European Journal of Clinical Nutrition (2012) 66, 119–129

0954-3007/12
www.nature.com/ejcn

ORIGINAL ARTICLE

# Estimating the global and regional burden of suboptimal nutrition on chronic disease: methods and inputs to the analysis

R Micha[1], S Kalantarian[1], P Wirojratana[1], T Byers[2], G Danaei[1], I Elmadfa[3], E Ding[4,5], E Giovannucci[1,4], J Powles[6], S Smith-Warner[1,4], M Ezzati[7,8] and D Mozaffarian[1,4,5] on behalf of the Global Burden of Diseases, Nutrition and Chronic Disease Expert Group

[1]*Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA;* [2]*Department of Epidemiology, University of Colorado School of Public Health, Aurora, CO, USA;* [3]*Institute of Nutritional Sciences, University of Vienna, Vienna, Austria;* [4]*Department of Nutrition, Harvard School of Public Health, Boston, MA, USA;* [5]*Division of Cardiovascular Medicine and Channing Laboratory, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA; USA;* [6]*Department of Public Health and Primary Care, Institute of Public Health, University of Cambridge, UK;* [7]*Department of Global Health and Population, Harvard School of Public Health, Boston, MA, USA and* [8]*Department of Environmental Health, Harvard School of Public Health, Boston, MA, USA*

**Background/Objectives:** Global burdens of cardiovascular disease (CVD), diabetes and cancer are on the rise. Little quantitative data are available on the global impact of diet on these conditions. The objective of this study was to develop systematic and comparable methods to quantitatively assess the impact of suboptimal dietary habits on CVD, diabetes and cancer burdens globally and in 21 world regions.
**Subjects/Methods:** Using a comparative risk assessment framework, we developed methods to establish for selected dietary risk factors the effect sizes of probable or convincing causal diet–disease relationships, the alternative minimum-risk exposure distributions and the exposure distributions. These inputs, together with disease-specific mortality rates, allow computation of the numbers of events attributable to each dietary factor.
**Results:** Using World Health Organization and similar evidence criteria for convincing/probable causal effects, we identified 14 potential diet–disease relationships. Effect sizes and ranges of uncertainty will be derived from systematic reviews and meta-analyses of trials or high-quality observational studies. Alternative minimum-risk distributions were identified based on amounts corresponding to the lowest disease rates in populations. Optimal and alternative definitions for each exposure were established based on the data used to quantify harmful or protective effects. We developed methods for identifying and obtaining data from nationally representative surveys. A ranking scale was developed to assess survey quality and validity of dietary assessment methods. Multi-level hierarchical models will be developed to impute missing data.
**Conclusions:** These new methods will allow, for the first time, assessment of the global impact of specific dietary factors on chronic disease mortality. Such global assessment is not only possible but is also imperative for priority setting and policy making.
*European Journal of Clinical Nutrition* (2012) **66,** 119–129; doi:10.1038/ejcn.2011.147; published online 14 September 2011

**Keywords:** global burden of disease; diet; cardiovascular disease; diabetes; cancer

Correspondence: Dr R Micha or D Mozaffarian, Department of Epidemiology, Harvard School of Public Health, 677 Huntington Avenue Building 3-913, Boston, MA 02115, USA.
E-mail: rmicha@hsph.harvard.edu or dmozaffa@hsph.harvard.edu
*Contributors*: RM, SK and PW contributed to the study concept and design; systematic searches; data collection; interpretation of data; critical revision of the manuscript for important intellectual content and approval of the final manuscript for submission. RM and DM contributed in drafting of manuscript. TB, GD, ED, IE, EG, JP, SS-W, ME and DM contributed to the study concept and design; interpretation of data; critical revision of the manuscript for important intellectual content and approval of the final manuscript for submission.
Received 6 April 2011; revised 15 July 2011; accepted 20 July 2011; published online 14 September 2011

## Introduction

The worldwide burdens of chronic diseases, including cardiovascular disease (CVD), type II diabetes and cancer, are on the rise. It is expected that by 2020, almost 75% of all deaths worldwide and 60% of all disability-adjusted life years will be attributed to chronic diseases, with largest increases in developing rather than in developed countries (Murray and Lopez, 1997; World Health Organization, 1998). Considering that most chronic diseases are premature and can be prevented or delayed (Doll and Peto, 1981; Stampfer

*et al*., 2000; Danaei *et al*., 2009), identifying and targeting the modifiable risk factors with the greatest potential for reducing risk is of major scientific and public health importance. Suboptimal dietary habits are a major preventable cause of many chronic diseases. However, the quantitative impact of diet on chronic disease deaths and disease burdens worldwide is unknown, mainly because little systematically assessed global data are available on dietary habits.

The Global Burden of Disease (GBD) study is an international collaborative effort since 1990 to produce comprehensive and comparable estimates of the burdens of diseases, injuries and risk factors around the world. Such global estimates are highly informative and often used in policy making. The 2010 GBD study is currently underway. A critically important strength is a new focus on diet as a risk factor for chronic diseases, which was not systematically assessed in previous GBD studies, causing serious underestimation of the impact of diet on global health. For example, the 1990 GBD study did not assess any dietary factors and chronic diseases, and the 2000 GBD study assessed only total fruit and vegetable intake and chronic diseases (Lock *et al*., 2000). Even so, the 2000 analysis concluded that low fruit and vegetable consumption contributes substantially to worldwide deaths and disease burdens, highlighting the importance of evaluating diet and the need for systematic updating and emphasis on its potential impact. The aim of the GBD Nutrition and Chronic Disease Expert Group is to produce comparable estimates of the overall burden of diet on coronary heart disease, stroke, type II diabetes and cancer in 21 world regions for all dietary factors for which convincing or probable evidence of a causal diet–disease relationship exists.

The compilation of epidemiological parameters and the burden of disease estimates will be generated for 2 time periods, namely 1990 (1980–1998) and 2005 (1999–2010), and by sex and 8 age subgroups (20–24, 25–34, 35–44, 45–54, 55–64, 65–74, 75–84 and 85 + years). The analysis of chronic diseases will focus on adults, given the much lower incidence of chronic diseases in children, but the methods developed and described here can be extended in future analyses to consider the impact of diet on disease mortality in children. Specific quantitative inputs to the analysis were developed for the GBD study and are known as comparative risk assessment analysis (Table 1) (Ezzati *et al*., 2002, 2004; Murray *et al*., 2003; Danaei *et al*., 2009). These include: (1) the best evidence-based effect size of the causal diet–disease relationship; (2) the observed dietary factor distribution in the population; (3) the alternative dietary distribution, known as the optimal or theoretical minimum-risk exposure distribution (TMRED) and (4) the absolute numbers of disease-specific deaths in the population. These four inputs are used to determine the population attributable fraction (PAF), which is the proportional reduction in deaths

