MS ritgerð
í heilsuhagfræði

Genetic Instruments for Body Mass Index

Inga Lára Karlsdóttir

Leiðbeinandi: Tinna Laufey Ásgeirsdóttir
Hagfræðideild
Febrúar 2015

HÁSKÓLI ÍSLANDS

Genetic Instruments for Body Mass Index

Inga Lára Karlsdóttir

Lokaverkefni til MS -gráðu í hagfræði

Leiðbeinandi: Tinna Laufey Ásgeirsdóttir

Hagfræðideild

Félagsvísindasvið Háskóla Íslands

Febrúar 2015

Genetic Instruments for Body Mass Index

Ritgerð þessi er 30 eininga lokaverkefni til MS prófs við Hagfræðideild, Félagsvísindasvið Háskóla Íslands.

Prentun: Háskólaprent ehf.

Reykjavík, 2015

Formáli

Verkefni þetta er meistararitgerð í heilsuhagfræði og er vægi hennar 30 ECTS einingar.

Leiðbeinandi minn er Tinna Laufey Ásgeirsdóttir, doktor í hagfræði og dósent við Hagfræðideild Háskóla Íslands. Ég vil þakka henni fyrir góðar ábendingar og samvinnu. Ég vil jafnframt þakka starfsfólki Hjartaverndar fyrir aðstoð við gerð verkefnisins, þá sérstaklega Gauta Kjartani Gíslasyni fyrir vinnu með gögn og Thor Aspelund fyrir ómetanlegan stuðning og kennslu. Jökull Úlfarsson, Anna Kristín Karlsdóttir og Rowan Cain fá þakkir fyrir yfirlestur.

Abstract

Background: Instrumental variable analysis has been used in health economics, e.g. to find causal effects of health on labor-market outcomes. With advanced knowledge in genetics there has been a growing interest of using genetic information as instruments. The aim of this paper is to examine the quality of molecular genetic variants as instruments for body mass index (BMI).

Data and Methods: The data used is from the Icelandic Heart Association, the Reykjavik Study (RS) established in 1967 and the Age, Gene/Environment Susceptibility–Reykjavik Study (AGES-RS) initiated in 2002. Participants are men and women born 1907-1935 in Reykjavík. Genetic variants or single nucleotide polymorphisms (SNPs) found to be robustly associated with BMI in a Genome Wide Association Study (GWAS) was used as instruments in the analysis. We made regressions where the SNPs were both used as a set of instruments and as a weighted genetic risk score (GRS).

Results: First stage regressions show the instruments to be too weak to serve as instruments for BMI. The F-statistics result in a value of around 2, far below the minnimum of 10 that is often used as a threshold.

Conclusion: This paper supports and further reinforces the literature in that far stronger genetic instruments are needed for BMI than are available to date. Therefore, results with such instruments need to be cautiously interpreted.

Table of contents

List of Figures and Tables

1 Introduction

Obesity[1] has risen dramatically in the last decades (WHO, 2000) with the OECD average of 18 percent among adults (OECD, 2014). Obesity is a risk factor for various diseases. For example, myocardial infarction, stroke, type 2 diabetes, hypertension, osteoarthritis and depression (Dixon, 2010). There are other economic effects of obesity as well. Health-care costs, for example have been estimated to be 41.5 percent higher for the obese, than for normal-weight individuals (Finkelstein, Trogdon, Cohen and Dietz, 2009) or even up to twice the amount when using instrumental-variables (IV) estimations (Cawley and Meyerhoefer, 2012).

The consequences of this change in body weight also include potential labor-market effects. Several previous studies have found, among females especially, a negative correlation between body weight and wages, but the findings among men are less apparent (Asgeirsdottir, 2011; Averett and Korenman, 1996; Baum and Ford, 2004; Cawley, 2000; Cawley, 2004; Johansson, Böckerman, Kiiskinen and Heliövaara, 2009 and Pagan and Davila, 1997). However, the causal direction of this relationship, let alone the specific reasons are still not fully worked out.

There are various pathways through which the observed relationship between weight and labor-market outcomes could work. Cawley (2000; 2004) identified three reasons for this correlation. First, obesity may affect the labor market, for example by lowering productivity due to health constraints or because of discrimination. The second explanation is that labor-market outcomes affect weight, for example, due to depression effects of unemployment or poor labor-market standing that could lead to weight gain. The third category of explanations is that weight and labor-market outcomes may be correlated with unobserved variables. One example of such a variable

[1] The body mass index (BMI) adjusts body weight for height and is calculated as weight in kilograms divided by height in meters squared. Weight categories are defined in the following way; BMI for normal weight individuals is equal to 18.5 and ranges up to 25.0, those with BMI equal to 25.0 and up to 30.0 are considered overweight, BMI equal to and above 30.0 is in the obese range and underweight individuals are those with BMI below 18.5 (NIH, 1998).

is the individual rate of time discounting. Those who have a higher marginal rate of time preference may invest little in their human capital, for example training and thus have low wages. At the same time they may be less concerned about long-term effects on health and be more prone to weight-gaining consumption. Some scientists believe that because of this obese workers might have a flatter earnings profile over time (Baum and Ford, 2004 and Lindeboom, Lundborg and van der Klaauw, 2010).

Due to the ethical and practical constraints of conducting experiments when examining the effects of obesity, this literature has been plagued with methodological challenges and previous findings may well be partly due to the endogeneity of weight. To combat this scientists have applied different methods to estimate the causal effect of obesity and labor-market outcomes. Some have made use of the fact that prior research in behavioral genetics suggest that roughly half of the variation in BMI is genetic in origin (Comuzzie and Allison, 1998) and there are even studies that find up to 72 percent of the variation in obesity to be due to genetic factors (Cutler and Glaeser, 2005).

