*UCD CENTRE FOR ECONOMIC RESEARCH*

*WORKING PAPER SERIES*

*2013*

# Self-Reported and Measured BMI in Ireland: Should We Adjust the Obesity Thresholds?

David Madden, University College Dublin

WP13/01

February 2013

**UCD SCHOOL OF ECONOMICS**
**UNIVERSITY COLLEGE DUBLIN**
**BELFIELD DUBLIN 4**

# Self-Reported and Measured BMI in Ireland: Should We Adjust the Obesity Thresholds?

**David Madden**

**(University College Dublin)**

**February 2013**

**Abstract:** Using the nationally representative Slan dataset of 2007 we analyse the relationship between self-reported and measured BMI. We find that self-reported BMI significantly underestimates obesity rates and suggest that the traditional threshold of 30 should be adjusted downwards. We outline a number of approaches to choose the optimal threshold and results suggest that the new obesity threshold for self-reported BMI could be as low as 26.

**Keywords:** body mass index, receiver operating characteristic, sensitivity, specificity.

**Corresponding Author:**

David Madden,

School of Economics,

University College Dublin,

Dublin 4,

Ireland.

**Phone:** +353-1-7168396

**Fax:** +353-1-2830068

**E-mail:** david.madden@ucd.ie

**Acknowledgements:** I am grateful to Donal O Neill and Olive Sweetman for helpful comments and to Karen Morgan for assistance with the data. The usual disclaimer applies.

# Self-Reported and Measured BMI in Ireland: Should We Adjust the Obesity Thresholds?

## 1. Introduction

Obesity has become one of the most important public health issues in Ireland. In 2005 a report was produced by the National Taskforce on Obesity and an extensive list of recommendations was provided. In 2009 a review of these recommendations was carried out and since the original recommendations were only partially implemented a Special Action Group was set up to work across Government Departments and agencies. The statistics motivating these concerns came from the nationally representative Slan data (which was collected in 2006) and showed rates of obesity of 25% and rates of overweight of 39% (both based on measured as opposed to self-reported data), while the National Adult Nutrition Survey (which looked at data from the 2008-2010 period) indicated that 24 per cent of adults were obese and 37 per cent were overweight.

Rates of obesity and overweight are typically measured via body mass index (BMI). BMI is obtained by dividing weight (in kilos) by height (in metres) squared. The World Health Organisation suggests a threshold BMI of 25 for “overweight”, a threshold of 30 for “obesity” and a threshold of 40 for “severely obese”.

It is important to note that there is criticism of BMI as a measure of obesity with some authors suggesting that other measures such as total body fat, percent body fat and waist circumference are superior measures of fatness (see Cawley and Burkhauser, 2006). However, while these measures may provide a more accurate indicator of obesity, they are expensive to produce and in terms of large-scale nationally representative samples, the likelihood is that BMI will remain the most commonly used indicator of obesity for the foreseeable future.

However, there is a further issue with BMI as it is frequently reported in large scale nationally representative samples. Once again, for reasons of economy, it is typically the case that BMI is calculated from self-reported height and weight. This clearly gives rise to scope for mis-reporting (compared to true measured height and weight). If mis-reporting was random (people being as likely to over and under report their height/weight) then reported mean BMI would still be unbiased, but reported variance would be higher than “true” variance. However if mis-reporting is *systematic*, then this represents a more serious problem, since it suggests that mean BMI as calculated from national samples may be *biased*, and further problems emerge if the degree of bias differs across categories such as age, gender and socio-economic background.

Evidence worldwide, and for Ireland (Connor Gorber et al, 2008, Shiely et al, 2010), suggests that mis-reporting in self-reported BMI is not random and that through a combination of over-statement of height and under-statement of weight, self-reported BMI will typically *underestimate* "true" (or measured) BMI. Moreover, Shiely et al demonstrate that this degree of mis-reporting appears to be increasing over time in Ireland. However the evidence for Ireland is relatively sparse as there are not many large scale datasets which include both self-reported and measured BMI.

An alternative perspective on this issue is provided by Dauphinot et al (2009). Using a Swiss sample with self-reported and clinically measured BMI they find as per the references above, evidence that self-reported BMI understates obesity levels. However using Receiver Operating Characteristic (ROC) curves, they calculate what the threshold level of self-reported BMI should be in order for it to provide the "optimal" signal of true underlying BMI. However, their revised thresholds have been criticized on the basis that they are relevant only for their specific dataset and for other datasets, different thresholds may be optimal (Shi et al, 2009, Bopp and Faeh, 2009).

This paper also examines the relationship between self-reported and measured BMI and discusses the role of ROC curves. However we employ a wider range of approaches to calculating the "optimal" threshold and show how calculated thresholds can vary quite substantially depending upon the approach adopted. In particular we show that some of the more popular approaches may lead to analysts unconsciously making value judgements regarding the relative costs of different types of mis-classification. We also examine whether the optimally calculated threshold differs according to characteristics such as age and gender.

In section 2 of the paper we explain the application of ROC curves to the relationship between self-reported and measured obesity and we also outline the different possible approaches to obtaining the optimal threshold. In section 3 we present our data and results while section 4 provides concluding comments.

## 2. Receiver Operating Characteristic (ROC) Curves

The ROC curve provides a useful procedure for analysing the extent to which a given signal can detect an underlying condition. In the application here, measured BMI is taken as the "true" or gold standard measure of obesity and a threshold of 30 for this measure partitions the population into the binary categories of obese and non-obese. We then assess the degree to which self-reported BMI (sometimes called the "marker") produces the "same" partition. If self-reported BMI assigns someone as obese who is also obese under the measured BMI definition then this is called a "true positive" (TP). If it signals someone as obese who is not obese under the measured definition it is a "false positive" (FP). If it signals someone as non-obese even though they are obese under the measured definition it is a "false negative" (FN). Finally "true negatives" (TN) are those who are classified as non-obese under both definitions.

