization, and total income. In addition, the SHARE dataset includes information about household expenditures on FAH and FAFH[3] as well as purchasing power parity coefficients in the various countries. Although FAH and FAFH are based on recall questions, Browning et al. (2003) provide evidence that respondents can accurately report such expenditures in recall questions. Recall questions on food expenditures are also common in several surveys like the US Consumer Expenditure Survey (CEX), the Canadian Out of Employment Panel (COEP) and Family Expenditure Survey (FAMEX), the Italian Survey on Household Income and Wealth (SHIW) and others.

Observations with incomplete information on the variables of interest were dropped from the analysis thus leaving us with 25,075 valid cases. Figure 1a and b presents mean food expenditures for at-home and away-from-home consumption, respectively, of older Europeans by weight status and regions of Europe. Scandinavians spend on

[2] Hand grip strength has been shown to be correlated with mental and physical health and is predictive of the incidence of functional limitations, disability and even mortality in old age (Christensen et al. 2001; Frederiksen et al. 2002; Rantanen et al. 1998, 1999). It is measured using a handheld dynamometer—where respondents are asked to press a lever as hard as they can. The dynamometer shows grip strength in kilograms.

[3] Food away from home includes food *eaten* away from home as well as food *prepared* away from home.

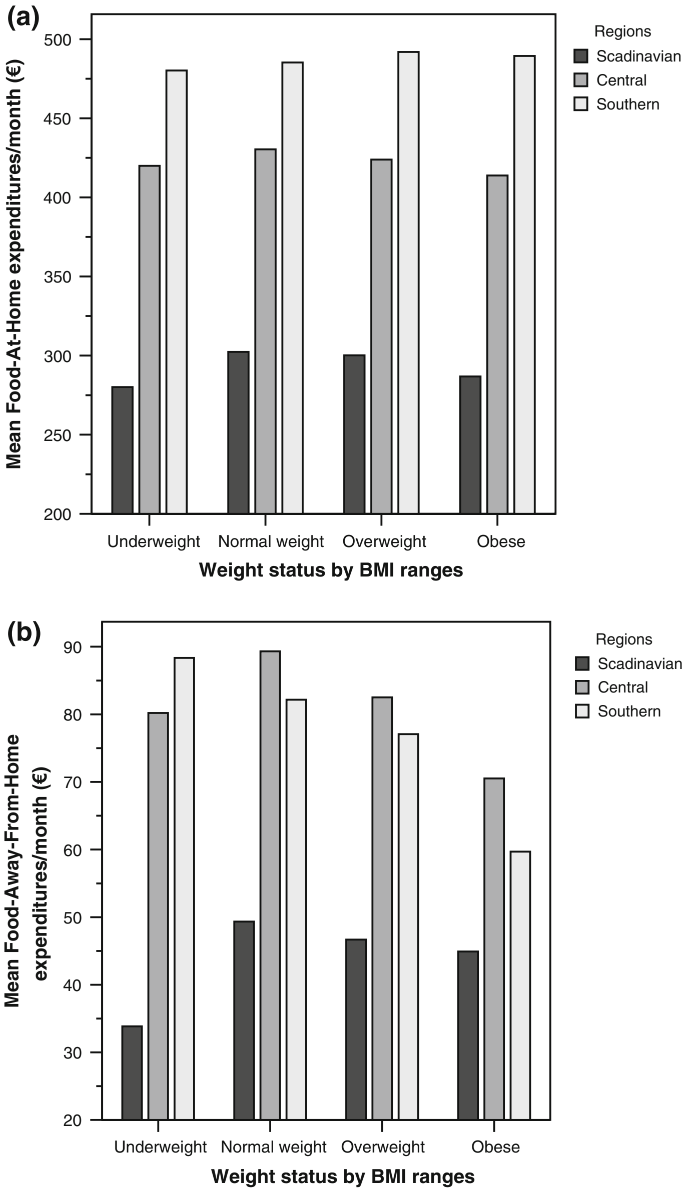

Fig. 1 Mean food expenditures by weight status and regions

average less for FAH and FAFH than Central and Southern Europeans. The Southern Europeans spend more than other Europeans on FAH but less than Central Europeans on FAFH. In terms of weight status, it appears that as weight increases, mean FAFH expenditures decrease while FAH expenditures are kept constant.

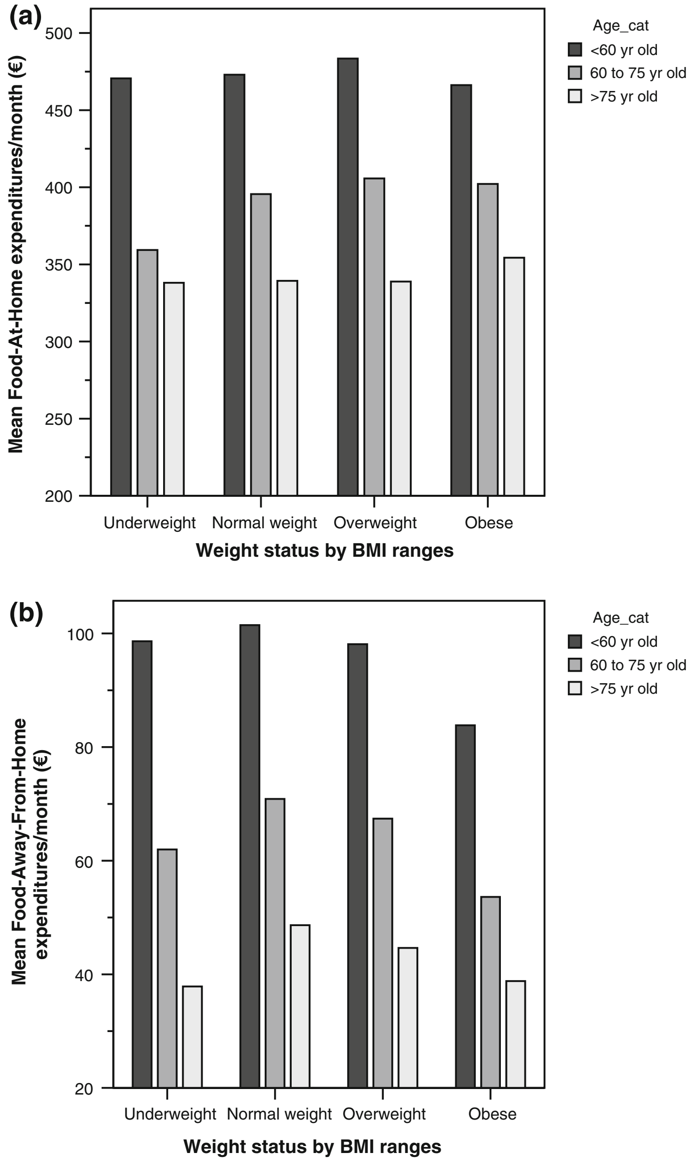

Fig. 2 Mean food expenditures by weight status and age

Figure 2a and b shows mean food expenditures by weight status and age categories. Lower mean expenditures for at-home and away-from-home consumption for older individuals may stem from declining energy needs that are related with aging. Heavier individuals also spend less on FAFH consumption but there is no clear trend for the FAH consumption.

4 Empirical specification

We model the relation between obesity and FAFH expenditures using the equation depicted below:

$$\text{BMI} = a_0 + a_1\text{TF} + a_2\text{PFAFH} + \mathbf{a_3X} + \mathbf{a_4Z} + u \quad (1)$$

where BMI is Body Mass Index, TF is total food expenditures, PFAFH is percentage of food spent on food-away-from-home, **X** is a vector of demographic variables like age, gender and household size, **Z** is a vector of health related determinants and u is the error term. The **Z** vector includes variables such as number of cigarettes smoked per day (NumCig), a dummy for physical inactivity (PHinactiv), the maximum grip strength of the respondent (MaxGrip), number of chronic diseases the person suffers from (Chronic), and a dummy that identifies excessive alcohol drinkers (Drinking). At a more fundamental level the reader can refer to the theoretical model of Huffman and Rizov (2010) that establishes the relation between food expenditures and BMI. Even though FAH expenditures are not included explicitly in the model, we can obviously draw inferences for FAH since FAH and FAFH comprise total food expenditures. We use this formulation because it allows us to examine how expenditures are allocated between FAH and FAFH when total food expenditures are kept constant.

