Teenagers' (and young adults') self body image: is it influenced by social factors?

Silvia Ceria, Elena Perazzi and Panos Protopapas

January 14, 2013

1 Introduction

The purpose of this paper is to examine the influence of social factors on an individual's own weight perception. While an individual's weight is one of his or hers objective characteristics, and easily measurable, it is common experience that many people misperceive whether their weight is "normal", "below normal", or "excessive".

Misperceptions can be relevant to economics in that they can lead individuals to suboptimal behaviors and choices. Misperceptions about one's own weight are also relevant to the medical field in that they can lead to such self-destructive behaviors as anorexia or bulimia, that are widespread especially among the adolescent population. Not surprisingly, we found many papers in the medical literature addressing the causes of weight misperception.

Most of the papers we found examine the influence of socio-demographic factors (for example gender)[1,2], cultural and social norms [3], psychopathology and biological factors[1,4]. In our research, we focus on the influence of an individual's social interactions on this misperception.

The paper is organized as follows. In Section 2 we will describe our data set and in particular the variables we use for our study. In Section 3 we will describe our estimation methods, and for each method we will illustrate our main results. Finally, Section 4 will discuss our conclusions.

2 Data

2.1 Add Health

The main source for our data is the National Longitudinal Study of Adolescent Health (Add Health), a nationally representative study of adolescent physical and mental health in the United States. The study consists of four rounds of interviews to the same group of adolescents. In the first round of interviews (Wave I), the Add health cohort was chosen among students in grades 7-12, during the 1994-95 school year. The same subjects were followed into young adulthood with three more interviews – the most recent in 2008 – when they were aged 24-32. In our study we use data from Wave I, Wave II and Wave III.

One of the unique aspects of Add Health is its focus on social networks as a means of understanding adolescent health. In Wave I, several questions allowed to identify aspects of the subjects' social network. We know for example the number of people the individual nominates as friends inside and outside his/her own school (out-degree), the number of people that nominate the individual as a friend inside and outside school (in-degree), whether the individual has a best male and/or female friend, whether the alleged best male/female friend reciprocates the nomination, the size of the individual's receiver and sender network[1]. Unfortunately, network-related data was not collected in the subsequent rounds of interviews (after Wave I).

[1]The individual's receiver network is defined as the set of the people that nominate him/her as a friend (in-friends), plus the set of the people that nominate as a friend one of his/her in-friends (the in-friends of the in-friends), plus the set of the in-friends of the in-friends of the in-friends, and so on. Similarly, the individual's sender network is the set of his/her out-friends (people he/she nominates as friends), plus the out-friends of the out-friends and so on

2.2 The relevant variables

Our dependent variable is the respondent's self weight perception. The respondents were asked whether they viewed themselves as being "extremely underweight", "underweight", "normal", "overweight" or "extremely overweight". Our variable takes values 1 to 5, where 1 is given to people that consider themselves "extremely underweight" and 5 to people that consider themselves "extremely overweight".

Among the dependent variables, the most natural one to use is a measure of the objective weight appropriateness of the respondents. We define a discrete variable taking values 1 to 4 , using the medical definition of "underweight" (1), "normal' (2), "overweight"(3) and "obese" (4). A widely used medical definition is based on of the BMI index. The latter is defined as the ratio of weight in kilograms over the square of the height in meters. For example, the BMI index of a person whose weight is 70 kilos and whose height is 1.70 meters is $70/1.70^2 \simeq 24.22$.

One problem we encountered is that the official medical definitions are somewhat inconsistent for people below and above age 20 (For example, taking the official definitions literally, a person could be considered "normal" the day before turning 20 and "overweight" the day after turning 20, without any change in his/her weight). To overcome this problem, we built our measure of "weight appropriateness" in 3 different ways: the first takes the medical definitions literally, and the other two are modified to correct the inconsistency between the definition for people below and under 20. More details about how we constructed these variables will be found in the Appendix. By comparing our results using the 3 different definitions of "objective weight appropriateness", we also get a sense of the robustness of our estimation.