**Table 1** Inputs to the analysis for estimating the burden of specific dietary factors on chronic disease

| *Input* | *Purpose* | *Data sources* |
|---|---|---|
| 1. Effect size (relative risk estimate) of the causal diet–disease relationship, by age | To quantify the diet–disease relationship for which probable or convincing evidence of a causal effect exists | Systematic reviews and meta-analyses of randomized controlled trials and/or high-quality observational studies of each diet–disease relationship |
| 2. Optimal or theoretical minimum-risk exposure distribution (TMRED) | To determine the optimal risk factor exposure distribution that is realistically attainable, and is associated with the lowest possible disease risk | The observed exposure distribution corresponding to lowest mortality/incidence rates in epidemiological studies |
| 3. Dietary risk factor exposure distribution in the population, by age and sex | To determine the current average (mean ± s.d., or categories) of usual exposure levels of the dietary risk factor in the population | Nationally representative nutrition surveys, which have been identified and from which data have been extracted using systematic and comparable methods, with missing and incomplete data imputed using a multi-level hierarchical model |
| 4. Total number of disease-specific deaths (plus non-fatal events, when available) in the population, by age and sex | To determine the absolute numbers of disease events caused by a certain disease in the population | World Health Organization mortality database, with adjustment for misattribution and miscoding to maximize validity and comparability |
| *Overall analysis:* Population attributable fraction = $\frac{\int_{x=0}^{m} RR(x)P(x)dx - \int_{x=0}^{m} RR(x)P'(x)dx}{\int_{x=0}^{m} RR(x)P(x)dx}$ | To determine the proportion of disease events attributable to a certain dietary risk factor in the population, corresponding to the proportional reduction in deaths that would occur if the current dietary risk factor exposure distribution were shifted to the TMRED | $x$: the exposure level<br>$P(x)$: the usual exposure distribution in the population<br>$P'(x)$: the TMRED<br>$RR(x)$: the relative risk of mortality or morbidity at exposure level $x$<br>$m$: the maximum exposure level |

For each world region, burdens of chronic diseases due to each causal diet–disease relationship of interest will be estimated for two time periods, namely 1990 (based on data from 1980–1998) and 2005 (based on data from 1999–2009), by sex (males, female) and by age group (20–24, 25–34, 35–44, 45–54, 55–64, 65–74, 75–84, 75–84 and 85 + years).

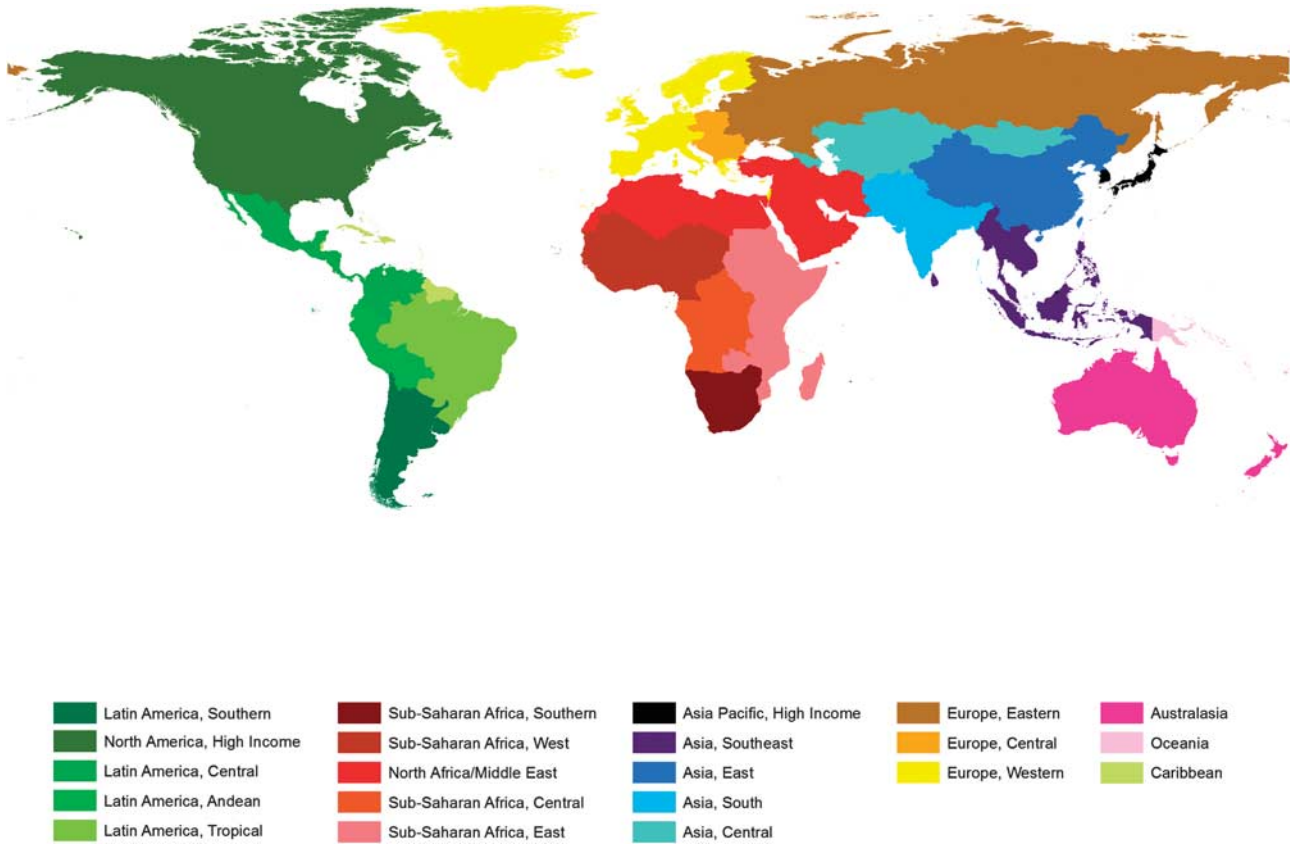


**Figure 1** Regionalization of the world for the purposes of the Global Burden of Diseases, Injuries, and Risk Factors (GBD) Study. The world was divided in 21 distinct geographic regions based on epidemiological similarities in childhood and adult mortality and major causes of death across broad geographical regions or continents. The end product of this regionalization includes low-income developing, mid-income developing and high-income developed regions.

that would occur if dietary exposure was shifted to optimal distribution. This paper presents GBD Nutrition and Chronic Disease Expert Group methods that were developed to assess the impact of suboptimal dietary habits on the global burden of chronic disease.