As previously discussed, we have to take into account the endogeneity issue since some unobservable characteristics that could affect obesity could also influence food expenditures. For example, individuals who prefer or care about good health may engage in activities that may limit weight gain and consumption of FAFH. Also, it is possible that heavier individuals may have higher food expenditures because they are bigger and are in need of more food than others. To address the possible endogeneity of the expenditure variables (TF and PFAFH), we use an instrumental variables approach where the endogenous variables are instrumented using a set of country-level socio-economic determinants that can influence food expenditures. The instrument list includes variables that are presumed to capture differences in expenditures between countries due to socio-economic conditions: the purchasing parity coefficient (PPP), GDP per capita (GDPpc, in thousands €), inflation rate (Inflat), unemployment rate (Unempl) and a socio-economic index (Socioecon).[4] PPP serves as a proxy for prices that are unavailable in the SHARE dataset. It equalizes the purchasing power of different currencies in home countries for a given basket of goods. PPP takes into account the relative cost of living and the inflation rate of the different countries which then gives a good proxy for between-countries comparison of prices. In the next sections, we thoroughly scrutinize this external set of instruments for exogeneity, relevance, and validity and find that this specific set of external instruments is weak. An alternative to external instruments is Lewbel's (2010) method which exploits heteroskedasticity in the first stage regression to achieve identification (instead of identification through

[4] GDP per capita was extracted from Penn tables (http://pwt.econ.upenn.edu/php_site/pwt62/pwt62_form.php), inflation and unemployment rate were extracted from EconStats (http://www.econstats.com/index_gl.htm), and the socio-economic index was taken from the International Country Risk Guide (http://www.prsgroup.com/ICRG.aspx).

standard exclusion restrictions). Instruments are then constructed from variables within the given set of covariates (**X** and **Z** vector in our case). This method is not of course without caveats. As Lewbel (2010) notes, "the resulting identification is based on higher moments, and so is likely to provide less reliable estimates than identification based on standard exclusion restrictions." However it "...may be useful in applications where traditional instruments are not available, or could be used along with traditional instruments to increase efficiency." Therefore, we first present the analysis using the external set of instruments and then complement this analysis with the use of additional instruments (i.e., internally constructed instruments) using Lewbel's method.

The analysis we undertake is conducted separately for males and females, as is common in studies that involve weight adjusted outcomes like BMI. This way we relax the constraint that any given variable has the same effect on BMI for both males and females. In the next section, we undertake a series of tests to assess the validity of our selected instruments. Table 2 exhibits the variables used in the analysis, their description, and some basic descriptive statistics. The table also reports population means derived using weights and information on the complex design of the survey.

5 Estimation

Two of the widely used single-equation estimation methods are the Instrumental Variables estimator (two-stage least squares/2SLS) and the general method of moments estimator (GMM). However, the conventional IV estimator is inefficient in the presence of heteroskedasticity and the usual approach when facing heteroskedasticity of an unknown form is to use the GMM estimator. If heteroskedasticity is indeed present, the GMM estimator is more efficient than the simple IV estimator, whereas if heteroskedasticity is not present the GMM estimator is no worse asymptotically than the IV estimator (Baum et al. 2003). However, the GMM estimator may have poor small sample properties and if in fact the errors are homoskedastic, IV would be preferable. We conducted tests for the presence of heteroskedasticity, as proposed by Pagan and Hall (1983)[5] (Schaffer 2002; Baum et al. 2003). The null of homoskedastic disturbances is rejected for both sub-samples (see Table A1 in the appendix).

We therefore estimate all our equations with the two-step feasible general method of moments (2S-GMM).[6] Probability weights are used in all our estimations and the estimated standard errors are robust to deviations from i.i.d. disturbances, i.e., arbitrary heteroskedasticity and arbitrary intragroup correlation (the latter is due to the fact that multiple observations may come from the same survey stratum). Before presenting the estimated results, we conduct endogeneity tests for the assumed endogenous regressors, tests on the relevance of the instruments (under- and weak-identification),

[5] We used the ivhettest command in Stata 10.

[6] Equations were estimated with the ivreg2 user-written module in Stata 10 (Baum et al. 2007). However, in a latter section since we find evidence of weak instruments we also use Limited Information Maximum Likelihood estimation (LIML) and Fuller's LIML which are partially robust to weak instruments.
In addition, since the survey commands (svy) are not available for ivreg2, the Huber/White sandwich estimator of variance is used with an adjustment for clustering within the SHARE survey strata (similar action was taken by Long 2008).

Table 2 Variable description

Variable names	Variable description	Full sample		Male subsample		Female subsample		Male population[b]		Female population[b]	
		Mean	SD	Mean	SD	Mean	SD	Mean	SE	Mean	SE
TF	Monthly household food expenditures in €, ppp-adjusted	495.38	313.48	514.87	308.98	479.22	316.27	539.20	9.61	472.18	8.58
PFAFH	Percentage of monthly household food away from home expenditure, ppp-adjusted	12.69	14.65	13.78	15.16	11.78	14.15	13.60	0.32	11.00	0.35
BMI	Body Mass Index	26.37	4.29	26.72	3.76	26.08	4.66	26.80	0.07	26.17	0.10
Scandinavian	Dummy for Scandinavian countries	0.17	0.38	0.17	0.38	0.17	0.37	0.05	0.02	0.05	0.02
Central	Dummy for Central European countries	0.56	0.50	0.56	0.50	0.55	0.50	0.58	0.03	0.58	0.03
Southern[a]	Dummy for Mediterranean countries	0.27	0.44	0.26	0.44	0.28	0.45	0.37	0.03	0.37	0.03
Gender	Dummy for males	0.45	0.50	–	–	–	–	–	–	–	–
EducYears	Years of education	10.15	4.46	10.66	4.47	9.73	4.40	10.48	0.17	9.28	0.17
Urb_1	Dummy, household resides in a big city or suburbs	0.32	0.47	0.32	0.46	0.33	0.47	0.28	0.03	0.29	0.03
Urb_2	Dummy, household resides in a large town	0.19	0.39	0.19	0.39	0.19	0.39	0.16	0.01	0.17	0.01
Urb_3	Dummy, household resides in a small town	0.25	0.44	0.25	0.44	0.25	0.44	0.27	0.02	0.27	0.02

Table 2 continued

Variable names	Variable description	Full sample		Male subsample		Female subsample		Male population[b]		Female population[b]	
		Mean	SD	Mean	SD	Mean	SD	Mean	SE	Mean	SE
Urb_4[a]	Dummy, household resides in a village	0.23	0.42	0.24	0.43	0.23	0.42	0.29	0.02	0.27	0.02
Hsize	Household size	2.21	1.00	2.30	0.97	2.14	1.02	2.30	0.02	2.01	0.02
Age	Age	63.68	10.19	64.14	9.60	63.30	10.63	64.09	0.19	65.75	0.22
Log(Inc)	Logarithm of total income, ppp-adjusted	10.06	2.32	10.19	2.12	9.95	2.46	10.03	0.05	9.78	0.05
NumCig	Number of cigarettes smoked per day	3.20	8.02	4.12	9.41	2.43	6.56	4.03	0.15	1.90	0.12
PHinactiv	Dummy for physical inactivity (never or almost never engaging in neither moderate nor vigorous physical activity)	0.08	0.27	0.06	0.24	0.09	0.29	0.08	0.01	0.12	0.01
MaxGrip	Maximum of grip strength measures	34.49	12.30	43.62	10.66	26.92	7.50	42.75	0.29	25.77	0.19
Chronic	Number of chronic diseases	1.45	1.39	1.37	1.31	1.52	1.45	1.40	0.03	1.67	0.03
Drinking	Dummy for drinking more than 2 glasses of alcohol every day	0.14	0.35	0.23	0.42	0.07	0.25	0.27	0.01	0.07	0.00
Homemaker	Dummy for current occupation being a homemaker	0.16	0.36	0.003	0.06	0.28	0.45	0.003	0.0006	0.29	0.01