If the BMI index captured weight approrpriateness perfectly, and if perception were perfectly in line with reality, perception of weight appropriateness (our dependent variable) and BMI index would be perfectly correlated and no other explanatory variable would be needed. This is clearly not the case: even a first look at the data reveals that "perception" and "reality" (at least, reality as captured by the BMI-based medical definitions) are not perfectly in line (see Figure 1)

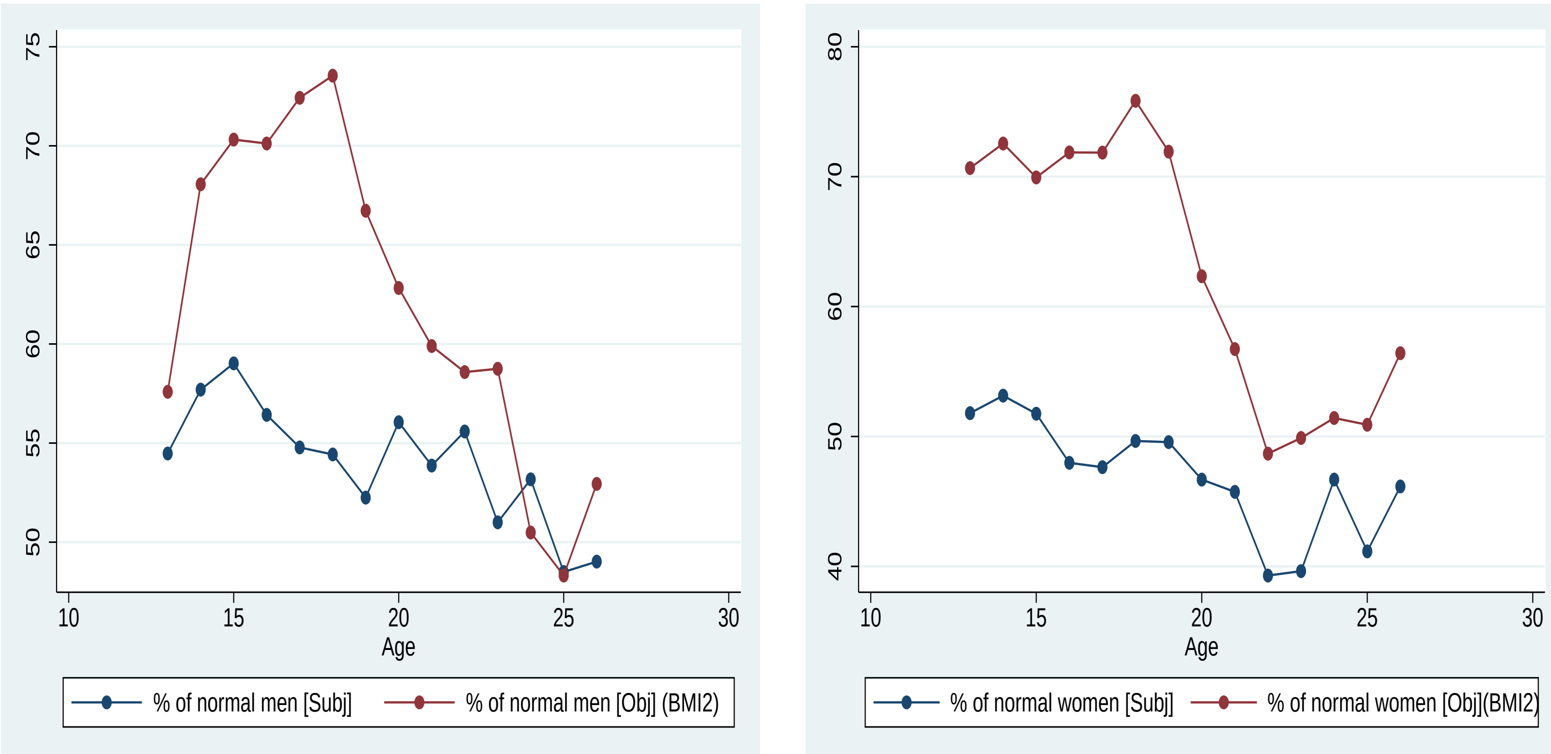

Figure 1: The picture on the left (right) shows the percentage of males (females) objectively"normal" (according to our measure "BMI2") versus the percentage of males (females) that consider themselves to be "normal"

To account for the fact that the BMI index is an imperfect indicator of weight appropriateness, we add "physical activity" (a discrete variable taking values 1 to 9) as an explanatory variable[2]. The idea is that if a person does a lot of physical activity, his/her BMI could be high due only to his/her muscular mass.

