# Work Placement Project Portfolio

# 2012


# Jason Wu

# Preface

## Introduction

The aim of this work place portfolio (WPP) was to carry out a systematic review and meta-analysis of the association of dietary seafood, dietary omega-3 polyunsaturated fatty acid (n-3 PUFA), and n-3 PUFA biomarkers with the risk of type 2 diabetes (T2D) in prospective cohort studies. The WPP was conducted between May 2011 and September 2011.

## Student's role

I led this project from start to finish in collaboration with my supervisor and colleagues within my own working group and the Harvard School of Public Health. I designed the study protocol with my supervisor and colleagues through a series of meetings and discussions. I carried out database searches to identify relevant studies; and extracted all data (including extensive communications with study authors to obtain un-published information) required for the systematic review and meta-analysis, the accuracy of which was independently verified by another co-investigator. I designed and input data into Microsoft Excel and Stata and conducted statistical analyses. Finally, working with all co-authors on this project, we interpreted the study findings and I drafted the manuscript for this project which was submitted to a peer-reviewed journal. Through out this process I had regular meeting and sought advice from Dr Dariush Mozaffarian (primary supervisor at my work place) and also received valuable project and statistical support from Dr Renata Micha (co-investigator) and Dr Eric Ding (a friend and an expert of meta-analyses)  Mr Kevin McGeechan (WPP advisor) provided extensive feedback on the final WPP portfolio.

## Reflections on learning

This was my first foray into a systematic review and meta-analysis. The WPP provided a valuable opportunity to consolidate key knowledge and statistical methodologies that I learnt during my BCA course work, but also enabled development of many new skills. I gained an appreciation that a well thought-out and prepared study protocol is a central ingredient for a meta-analysis, which involved multiple elements such as expert knowledge of the subject matter; understanding of study design issues (as many of the weaknesses in the original published studies naturally carries over into the pooled data); and the importance of setting *a priori* goals (e.g. consideration of potential sources of heterogeneity) to avoid multiple hypothesis testing and data dredging. I developed skills in electronic database searching to find all studies relevant to our analyses. I became aware that (sometimes through trial and error) having a carefully planned data extraction database can make a significant difference in saving time and effort (e.g. by collecting all the key data needed for the meta-analysis in the same area, order and orientation in the original Excel data extraction sheet makes it a lot easier for subsequent

importing into Stata). Furthermore, regular updating of records was essential to track correspondence with authors of the identified relevant studies and data received.

I implemented some of the key meta-analysis skills I learnt in my BCA course work (e.g. random and fixed effects meta-analyses, investigation of possible sources of heterogeneity by meta-regression and publication bias by funnel plots), and importantly developed new analytical skills by employing the generalized least-squares trend (GLST) model to allow calculation and pooling of dose-response effects. These efforts also allowed continued improvement in my ability to efficiently use Stata as an powerful analytic software.

## Team work and communication skills

This WPP was the result of direct collaboration with a close group of colleagues and there were multiple opportunities that allowed sharpening of my communication skills. The initial study protocol was developed in conjunction with Drs Mozaffarian and Micha. I was able to develop an effective data request form and directly corresponded with study authors to explain our investigations and obtain unpublished data. During each step of the study identification and data extraction process, I again coordinated with Dr Micha to independently obtain and confirm data accuracy, and ensured we kept to a pre-defined time-line. During the data analysis stage, I regularly presented the latest results in detailed tables and graphs during regular group meetings. These were thoroughly reviewed, and my colleagues provided criticism and feedback, as well as thoughts on additional analyses. These interactions were effective to improve knowledge and skills in systematic review and meta-analysis for myself and those involved in the discussions. The drafting of a manuscript for submission to a peer-reviewed journal also aided development of my writing skills.

## Ethical considerations

All studies that were included in the meta-analysis had human ethics approval by their institutional review committee, and participants gave written informed consent. We gratefully acknowledge the generous assistance of all authors who provided clarification on their published articles and/or additional unpublished data.

# Project report

## Project title

Omega-3 Fatty Acids and incident Type 2 Diabetes: A Systematic Review and Meta-Analysis

## Location and dates

Harvard School of Public Health, Harvard University, USA.

