Absolute Mortality Risk as a Function of BMI

Methodology background note for the Healthy Skepticism course

Overview

This note documents how the absolute mortality risk curves in the BMI module are built, and why each step holds up. The curves combine three published sources, each doing one job. A relative risk model supplies the shape of the mortality response to BMI, a life table supplies the absolute level, and a population distribution supplies context. Nothing in the figures is fitted to a private dataset, and every quantity that is illustrative rather than measured is labeled as such.

The problem: shape without level

Relative risk carries a built-in reference point. When the healthy band is assigned a hazard ratio of one, every other BMI is expressed as a multiple of it, and the vertical position of the curve is fixed by construction. Absolute risk has no such free reference. The probability that a given person dies within ten years is governed mainly by age and by competing causes, not by BMI, so a curve of absolute risk against BMI has to obtain its height from outside the BMI signal itself. If that height is chosen by hand, the result is a relative risk curve wearing absolute units, since the shape comes from the hazard ratio and the level is an assumption. The method below removes the assumption by reading the level from a measured mortality rate.

Three sources, three jobs

The category thresholds come from the World Health Organization classification of adult BMI.[1] They mark the familiar bands: underweight below 18.5, normal weight from 18.5 to 25, overweight from 25 to 30, and obesity grades at 30, 35, and 40. In this module the thresholds only shade and label the figures. They do not enter the risk calculation, which treats BMI as continuous, because the bands are a lossy summary of a smooth signal rather than biological boundaries.

The mortality response comes from the Global BMI Mortality Collaboration, a pooled analysis of 239 prospective studies covering about four million never-smokers who were free of chronic disease at entry and who survived the first five years of follow-up.[2] Restricting to that group is the study's way of limiting reverse causation and confounding by smoking. Relative to the reference band of 22.5 to 25, the all-cause mortality hazard ratios rise on both sides of the healthy range, reaching 1.51 at the low end and 2.76 in the highest obesity grade. The lowest mortality sitting near 22.5 to 25 is consistent with earlier pooled work.[3]

---

[1] World Health Organization. Obesity: preventing and managing the global epidemic. WHO Technical Report Series 894. Geneva: WHO; 2000. Adult cut points: underweight below 18.5, normal 18.5 to 24.9, overweight 25.0 to 29.9, obesity at or above 30, grades 1 to 3 at 30, 35, and 40 kg/m2.

[2] Global BMI Mortality Collaboration (Di Angelantonio E, Bhupathiraju SN, Wormser D, et al.). Body-mass index and all-cause mortality: individual-participant-data meta-analysis of 239 prospective studies in four continents. Lancet. 2016;388(10046):776 to 786. Hazard ratios relative to the 22.5 to under 25.0 band, never-smokers without chronic disease, first 5 years of follow-up excluded: 15.0 to 18.5, 1.51; 18.5 to 20.0, 1.13; 20.0 to 22.5, 1.00; 22.5 to 25.0, 1.00 (reference); 25.0 to 27.5, 1.07; 27.5 to 30.0, 1.20; 30.0 to 35.0, 1.45; 35.0 to 40.0, 1.94; 40.0 to 60.0, 2.76.

[3] Prospective Studies Collaboration. Body-mass index and cause-specific mortality in 900,000 adults: collaborative analyses of 57 prospective studies. Lancet. 2009;373(9669):1083 to 1096. All-cause mortality was lowest at about 22.5 to 25 kg/m2, with each 5 kg/m2 higher BMI associated with roughly 30 percent higher mortality above that range.

The absolute level comes from the Social Security Administration period life table for 2019.[4] Ten-year all-cause mortality at a given age is one minus the product of the single-year survival probabilities across the decade. Averaged across the two sexes, these rates are about 5.8 percent at age 50, 12.0 percent at age 60, and 25.6 percent at age 70.

The population distribution

The density drawn behind the population figure is a teaching model, a right-skewed skew-normal with a mean near 29 and a standard deviation near 7, not fitted to a named dataset. It is used for context and for the optional overlay of your own readings, and it also supplies the weighting logic referred to below when the anchor is discussed. It is labeled illustrative wherever it appears.

The U-shaped relative-risk model

The nine band hazard ratios are placed at their band midpoints and interpolated in log-hazard space, which preserves the multiplicative meaning of a hazard ratio and avoids kinks in the rate of change. The open topmost band, 40 to 60, is anchored at 45 rather than at its midpoint, so the curve is not extended past where the data mass sits. Because the two central bands both carry a hazard ratio of one, the interpolated curve has a flat healthy nadir near 22 to 24 and turns upward in both directions. This U shape is what distinguishes BMI from monotone measures such as blood pressure or LDL, where a single exponential term is enough. Writing the anchors as pairs of BMI and log hazard, the model is

$$HR(x) = \exp(\text{ linear interpolation of } \log HR^i \text{ evaluated at } x\ ).$$

From ratio to probability

Under the proportional-hazards assumption a hazard ratio scales the cumulative hazard, so a baseline ten-year risk becomes, at another BMI,

$$risk(x) = 1 - (1 - p^0)^{HR(x)}.$$

This stays within zero and one, reduces to the baseline times the hazard ratio when the baseline is small (the rare-event regime, where the risk ratio and the hazard ratio agree), and attenuates rather than exceeding one when the baseline is large. It is the standard survival identity, the survivor function raised to the hazard ratio, applied at the ten-year horizon.