Following this general approach, some have used differences to another individual with highly correlated genes, like a parent, same sex sibling or a twin (for an example see: Averett and Korenman (1996), Baum and Ford (2004) and Cawley (2004)). This is done in order to difference away unobserved heterogeneity assuming that it is constant among the two family members. Others have corrected for endogeneity by directly or indirectly applying IV models. Some researchers have used BMI of a closely related individual as an instrument (for an example see: Cawley (2000; 2004), Cawley and Meyerhoefer (2012) Gregory and Ruhm (2009) and Lindeboom et al. (2010)) while others have used molecular genetic variation as instruments.

IV analysis is a statistical method that is used substantially in various fields of health economics (Cawley, Han and Norton, 2011 and Wehby, Ohsfeldt and Murray, 2008). Instruments (Z) are variables that are known to be related to the endogenous variable (X) but are assumed to have no direct or undirect connection to the outcome (Y), besides through X (Conley, 2009 and Wehby et al., 2008). The technique became known within epidemiology as ´Mendelian randomization‘ because of the random assignment of alleles parents pass on to their children (Wehby et al., 2008). Wehby et al. (2008)

suggested that instead of the term ´Mendelian randomization‘ researchers use the term ´instrumental-variable analysis with a genetic variant´ to avoid communication barriers across fields. We thus use the term instrumental variables here, the idea being that using single nucleotide polymorphism (SNP) as IVs serves as a natural experiment to find the causal impact of weight on labor-market outcomes since some participants have obesity markers while others don´t .

The term genoeconomics was introduced in 2007 in a paper by Benjamin et al. (2007). Ding, Lehrer, Rosenquist and Audrain-McGovern (2006) were the first to use molecular genetic instruments in economic analysis to study how health affects education. Subsequently many thought extending genetic data to economics in order to identify causal pathways to be promising (Beauchamp et al., 2011; Benjamin et al., 2007; Benjamin et al., 2012a; Cawley et al., 2011 and Ding et al., 2006) and believed that the use of genes as IVs would become widespread (Cawley et al., 2011). At the same time other believe that most research has reached false positive genetic associations or overestimates of true effect sizes (Benjamin et al., 2012a; Chabris et al., 2012 and Chabris et al., 2013). As an example Beauchamp et al. (2011) failed to replicate results from a Genome Wide Association Study (GWAS) of educational attainment. Similarly Benjamin et al. (2012a) failed to replicate associations between a particular genetic variant and educational attainment as well as cognitive function in three samples. In an attempt to replicate published associations of 12 genotypes of general intelligence, only one variant replicated in one out of three samples (Chabris et al., 2012) and in a study of economic and political preferences no significant associations with any set of traits were found (Benjamin et al., 2012b). The results are similar for self-employment where van der Loos et al. (2013) failed to replicate reported associations.

For socioeconomic traits the causal chain from genetic to phenotypic variation is thought to be long, explaining the terms distal and proximal phenotypes. Distal phenotypes are most likely affected by many genotypes, each one with a very small influence. Another problem is that only a small fraction of the variation in the phenotypes can be explained by the genetic markers that have been found. With instrumental-variable analysis, associations of such tiny effect sizes raise the concern of

weak instruments and therefore underpowered research (Beauchamp et al., 2011; Benjamin et al., 2012a and Burgess and Thompson, 2011). The mentioned concerns are among those that have led some scientists to believe that it is unlikely that genetic variants of interest to economists exist or will ever be found (Beauchamp et al., 2011; Chabris et al., 2013 and Conley, 2009).

The aim of this paper is to examine the quality of molecular genetic information as instruments for BMI. The first locus found to affect BMI was the fat mass and obesity associated (FTO) genotype, also explaining the biggest part of the known variance in BMI (Frayling et al., 2007; Li et al., 2010 and Speliotes et al., 2010). Doubt has already been cast on the FTO locus as a genetic instrument for BMI because of violation of assumption for the IV approach, that the instrument should only affect the outcome variable (mental disorder) through BMI. The association of the FTO alleles and obesity were strong and significant for men but not for women (Kivimaki et al., 2011). Results differed in a Swedish cohort study of men born 1920-1924 where the FTO variant was not associated with BMI (Jacobsson et al., 2009). For a physical trait like BMI there are also influential behavioral factors such as what a person likes to eat. There is some evidence that a genetic variant in the FTO gene affects appetite or preference for energy dense foods (Cecil, Tavendale, Watt, Hetherington and Palmer, 2008).

Other loci have been uncovered in GWAS and found to be robustly associated with BMI. The large GWAS meta-analysis by Speliotes et al. (2010) with 32 BMI-related SNPs is a leading paper. Together the SNPs have been found to account for 1.45% of the variance in BMI (Speliotes et al., 2010), which may be considered small, but genetic effects on phenotypes like BMI are most often small (Chabris et al., 2013 and Li et al., 2010).

In a GWAS of Chabris et al. (2013) the wrong-signed SNPs were the most statistically significant, but out of 17 known BMI loci included in their data 11 had estimated effect sizes of the correct sign. A genetic risk score (GRS) of 32 SNPs was a statistically significant predictor of BMI among whites, with R^2 equal to 0.13 (Belsky et al., 2013). The GRS was constructed from published GWAS through 2010, including the one of Speliotes et al. (2010), a paper which was also used for the effect-size weighting for the score. The GRS from Belsky et al. (2013) was at least as predictive as a GRS generated

from a GWAS meta-analysis on BMI of Speliotes et al. (2010) and it indicates that the GRS performed similarly among men and women (Belsky et al., 2013). Another study using the previously mentioned meta-analysis by Speliotes et al. (2010), among other GWAS, managed to replicate association of 12 SNPs and BMI with a large dataset. The same meta-analysis was used for effect size weighting of a score that could explain 0.97 percent of the variation in BMI (Vimaleswaran et al., 2013). Yet another study using effect sizes from Speliotes et al. (2010) research for a score of three SNPs affecting BMI proved to be statically significant (Nordestgaard et al., 2012). Similarly Peterson et al. (2014) found 7 of the 32 SNPs to be associated with BMI in their data and they found a GRS to account for 3.1% of the variance in BMI (the sum score but the weighted score resulted in 0.6-0.9% of the variance accounted for).

