Appendix 2. The Bayesian hierarchical model for country-level and regional estimates.

To account for missing data and estimate mean intake of selected dietary fats and dietary cholesterol, we developed and used a Bayesian hierarchical imputation model that accounted for both country- and region-level data and multiple levels of missingness. This model has been described in detail.[1] The model generated estimates for each sex-age-country-time period unit; the estimates were informed by data from that unit itself, if available, and by data from other units. Surveys carried out between 1980 to 1997 were used to inform the 1990 time period, and surveys carried out between 1997 to 2010 the 2010 periods. The effects of changes over time were taken into account through the country-specific covariates specified in the models.

Data from individual countries were used simultaneously as inputs to country and regional estimates. Dietary risk factor levels and trends over time in individual countries were nested within regional levels, which were then nested within global levels. This structure allows the model to borrow information across countries and regions as necessary depending on the extent of data that are missing or less informative (e.g., have large uncertainty). The larger the uncertainty, the greater the extent to which the model borrows information. Time-varying country-level covariates were used to inform the estimates, including lagged distributed income(LDI) [2] and relevant covariates derived and standardized from United Nations Food and Agricultural Organization food disappearance balance sheets, [3] including 17 nutrients or food groups and 4 factors derived from principal components analysis of these 17 variables. LDI per capita is gross domestic product (GDP) per capita smoothed with a 10-year weighted moving average. We use LDI instead of straight GDP because it is more predictive of changes in health. Although GDP can experience drastic changes from year to year, changes in health and related outcomes are slower to change, meaning that the smoother measure of LDI more accurately predicts changes in health. The sources of GDP used are the Penn World Tables, The Maddison Project, the UN Statistics Agency, the World Bank, and the IMF.

The equation used to generate LDI: LDI_y = (GDP per $capita_y$ + 0.9(GDP per $capita_{y-1}$) + 0.8(GDP per $capita_{y-2}$) + 0.7(GDP per $capita_{y-3}$) + 0.6(GDP per $capita_{y-4}$) + 0.5(GDP per $capita_{y-5}$) + 0.4(GDP per $capita_{y-6}$) + 0.3(GDP per $capita_{y-7}$) + 0.2(GDP per $capita_{y-8}$) + 0.1(GDP per $capita_{y-9}$))/5.5 (5.5 = the sum of the weights).

The model covariates by dietary risk factor of interest are presented in **Table 1**. Study-level covariates were also used to inform the estimates, including metric type, which was used as an indicator category (optimal metric was indicated as the reference category), representativeness (national with or without representative sampling as the reference category, followed sub-national, predominantly urban, sub-national, predominantly rural, sub-national, both urban and rural, and sub-national, urban/rural/both coverage unknown), and diet assessment method (dichotomized to diet recalls/records or FFQs as the reference category vs. household budget surveys). Each dietary risk factor of interest was imputed using energy adjusted estimates standardized to 2,000 kcal or % of energy intake, and similarly imputed dietary covariates used were also energy-adjusted.

For two out of six dietary risk factors, omega-6 PUFA and seafood omega-3 fats, alternative metrics were provided for a few data sources instead of optimal ones. Mean intakes for alternative metrics may differ systematically from those of optimal ones (i.e., total PUFA intake is expected to be higher and linoleic acid intake is expected to be lower than the optimal measure of omega-6 PUFA intake, and EPA+DHA+DPA intake is expected to be higher than EPA+DHA intake (optimal)). Thus, the model included additional variance components to account for differences between different metric types, and to allow for optimal metrics to have more influence on estimates than alternative metrics. Sub-national (urban, rural, or both) studies might be undertaken in areas with low or high dietary fat and cholesterol intake, and thus may differ systematically from national surveys (with or without representative sampling). They might also have larger variation than national surveys. The model thus further included additional variance components to account for differences between national and sub-national studies. These variance components were estimated empirically to allow for non-national data to have less influence on estimates than national data. Mean dietary fat and cholesterol intakes may also differ

systematically according to the diet assessment method, with the difference mainly depending on whether the data were collected at the individual (diet recall/records, FFQs) vs. the household-level (household budget surveys). Therefore, our model also accounted for differences between individual-level and household-level dietary data, by allowing individual-level data to have more influence on estimates than household-level data.

Mean dietary fat and cholesterol intakes may be non-linearly associated with age; for example, the age relationship might flatten or even decrease in older ages. The age association may further vary across countries. Therefore, we used a cubic spline model, with parameters estimated as a function of dietary fat and cholesterol intakes at a baseline age. Mean dietary fats and cholesterol were estimated from the model for each one year age interval, by country, sex, and time period, for adults 20 years of age and older. Regional estimates for each time period were calculated as population-weighted averages of the constituent country estimates by age and sex. Sources of uncertainty were also quantified, including sampling uncertainty of original data sources, uncertainty associated with data sources that were not national, uncertainty due to data collected at the household- (rather than individual-) level or that did not provide optimal metrics, and uncertainty due to use of a model to estimate mean dietary fats and cholesterol by age, sex, and country when data were missing.

We fitted the Bayesian model with the Markov chain Monte Carlo (MCMC) algorithm and obtained samples from the posterior distribution of model parameters, which were in turn used to obtain the posterior distribution of mean dietary saturated fats, omega-6 polyunsaturated fats, seafood and plant omega-3 fats, trans fats, and of cholesterol for each age, sex, and country, reflecting all the above sources of uncertainty. The uncertainty intervals represent the 2.5 – 97.5 percentiles of the posterior distribution of estimated mean dietary fats and cholesterol. We further report the posterior probability that an estimated increase or decrease corresponds to a truly increasing or decreasing trend, referred to simply as posterior probability. Of note, the posterior probability is not a p value; the posterior probability would be 0·50 in a country or region in which an increase is statistically indistinguishable from a decrease, and a larger posterior probability indicated more certainty.

References

1. Lim SS, Vos T, Flaxman AD, et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2013;380(9859):2224-60.

2. James SL, Gubbins P, Murray CJ, Gakidou E. Developing a comprehensive time series of GDP per capita for 210 countries from 1950 to 2015. Popul Health Metr 2012;10(1):12.

3. FAO. Food Availability Data. http://faostat.fao.org/.