Table 2 continued

Variable names	Variable description	Full sample		Male subsample		Female subsample		Male population[b]		Female population[b]	
		Mean	SD	Mean	SD	Mean	SD	Mean	SE	Mean	SE
Retired	Dummy for current occupation being retired	0.47	0.50	0.58	0.49	0.39	0.48	0.10	0.01	0.44	0.01
PPP	Purchase power parity coefficient	1.02	0.12								
GDPpc	GDP per capita (in thousands of euros)	27.24	3.64								
Inflat	Inflation rate	1.94	0.68								
Unempl	Unemployment rate	7.80	2.22								
Socioecon	Socioeconomic conditions index	8.82	1.02								

[a] Variables excluded for estimation purposes
[b] Estimates obtained using probability weights

tests for over-identifying restrictions, and tests for the exogeneity of the instruments (Murray 2006).

5.1 Endogeneity tests

The first step we undertake in this series of tests is to test whether the assumed endogenous regressors are actually endogenous. The null hypothesis is that the specified endogenous variables can be treated as exogenous. The test statistic is distributed as χ^2 with degrees of freedom equal to the number of variables tested. Under i.i.d. assumptions, this endogeneity test statistic is numerically equal to a Hausman test statistic (Baum et al. 2003). The reported test statistic in Table A2 is robust to violations of homoskedasticity. The null is rejected at the 10% significance level in both cases.

5.2 Testing the relevance of the instruments

To test if our instruments are relevant, i.e., sufficiently correlated with the included endogenous variables, we use several statistics. The partial R^2 of Bound et al. (1995) is a statistic commonly used for this purpose.[7] However, when multiple endogenous regressors are used other statistics may also be required. The Shea's partial R^2 (Shea 1997) takes into account the inter-correlations among the instruments.[8]

Table A3 shows that the F test always rejects the null that the coefficients of the instruments from the first stage regressions are zero. Hence, the instruments pass the joint significance test. However, the partial R^2 values are small, making further inferences problematic. Other tests are then required. Table A4 reports the LM test of the Kleibergen and Paap (2006) rk statistic[9] as well as the Angrist–Pischke (Angrist and Pischke 2009) first stage χ^2 statistics.[10] The null that the equation is under-identified is rejected based on the LM test (only at the 10% level in the case of the female sub-sample), i.e., all equations are identified. The Angrist–Pischke test of whether a particular endogenous regressor alone is unidentified is rejected in all cases.

In Table A5, we report the Wald F statistic of the Kleibergen and Paap (2006) rk statistic which is an identification test of whether the excluded instruments are weakly correlated with the endogenous regressors. The statistic can be compared to the

[7] The partial R^2 is the R^2 of the first stage regression with the included instruments "partialled out," which is equivalent to an F test of the joint significance of the excluded instruments (see Baum et al. 2003 for more details).

[8] The Shea's partial R^2 and the standard partial R^2 are equivalent when the model contains only one endogenous regressor.

[9] The statistic is distributed as χ^2 with $(L_1 - k_1 + 1)$ degrees of freedom where L_1 is the number of excluded instruments and k_1 the number of endogenous regressors. The LM test is a test of under-identification, i.e., the excluded instruments are correlated with the endogenous variables.

[10] The Angrist–Pischke χ^2 statistic is distributed as χ^2 with $(L_1 - k_1 + 1)$ degrees of freedom under the null that the particular endogenous regressor in question is unidentified. In our case, where we have two endogenous regressors, the Angrist–Pischke test fails to reject if a *particular* endogenous regressor is unidentified, while the Kleibergen–Paap test fails to reject if *any* of the endogenous regressors is unidentified.

Stock and Yogo (2005) compilation of IV critical values.[11,12] The null hypothesis of Stock and Yogo (2005) is that the given group of instruments is weak against the alternative that it is strong. The test rejects if the Wald F statistic exceeds the critical value. The null is not rejected in our case. The analogous F statistic for testing particular endogenous regressors is the Angrist–Pischke F statistic (reported in Table A5). This test does not reject the null hypothesis for the case of the PFAFH variable.

Although the critical values of Stock and Yogo (2005) are only for the i.i.d. there is some indication that the group of instruments we used is weak, particularly for the case of the PFAFH variable. To cope with weak instruments, we complement our estimations with the Limited Information Maximum Likelihood (LIML) and the Fuller's modified LIML (FULL) estimators.[13] In addition, as an alternative to the external instrument list for achieving identification, we exploit the heteroskedasticity in the first stage regressions[14] and follow Lewbel's (2010) method in developing additional instruments for the estimation. Specifically, we construct instruments in the form of $(w - \bar{w})\hat{e}_i$, where w is any continuous variable of the $\mathbf{X}$ or $\mathbf{Z}$ vector in Eq. 1 ($\bar{w}$ being the sample mean) and $\hat{e}_i$ are the residuals from the first stage regression for i = TF or PFAFH. Instruments are constructed separately for the two sub-samples. We use this instrument list along with the external instrument list to complement our estimations.

Given that the external set of instruments is at the country level, clustering by country (along with the stratum) is necessary. Unfortunately, since the number of clusters (11 countries/clusters) is less than the number of exogenous regressors plus the number of excluded instruments, the covariance matrix of orthogonality conditions S is not of full rank and GMM is infeasible since the optimal weighting matrix $W = S^{-1}$ cannot be calculated. Even if we dropped most of the demographic variables (to get less exogenous regressors plus excluded instruments than clusters), we would still have bias from the few clusters. And in this case, the use of double clustered standard errors can do more harm than good (Thompson 2011). A possible but not pragmatic solution to this problem would be to increase the number of clusters by collecting data for more countries (Angrist and Pischke 2009). Given that the number of countries is fixed in the SHARE dataset, we attempt to demonstrate that the bias of robust standard errors is not severe in our case by dismissing the country level instruments in Lewbel's method. The results, exhibited along with the rest of the estimations, generally show that our inferences are robust.

[11] The Stock–Yogo weak instruments tests come in two variants: maximal relative bias and maximal size. The first variant is based on the ratio of the bias of the estimator to the bias of OLS and the second variant is based on the performance of the Wald test statistic (rejection rate of the test). See Baum and Schaffer (2007) for more details.

[12] Baum and Schaffer (2007) suggest this comparison even though they acknowledged that the critical values of Stock and Yogo (2005) are for the i.i.d. case. They suggest this as a sensible choice since no studies on testing for weak instruments in the presence of non-i.i.d. errors exists.

[13] The advantage of LIML estimation is that it is median unbiased: the median of its sampling distribution is generally close to the population parameter under the assumption of normal distribution of the error term (see Bascle 2008).

[14] Pagan and Hall (1983) heteroskedasticity tests for both first stage regressions (TF, PFAFH) and both subsamples reject the null with a P value of 0 in all cases.

5.3 Testing over-identifying restrictions

We validate our exclusion restrictions by employing over-identification tests using Hansen's J statistic (Hansen 1982; Baum and Schaffer 2007). The joint null hypothesis is that the instruments are valid instruments, i.e., uncorrelated with the error term.[15] Hansen's J statistic for the external instrument vector is reported in Table A6. The null hypothesis is rejected in the case of the female sub-sample, which casts doubts on the exogeneity of one or more instruments, but is not rejected in the case of the male sub-sample. Therefore, for the male subsample, the decision on which instruments to keep is easy, i.e., we keep the whole external instrument vector.