Next, we add several independent variables with the purpose of explaining any divergence between perceived and objective weight appropriateness. We add some socio-demographic variables: age, age square, gender and health condition[3], as well as some variables related

[2]This variable was constructed combining the answer to 3 different questions. The questions were: "During the past week, how many times did you go roller-blading, skate-boarding or bycicling?", ""During the past week, how many times did you play an active sport, such as baseball, soccer, swimming etc?", "During the past week, how many times did you do exercise, such as jogging or gymnastics?" Each question could be answered with a variable running from 0 to 3, increasing in the frequency of the described activity

[3]Health condition is a self-reported view of personal health taking values from 1 to 5, where a value of 1

to the social environment of the respondents: best male/female friend quality, and two network-related variables :"geodesic mean", and "out-degree".

The variables "Best male/female friend quality" try to capture if the type of interaction the respondent has with at least one male/female friend is a truly personal one, or if it is just a sporadic relationship, that mostly stays within the confines of the school routine[4].

"Out-degree" is the number of people the respondent nominates as friends.

"Geodesic mean" is defined as the average inward path length between an individual and the set of people he/she nominates as friends. For example, if Peter nominates Paul as his only friend, and Paul does not nominate Peter as a friend, but nominates John that in turn nominates Peter, then the path length from Paul to Peter is 2, and, being Paul the only person Peter nominated as friend, Peter's geodesic mean is 2.

3 Estimation methods and results

3.1 OLS

First we performed a "pooled OLS" estimation (we pool all the data from the 3 waves together).The model in this case is simply defined by the equation

$$y_i = \mathbf{x_i}\beta + \epsilon_\mathbf{i} \tag{1}$$

where x_i is a vector $(1 \times k)$ of all the regressors an β is the $k \times 1$ vector of coefficients. We don't have a time index in eq(1) because an individual i interviewed at different times is counted as a different individual. Since we use 3 waves, the index i runs from 1 to $3N$ where

corresponds to "Excellent" health while a value of 5 to "Poor" health.

[4]To construct this variable we first take the average of the answered questions: In the last week, "Did you go to X's house?", "Did you meet X after school to hang out?", "Did you spend time with X in the weekend?", "Did you talk to X about a personal problem?", "Did you talk to X on the phone?", where X is the male/female friend. Following this, the highest average of male and female friends is chosen as the best male and female friend quality respectively.

N is the number of respondents.

As can be seen from Table 1, some of the socio-demographic variables are highly significant, especially gender and health condition (and, obviously, objective weight, in all the 3 versions "BMI1", "BMI2", and "BMI3"). Age and age square are significant only when we omit some socially-related variables (which seems reasonable, since social activity becomes more intense and more important for an individual the more he/she progresses into adolescence). Physical activity is also only significant when socially-related variables are excluded. Also, in line with previous results in the literature, females tend to feel more overweight than men (gender is coded as a dummy variable that takes value 1 if the individual is a female. Therefore a positive coefficient of the "gender" dummy variable points to increased weight perception when the respondent is a female). As for the other variables, while best male/female friend quality and out-degree are insignificant, geodesic mean seems to be highly significant.

3.2 Fixed Effects

As is well known, OLS estimation is unbiased only if the regressors are uncorrelated with the error term of the regression. This is not be the case if the intercept (constant term in the regression) differs across individuals, and is correlated with some of the regressors. In this case, the model can be described by the equation

$$y_{it} = \alpha_i + \mathbf{x_{it}}\beta + \mathbf{z_i}\delta + \epsilon_{\mathbf{it}} \tag{2}$$

Now i runs from 1 to N, t runs from 1 to 3. x_{it} is the vector of the time-dependent regressors (in our case, age, health condition, physical activity, objective weight appropriateness, best male/female friend quality), z_i is the vector of the time-independent regressors (in our case, gender, outdegree, geomean)

The standard way to overcome the problem of the correlation between some of the regressors and α_i is to perform a fixed-effects (FE) estimation, in which, for every individual, we transform each (dependent or independent) variable by taking the difference between the