The TP rate is sometimes called the *sensitivity* (Se) of the signal and is TP/(TP+FN), while the corresponding concept for the TN rate is known as *specificity* (Sp) and is TN/(FP+TN), which in turn is equal to one minus the FP rate. The ROC curve then graphs the TP rate (on the vertical axis) against the FP rate (one minus the specificity rate) for all possible values of the self-reported obesity threshold. Thus as the threshold goes from its lowest to its highest level the ROC curve traces out from (0,0) to (1,1) and the better the signal the further above and to the left (or north-west) of the $45^0$ line will be the curve. The less accurate the signal the nearer the curve will be to the $45^0$ line. If the curve lies below the $45^0$ line then it is effectively acting as a contra-indicator and paradoxically the further to the south-east the curve lies the better, since the ROC curve for the negative of the indicator is simply the mirror image of the ROC curve for the original indicator. Figure 1 shows an example of a typical ROC curve.

For a given marker, each point on the ROC curve will correspond to a particular threshold and the ROC curve shows the combination of sensitivity and (one minus) specificity which are associated with that threshold. Clearly a very low threshold will provide very high levels of sensitivity (lots of TP and very few TN), but at the cost of low specificity since a low threshold will likely also have high rates of FP. Likewise a very high threshold will produce high levels of specificity but at the risk of low levels of sensitivity.

If we have a number of different possible markers for the same underlying condition then the ROC curve can be used to make a comparison between these markers and their usefulness as a signal. Clearly if the ROC curve for one marker always lies above and to the left of that of another, then the former marker acts as a better signal for all values of the threshold and can be said to "dominate" (since it will have higher levels of both sensitivity and specificity). However there is no guarantee that dominance will be found when comparing any two markers. In that case a summary index may be used. Probably the most popular one is the area under the ROC curve (AUC). If the ROC curve lies on the $45^0$ line then this area equals 0.5 and this corresponds to the situation where the marker effectively gives no signal. If the ROC curve corresponds to the vertical line from (0,0) to (0,1) and then across to (1,1) the AUC is one and the marker gives a perfect signal. Intuitively the AUC corresponds to the probability that self-reported BMI for a randomly chosen obese person is higher than the self-reported BMI for a (randomly chosen) non obese person.

One criticism which has been made of the AUC as a summary index of a marker's ability to detect the underlying condition is that it will to some extent be determined by areas corresponding to either very high or very low thresholds, values which are very unlikely to be chosen by the analyst, but which yet might still influence the ranking of two markers by AUC. To overcome this, some analysts have suggested instead the use of the partial AUC whereby the area is measured for only a limited range of the threshold, a range which would not include clearly unreasonably high (or low) values of the threshold.

The ROC curve is clearly a very useful graphical device when making a comparison between two different markers for an underlying condition. As illustrated by Dauphinot et al (2008) it can also be of use in the case where we have only one marker which is continuous, but where we wish to choose the optimal threshold, so that the partitioning of the population into obese and non-obese by the marker (self-reported BMI) is in some sense "closest" to the partitioning by the true measure (clinically measured BMI). But how do we choose the optimal level of the threshold?[1] As discussed in Greiner et al (2000) there are a number of approaches we can take. One approach is to utilise the ROC curve and to choose that point which maximises the Youden J index i.e. the point which gives the maximal vertical distance from the ROC curve to the main diagonal, in other words the point which is most "north-west" on the ROC curve, as illustrated in figure 2. Intuitively the J index is Se+Sp-1 i.e. the sum of the sensitivity and specificity rates (minus one).

However, there are other possible and arguably equally plausible criteria for choosing the optimal threshold. For example, we could choose the threshold which maximises the percentage of cases which are correctly classified (or minimises those mis-classified). Greiner et al label this *efficiency*, and it is that value of the threshold, *t**, which maximises $P.Se(t)+(1-P).Sp(t)$, where *P* represents the prevalence of obesity (in proportional terms).

Another approach is to choose that threshold which maximises the odds ratio. Suppose the 2x2 table of self-reported and clinically measured obese for any given threshold point *t* is given by the following table, where *a, b, c* and *d* are the numbers in each cell and "1" and "0" refer to diagnosis and non-diagnosis of obesity respectively.

| | Clinically Measured Obesity | | |
|---|---|---|---|
| Self Reported Obesity | | 0 | 1 |
| | 0 | *a* | *b* |
| | 1 | *c* | *d* |

In this instance the odds ratio is $\frac{ad}{bc}$. For each value of *t* there will be a corresponding odds ratio. The optimal threshold is that value of *t, t**, which maximises the odds ratio, which is effectively the ratio of correct to incorrect classifications.

It is important to note that the efficiency and Youden J approach are both specific cases of a more generalised approach. The rate of false negatives for any given threshold *t*, will be *P.(1-Se(t))*, while that of false positives is *(1-P).(1-Sp(t))*. Note that in this case we are referring to the rate of FN relative to the total population (hence we multiply by *P*), as opposed to the rate relative to those who are truly obese. However the analyst may associate different costs with different types of mis-classification. For example, it seems reasonable in the case of obesity that analysts would assign a higher weight to FN rather than FP, since if someone is diagnosed FN they may not take precautions in terms of diet and lifestyle which they probably should. A diagnosis of FP on the other hand may lead them to consult their GP where their "true" BMI will presumably become known.

[1] Note that while the AUC could be used as a criterion when choosing between different "markers" for BMI, such as a choice between self-reported BMI versus percentage total body fat, it is not used when choosing an optimal threshold for a given marker, since the AUC will be determined by all points on the curve and each point corresponds to a different threshold.