On the other hand, the instruments for the female sub-sample require further scrutiny. We can test for the exogeneity of one or more instruments using the C statistic. The C statistic is defined as the difference of the Hansen's J statistic of the equation with the smaller set of instruments and the equation with the full set of instruments, i.e., including the instruments whose validity is suspect.[16]

We start by testing separately each of the instruments of the instrument vector. Table 3a reports the Hansen's J statistic for the female sub-sample (where the suspect instrument is excluded) and the C statistic (where under the null hypothesis, both the smaller set of instruments and the additional suspect instruments are valid). Rejection of the null of the C statistic requires that the full set of instruments is not valid. Basically, we are looking for cases where when the suspect instrument is excluded from the set of full instruments, we fail to reject the null of the Hansen's J statistic but reject the null of the C statistic.[17]

For the female sub-sample, we keep the instrument in bold (i.e., GDPpc) for which the C statistic does not reject (i.e., P value is much lower than conventional significance levels) and the Hansen's J statistic rejects. For the rest of the instruments, we find that they are indeed suspect instruments (C statistic) and that when we exclude them from the set of full instruments there is no big improvement in Hansen's J statistic (the rest of the instruments are not uncorrelated with the error term).[18] This calls for further scrutiny of the instrument list.

In a second step, we do the same tests by excluding each time all possible pairs of the instruments (see Table 3b). We find that among all pairs tested, highest improvement in Hansen's J statistic (lowest values) is produced by the two pairs in bold. Therefore, we dismiss the instrument pair Inflat–Socioecon since excluding this produces the lowest

[15] Failing to reject the null hypothesis provides some confidence in the identification assumption. Under the null, the test statistic is distributed as χ^2 in the number of $(L - K)$ over-identifying restrictions where L is the number of instruments (excluded and included) and K is the number of regressors of the equation (endogenous and exogenous).

[16] This is equivalent to dropping excluded instruments from the instrument list.

[17] This would in essence mean that the suspect instrument is indeed suspect (C statistic) and that by excluding it from the set of full instruments the rest of the instruments are valid (Hansen's J statistic), or that we fail to reject the null that the smaller set of instruments is valid.

[18] Although, the Socioecon and Unempl instruments seem to pass the test in terms of Hansen's J statistic, P-values are very close to conventional rejection levels. In fact, dismissing both instruments from the instrument list leads to a rejection of the null hypothesis of the instruments being uncorrelated with the error term.

Table 3 Test of subsets of regressors (Hansen's J statistics and C statistic)

Females	Instrument tested				
	PPP	GDPpc	Inflat	Socioecon	Unempl
(a)					
Hansen's J statistic[a]	6.967	**11.464**	8.127	3.720	3.922
(P-value)	(0.031)	**(0.003)**	(0.017)	(0.156)	(0.141)
C statistic	4.529	**0.032**	3.369	7.776	7.574
(P-value)	(0.033)	**(0.857)**	(0.066)	(0.005)	(0.006)

Females	Instrument tested					
	PPP, Inflat	PPP, Socioecon	PPP, Unempl	Inflat, Socioecon	Inflat, Unempl	Socioecon, Unempl
(b)						
Hansen's J statistic[a]	6.919	2.999	3.871	**0.949**	**1.463**	3.604
(P-value)	(0.009)	(0.083)	(0.049)	**(0.330)**	**(0.226)**	(0.058)
C statistic	4.577	8.497	7.625	**10.547**	**10.033**	7.892
(P-value)	(0.101)	(0.014)	(0.022)	**(0.005)**	**(0.007)**	(0.019)

[a] Excluding suspect instrument

Table 4 Instruments used by subsample

Instrumented variables		Instruments used
TF, PFAFH		
	Males	PPP, GDPpc, Inflat, Socioecon, Unempl
	Females	PPP, GDPpc, Unempl
TF, PFAFH		
	Males Females	PPP, GDPpc, Inflat, Socioecon, Unempl + constructed instruments

value in Hansen's J statistic. Table 4 summarizes the instruments used by sub-sample. We note that our strategy to use separate instruments for the two sub-samples is based purely on statistical grounds.

5.4 Tests after correct specification of instrument list

After coming up with the correct specification of the external instrument list, we re-run all tests reported in Tables A1, A2, A3, A4, A5, and A6 for the female sub-sample and report them all in Table 5. Table 5 also summarizes all tests when using the constructed instruments using the Lewbel method along with the external instrument list. The "males" column summarizes Tables A1, A2, A3, A4, A5, and A6 and is exhibited for comparative purposes. Three points are worth noting: (a) the null hypothesis of homoskedastic disturbances is highly rejected, (b) overidentification tests show that instruments are uncorrelated with the error terms,[19] and (c) the weak identification tests show that the two instrument sets for the male and female sub-samples respectively, are weak (Kleibergen and Paap rk Wald F statistic) but only as instruments of the PFAFH variable (Angrist–Pischke F statistic). However, when adding the constructed instruments, tests for the full instrument list show that the instruments are particularly strong. To cope with weak instruments in the case where we only use the external instruments, we report the FULL and LIML estimators as well. Moreira's conditional likelihood ratio (CLR) approach (Moreira 2003) cannot be employed since we have two endogenous variables.

6 Results

Table 6a and b shows the results for the two subsamples where 2S-GMM, LIML and FULL methods are reported. OLS is reported as well for comparison purposes. There are some notable differences between the two sub-samples. For example, the coefficient of the endogenous variable PFAFH has a positive sign for the male sub-sample when using the external instrument list and gets very close to zero when using the external instrument list along with the constructed Lewbel instruments. However, the coefficients in most cases are not significant given the dispersions. On the other hand, the coefficient of PFAFH is negative and statistically significant for the female sub-sample. The coefficient of the TF variable is negative and statistically significant in both sub-samples. But what is the interpretation of these variables given the ceteris paribus context? The simple case is the interpretation of an increase of the percentage of FAFH (PFAFH): given that total food expenditures (TF) are held constant, an increase of PFAFH indicates an increase in the amount of FAFH spent and a decrease in the amount of FAH spent.[20] On the other hand, an increase of total food expenditures (TF) can be interpreted as follows: given that PFAFH is held constant an increase in TF by a factor of k will result in an increase in FAFH and FAH by k as well.[21]

Obviously, expenditures and quantities mean two different things since the former also includes the concept of quality. Therefore, increases in food expenditures do not necessarily imply increases in consumed quantity. It could also imply that the household is spending more on more expensive healthier foods. Two results come out

[19] Note that we did not have to scrutinize the instrument list when adding the constructed Lewbel instruments to the external instrument list since Hansen's J statistics are such that we fail to reject the null of no correlation with the error term.

[20] This can be illustrated as follows: assume that the percentage of expenditures on FAFH is increased from PFAFH to PFAFH$'$. Since TF is held constant then $\mathrm{TF}' = \mathrm{TF}$ so that we can write: $\mathrm{PFAFH}' = \mathrm{FAFH}'/\mathrm{TF}$. It follows that if $\mathrm{PFAFH}' > \mathrm{PFAFH}$ then $\mathrm{FAFH}' > \mathrm{FAFH}$.
Note, however that the following equality must hold: $\mathrm{FAFH}' + \mathrm{FAH}' = \mathrm{TF}' = \mathrm{TF}$ or that $\mathrm{FAFH}' = \mathrm{TF} - \mathrm{FAH}'$. Since we showed that $\mathrm{FAFH}' > \mathrm{FAFH}$ we can rewrite this inequality as $\mathrm{TF} - \mathrm{FAH}' > \mathrm{TF} - \mathrm{FAH}$, which is equivalent to $\mathrm{FAH}' < \mathrm{FAH}$.