## Selection of 21 GBD world regions

The 2010 GBD study defines 21 distinct geographic regions for analysis (Figure 1) comprising 199 countries and territories (GBD Study, 2008). The objectives of regionalization were: (1) to define regions that were epidemiologically relatively homogeneous, allowing for detailed information from one country to be extrapolated to other countries within the same region and (2) related to the first, to generate regional burden estimates, which individual countries could use to inform public health policies. The regions were chosen based on mortality estimates from the World Health Organization (WHO) and the United Nations, in addition to current knowledge on country-specific epidemiological conditions. All regions were defined and selected to incorporate broad geographical regions or continents and comprise no fewer than two countries. Countries were grouped based on mortality levels (child and adult) and major causes of death, rather than national population or income, despite a clear relation of the latter to epidemiological profiles. The end product of regionalization includes low-income developing regions, such as those in sub-Saharan Africa and South Asia, mid-income developing regions, such as those in East Asia and Latin America and high-income regions, such as those in North America, Asia Pacific and Western Europe. At present, several other regional classifications are in use. The World Bank uses economic classifications, which change over time; the United Nations uses geographic classifications; and the WHO uses geographical and political classifications; each of these serve different purposes than a focus on health. To our knowledge, a cross-mapping of these different regional classifications is not currently available. Ideally, all analyses should move to country level, once sufficient data become available; at this stage, combinations of geographical, income and epidemiological data are practical.

## Diet–disease relationships

*Grading evidence*

To identify causal diet–disease relationships, we assessed current major dietary risk factors, identified chronic diseases that may be affected, evaluated convincing or probable evidence for a causal effect and identified unbiased effect estimates on chronic disease risk. Several criteria were taken into account to establish a causal effect, including the Bradford Hill criteria for causation (Hill, 1965; GBD Study, 2008), the WHO criteria for grading evidence (World Health Organization, 2003) and the similar World Cancer Research Fund criteria (World Cancer Research Fund/American Institute for Cancer Research, 2007). These criteria do not define indisputable rules for causation. Rather, overall evidence is graded by expert opinion as convincing, probable, possible or insufficient for either the presence or absence of a clinical effect (World Health Organization, 2003; World Cancer Research Fund/American Institute for Cancer Research, 2007) (Table 2). We focused on those diet–disease relationships with either convincing or probable evidence for a causal effect on coronary heart disease, stroke, type II diabetes or cancers (Table 3). We also considered whether this effect varied depending on the replacement nutrient, when appropriate (for example, the replacement of saturated fats with polyunsaturated fats) (Micha and Mozaffarian, 2010).

*Quantifying relationships*

A major strength of this study will be the use of systematic reviews and meta-analyses of high-quality observational studies and, when available, randomized controlled trials to derive the best current evidence of etiological effects of diet on CVD, diabetes and cancer risk. The best available estimates of the effect size (relative risk, RR) of the causal diet–disease relationships and supporting evidence will be the subject of a forthcoming paper, drawing on, for example, several recent or ongoing meta-analyses (Mozaffarian *et al*., 2006, 2010; World Cancer Research Fund/American Institute for Cancer Research, 2007; Micha *et al*., 2010). For each diet–disease relationship, we will also assess the potential for differential effects on incidence versus cause-specific mortality.

Relative risks will be obtained ideally per unit of exposure for risks measured continuously or, when the former is unavailable or evidence exists for non-linear effects, for each exposure category. In the absence of data from randomized controlled trials, we will select long-term prospective cohorts that adjust for major potential confounding factors and, when available, for bias introduced by measurement error in the exposure (Rosner *et al*., 1992; Lewington *et al*., 2002, 2007; Wald and Law, 2003; Lawes *et al*., 2004; Fleming *et al*., 2005), which generally results in underestimation of the true RR (Willett, 1998a, b). Data from retrospective case–control studies will not be used, given the potential for substantial recall and control selection bias.

*Effect modification*

Data are expected to be limited to evaluate the potential for effect modification of diet–disease relationships by age, sex or region. When data are scarce, we will use documented interactions of metabolic risk factors (such as blood pressure,

**Table 2** FAO/WHO criteria for grading evidence for causality of diet–disease relationships

| *Grading* | *Evidence* |
|---|---|
| Convincing | Evidence based on epidemiological studies showing consistent associations between exposure and disease, with little or no evidence to the contrary. The available evidence is based on a substantial number of studies including prospective observational studies and where relevant, randomized controlled trials of sufficient size, duration and quality showing consistent effects. The association should be biologically plausible |
| Probable | Evidence based on epidemiological studies showing fairly consistent associations between exposure and disease, but where there are perceived shortcomings in the available evidence or some evidence to the contrary, which precludes a more definite judgment. Shortcomings in the evidence may be any of the following: insufficient duration of trials (or studies), insufficient trials (or studies) available, inadequate sample sizes and incomplete follow-up. Laboratory evidence is usually supportive. Again, the association should be biologically plausible |
| Possible | Evidence based mainly on findings from case–control and cross-sectional studies. Insufficient randomized controlled trials, observational studies or non-randomized controlled trials are available. Evidence based on non-epidemiological studies, such as clinical and laboratory investigations, is supportive. More trials are required to support the tentative associations, which should also be biologically plausible |
| Insufficient | Evidence based on findings of a few studies which are suggestive, but are insufficient to establish an association between exposure and disease. Limited or no evidence is available from randomized controlled trials. More well-designed research is required to support the tentative associations |

Abbreviations: FAO, Food and Agriculture Organization; WHO, World Health Organization.
Adapted from the World Health Organization, Diet, Nutrition and the Prevention of Chronic Diseases: report of a joint WHO/FAO expert consultation (World Health Organization and FAO, 2003). Advances in nutritional science now provide a substantial body of evidence to evaluate causality of various diet–disease relationships (Harris *et al*., 2009; Smit *et al*., 2009). Optimal evidence is provided by well-conducted randomized clinical trials and/or prospective cohort studies of dietary risk factors and disease outcomes in humans, which provide direct evidence for effects on disease in comparison with studies of risk factors (such as blood lipids, glucose levels). Controlled trials of multiple risk factors provide further supporting evidence. These research paradigms each have complementary strengths and limitations, and conclusions can be considered most robust when evidence is consistent across paradigms. Evidence from retrospective, cross-sectional, ecological and animal studies are important for further support or hypothesis generation, but not for establishing causality.

