The heritability of body mass index among an international sample of monozygotic twins reared apart

DB Allison1, J Kaprio2, M Korkeila2, M Koskenvuo3, MC Neale4 and K Hayakawa5

1 Obesity Research Center, St. Luke’s/Roosevelt Hospital Center, Columbia University College of Physicians and Surgeons, NY, USA; 2 Department of Public Health, University of Helsinki, Helsinki, Finland; 3 Department of Public Health, University of Turku, Finland; 4 Department of Psychiatric Genetics, Medical College of Virginia, USA; 5 Department of Public Health, Kinki University School of Medicine, Osaka, Japan

BACKGROUND: Published heritability estimates ($h^2$) for body mass index (BMI) range from as low as 0.05 to as high as 0.90. The purpose of this paper is to introduce new data to help narrow the range of plausible estimates.

SUBJECTS: Subjects were 53 pairs (23 M; 30 F) of monozygotic twins reared apart (MZAs), whose mean BMI was 24.2 (SD = 4.7). BMI’s were transformed to approximate normality via the Box–Cox transformation. Twin pairs came from the Finnish Twin Cohort (17 pairs), a data base of Japanese twins (10 pairs) and published case histories of primarily American twins (26 pairs).

RESULTS: The $h^2$ for MZAs is given by the correlation among the twin pairs. For the transformed data, the zero-order correlation of twins’ BMIs was 0.79 for all twins, 0.63 for the Finnish twins, 0.73 for the Japanese twins and 0.85 for the ‘archival’ twins. When modeled with regression to control for relevant covariates, the estimate of $h^2$ is either 0.50 or 0.70, depending on one’s definition. The semipartial $r$ was 0.50, suggesting that 50% of the total variance in BMI appears to be genetic in origin after controlling the covariates. The partial $r$ was 0.70, suggesting that 70% of the variance in BMI that is not accounted for by the covariates can be attributed to genetic variation. Separation age had a small positive correlation with the change in intra-pair difference in BMI, suggesting that these estimates of $h^2$ are not biased upwards due to early shared environment.

CONCLUSIONS: Findings are consistent with past studies of MZAs and suggest that $h^2$ estimates between 0.50 and 0.70 are reasonable. Implications of this finding are discussed.

Keywords: obesity; twins; separated twins; heritability; genetics; body mass index

Introduction

It is now generally accepted that human adiposity, as operationalized by relative weight, has a statistically significant genetic component. However, previous investigations have yielded estimates of $h^2$ which have ranged from as low as 0.05 to as high as 0.90. Thus, there is continued debate regarding the proportion of between person phenotypic variance that is attributable to between person genotypic variance (i.e. heritability or $h^2$).

Why is it valuable to provide precise estimates of $h^2$?

It might be argued that, although the demonstration of adiposity’s heritability (i.e. that $h^2$ is not 0) is important, an exact point estimate is of less relevance. However, the accurate estimation of $h^2$ is advantageous for at least two reasons.

Genetic effects do not mean predetermination, but they do mean predisposition. The larger $h^2$ is, the smaller the impact one can expect conventional environmental manipulations to have. ('Conventional' environmental manipulations refers to those aspects of the environment that typically vary among individuals. In other words, $h^2$ provides little information about the likely effects of novel environmental events such as the introduction of a new drug, foodstuff, virus, law or incentive plan into a population.) This obviously has important implications for public health obesity prevention proposals.

Second, many investigators are currently searching for specific molecular markers that are linked to or associated with human obesity. The power of such studies can be greatly affected by $h^2$. All other things being equal, the greater $h^2$ is, the more power one has to detect significant effects in such studies.

Possible reasons for varying estimates of $h^2$

There are several plausible explanations for the range of $h^2$ estimates observed. These include sampling error, differences in the populations observed, the type of data modeled (e.g. twin, family or adoptee), and the type of statistical model fit to the data.

