Online-only material

Contents:
Supplementary Table 1. Participants, dates, and data availability in each HUNT and YoungHUNT (YH) survey
Supplementary Table 2. International classification of diseases (ICD) codes included for each cause of death
Supplementary Table 3. Characteristics of parents and offspring according to quintiles of mothers' BMI
Supplementary Table 4. Characteristics of parents and offspring according to quintiles of fathers' BMI

Supplementary Figure 1. Parent-offspring associations in BMI
Supplementary Figure 2. Fitted associations of all-cause and cause-specific mortality with own and offspring BMI (kg m⁻²)
Supplementary Figure 3. Bias components from measured covariates in the association of all-cause mortality with (i) own BMI and (ii) offspring BMI as an instrument for own BMI
Supplementary Figure 4. Directed acyclic graph illustrating the role of potential confounders and mediators of the association between an exposure (e.g. BMI) and an outcome (e.g. mortality) when an instrument (e.g. offspring BMI) is used

Supplementary Methods 1. Details of estimation using instrumental variables
Supplementary Methods 2. Stata do-file used to estimate linear associations of mortality with BMI
Supplementary Methods 3. Stata do-file used to estimate and plot cubic spline associations of mortality with BMI
## Supplementary Table 1. Participants, dates, and data availability in each HUNT and YoungHUNT (YH) survey

<table>
<thead>
<tr>
<th>Dates</th>
<th>HUNT1</th>
<th>HUNT2</th>
<th>YH1</th>
<th>YH2</th>
<th>HUNT3</th>
<th>YH3</th>
<th>Coding details of confounders</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total participants</td>
<td>77 212</td>
<td>65 237</td>
<td>8983</td>
<td>2399</td>
<td>50 807</td>
<td>8200</td>
<td></td>
</tr>
<tr>
<td>BMI</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Continuous</td>
</tr>
<tr>
<td>Blood pressure</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Continuous</td>
</tr>
<tr>
<td>LDL cholesterol</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>Continuous</td>
</tr>
<tr>
<td>Smoking (own &amp; offspring)</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Never-smoker, Current-smoker, or Ex-smoker</td>
</tr>
<tr>
<td>Alcohol</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>1, 1-4, or &gt;4 consumption occasions per fortnight</td>
</tr>
<tr>
<td>Education</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>&lt;10, 10-12, or &gt;12 years' education</td>
</tr>
<tr>
<td>Employment (own &amp; spouse)</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>Never worked/Unskilled, Skilled/Driver, Clerical, Professional, Senior, Farmer/Fisher, or Self-employed</td>
</tr>
<tr>
<td>Exercise</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>None, Low intensity, or High intensity</td>
</tr>
<tr>
<td>General good health</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Binary</td>
</tr>
<tr>
<td>Functional impairment</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>Binary</td>
</tr>
<tr>
<td>Angina pectoris</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>Binary</td>
</tr>
<tr>
<td>Diabetes</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>Binary</td>
</tr>
<tr>
<td>MI history</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>Binary</td>
</tr>
<tr>
<td>Antihypertensive use</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>Binary</td>
</tr>
</tbody>
</table>

Missing data were included as a category for each categorical variable when used in adjustment. In the covariate balance tests, participants lacking data on the variable being used as a proxy for unmeasured covariates were omitted and categories were combined to render all categorical variables binary.
## Supplementary Table 2. International classification of diseases (ICD) codes included for each cause of death

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>All cause</td>
<td>all</td>
<td>all</td>
<td>all</td>
</tr>
<tr>
<td>Cardiovascular disease</td>
<td>3900-4441; 4444-4589; 3900-4599</td>
<td>4100-4149; 4299; 4100-4149; 4292</td>
<td>I(^a); G450-G459</td>
</tr>
<tr>
<td>Coronary heart disease</td>
<td>7820-7829</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stroke</td>
<td>2930-2931; 3440-3449; 2904; 3420-3429; 3440-4389</td>
<td>2904; 3420-3429; 3440-4389; 4300-4389</td>
<td>I200-I259; I516</td>
</tr>
<tr>
<td>Diabetes</td>
<td>2500-2509</td>
<td>2500-2509</td>
<td>E100-E149</td>
</tr>
<tr>
<td>Respiratory diseases</td>
<td>4600-5199</td>
<td>4600-5199</td>
<td>J(^a)</td>
</tr>
<tr>
<td>External causes</td>
<td>8000-9999</td>
<td>8000-9999</td>
<td>V(^a); W(^a); X(^a); Y(^a)</td>
</tr>
<tr>
<td>Lung cancer</td>
<td>1620-1639</td>
<td>1620-1639</td>
<td>C330-C349</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>1740-1749</td>
<td>1740-1759</td>
<td>C500-C509</td>
</tr>
<tr>
<td>Prostate cancer</td>
<td>1850-1859</td>
<td>1850-1859</td>
<td>C610-C619</td>
</tr>
<tr>
<td>Colorectal cancer</td>
<td>1530-1549</td>
<td>1530-1549</td>
<td>C180-C219</td>
</tr>
<tr>
<td>Pancreatic cancer</td>
<td>1570-1579</td>
<td>1570-1579</td>
<td>C250-C259</td>
</tr>
<tr>
<td>Stomach cancer</td>
<td>1510-1519</td>
<td>1510-1519</td>
<td>C160-C169</td>
</tr>
<tr>
<td>Ovarian cancer</td>
<td>1830</td>
<td>1830</td>
<td>C560-C569</td>
</tr>
</tbody>
</table>

\(^{a}\)Any code beginning with the indicated letter.
## Supplementary Table 3. Characteristics of parents and offspring according to quintiles of mothers' BMI

<table>
<thead>
<tr>
<th>Person, Measurement</th>
<th>Quintile of mother’s BMI</th>
<th>Linear or logistic regression per SD</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>2&lt;sup&gt;nd&lt;/sup&gt;</td>
<td>3&lt;sup&gt;rd&lt;/sup&gt;</td>
</tr>
<tr>
<td><strong>Offspring</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean BMI (kg m&lt;sup&gt;-2&lt;/sup&gt;)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>23.0</td>
<td>23.5</td>
<td>24.0</td>
</tr>
<tr>
<td>Mean systolic blood pressure (mm Hg)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>124.9</td>
<td>125.2</td>
<td>125.6</td>
</tr>
<tr>
<td>Mean diastolic blood pressure (mm Hg)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>73.8</td>
<td>73.8</td>
<td>74.1</td>
</tr>
<tr>
<td>Mean age at BMI measurement (years)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>28.1</td>
<td>28.1</td>
<td>28.1</td>
</tr>
<tr>
<td>Proportion ever smoked (%)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>37.7</td>
<td>37.3</td>
<td>37.6</td>
</tr>
<tr>
<td>Proportion drinking &gt;= 5 times fortnightly (%)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>4.7</td>
<td>3.9</td>
<td>4.3</td>
</tr>
<tr>
<td>Proportion educated &gt;=10 years (%)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>76.3</td>
<td>77.9</td>
<td>76.5</td>
</tr>
<tr>
<td>Proportion physically active (%)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>89.8</td>
<td>92.3</td>
<td>92.3</td>
</tr>
<tr>
<td><strong>Mothers</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean BMI (kg m&lt;sup&gt;-2&lt;/sup&gt;)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>20.5</td>
<td>22.9</td>
<td>24.6</td>
</tr>
<tr>
<td>Mean systolic blood pressure (mm Hg)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>129.8</td>
<td>131.5</td>
<td>133.6</td>
</tr>
<tr>
<td>Mean diastolic blood pressure (mm Hg)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>78.7</td>
<td>79.8</td>
<td>81.1</td>
</tr>
<tr>
<td>Mean age at offspring's birth (years)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>27.4</td>
<td>27.4</td>
<td>27.3</td>
</tr>
<tr>
<td>Mean age at BMI measurement (years)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>47.6</td>
<td>47.4</td>
<td>47.4</td>
</tr>
<tr>
<td>Proportion ever smoked (%)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>54.1</td>
<td>49.1</td>
<td>46.5</td>
</tr>
<tr>
<td>Proportion drinking &gt;= 5 times fortnightly (%)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>3.4</td>
<td>3.3</td>
<td>2.8</td>
</tr>
<tr>
<td>Proportion educated &gt;=10 years (%)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>47.6</td>
<td>50.9</td>
<td>49.4</td>
</tr>
<tr>
<td>Proportion physically active (%)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>85.3</td>
<td>88.1</td>
<td>87.6</td>
</tr>
<tr>
<td><strong>Fathers</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean BMI (kg m&lt;sup&gt;-2&lt;/sup&gt;)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>25.0</td>
<td>25.3</td>
<td>25.5</td>
</tr>
<tr>
<td>Mean systolic blood pressure (mm Hg)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>137.0</td>
<td>137.7</td>
<td>138.5</td>
</tr>
<tr>
<td>Mean diastolic blood pressure (mm Hg)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>84.5</td>
<td>84.8</td>
<td>84.9</td>
</tr>
<tr>
<td>Mean age at offspring's birth (years)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>30.0</td>
<td>30.3</td>
<td>30.3</td>
</tr>
<tr>
<td>Mean age at BMI measurement (years)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>47.0</td>
<td>47.3</td>
<td>47.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>Proportion ever smoked (%)(^b)</td>
<td>64.7</td>
<td>62.9</td>
<td>63.6</td>
</tr>
<tr>
<td>Proportion drinking &gt;= 5 times fortnightly (%)(^b)</td>
<td>9.1</td>
<td>8.4</td>
<td>8.7</td>
</tr>
<tr>
<td>Proportion educated &gt;=10 years (%)(^b)</td>
<td>56.1</td>
<td>56.0</td>
<td>52.4</td>
</tr>
<tr>
<td>Proportion in non-manual employment (%)(^b)</td>
<td>44.2</td>
<td>42.6</td>
<td>40.2</td>
</tr>
<tr>
<td>Proportion physically active (%)(^b)</td>
<td>87.2</td>
<td>86.7</td>
<td>86.6</td>
</tr>
</tbody>
</table>