**Table 3** Diet–disease relationships identified to date based on either convincing or probable evidence for a causal effect[a]

| *Dietary risk factors* | *CVD outcomes* | *Cancer outcomes* |
|---|---|---|
| *Foods* | | |
| Fruits | CHD, stroke | Mouth, pharynx, larynx, esophagus, lung |
| Vegetables | CHD, stroke | Mouth, pharynx, larynx |
| Whole grains | CHD, diabetes | |
| Nuts | CHD | |
| Red meats, unprocessed | Diabetes | Colorectal |
| Processed meats | CHD, diabetes | Colorectal |
| Milk | Diabetes | Colorectal |
| Sugar-sweetened beverages | Body mass index, diabetes | |
| *Nutrients* | | |
| Polyunsaturated fat replacing saturated fat | CHD | |
| Seafood omega-3 fatty acids | CHD, stroke | |
| Trans fats | CHD | |
| Dietary fiber | CHD | Colorectal |
| Dietary sodium | Blood pressure, stroke | Stomach |
| Dietary calcium | | Colorectal, prostate |

Abbreviations: CHD, coronary heart disease; CVD, cardiovascular disease.
[a]Future reports will present the full supporting evidence and best available estimates of the effect size (relative risk) of the causal diet–disease relationships identified. Additional dietary risk factors currently being evaluated include beans/legumes, plant omega-3 fatty acids and dietary cholesterol; as well as other potential causal diet-disease relations for the foods and nutrients listed above.

blood glucose, blood cholesterol) and disease risk to make inference about dietary factors. For example, evidence from several previous studies suggests that per unit of exposure, the proportional effects of many metabolic risk factors on chronic disease may be similar by sex (Lewington *et al*., 2002, 2007; Lawes *et al*., 2004; Ni Mhurchu *et al*., 2004; Fleming *et al*., 2005). Therefore, unless compelling evidence to the contrary is found, we will assume a lack of proportional effect size modification by sex for the dietary factors of interest. In contrast, our preliminary work indicates that for most metabolic and other risk factors, age influences the proportional and absolute effects on CVD and diabetes. In general, an inverse age association is seen for RR differences, principally because of the phenomenon of competing risks; and a positive age association is seen for absolute risk differences because of increasing overall risk with age (Danaei *et al*., 2009).

Accounting for such age associations will be critical given the differences in population age distributions across different countries and regions, particularly in the developing world. We will incorporate this pattern of effect modification by age in a consistent manner across all dietary risk factors. For most risk factors, the age interaction for RRs is expected to follow a log-linear relationship (Danaei *et al*., 2009). When pooling RRs from individual epidemiological studies, such data will be converted to similar age groups. For those studies providing RRs by age, conversion to GBD age groups will be performed by interpolation using a log-linear relationship. For studies providing a single RR for all ages combined, the overall RR will be redistributed into age-specific RRs using a log-linear relationship centered at the median age at events in the study (Lock *et al*., 2000). To evaluate potential effect modification by region, mainly comparing Asian versus Western regions, we will compare RRs from meta-analyses of international regional cohorts with subjects of different race groups (for example, the Asia Pacific Cohort Studies Collaboration versus the Prospective Cohort Studies Collaboration).

## Theoretical minimum-risk exposure distribution

The TMRED is the alternative distribution that would be expected to produce the lowest possible disease risk associated with an exposure, irrespective of whether currently available or feasible in practice (Ezzati *et al*., 2004) (Figure 2). So that our estimates can directly inform policies and priorities, for each dietary factor, we will evaluate the optimal minimum-risk exposure distribution, a closely related concept to the TMRED that also requires the alternative distribution to be feasible or observed in some populations. For consistency with other GBD publications, we will refer to this optimal minimum-risk exposure distribution as the TMRED hereafter.

For some harmful exposures (such as tobacco smoking), the TMRED can be zero. However, zero exposure is implausible or impossible for most dietary factors. Therefore, we will determine the TMRED (minimum for harmful exposures and maximum for protective exposures) that can be feasibly achieved based on observed levels associated with the lowest mortality rates in epidemiological studies and levels to which beneficial effects may plausibly continue. When appropriate, we will also consider simple intake levels versus 'replacement' modeling (Mozaffarian *et al*., 2010). As all individuals in a population cannot be brought to precisely the same exposure level, a plausible s.d. around

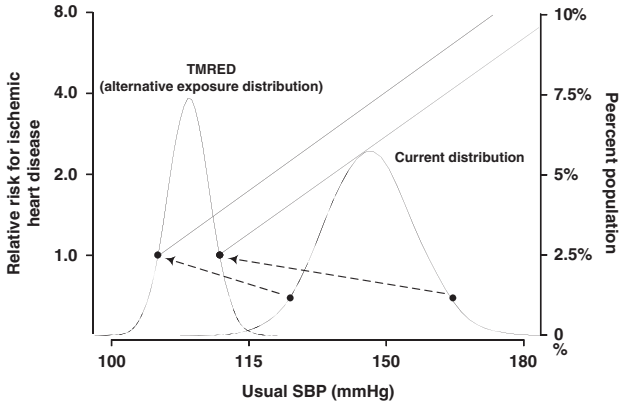


**Figure 2** Theoretical minimum-risk exposure distribution for continuous risk factors using systolic blood pressure (SBP) as an example. Each point represents a hypothetical individual or small group of individuals in the population. The solid lines represent the increasing RR, on a log scale, for ischemic heart disease with increasing SBP. Adapted with permission from the study by Ezzati *et al*. (2004).

the TMRED will also be determined, reflecting the residual between-person variation within the population after implementation of hypothetical changes in mean population exposure. When such data are not directly available, we will make inference from s.d. of TMREDs of metabolic risk factors, For example, in our preliminary work (Danaei *et al*., 2009), we defined the TMRED ± s.d. for fruits (300 ± 30 g/day), vegetables (300 ± 30 g/day), seafood omega-3 fatty acids (250 mg/day), polyunsaturated fatty acid as a replacement for saturated fats (10 ± 1% energy), trans fats (0.5 ± 0.05% energy) and sodium (500 ± 50 mg/day).

## Dietary risk factor exposure distribution

*Definition of dietary risk factors*

To facilitate comparability across countries, regions, time periods and diet assessment methods, we created standardized optimal and alternative exposure definitions (metrics) and units of measure for each dietary factor (Table 4). We selected these metrics and units of measure to be (1) as similar as possible to the corresponding values in epidemiological studies or trials used to quantify the harmful or protective effects of each dietary factor on CVD, diabetes and cancer risk; (2) most interpretable from biologic and policy perspectives when considering diet and (3) as similar as possible to common definitions used to ascertain exposure data in the regions and countries of interest.

*Data sources and retrieval*

Systematic searches and reviews were performed to identify country-specific nationally representative surveys and to determine mean and s.d. of dietary exposures by time periods and by sex- and age-specific groups. Standardized protocols were developed to allow collection, analysis and extraction of data in a systematic and consistent manner across countries and regions, including standardized systematic literature searches, identification and inclusion of relevant surveys, survey quality assessment and data extraction.

*Systematic literature searches.* We searched for nationally representative nutrition surveys that provided exposure data on the dietary risk factors of interest. If nationally representative surveys were not available for any country, then we also considered national surveys without representative sampling, followed by regional, urban or rural surveys, and local selected cohorts, provided that selection and measurement bias were not apparent limitations. Multiple sources were searched including MEDLINE, Embase, CAB abstracts, WHO library (WHOLIST) and SIGLE (gray literature database), hand searching of reference lists of identified studies and direct contact with authors. Searches were performed from March 2008 to September 2010. Search terms included: 'nutrition' OR 'diet' OR 'food habits' OR 'nutrition surveys' OR 'diet surveys' OR 'food habits'[mesh] OR 'diet'[mesh] OR 'nutrition surveys'[mesh] OR 'diet surveys'[mesh] AND ('country of interest'). Additional search terms were applied to refine or expand each country search, as appropriate (see Supplementary Figure 1).