Sampling error is an unlikely explanation for the variation of estimates given that many of the studies have employed rather large samples (frequently several
thousand subjects).}\textsuperscript{3,5–7} Furthermore, no systematic differences (i.e. consistently higher or lower estimates of $h^2$) between North American\textsuperscript{5–9,11} and European\textsuperscript{6–8} studies have been demonstrated. This is contrary to what would be expected if systematic differences among the populations being assessed significantly influenced the range of $h^2$ estimates yielded. However, it is also recognized that there is no necessity that $h^2$ must be the same across all populations.

The lowest estimates observed\textsuperscript{3} may have been spuriously low due to the statistical model employed (see further discussion of the statistical model, Allison, 1995).\textsuperscript{12} However, following the elimination of the estimate produced by this model, estimates of $h^2$ continued to range from about 0.25 to about 0.90. Thus, this leaves the type of study employed as the most likely explanation for the varying estimates.

Twin studies typically yield higher heritability estimates than adoption or family studies.\textsuperscript{13–17} Each type of study potentially suffers from biases that may distort results. With regard to twin studies, a central assumption is the ‘equal environment assumption’ which implies that monozygotic (MZ) twins are no more similar with respect to the environmental causes of the phenotype under study than are dizygotic (DZ) twins. If the equal environments assumption does not hold and MZ twins are more similar for their environmental causes of adiposity than are DZ twins, then this would cause twin studies to overestimate $h^2$. Having observed the greater estimates of $h^2$ for BMI obtained among twins, some authors have suggested that the equal environments assumption may not be valid in twin studies of BMI.\textsuperscript{17} However, with respect to cognitive, personality and psychopathology variables, this assumption has been repeatedly tested and found to be valid.\textsuperscript{18,19} Similar tests involving obesity-related phenotypes have also supported the validity of the equal environments assumption.\textsuperscript{20} Moreover, to the extent that some genetic effects are non-additive (e.g. genetic dominance) twin studies correctly yield higher estimates.

Just as twin studies are susceptible to potential biases that can inflate $h^2$, adoption and family studies are susceptible to potential biases that can cause $h^2$ to be underestimated. These include false paternity (which can be as high as 30% in some samples but is usually closer to 5%),\textsuperscript{21} failure to account for age-specific genetic effects\textsuperscript{15} which have been shown to exist,\textsuperscript{22} and failure to account for epistatic effects.\textsuperscript{23}

Studies of monozygotic twins reared apart

One alternative to traditional twin, family and adoption studies is the study of monozygotic twins reared apart (MZAs). The study of MZAs has all of the advantages of the traditional twin study, but does not rely on the equal environments assumption. Rather, it assumes only that the co-twins are uncorrelated for their environments. It does, however, rely on random placement. Of course, this cannot be strictly true because twins share the intra-uterine environment. However, it does not appear that ordinary variations in the intra-uterine environment have an enduring impact on the BMI of twins.\textsuperscript{24}

To date, two data sets of MZAs have been analyzed with respect to BMI. Using 93 pairs of Swedish MZAs, Stunkard et al.\textsuperscript{8} estimated $h^2$ to be 0.70 for men and 0.66 for women. Using 34 pairs of male and female British MZAs, Price and Gottesman\textsuperscript{25} estimated $h^2$ to be 0.61.\textsuperscript{26}

The purpose of the present study was to add to the existing literature by estimating $h^2$ for BMI using additional sources of MZAs. Five additional samples of MZAs were considered: the Finnish cohort,\textsuperscript{27,28} a Japanese twin study,\textsuperscript{29} numerous published case histories,\textsuperscript{30} Cyril Burt’s data\textsuperscript{31} and the Minnesota twins.\textsuperscript{32} Cyril Burt’s data was excluded from the analyses in light of recent reviews suggesting possible concerns about the data collection.\textsuperscript{33} Furthermore, the data on the Minnesota MZAs is planned to be published separately (T Bouchard, personal communication). Thus, the present study will report on the remaining three sources of MZAs.

### Methods

#### Subjects

The subjects in this study were 53 pairs of MZ twins reared apart from three sources: (1) the Finnish twin cohort (17 pairs); (2) a Japanese twin cohort (10 pairs) and (3) archival published case histories (26 pairs). Table 1 provides descriptive statistics on the combined sample as well as each individual sample.