Suppose then that the cost of a false negative is given by $C_{FN}$ and that of a false positive by $C_{FP}$. Then the total cost associated with any given threshold is $C_{FN}.P.(1-Se(t))+C_{FP}.(1-P).(1-Sp(t))$. A decision rule could then be adopted to choose that threshold, *t**, which minimises the above expression or equivalently which minimises $r.P.(1-Se(t))+(1-P).(1-Sp(t))$ , where $r=\frac{C_{FN}}{C_{FP}}$ is the relative cost of FN compared to FP.

As pointed out by Smits (2010), the choice of a threshold based upon the maximisation of Youden's J is equivalent to a choice based on a minimisation of cost where *r*, the ratio of the cost of FN to that of FP is set equal to $\frac{1-P}{P}$. Thus Youden's J is a specific case of a more general decision-based approach. Another way of looking at this is should an analyst choose that threshold which maximises the value of Youden's J, they are implicitly (and perhaps unknowingly) imposing a relative cost of FN to FP equal to $\frac{1-P}{P}$, a ratio which may or may not conform to the actual values or beliefs of the analyst.

It is also clear that the value of *t* which maximises efficiency is also that which minimises $r.P(1-Se(t))+(1-P)(1-Sp(t))$ where *r*=1. Thus both efficiency and Youden's J can be regarded as special cases of a more general decision-based approach.

The approaches we have described above essentially involve choosing that threshold which minimises a weighted average of the cost of FP and FN, where the weights can either be chosen explicitly by the analyst or may be implicitly chosen by the choice of an index such as the Youden J index. However, it is also possible that the analyst may take what we can call a lexicographic or constrained optimisation approach. Suppose, as would seem natural in the application here, that the analyst regards FN as more costly than FP. The analyst could then choose a benchmark level of FN above which he is not prepared to go. The threshold is then that level which minimises the FP rate subject to attaining the given level of FN. It is lexicographic as priority is first given to attaining a certain level of FN and then the threshold is chosen which optimises FP. It can also be regarded as a constrained optimisation approach in that FP is minimised subject to attaining a given level of FN.

Thus there are a number of criteria which could be applied to choose the optimal threshold. The degree to which the different criteria give different values of *t**, and also the degree to which these different values of *t** differs between different populations is ultimately an empirical matter which we now investigate. We examine how *t** varies according to the following different criteria: efficiency, Youden's J, maximum value of the odds ratio and the minimum cost basis where we choose a range of *r* (some values of *r*, of course, having already been included in efficiency and the J index), and a lexicographic approach where we choose three values of FN (1%, 5% and 10%). The latter is equivalent to choosing sensitivity levels of 99%, 95% and 90%. We also examine how *t** varies according to age and gender.

We now discuss our data and present our results.

## 3. Data and Results

Our data comes from the Survey of Lifestyle, Attitudes and Nutrition in Ireland, usually known as the Slán survey. The Slan surveys were carried out in 1998, 2002 and 2007. For this paper we use the 2007 data, since as well as providing information on self-reported BMI it also provides information on clinically measured BMI for a reasonable sized subset of the sample (data on measured BMI was also provided for 1998 and 2002 Slan but proportionally these sub-samples were only half as large as that for 2007). The Slan 2007 survey is a comprehensive, nationally representative survey carried out by face-to-face interview in the respondent's house with a sample size of 10364. The 2007 sample was provided by the Irish Social Science Data Archive (ISSDA) with the Geodirectory (a listing of all residential addresses in Ireland compiled by the postal service) used as the sampling frame and weights supplied with the data (in all subsequent analysis sampling weights are applied). Morgan et al (2008) provide greater detail.

Self-reported BMI was collected as all respondents were asked to self-report their weight without clothes and their height without shoes. In addition about 20 per cent of the sample (2174) also underwent a medical examination, which included height and weight measurement. Respondents provided the self-reported data before their examination, and weight and height were measured in light clothing without shoes. Weight was measured to the nearest 0.1 kg using electronic platform scales and height was measured to the nearest 0.1 cm using measuring rods.

Since the purpose of this paper is to examine the relationship between self-reported and measured BMI we are forced to restrict our sample to those who provided data on both. In the version of Slan provided to us by ISSDA we initially had 2171 observations where measured BMI was available. We then had to discard those observations where self-reported BMI was not available and this brought the sample size to 1976. When examining summary BMI statistics for this group, it became clear that there were a small number of cases which appeared to suffer from measurement error (e.g. recorded self-reported BMI of zero), and so it was decided to trim the data by removing all observations with BMI (either self-reported or measured) less than 15 or greater than 50. This brought the sample size to 1874.

Given the adjustments which had to be made to the data it is important to check that the remaining sample is reasonably representative. Table 1 gives summary statistics for our sample and for the complete Slan 2007 sample (the latter figures were obtained from the Slan 2007 report, see Morgan et al, 2008). The discrepancy between self-reported and measured BMI is clear. There is a gap of over 9 per cent between measured obesity and self-reported obesity i.e. "true" obesity is higher than self-reported by almost two-thirds and the t statistic for the paired t test is 12.9, with a p value of 0.000. In terms of actual BMI (as opposed to BMI categories) self-reported BMI is about 1.4 below measured BMI and a paired t test of the null hypothesis of equality of measured and self-reported BMI gives a t statistic of 26.3 and a p-value of 0.000.

Table 1 also shows that the data used in our analysis has a slightly younger age profile and correspondingly a slightly higher education profile. Nevertheless, on the basis of table 1 it seems reasonable to suggest that the sample analysed in this paper is close enough to the overall Slan sample to permit us to calculate revised optimal thresholds for self-reported BMI that should prove useful to policy-makers.