May 2011 – September 2011.

## Context

I am currently working as a post-doctoral fellow at the Harvard School of Public Health, under the supervision of Dr Dariush Mozaffarian. The research for our group focuses on the impact of nutrition and physical activity on cardiovascular diseases. Long chain omega-3 polyunsaturated fatty acids (n-3 PUFA) have been shown to positively influence glucose and insulin homeostasis in a variety of experimental studies. However, whether dietary intake of n-3 PUFA influence risk of type 2 diabetes remain uncertain. We therefore carried out a systematic review and meta-analysis to pool data from prospective cohort studies to address this question.

## Student contribution

I was responsible for designing the study, carried out all statistical analysis and drafted the manuscript for publication. Throughout each stage of the project, I discussed my ideas and work progress with my supervisor and co-investigators (principally Dr Mozaffarian and Dr Renata Micha) and WPP supervisor (Mr Kevin McGeechan).

## Statistical issues

The primary analysis involved using generalized least-squares trend (GLST) model to allow calculation and pooling of dose-response effects. A range of modeling issues were considered such as random vs fixed effect modeling, secondary analyses (e.g. pooling relative risk estimates of the top compared to the bottom category in each study), testing for heterogeneity between study results, investigation of potential sources of heterogeneity by meta-regression, and assessment of publication bias.

## Acknowledgements

I would like to thank all authors who provided clarification on their published articles and/or additional unpublished data. I am also most grateful for all the guidance and feedback I received from Drs Mozaffarian, Micha and Ding, Mr McGeechan and other study collaborators.

## Student declaration

I declare this project is my own work, with direction and assistance provided by Dr Mozaffarian and Mr McGeechan, and I have not previously submitted it for academic credit.

Jason Wu

Date 27.10.11

## Supervisor declaration

I declare that Jason worked independently on this project. I have been very impressed by Jason's ability to apply the statistical methods that he learned in the Master of Biostatistics and his understanding and application of new methods. Jason required minimal input from me on the choice of statistical methods and the interpretation of results.

Kevin McGeechan

Date 27.10.11

# Omega-3 Fatty Acids and incident Type 2 Diabetes: A Systematic Review and Meta-Analysis

## Introduction

Type 2 diabetes mellitus (DM) accounts for 90-95% of all diabetes cases and has reached epidemic proportions globally, including in both developed and developing countries [1]. As a major risk factor for coronary heart disease, stroke, blindness, kidney failure, and peripheral arterial disease, DM poses tremendous public health burdens. Epidemiological and clinical trial evidence demonstrate that lifestyle including diet plays a major role in the development of DM [2]. Further understanding of the role of specific foods and nutrients in the pathogenesis of DM is of paramount importance.

Omega-3 polyunsaturated fatty acids (n-3 PUFA), include eicosapentaenoic acid (EPA, 20:5n-3) and docosahexaenoic acid (DHA, 22:6n-3) from seafood, and alpha-linolenic acid (ALA, 18:3n-3) from plant sources. Based on animal experimental studies, n-3 PUFA improve several metabolic abnormalities underlying the development of DM. Such effects include insulin-sensitizing effects via increased production and secretion of adipocytokines such as adiponectin and leptin [3-6]; and potential prevention of insulin resistance via anti-inflammatory effects mediated directly [7] or through conversion to specialized pro-resolution mediators such as resolvins and protectins [8; 9]. Through modulation of transcription factors (e.g. sterol regulatory element binding protein-1c), n-3 PUFA could also enhance fatty acid oxidation and reduce de novo lipogenesis, effects which could reduce hepatic fat accumulation and preserve hepatic insulin sensitivity [10-13].

Despite metabolic benefits in animal experiments, the impact of n-3 PUFA consumption on risk of DM in humans remain uncertain. In meta-analyses of controlled supplementation trials, n-3 PUFA supplementation does not produce major changes in biomarkers of glucose-insulin homeostasis in subjects with DM [14-16]; similar trials in healthy subjects have reported conflicting findings [17]. In addition to these short-term trials, which generally tested high supplemental doses of n-3 PUFA, several long-term prospective studies have assessed how habitual dietary consumption of n-3 PUFA or seafood, or circulating biomarkers of consumption, relate to incidence of DM, but with mixed findings. Therefore, whether n-3 PUFA influence risk of incident DM and, if so, the direction and magnitude of effect remain unknown. To address these important scientific and public health questions, we carried out a systematic review and meta-analysis of prospective studies that assessed the relation of dietary n-3 PUFA, fish and/or seafood consumption, and biomarker levels of n-3 PUFA with the incidence of DM.