This paper thus reveals that it´s not the same molecular genetic variants that prove to be significantly associated with BMI across the studies. As expected from the GWAS of Speliotes et al. (2010) the variants don´t explain a big part of the variance in BMI and thus it is likely to result in weak instruments when applied as instrument in IV analysis. Weather genetic variation is a good instrument for body weight, thus remains an open hypothesis.

Before substantial research is carried out using this methodology and policy action taken on its basis, it may be worth a while to give the foundation on which this literature is based some thought, as is done in the current paper. In our study with Icelandic data we tested if SNPs robustly associated with BMI from the GWAS of Speliotes et al. (2010) hold up against the criteria needed to serve as good (strong and valid) instruments for BMI. The results show it to be too weakly associated with BMI to meet the standard of a good IV.

2 Data and Methods

The Icelandic Heart Association started the Reykjavik study (RS) in 1967. It is a population-based cohort of 30,795 men and women born in 1907-1935 and living in Reykjavik. Participants were surveyed between 1967 and 1991. The Age, Gene/Environment Susceptibility–Reykjavik Study (AGES-RS) is a sample of surviving participants from the RS, initiated in 2002. Genetic information was gathered from 3,200 people in the AGES-RS while BMI measures are from RS and collected as close to participants middle age as possible. Figure 1 shows the number of participants with respect to datasets and variables. The number of participants narrows down when variables are added to the analysis. Numbers are for the full sample and the figure shows number of participants with necessary data for single stage and two stage regressions. The data is thoroughly described in Harris et al. (2007).

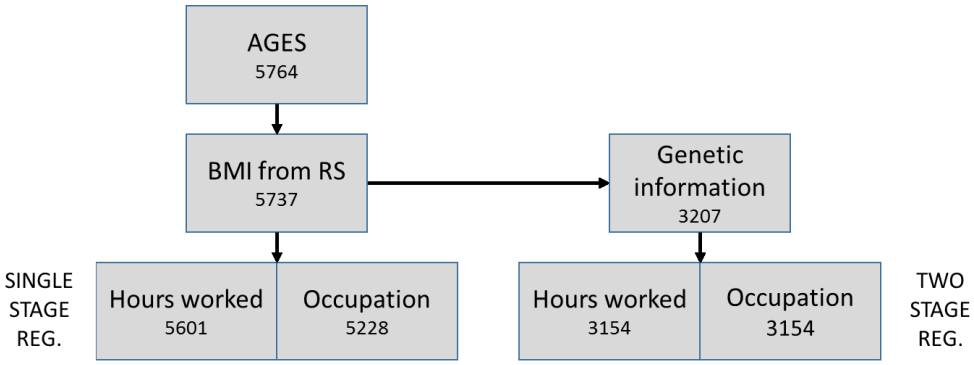

Figure 1. Number of participants with respect to datasets and variables.

The measure of body composition is body max index. Despite its shortcomings, BMI is widely used in social science research for classification, as it is inexpensive to measure in large samples with self-reports. Although it is always a shortcoming of this measure, that it doesn´t distinguish between fat and other tissue, it is still favorable that the measurements from RS are not based on self-reported height and weight as is often the case, but was measured by professionals. In this paper BMI is used as a continuous variable.

The focus in the current examination is on the quality of molecular genetic variation as instruments, rather than the effects of the endogenous variable BMI on the outcome. The appropriate quality measures are obtained using traditional two-stage IV methods but only through the effect of the genetic variants chosen. In the first stage regression we regress the phenotype BMI on the SNPs and in the second stage we regress an outcome variable on the fitted values of BMI from the first stage (Burgess and Thompson, 2011 and Lawlor, Harbord, Sterne, Timpson and Smith, 2008).

Although the focus is on the first stage regressions, we employ two second-stage labor-market outcomes as cases in point. Specifically those are hours worked and occupation. The former outcome variable, hours worked per week is based on two questions from the RS cohort. In those questions, the participants chose one of four possible categories, indicating how many hours per week they work on average in a main job and, if applicable, an extra job. As the distribution of hours worked is positively skewed we take its logarithm and assume a linear association of log hours and BMI.

The latter outcome variable is people's occupation coded into a binary variable, taking the value one for managers and professionals, and zero for all other occupations. The data comes from the AGES-RS where people defined their career after retirement. We apply logistic regression models to examine odds ratios (OR) of the set of instruments and GRS for BMI. The categories of occupation are based on International Standard Classification of Occupation (ILO, 1990). Since labor-market behavior and the accumulation and distribution of fat on the body differ considerably by gender, all estimations are performed for females and males separately, although also reported for the full sample in the first stage regressions. The only control variable used in the regressions is age but the Icelandic population is unusually homogenous, thus not calling for the use of other controls, such as race. Summary statistics on those main variables are reported in table 1. We control for heteroskedasticity and report robust standard errors for the OLS models and 2SLS models with the outcome variable.

Table 1. Summary statistics.

	MEAN/FREQUENCY	
VARIABLE	**Female**	**Male**
AGE (N= 1856; 1333)	52.0 (6.6)	49.7 (5.9)
BMI (N= 1856; 1351)	24.9 (3.8)	25.6 (3.1)
DEPENDENT VARIABLES		
LOG HOURS (N=1821;1333)	3.9 (0.2)	4.1 (0.3)
HOURS WORKED (N=1821;1333)	62.2 (14.2)	51.4 (9.2)
OCCUPATION (N=1820; 1345)		
SECTOR 1	**13.7%**	**25.4%**
SECTOR 0	86.3%	74.6%

Standard deviations are in parenthesis.