Figure 3 confirms the summary statistics, showing the kernel density for measured and self-reported BMI, while figure 4 shows the ROC curve. The density for measured BMI shows more weight in the right hand side of the distribution.

We now look at a cross-tabulation between self-reported and measured obesity. Table 2 shows this cross-tabulation on the basis of a threshold of 30 for both measures. This table shows that if we use a threshold of 30 for both measures then self-reported BMI will correctly classify about 87% of observations i.e. (1386+250)/1874. This corresponds to a sensitivity rate of about 55% and a specificity rate of about 98%.

Before calculating optimal thresholds under the different criteria outlined above, it might be worth checking the type of factors which might influence the difference between self-reported and measured BMI. This would be helpful in terms of identifying different sub-groups who might have a different optimal threshold. In table A1 we present results from a simple linear regression, where the dependent variable is measured BMI less self-reported BMI and we have regressed it against a number of demographic and lifestyle factors. The demographic factors we choose are age, gender, education and marital status, while the lifestyle factors are self-assessed general health, smoking and drinking.

The results in table A1 show that the difference between measured and self-reported BMI is influenced by age, gender, BMI category, marital status and drinking alcohol. In terms of identifying sub-groups for whom we might wish to calculate different optimal thresholds, it seems most useful to concentrate upon variables which are exogenous, in this case age and gender. Thus in the analysis which follows we estimate $t^*$ using the different methods outlined above and also by age and gender.

Table 3 shows the value of $t^*$ for different criteria and for the whole of our sample as well as specific sub-groups and it also provides rates of sensitivity and specificity. By reading down the column we can see how $t^*$ varies according to the different criteria. Taking the column for the total sample initially, we first of all see that the values of $t^*$ essentially fall into three bands. First of all, if we employ the efficiency criterion we obtain a $t^*$ of 29.1, quite close to the typically adopted threshold of 30. Thus 30 is only likely to be close to the optimal value of the threshold if the “efficiency” criterion is used i.e. equal costs are assigned to FN as to FP.

The values of $t^*$ for the other criteria can be assigned into two bands, both of which differ quite substantially from 30. Using the criteria of Youden’s J, maximising the odds ratio or minimising the MCF for “low” values of $r$ (i.e. 2-5) we obtain a range of $t^*$ from 27.1 to 27.5. It is worth noting that $t^*$ as chosen by the Youden J index is the same as $t^*$ for $r$= 3. This is to be expected since with $P$=0.24, $\frac{1-P}{P}$ = 3.17.Clearly the higher is the value of $r$, and hence the higher is the relative cost of FN, then the lower becomes $t^*$, since in the limit, a very low value of $t^*$ would ensure no FN, though at the expense of a very high rate of FP. This is essentially what is happening with respect to the third band of values of $t^*$, those chosen using $r$=10 and the constrained optimisation criterion whereby we choose “standard” sensitivity values of 99%, 95% and 90% (corresponding to FN rates of 1%, 5% and 10% respectively). This gives a range of $t^*$ of 22.4-27.1, considerably lower than the other ranges. However, this high rate of sensitivity comes at the expense of low rates of specificity, in the region of only 30%.

It is also clear that choosing “high” values of $r$, i.e. 10 or above, provides values of $t^*$ which are very similar to those when we choose “conventional” levels of significance of, say, 5%.

The pattern of three “bands” of $t^*$ persists when we look at $t^*$ by age and gender and as before the values of $t^*$ for the efficiency criterion are highest, while those using the constrained optimisation criteria are lowest. In general recommended $t^*$ for females is lower than for males. The pattern with respect to age is not so clearcut. For the constrained optimisation approach with a FN rate of 1% the recommended $t^*$ for young is over 2 units lower than for old, but for other criteria there is not so much difference.

We also provide summary information as to how $t^*$ varies by criterion and by demographic groups by calculating the coefficient of variation. Thus the variation within each criterion by demographic group can be examined by looking at the values of the CV in the right-hand column. This shows that Youden’s J and the odds ratio shows the least variation and the greatest variation is for the constrained optimisation with FN set at 1%.

We can also look at variation within each demographic group, by examining the CV in the third last row of table 3. The greatest variation is seen amongst young people, mainly driven by the very low $t^*$ for the constrained optimisation case where FN is set to 1%

So, are there any general rules of thumb which we can draw from table 3? First of all, in the case of self-reported and measured BMI, it appears likely that for any population, or for any approach to calculating $t^*$, with the exception of the efficiency criterion where the cost of FN and FP are equivalent, then the optimal threshold will differ from 30. Quite how far from 30 however depends upon what optimisation criterion is chosen. For relatively low values of $r$, the relative costs of FN to FP, then a threshold self-reported BMI of around 27-27.5 seems appropriate, indicating a downward adjustment of the current threshold for self-reported BMI of 2-2.5 units. Given the implicit weighting of FN and FP in Youdens J index, then with prevalence rates in the region of 24%, $t^*$ the downward adjustment as chosen by this criterion will be of the same magnitude. However if the analyst wishes to be guaranteed a sensitivity rate of 95% (or higher), then an adjustment of 4 or maybe more units would seem to be required.

Which of these adjustments would be warranted depends upon a number of factors. The desired sensitivity of the test (and also the ratio of costs of FN to FP) will depend upon the nature of treatment. In the case of obesity, a choice of a low threshold will ensure a low rate of FN but perhaps a relatively high rate of FP. However, since the treatment for obesity (in terms of changed lifestyle etc) is relatively non-intrusive and easily reversible, once the "true" diagnosis becomes known, then for self-reported BMI there does seem to be a case for a low threshold. This might not be the case if treatment was invasive and with potentially harmful side-effects.