	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) OLS	(8) OLS	(9) OLS	(10) OLS
Age	.0462***	.08***	.0635	.1628**	.1013	.055	.0657	.1448	.0965	.0657
Age^2	-.0013***	-.0021***	-.0017	-.0045**	-.0026	-.0012	-.0015	-.0037	-.0022	-.0015
Gender	.3386***	.3343***	.3345***	.3248***	.3242***	.3238***	.3739***	.3119***	.3114***	.3739***
Bmi 1	.4996***	.5109***	.5099***	.5065***	.5244***			.5260***		
Bmi 2						.5240***			.5256***	
Bmi 3							.5275***			.5321***
Physical Activity	.0044*	.0031	.0019	.0022	-.0021	-.0021	-.0026	-.0022	-.0023	-.0026
Health Condition	.0817***	.0851***	.087***	.082***	.083***	.0823***	.0763***	.0839***	.0843***	.0763***
Best Male Friend Quality			.0232		.0212	.0217	.0183			
Best Female Friend Quality				.023				.0283	.0277	.0183
Out-degree		.0020			-.0013	-.0014	-.0011	-.0006	-.0007	-.0011
Geodesic Mean		-.0104**			-.02***	-.02***	-.0174***	-.0194***	-.0195***	-.0174***
R^2	.3355	.3567	.3034	.2992	.3240	0.3228	.3685	.3242	.3231	.3685

Table 1: *** p=0.99, ** p=0.95, * p=0.90

original variable and its average over time. The model becomes

$$(y_{it} - \overline{y}_i) = (\mathbf{x_{it}} - \overline{\mathbf{x}}_\mathbf{i})\beta + \tilde{\epsilon}_\mathbf{it} \tag{3}$$

As we see from eq(3), the drawback of this procedure is that it only allows the estimation of the effect of time-dependent variables. For example, it does not allow to estimate the effect of gender. In our study, this is particularly problematic because we have data for the network-related variables (which are the focus of our study) only in Wave I. This estimation method, therefore, is unbiased but is only useful to estimate the effect of the variables for which we have time-varying information.

Looking at the results in Table 2, the most important change we notice is that health condition, while highly significant under OLS, now becomes insignificant.

	(1) FE	**(2)** FE	**(3)** FE	**(4)** FE	**(5)** FE	**(6)** FE
Age	.08406	.07972	.0858	.1261	.1200	.1271
Age^2	-.0024	-.0022	-.0024	-.0037	-.0035	-.0037
Gender	0	0	0	0	0	0
Bmi 1	.05872***			.0759***		
Bmi 2		.05599***			.07602***	
Bmi 3			.07496***			.0950***
Physical Activity	-.0019	-.0018	-.0017	-.0045	-.0045	-.0045
Health Condition	.01288	.0129	.0129	.0121	.01218	.01170
Best Male Friend Quality	-.0189	-.0194	-.0195			
Best Female Friend Quality				.0200	.0198	.0212
Out-degree						
Geodesic Mean						

Table 2: *** p=0.99, ** p=0.95, * p=0.90

3.3 Hausman-Taylor

Another way to overcome the biasedness of OLS is to perform a Hausman-Taylor estimation [5]. The Hausman-Taylor method consists in identifying the regressors that are correlated with the individual intercepts (i.e. the endogenous regressors) and instrument them with the exogenous regressors (the regressors that are uncorrelated with the individual effects). The model can be written as

$$y_{it} = \alpha_i + \mathbf{x_{it}^{(end)}}\beta + \mathbf{x_{it}^{(exo)}}\beta' + \mathbf{z_i^{(end)}}\delta + \mathbf{z_i^{(exo)}}\delta' + \epsilon_{\mathbf{it}} \tag{4}$$

where we split the vector of time-dependent and time-independent variables into two subvectors each, containing the endogenous and the exogenous ones respectively. We assume that health condition is the only endogenous variable (since this is a time-dependent variable, in our case the vector $z_i^{(end)}$ is empty). The idea is that α_i captures some "deep" body image, an immutable component not affacted by life circumstances, and therefore is likely to be correlated only with other "innate" characteristics. Gender is an innate characteristic, but since its coefficient already captures the extent to which body image is due to gender, there is no need to assume further correlation between α_i and gender. Health condition can depend on life circumstances, but can be also partly due to innate conditions (and it is reasonable that an innate health condition can have an impact on body image)[5]. The other variables we use (physical activity, social variables) are more related to life events, therefore we assume that they are not correlated with α_i.