We used genetic variants as instruments for BMI, more specifically different genetic variants of SNPs called alleles. In a GWAS by Speliotes et al. (2010), a total of 249,769 individuals of European ancestry, which the Icelandic data was part of, 32 SNPs were found to be robustly associated with BMI. To guarantee a positive relation of BMI and the SNPs, a linear regression was performed for each SNP in order to turn around the variants which showed a negative relation to BMI. The SNPs were both used as a set of instruments, where all 32 SNPs were combined in a single regression and as a weighted GRS. A GRS is a single variable with cumulative effects of the SNPs weighted with BMI effect sizes from the GWAS mentioned earlier (Belsky et al., 2013 and Speliotes et al., 2010). The research by Belsky et al. (2013) supported a linear association between the GRS and BMI.

There are two requirements for an instrument, it must be powerful and valid (Baiocchi, Cheng and Small, 2014 and Cawley and Meyerhoefer, 2012). We follow Bound, Jaeger and Baker (1995) suggestion and report both the F-statistic from the first stage regressions and R^2 to estimate the instruments strength. For an instrument to be powerful it must explain the variation in the phenotype. If the correlation of the instruments and the endogenous variable is weak, then only a weak correlation

between the IVs and the error in the main equation can result in largely biased IV estimates, even larger than in the OLS estimate (Bound et al., 1995). For evaluation we used F-statistics from the first-stage regression, with values greater than 10 taken as evidence against weak instruments (Staiger and Stock, 1997 and Stock and Yogo, 2002). The F-statistic depends on the sample size, therefore increasing the sample size can reduce bias (Bound et al., 1995 and Burgess and Thompson, 2011).

For instrumental variables to be valid the following assumption must hold. In context with our model it means that: (1) the SNPs (the genotype) and the BMI (the phenotype) must be consistently associated, (2) for the exclusion restriction to hold, the SNPs can affect the outcome only through BMI and (3) the SNPs have to be independent of unmeasured confounders affecting the outcome (Angrist, Imbens and Rubin, 1996; Baiocchi et al., 2014; Didelez and Sheehan, 2007; Lawlor et al., 2008; Taylor et al., 2014 and Wehby et al., 2008). The first assumption, the association of BMI and the SNPs can be easily evaluated (Glymour, Tchetgen and Robins, 2012) but it is not a requirement that the instrument is causally related to the phenotype (Didelez and Sheehan, 2007 and Lawlor et al., 2008). The second and third assumptions are harder to prove (Glymour et al., 2012 and Lawlor et al., 2008) and isn´t done in this paper.

Statistical procedures are carried out using Stata 12. Participants in both RS and AGES-RS signed informed consent before they enrolled in the study. The AGES-RS study was approved by the National Bioethics committee and the Data Protection Authority (VSN AGES 00-063 and PV AGES 2002050228 MS/-- resepectively).

3 Results

First stage regression results (from an instrumental variable regression on log(hours)) are presented in table 2, with coefficients for SNPs. Some SNPs have statistically significant OLS coefficients, more for females than males and most often for the full sample. However, only one SNP, rs2815752 near NEGR1 is significant at the 5% level for both genders and the full sample. Coefficients have both positive and negative signs, despite having been coded in such a way that if estimated in accordance with the previous literature, they were all expected to be positive. Having said that, the statistically significant SNPs are all positive. The same analysis was conducted with a genetic risk score (GRS) as an instrument, resulting in non-significant coefficients for the GRS as seen in table 3.

Table 2. Coefficients from first stage regressions with log(hours) as the outcome variable and 32 SNPs as set of instruments.

VARIABLE	NEAREST GENE	FULL SAMPLE	FEMALES	MALES
		dy/dx	dy/dx	dy/dx
MALE		0.921 (0.12)		
AGE		0.064 0.01	0.085 (0.01)***	0.027 (0.01)*
RS4929949	RPL27A	0.023 (0.08)	-0.009 (0.12)	0.044 (0.12)
RS10767664	BDNF	0.207 (0.10)**	0.193 (0.14)	0.215 (0.15)
RS4771122	MTIF3	0.008 (0.10)	-0.119 (0.14)	0.234 (0.15)
RS2241423	MAP2K5	0.112 (0.10)	0.267 (0.14)*	0.234 (0.15)
RS12444979	GPRC5B	0.214 (0.14)	0.363 (0.19)*	0.039 (0.21)
RS571312	MC4R	0.274 (0.10)**	0.193 (0.14)	0.351 (0.14)*
RS2287019	QPCTL	0.118 (0.11)	0.153 (0.15)	0.044 (0.14)
RS543874	SEC16	0.092 (0.11)	0.016 (0.16)	0.172 (0.15)
RS1514175	TNN13K	0.031 (0.09)	0.116 (0.12)	-0.074 (0.12)
RS1555543	PTBP2	0.041 (0.09)	-0.068 (0.13)	0.162 (0.12)
RS2867125	TMEM18	0.197 (0.11)*	0.353 (0.16)**	-0.012 (0.16)
RS713586	RBJ	0.101 (0.09)	0.087 (0.13)	0.101 (0.12)
RS887912	FANCL	0.105 (0.10)	0.041 (0.14)	0.182 (0.14)
RS9816226	ETV5	0.216 (0.10)**	0.319 (0.15)**	0.079 (0.15)
RS2112347	FLJ35779	0.151 (0.09)	0.151 (0.13)	0.148 (0.13)
RS4836133	ZNF608	0.232 (0.09)**	0.323 (0.13)**	0.130 (0.13)
RS10968576	LRRN6C	-0.030 (0.09)	-0.060 (0.13)	0.004 (0.13)
RS1558902	FTO	0.249 (0.10)***	0.309 (0.14)**	0.135 (0.13)
RS10938397	GNPDA2	0.163 (0.09)*	0.218 (0.13)	0.128 (0.13)
RS2815752	NEGR1	0.404 (0.09)***	0.278 (0.13)**	0.573 (0.12)***
RS7359397	SH2B1	0.189 (0.09)**	0.219 (0.13)*	0.210 (0.12)*
RS3817334	MTCH2	-0.025 (0.09)	-0.011 (0.12)	-0.003 (0.12)
RS987237	TFAP2B	0.129 (0.12)	0.0435 (0.16)	0.228 (0.17)
RS7138803	FAIM2	0.224 (0.09)**	0.281 (0.14)**	0.156 (0.13)
RS11847697	PRKD1	0.200 (0.23)	0.241 (0.33)	0.162 (0.34)
RS13107325	SLC39A8	0.102 (0.35)	-0.181 (0.43)	0.718 (0.62)
RS3810291	TMEM160	0.219 (0.11)**	0.203 (0.16)	0.282 (0.14)**
RS13078807	CADM2	0.135 (0.11)	0.198 (0.16)	0.075 (0.14)
RS2890652	LRP1B	0.083 (0.11)	-0.117 (0.16)	0.296 (0.16)*
RS206936	NUDT3	0.151 (0.11)	0.203 (0.17)	0.064 (0.15)
RS10150332	NRXN3	0.105 (0.11)	0.198 (0.15)	-0.010 (0.16)
RS29941	KCTD15	-0.174 (0.09)*	-0.258 (0.13)**	-0.036 (0.12)
CONS		17.29 (0.79)	15.54 (1.10)***	20.82(1.07)***
		N = 3154	N = 1821	N = 1333
		R^2 = 0.051	R^2 = 0.057	R^2 = 0.045
		F = 2.927	F = 2.126	F = 2.053