## Methods

*Search Strategy and Eligibility Criteria*

We followed the Meta-analysis of Observational Studies in Epidemiology guidelines for the design, implementation, analysis and reporting of this study [18]. We searched for all prospective cohort studies, including nested prospective studies, that assessed the association of dietary n-3 PUFA intake (ALA, EPA+DHA, or EPA and DHA individually), dietary fish and/or seafood intake, and biomarkers of n-3 PUFA (ALA, EPA+DHA, or EPA and DHA individually) with incidence of DM. Searches were performed electronically through MEDLINE, EMBASE, Latin American and Caribbean Health Sciences Literature (LILACS), related articles, hand-searching of references, and direct author contact. Key words included (among others) *omega-3*, *alpha-linolenic acid*, *eicosapentaenoic acid*, *docosahexaenoic acid*, *fish oils*, *fishes*, *diabetes mellitus*, *cohort studies*, *prospective studies*, and *nested case control*; the full search terms are available on request. Searches included the earliest available online indexing year through to June 30, 2011, with no language restrictions.

Studies were eligible for inclusion if they were prospective cohort studies that provided a multivariate-adjusted effect estimate (odds ratio, relative risk [RR], or hazard ratio) and information about its variance for any of the exposures of interest and incident DM. Exclusion criteria included studies of pregnant women or children (<19y), studies of type 1 diabetes, retrospective case-control studies, cross-sectional and ecological studies, literature reviews, commentaries, editorials, letters, case reports, studies which investigated only shellfish or selected subtypes of fish intake (e.g., fried fish), and studies which provided only crude risk estimates. If multiple manuscripts were published from the same cohort, we included the most up-to-date analyses that had accrued the highest number of DM cases.

*Selection of Articles*

The titles and abstracts of all identified articles were screened by one investigator (J.W.) for eligibility. Two investigators (J.W., R.M.) assessed independently and in duplicate the full texts of the remaining articles to reach a final decision on inclusion or exclusion, with differences resolved by consensus. Of 288 initially identified articles, 262 were excluded based on title and abstract (**Figure 1**). Of 26 full text articles, 10 were excluded because they were duplicate publications from the same cohort (n = 3)[19-21], assessed consumption of restaurant fried fish intake but not overall fish intake (n = 1)[22], lacked RR estimates for fish or n-3 PUFA exposures (n=4)[23-26], or did not separately report results for DM vs. impaired glucose tolerance (n = 2).[27; 28] After final exclusions, 16 studies were identified for inclusion in the meta-analysis [29-44].

*Data Extraction*

For each included study, data were extracted independently and in duplicate by two investigators (J.W., R.M.) using a standardized electronic form, including information on study design, study location (North America, Europe, or Asia/Australia), whether the analysis was pre-specified or post hoc, subject inclusion and exclusion criteria, sample size, subject age, body mass index (BMI), co-morbidities, gender, race, duration of follow-up (mean, median or max number of years), number of events, methods for dietary assessment and diagnosis of DM, laboratory procedure for fatty acid biomarkers (e.g. plasma phospholipid measurements), covariates adjusted for, and multivariable-adjusted risk estimates including data needed to calculate its variance (e.g., CI, SE, or *P*-value). For studies reporting risk across categories of exposure, we further recorded for each study the exposure-category-specific data on person-years of follow up, number of subjects, number of DM cases, median value of exposures, and risk estimate including its variance. For studies reporting on both fish and total seafood intake, we extracted both estimates. When more than one multivariable model was assessed, we extracted risk estimates having the greatest adjustment for potential confounders that did not also include potential intermediates (blood glucose, triglyceride, or inflammatory biomarkers). Whereas no single accepted criteria for grading quality of cohort exists, such grading can be useful for exploring quality and heterogeneity. We assessed quality using previously reported methods by considering 5 criteria: appropriateness and reporting of inclusion and exclusion criteria, methods for assessment of exposure, methods for assessment of outcome, adjustment for confounding, and evidence of bias [45].