The Finnish twin cohort consists of all Finnish adult like-sexed twin pairs ($N = 17,357$) born between 1958 and with both members alive in 1967. The Finnish MZAs were identified by means of a two-stage ques-

<table>
<thead>
<tr>
<th>Source</th>
<th>Years at separation</th>
<th>cm (SD)</th>
<th>kg (SD)</th>
<th>BMI (SD)</th>
<th>Age (SD)</th>
<th>% Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Finland</td>
<td>3.80 (2.57)</td>
<td>169.24  (8.92)</td>
<td>71.21  (13.57)</td>
<td>24.76  (3.72)</td>
<td>36.39  (11.90)</td>
<td>41.2</td>
</tr>
<tr>
<td>Japan</td>
<td>0.90 (1.73)</td>
<td>154.41  (6.22)</td>
<td>53.17  (8.75)</td>
<td>22.23  (3.00)</td>
<td>64.70  (6.27)</td>
<td>40.0</td>
</tr>
<tr>
<td>Archive</td>
<td>1.20 (1.40)</td>
<td>160.78  (7.35)</td>
<td>63.67  (16.42)</td>
<td>24.55  (5.91)</td>
<td>39.88  (18.63)</td>
<td>73.1</td>
</tr>
<tr>
<td>Combined</td>
<td>1.98 (2.26)</td>
<td>162.29  (9.28)</td>
<td>64.11  (15.49)</td>
<td>24.18  (4.86)</td>
<td>43.44  (18.07)</td>
<td>56.6</td>
</tr>
</tbody>
</table>
tionnaire process. Initially, subjects were assessed for cohabitation and age at a separation by a baseline (1975) survey of the older Finnish twin cohort with an overall response rate of 89%. Second, a follow-up questionnaire was sent to potential MZAs concerning the circumstances of separation. Subjects were invited to participate in a project investigating genetic and environmental determinants of neuropsychological and neurophysiological traits. During a two-day study period at the Institute of Occupational Health in Helsinki, the MZA twins were interviewed about their lives pre and post separation, the age at separation and more detailed reasons for separation than in the earlier questionnaire studies. (This sampling procedure yielded 18 MZA pairs. Complete data were available for 17 pairs (95%). When this information is combined with the 89% response rate to the 1975 survey, the best estimate is that this sample includes 85% of all living Finnish MZAs born before 1958.) The transcripts of the taped interviews were used to define the age of separation of the MZA twins. Zygosity was determined by 11 blood groups as described by Langinaino et al.

Height and weight were determined by self-report measures. Such measures have been shown to be highly correlated \( r^2 > 0.90 \) with actual height and weight in numerous studies. Specifically, a thorough review of studies reporting correlations between adult self-report values and actual measured values of height and weight indicated the percentage of correlations exceeding 0.90 were 81% and 94%, for height and weight, respectively.

The Japanese twins come from the Kinki University Adult Twin Registry which consists of 1314 adult twin pairs and was begun in 1981. A total of 543 like-sex pairs of twins (454 male, 89 female) in an age range of 47 to 87 years were included in a survey study from which possible MZAs were identified. All twins are from Japan and are predominantly from Western Japan. Height and weight and zygosity were determined by questionnaire. The establishment of zygosity was shown to be 97% accurate when cross-checked against serological testing.

The archival twins were collected from several sources. Farber compiled what appears to be an exhaustive listing of published sets of MZAs (not including the Minnesota twins which were published later). From Farber’s listing, 11 pairs were selected from Newman et al., from Juel-Nielsen and three from Burks. Only pairs for whom weight and height were available were used. Twins were from the United States and Denmark and all are believed to be white. Methods by which the twins were ascertained varied from case to case. Zygosity was generally determined by extensive physical comparisons including fingerprint analyses.

### Variables
Variables used were age, height (cm), weight (kg), source of data and sex. Height and weight were converted to BMI (kg/m²).