Robust standard errors are in parentheses.

R^2 and F are first stage summary statistics.

*Significant at 10% level, **significant at 5% level, ***significant at 1% level.

Table 3. Coefficients from first stage regression with log hours as the outcome variable and GRS as an instrument.

VARIABLE	FULL SAMPLE	FEMALES	MALES
	dy/dx	dy/dx	dy/dx
AGE	0.066(0.01)***	0.088 (0.01)***	0.028 (0.01)**
GRS	-0.067 (0.12)	-0.044 (0.17)	-0.111 (0.16)
CONS.	21.55(0.55)***	20.34 (0.738305)***	24.41 (0.76)***
	N =3154	N = 1821	N = 1333
	R^2 = 0.025	R^2 = 0.023	R^2 = 0.003
	F =0.314	F =0.069	F = 0.471

Robust standard errors are in parentheses.
R^2 and F are first stage summary statistics.
*Significant at 10% level, **significant at 5% level, ***significant at 1% level.

The F-statistics, reported in table 2, relating to the hypothesis that the first-stage coefficients of the instruments are jointly equal to zero were 2.126 for females, 2.053 for males and 2.927 for the full sample. None of the F-values for either the set of instruments or GRS, in table 3 exceed the minimum of 10 as suggested by Staiger and Stock (1997) and should thus be declared as weak (Staiger and Stock, 1997 and Stock and Yogo, 2002). The R^2 are likewise very low.

To give examples we performed regressions with two outcome variables, log(hours) and occupation. Results from those estimations can be found in table 4. For both outcome variables we checked non-linear specifications of BMI but they all proved to be weakly associated with the outcome variables, resulting in opposite the sign from what was expected or no correlation at all. Other adiposity indicators where tested, such as clinical classifications of BMI, but they did not show any correlation with the outcome variables. BMI coefficients from regression of log(hours) on BMI are reported in table 3; both OLS estimates from single-stage regressions and 2SLS estimates from IV-regressions, with SNPs as a set of instruments and the GRS. Coefficients from the single stage are small but statistically significant at the 5% level for both genders, showing the opposite sign from what was expected based on the previous literature, i.e. those who work more tend to have higher BMI in this sample. The BMI coefficients are not statistically significant for either type of instruments.

The latter example with occupation as a binary variable is also regressed on BMI, both in a single stage and in a two-stage estimation with the instruments. No odds

ratios, presented in table 4, were statistically significant. There was furthermore no association detectable between occupation and BMI.

Table 4. Marginal effects of BMI on log hours and occupation.

VARIABLE		SINGLE STAGE	TWO STAGE IV- GRS	TWO STAGE IV-SNP
		Log hours†		
	OLS	dy/dx	dy/dx	dy/dx
FEMALES		0.003 (0.002)**	0.218 (0.88)	-0.005 (0.01)
		N= 3221	N= 1821	N= 1821
		R^2=0.065	R^2=0.023	R^2 = 0.057
MALES		0.003 (0.001)**	0.007 (0.09)	0.002 (0.01)
		N=2380	N=1333	N=1333
		R^2=0.041	R^2=0.003	R^2=0.045
		Occupation††		
	LOGISTIC REGRESSION	OR	OR	OR
FEMALES		0.990 (0.01)	0.156 (0.38)	1.038 (0.10)
95% CI		(0.962 1.019)	(0.001 18.78)	(0.858 1.256)
		N=2966	N=1810	N=1810
MALES		1.003 (0.02)	1.480 (1.92)	1.088 (0.11)
95% CI		(0.097 1.03)	(0.117 18.73)	(0.893 1.326)
		N=2262	N=1344	N=1344

Robust standard errors are in parentheses for log hours.
Standard errors are in parentheses for occupation.
†Logarithm of hours worked.
†† Occupation is coded into two sectors from the international standard classification of occupation, 1 for managers and professionals and 0 for others.
*Significant at 10% level, **significant at 5% level, ***significant at 1% level.