*Survey identification.* Surveys were included in the initial screening phase if the survey was reasonably population based and representative, exposure data were reported or could be plausibly obtained and sample size was at least 100 individuals. Additional issues were considered such as whether the survey provided age- and gender-specific estimates, covered a wide age range of adults or had used a validated diet assessment tool. For countries with no surveys identified, other sources of potential data were considered, including household budget survey data, FAO (Food and Agriculture Organization) food availability data, large cohorts, the WHO infobase and the STEP database.

*Survey quality assessment.* Survey quality assessment was performed by review of evidence for selection bias, sample representativeness, response rate, sample size and validity of diet assessment methods. A ranking scale was developed (see Supplementary Table 1) to assess quality of identified surveys and validity of the diet assessment method used. Sample size was initially considered as a measure of quality, but we concluded that such a metric is already captured in each study-specific s.e. Although we attempted to ensure that quality assessment was as unbiased and representative as possible, we recognized that assignment of survey quality and validity of diet assessment methods could be subjective. Nonetheless, the proposed ranking scale was our best available method to allow and to assess comparability of collected data within and across regions.

**Table 4** Dietary risk factors to be evaluated, including optimal and alternative metrics and units of measurement for data collection[a]

| *Food/nutrient* | *Optimal metric* | *Acceptable alternative metrics* | *Optimal unit of measure* | *Alternative unit of measure* |
|---|---|---|---|---|
| 1. Energy | Total energy intake from all dietary sources | NA | kcal/day | kJ/day |
| 2. Fruits | Total fruit intake, including fresh, frozen, cooked, canned or dried fruit. Exclude fruit juices and salted or pickled fruits | 1. Total fruit intake, including fruit juices 2. Total fruit and vegetable intake (f & v), excluding f & v juices. 3. Total f & v intake, including f & v juices | g/day (ideally energy adjusted) | servings/day |
| 3. Fruit juices | Total fruit juices intake, provide separately | NA | g/day (ideally energy adjusted) | servings/day |
| 4. Vegetables | Total vegetable intake, including fresh, frozen, cooked, canned or dried vegetables. Exclude salted or pickled vegetables, vegetable juices, starchy vegetables (such as potatoes, corn), legumes, nuts and seeds | 1. Total vegetable intake, including vegetable juices 2. Total f & v intake, excluding f & v juices 3. Total f & v intake, including f & v juices | g/day (ideally energy adjusted) | servings/day |
| 5. Beans, legumes | Total intake of beans and legumes, including tofu. Exclude soy milk | NA | g/day (ideally energy adjusted) | servings/day |
| 6. Nuts, seeds | Total intake of nuts and seeds | NA | g/day (ideally energy adjusted) | servings/day |
| 7. Whole grains/ whole grain foods | Total intake of whole grain foods, including breakfast cereals, bread, rice, pasta, biscuits, muffins, tortilla, pancake, etc. A whole grain is defined as a food with $\geqslant$1.0 g of fiber per 10 g of carbohydrate (reference to the fiber content of whole wheat)[b] | 1. Any other definition of whole grains as used in your survey (for example, based on food or products names, other fiber content, etc.) | g/day (ideally energy adjusted) | servings/day (that is, ounce equivalents) |
| 8. Red meats | Total red meat intake from all livestock, both domesticated and non-domesticated (that is, game), excluding poultry, fish, eggs and all processed meats | 1. Total red meat as previously defined, but also including processed red meats 2. Total red meat as previously defined, but also including poultry | g/day (ideally energy adjusted) | servings/day |
| 9. Processed meats | Total processed meat intake (for example, processed deli or luncheon meats (ham, turkey, pastrami, etc.), bacon, salami, sausages, bratwursts, frankfurters, hot dogs) | NA | g/day (ideally energy adjusted) | servings/day |
| 10. Milk | Total milk intake (combined non-fat, low-fat and full-fat milk). Exclude soya milk or other plant-derived alternatives | NA | g/day (ideally energy adjusted) | servings/day |
| 11. Sugar-sweetened beverages | Total sugar-sweetened beverages intake defined as any sugar-sweetened beverage with $\geqslant$50 kcal per 8 oz (226.8 g) serving, including carbonated beverages, soft drinks, sodas, energy drinks, fruit drinks, etc.[b] Exclude 100% fruit and vegetable juices | NA | g/day (ideally energy adjusted) | servings/day |
| 12. Saturated fat | Total saturated fat intake from all dietary sources (primarily meat and dairy products and tropical oils) | NA | % kcal (energy contribution) | g/day (ideally energy adjusted) |
| 14. Omega-6 (*n*-6) polyunsaturated fat | Total omega-6 fatty acid intake from all dietary sources (primarily liquid vegetable oils, including soybean oil, corn oil and safflower oil) | 1. Total PUFA intake, as n-6 consists $\sim$90% of PUFA ($r>0.95$) 2. Total LA (linoleic) intake | % kcal (energy contribution) | g/day (ideally energy adjusted) |
| 14. Seafood omega-3 (*n*-3) fat | Total dietary EPA + DHA (eicosapentaenoic + docosahexaenoic) intake. Exclude supplements | 1. Total dietary EPA + DPA + DHA 2. Total seafood intake 3. Total fish intake | mg/day (ideally energy adjusted) | % kcal (if seafood/fish: servings/day) |
| 15. Plant omega-3 (*n*-3) fat | Total dietary ALA ($\alpha$-linolenic acid) intake. Exclude supplements | NA | mg/day (ideally energy adjusted) | % kcal |
| 16. Trans fat | Total trans fatty acid intake from all dietary sources (mainly partially hydrogenated vegetable oils and ruminant products) | NA | % kcal (energy contribution) | g/day (ideally energy adjusted) |
| 17. Dietary cholesterol | Total dietary cholesterol from all dietary sources | NA | mg/day (ideally energy adjusted) | NA |
| 18. Dietary fiber | Total dietary fiber intake from all dietary sources (fruits, vegetables, grains, legumes, pulses). Exclude supplements | NA | g/day (ideally energy adjusted) | NA |

**Table 4** *Continued*

| *Food/nutrient* | *Optimal metric* | *Acceptable alternative metrics* | *Optimal unit of measure* | *Alternative unit of measure* |
|---|---|---|---|---|
| 19. Dietary sodium | Total dietary sodium intake from all dietary sources | NA | mg/day (ideally energy adjusted) | NA |
| 20. Dietary calcium | Total dietary calcium intake. Exclude supplements | NA | mg/day (ideally energy adjusted) | NA |