[21] The relationships can be illustrated as follows: Assume that TF is increased by a factor of k so that $\mathrm{TF}' = k\mathrm{TF}$. Given that PFAFH is held constant then we can write $\mathrm{PFAFH}' = \mathrm{PFAFH}$ or that $\mathrm{FAFH}'/\mathrm{TF}' = \mathrm{FAFH}/\mathrm{TF}$. This last equality can be written as: $\mathrm{FAFH}'/k\mathrm{TF} = \mathrm{FAFH}/\mathrm{TF}$ or $\mathrm{FAFH}' = k\mathrm{FAFH}$.
From the equality $\mathrm{FAFH}' + \mathrm{FAH}' = \mathrm{TF}'$ we get $k\mathrm{FAFH} + \mathrm{FAH}' = k\mathrm{TF}$ or that $\mathrm{FAH}' = k\mathrm{FAH}$.

Table 5 Tests for equations after instrument specification

	External instruments				External plus constructed instruments			
	Males		Females		Males		Females	
	TF	PFAFH	TF	PFAFH	TF	PFAFH	TF	PFAFH
Heteroskedasticity tests								
Pagan Hall statistic (P-value)	105.89		42.05		123.41		197.00	
	(0.00)		(0.00)		0.00		(0.00)	
Overidentification test of all instruments								
Hansen's J statistic (P-value)	5.819		0.237		12.822		13.444	
	(0.12)		(0.63)		(0.31)		(0.27)	
Endogeneity tests								
χ^2 (P-value)	5.223		13.152		4.914		5.214	
	(0.07)		(0.000)		(0.086)		(0.074)	
Relevance of the instruments								
Shea's partial R^2	0.021	0.002	0.008	0.001	0.062	0.047	0.056	0.025
Standard partial R^2	0.024	0.003	0.009	0.001	0.061	0.046	0.056	0.025
F test (P-value)	42.41	4.56	15.09	3.24	42.77	19.92	94.84	9.95
	(0.01)	(0.04)	(0.00)	(0.03)	(0.00)	(0.00)	(0.00)	(0.00)
Kleibergen and Paap (2006) rk LM statistic (P-value)	10.43		3.124		21.65		22.06	
	(0.03)		(0.21)		(0.04)		(0.04)	
Angrist–Pischke χ^2 (P-value)	210.30	22.88	38.00	9.93	569.24	259.51	1138.69	98.67
	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
Weak identification test								
Kleibergen and Paap (2006) rk Wald F statistic	3.373 [a]		3.182 [b]		19.99 [c]		11.83 [c]	
Angrist–Pischke F statistic [c]	51.38 [d]	5.59 [d]	18.58 [e]	4.85 [e]	46.33 [f]	21.22 [f]	92.72 [f]	8.03 [f]

Stock-Yogo weak ID test critical values:

[a] 5% maximal IV relative bias 13.97; 10% maximal IV relative bias 8.78; 20% maximal IV relative bias 5.91; 30% maximal IV relative bias 4.79; 10% maximal IV size 19.45; 15% maximal IV size 11.22; 20% maximal IV size 8.38; 25% maximal IV size 6.89

[b] 10% maximal IV size 13.43; 15% maximal IV size 8.18; 20% maximal IV size 6.40; 25% maximal IV size 5.45

[c] 5% maximal IV relative bias 19.64; 10% maximal IV relative bias 10.84; 20% maximal IV relative bias 6.21; 30% maximal IV relative bias 4.56; 10% maximal IV size 34.62; 15% maximal IV size 18.84; 20% maximal IV size 13.45; 25% maximal IV size 10.68

[d] 5% maximal IV relative bias 18.37; 10% maximal IV relative bias 10.27; 20% maximal IV relative bias 6.71; 30% maximal IV relative bias 5.34; 10% maximal IV size 24.58; 15% maximal IV size 13.96; 20% maximal IV size 10.26; 25% maximal IV size 8.31

[e] 5% maximal IV relative bias 13.91; 10% maximal IV size 19.93; 15% maximal IV size 11.59; 20% maximal IV size 8.75; 25% maximal IV size 7.25

[f] 5% maximal IV relative bias 21.10; 10% maximal IV relative bias 11.52; 20% maximal IV relative bias 6.53; 30% maximal IV relative bias 4.75; 10% maximal IV size 43.27; 15% maximal IV size 23.24; 20% maximal IV size 16.35; 25% maximal IV size 12.82

Table 6 Estimation results

	External instruments			External plus constructed instruments	Constructed instruments	
	2S-GMM	LIML	FULL	2S-GMM		OLS
(a) Males						
Constant	21.958*** (3.949)	20.712** (10.032)	24.415*** (2.578)	28.283*** (0.595)	28.532*** (0.702)	28.394*** (0.918)
TF	−0.011** (0.005)	−0.013 (0.011)	−0.009** (0.004)	−0.002*** (0.0004)	0.001 (0.001)	−0.0001 (0.0001)
PFAFH	0.333 (0.204)	0.406 (0.501)	0.221* (0.124)	−0.009 (0.009)	−0.007 (0.017)	−0.010*** (0.003)
Scandinavian	−2.804*** (0.844)	−2.782 (1.905)	−2.469** (1.033)	−1.198*** (0.130)	−0.574** (0.259)	−0.835*** (0.164)
Central	−1.289*** (0.480)	−1.422 (0.884)	−1.159*** (0.441)	−0.344** (0.140)	−0.171 (0.226)	−0.257 (0.200)
EducYears	−0.124 (0.079)	−0.157 (0.189)	−0.090 (0.055)	−0.036*** (0.009)	−0.065*** (0.020)	−0.058*** (0.016)
Urb_1	−0.422 (0.434)	−0.453 (1.083)	−0.210 (0.539)	−0.077 (0.120)	−0.478*** (0.184)	−0.307* (0.176)
Urb_2	−0.052 (0.176)	−0.040 (0.280)	0.005 (0.173)	0.003 (0.126)	−0.165 (0.158)	−0.090 (0.166)
Urb_3	0.364* (0.196)	0.342 (0.385)	0.239 (0.209)	0.028 (0.075)	−0.141 (0.119)	0.000 (0.139)
Hsize	1.868** (0.955)	2.144 (2.230)	1.356** (0.664)	0.098 (0.085)	−0.144 (0.112)	−0.051 (0.055)
Age	−0.009 (0.032)	−0.002 (0.077)	−0.029 (0.022)	−0.056*** (0.005)	−0.055*** (0.007)	−0.054*** (0.009)
Log(Inc)	0.037 (0.057)	0.052 (0.102)	0.060 (0.065)	0.008 (0.017)	−0.024 (0.020)	0.000 (0.028)
NumCig	−0.026*** (0.009)	−0.024* (0.013)	−0.023*** (0.009)	−0.020*** (0.006)	−0.019*** (0.006)	−0.020*** (0.007)
PHinactiv	0.907*** (0.435)	1.113 (0.873)	0.756** (0.338)	0.414*** (0.128)	0.274 (0.172)	0.204 (0.246)
MaxGrip	0.063*** (0.009)	0.065*** (0.012)	0.062*** (0.009)	0.055*** (0.004)	0.045*** (0.005)	0.050*** (0.006)
Chronic	0.769*** (0.157)	0.796* (0.420)	0.651*** (0.106)	0.518*** (0.025)	0.496*** (0.035)	0.486*** (0.044)
Drinking	0.659** (0.312)	0.720 (0.635)	0.518* (0.286)	0.230*** (0.075)	0.084 (0.094)	0.104 (0.121)
Homemaker	3.521 (2.814)	4.120 (4.536)	2.944 (2.655)	0.735 (0.886)	0.028 (0.845)	0.756 (0.987)
Retired	1.178*** (0.380)	1.282 (0.865)	0.981*** (0.261)	0.566*** (0.094)	0.610*** (0.111)	0.615*** (0.134)