Quintiles were calculated among women of similar age, measured at the same survey occasion.

\(^a\)Linear regression coefficients, calculated per standard deviation \((4.33 \text{ kg m}^{-2})\) of maternal BMI, adjusted for age and survey occasion.

\(^b\)Logistic regression odds ratios, calculated per standard deviation \((4.33 \text{ kg m}^{-2})\) of maternal BMI, adjusted for age and survey occasion.
### Supplementary Table 4. Characteristics of parents and offspring according to quintiles of fathers’ BMI

<table>
<thead>
<tr>
<th>Person, Measurement</th>
<th>Quintile of father’s BMI</th>
<th>Linear or logistic regression per SD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>2&lt;sup&gt;nd&lt;/sup&gt;</td>
</tr>
<tr>
<td>Offspring</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean BMI (kg m&lt;sup&gt;2&lt;/sup&gt;)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>22.9</td>
<td>23.5</td>
</tr>
<tr>
<td>Mean systolic blood pressure (mm Hg)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>124.8</td>
<td>124.8</td>
</tr>
<tr>
<td>Mean diastolic blood pressure (mm Hg)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>73.2</td>
<td>73.2</td>
</tr>
<tr>
<td>Mean age at BMI measurement (years)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>26.9</td>
<td>27.2</td>
</tr>
<tr>
<td>Proportion ever smoked (%)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>36.1</td>
<td>35.2</td>
</tr>
<tr>
<td>Proportion drinking &gt;= 5 times fortnightly (%)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>3.7</td>
<td>3.6</td>
</tr>
<tr>
<td>Proportion educated &gt;=10 years (%)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>78.5</td>
<td>80.0</td>
</tr>
<tr>
<td>Proportion in non-manual employment (%)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>47.2</td>
<td>49.6</td>
</tr>
<tr>
<td>Proportion physically active (%)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>91.6</td>
<td>92.2</td>
</tr>
<tr>
<td>Mothers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean BMI (kg m&lt;sup&gt;2&lt;/sup&gt;)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>24.6</td>
<td>24.7</td>
</tr>
<tr>
<td>Mean systolic blood pressure (mm Hg)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>130.8</td>
<td>131.0</td>
</tr>
<tr>
<td>Mean diastolic blood pressure (mm Hg)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>80.6</td>
<td>80.6</td>
</tr>
<tr>
<td>Mean age at offspring’s birth (years)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>27.3</td>
<td>27.2</td>
</tr>
<tr>
<td>Mean age at BMI measurement (years)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>44.3</td>
<td>44.5</td>
</tr>
<tr>
<td>Proportion ever smoked (%)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>52.0</td>
<td>48.6</td>
</tr>
<tr>
<td>Proportion drinking &gt;= 5 times fortnightly (%)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>2.5</td>
<td>2.7</td>
</tr>
<tr>
<td>Proportion educated &gt;=10 years (%)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>49.7</td>
<td>49.3</td>
</tr>
<tr>
<td>Proportion in non-manual employment (%)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>56.1</td>
<td>56.4</td>
</tr>
<tr>
<td>Proportion physically active (%)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>87.5</td>
<td>88.4</td>
</tr>
<tr>
<td>Fathers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean BMI (kg m&lt;sup&gt;2&lt;/sup&gt;)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>21.6</td>
<td>23.8</td>
</tr>
<tr>
<td>Mean systolic blood pressure (mm Hg)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>135.1</td>
<td>137.3</td>
</tr>
<tr>
<td>Mean diastolic blood pressure (mm Hg)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>81.9</td>
<td>83.7</td>
</tr>
<tr>
<td>Mean age at offspring’s birth (years)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>30.9</td>
<td>30.5</td>
</tr>
<tr>
<td>Mean age at BMI measurement (years)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>48.7</td>
<td>48.6</td>
</tr>
<tr>
<td>Measure</td>
<td>Quintiles</td>
<td>Linear regression coefficients, calculated per standard deviation (3.39 kg m⁻²) of paternal BMI, adjusted for age and survey occasion.</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>-----------</td>
<td>-----------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Proportion ever smoked (%)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>67.3</td>
<td>62.2 61.3 62.3 63.6</td>
</tr>
<tr>
<td>Proportion drinking &gt;= 5 times fortnightly (%)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>7.5</td>
<td>8.3 7.6 8.2 8.6</td>
</tr>
<tr>
<td>Proportion educated &gt;=10 years (%)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>50.3</td>
<td>52.4 52.6 49.9 44.5</td>
</tr>
<tr>
<td>Proportion in non-manual employment (%)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>37.9</td>
<td>40.0 38.7 35.9 35.8</td>
</tr>
<tr>
<td>Proportion physically active (%)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>83.6</td>
<td>87.9 87.0 85.7 82.0</td>
</tr>
</tbody>
</table>

<sup>a</sup> Quintiles were calculated among men of similar age, measured at the same survey occasion.

<sup>b</sup> Linear regression coefficients, calculated per standard deviation (3.39 kg m⁻²) of paternal BMI, adjusted for age and survey occasion.

<sup>c</sup> Logistic regression odds ratios, calculated per standard deviation (3.39 kg m⁻²) of paternal BMI, adjusted for age and survey occasion.
Supplementary Figure 1. Parent-offspring associations in BMI

All BMI measurements were converted to sex-specific Z scores adjusted for HUNT survey (categorical, with contemporary HUNT and YoungHUNT combined), a cubic spline of age with five knots at percentiles of 5, 27.5, 50, 72.5, 95, and an interaction between them. The residual standard deviations were 4.33 kg m$^{-2}$ for women and 3.39 kg m$^{-2}$ for men. Offspring BMI Z scores used in the main analysis were divided into 50 quantiles, each plotted at its mean value.
Supplementary Figure 2. Fitted associations of all-cause and cause-specific mortality with own and offspring BMI (kg m\(^{-2}\)).

Hazard ratios were calculated per standard deviation of BMI (4.33 kg m\(^{-2}\) in women and 3.39 kg m\(^{-2}\) in men) and back-converted to the original units (upper x-axis for men, lower x-axis for women). Hazard ratios are relative to a person of mean BMI for their group (fathers, mothers, offspring) and are adjusted for parental age, date of birth, alcohol use, education, employment (own and spouse’s), exercise levels and smoking (own and offspring). BMI was pre-adjusted for age, sex and HUNT survey. Data were restricted to those with valid data on parent and offspring (but not necessarily both parents). Plotted data were truncated at the 1\(^{st}\) and 99\(^{th}\) percentiles of BMI to improve resolution in the main part of the distribution.
Supplementary Figure 2 (continued). Fitted associations of all-cause and cause-specific mortality with own and offspring BMI (kg m$^{-2}$).