Among the 16 studies included in the meta-analysis, we contacted and received responses from authors of 11 studies for relevant missing information, such as baseline characteristics (age, BMI, exposure distributions) or exposure category data (number of participants, person-years of follow-up, number of cases, median level of exposure, or risk estimates and 95% CI).

*Statistical Analysis*

Because nearly all studies reported risk across exposure categories, risk estimates were meta-analyzed using the 2-step generalized least-squares trend (GLST) model [46; 47]. This method utilizes all information from all exposure categories to estimate the log-linear dose-response slope within each study, which are then pooled to derive an overall risk estimate. For studies already reporting risk estimates for linear differences in exposure (e.g., per SD change in biomarker levels[34]) rather than in categories, data were added to the GLST model at the second stage. Because outcomes were relatively rare in all studies, we considered odds ratios and hazard ratios to approximate RRs. Necessary data for the first step of GLST included, for each exposure category in each study, the multivariable-adjusted risk estimate and its corresponding standard errors, person-years of follow-up (for prospective cohort studies) or number of subjects (for nested case-cohort studies), median level of exposure, and number of cases in each exposure category. When category-specific median intakes were not reported, they were imputed based on the midpoint of intake in the range in each category[42], or, for estimating category-specific EPA+DHA intake from fish or seafood intakes [37; 39; 42], by using a conversion factor calculated from nationally

representative dietary surveys as part of our work in the Global Burden of Diseases study [48]. The $I^2$ statistic was used to assess heterogeneity [49]. Fixed effect models were used to pool results with low $I^2$ (<35%), whereas random effect models were used to pool studies with moderate or higher $I^2$ (≥35%). We conducted secondary meta-analyses without GLST, by pooling the risk estimates comparing the highest to the lowest category in each study.

Potential heterogeneity was explored by meta-regression to examine whether risk estimates were significantly different according to several prespecified factors, each modeled individually and then, if significant in univariate models, together with other significant factors to determine independence. Pre-specified potential sources of heterogeneity included age (continuous, or < or ≥ median of all cohorts), BMI (continuous, or < or ≥ median of all cohorts), study quality score (low, 0-3 or high, 4-5), whether the analysis was pre-specified or post hoc (yes or no), and duration of follow-up (continuous, or < or ≥ median of all cohorts). Publication bias was assessed by visual inspection of funnel plots as well as with Egger's test [50; 51]. All analyses were performed using STATA 10.0 (StataCorp, College Station, Tex), with 2-sided alpha=0.05.

## Results

*Study populations*

The 16 studies covered 4 continents, including 9 studies in the US, 5 in Europe, 3 in Asia, and 1 in Australia (**Table 1**). In sum, 25,670 cases of incident diabetes were identified among 540,184 participants. The average participant was middle-aged, although ranges of ages within each cohort were relatively broad. The average follow-up duration in each study varied from 4.0 to 16.7 y among the 15 studies that reported this value. The majority of studies (14 of 16) were judged to be of high quality (quality score 4 or 5). Among dietary exposures, 13 cohorts provided risk estimates for fish and/or seafood, 12 cohorts for EPA+DHA with 4 additional cohorts having data allowing imputation of EPA+DHA, and 7 cohorts for ALA. Among biomarker exposures, 5 cohorts provided risk estimates for EPA alone, DHA alone, and EPA+DHA, and 6 cohorts for ALA. The ranges of dietary and biomarker exposure distributions for each study were broad (**Supplementary Tables 1-2**). All studies provided multivariable-adjusted risk estimates without adjustment for the potential intermediates of blood glucose, triglycerides, or inflammatory biomarkers.