Table 6 continued

	External instruments			External plus constructed instruments	Constructed instruments	
	2S-GMM	LIML	FULL	2S-GMM		OLS
R^2	0.9462	0.9314	0.9636	0.9809	0.9809	0.06
Number of observations	11364					
(b) Females						
Constant	32.245*** (1.245)	32.257*** (1.490)	31.692*** (1.010)	28.479*** (0.430)	28.041*** (0.551)	27.672*** (0.678)
TF	−0.009*** (0.003)	−0.009*** (0.003)	−0.009*** (0.003)	−0.004*** (0.001)	−0.002** (0.001)	−0.0003 (0.0002)
PFAFH	−0.403*** (0.122)	−0.380* (0.202)	−0.286*** (0.105)	−0.044*** (0.017)	−0.052*** (0.019)	−0.018*** (0.005)
Scandinavian	−2.935*** (0.473)	−2.989*** (0.506)	−2.831*** (0.424)	−1.270*** (0.163)	−0.984*** (0.232)	−0.492** (0.234)
Central	−0.481 (0.299)	−0.464 (0.312)	−0.503* (0.277)	−0.236** (0.104)	−0.182 (0.172)	0.054 (0.250)
EducYears	0.149** (0.068)	0.129 (0.119)	0.082 (0.067)	−0.070*** (0.016)	−0.095*** (0.017)	−0.127*** (0.018)
Urb_1	1.881*** (0.459)	1.698** (0.805)	1.469*** (0.531)	0.496*** (0.101)	0.246 (0.158)	0.218 (0.208)
Urb_2	0.431 (0.310)	0.379 (0.330)	0.418 (0.301)	0.285** (0.127)	0.225 (0.173)	0.301 (0.210)
Urb_3	0.497* (0.297)	0.523 (0.326)	0.550** (0.265)	0.366*** (0.105)	0.399*** (0.113)	0.448** (0.200)
Hsize	0.813** (0.369)	0.756* (0.388)	0.810** (0.352)	0.466*** (0.090)	0.191* (0.114)	0.024 (0.060)
Age	−0.087*** (0.015)	−0.087*** (0.016)	−0.084*** (0.013)	−0.064*** (0.004)	−0.058*** (0.005)	−0.054*** (0.007)
Log(Inc)	0.126** (0.056)	0.123* (0.067)	0.102** (0.046)	−0.012 (0.014)	-0.002 (0.019)	−0.042 (0.025)
NumCig	−0.056*** (0.017)	−0.056*** (0.018)	−0.053*** (0.016)	−0.042*** (0.006)	−0.044*** (0.007)	−0.048*** (0.013)
PHinactiv	1.560*** (0.370)	1.560*** (0.403)	1.460*** (0.333)	1.292*** (0.133)	1.013*** (0.178)	1.019*** (0.189)
MaxGrip	0.118*** (0.019)	0.112*** (0.031)	0.102*** (0.022)	0.091*** (0.008)	0.084*** (0.009)	0.077*** (0.010)
Chronic	0.365*** (0.131)	0.383** (0.188)	0.460*** (0.114)	0.726*** (0.025)	0.734*** (0.025)	0.721*** (0.042)
Drinking	−0.044 (0.379)	−0.060 (0.402)	−0.077 (0.340)	−0.366*** (0.123)	−0.440*** (0.134)	−0.656*** (0.209)
Homemaker	0.117 (0.224)	0.159 (0.330)	0.288 (0.216)	0.685*** (0.057)	0.588*** (0.078)	0.587*** (0.110)

Table 6 continued

	External instruments			External plus constructed instruments	Constructed instruments	
	2S-GMM	LIML	FULL	2S-GMM		OLS
Retired	0.090 (0.260)	0.167 (0.406)	0.293 (0.280)	0.691*** (0.087)	0.702*** (0.092)	0.732*** (0.115)
R^2	0.9137	0.9199	0.9369	0.9678	0.9694	0.09
Number of observations	13711					

*,**,*** Significance at 10, 5, and 1% level, respectively
Standard errors in *parenthesis*

of these coefficients: (a) proportional increases in the FAH and FAFH expenditures are related to lower BMI's (interpretation of the sign of the TF coefficient), (b) male individuals with higher FAFH expenditures do not have statistically significant differences in BMI while female individuals with higher FAFH expenditures have lower BMI (this interpretation comes from the PFAFH coefficient).

The reasons for the different results between males and females are not entirely clear. However, given that we know nothing about the quantity of foods consumed and that the demand for most foods is inelastic, it is possible that this is the result of women buying healthier FAFH. This is also not surprising given that a number of studies have found that women are more interested in nutrition and health issues than men (e.g., Kim et al. 2001; McLean-Meyinsse 2001). These findings may imply that cries for policy intervention that would mandate labeling or some form of warning signs in the FAFH market similar to the situation in the smoking industry may not be entirely appropriate. Initiatives should perhaps target promotion of the nutritional quality of restaurant foods rather than just requiring some form of labeling that would inform consumers about the nutritional content of the foods. Promotion may then be more appropriately targeted at men based on our findings.

We also find differences in results based on geographic location. Specifically, our results suggest that older Northern and Central European males have lower BMIs (Scandinavian, Central) than Southern European males and that older Northern European females have lower BMI's than Southern females. For an average height male, Scandinavians and Central Europeans weigh 5.7 and 2.1 kg less than Southern Europeans, respectively (based on 2S-GMM results with external instruments only). For an average height female, Scandinavian women weigh roughly 7.7 kg less than Southern European women.

Household size and income both positively affect BMI while age negatively affects BMI level. The effect on age is, however, small since males (females) with an age difference of about 10 years differ only by about 0.1–0.5 (0.6–0.9) BMI units. Number of cigarettes smoked per day is negatively related to BMI; while physical inactivity, grip strength, and chronic diseases positively affect BMI level. Drinking behavior positively affects BMI levels of males. Note that there is a negative effect of drinking behavior for females based on the 2S-GMM results (with the constructed instruments). Retired males and females have higher BMI; although, the effect for females is statistically significant in only one of the models.

7 Discussion and conclusions

The primary contribution of this article is the exploration of the relationship between body weight outcomes and FAFH expenditures. To the best of our knowledge, this is the first time this issue is being explored among an increasingly very important segment of the population, the older Europeans who are expected to account for more than 40% of Europe's population by 2050.

To account for the endogeneity of food expenditures when using BMI as the outcome of interest, we estimate a single-equation model using a two-step efficient Generalized Method of Moments. We undertake a series of tests to validate the use of the instrument list and also use LIML and Fuller's modified LIML to cope with weak instruments. We further complement our analysis by exploiting the heteroskedasticity in the first stage regressions to achieve identification and construct instruments following the Lewbel (2010) method. Our results suggest that FAFH expenditures negatively affect BMI for females but has no effect for males. The reasons for these gender differences in the results are not clear. Data on the quantity and quality of food consumed in the FAFH market are needed to assess whether it is the quantity or the quality of the food or both that is making the difference in body weight outcomes of men and women. This food composition issue is important to consider in the future especially since Huffman and Rizov (2010) found a significant effect of composition of expenditures on weight as well as significant gender differences across food groups. For example, they reported that higher expenditures on dairy/eggs and sugars contribute more to females' weight than males, while higher expenditures on meat/fish significantly contribute to weight gain for men but not for women.

Our findings have significant implications given the increasing scrutiny that the food away from home sector is getting from governments and consumer interest groups with regards to the obesity issue. So is this scrutiny justified and thus warrant policy intervention? As previously mentioned, the behavioral economics argument is that individuals have self-control problems and do not make decisions that serve their own interests. This view, however, is complex normatively since it requires that we use something other than the Pareto approach to evaluating welfare. Given our findings, the implication seems to be that a policy intervention related to the FAFH and obesity issue is not warranted, at least in relation to older Europeans.

A limitation of our study is that we are not able to determine if it is the quantity or quality of FAFH that is making the difference between the body weight outcomes of men and women. We currently do not know of any comprehensive dataset that would provide us these data for the entire FAFH sector. Also, given the restrictive time dimension of our dataset, we are unable to explore the dynamics of the relationship between FAFH expenditures and weight. Given data availability, future research should consider these issues, in addition to finding additional instruments to overcome the endogeneity issues.