### Statistical methods
Among MZAs, an estimate of \( h^2 \) is given by the phenotypic correlation between twins. Furthermore, multiple partial and semi-partial correlations between twin 1’s BMI and twin 2’s BMI after controlling for age, sex, their interactive effect and source of data were computed. Although, several indices of relationship magnitude are provided, (i.e. partial and semi-partial correlations) neither one is preferable to the other. Specifically, each value represents a different quantity and can be of interest depending on the particular perspective one is adopting. The choice of which twin to use as the dependent variable was irrelevant as the partial correlation is equivalent either way. (Although the semi-partial correlation can be different depending on which twin is used as the dependent variable (DV), analyses were repeated with each twin as the DV and there were no substantial differences in the results. The semi-partial (or part) \( r^2 \) is the proportion of the total variance in the dependent variable accounted for by the variable of interest after controlling for the covariates. The partial \( r^2 \) is the proportion of the remaining variance in the DV accounted for by the variable of interest after controlling for the covariates.) Analyses were conducted with both raw BMI and BMI transformed to approximate normality via a Box-Cox transformation as implemented by the UNICORN software. The use of the transformed estimates may be more appropriate because they more closely meet the assumptions of the analyses. However, estimates are presented for both transformed and untransformed data. The transformation entailed raising BMI to the \(-0.854\) power.

### Results
Table 2 contains the variance-covariance and correlation matrices for the twins. The simplest way to estimate \( h^2 \) of MZAs is by the correlation between pairs. For the total sample, the raw data correlation was 0.81 \((p < 0.0005; 95\% CI = 0.69–0.88)\) and the transformed data correlation was 0.79 \((p < 0.0005; 95\% CI = 0.66–0.87)\). For the Finnish, Japanese and archival samples, the \( h^2 \)’s estimated by this method were 0.54 \((95\% CI = 0.08–0.80, 0.77 (95\% CI = 0.25–0.93)\) and 0.89 \((95\% CI = 0.76–0.95)\), respectively using untransformed BMI. When transformed BMI was used, the corresponding correlations were 0.65 \((95\% CI = 0.24–0.85), 0.73 (95\% CI = 0.17–0.92)\) and 0.85 \((95\% CI = 0.68–0.93)\), respectively.

Differences between the correlations were tested via the Z-test for the equality of independent correlations. There was no significant difference between the Finnish and Japanese correlations \((p = 0.386 \text{ raw}; p = 0.745 \text{ transformed})\) or the Japanese and archival correlations \((p = 0.328 \text{ raw}; p = 0.460 \text{ transformed})\). A significant difference was observed between the Finnish and archival correlations for the raw BMI data \((p = 0.015)\).
Table 2 Variance–covariance and correlation matrices

<table>
<thead>
<tr>
<th></th>
<th>Sex</th>
<th>Age sep.</th>
<th>Age</th>
<th>CM1</th>
<th>KG1</th>
<th>CM2</th>
<th>KG2</th>
<th>BM11</th>
<th>BM12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>0.250</td>
<td>-0.194</td>
<td>0.039</td>
<td>-0.419</td>
<td>-0.154</td>
<td>-0.410</td>
<td>-0.221</td>
<td>0.069</td>
<td>-0.023</td>
</tr>
<tr>
<td>Age sep.</td>
<td>-0.219</td>
<td>5.107</td>
<td>0.060</td>
<td>0.334</td>
<td>0.273</td>
<td>0.305</td>
<td>0.338</td>
<td>0.131</td>
<td>0.239</td>
</tr>
<tr>
<td>Age</td>
<td>0.351</td>
<td>2.431</td>
<td>326.668</td>
<td>-0.378</td>
<td>0.099</td>
<td>-0.421</td>
<td>0.047</td>
<td>0.337</td>
<td>0.298</td>
</tr>
<tr>
<td>CM1</td>
<td>-1.977</td>
<td>7.119</td>
<td>-64.444</td>
<td>88.900</td>
<td>-0.543</td>
<td>0.966</td>
<td>0.615</td>
<td>0.085</td>
<td>0.232</td>
</tr>
<tr>
<td>KG1</td>
<td>-1.182</td>
<td>9.508</td>
<td>27.622</td>
<td>78.744</td>
<td>236.999</td>
<td>0.552</td>
<td>0.873</td>
<td>0.879</td>
<td>0.797</td>
</tr>
<tr>
<td>CM2</td>
<td>-1.870</td>
<td>6.285</td>
<td>-69.393</td>
<td>83.043</td>
<td>77.485</td>
<td>83.155</td>
<td>0.640</td>
<td>0.109</td>
<td>0.236</td>
</tr>
<tr>
<td>KG2</td>
<td>-1.726</td>
<td>11.886</td>
<td>13.099</td>
<td>90.275</td>
<td>209.357</td>
<td>90.899</td>
<td>242.737</td>
<td>0.681</td>
<td>0.894</td>
</tr>
<tr>
<td>BM1</td>
<td>0.173</td>
<td>1.489</td>
<td>30.663</td>
<td>4.006</td>
<td>68.077</td>
<td>5.009</td>
<td>53.342</td>
<td>25.299</td>
<td>0.813</td>
</tr>
<tr>
<td>BM2</td>
<td>-0.053</td>
<td>2.530</td>
<td>25.262</td>
<td>10.239</td>
<td>57.570</td>
<td>10.102</td>
<td>65.305</td>
<td>19.185</td>
<td>21.997</td>
</tr>
</tbody>
</table>