*Fish /Seafood Consumption*

Across 13 cohorts, consumption of fish and/or seafood was not significantly associated with incidence of DM (per 100g/day, RR=1.12, 95% CI=0.94, 1.34, P=0.21, **Figure 2**). Substantial between-study heterogeneity was evident ($I^2$=82.9%). In multivariable meta-regression including both BMI and mean follow-up duration as predictors, follow-up duration was identified as a significant modifier of the association between fish/seafood intake and risk of DM (*P*=0.03). Among cohorts with follow-up ≥10.2y (median follow-up of all included cohorts), consumption of fish/seafood was associated with 46% higher risk of DM (per 100g/day, RR=1.46, 95% CI=1.23, 1.73, *P*<0.001), but still with substantial unexplained heterogeneity across studies ($I^2$=65.5%) (**Supplementary Figure 1**). In contrast, among cohorts with follow-up <10.2y, fish/seafood consumption was not associated with the risk of DM (per 100g/day, RR=0.88, 95% CI=0.78, 1.01, *P*=0.06, $I^2$=29.7%, Supplementary Figure 1).

Meta-analysis of risk estimates comparing the top to the bottom exposure category in each study gave similar results (RR=1.07, 95% CI=0.94, 1.22, P=0.29, $I^2$=77.7%). Follow-up duration was again the only factor identified as a significant modifier of the association between fish/seafood intake and risk of DM (*P*=0.001). Overall findings were similar when restricted to the 8 risk estimates for fish consumption alone or the 8 risk estimates for seafood consumption alone (data not shown).

*Dietary EPA and DHA*

Across 16 cohorts, estimated EPA+DHA consumption was not associated with DM risk (per 250mg/day, RR=1.04, 95% CI = 0.97-1.10, P = 0.27, **Figure 3**). Substantial heterogeneity was evident ($I^2$=82%), which in

meta-regression analyses again appeared related to varying duration of follow-up ($P$ = 0.03, **Supplementary Figure 2**). In cohorts with follow-up >9.6y (the median follow-up of all included cohorts), EPA+DHA consumption was associated with 12% higher risk of DM (per 250mg/day, RR=1.12, 95% CI=1.04, 1.20, $I^2$=71%) whereas in cohorts with follow-up <9.6y, EPA+DHA consumption was not associated with DM risk (per 250mg/day, RR=0.96, 95% CI=0.90, 1.03, $I^2$=56%). Analysis by pooling RR for the top compared to the bottom exposure category in each study gave similar results (RR=1.04, 95% CI=0.94, 1.16, $P$=0.46, $I^2$=79%).

Among the 5 cohorts reporting risk estimates for estimated dietary EPA or DHA separately, neither EPA (per 125mg/day, RR=1.07, 95% CI=0.85, 1.34, $P$=0.58) nor DHA (per 125mg/day, RR=1.04, 95% CI=0.90, 1.21, $P$=0.59) were associated with incidence of DM, although substantial heterogeneity was present for both estimates ($I^2$=77.4% and 80.2%, respectively).

*Circulating EPA+DHA*

Among 5 cohorts that evaluated circulating EPA+DHA biomarkers, no association was seen between EPA+DHA concentrations and incidence of DM (per 3% of total fatty acids, RR=0.94, 95% CI=0.75, 1.17, $P$=0.56, **Figure 4**). Moderate heterogeneity was evident ($I^2$=40.4%). Meta-regression did not identify any statistically significant sources of heterogeneity, although statistical power was limited due to only 5 studies. Meta-analysis of risk estimates for the top compared to the bottom exposure category in each study gave similar results (RR=0.94, 95% CI=0.82, 1.09, $I^2$=47%). When circulating EPA and DHA were evaluated separately, neither EPA (per 1% of total fatty acids, RR= 0.96, 95% CI=0.86, 1.07, $P$=0.48, $I^2$=0%) nor DHA (per 1% of total fatty acids, RR=1.00, 95% CI=0.91, 1.10, $P$=0.97, $I^2$=0%) were associated with risk of DM.

*Dietary ALA*

Across 7 cohort studies, estimated ALA consumption was not associated with risk of DM (per 0.5g/day, RR=0.93, 95% CI=0.83, 1.04, $P$=0.20, **Figure 5**). Heterogeneity was present ($I^2$=54.3%), but no statistically significant sources of heterogeneity were identified by meta-regression, although power was limited. Similar results were obtained in meta-analyses of risk estimates comparing the top to the bottom exposure category in each study (RR=0.92, 95% CI=0.80, 1.05, $P$=0.21, $I^2$=40.3%).