Hazard ratios were calculated per standard deviation of BMI (4.33 kg m$^{-2}$ in women and 3.39 kg m$^{-2}$ in men) and back-converted to the original units (upper x-axis for men, lower x-axis for women). Hazard ratios are relative to a person of mean BMI for their group (fathers, mothers, offspring) and are adjusted for parental age, date of birth, alcohol use, education, employment (own and spouse’s), exercise levels and smoking (own and offspring). BMI was pre-adjusted for age, sex and HUNT survey. Data were restricted to those with valid data on parent and offspring (but not necessarily both parents). Plotted data were truncated at the 1st and 99th percentiles of BMI to improve resolution in the main part of the distribution.
Supplementary Figure 2 (continued). Fitted associations of all-cause and cause-specific mortality with own and offspring BMI (kg m$^{-2}$).

Hazard ratios were calculated per standard deviation of BMI (4.33 kg m$^{-2}$ in women and 3.39 kg m$^{-2}$ in men) and back-converted to the original units (upper x-axis for men, lower x-axis for women). Hazard ratios are relative to a person of mean BMI for their group (fathers, mothers, offspring) and are adjusted for parental age, date of birth, alcohol use, education, employment (own and spouse’s), exercise levels and smoking (own and offspring). BMI was pre-adjusted for age, sex and HUNT survey. Data were restricted to those with valid data on parent and offspring (but not necessarily both parents). Plotted data were truncated at the 1st and 99th percentiles of BMI to improve resolution in the main part of the distribution.
Supplementary Figure 3. Bias components from measured covariates with (i) own BMI (empty circles) and (ii) offspring BMI as an instrument for own BMI (filled circles).

The generalized method of moments with robust standard errors and no adjustment was used to estimate the association of each listed covariate with parental BMI with two methods simultaneously: (i) using parental BMI directly (empty circles) and (ii) using offspring BMI as an instrument for parental BMI (filled circles). The bias components are the estimated associations using each method, which are proportional to the bias if a mortality outcome were to be estimated by each method, with the covariate omitted. Participants with missing data for the covariate in question were omitted from these analyses. Some covariates may be mediators, rather than confounders, of the association between exposure and outcome, in which case the components do not represent bias. Bias components are on an arbitrary, relative scale and are comparable between the two methods for each covariate, but not across covariates. They were therefore scaled by the absolute magnitude of the larger of each pair for ease of presentation. Covariates are ordered by the relative bias (absolute bias component using offspring BMI as an instrument / absolute bias component using own BMI) for mothers.
Supplementary Figure 4. Directed acyclic graph illustrating the role of potential confounders and mediators of the association between an exposure (e.g. BMI) and an outcome (e.g. mortality) when an instrument (e.g. offspring BMI) is used

A conventional estimate of \( f \) (the effect of the exposure on the outcome) may be biased due to confounding via pathway \( bc \). An IV estimate of the same effect may be biased by confounding via pathway \( ac \), and any such bias is magnified by the reciprocal of the association \( de + ab \) between instrument and exposure. Bias from measured confounders may be adjusted for, but its relative magnitude in conventional and instrumental variables analyses may represent the likely relative magnitude of bias from unmeasured confounders in the two analytical methods. An unbiased instrumental variables analysis also requires that there be no causal pathway \( g \) from instrument to outcome (except via the exposure) and that the association between instrument and exposure is non-null. In most instrumental variable applications, the instrument is a cause of the exposure. In the present case, a causal effect of offspring BMI on BMI is implausible; we must further assume that there is no causal effect of BMI on offspring BMI and that the common genetic and environmental factors (G/E) causing the instrument and exposure to be associated \((de)\) are distinct from those confounding the exposure and outcome.
Supplementary Methods 1. Details of estimation using instrumental variables

The step-by-step process by which instrumental variable estimates were made is given below. As an example, we estimate a fully adjusted hazard ratio for all-cause mortality per standard deviation of BMI among fathers, using offspring BMI as an instrument. BMI values in fathers and sons were already converted to standard deviation units.

1. Regress paternal BMI against offspring BMI with full adjustment. The regression coefficient is the denominator of the instrumental variable ratio. Retain the regression coefficient ($\beta_D = 0.2221$) and its standard error ($SE_D = 0.0055$).

2. Conduct Cox proportional hazards regression for paternal all-cause mortality against offspring BMI with full regression. The natural logarithm of the hazard ratio is the numerator of the instrumental variable ratio. Retain the natural logarithm of the hazard ratio ($\beta_N = 0.0202$) and its standard error ($SE_N = 0.0110$).

3. Divide the natural logarithm of the hazard ratio from (2) by the regression coefficient from (1) to get the natural logarithm of the instrumental variable hazard ratio ($\beta_{IV} = 0.0911$). Exponentiate this to obtain the instrumental variable hazard ratio (1.0952).

4. Estimate the standard error of the natural logarithm of the instrumental variable hazard ratio using Taylor series expansion:

$$SE_{IV} = \sqrt{SE_N^2 / \beta_D^2 + \beta_N^2 / \beta_D^4 * SE_D^2 - 2*\beta_N / \beta_D^3 * Cov(ND)} = 0.0496$$

where Cov(ND) is the covariance between the numerator and denominator which is assumed to be zero.

5. For comparison with the instrumental variable estimate, use Cox proportional hazards regression for paternal all-cause mortality against paternal BMI with full regression to make a conventional estimate of the hazard ratio (1.0491). Retain the natural logarithm of the hazard ratio ($\beta_{CE} = 0.0479$) and its standard error ($SE_{CE} = 0.0110$).

6. Calculate the Durbin-Wu-Hausman statistic $H = |(\beta_{IV}-\beta_{CE})^2 / (SE_{IV}^2-SE_{CE}^2)| = 0.7950$. Take the corresponding $P$ value from a Chi-squared distribution with one degree of freedom ($P=0.3726$).
Supplementary Methods 2. Stata do-file used to estimate linear associations of mortality with BMI
**********************************************************************************************
* Offspring-as-IV analysis of mortality against BMI in the HUNT data
* David Carslake, July 2017
* Requires HUNT_bmi_data.dta
* Analyses parents separately (B) and together (F)
* Outputs: (i) HUNT_bmi_analysis_`outcome'.log  
* Outputs: (i) HUNT_bmi_analysisB_`outcome'.dta  
* Outputs: (i) HUNT_bmi_analysisF_`outcome'.dta
* Takes about 25 minutes on my desktop.
**********************************************************************************************
* Am I running this on my desktop or on Blue Crystal?
local run_on_BC = "Y"
if "`run_on_BC'"=="Y"{
    args BCoutcome WORKDIR
    local outcome = "BCoutcome"
    local working_dir = "WORKDIR"
    display "Using Blue Crystal; outcome is `BCoutcome', working_dir is `WORKDIR'"
} else{
    local outcome = "allcause"
    local working_dir = "H:\\Davids Bristol Files\\Working_Directory"
    local data_dir = "H:\\Davids Bristol Files\\Datasets\\HUNT\\Processed data\\Offspring_BMI"
    display "Using Desktop; outcome is `BCoutcome', working_dir is `working_dir', data directory is `data_dir'"
    capture log close
    cd ""working_dir"
    log using "HUNT_bmi_analysis_`outcome'.log", replace
}

cd ""working_dir"
display "HUNT_mortality_analysis started at `c(current_date)', `c(current_time)''

*------------
* Settings:
*------------
local pm = 6
local include_B = "Y"  
local include_F = "Y"
local maximum_iterations = 200
* Option: Exclude extreme values of (raw) exposure:
local bmi_min = 0
local bmi_max = 100000
* Option: Exclude the first `lag' years of follow-up:
local lag = 0
*Define the confounder sets:
foreach p in "m" "f" "p"
    local adjlist "a b"
    local p\'_confounders_a = ""
    local p\'_confounders_b = "p\'_bmi_smokerD* p\'_bmi_alcoholD* p\'_bmi_educD* p\'_bmi_workselfD* p\'_bmi_workspsD* p\'_bmi_exerD* b\'_bmi_smokerD"
}