Note: Variances are on the leading diagonal, covariances are below the diagonal, and correlations are above the diagonal. Variables are sex, age at separation (Age sep.), age at measurement (Age), height of twin 1 (CM1), weight of twin 1 (KG1), height of twin 2 (CM2), weight of twin 2 (KG2), body-mass index of twin 1 (BM1), and body-mass index of twin 2 (BM2).

However, no significant difference was noted for the transformed BMI data (p = 0.166).

Following these preliminary analyses, the data were modeled using multiple correlation/regression. The purpose of this analysis was to estimate the twin correlation after controlling for variables that twins share and that affect their BMIs (i.e., sex, age, source of data). The regression results are detailed in Table 3. For the untransformed data, the semipartial correlation was 0.55 (95% CI = 0.31–0.72), suggesting that 55% of the total variance in BMI appears to be genetic in origin after controlling the covariates. The partial r was 0.74 (95% CI = 0.57–0.84), suggesting that 74% of the variance in BMI that is not accounted for by the covariates can be attributed to genetic variation. For the transformed data, the semipartial correlation was 0.50 (95% CI = 0.25–0.68), suggesting that 50% of the total variance in BMI appears to be genetic in origin after controlling the covariates. The partial r was 0.70 (95% CI = 0.52–0.82), suggesting that 70% of the variance in BMI that is not accounted for by the covariates can be attributed to genetic variation.

Finally, the correlation between age at separation and absolute intra-pair difference in BMI was computed in an effort to assess whether twins who remained together longer tended to be more similar. Results showed that there was a small positive correlation between age at separation and intra-pair difference in BMI indicating that twins who remained together longer were less similar in BMI (r = 0.24, p = 0.039). However, this correlation was not significant when transformed BMIs were used (r = 0.09; p = 0.261).

Discussion

Previous studies have offered varying estimates of the heritability of BMI with adoption and family studies suggesting relatively low estimates and twin studies suggesting relatively high estimates. The present study avoids the equal environments assumption of the classical twin study and found heritability estimates of approximately 0.50–0.70 depending on one’s definition of heritability. A significant difference was demonstrated between the Finnish and archival samples’ BMI heritability estimates. However, the considerably greater age range within the published case histories may account for this difference between estimates based on the original data. Furthermore, evaluation of the estimates based on the transformed data indicated no significant differences between the groups.