Abbreviations: NA, not applicable; PUFA, polyunsaturated fatty acid.
[a]Standardization of metrics and units is necessary to maximize comparability across surveys, countries, regions and time periods. The optimal metric for each dietary risk factor was selected to correspond to those used in studies to derive the effect size of the causal diet–disease relationship, with additional consideration of biologic and policy perspectives for interpretation and common metrics available in surveys from the countries and regions of interest. Alternative metrics were also developed to facilitate data collection across multiple survey methods, as well as to account for potential differences across regions. Optimal and alternative measurements units were developed based on similar considerations. For foods, g/day was preferred to servings/day to account for differences in serving sizes across surveys. Energy adjustment was considered optimal, either using residual or nutrient density methods, to reduce measurement error and at least partly account for differences in body size, metabolism and physical activity between individuals (Willett, 1998a, b).
[b]For some foods, such as whole grain foods and sugar-sweetened beverages, no established operational definition exists. We selected pragmatic definitions used in the American Heart Association 2020 Strategic Impact Goals (Lloyd-Jones *et al*., 2010). For each survey, we asked for data to be re-analyzed using these standardized definitions.

*Expert identification*. We recognized early in the process that although nutrition surveys had been carried out in many countries, published exposure data were either very limited or not in the required format. We thus relied almost entirely on direct author contacts in each country to provide us with exposure data directly. To assist in this process, we developed a standardized protocol to contact experts and request data, including standardized invitation e-mails and timelines, survey information request forms, data analysis instructions and data extraction sheets (see Supplementary Figure 2). Contacts were identified from each relevant publication. Identified experts were contacted and invited to become corresponding members of the GBD Nutrition and Chronic Disease Expert Group and to share their generated estimates of the specific foods and nutrients of interest according to the GBD age and gender subgroups. Each corresponding member could opt to send us the raw data or perform the analysis and send us the data in the necessary format. The standardized data analysis approach accounted for sampling strategies within the survey including sampling weights, if available, used the average of all days to estimate dietary intake and correct the population s.d. for random within-person variation, if multiple dietary assessments were available (see Identification of usual exposure distribution, below) and adjusted for total energy intake using either residual or nutrient density methods.

*Data extraction*. For each identified survey, published or directly obtained data were extracted using a standardized electronic extraction sheet, including survey name, country, years performed, sampling design, response rate, age range, type of data (individual versus household), national representativeness, diet assessment method and validation, sample size, definitions and measurement units of dietary risk factors of interest, as well as mean and s.d. of intake (exposure distributions). To ensure correct extraction, random double checks were performed.

*Evaluation of dietary assessment methods*
Dietary habits were commonly assessed by 24-h diet recalls/records, food frequency questionnaires (FFQs) and household surveys. Diet records, recalls, and FFQs provided estimates of individual-level consumption, in comparison with household surveys, which estimated food intake/expenditure at the household level. Nutrient intakes were subsequently estimated in each survey by means of country-specific food composition tables and summing of nutrients across all foods consumed.

Diet records/recalls and FFQs have different strengths and limitations (Willett, 1998a, b). Diet recalls/records estimate individual dietary intake on one or more days, whereas FFQs estimate habitual dietary intake, commonly over the past 1 year. FFQs are practical and cost-efficient and allow relatively accurate between-individual rankings of usual long-term diet. Diet records/recalls are often more quantitatively accurate for absolute intakes during the brief periods (for example, 24 h) of assessments and allow greater flexibility and detail in describing foods and preparation methods. However, in comparison with FFQs, such methods are more limited for capturing long-term diet due to day-to-day and seasonal within-person variation, especially for foods that are not consumed every day. Repeated short-term recalls can be used to statistically correct but not eliminate this limitation. To account for these different strengths and limitations, we developed a ranking scale to assess appropriateness and validity of the diet assessment method used (see Supplementary Table 1).

*Identification of 'usual' exposure distributions*
When multiple short-term diet assessments are available, the population mean estimated from the average of two single measures will be unbiased, but the population s.d. will overestimate the s.d. of the true 'usual' population exposure distribution (Willett, 1998a, b). Correcting for this

overestimation is important to identify true population exposure distributions. When multiple short-term diet measures were available, the within- versus between-person variation in intake of foods and nutrients were quantified and partitioned using established methods (Willett, 1998a, b). This corrects population s.d. to account for within-individual variation in short-term (for example, day-to-day) dietary intake.

## Missing exposure data and imputation

Although our approaches to data retrieval will lead to the most comprehensive data set of global dietary factors, we expect data gaps both within countries and within regions. For example, data on each of the dietary factors of interest may not be available in all individual countries; certain age groups may not be separately reported in some surveys; the survey assessment may be performed in different years than the analysis years of interest; and some data may be missing entirely for a country or multiple countries in a region.

On the basis of previous work and data currently identified, we are developing novel methods to impute missing country and regional exposure data, based on non-missing exposure data from other regions and available country-level covariates on other characteristics (Ezzati *et al*., 2005; Stevens *et al*., 2008). A hierarchical imputation model is anticipated that will account for both country- and region-level data and multiple levels of missingness. The primary analysis will focus on generating regional and global estimates. Data from individual countries will be used simultaneously as inputs to country and regional estimates. Dietary risk factor levels and trends over time in individual countries will be nested within regional levels, which will be in turn nested within global levels. This structure will allow the model to borrow information across countries and regions as necessary, depending on the extent of data which are missing or less informative (for example, having large uncertainty). Time-varying country-level covariates, such as national income, population, distributions of age, gender and other lifestyle habits, economic variables and FAO food supply data, will serve to further inform the estimates. The model will also include additional variance components to account for differences between subnational and nationally representative studies, so that non-national data have less influence on estimates than do national data.

This approach to handling missing and less informative data, incorporating our methods for collecting and maximizing comparability of all existing data into a hierarchical prediction model with multiple informative covariates, will provide the most robust global database to date on nutritional risk factors for chronic disease. A full description of this model and its outputs will be the subject of a separate report upon the conclusion of this study.

## Disease-specific mortality and morbidity statistics

To determine the absolute and proportional impact of each dietary factor on disease burdens, cause-specific disease information is required according to age- and sex-specific subgroups. Data on disease-specific mortality will be first obtained from the WHO mortality database (World Health Organization, 2009), which provides annual reports by age and sex. The WHO mortality database uses ICD (International Classification of Diseases)-10 codes from countries' civil registration systems to assign a single underlying cause of death. Our previous work has demonstrated that validity and comparability of such cause-of-death statistics may vary across countries, especially for CVD and diabetes (Lu *et al*., 2006; Murray *et al*., 2006, 2008; Naghavi *et al*., 2010). Misattribution or miscoding can occur because of incorrect or systematic biases in diagnosis, incorrect or incomplete death certificates, misinterpretation of ICD rules for selection of the underlying cause and variations in the use of coding categories for poorly defined or unknown causes. We will use previously described methods to adjust for the lack of comparability in cause-of-death data (Murray *et al*., 2006, 2008; Danaei *et al*., 2009), based on multiple contributing causes of death and country of residence (Danaei *et al*., 2009; Naghavi *et al*., 2010). We will also use GBD data obtained from the current disease expert groups on incidence, prevalence, case fatality and survival for non-fatal disease cases and their sequelae to make internally consistent estimates of disease incidence and duration. This information will be used to make estimates of years of life lived with disability.