*-------------
*Preliminaries:
*-------------
clear all
set more off
use "`data_dir'/HUNT_bmi_data.dta", clear
cd ""working_dir"
save delme.dta,replace
* Exclude those parents with exposure outside the permitted range:
foreach p in "m" "f"
quietly count if (`p'_bmi_use<`bmi_min' | `p'_bmi_use>`bmi_max') & `p'_useforB==1
display "r(N)" parents `p' dropped from analysis B because their BMI was outside the permitted
range"
replace `p'_useforB = 0 if `p'_bmi_use<`bmi_min' | `p'_bmi_use>`bmi_max'
}
* Adapt the censoring and inclusion variables to exclude the first `lag' years of follow-up:
foreach p in "m" "f"
  replace `p'_startageB = max(`p'_startageB,`p'_bmi_age+`lag') if `p'_useforB == 1
  quietly count if `p'_startageB >= `p'_finalageB & `p'_useforB == 1
display "r(N)" parents `p' dropped from analysis B because they died within `lag' years of
recruitment"
  replace `p'_useforB = 0 if `p'_startageB >= `p'_finalageB & `p'_useforB == 1
}
* Establish Tom's program "tsci":
*-----------------------------------------------
*! 1.0.0 Tom Palmer 8jan2010
capture program drop tsci
program tsci, rclass
syntax anything, [eform]
local n = wordcount("anything")
tokenize "anything"
tempname gd segd gp segp cov ratio seratio z p
sca `gd' = '1'
sca `segd' = '2'
sca `gp' = '3'
sca `segp' = '4'
if ""n"" == ""4"" {
  sca `cov' = 0
}
else {
  sca `cov' = '5'
}
}sca `ratio' = `gd'/`gp'
sca `seratio' = sqrt((`segd'^2/`gp'^2) + (`gd'^2/`gp'^4)*`segp'^2
- 2*(`gd'/`gp'^3)*`cov')
sca `z' = abs(`ratio'/`seratio')
di as res `ratio', `seratio', "Z=" `z', "P=" `p'
if ""eform"" != ""
  di as res exp(`ratio'), "exp(`ratio' + invnormal(0.975)*`seratio')", "exp(`ratio' + invnormal(0.975)*`seratio')")", "P=" `p'
} ret sca ratio = `ratio'
ret sca seratio = `seratio'
end
* Analysis B: Mothers and Fathers separately, offspring and own exposure (including 2SIV)
*----------------------------------------------------------------------------------------------
if ""include_B""==""Y"
  display "Analysis B started at `c(current_date)', `c(current_time)"
  * Set up a place to store the analysis results:
tempname memholdB
  postfile 'memholdB' str30 outcome str10 analysis str10 method str10 parent str10 exposure str10
  adjustment str150 confounders N deaths X power coeff SEcoeff HR LCLHR UCLHR P PHtest using
  "HUNT_bmi_analysisB_`outcome'", replace
  * Do the analyses:
  foreach p in "m" "f"
    if ("p"=="m" & "outcome"=="prostate_cancer")|("p"=="f" & "outcome"=="ovarian_cancer")
      display "Outcome 'outcome' not applicable in parent 'p'"
    else{
      local P = upper("p")
stset `p'_finalageB if `p'_useforB==1, enter(`p'_startageB) failure(`p'_`outcome')
id(`p'_id)
```

generate logt = log(_t)
summarize _t if _d==1

foreach adj of local adjlist{
    * Analysis B00`P'`adj' (Parent exposure-offspring exposure):
        local analysis "B00`P'`adj'"
        local method "LR"
        display _n "Analysis `analysis', parent's bmi-offspring bmi in parents `P':"
        regress `p'_bmi_use_Z`pm' b_bmi_use_Z`pm' `p'_dob_cspl* `p'_confounders_`adj''
        if `p'_useforB==1
        post `memholdB' ("NA") ("`analysis'") ("`method'") ("`P'") ("b_bmi") ("`adj'")
        ("``p'_confounders_`adj''") (e(N)) (. .) (. .) (. [b_bmi_use_Z`pm']) (. .) (. .) (. .)
        (.)
        scalar B00`P'`adj'_coef = _b[ b_bmi_use_Z`pm']
        scalar B00`P'`adj'_se = _se[ b_bmi_use_Z`pm']
        *
        Analysis B01`P'`adj' (Cox regression against offspring BMI):
        local analysis "B01`P'`adj'"
        local method "Cox"
        display _n "Analysis `analysis', `outcome'~offspring bmi in parents `P':"
        stcox b_bmi_use_Z`pm' `p'_dob_cspl* `p'_confounders_`adj'' if `p'_useforB==1
        matrix M = r(table)
        correlate scaled_sch1 logt
        local PHtest1 = 2*ttail(r(N)-2,abs(r(zho))*sqrt(r(N)-2)/sqrt(1-r(zho)^2))
        drop scaled_sch*
        post `memholdB' ("outcome") ("`analysis'") ("`method'") ("`P'") ("b_bmi") ("`adj'")
        ("``p'_confounders_`adj''") (e(N) (. .)) (. .) (. [b_bmi_use_Z`pm']) (. .) (. .) (. .)
        (.)
        (exp([M[1,1]]) (exp([M[6,1]])) (M[1,1]) (exp([M[1,1]]))
       .scalar(B01`P'`adj'_coef) scalar(B01`P'`adj'_se) scalar(B00`P'`adj'_coef)
        scalar(B00`P'`adj'_se), eform
        local z = r(ratio)/r(seratio)
        post `memholdB' ("outcome") ("`analysis'") ("`method'") ("`P'") ("b_bmi") ("`adj'")
        ("``p'_confounders_`adj''") (. .) (. .) (. .) (. [r(ratio)]) (. [seratio]) (. [exp(r(ratio))])
        (. [exp(r(ratio)+invnorm(0.025)*r(seratio))]) (. [exp(r(ratio)+invnorm(0.975)*r(seratio))]) (. [2*min(1-
        normal(`z'),normal(`z'))]) .
        *
        Analysis B02`P'`adj' (Cox regression against offspring BMI with split follow-up and varying HR):
        local analysis "B02`P'`adj'"
        local method "Split Cox"
        display _n "Analysis `analysis', `outcome'~offspring bmi (with split follow-up)" :
        local agebands = "0 60 70 80 90"
        stsplit Ageband, at(`agebands')
        foreach ageband of local agebands{
            generate X_`ageband' = 0
            replace X_`ageband' = b_bmi_use_Z`pm' if Ageband==`ageband'
        }
        stcox X_* `p'_dob_cspl* `p'_confounders_`adj'' if `p'_useforB==1, nohr
        matrix M = r(table)
    }
```
test X_0 = X_60 = X_70 = X_80 = X_90
local PHtest2 = r(p)
local col = 0
foreach ageband of local agebands{
  local col = `col'+1
  assert M[1,`col'] == b[X_`ageband']
  quietly count if `p'_useforB==1 & Ageband==`ageband' & _d==1
  local Ndeaths = r(N)
  quietly count if `p'_useforB==1 & Ageband==`ageband'
  local Ntotal = r(N)
  post `memholdB' ("outcome") ("analysis") ("method") ("p")
  ("b_bmi_`ageband'") ("adj") ("`p'_confounders_adj") ("Ntotal") (Ndeaths) (. .) (M[1,`col'])
  (M[2,`col']) (exp(M[1,`col']) (exp(M[5,`col']) (exp(M[6,`col']) (M[4,`col']) (PHtest2')
}
drop Ageband X_*
stjoin
* Analysis B12.'P'.adj' (Cox regression against own BMI with split follow-up and varying HR):
  local analysis "B12.'P'.adj"
  local method "Split Cox"
  display _n "Analysis `analysis', `outcome'~own bmi (with split follow-up) in parents `P':"
  local agebands = "0 60 70 80 90"
  stsplit Ageband, at(`agebands')
  foreach ageband of local agebands{
    generate X_`ageband' = 0
    replace X_`ageband' = `p'_bmi_use_Z`pm' if Ageband==`ageband'
  }
stcox X_* `p'_dob_cspl* `.`p'_confounders_`adj'" if `p'_useforB==1, nohr
  matrix M = r(table)
test X_0 = X_60 = X_70 = X_80 = X_90
  local PHtest2 = r(p)
  local col = 0
  foreach ageband of local agebands{
    local col = `col'+1
    assert M[1,`col'] == b[X_`ageband']
    quietly count if `p'_useforB==1 & Ageband==`ageband' & _d==1
    local Ndeaths = r(N)
    quietly count if `p'_useforB==1 & Ageband==`ageband'
    local Ntotal = r(N)
    post `memholdB' ("outcome") ("analysis") ("method") ("p")
    ("b_bmi_`ageband'") ("adj") ("`p'_confounders_adj") ("Ntotal") (Ndeaths) (. .) (M[1,`col'])
    (M[2,`col']) (exp(M[1,`col']) (exp(M[5,`col']) (exp(M[6,`col']) (M[4,`col']) (PHtest2')
  }
drop Ageband X_*
stjoin
* Analysis F: Combined analysis of mothers and fathers
  *-----------------------------------------------------
  * Analysis F started at `c(current_date)', `c(current_time)'*
  * Set up a place to store the analysis results:
  *-----------------------------------------------------
  * Differences from analysis B are i) robust SE clustered by b_id ii) p_male as predictor (and interacting
  * with confounders) iii) additional analysis includes interaction p_male:bmi
  *-----------------------------------------------------
  * Differences from analysis B are i) robust SE clustered by b_id ii) p_male as predictor (and interacting
  * with confounders) iii) additional analysis includes interaction p_male:bmi
  local MF_pairs_only = "N"
  if ".include_F":="Y" & "outcome":="breast_cancer" & "outcome":="prostate_cancer" & "outcome":="ovarian_cancer"
  display "Analysis F started at `c(current_date)', `c(current_time)'"
  * Set up a place to store the analysis results:
  *-----------------------------------------------------
save delme.dta, replace
define p in "m" "f"
foreach p in "m" "f"
    use delme.dta, clear
    local other_p = cond("p"=="m","f","m")
    keep if `p'_useforB==1
    if "MF_pairs_only"=="Y"{keep if `other_p'_useforB==1}
generate p_male = cond("p"=="m",0,1)
rename `p'_* p_*
drop `other_p'_*
save `p'_delme.dta,replace
}
use m_delme.dta, clear
append using f_delme.dta
* Re-do the cubic splines for parents' date of birth (so they're for parents overall):
drop p_dob_cspl*
summarize p_dob_use
generate p_dob_demean = p_dob_use-r(mean)
mkspline "p_dob_cspl" = p_dob_demean, cubic nknots(5) displayknots
drop p_dob_demean
stset p_finalageB if p_useforB==1, enter(p_startageB) failure(p_'outcome') id(p_id)
generate logt = log(_t)
summaries _t if _d==1
foreach adj of local adjlist{
    * Analysis F00P`adj' (Parent exposure-offspring exposure):
       local analysis "F00P`adj''
       local method "LR"
       display _n "Analysis `analysis', parent's bmi-offspring bmi in both parents:
       quietly regress p_bmi_use_Z`pm' b_bmi_use_Z`pm' p_dob_cspl* p_male c.b_bmi_use_Z`pm'#c.p_male
       `p_confounders_`adj'' if p_useforB==1,vce(cluster b_id)
       local zi = _b[b_bmi_use_Z`pm'#p_male]/_se[b_bmi_use_Z`pm'#p_male]
       local Pint = 2*min(1-normal(`zi'),normal(`zi'))
       regress p_bmi_use_Z`pm' b_bmi_use_Z`pm' p_dob_cspl* p_male `p_confounders_`adj'' if p_useforB==1,vce(cluster b_id)
       post `memholdF' ("NA") ("analysis") ("method") ("P") ("b_bmi") ("adj")