The underlying seriousness of the condition in terms of increased morbidity and mortality will also be relevant. There is some recent evidence suggesting that the relationship between BMI and mortality may not be monotonic, with higher BMI over some ranges (in particular 25-30) appearing to have a protective effect in terms of mortality and BMI for grade 1 levels of obesity (i.e. BMI from 30 to 35) having no significant impact upon mortality (Flegel et al, 2013). In that case, the relative cost of FN would presumably become lower. However, regardless of how this issue eventually resolves, it seems desirable that BMI should be measured accurately and the evidence presented here suggests some adjustment is necessary.

## 4. Conclusions

This paper has addressed the issue of the use of self-reported BMI as a marker for clinically measured or "true" BMI. It is generally found that use of the threshold of 30 for self-reported BMI leads to quite substantial under-measurement of obesity. This paper has discussed different criteria which might be applied in order to arrive at an optimal threshold value. As an illustration, the optimal threshold has been calculated for a representative sample of Irish adults and the paper also investigates the extent to which this optimal threshold might differ according to age and gender. The results suggest that the optimal threshold value of self-reported BMI can vary according to the choice criterion and that a threshold of as low as 26 could be justified, depending upon the weighting the analyst applies to sensitivity compared to specificity. The paper also shows that the optimal threshold can vary by demographic group and that it may be advisable to have a lower threshold for women. Of course the benefit from a measurement perspective of having a number of different optimal thresholds may have to be balanced in terms of the simplicity of whatever public health message it is desired to deliver.

It should also be borne in mind that the optimal thresholds calculated in this paper may be specific to the sample analysed and that these thresholds may differ for different samples e.g. for different countries or time periods. However, the paper does illustrate that the degree of adjustment which may be required for self-reported BMI may be quite substantial and it seems advisable that care should be taken in all cases where public health decisions in the area of obesity rely on self-reporting.

**Table 1: Self-Reported and Measured BMI**

| | **Our sample (reweighted, n=1874)** | **Slan 2007 Main Report** |
|---|---|---|
| Self-rep BMI <18.5 | 1.66 | 2.12 |
| Self-rep BMI, 18.5-24.99 | 45.28 | 47.87 |
| Self-rep BMI 25-29.99 | 38.10 | 35.11 |
| Self-rep BMI >30 | 14.96 | 14.89 |
| Measured BMI <18.5 | 1.31 | |
| Measured BMI, 18.5-24.99 | 35.44 | |
| Measured BMI 25-29.99 | 38.82 | |
| Measured BMI >30 | 24.44 | |
| Gender (% female) | 51 | 50 |
| Age 18-29 | 20 | 25 |
| Age 30-44 | 28 | 31 |
| Age 44-65 | 36 | 29 |
| Age 65 and over | 16 | 15 |
| Primary or below (%) | 17 | 19 |
| Lower Secondary (%) | 18 | 17 |
| Leaving Certificate (%) | 24 | 27 |
| Cert/Diploma (%) | 21 | 19 |
| 3rd level degree (%) | 19 | 18 |

**Table 2: Cross Tabulation between Self-Reported and Measured Obesity**

| | | Measured Obese | | Total |
|---|---|---|---|---|
| | | 0 | 1 | |
| Self-Reported Obese | 0 | **1386** | **208** | 1594 |
| | 1 | **30** | **250** | 280 |
| Total | | 1416 | 458 | **1874** |

**Table 3: Optimal Values of Self-Reported BMI Thresholds (with percentages of sensitivity and specificity in brackets)**

| Criterion | Total | Male | Female | Young | Old | CV |
|---|---|---|---|---|---|---|
| **Efficiency** | 29.1 (68.3, 95.8) | 29.3 (66.5, 94.8) | 28.3 (76.9, 96.1) | 29.5 (71.3, 96.9) | 28.1 (77.9, 91.1) | 0.022 |
| **Youden J** | 27.5 (87.6, 88.1) | 27.5 (91.3, 84.5) | 27.1 (86.6, 90.0) | 26.9 (91.9, 85.7) | 27.5 (88.0, 85.9) | 0.010 |
| **OR** | 27.5 (87.6, 88.1) | 27.5 (91.3, 84.5) | 27.1 (86.6, 90.0) | 26.9 (91.9, 85.7) | 27.5 (88.0, 85.9) | 0.010 |
| **MCF, r=10** | 26.0 (95, 72.2) | 26.9 (96.4, 75.4) | 26.0 (92.1, 82.0) | 26.9 (91.9, 85.7) | 26.0 (95.5, 65.6) | 0.019 |
| **MCF, r=5** | 27.1 (89.9, 85.4) | 27.5 (91.3, 84.5) | 26.0 (92.1, 82.0) | 26.9 (91.9, 85.7) | 27.2 (90.1, 82.7) | 0.021 |
| **MCF, r=2** | 27.5 (87.6, 88.1) | 27.5 (91.3, 84.5) | 27.9 (81.1, 94.4) | 29.2 (74.5, 95.9) | 27.5 (88.0, 85.9) | 0.026 |
| **FN rate=1%** | 22.4 (99, 30.5) | 23.6 (99, 35.6) | 21.2 (99, 25.1) | 20.9 (99, 20.1) | 23.0 (99, 26.5) | 0.052 |
| **FN rate=5%** | 26.0 (95, 72.2) | 26.9 (95, 75.4) | 24.8 (95, 67.8) | 25.6 (95, 75.0) | 26.0 (95, 65.6) | 0.029 |
| **FN rate=10%** | 27.1 (90, 84.4) | 27.5 (90, 84.5) | 26.3 (90, 83.2) | 27.1 (90, 86.7) | 27.2 (90, 82.7) | 0.016 |
| **CV** | 0.069 | 0.055 | 0.081 | 0.093 | 0.058 | |
| **P** | 0.24 | 0.24 | 0.25 | 0.17 | 0.32 | |
| **(1-P)/P** | 3.17 | 3.13 | 3.03 | 4.99 | 2.15 | |