Acknowledgments This article uses data from release 2 of SHARE 2004. The SHARE data collection has been primarily funded by the European Commission through the 5th framework programme (project QLK6-CT-2001-00360 in the thematic programme Quality of Life). Additional funding came from the US National Institute on Ageing (U01 AG09740-13S2, P01 AG005842, P01 AG08291, P30 AG12815, Y1-AG-4553-01 and OGHA 04-064). Data collection in Austria (through the Austrian Science Foundation, FWF), Belgium (through the Belgian Science Policy Office) and Switzerland (through BBW/OFES/UFES) was nationally funded. The SHARE data collection in Israel was funded by the US National Institute on Aging (R21 AG025169), by the German-Israeli Foundation for Scientific Research and Development (G.I.F.), and by the National Insurance Institute of Israel. Further support by the European Commission through the 6th framework program (projects SHARE-I3, RII-CT-2006-062193, and COMPARE, CIT5-CT-2005-028857) is gratefully acknowledged. For methodological details see Börsch-Supan and Jürges (2005).

Appendix

Table A1 Pagan and Hall (1983) heteroskedasticity tests

	Males	Females
Pagan Hall statistic (*P*-value)	105.890 (0.00)	170.431 (0.00)

Table A2 Endogeneity tests

Endogenous variables	χ^2 (*P*-value)
TF	
Males	5.22 (0.07)
PFAFH	
Females	5.65 (0.06)

Table A3 Standard partial R^2, Shea's partial R^2 and F test of the joint significance of the excluded instruments

Endogenous variables	Shea's partial R^2	Standard partial R^2	F test (*P*-value)
TF			
Males	0.0219	0.0246	42.41 (0.00)
Females	0.0010	0.0200	58.59 (0.00)
PFAFH			
Males	0.0022	0.0030	4.56 (0.00)
Females	0.0014	0.0029	6.24 (0.00)

Table A4 Under-identification tests Kleibergen and Paap (2006) rk LM statistic

	Kleibergen and Paap rk LM statistic (*P*-value)	Angrist–Pischke χ^2 (*P*-value)	
		TF	PFAFH
Males	10.428 (0.03)	210.297 (0.00)	22.884 (0.00)
Females	8.466 (0.07)	58.843 (0.00)	17.084 (0.00)

Table A5 Weak identification tests

	Kleibergen and Paap rk Wald F statistic[a]	Angrist–Pischke F statistic[b]	
		TF	PFAFH
Males	3.373	51.38	5.59
Females	2.716	14.38	4.18

Stock and Yogo (2005) *weak identification test critical values*:

[a] 5% maximal IV relative bias 13.97; 10% maximal IV relative bias 8.78; 20% maximal IV relative bias 5.91; 30% maximal IV relative bias 4.79; 10% maximal IV size 19.45; 15% maximal IV size 11.22; 20% maximal IV size 8.38; 25% maximal IV size 6.89

[b] 5% maximal IV relative bias 18.37; 10% maximal IV relative bias 10.27; 20% maximal IV relative bias 6.71; 30% maximal IV relative bias 5.34; 10% maximal IV size 24.58; 15% maximal IV size 13.96; 20% maximal IV size 10.26; 25% maximal IV size 8.31

Table A6 Hansen's *J* statistic of overidentifying restrictions

	Males	Females
Hansen's *J* statistic	5.819	11.496
(*P*-value)	(0.12)	(0.01)

References

Anderson M, Matsa DA (2009) Are restaurants really supersizing America?UC Berkeley and Northwestern University, Mimeo

Anderson PM, Butcher KF, Levine PB (2004) Maternal employment and childhood obesity. In: Philipson T, Dai C, Helmchen L, Variyam JN (eds) The economics of obesity: A Report on the Workshop Held at USDA's Economic Research Service. Economic Research Service/USDA, pp 14–17

Angrist JD, Pischke J-S (2009) Mostly harmless econometrics: an empiricist's companion. Princeton University Press, Princeton

Bascle G (2008) Controlling for endogeneity with instrumental variables in strategic management research. Strateg Organ 6(3):285–327. doi:10.1177/1476127008094339

Baum CF, Schaffer ME (2007) Enhanced routines for instrumental variables/generalized method of moments estimation and testing. Stata J 7(4):465–506

Baum CF, Schaffer ME, Stillman S (2003) Instrumental variables and GMM: estimation and testing. Stata J 3(1):1–31

Baum CF, Schaffer ME, Stillman S (2007) ivreg2: Stata module for extended instrumental variables/2SLS, GMM and AC/HAC, LIML, and k-clas regression. Boston College Department of Economics, Statistical Software Components S425401 Downloadable from http://ideasrepecorg/c/boc/bocode/s425401.html

Becker GS, Murphy KM (1988) A theory of rational addiction. J Polit Econ 96(4):675–700

Börsch-Supan A (2005) Introduction. In: Börsch-Supan A, Brugiavini A, Jürges H, Mackenbach J, Siegrist J, Weber G (eds) Health, Ageing and Retirement in Europe—first results from the Survey of Health, Ageing and Retirement in Europe. Mannheim Research Institute for the Economics of Aging, Mannheim

Börsch-Supan A, Jürges H (2005) The Survey of Health, Aging, and Retirement in Europe—methodology. Strauss GmbH, Mannheim Research Institute for the Economics of Aging (MEA), Manheim

Bound J, Jaeger DA, Baker RM (1995) Problems with instrumental variables estimation when the correlation between the instruments and the endogeneous explanatory variable is weak. J Am Stat Assoc 90(430):443–450. doi:10.2307/2291055

Bray GA (2004) Medical consequences of obesity. J Clin Endocrinol & Metab 89(6):2583–2589. doi:10.1210/jc.2004-0535

Browning M, Crossley TF, Weber G (2003) Asking consumption questions in general purpose surveys. Econ J 113(491):F540–F567. doi:10.1046/j.0013-0133.2003.00168.x

Cawley J (1999) Rational addiction, the consumption of calories, and body weight. PhD Dissertation, The University of Chicago

Chou S-Y, Grossman M, Saffer H (2004) An economic analysis of adult obesity: results from the behavioral risk factor surveillance system. J Health Econ 23:565–587. doi:10.1016/j.jhealeco.2003.10.003

Chou S-Y, Rashad I, Grossman M (2008) Fast-food restaurant advertising on television and its influence on childhood obesity. J Law & Econ 51(4):599–618. doi:10.1086/590132

Christensen H, Mackinnon AJ, Korten A, Jorm AF (2001) The "common cause hypothesis" of cognitive aging: Evidence for not only a common factor but also specific associations of age with vision and grip strength in a cross-sectional analysis. Psychol Aging 16(4):588–599. doi:10.1037/0882-7974.16.4.588

Currie J, Dellavigna S, Moretti E, Pathania V (2010) The effect of fast food restaurants on obesity and weight gain. Am Econ J Econ Pol 2(3):32–63. doi:10.1257/pol.2.3.32

Cutler DM, Glaeser EL, Shapiro JM (2003) Why have Americans become more obese? J Econ Perspect 17(3):93–118. doi:10.1257/089533003769204371

Dockner EJ, Feichtinger G (1993) Cyclical consumption patterns and rational addiction. Am Econ Rev 83(1):256–263