       ("p_confounders_`adj''") (e(N)) (.) (.) (.) (.) (.) (.) (.) (.) (.) (.)

       ("Pint") (.)
       scalar F00P`adj'_coef = _b[b_bmi_use_Z`pm']
       scalar F00P`adj'_se = _se[b_bmi_use_Z`pm']
       * Analysis F01P`adj' (Cox regression against offspring BMI):
       local analysis "F01P`adj''
       local method "Cox"
       display _n "Analysis `analysis', `outcome'~offspring bmi in both parents:
       quietly stcox b_bmi_use_Z`pm' p_dob_cspl* p_male c.b_bmi_use_Z`pm'#c.p_male
       `p_confounders_`adj'' if p_useforB==1, nohr vce(cluster b_id)
       local zi = _b[b_bmi_use_Z`pm'#p_male]/_se[b_bmi_use_Z`pm'#p_male]
       local Pint = 2*min(1-normal(`zi'),normal(`zi'))
       stcox b_bmi_use_Z`pm' p_dob_cspl* p_male `p_confounders_`adj'' if p_useforB==1, nohr
       vce(cluster b_id)
       matrix M = r(table)
correlate scaled_sch logt
local PHtest1 = 2*ttail(r(N)-2,abs(r(rho))*sqrt(r(N)-2)/sqrt(1-r(rho)^2))
drop scaled_sch*
post `memholdF' ("outcome") ("analysis") ("method") ("P") ("b_bmi") ("adj")
("p_confounders_`adj''") (e(N_sub)) (e(N_fail)) (.) (.) (.) (.) (M[1,1]) (M[2,1]) (exp(M[1,1])) (exp(M[5,1]))
(exp(M[6,1])) (M[4,1]) ("Pint") ("PHtest1")

scalar F01P`adj'_coef = _b[b_bmi_use_Z`pm']
scalar F01P`adj'_se = _se[b_bmi_use_Z`pm']
* Analysis F11P`adj' (Cox regression against own BMI):
local analysis "F11P`adj''
local method "Cox"
display _n "Analysis `analysis', `outcome'~own bmi in both parents:
quietly stcox p_bmi_use_Z`pm' p_dob_cspl* p_male c.p_bmi_use_Z`pm'#c.p_male
`p_confounders_`adj'' if p_useforB==1, nohr vce(cluster b_id)
local zi = _b[p_bmi_use_Z`pm'#p_male]/_se[p_bmi_use_Z`pm'#p_male]
local Pint = 2*min(1-normal(`zi'),normal(`zi'))
* stcox p_bmi_use Z`pm' p_dob_cspl* p_male `p_confounders_`adj'' if p_useforB==1, nohr vce(cluster b_id) scaledsch(scaled_sch*)
  * matrix M = r(table)
  * correlate scaled_sch logt
  * local PHtest1 = 2*ttail(r(N) - 2, abs(r(rho))*sqrt(r(N) - 2)/sqrt(1 - r(rho)^2))
  * drop scaled_sch*
  * post 'memholdF' ("'outcome'") ("'analysis'") ("'method'") ("'P'") ("'p_bmi'") ("'adj'")
  ("'p_confounders_`adj''") (e(N_sub)) (e(N_fail)) (. .) (M[1,1]) (M[2,1]) (exp(M[1,1])) (exp(M[5,1]))
  (exp(M[6,1])) (M[4,1]) (Pint') (PHtest1')

  * Analysis F99P'adj' (2SIV estimate):
  * local analysis "F99P'adj'"
  * local method "2SIV"
  * display _n 'Analysis 'analysis', 'outcome'-own bmi (offspring BMI as IV) in both parents:'
    * tsci scalar(F01P'adj'_coef) scalar(F01P'adj'_se) scalar(F00P'adj'_coef) scalar(F00P'adj'_se),
    * eform
    * local z = r(ratio)/r(seratio)
    * post 'memholdF' ("'outcome'") ("'analysis'") ("'method'") ("'P'") ("'p_bmi'") ("'adj'")
      ("'p_confounders_`adj''") (. .) (. .) (r(ratio)) (r(seratio)) (exp(r(ratio)) + invnorm(0.025)*r(seratio))) (2*min(1-normal('z'),normal('z')))
  * Analysis F02P'adj' (Cox regression against offspring BMI with split follow-up and varying HR):
    * local analysis "F02P'adj'"
    * local method "Split Cox"
    * display _n 'Analysis 'analysis', 'outcome'-offspr

  * Analysis F12P'adj' (Cox regression against own BMI with split follow-up and varying HR):
    * local analysis "F12P'adj'"
    * local method "Split Cox"

  stsplit Ageband, at(\"agebands\")
  foreach ageband of local agebands{
    generate X_`ageband' = 0
    replace X_`ageband' = b_bmi_use Z`pm' if Ageband==\"ageband\"
}

  stcox X_* p_dob_cspl* p_male `p_confounders_`adj'' if p_useforB==1, nohr vce(cluster b_id)
  * matrix M = r(table)
  * test X_0 = X_60 = X_70 = X_80 = X_90
  * local PHtest2 = r(p)
  * local col = 0
  * foreach ageband of local agebands{
    local col = `col'+1
    assert M[1,`col'] == _b[X_`ageband']
    quietly count if p_useforB==1 & Ageband==\"ageband\" & _d==1
    local Ndeaths = r(N)
    quietly count if p_useforB==1 & Ageband==\"ageband\" & _d==0
    local Ntotal = r(N)
    post 'memholdF' ("'outcome'") ("'analysis'") ("'method'") ("'P'") ("'b_bmi_ageband'")
      ("'adj'") ("'p_confounders_`adj''") (Ntotal) (Ndeaths) (. .) (M[1,`col']) (M[2,`col'])
      (exp(M[1,`col']) (exp(M[5,`col']) (exp(M[6,`col']) (exp(M[4,`col']) (PHtest2')
  }
  drop Ageband X_*
  stjoin