Table 3 Hierarchical regression results

<table>
<thead>
<tr>
<th>Variables</th>
<th>F</th>
<th>d.f.</th>
<th>p-value</th>
<th>R²</th>
<th>Incremental R²</th>
<th>Partial r for BMI</th>
<th>Semi-partial r for BMI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Untransformed data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Age, age², sex, age x sex</td>
<td>3.51</td>
<td>4, 48</td>
<td>0.014</td>
<td>0.226</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>2. Data source (Japan vs Finland vs archival)</td>
<td>6.33</td>
<td>6, 46</td>
<td>&lt;0.0001</td>
<td>0.452</td>
<td>0.226</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>3. BMI</td>
<td>19.70</td>
<td>7, 45</td>
<td>&lt;0.0001</td>
<td>0.754</td>
<td>0.302</td>
<td>0.74</td>
<td>0.55</td>
</tr>
<tr>
<td>Transformed data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Age, age², sex, age x sex</td>
<td>3.90</td>
<td>4, 48</td>
<td>0.008</td>
<td>0.245</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>2. Data source (Japan vs Finland vs archival)</td>
<td>7.49</td>
<td>6, 46</td>
<td>&lt;0.0001</td>
<td>0.494</td>
<td>0.249</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>3. BMI</td>
<td>18.30</td>
<td>7, 45</td>
<td>&lt;0.0001</td>
<td>0.740</td>
<td>0.246</td>
<td>0.70</td>
<td>0.50</td>
</tr>
</tbody>
</table>
The use of MZAs to estimate $h^2$ assumes that the twins were truly separated and did not share environmental influences. Theoretically, to the extent that this assumption is incorrect, this approach would overestimate $h^2$. However, the lack of a negative correlation between age at separation and degree of similarity suggests that this is not a concern. This is consistent with Price and Gottesman’s finding that MZAs placed in more similar rearing environments were no more similar in BMI than were MZAs placed in less similar environments. It is also consistent with the general finding from both twin and adoption studies that shared common environmental factors have no significant effects on BMI. Another assumption of the MZA approach is that the MZAs sampled are representative of the population, with the greatest concern being that only unusually similar twins are located for inclusion. However, the Finnish data reported herein and the Swedish data reported by Stunkard et al. are derived from population-based registries and the results of these samples are quite consistent with the results from the other samples of MZAs.

There are several implications of these findings. As pointed out earlier, the degree of $h^2$ may provide some sense of the malleability of BMI in response to manipulations of conventional environmental factors. The magnitude of $h^2$ also has implications for developing models of the inheritance of BMI. For example, if a putative intermediary phenotype explains a proportion of variance in BMI that is significantly less than the $h^2$ for BMI, then this indicates other intermediary phenotypes must also be involved. This type of thinking leads to more complex path analytic approaches that may be useful in defining intermediary phenotypes between genotype and BMI. Planning, and in some cases, fitting such models requires an awareness of the magnitude of $h^2$.

Additionally, $h^2$ is a function of both the number of genes influencing a phenotype and the proportion of phenotypic variance accounted for by each of these genes. Both of these factors will influence the power of tests for specific gene effects or associations. The effect of $h^2$ on the power of certain nonparametric methods of locating chromosomal areas of influence was well illustrated by Goldgar.

The results of the present study are quite consistent with estimates $h^2$ provided by two past studies of MZA’s. Stunkard et al. estimated $h^2$ at 0.70 among men and 0.66 among women, while Price and Gottesman estimated $h^2$ at 0.61. Combining these estimates with the partial correlation obtained herein by taking a weighted average, one obtains a correlation of 0.675 (95% CI = 0.586–0.747). (Correlations were weighted by the inverse of the variances after being converted to Fisher’s Z’s and the weighted average Z was then back transformed to r.) It should be noted that the correlation from Price and Gottesman is a partial correlation adjusted for age. The estimates from Stunkard et al. are zero-order correlations. Collectively, these data are remarkably consistent and suggest that estimates of $h^2$ between 0.59 and 0.75 are quite reasonable. They suggest further that past twin studies reporting estimates of $h^2$ much above 0.75 may have been biased upwards and that family and adoption studies reporting estimates much below 0.55 may have been biased downwards.

Acknowledgements
Supported in part by grants P30DK26687 and R29DK47256-01A1 from the National Institutes of Health and by the Academy of Finland. We are grateful to Dr. David Goldgar for his helpful suggestions on this manuscript and to Kenneth M. Carpenter, Melissa I. Kezis and Rachel Weber for their assistance in its preparation.

References
16. Vogler GP, Sorensen TIA, Stunkard AJ, Sriniwasan MR, Rao DC. Influences of genes and shared family environment on adult body mass index assessed in an adoption