4 Discussion

In this study we tested the legitimacy of genetic variation as an instrument for BMI. We failed to replicate the robust associations found by Speliotes et al. (2010), but recently scientists have also failed to replicate associations of molecular genetic variants and economic traits (Beauchamp et al., 2011; Benjamin et al., 2012a; Benjamin et al., 2012b; Chabris et al., 2012 and van der Loos et al., 2013). The reason may be that while genetic factors science has uncovered are many, they collectively explain only a small fraction of the total variance in the phenotype, as is the case for BMI. Thus weak instruments with corresponding biases are a concern in genoeconomics.

Results from different data sets of the association of BMI and genetic markers are not conclusive, which raises the concern if other researchers find the association to be significant, just by a coincidence. Publication biases, where estimations in which genetic instruments don´t hold up, don´t find their way to journal pages, may exacerbate the problem. Most likely there are researchers who have embarked on projects to estimate BMI effects with data similar to what is used in this study, who have bias aborted the projects when they found their instruments not to hold up.

The time of the current BMI data collection was earlier than what has been seen by others who found the relationship to be statistically significant (see for example: Belsky et al. (2013) and Nordestgaard et al. (2012)). This may be taken as an example of suspicious patterns. If the genetic instruments really affect BMI directly through biological mechanisms they should not be confounded by social, cultural or institutional factors. Valid instruments of this kind should hold up against changes in context, such as for different generational cohorts regardless of the prevalence of obesity. However, it could be that the genetic variation measured affects other variables that have become increasingly related to BMI over the years.

Many socioeconomic studies are troubled with endogeneity. IV analysis is supposed to correct this, but the use of genetic information for instrumentation needs a closer attention. Finding a good instrumental variable is desirable although difficult, since strong instruments are not the only requirement, they also need to be valid. Our

instruments didn´t proof to be powerful, but even if it's possible to come around the weak instrument bias there are problems one needs to be aware of when genetic information is applied. For instruments to be valid they also need to fulfill the core assumptions as previously described. Although the SNPs we used as instruments are detected in genome-wide meta-analysis for BMI, we don´t know the mechanism through which they work. That is, we do not know how the genotypes affect the phenotype BMI. It is important to keep in mind the possibility that there are many ways that genotypes can directly affect phenotypes for both physical- and behavioral traits (Chabris et al., 2013 and Glymour et al., 2012). The SNPs might be in linkage disequilibrium (LD) with the causal variant affecting BMI, i.e. it is physically close allele and is co-inherited with the SNP used in the study (Lawlor et al., 2008). Despite LD, the chosen variants might fulfill the assumptions to be valid instruments if there are no other pathways from the genetic variant in LD with the instrumental variable to the outcome variable (Glymour et al., 2012 and Lawlor et al., 2008). If the genetic instrument is in LD with another genetic variant that directly or indirectly affects the outcome variable it violates the exclusion restriction (Didelez and Sheehan, 2007 and VanderWeele, Tchetgen Tchetgen, Cornelis and Kraft, 2014), as well as the third assumption that the instrument has to be independent of any confounders affecting the outcome variable (Didelez and Sheehan, 2007). We are not aware of any research that can erase the concern of LD problem with the outcome variables used as an example in this paper.

Another problem that may arise is population stratification, possibly violating the exclusion restriction (VanderWeele et al., 2014) and the third assumption (Didelez and Sheehan, 2007 and Lawlor et al., 2008). The concept refers to populations subgroups that experience different distribution of BMI and different frequencies of the genetic variants (Didelez and Sheehan, 2007 and Lawlor et al., 2008). In a paper by Harris et al. (2007) on the Icelandic Heart association data this problem is thought to be improbable, but Benjamin et al. (2012a) try to deal with population stratification in the same data since they believe that even in Iceland there may be ethnic stratification.

Yet another concern is that most genetic variants have multiple functions, called pleiotrophy in the genetic literature. This will not violate any of the core assumptions if

the genetic variant is associated with pleiotrophic effects that do not affect the outcome. If the variant affects an unmeasured phenotype that also affects the outcome variable it results in invalidation of the IV approach, both the ER and the third assumption (Glymour et al., 2012 and Lawlor et al., 2008). This is very likely one of our problems. However, despite our awareness of it, we cannot verify if that is the case.

This paper is in line with findings from Chabris et al. (2013) suggesting that far stronger genetic instruments are needed for IV analysis with BMI as a phenotype than have been available to date. As the discussion reveals, we don´t know the function of the genetic variants in this study rather than many others who apply genetic markers as instruments with different phenotypes. We aren´t optimistic that analysis with molecular genetic instruments will benefit economists or other science in the near future or until genetic knowledge has advanced. In the early days of genetic instrumentation in economics Conley (2009) was one of those who were pessimistic of ever finding a good instrument of interest to economists. A recent paper with consideration on the method suggest that it might be more reliable when applied to provide evidence that a variable doesn´t affect an outcome instead that it does (VanderWeele et al., 2014). Baiocchi et al. (2014) didn´t express it so forcefully, but indicated that even if there is no such thing as a perfectly valid IV the analysis may still provide useful information about the treatment effect or phenotype. Anyhow, there are obstacles in the way scientists need to get passed before genetic data will be promising as instruments. Whatever the case, it is important to be careful when results with genetic instruments are interpreted.