local Nfails = _N
local Ndeaths = _N
local Ntotal = _N
local PHtest = 2*ttail(r(N) - 2, abs(r(rho))*sqrt(r(N) - 2)/sqrt(1 - r(rho)^2))


drop scaled_sch*

* Analysis F99P'adj' (2SIV estimate):
local analysis "F99P'adj'"
local method "2SIV"
display _n 'Analysis 'analysis', 'outcome'-own bmi (offspring BMI as IV) in both parents:'
tsci scalar(F01P'adj'_coef) scalar(F01P'adj'_se) scalar(F00P'adj'_coef) scalar(F00P'adj'_se),
eform
local z = r(ratio)/r(seratio)
post 'memholdF' ("'outcome'") ("'analysis'") ("'method'") ("'P'") ("'p_bmi'") ("'adj'")
("'p_confounders_`adj''") (. .) (. .) (r(ratio)) (r(seratio)) (exp(r(ratio)) + invnorm(0.025)*r(seratio))) (2*min(1-normal('z'),normal('z')))
* Analysis F02P'adj' (Cox regression against offspring BMI with split follow-up and varying
HR):
local analysis "F02P'adj'"
local method "Split Cox"
display _n 'Analysis 'analysis', 'outcome'-offspr

* Analysis F12P'adj' (Cox regression against own BMI with split follow-up and varying
HR):
local analysis "F12P'adj'"
local method "Split Cox"
display _n 'Analysis 'analysis', 'outcome'-own bmi (with split follow-up) in both
parents:'
local agebands = "0 60 70 80 90"
stsplit Ageband, at(\"agebands\")
foreach ageband of local agebands{
generate X_`ageband' = 0
replace X_`ageband' = b_bmi_use Z`pm' if Ageband==\"ageband\"
}

stcox X_* p_dob_cspl* p_male `p_confounders_`adj'' if p_useforB==1, nohr vce(cluster b_id)
* matrix M = r(table)
* test X_0 = X_60 = X_70 = X_80 = X_90
* local PHtest2 = r(p)
* local col = 0
* foreach ageband of local agebands{
  local col = `col'+1
  assert M[1,`col'] == _b[X_`ageband']
  quietly count if p_useforB==1 & Ageband==\"ageband\" & _d==1
  local Ndeaths = r(N)
  quietly count if p_useforB==1 & Ageband==\"ageband\" & _d==0
  local Ntotal = r(N)
  post 'memholdF' ("'outcome'") ("'analysis'") ("'method'") ("'P'") ("'b_bmi_ageband'")
  ("'adj'") ("'p_confounders_`adj''") (Ntotal) (Ndeaths) (. .) (M[1,`col']) (M[2,`col'])
  (exp(M[1,`col']) (exp(M[5,`col']) (exp(M[6,`col']) (exp(M[4,`col']) (PHtest2')
}
drop Ageband X_*
stjoin
quietly count if p_useforB==1 & Ageband==`ageband' & _d==1
local Ndeaths = r(N)
quietly count if p_useforB==1 & Ageband==`ageband'
local Ntotal = r(N)
post `memholdF' ("outcome") ("analysis") ("method") ("P") ("p_bmi_`ageband'")
("adj") ("p_confounders_`adj''") ("Ntotal") ("Ndeaths") (.) (.). (M[1,"col'"]) (M[2,"col'"])
(exp(M[1,"col'"]) (exp(M[5,"col'"]) (exp(M[6,"col'"]) (M[4,"col'"]) (.) (PHtest2'))
}
drop Ageband X_*
stjoin
}
drop _st _d _t _t0
postclose `memholdF'
use delme.dta, clear
erase m_delme.dta
erase f_delme.dta
}

*--------------
* Wrap it up:
*--------------
clear all
erase delme.dta
display "HUNT_mortality_analysis ended at `c(current_date)', `c(current_time)"
capture log close
Supplementary Methods 3. Stata do-file used to estimate and plot cubic spline associations of mortality with BMI

* Makes Figure 2 (selected causes) and Figure S2 (all causes) for the HUNT paper
* Runs cubic spline Cox models and plots mortality against own & offspring BMI
* Has twin X-axes (upper for men, lower for women)
* Vertical lines show the 1st & 99th %iles of exposure (unless truncated)
* Takes about 10 minutes on my desktop
* David Carslake, June 2017

*----------
* Settings:
*----------
local data_dir = "H:\Davids Bristol Files\Datasets\HUNT\Processed data\Offspring_BMI"
local working_dir = "H:\Davids Bristol Files\Working_Directory"
local m_confounders = "m\_bmi\_smokerD* m\_bmi\_alcoholD* m\_bmi\_educD* m\_bmi\_workselfD* m\_bmi\_workspSD* m\_bmi\_exerD* b\_bmi\_smokerD*"
local f_confounders = "f\_bmi\_smokerD* f\_bmi\_alcoholD* f\_bmi\_educD* f\_bmi\_workselfD* f\_bmi\_workspSD* f\_bmi\_exerD* b\_bmi\_smokerD*"
local outcomes = "allcause cardiovascular chd stroke diabetes respiratory cancer lung_cancer breast_cancer prostate_cancer colorectal_cancer pancreatic_cancer stomach_cancer ovarian_cancer external"
local m_only_outcomes = "breast_cancer ovarian_cancer"
local f_only_outcomes = "prostate_cancer"
local plotpoints = 100
* Define the pre-adjustment of BMI (normally 6; HUNT survey * age)
local pm = 6
* Allow a lag after participation (normally set to zero):
local lag = 0
* Allow truncation of the X-axis beyond percentiles (1 & 100 is no truncation):
local X_pctile_min_plotted = 2
local X_pctile_max_plotted = 99
* Allow corresponding plots of own and offspring BMI to have a common Y scale:
local common_Y = ”Y”

*----------
* Preliminaries:
*----------
clear all
cd ""working_dir"
capture log close
log using "HUNT_bmi_plots.log", replace
display c(current_date),”, “c(current_time)
set more off
use ""data_dir\HUNT_bmi_data.dta", clear
save delme.dta,replace
set graphics off
net sj 11-3 st0215_1
net install st0215_1

*----------
* Adjust follow-up and inclusion variables for any lag:
*----------
foreach p in "m" "f"
    generate `p'\_startageE = max('p'\_startageB,`p'\_bmi\_age+`lag')
genenerate `p'\_finalageE = `p'\_finalageB
generate `p'\_useforE = `p'\_useforB
count if `p'\_startageE >= `p'\_finalageE & `p'\_useforE == 1
    display "r(N)" parents `p' dropped because they died within `lag' years of recruitment"
    replace `p'\_useforE = 0 if `p'\_startageE >= `p'\_finalageE & `p'\_useforE == 1
}

*--------------------------*
* Determine the BMI values to be plotted:
*--------------------------*
foreach p in "m" "f"
* Plotted values of parent's own BMI:
  generate `p'_plotme_`p' = 0
  xtile Percentile = `p'_bmi_use_Z`pm' if `p'_useforE==1, nq(100)
  summarize `p'_bmi_use_Z`pm' if `p'_useforE==1 & Percentile>`X_pctile_min_plotted' & Percentile<=`X_pctile_max_plotted'
  local Xmin = r(min)
  local Xmax = r(max)
  generate Band = ceil((`plotpoints' - 2)*(`p'_bmi_use_Z`pm' - `Xmin')/(`Xmax' - `Xmin')) if `p'_useforE==1 & Percentile>`X_pctile_min_plotted' & Percentile<=`X_pctile_max_plotted'
  bysort `p'_useforE Band (`p'_bmi_use_Z`pm'): replace `p'_plotme_`p' = 1 if _n==1 & `p'_useforE==1 & Band<.
  bysort `p'_useforE Band (`p'_bmi_use_Z`pm'): replace `p'_plotme_`p' = 1 if _n==_N & `p'_useforE==1 & Band==(`plotpoints' - 2)
  drop Percentile Band
* Plotted values of offspring BMI:
  generate `p'_plotme_b = 0
  xtile Percentile = b_bmi_use_Z`pm' if `p'_useforE==1, nq(100)
  summarize b_bmi_use_Z`pm' if `p'_useforE==1 & Percentile>`X_pctile_min_plotted' & Percentile<=`X_pctile_max_plotted'
  local Xmin = r(min)
  local Xmax = r(max)
  generate Band = ceil((`plotpoints' - 2)*(b_bmi_use_Z`pm' - `Xmin')/(`Xmax' - `Xmin')) if `p'_useforE==1 & Percentile>`X_pctile_min_plotted' & Percentile<=`X_pctile_max_plotted'
  bysort `p'_useforE Band (b_bmi_use_Z`pm'): replace `p'_plotme_b = 1 if _n==1 & `p'_useforE==1 & Band<.
  bysort `p'_useforE Band (b_bmi_use_Z`pm'): replace `p'_plotme_b = 1 if _n==_N & `p'_useforE==1 & Band==(`plotpoints' - 2)
  drop Percentile Band
}