Dunn RA (2007) Obesity and the availability of fast-food: an instrumental variables approach. In: International health economics association 6th world congress: explorations in health economics, Compenhagen, Denmark
Dunn RA (2010) The effect of fast-food availability on obesity: an analysis by gender, race, and residential location. Am J Agr Econ 92(4):1149–1164. doi:10.1093/ajae/aaq041
Ejerblad E, Fored CM, Lindblad P, Fryzek J, McLaughlin JK, Nyren O (2006) Obesity and risk for chronic renal failure. J Am Soc Nephrol 17(6):1695–1702. doi:10.1681/asn.2005060638
Esposito K, Giugliano F, Di Palo C, Giugliano G, Marfella R, D'Andrea F, D'Armiento M, Giugliano D (2004) Effect of ifestyle changes on erectile dysfunction in obese men: a randomized controlled trial. J Am Med Assoc 291(24):2978–2984. doi:10.1001/jama.291.24.2978
Ewing R, Schmid T, Killingsworth R, Zlot A, Raudenbush S (2003) Relationship between urban sprawl and physical activity, obesity, and morbidity. Am J Health Promot 18(1):47–57. doi:10.1007/978-0-387-73412-5_37
Frank TH (2006) A taxonomy of obesity litigation. U Arkansas Little Rock Law Rev 28(3):427–441
Frederiksen H, Gaist D, Petersen HC, Hjelmborg J, McGue M, Vaupel JW, Christensen K (2002) Hand grip strength: A phenotype suitable for identifying genetic variants affecting mid- and late-life physical functioning. Genet Epidemiol 23(2):110–122. doi:10.1002/gepi.1127
Goldfarb RS, Leonard TC, Suranovic S (2006) Modeling alternative motives for dieting. Eastern Econ J 32(1):115–131
Grundy SM (2004) Obesity, metabolic syndrome, and cardiovascular disease. J Clin Endocr & Metab 89(6):2595–2600. doi:10.1210/jc.2004-0372
Hansen LP (1982) Large sample properties of generalized method of moments estimators. Econometrica 50(4):1029–1054
Huffman SK, Rizov M (2010) The rise of obesity in transition: theory and empirical evidence from Russia. J Dev Stud 46(3):574–594. doi:10.1080/00220380903383230
Kim S-Y, Nayga RM Jr, Capps O Jr (2001) Food label use, self-selectivity, and diet quality. J Consum Aff 35(2):346–363. doi:10.1111/j.1745-6606.2001.tb00118.x
Kleibergen F, Paap R (2006) Generalized reduced rank tests using the singular value decomposition. J Econometrics 133(1):97–126. doi:10.1016/j.jeconom.2005.02.011
Kyureghian G, Nayga RM Jr, Davis GC, Lin B-H (2007) Food away from home consumption and obesity: An analysis by service type and by meal occasion. Selected paper prepared for presentation at the American Agricultural Economics Association Annual Meeting, July 29–August 1, Portland, Oregon
Lakdawalla DN, Philipson TJ (2002) The growth of obesity and technological change: A theoretical and empirical examination. National Bureau of Economic Research Working Paper No W8946
Lakdawalla DN, Philipson TJ (2006) Economics of obesity. In: Jones AM (ed) The Elgar companion to health economics. Edward Elgar Publishing Limited, Northampton pp 72–82
Levy A (2002) Rational eating: can it lead to overweighhtness or underweightness? J Health Econ 21:887–899. doi:10.1016/S0167-6296(02)00036-X
Levy A (2003) A theory of LTR Junk-food consumption. University of Wollongong, Working paper 03-05
Lewbel A (2010) Using heteroskedasticity to identify and estimate mismeasured and endogenous regressor models. Working paper, Department of Economics, Boston College
Long SK (2008) Do for-profit health plans restrict access to care under medicaid managed care? Med Care Res Rev 65(5):638–648. doi:10.1177/1077558708316930
Loureiro ML, Nayga RM Jr (2005) International dimensions of obesity and overweight related problems: an economics perspective. Am J Agr Econ 87(5):1147–1153. doi:10.1111/j.1467-8276.2005.00800.x
McLean-Meyinsse PE (2001) An analysis of nutritional label use in the Southern United States. J Food Distr Res 32(1):110–114
Moreira MJ (2003) A conditional likelihood ratio test for structural models. Econometrica 71(4):1027–1048. doi:10.1111/1468-0262.00438
Murray MP (2006) Avoiding invalid instruments and coping with weak instruments. J Econ Perspect 20(4):111–132. doi:10.1257/jep.20.4.111
Nayga RM Jr (2008) Nutrition, obesity, and health: policies and economic research challenges. Eur Rev Agr Econ 35(3):281–302. doi:10.1093/erae/jbn013
Pagan AR, Hall D (1983) Diagnostic tests as residual analysis. Economet Rev 2(2):159–218
Palkhivala A (2002) Obesity hits elderly harder than younger patients. Special to Doctor's Guide news. http://www.pslgroupcom/dg/220d4e.htm

Philipson TJ, Posner RA (2003) The long-run growth in obesity as a function of technological change. Perspect Biol Med 46(S3):S87–S107

Plantinga AJ, Bernell S (2007) The association between urban sprawl and obesity: is it a two-way street? J Reg Sci 47(5):857–879. doi:10.1111/j.1467-9787.2007.00533.x

Rantanen T, Masaki K, Foley D, Izmirlian G, White L, Guralnik JM (1998) Grip strength changes over 27 yr in Japanese-American men. J Appl Physiol 85(6):2047–2053

Rantanen T, Guralnik JM, Foley D, Masaki K, Leveille S, Curb JD, White L (1999) Midlife hand grip strength as a predictor of old age disability. JAMA 281(6):558–560. doi:10.1001/jama.281.6.558

Rashad I, Grossman M, Chou S-Y (2006) The supersize of America: an economic estimation of body mass index and obesity in adults. Eastern Econ J 32:133–148

Schaffer ME (2002) ivhettest: Stata module to perform Pagan-Hall and related heteroskedasticity tests after IV. Department of Economics, School of Management and Languages, Heriot-Watt University, Edinburgh, Statistical Software Components S428801 Downloadable from http://ideasrepecorg/c/boc/bocode/s428801.html

Shea J (1997) Instrument relevance in multivariate linear models: a simple measure. Rev Econ Stat 79(2):348–352

Smith PK, Bogin B, Bishai D (2005) Are time preference and body mass index associated?: Evidence from the National Longitudinal Survey of Youth. Econ Hum Biol 3(2):259–270. doi:10.1016/j.ehb.2005.05.001

Stock JH, Yogo M (2005) Testing for weak instruments in linear IV regression. In: Andrews DWK, Stock JH (eds) Identification and inference for econometric models: essays in honor of Thomas Rothenberg. Cambridge University Press, New York pp 80–108

Suranovic SM, Goldfarb RS (2006) A behavioral model of cyclical dieting. In: Bolin K, Cawley J (eds) The economics of obesity. Advances in Health Economics and Health Services Research. JAI Press, Amsterdam pp 49–78

Thompson SB (2011) Simple formulas for standard errors that cluster by both firm and time. J Financ Econ 99(1):1–10. doi:10.1016/j.jfineco.2010.08.016

van der Steeg JW, Steures P, Eijkemans MJC, Habbema JDF, Hompes PGA, Burggraaff JM, Oosterhuis GJE, Bossuyt PMM, van der Veen F, Mol BWJ (2008) Obesity affects spontaneous pregnancy chances in subfertile, ovulatory women. Hum Reprod 23(2):324–328. doi:10.1093/humrep/dem371

Wansink B, Huckabee M (2005) De-marketing obesity. Calif Manage Rev 47(4):6–18

Whitmer RA, Gunderson EP, Barrett-Connor E, Quesenberry CP Jr, Yaffe K (2005) Obesity in middle age and future risk of dementia: a 27 year longitudinal population based study. Brit Med J 330(7504):1360–1364. doi:10.1136/bmj.38446.466238.E0

WHO (2006) Obesity and Overweight. Fact Sheet No 311. World Health Organization

Yang Z, Hall AG (2008) The financial burden of overweight and obesity among elderly Americans: The dynamics of weight, longevity, and health care cost. Health Serv Res 43(3):849–868. doi:10.1111/j.1475-6773.2007.00801.x

You W, Davis GC (2010) Household food expenditures, parental time allocation, and childhood overweight: An integrated two-stage collective model with an empirical application and test. Am J Agr Econ 92(3):859–872. doi:10.1093/ajae/aap031

Zhang L, Rashad I (2008) Obesity and time preference: The health consequences of discounting the future. J Biosoc Sci 40(1):97–113. doi:10.1017/S0021932007002039