**Figure 1: ROC Curve**

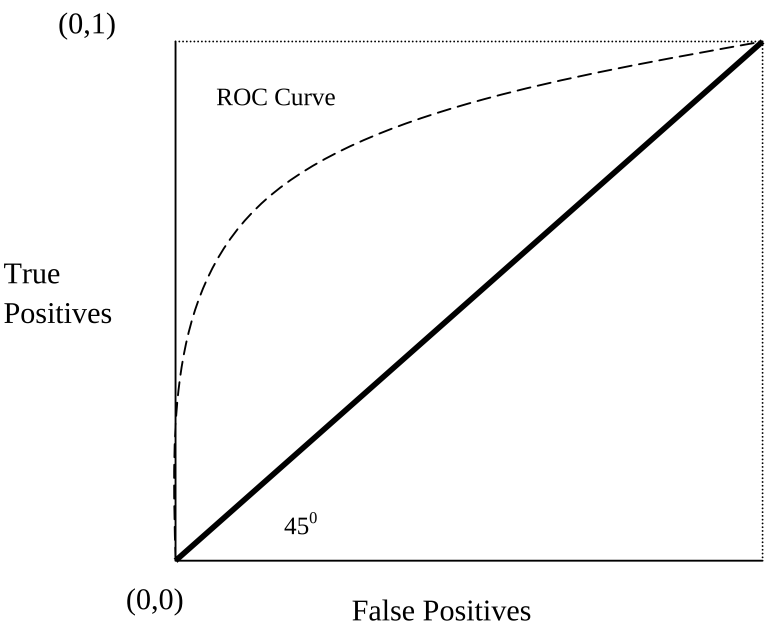



**Figure 2: Youdens J on ROC Curve**

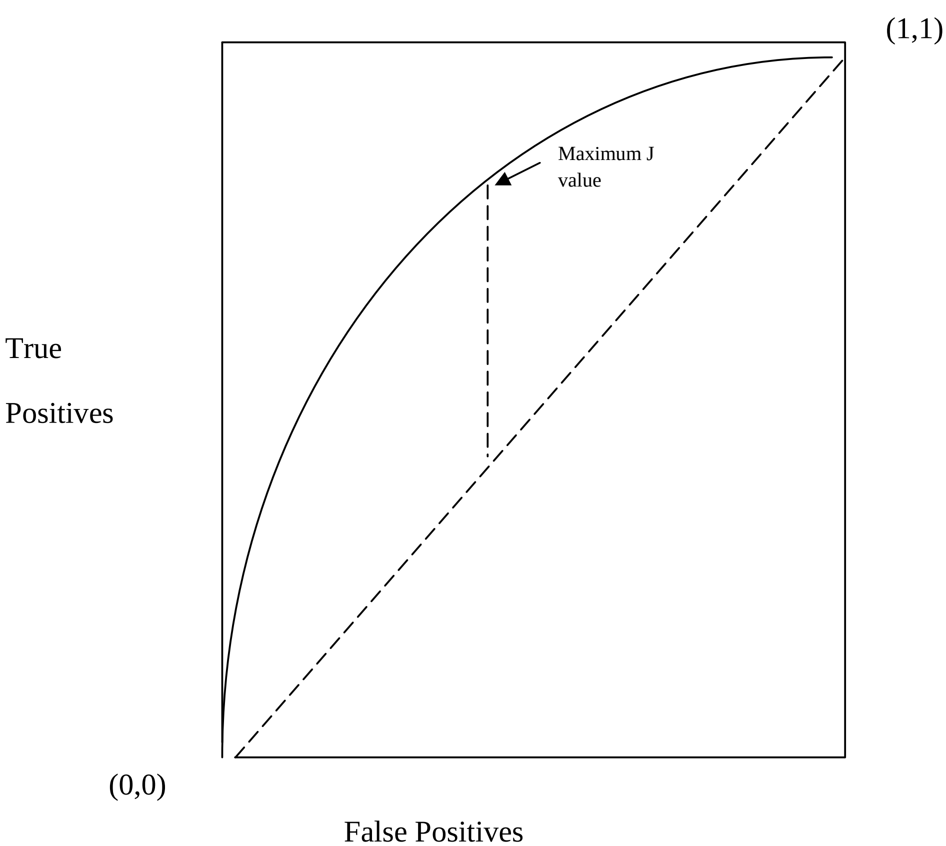


Figure 3: Kernel Density of measured and self-reported BMI

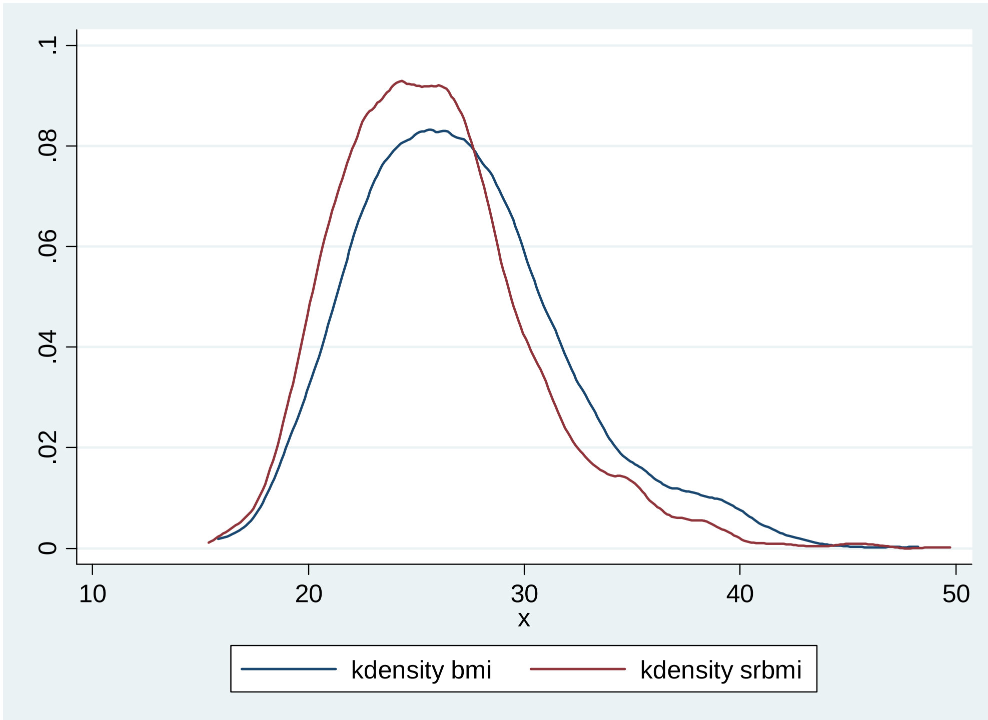
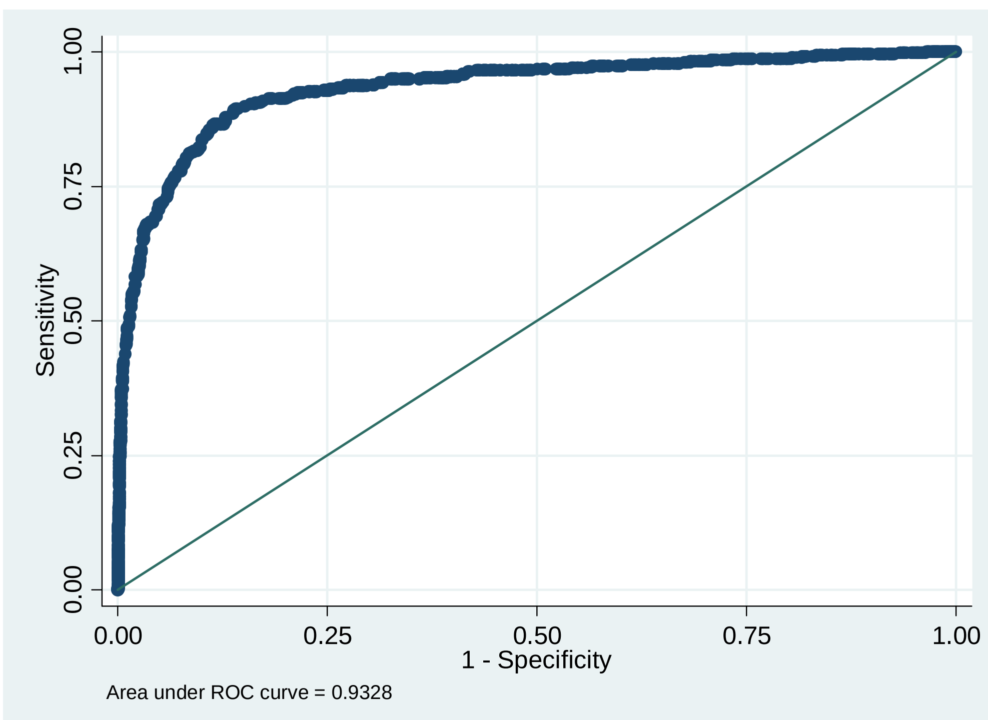



Figure 4: ROC Curve for Self-Reported BMI



**References**

**Bopp, M. and D. Faeh (2009):** "Proposed Obesity Body Mass Index Correction for Self-Reported Data May Not Be Appropriate", *Journal of Epidemiology and Community Health*, Vol. 63.

**Cawley, J., and R. Burkhauser. (2006).** Beyond BMI: The Value of More Accurate Measures of Fatness and Obesity in Social Science Research. *National Bureau of Economic Research*. Working Paper No. 12291.

**Cohen, J. (1968):** "Weighted Kappa: Nominal Scale Agreement with Provision for Scaled Disagreement or Partial Credit", *Psychological Bulletin*, Vol. 70, pp. 213-220.