* Do the analyses:

foreach p in "m" "f"{ local otherp = cond("p"=="m","f","m")
  local o = 0
  * Make cubic splines for parents' & offspring exposures (at Harrell's percentiles):
    mkspline "`p'_bmicspl" = `p'_bmi_use_Z`pm' if `p'_useforE==1, cubic nknots(5) displayknots
    mkspline "b_bmicspl" = b_bmi_use_Z`pm' if `p'_useforE==1, cubic nknots(5) displayknots
  foreach outcome of local outcomes{
    local o = `o'+1
    * Certain outcomes in certain parents shouldn't be analysed or plotted:
      if regexm(" `otherp'_only_outcomes' ",`outcome')=1 display "Not enough deaths from `outcome' in parent `p'"
      else{
        * Define a survival data set of (respectively) `p'_finalageE(_t), `p'_startageE(t0), `p'_outcome(_d) and survival time(_st):
          stset `p'_finalageE if `p'_useforE==1, enter(`p'_startageE) failure(`p'_`outcome') id(`p'_id)
        id(`p'_id)
          summarize _t if _d==1
          * Do the analysis for offspring exposure (rounding avoids problems with xblc):
            generate XR = round(b_bmi_use_Z`pm',0.001)
            display _n "Doing outcome `outcome' against offspring BMI in parent `p':"
            stcox b_bmicspl* `p'_dob_cspl* `p'_confounders' if `p'_useforE==1
            levelsof XR if `p'_plotme_b==1, local(levels)
            xblc b_bmicspl*, covname(XR) at(levels) eform reference(0)
            generate(plotX_`p'b`o' plotY_`p'b`o' LCL_`p'b`o' UCL_`p'b`o')
          drop XR
          * Do the analysis for own exposure (rounding avoids problems with xblc):
            generate XR = round(`p'_bmi_use_Z`pm',0.001)
            display _n "Doing outcome `outcome' against own BMI in parent `p':"
            stcox `p'_bmicspl* `p'_dob_cspl* `p'_confounders' if `p'_useforE==1
            levelsof XR if `p'_plotme_`p'==1, local(levels)
            xblc `p'_bmicspl*, covname(XR) at(levels) eform reference(0)
            generate(plotX_`p'`p'`o' plotY_`p'`p'`o' LCL_`p'`p'`o' UCL_`p'`p'`o')
          drop XR
          drop _st _d _t _t0
        }
      }
    }
  }
}

23
drop `p'_bmicspl* b_bmicspl*

* Confirm plotX is the same for all exposures within a parent and own/offspring:

foreach p in "m" "f"
    local otherp = cond("p"=="m","f","m")
    foreach g in "b" "p"
        local o = 0
        foreach outcome of local outcomes
            local o = `o'+1
            if regexm(" `otherp'_only_outcomes' "," `outcome' ")!=1{
                assert plotX_`p'`g'`o'==plotX_`p'`g'1
            }
        }
    }

* Convert plotX back to original units:

foreach sex in 0 1{
    generate X_`p'b`sex' = .
    summarize b_bmi_standard if b_male==`sex'
    assert r(sd)==0
    local Mean`p'b`sex' = r(mean)
    summarize b_bmi_SD_Z`pm' if b_male==`sex'
    assert r(sd)==0
    local SD`p'b`sex' = r(mean)
    replace X_`p'b`sex' = `Mean`p'b`sex'' + plotX_`p'b1*`SD`p'b`sex''
    quietly summarize plotX_`p'b1
    local Xmin_`p'b`sex'_Z = r(min)
    local Xmax_`p'b`sex'_Z = r(max)
}

* Parental BMI:
    generate X_`p'`p' = .
    summarize `p'_bmi_standard
    assert r(sd)==0
    local Mean`p'`p' = r(mean)
    summarize `p'_bmi_SD_Z`pm'
    assert r(sd)==0
    local SD`p'`p' = r(mean)
    replace X_`p'`p' =  `Mean`p'`p'' + plotX_`p'`p'1*`SD`p'`p''
    quietly summarize plotX_`p'`p'1
    local Xmin_`p'`p'_Z = r(min)
    local Xmax_`p'`p'_Z = r(max)

* Convert some percentiles back to (female scale) BMI:
    summarize b_bmi_use_Z`pm' if `p'_useforE==1, detail
    local X_LC_`p'b = `Mean`p'b0''+r(p1)*`SD`p'b0''
    local X_UC_`p'b = `Mean`p'b0''+r(p99)*`SD`p'b0''
    summarize `p'_bmi_use_Z`pm' if `p'_useforE==1, detail
    local X_LC_`p'`p' = `Mean`p'`p''+r(p1)*`SD`p'`p''
    local X_UC_`p'`p' = `Mean`p'`p''+r(p99)*`SD`p'`p''

* The range for all X-axes must be the overall min and max plotted Z score BMI:
    local Xmin_Z = min(`Xmin_mb0_Z', `Xmin_mb1_Z', `Xmin_fb0_Z', `Xmin_fb1_Z', `Xmin_mm_Z', `Xmin_ff_Z')
    local Xmax_Z = max(`Xmax_mb0_Z', `Xmax_mb1_Z', `Xmax_fb0_Z', `Xmax_fb1_Z', `Xmax_mm_Z', `Xmax_ff_Z')

* This is then converted into the appropriate units for each axis:
    local Xmin_mb0 = 'Meanmb0' + 'Xmin_Z'*'SDmb0'
    local Xmin_mb1 = 'Meanmb1' + 'Xmin_Z'*'SDmb1'
    local Xmin_fb0 = 'Meanfb0' + 'Xmin_Z'*'SDFb0'
    local Xmin_fb1 = 'Meanfb1' + 'Xmin_Z'*'SDFb1'
local Xmin_mm = `Meanmm' + `Xmin_Z'*`SDmm'
local Xmin_ff = `Meanff' + `Xmin_Z'*`SDff'
local Xmax_mb0 = `Meanmb0' + `Xmax_Z'*`SDmb0'
local Xmax_mb1 = `Meanmb1' + `Xmax_Z'*`SDmb1'
local Xmax_fb0 = `Meanfb0' + `Xmax_Z'*`SDfb0'
local Xmax_fb1 = `Meanfb1' + `Xmax_Z'*`SDfb1'
local Xmax_mm = `Meanmm' + `Xmax_Z'*`SDmm'
local Xmax_ff = `Meanff' + `Xmax_Z'*`SDff'