Looking at the results of the Hausman-Taylor estimation (Table 3) we notice two main things: first, the geodesic mean remains highly significant, giving further robustness to the hypothesis that there is some network effect that affects body image. Second, health condition, that had disappeared into insignificance with the FE estimation, becomes highly significant again. This suggest that maybe the assumption that the variables other than health condition are

[5]The endogenousness of health condition becomes obvious when we compare the results of OLS and FE: using FE, health condition becomes insignificant, suggesting that the coefficient we had estimated with OLS was mostly due to the constant term.

uncorrelated with the constant term requires more caution. It's possible that also other variables (that we use as instruments for health condition) are correlated with the constant term, and that the coefficient of health condition estimated with HT is also partly due to the constant term.

3.4 Probit

To check the integrity of our previous regressions we also performed a binomial Probit model on our data. To perform this regression we first introduced a new variable "misperception" in our data taken the value of 0 if a person's view of his weight was correct according to his BMI, or 1 otherwise[6]. More specifically what this model does is the following. Given the set of individual characteristics x_i, it measures the probability that the individual will have, or not, a misperception $[y]$ about his weight $P(y = j|x_i)$, $j = \langle 0, 1\rangle$. What we are then interested in observing is the partial effects of the model, depicted in Table 3 as the dy/dx coefficients, which show the change in these probabilities if we change the characteristics x_i, i.e. $\gamma(x_0, x_1) = P(y = 1|x_1) - P(y = 1|x_0)$.

	(1) BP dy/dx	**(2)** BP dy/dx	**(3)** BP dy/dx
Age	.0198	.0175	.0001
Age^2	-.0031	-.0040	-.0009
Gender	-.0675***	-.0482***	-.1023***
Bmi 1	.1781***		
Bmi 2		.1698***	
Bmi 3			.1894***
Physical Activity	-.0059**	-.0047*	-.0030
Health Condition	.0129**	.0184***	.0018
Out-degree	-.0013	-.0005	-.0006
Geodesic Mean	.0077**	.0081***	.0048

Table 3: *** p=0.99, ** p=0.95, * p=0.90

[6] For example a person believing to be overweight and also having an overweight BMI was coded as 0

	(1) HT	(2) HT	(3) HT	(4) HT	(5) HT	(6) HT	(7) HT	(8) HT	(9) HT
Time-Variant Exogenous									
Age	-.0002	-.0089	.0004	.1176	.1065	.1316	.1209	.1114	.1401
Age^2	-.0000	.0004	.0001	-.0033	-.0029	-.0036	-.0032	-.0029	-.0038
Bmi 1	.3772***			.3814***			.3916***		
Bmi 2		.3746***			.3807***			.3913***	
Bmi 3			.4051***			.3960***			.4141***
Physical Activity	-.0035	-.0039	-.0039	-.0042	-.0041	-.0039	-.0060	-.0059	-.0060
Best Male Friend Quality				-.0033	-.0042	-.0038			
Best Female Friend Quality							-.0025	-.0030	-.0026
Time-Variant Endogenous									
Health Condition	.0621***	.0631***	.05671***	.0537***	.0537***	.0538***	.0526***	.0524***	.0484***
Time-Invariant Exogenous									
Gender	.3323***	.3241***	.3592***	.3249***	.3246***	.3621***	.3189***	.3189***	.3611***
Out-degree	-.0004	-.0005	.0002	-.0035	-.0036	-.0031	-.0031	-.0032	-.0025
Geodesic Mean	-.0155***	-.0165***	-.0143**	-.0233***	-.0233***	-.0213***	-.0225***	-.0225***	-.0198 ***

Table 4: *** p=0.99, ** p=0.95, * p=0.90

As it can be seen, the results we get are similar with the previous part of our analysis. More specifically, we observe that women, coded as 1, are more likely to have a misperception about their weight since, ceteris paribus, being a female decreases the probability of correctly perceiving your weight, the latter being the base outcome. Moreover, the more overweight someone is the greater the chances of having a wrong perception of his weight. Finally, under the first two BMI specifications we find that by decreasing one's geodesic mean, by improving his health[7] or increasing his physical activity in effect decreases the probability of this misperception.

4 Conclusions

The purpose of our work was to study the determinants of self-weight perception among adolescents and young adults, using data from the Add Health project. Specifically, we wanted to test the significance of social and network-related aspects. For our estimation we have used several econometric methods: OLS, Fixed Effects (to account for the correlations between regressors and individual intercept), Hausman-Taylor in order to try to estimate both time-independent and time-dependent variables in an unbiased way. Furthermore, after estimating a binomial Probit model we have successfully cross-checked our results as discussed below.