References

Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, *91*(434), 444-455. doi: 10.2307/2291629

Asgeirsdottir, T. L. (2011). Do body weight and gender shape the work force? The case of Iceland. *Economics & Human Biology, 9*(2), 148-156. doi: http://dx.doi.org/10.1016/j.ehb.2010.12.001

Averett, S., and Korenman, S. (1996). The economic reality of the beauty myth. *Journal of Human Resources, 31*(2), 304-330. doi: 10.2307/146065

Baiocchi, M., Cheng, J., and Small, D. S. (2014). Instrumental variable methods for causal inference. *Statistics in Medicine, 33(*13), 2297-2340. doi: 10.1002/Sim.6128

Baum, C. L., and Ford, W. F. (2004). The wage effects of obesity: a longitudinal study. *Health Economics, 13*(9), 885-899. doi: 10.1002/hec.881

Beauchamp, J. P., Cesarini, D., Johannesson, M., van der Loos, M. J. H. M., Koellinger, P. D., Groenen, P. J. F., et al. (2011). Molecular Genetics and Economics. *Journal of Economic Perspectives, 25*(4), 57-82. doi: 10.1257/jep.25.4.57

Belsky, D. W., Moffitt, T. E., Sugden, K., Williams, B., Houts, R., McCarthy, J., and Caspi, A. (2013). Development and Evaluation of a Genetic Risk Score for Obesity. *Biodemography and Social Biology, 59*(1), 85-100. doi: 10.1080/19485565.2013.774628

Benjamin, D. J., Cesarini, D., Chabris, C. F., Glaeser, E. L., Laibson, D. I., Guðnason, V., et al. (2012a). The Promises and Pitfalls of Genoeconomics. *Annual Review of Economics, 4*(1), 627-662. doi: 10.1146/annurev-economics-080511-110939

Benjamin, D. J., Cesarini, D., van der Loos, M. J. H. M., Dawes, C. T., Koellinger, P. D., Magnusson, P. K. E., et al. (2012b). The genetic architecture of economic and political preferences*. Proceedings of the National Academy of Sciences.* doi: 10.1073/pnas.1120666109

Benjamin, D. J., Chabris, C. F., Glaeser, E. L., Gudnason, V., Harris, T. B., Laibson, D. I., et al. (2007). *Biosocial Surveys:* The National Academies Press.

Bound, J., Jaeger, D. A., and Baker, R. M. (1995). Problems with Instrumental Variables Estimation When the Correlation between the Instruments and the Endogenous Explanatory Variable Is Weak. *Journal of the American Statistical Association, 90*(430), 443-450. doi: 10.2307/2291055

Burgess, S., Thompson, S. G., and Collaboration, C. C. G. (2011). Avoiding bias from weak instruments in Mendelian randomization studies. *International Journal of Epidemiology, 40*(3), 755-764. doi: 10.1093/Ije/Dyr036

Cawley, J. (2000). Body Weight and Women's Labor Market Outcomes. *National Bureau of Economic Research Working Paper Series, No. 7841*.

Cawley, J. (2004). The impact of obesity on wages. *Journal of Human Resources, 39*(2), 451-474. doi: 10.2307/3559022

Cawley, J., Han, E., and Norton, E. C. (2011). The validity of genes related to neurotransmitters as instrumental variables. *Health Economics, 20*(8), 884-888. doi: 10.1002/hec.1744

Cawley, J., and Meyerhoefer, C. (2012). The medical care costs of obesity: An instrumental variables approach. *Journal of Health Economics, 31*(1), 219-230. doi: http://dx.doi.org/10.1016/j.jhealeco.2011.10.003

Cecil, J. E., Tavendale, R., Watt, P., Hetherington, M. M., and Palmer, C. N. A. (2008). An Obesity-Associated FTO Gene Variant and Increased Energy Intake in Children. *New England Journal of Medicine, 359*(24), 2558-2566. doi: 10.1056/NEJMoa0803839

Chabris, C. F., Hebert, B. M., Benjamin, D. J., Beauchamp, J., Cesarini, D., van der Loos, M., et al. (2012). Most Reported Genetic Associations With General Intelligence Are Probably False Positives. *Psychological Science, 23*(11), 1314-1323. doi: 10.1177/0956797611435528

Chabris, C. F., Lee, J. J., Benjamin, D. J., Beauchamp, J. P., Glaeser, E. L., Borst, G., et al. (2013). Why It Is Hard to Find Genes Associated With Social Science Traits: Theoretical and Empirical Considerations. *American Journal of Public Health*, *103*, 152-166. doi: 10.2105/Ajph.2013.301327

Comuzzie, A. G., and Allison, D. B. (1998). The Search for Human Obesity Genes. *Science, 280(*5368), 1374-1377. doi: 10.1126/science.280.5368.1374

Conley, D. (2009). The Promise and Challenges of Incorporating Genetic Data into Longitudinal Social Science Surveys and Research. *Biodemography and Social Biology, 55*(2), 238-251. doi: 10.1080/19485560903415807

Cutler, D. M., and Glaeser, E. (2005). What Explains Differences in Smoking, Drinking, and Other Health-Related Behaviors? *American Economic Review, 95*(2), 238-242. doi: 10.1257/000282805774670464

Didelez, V., and Sheehan, N. (2007). Mendelian randomization as an instrumental variable approach to causal inference. *Statistical Methods in Medical Research, 16*(4), 309-330. doi: 10.1177/0962280206077743

Ding, W., Lehrer, S. F., Rosenquist, J. N., and Audrain-McGovern, J. (2006). The Impact of Poor Health on Education: New Evidence Using Genetic Markers. *National Bureau of Economic Research Working Paper Series, No. 12304*. doi: 10.3386/w12304

Finkelstein, E. A., Trogdon, J. G., Cohen, J. W., and Dietz, W. (2009). Annual Medical Spending Attributable To Obesity: Payer-And Service-Specific Estimates. *Health Affairs, 28*(5), w822-w831. doi: 10.1377/hlthaff.28.5.w822

Frayling, T. M., Timpson, N. J., Weedon, M. N., Zeggini, E., Freathy, R. M., Lindgren, C. M., et al. (2007). A Common Variant in the FTO Gene Is Associated with Body Mass Index and Predisposes to Childhood and Adult Obesity. *Science, 316*(5826), 889-894. doi: 10.1126/science.1141634