* Make paired plots of offspring & own BMI:
foreach p in "m" "f"
local otherp = cond("p'=="m","f","m")
local posh_p = cond("p'=="m","mother","father")
local o = 0
foreach outcome of local outcomes{
local o = `o'+1
local posh_outcome = ""
local posh_outcome = cond("outcome'=="allcause","All-cause","posh_outcome'")
local posh_outcome = cond("outcome'=="cardiovascular","Cardiovascular
disease","posh_outcome'")
local posh_outcome = cond("outcome'=="chd","Coronary heart disease","posh_outcome'")
local posh_outcome = cond("outcome'=="diabetes","Diabetes","posh_outcome'")
local posh_outcome = cond("outcome'=="respiratory","Respiratory disease","posh_outcome'")
local posh_outcome = cond("outcome'=="cancer","Cancer","posh_outcome'")
local posh_outcome = cond("outcome'=="lung_cancer","Lung cancer","posh_outcome'")
local posh_outcome = cond("outcome'=="breast_cancer","Breast cancer","posh_outcome'")
local posh_outcome = cond("outcome'=="prostate_cancer","Prostate cancer","posh_outcome'")
local posh_outcome = cond("outcome'=="colorectal_cancer","Colorectal
cancer","posh_outcome'")
local posh_outcome = cond("outcome'=="pancreatic_cancer","Pancreatic
cancer","posh_outcome'")
local posh_outcome = cond("outcome'=="stomach_cancer","Stomach cancer","posh_outcome'")
local posh_outcome = cond("outcome'=="ovarian_cancer","Ovarian cancer","posh_outcome'")
local posh_outcome = cond("outcome'=="external","External causes","posh_outcome'")
* Certain outcomes in certain parents shouldn't be analysed or plotted:
if regexm("otherp'_only_outcomes'","outcome'")==1 display "Not enough deaths from
'outcome' in parent 'p'"
else{
display 'Plotting 'outcome' in parent 'p'"
* Determine the scale on the Y axis:
quietly summarize LCL_`p'b`o'
local ymin_b = r(min)
quietly summarize UCL_`p'b`o'
local ymax_b = r(max)
quietly summarize LCL_`p'b`o'
local ymin_`p' = r(min)
quietly summarize UCL_`p'b`o'
local ymax_`p' = r(max)
foreach g in "b" "p"{
if "common Y'"=="Y"
local ymin_g' = min('ymin_b', 'ymin_p')
local ymax_g' = max('ymax_b', 'ymax_p')
}
local yrange_g' = "'ymin_g' - 'ymax_g'
local ytick_g' = "1 1.5"
if 'ymax_g'>2 local ytick_g = "1 2"
if 'ymax_g'>3 local ytick_g = "1 2 3"
if 'ymax_g'>5 local ytick_g = "1 3 5"
if 'ymax_g'>10 local ytick_g = "1 5 10"
if 'ymax_g'>20 local ytick_g = "1 5 20"
if 'ymax_g'>50 local ytick_g = "1 5 50"
if 'ymax_g'>100 local ytick_g = "1 10 100"
if 'ymax_g'>200 local ytick_g = "1 10 200"
if 'ymax_g'>500 local ytick_g = "1 20 500"
if 'ymax_g'>1000 local ytick_g = "1 1000"
if `ymax_`g''>5000 local ytix_`g' = "1 5000"
if `ymax_`g''>10000 local ytix_`g' = "1 10000"
if `ymax_`g''>50000 local ytix_`g' = "1 50000"
if (`ymin_`g''<0.5 & `ymax_`g''<20) local ytix_`g' = "0.5 `ytix_`g''"
}

* Plot outcome against offspring BMI (use the female axis for CI and male axis for HR):
local Xrange0 = "`Xmin_`p'b0' `Xmax_`p'b0'"
local Xrange1 = "`Xmin_`p'b1' `Xmax_`p'b1'"
local X1_options = "'xtitle("") xscale(range(`Xrange0')) xlabel("20(5)35", labgap(*0.5))"
local X2_options = "'xtitle("", axis(2)) xscale(range(`Xrange1')) xlabel("20(5)35", labgap(*0.5)) axis(2)"
local Y_options = "ytitle() yscale(log range(`yrange_b')) ylabel(`ytix_b', glcolor(gs12) glwidth(medium))"
local other_options = "'legend(off) title("") name('p'b'o',replace) scheme(s2mono) graphregion(color(none))"
local xline_options = ""
if `X_pctile_min_plotted'==1 & `X_pctile_max_plotted'==100 local xline_options = "xline(`X_LC_`p'b0' `X_UC_`p'b0')"

graph twoway (rarea LCL_`p'b0' UCL_`p'b0' X_`p'b0,bcolor(gs8) lwidth(none) xaxis(1)) (line plotY_`p'b0' X_`p'b1,lpattern(solid) lcolor(gs4) xaxis(2)), `X1_options' `X2_options' `Y_options' `xline_options' `other_options'  

* Plot outcome against own BMI (blank out the axis for the "other" sex):
local toptix_color = cond("`p'"=="m","none","black")
local bottix_color = cond("`p'"=="m","black","none")
local toplab_color = cond("`p'"=="m","white","black")
local botlab_color = cond("`p'"=="m","black","white")
local Xrange0 = "`Xmin_`p'`p'' `Xmax_`p'`p''"
local Xrange1 = "`Xmin_`p'`p'' `Xmax_`p'`p''"
local X1_options = "'xtitle("") xscale(range(`Xrange0')) xlabel("20(5)35", labgap(*0.5) tlcolor(`bottix_color') labcolor(`botlab_color'))"
local X2_options = "'xtitle("", axis(2)) xscale(range(`Xrange1')) xlabel("20(5)35", labgap(*0.5) tlcolor(`toptix_color') labcolor(`toplab_color') axis(2))"
local Y_options = "ytitle() yscale(log range(`yrange_`p'')) ylabel(`ytix_`p'`, glcolor(gs12) glwidth(medium))"
local other_options = "'legend(off) title("") name('p`p'`o',replace) scheme(s2mono) graphregion(color(none))"
local xline_options = ""
if `X_pctile_min_plotted'==1 & `X_pctile_max_plotted'==100 local xline_options = "xline(`X_LC_`p'`p'' `X_UC_`p'`p''")"

graph twoway (rarea LCL_`p'`p'' UCL_`p'`p'' X_`p'`p',bcolor(gs8) lwidth(none) xaxis(1)) (line plotY_`p'`p'' X_`p'`p',lpattern(solid) lcolor(gs4) xaxis(2)), `X1_options' `X2_options' `Y_options' `xline_options' `other_options' 

* Combine the plots for own and offspring BMI:
local plotlist = "`p'`p'`o' `p'b`o'"
local xline_options = ""

graph combine plotlist, cols(2) ysize(1) xsize(3) name('p`p'`o',replace) holes() scheme(smono) ltitle("") btitle("") iscale(1) title("'posh_outcome', 'posh_p's",position(11) ring(1) size(large)) imargin(tiny) 


* Combine all causes of death into manuscript figures:
*-----------------------------------------------------
*
* For the main figure (selected outcomes):
local plotlist1 = "m1 f1 m2 f2 m4 f4 m6 f6 m15 f15 m7 f7"
graph combine plotlist1, cols(2) ysize(9) xsize(6.5) name(Figure_2, replace) holes() scheme(smono) ltitle("Mortality hazard ratio (95% CI)") btitle(" Own BMI Offspring BMI") iscale(0.5) imargin(vsmall) b2title("Exposure (kg m{superscript:-2})") size(medium) graph save Figure_2 Figure_2, replace 

* Make multiplots for the appendix figure with all outcomes:
local plotlist1 "m1 f1 m2 f2 m3 f3 m4 f4 m5 f5"
graph combine `plotlist1', cols(2) ysize(8) xsize(6.5) name(Figure_S2a, replace) holes() scheme(s1mono)
  l1title("Mortality hazard ratio (95% CI)", size(medium)) b1title("Own BMI                Offspring BMI"
  Own BMI                Offspring BMI", size(small)) b2title("Exposure (kg m\(^{-2}\))", size(medium))
  iscale(0.5) imargin(vsmall)
graph save Figure_S2a Figure_S2a, replace

local plotlist2 "m6 f6 m15 f15 m7 f7 m8 f8 m9"
graph combine `plotlist2', cols(2) ysize(8) xsize(6.5) name(Figure_S2b, replace) holes(10) scheme(s1mono)
  l1title("Mortality hazard ratio (95% CI)", size(medium)) b1title("Own BMI                Offspring BMI"
  Own BMI                Offspring BMI", size(small)) b2title("Exposure (kg m\(^{-2}\))", size(medium))
  iscale(0.5) imargin(vsmall)
graph save Figure_S2b Figure_S2b, replace

local plotlist3 "f10 m11 f11 m12 f12 m13 f13 m14"
graph combine `plotlist3', cols(2) ysize(8) xsize(6.5) name(Figure_S2c, replace) holes(1 10) scheme(s1mono)
  l1title("Mortality hazard ratio (95% CI)", size(medium)) b1title("Own BMI                Offspring BMI"
  Own BMI                Offspring BMI", size(small)) b2title("Exposure (kg m\(^{-2}\))", size(medium))
  iscale(0.5) imargin(vsmall)
graph save Figure_S2c Figure_S2c, replace

*-----------------------------
* Export the graphs as 300 dpi tiffs:
*-----------------------------
local raw_width = 215
local filenames = "Figure_2 Figure_S2a Figure_S2b Figure_S2c"
local new_width = ceil(`raw_width'*300/72)
display `new_width'
foreach name of local filenames{
  graph use `name'.gph
  graph export `name'.tif, width(`new_width') replace
}
* You can then open them in gimp and save as tif with LZW compression to get the size down.

*-----------------------------
* Finish it off:
*-----------------------------
clear all
erase delme.dta
display c(current_date) "", c(current_time)
log close