## Estimating mortality and morbidity attributable to risk factors

The impact of each dietary factor on disease burdens will be calculated using previously described inputs to compute the PAF for each specific diet–disease relationship. The PAF reflects the proportional reduction in deaths for each disease causally associated with the exposure that would occur if the usual exposure distribution had been reduced (for harmful exposures) or increased (for protective exposures) to the optimal minimum-risk exposure distribution. The PAF estimates the total effects of a risk factor and is computed as we have previously described, taking into account both age- and sex-specific exposure distributions and disease rates (Danaei *et al*., 2009) (Table 1).

To obtain the absolute number of disease-specific deaths attributed to a dietary factor, the PAF is multiplied by the total deaths from that disease. Deaths are always assigned to a single underlying cause (for example, a death due to coronary disease is not also counted as a death due to other causes). Thus, the number of attributable deaths from different diseases across a single dietary factor can be summed to derive the best estimate of total (all-cause) mortality attributable to that dietary factor.

In contrast, risk factors for any single death will often overlap, with potentially competing risks (for example, from two or more dietary or other risk factors) increasing the likelihood of death. In particular, chronic diseases are rarely caused by any single risk factor. Distributions of competing risks may also vary across population subgroups, with some exposed to many and others to fewer competing risks. Furthermore, some risk factors may directly mediate the effects of more upstream risk factors. For instance, the total deaths attributable to sodium consumption and to hypertension cannot simply be summed to determine their joint effects, as a proportion of hypertension-caused deaths will be attributable to effects of sodium consumption on hypertension. On the other hand, such mediated effects do not alter the accuracy of the total mortality attributable to any single risk factor, which reflects the total effects, both direct and causally mediated through other factors, on disease.

Owing to competing risks, multi-causality and mediating effects, the numbers of deaths across different risk factors cannot simply be summed. Accurate estimation of joint effects of two or more risk factors requires knowledge of their (1) joint causal effects (joint RRs) on each disease, which accounts for multi-causality and mediated effects and (2) joint distributions in different population subgroups. We have found that reliable data on joint RRs and joint exposure distributions are rarely, if ever, available for two dietary (or non-dietary) risk factors. Our group plans to pursue evaluation and modeling of such joint effects in the future for both dietary and non-dietary risks.

The effects of one risk factor may also vary depending on the distribution of another risk factor. Our methods incorporate estimates of such effect modification by age and (when appropriate) by sex or world region. Current GBD methods do not quantify potential non-additivity (interaction) across effects of different modifiable risk factors, largely because of limited reliable data on such interactions for most risk factors. This is an active area of research within our Expert Group and the GBD in general, and we hope to develop and use these methods in the future as possible.

## Conclusions

We know that overall diet quality has a major impact on chronic diseases (Mozaffarian *et al.*, 2011). In the United States and similar nations, specific quantitative estimates of this impact for different dietary factors inform priorities for prevention (Stampfer *et al.*, 2000; Danaei *et al.*, 2009). In comparison, little is known about distributions and quantitative effects on chronic diseases of specific dietary factors in other world regions. The new methods presented herein will systematically assess and compile nationally representative data on exposure distributions and heterogeneity of major dietary factors across regions and by sex- and age-specific groups, and quantitative estimates of causal effects of these dietary factors on specific major chronic diseases. This will allow, for the first time, comparable and quantitative assessment of the global impact of specific diet factors on CVD, diabetes and cancers by sex- and age-specific groups in both developed and developing countries.

Such global assessment is imperative for priority setting and policy making in different countries and regions, allowing identification and targeting of the dietary factors with greatest impact on disease. Such global assessment is also critical to understand reasons for and thus reduce health disparities across nations. The results are expected to inform future epidemiological, population-based and interventional studies targeting these modifiable risk factors, as well as prevention initiatives and public health policies.

Overall, this project will produce the first global nutrition database and first global quantitative estimates of impact on multiple chronic diseases. This will increase worldwide awareness on the potential of diet to prevent disease and improve health. This study will also highlight the importance of having and systematically updating dietary consumption data in all nations to facilitate the global scientific and health communities to base their research and policies on the most recent, robust and reliable evidence possible. Our results will directly inform both regional and global priorities for implementation of a comprehensive strategy to target diet, including increased intake of the most protective dietary factors and decreased intake of the most harmful, to prevent millions of CVD, diabetes and cancer events and deaths worldwide.

## Conflict of interest

The authors declare no conflict of interest.

## Acknowledgements

We thank Barbara Bowman, Patricia Constante Jamie, Saman Fahimi, Karen Lock, Verena Nowak and Joceline Pomerleau for their expert advice on developing the methodology. We also thank Louise Dekker, Jenna Golan, Shahab Khatibzadeh, Mayuree Rao, Peilin Shi, Liesbeth Smit and Georgina Waweru for providing analytic and administrative support. We also are enormously grateful to each of our corresponding authors from all involved nations for providing unpublished data—their names are listed in the Supplementary Appendix. This work was undertaken as a part of the Global Burden of Diseases, Injuries, and Risk Factors Study. A grant from the Bill and Melinda Gates Foundation supported the Study's core activities and partially supported the epidemiological reviews in this paper.

## References

Danaei G, Ding EL, Mozaffarian D, Taylor B, Rehm J, Murray CJ *et al.* (2009). The preventable causes of death in the United States:

comparative risk assessment of dietary, lifestyle, and metabolic risk factors. *PLoS Med* **6**, e1000058.
Doll R, Peto R (1981). The causes of cancer: quantitative estimates of avoidable risks of cancer in the United States today. *J Natl Cancer Inst* **66**, 1191–1308.
Ezzati M, Lopez AD, Rodgers A, Murray CJ (2004). *Comparative Quantification of Health Risks: Global and Regional Burden of Disease Attributable to Selected Major Risk Factors (Volumes 1 and 2)*. World Health Organization: Geneva.
Ezzati M, Lopez AD, Rodgers A, Vander Hoorn S, Murray CJ (2002). Selected major risk factors and global and regional burden of disease. *Lancet* **360**, 1347–1360.
Ezzati M, Vander Hoorn S, Lawes CM, Leach R, James WP, Lopez AD *et al*. (2005). Rethinking the 'diseases of affluence' paradigm: global patterns of nutritional risks in relation to economic development. *PLoS Med* **2**, e133.
Fleming C, Whitlock EP, Beil TL, Lederle FA (2005). Screening for abdominal aortic aneurysm: a best-evidence systematic review for the U.S. Preventive Services Task Force. *Ann Intern Med* **142**, 203–211.
GBD Study (2008). *The Global Burden of Diseases, Injuries, and Risk Factors*. Operations manual. http://www.globalburden.org/GBD_Study_Operations_Manual_Jan_20_2009.pdf.
Harris WS, Mozaffarian D, Lefevre M, Toner CD, Colombo J, Cunnane SC *et al*. (2009). Towards establishing dietary reference intakes for eicosapentaenoic and docosahexaenoic acids. *J Nutr* **139**, 804S–819S.
Hill AB (1965). The environment and disease: association or causation? *Proc R Soc Med* **58**, 295–300.
Lawes CM, Parag V, Bennett DA, Suh I, Lam TH, Whitlock G *et al*. (2004). Blood glucose and risk of cardiovascular disease in the Asia Pacific region. *Diabetes Care* **27**, 2836–2842.
Lewington S, Clarke R, Qizilbash N, Peto R, Collins R (2002). Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. *Lancet* **360**, 1903–1913.
Lewington S, Whitlock G, Clarke R, Sherliker P, Emberson J, Halsey J *et al*. (2007). Blood cholesterol and vascular mortality by age, sex, and blood pressure: a meta-analysis of individual data from 61 prospective studies with 55 000 vascular deaths. *Lancet* **370**, 1829–1839.
Lloyd-Jones DM, Hong Y, Labarthe D, Mozaffarian D, Appel LJ, VanHorn L *et al*. (2010). Defining and setting national goals for cardiovascular health promotion and disease reduction: the American Heart Association's strategic Impact Goal through 2020 and beyond. *Circulation* **121**, 586–613.
Lock K, Pomerleau J, Causer L, McKee M (2000). Chapter 9: Low fruit and vegetable consumption. *Comparative Quantification of Health Risks* **1**, 597–728.
Lu TH, Hsu PY, Bjorkenstam C, Anderson RN (2006). Certifying diabetes-related cause-of-death: a comparison of inappropriate certification statements in Sweden, Taiwan and the USA. *Diabetologia* **49**, 2878–2881.
Micha R, Mozaffarian D (2010). Saturated fat and cardiometabolic risk factors, coronary heart disease, stroke, and diabetes: a fresh look at the evidence. *Lipids* **45**, 893–905.
Micha R, Wallace SK, Mozaffarian D (2010). Red and processed meat consumption and risk of incident coronary heart disease, stroke, and diabetes mellitus: a systematic review and meta-analysis. *Circulation* **121**, 2271–2283.
Mozaffarian D, Appel LJ, Van Horn L (2011). Components of a cardioprotective diet: new insights. *Circulation* **123**, 2870–2891.
Mozaffarian D, Katan MB, Ascherio A, Stampfer W, Willett MJC (2006). Trans fatty acids and cardiovascular disease. *N Engl J Med* **354**, 1601–1613.
Mozaffarian D, Micha S, Wallace R (2010). Effects on coronary heart disease of increasing polyunsaturated fat in place of saturated fat: a systematic review and meta-analysis of randomized controlled trials. *PLoS Med* **7**, e1000252.
Murray CJ, Dias RH, Kulkarni SC, Lozano R, Stevens GA, Ezzati M (2008). Improving the comparability of diabetes mortality statistics in the US and Mexico. *Diabetes Care* **31**, 451–458.
Murray CJ, Ezzati M, Lopez AD, Rodgers A, Vander Hoorn S (2003). Comparative quantification of health risks conceptual framework and methodological issues. *Popul Health Metr* **1**, 1.
Murray CJ, Kulkarni SC, Ezzati M (2006). Understanding the coronary heart disease versus total cardiovascular mortality paradox: a method to enhance the comparability of cardiovascular death statistics in the United States. *Circulation* **113**, 2071–2081.
Murray CJ, Lopez AD (1997). Alternative projections of mortality and disability by cause 1990–2020: Global Burden of Disease Study. *Lancet* **349**, 1498–1504.
Naghavi M, Makela S, Foreman K, O'Brien J, Pourmalek F, Lozano R (2010). Algorithms for enhancing public health utility of national causes-of-death data. *Popul Health Metr* **8**, 9.
Ni Mhurchu C, Rodgers A, Pan WH, Gu DF, Woodward M (2004). Body mass index and cardiovascular disease in the Asia-Pacific Region: an overview of 33 cohorts involving 310 000 participants. *Int J Epidemiol* **33**, 751–758.
Rosner B, Spiegelman D, Willett WC (1992). Correction of logistic regression relative risk estimates and confidence intervals for random within-person measurement error. *Am J Epidemiol* **136**, 1400–1413.
Smit LA, Mozaffarian D, Willett W (2009). Review of fat and fatty acid requirements and criteria for developing dietary guidelines. *Ann Nutr Metab* **55**, 44–55.
Stampfer MJ, Hu FB, Manson JE, Rimm EB, Willett WC (2000). Primary prevention of coronary heart disease in women through diet and lifestyle. *N Engl J Med* **343**, 16–22.
Stevens G, Dias RH, Thomas KJ, Rivera JA, Carvalho N, Barquera S. *et al*. (2008). Characterizing the epidemiological transition in Mexico: national and subnational burden of diseases, injuries, and risk factors. *PLoS Med* **5**, e125.
Wald NJ, Law MR (2003). A strategy to reduce cardiovascular disease by more than 80%. *BMJ* **326**, 1419.
Willet WC (1998a). Nutritional epidemiology. *Nature of Variation in Diet*. Oxford University Press: Oxford.
Willett WC (1998b). Nutritional epidemiology. Oxford University Press: New York.
World Cancer Research Fund/American Institute for Cancer Research (2007). *Food, Nutrition, Physical Activity, and the Prevention of Cancer: a Global Perspective*. AICR: Washington, DC.
World Health Organization (1998). *The World Health Report 1998. Life in the 21st Century: A Vision for All*. World Health Organization: Geneva.
World Health Organization (2003). Diet, nutrition and the prevention of chronic diseases: report of a joint WHO/FAO expert consultation. *World Health Organ Tech Rep Ser. 916: i-viii*. Geneva: 1–149.
World Health Organization. (2009) 'Mortality Data.' from http://www.who.int/healthinfo/statistics/mortality/en/index.html.
World Health Organization and FAO (2003). *Diet, Nutrition and the Prevention of Chronic Diseases: Report of a Joint WHO/FAO Expert Consultation (Report 916)*. World Health Organization: Geneva.

Supplementary Information accompanies the paper on European Journal of Clinical Nutrition website (http://www.nature.com/ejcn)