**Connor Gorber, S, M. Tremblay, D. Moher and B. Gorber (2007):** "A comparison of direct vs. self-report measures for assessing height, weight and body mass index: a systematic review", in *Obesity Review*, Vol. 8, pp. 307-326.

**Dauphinot, V., H. Wolff, F. Naudin, R. Gueguen, C. Sermet, J-M Gaspoz and M Kossovsky (2009):** "New Obesity Body Mass Index Threshold for Self-Reported Data", *Journal of Epidemiology and Community Health*, Vol. 63, pp. 128-132.

**Flegel, K., B. Kit, H. Orpana and B. Graubard (2013):** "Association of All-Cause Mortality with Overweight and Obesity Using Standard Body Mass Index Categories", *Journal of the American Medical Association*, Vol. 309, pp. 71-82.

**Greiner, M., D. Pfeiffer and R.D. Smith (2000):** "Principles and Practical Application of the Receiver-Operating Characteristic Analysis for Diagnostic Tests", *Preventive Veterinary Medicine*, Vol. 45, pp. 23-41.

**Johnson F., L. Cooke, H. Croker and J. Wardle (2008):** "Changing Perceptions of Weight in Great Britain: Comparison of Two Population Surveys", *British Medical Journal*, Vol. 337 (a494), pp. 1-5.