Glymour, M. M., Tchetgen, E. J. T., and Robins, J. M. (2012). Credible Mendelian Randomization Studies: Approaches for Evaluating the Instrumental Variable Assumptions. *American Journal of Epidemiology, 175*(4), 332-339. doi: 10.1093/Aje/Kwr323

Gregory, C. A., and Ruhm, C. J. (2009). Where Does the Wage Penalty Bite? *National Bureau of Economic Research Working Paper Series, No. 14984*

Harris, T. B., Launer, L. J., Eiriksdottir, G., Kjartansson, O., Jonsson, P. V., Sigurdsson, G., et al. (2007). Age, Gene/Environment Susceptibility-Reykjavik Study: Multidisciplinary applied phenomics. *American Journal of Epidemiology, 165*(9), 1076-1087. doi: 10.1093/Aje/Kwk115

International Labor Office. (2008), International Standard Classification of Occupations:ISCO–08, International Labor Office, Geneva.

Jacobsson, J. A., Riserus, U., Axelsson, T., Lannfelt, L., Schioth, H. B., and Fredriksson, R. (2009). The common FTO variant rs9939609 is not associated with BMI in a longitudinal study on a cohort of Swedish men born 1920-1924. *Bmc Medical Genetics, 10*. doi: 10.1186/1471-2350-10-131

Johansson, E., Böckerman, P., Kiiskinen, U., and Heliövaara, M. (2009). Obesity and labour market success in Finland: The difference between having a high BMI and being fat. *Economics & Human Biology, 7*(1), 36-45. doi: http://dx.doi.org/10.1016/j.ehb.2009.01.008

Kivimaki, M., Jokela, M., Hamer, M., Geddes, J., Ebmeier, K., Kumari, M., et al. (2011). Examining Overweight and Obesity as Risk Factors for Common Mental Disorders Using Fat Mass and Obesity-Associated (FTO) Genotype-Instrumented Analysis. *American Journal of Epidemiology, 173*(4), 421-429. doi: 10.1093/Aje/Kwq444

Lawlor, D. A., Harbord, R. M., Sterne, J. A. C., Timpson, N., and Smith, G. D. (2008). Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. *Statistics in Medicine, 27*(8), 1133-1163. doi: 10.1002/Sim.3034

Li, S. X., Zhao, J. H., Luan, J. A., Luben, R. N., Rodwell, S. A., Khaw, K. T., et al. (2010). Cumulative effects and predictive value of common obesity-susceptibility variants identified by genome-wide association studies. *American Journal of Clinical Nutrition, 91*(1), 184-190. doi: 10.3945/ajcn.2009.28403

Lindeboom, M., Lundborg, P., and van der Klaauw, B. (2010). Assessing the impact of obesity on labor market outcomes. *Economics & Human Biology, 8*(3), 309-319. doi: http://dx.doi.org/10.1016/j.ehb.2010.08.004

Nordestgaard, B. G., Palmer, T. M., Benn, M., Zacho, J., Tybaerg-Hansen, A., Smith, G. D., and Timpson, N. J. (2012). The Effect of Elevated Body Mass Index on Ischemic Heart Disease Risk: Causal Estimates from a Mendelian Randomisation Approach. *Plos Medicine, 9*(5). doi: 10.1371/journal.pmed.1001212

Pagan, J. A., and Davila, A. (1997). Obesity, occupational attainment, and earnings. *Social Science Quarterly, 78*(3), 756-770.

Peterson, R. E., Maes, H. H., Lin, P., Kramer, J. R., Hesselbrock, V. M., Bauer, L. O., et al. (2014). On the association of common and rare genetic variation influencing body mass index: a combined SNP and CNV analysis. *Bmc Genomics*. doi: 10.1186/1471-2164-15-368

Speliotes, E. K., Willer, C. J., Berndt, S. I., Monda, K. L., Thorleifsson, G., Jackson, A. U., et al. (2010). Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nature Genetics, 42*(11), 937-U953. doi: 10.1038/Ng.686

Staiger, D., and Stock, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica, 65*(3), 557-586. doi: 10.2307/2171753

Stock, J. H., and Yogo, M. (2002). Testing for Weak Instruments in Linear IV Regression. *National Bureau of Economic Research Technical Working Paper Series, No. 284*. doi: 10.3386/t0284

Taylor, A. E., Davies, N. M., Ware, J. J., VanderWeele, T., Smith, G. D., and Munafo, M. R. (2014). Mendelian randomization in health research: Using appropriate genetic variants and avoiding biased estimates. *Economics & Human Biology, 13*, 99-106. doi: 10.1016/j.ehb.2013.12.002

van der Loos, M. J. H. M., Rietveld, C. A., Eklund, N., Koellinger, P. D., Rivadeneira, F., Abecasis, G. R., et al. (2013). *The Molecular Genetic Architecture of Self-Employment. PLoS ONE, 8*(4), e60542. doi: 10.1371/journal.pone.0060542

VanderWeele, T. J., Tchetgen Tchetgen, E. J., Cornelis, M., and Kraft, P. (2014). Methodological Challenges in Mendelian Randomization. *Epidemiology, 25*(3), 427-435. doi: 10.1097/EDE.0000000000000081.

Vimaleswaran, K. S., Berry, D. J., Lu, C., Tikkanen, E., Pilz, S., Hiraki, L. T., et al. (2013). Causal Relationship between Obesity and Vitamin D Status: Bi-Directional

Mendelian Randomization Analysis of Multiple Cohorts. *PLoS Med, 10*(2), e1001383. doi: 10.1371/journal.pmed.1001383

Wehby, G. L., Ohsfeldt, R. L., and Murray, J. C. (2008). 'Mendelian randomization' equals instrumental variable analysis with genetic instruments. *Statistics in Medicine, 27*(15), 2745-2749. doi: 10.1002/sim.3255