Anchoring to a measured rate

The baseline is where the absolute level enters, and it must be read from data rather than assumed. The life-table rate is a population average across all body sizes for a given age, so it applies at the average body size. Setting the exponent to one at the average adult BMI, which recent NHANES data place near 29 to 30,[5] attaches the measured rate at the place it describes:

$$risk(x) = 1 - (1 - m)^{HR(x)\,/\,HR(29)}.$$

[4] Social Security Administration, Office of the Chief Actuary. Period Life Table, 2019, as used in the 2021 Trustees Report. Ten-year all-cause mortality is one minus the product of single-year survival probabilities over the decade. Sex-averaged rates used here: age 50, 5.8 percent; age 60, 12.0 percent; age 70, 25.6 percent.

[5] National Health and Nutrition Examination Survey (NCHS). Age-adjusted mean adult BMI rose from about 28.7 kg/m2 in 2011 to 2012 to about 29.8 kg/m2 in 2017 to 2018. The anchor uses 29 as a round representative value.

By construction the risk at BMI 29 equals the measured rate exactly, the curve falls below it on the healthy plateau, and it rises above it toward both tails. Each age produces a separate curve because each supplies a different measured rate, not a different shape. This is the step that makes the figures absolute. Remove the anchor and the curve floats to whatever level is imposed, which is the failure the method is designed to prevent.

The uncertainty band

The shaded band is plus or minus two standard deviations of an illustrative spread, taken as a lognormal factor on the hazard with a log-scale sigma of 0.35, then carried through the same transform. It represents individual variation around the mean response, not the sampling uncertainty of the published hazard ratios, which is far tighter. The highest obesity grade, for instance, has a 95 percent interval of roughly 2.60 to 2.92 in the same analysis. Because a symmetric spread in log hazard becomes asymmetric once it passes through the probability transform, the band is wider above the mean than below, which is the honest shape for a bounded quantity. It is labeled illustrative in the figures for exactly this reason.

Honest seams

Several assumptions deserve to be stated plainly, along with the direction in which each is likely to mislead.

The first is a transfer across cohorts. The hazard ratio is measured in never-smokers without chronic disease, while the anchor rate is a general-population figure that includes smokers and the already ill. Applying the never-smoker ratio to the general rate assumes the shape of the BMI response is similar in the two groups. That is reasonable for teaching but is the softest joint in the construction, and it most likely understates risk at the low-BMI end, where smoking and illness concentrate.

The second is a convexity gap. The anchor is attached at the mean BMI, yet the average of the risk over the BMI distribution is slightly higher than the risk at the mean BMI, because the curve bends upward and the distribution has a long right tail. This is an instance of Jensen's inequality. The curves therefore understate the true population average by a small amount, on the order of half a percentage point at the ages shown. A fully consistent alternative solves for the nadir so that the distribution-weighted average of the curve matches the life-table rate. That version lifts every curve slightly and is available if exactness is preferred over the simpler anchored form.

The third concerns the low-BMI arm. The upturn below about 20 is the part of the curve the source authors themselves treat with the most caution, since low body weight is often a consequence of illness rather than a cause of death.[6] Even after the never-smoker and five-year restrictions, residual reverse causation is likely, so the left arm should be read as an association whose causal direction is unsettled, not as evidence that leanness is itself hazardous.

The fourth concerns BMI as a measure. It does not separate fat from muscle, and it is often derived from self-reported height and weight, which biases category assignment.[7] Both limitations argue for treating

---

[6]Bhaskaran K, Dos-Santos-Silva I, Leon DA, Douglas IJ, Smeeth L. Association of BMI with overall and cause-specific mortality: a population-based cohort study of 3.6 million adults in the UK. Lancet Diabetes Endocrinol. 2018;6(12):944 to 953. Confirms the J-shaped association and discusses residual reverse causation at low BMI.

[7]Flegal KM, Ioannidis JPA, Doehner W. Flawed methods and inappropriate conclusions for health policy on overweight and obesity: the Global BMI Mortality Collaboration meta-analysis. J Cachexia Sarcopenia Muscle. 2019;10(1):9 to 13. Notes systematic bias from self-reported height and weight and consequent category misclassification.

the curves as descriptions of a crude population signal rather than as individual predictions, which is the same caution the course applies to every clinical band.

What the curves are for

Placed side by side on a shared vertical scale, the three age curves make the course's recurring point visible in one image. The gap between the age panels is far larger than the bend within any single panel, so age, the baseline, dominates the absolute risk while BMI modulates it. A person of seventy at the healthy nadir carries a higher ten-year mortality than a person of fifty at severe obesity. The relative risk of a high BMI can be genuinely large while its absolute consequence at a young age stays modest, and the same hazard ratio opens a much wider absolute gap once the baseline is high. That relationship between relative and absolute risk is what the module exists to teach, expressed here in the one common measure whose risk curve happens to be U-shaped.