**Morgan, K., H. McGee, D. Watson, I. Perry, M. Barry, E. Shelley, J. Harrington, M. Molcho, R. Layte, N. Tully, E. van Lente, M. Ward, J. Lutomski, R. Conroy, R. Brugha. (2008).** *SLAN 2007: Survey of Lifestyle, Attitudes and Nutrition in Ireland.* Main Report, Dublin. Department of Health and Children.

**Perkins N., and E. Schisterman (2006):** "The Inconsistency of "Optimal Cut-points Using Two ROC Based Criteria", *American Journal of Epidemiology*, Vol. 163, pp. 670-675.

**Shi, Z., A. Taylor, T. Gill, J. Grant, R. Adams and D. Wilson (2009):** "Body Mass Index Threshold for Obesity Using Self-Reported Data amongst Australian Population", *Journal of Epidemiology and Community Health*, Vol. 63.

**Shiely, F., I. Perry, J. Lutomski, J. Harrington, C. Kelleher, H. McGee and K. Hayes (2010):** "Temporal Change in Misclassification of Measured and Self-report Based Body Mass Index Categories - Findings from Three Population Surveys in Ireland", *BMC Public Health*, Vol. 10:560.

**Smits, J (2010):** "A Note on Youden's *J* and its Cost Ratio", *BMC Medical Research Methodology*, Vol. 10:89.

**Table A1: Measured less self-reported obesity (N=1874), standard errors in parenthesis**

| | |
|---|---|
| Age | 0.0412* (0.0205) |
| Gender | 0.445** (0.102) |
| Intermediate 2nd lev | 0.0879 (0.173) |
| Completed 2nd level | 0.205 (0.168) |
| Diploma/Cert | -0.00467 (0.172) |
| 3rd level | 0.168 (0.178) |
| BMI Category | 1.144** (0.0674) |
| Self-assessed health | -0.00306 (0.0535) |
| Married | -0.267* (0.135) |
| Separated/Divorce | -0.212 (0.279) |
| Widow | 0.233 (0.284) |
| Drinker | 0.281* (0.140) |
| Smoker | -0.0810 (0.0934) |
| Constant | -2.422** (0.287) |
| | **Significant at 99%, *Significant at 95% |
| | |
| Observations | 1,874 |
| R-squared | 0.170 |

Omitted category is single, with primary school or less education.

*UCD CENTRE FOR ECONOMIC RESEARCH – RECENT WORKING PAPERS*

WP12/08 Brendan Walsh: 'The Influence of Macroeconomic Conditions and Institutional Quality on National Levels of Life Satisfaction' March 2012
WP12/09 Ronald B Davies and Rodolphe Desbordesz: 'Greenfield FDI and Skill Upgrading' March 2012
WP12/10 Morgan Kelly and Cormac Ó Gráda: 'Change Points and Temporal Dependence in Reconstructions of Annual Temperature: Did Europe Experience a Little Ice Age?' March 2012
WP12/11 Morgan Kelly and Cormac Ó Gráda: 'The Waning of the Little Ice Age' April 2012
WP12/12 Morgan Kelly and Cormac Ó Gráda: 'Agricultural Output, Calories and Living Standards in England before and during The Industrial Revolution' April 2012
WP12/13 Arnaud Chevalier and Orla Doyle: 'Schooling and Voter Turnout - Is there an American Exception?' April 2012
WP12/14 David Madden: 'The Relationship Between Low Birthweight and Socioeconomic Status in Ireland' April 2012
WP12/15 Robert W Fairlie, Kanika Kapur and Susan Gates: 'Job Lock: Evidence from a Regression Discontinuity Design' April 2012
WP12/16 John Regan: 'Ballot Order effects: An analysis of Irish General Elections' April 2012
WP12/17 Orla Doyle: 'Breaking the Cycle of Deprivation: An Experimental Evaluation of an Early Childhood Intervention' April 2012
WP12/18 Ivan Pastine, Tuvana Pastine and Paul Redmond: 'Incumbent-Quality Advantage and Counterfactual Electoral Stagnation in the U.S. Senate' May 2012
WP12/19 Adrian Corcoran and Robert Gillanders: 'Foreign Direct Investment and The Ease of Doing Business' July 2012
WP12/20 Wen Fan and Yuanyuan Ma: 'Estimating the External Returns to Education: Evidence from China' August 2012
WP12/21 David Madden: 'The Socioeconomic Determinants of Mental Stress in Ireland' August 2012
WP12/22 Sinéad Kelleher: 'Playing by the Rules? The Development of an Amended Index to Measure the Impact of Rules of Origin on Intra-PTA Trade Flows' September 2012
WP12/23 Christopher Jepsen, Kenneth Troske and Paul Coomes: 'The Labor-Market Returns to Community College Degrees, Diplomas, and Certificates' September 2012
WP12/24 Ronald B Davies: 'CCCTB 4 EU? SA vs. FA w/ FTA' October 2012
WP12/25 Ronald B Davies and Tine Jeppesen: 'Export mode, Trade Costs, and Productivity Sorting' October 2012
WP12/26 Vincent Hogan, Patrick Massey and Shane Massey: 'Competitive Balance and Match Attendance in European Rugby Union Leagues' October 2012
WP12/27 Mark E McGovern: 'Don't Stress: Early Life Conditions, Hypertension, and Selection into Associated Risk Factors' October 2012
WP12/28 Vincent Hogan, Patrick Massey and Shane Massey: 'Analysing Determinants of Match Attendance in the European Rugby Cup' October 2012
WP12/29 Karl Whelan: 'TARGET2 and Central Bank Balance Sheets' November 2012
WP12/30 Wasiu Adekunle Are: 'Poverty-Reducing Directions of Indirect Marginal Tax Reforms in Ireland' December 2012
WP12/31 Wasiu Adekunle Are: 'Growth and Income Redistribution Components of Changes in Poverty: A Decomposition Analysis for Ireland, 1987-2005' December 2012