Our results confirm the importance of gender as a determinant of self-weight perception, already found before in the literature. Women perceive themselves as being "fatter" than men when belonging to the same medical category.

The main new result is the importance of "Geodesic mean", that is the average inward path length between a person and the people he nominates as friends. Geodesic mean (or, more precisely, the inverse of geodesic mean) can be interpreted as an indicator of popularity in an individual's network: if all the people that I nominate as friends reciprocate the nomination, my geodesic mean takes the lowest possible value, i.e. 1. So we could say that the lower my

[7] Note that this corresponds to decreasing the relevant explanatory variable for health condition.

geodesic mean, the higher is my popularity in my circle. A little bit surprisingly, we find that geodesic mean has a very significant and *negative* effect on self-weight perception (the more I am popular, the heavier I think I am). Assuming that "fatter" is by and large a negative attribute, this points to a negative effect of social interactions: as I become more popular and more integrated with my network, I become more critical about my body image. This could be possibly attributed to a pressure of the social group to be thinner and fitter.

One reason to be cautious about this result could be that individual A could have a higher geodesic mean than individual B only because A's school is bigger than B's school. (In a bigger school, typical path lengths between students are longer). We don't control for this because we do not have data about school size[8]. Nonetheless, we would like to mention that unless school size is a determinant of body image (which we think is not likely to be the case), the coefficient of the geodesic mean is still unbiased.

One thing that we would have liked to do but could not due to the limitations of our data set is to look at the actual network and analyze whether one individual's body image is correlated with his/her friends' body image. We could not do this since our data is restricted and for example only the number of friends of the respondents are reported but not their identities.

References

[1] Chang, V.W., Christakis, N.A. (2003), "Self-Perception of Weight Appropriateness in the United States", *American Journal of Preventive Medicine*,**24**, 332-339

[2] Grey, S. (1994), "Social aspects of body image: perception of normalcy of weight and

[8]We have thought about using "network density", the density of the network whose vertices are an individual's friends, as a proxy for school size. Since in larger schools not everybody is friend with everybody else, networks are typically less dense in larger schools. However, an individual's network density is very imperfectly correlated with his/her school's size. For example, the larger is an individual's set of friends, the less likely it is that his/her friends will all be friends of one another, i.e. more popular people might have less dense networks. As another example, individuals with friends from different school activities, e.g. sports and French classes, will probably have a smaller density value without it being indicative of their school size. In any case, we checked that this variable is not significantly correlated with body image.

affect of college undergraduates", *Perceptual and Motor Skills*,**45**, 1035-1040

[3] Slade, P.D. (1994), "What is body image?", *Behaviour Research and Therapy*, **32**(5), 497-502

[4] Blokstra, A., Burns, C.M., Seidell, J.C. (1999), "Perception of Weight Status and Dieting Behaviour in Dutch Men and Women", *International Journal of Obesity and Related Metabolical Disorders* , **23**(**1**) , 7-17

[5]Hausman, J.A. and Taylor, W.E. (1981), "Panel Data and Unobservable Individual Effects", *Econometrica*, **50**, 1029-1054

Appendix: Our definitions of objective weight appropriateness

It is costumary in the medical community to define “underweight”, “normal”, “overweight” of “obese” people in terms of their BMI index. For children and adolescents under 20 years old, standard BMI charts are provided by CDC, an American governmental agency.

According to the medical community,’ a child or adolescent is “underweight” if his/her BMI is below the 10^{th} percentile (for his/her age and gender), “normal” if his/her BMI is between the 10^{th} and the 85^{th} percentile, “overweight”if his/her BMI is between the 85^{th} and the 95^{th} percentile, “obese” if his/her BMI is above the 95^{th} percentile. However, for young adults of more than 20 years old, the definition changes: they are underweight when their BMI index is below 18.5, normal when their BMI index is between 18.5 and 25, overweight when their BMI index is between 25 and 30, obese when their BMI index is above 30, independently of their age and gender. The classification for people below 20 is somewhat at odds with the classification for people above 20. As can be seen from the charts in Figures 2 and 3, a male or a female could be considered “normal” just before turning 20 and “overweight” the day they turn 20 with no change in their weight, or could be considered “overweight” just before turning 20 and “obese” just after. Also, a male could be considered “underweight” just before turning 20 and “normal” just after.

Our first measure, “BMI1”, takes value 1 to 4 (1 for “underweight”, 4 for “obese”) following the medical definition literally, so this measure is discontinuous in weight across the 20-year-old threshold.

To construct the second measure, “BMI2”, we modify the medical definitions after 20, so as to eliminate the discontinuity. If, for example, just before 20 “normality” corresponds the range between BMI b_1 and BMI b_2, people over 20 will be defined as “normal” if they fall in the same BMI range.

To construct the third measure, BMI3, we modify the medical definitions for people under

20, again so as to eliminate the discontinuity at the 20-year-old threshold. For example, just before 20, the percentiles defining the thresholds between two weight categories (e.g. "underweight" and "normal") are selected so that the resulting BMI ranges for each category match the medical definitions for people over 20.

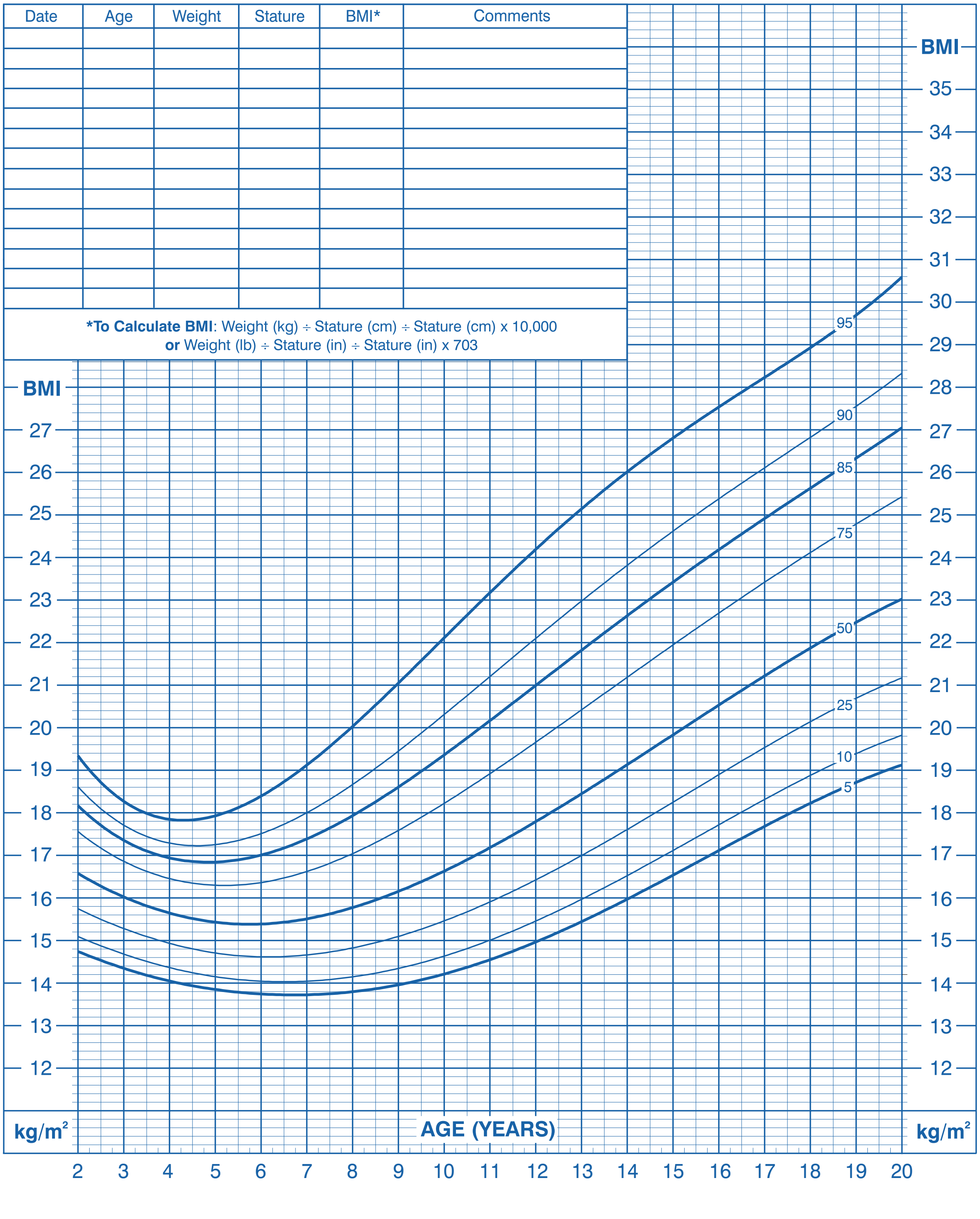

Figure 2: BMI chart for boys from 2 to 20 years old. Source: CDC

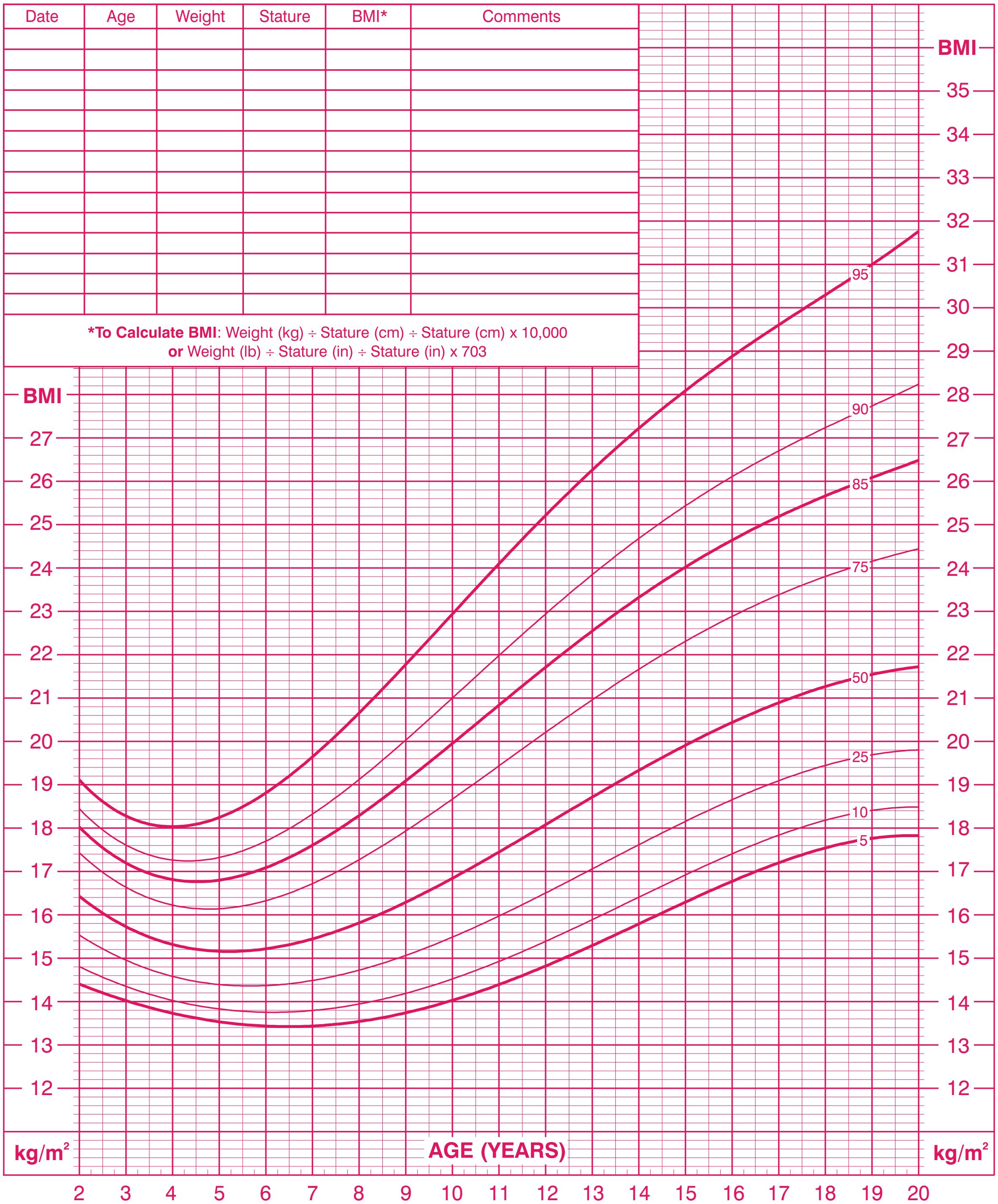

Figure 3: BMI chart for girls from 2 to 20 years old. Source: CDC