*Circulating ALA*

In 6 cohorts that assessed circulating ALA biomarkers, higher ALA levels were not associated risk of DM (per 0.1% of total fatty acids, RR=0.90, 95% CI=0.80, 1.00, $P$=0.06, **Figure 6**). Between-study heterogeneity was not evident ($I^2$ =17.9%). Results were similar for pooling RR's in the top compared to the bottom category of exposure in each study (RR=0.87, 95% CI=0.74, 1.01, $P$=0.07, $I^2$=18.9%).

*Publication bias*

For dietary fish/seafood intake, visual inspection of the funnel plot found that individual study RR estimates were reasonably symmetrical about the pooled effect estimate, suggesting no evidence of publication bias (**Supplementary Figure 3**, top panel). This was supported by a null Egger's test for publication bias ($P = 0.55$). For dietary EPA+DHA intake, the funnel plot showed slightly more data points among the smaller studies to the left of the pooled estimate of RR, indicating possible publication bias in favor of a protective association (**Supplementary Figure 3**, bottom panel). Egger's test ($P = 0.60$) suggested no significant evidence of publication bias. There was also little evidence for publication bias for all other exposure-outcome relationships based on either visual inspection of funnel plots or Egger's test (data not shown).

## Discussion

Findings from this systematic review and meta-analysis suggest that dietary EPA+DHA and fish/seafood consumption do not have either major harmful or beneficial associations with the development of DM. However, we identified substantial heterogeneity in findings across individual studies. Our findings also suggest that plant-derived ALA could be protective.

*Fish/seafood intake, n-3 PUFA, and DM risk*

We found little evidence that fish or seafood, dietary EPA+DHA, or circulating EPA+DHA biomarkers were associated with risk of DM. These studies typically comprised moderately overweight but otherwise generally healthy participants at baseline. Prior short-term randomized controlled trials have found very little effect of n-3 PUFA supplementation on glucose metabolism or indices of insulin resistance in healthy subjects [5; 52-54]. Therefore, our findings are consistent with these prior metabolic trials and suggest that, at typical dietary levels of consumption in generally healthy subjects, fish, seafood, or dietary EPA+DHA have minimal effects on development of DM.

Our findings suggest there is large heterogeneity between study results. Such heterogeneity could be due to chance, or, related to variations in population characteristics, geographical locations, or study methodology. We did not observe significant association between most of our *a priori* selected potential sources of heterogeneity including age, BMI, study quality score, and whether analyses were pre-specified or post-hoc. EPA+DHA and fish/seafood consumption were associated with higher risk of DM among studies with mean/median follow-up duration >~10y, and trends toward lower risk of DM among studies with median follow-up duration < 10y. Such potential time varying effects could represent true biologic effects. For example, some experimental studies suggest environmental contaminants in fish/seafood could adversely affect metabolic pathways underlying the development of DM [55-57], and it could be hypothesized that the effect may become more evident over prolonged periods of time and counteract any potential metabolic benefits of EPA+DHA. However, data on effects of environmental contaminants found in fish/seafood on glucose and insulin homeostasis remain very limited. From a methodological perspective, nearly all of these cohorts assessed diet or biomarkers only once at baseline. Thus, the most accurate estimates of relations with disease should be in the initial years of follow-up, as with longer follow-up, unmeasured dietary changes will lead to exposure misclassification over time. Overall, this secondary finding may be due to chance and also should be interpreted with caution as the analysis was based on average study-level, rather than individual-level, follow-up duration., Our findings suggest that future studies should assess duration of follow-up at studies or an individual-level to as a potential effect modifier of the relation between fish, EPA+DHA, and incident DM.

*ALA and DM risk*

Compared to seafood sources, we identified fewer prospective studies of ALA and incident DM. We found that both estimated dietary consumption and circulating biomarkers of ALA were not associated with risk of DM. However, the risk estimate and 95% confidence interval also do not exclude a possible, minor benefit of ALA. These pooled estimates demonstrated modest or low heterogeneity between-studies, suggesting relatively consistent findings among studies. In several animal models, dietary ALA or flaxseed oil (a rich source of ALA) improved insulin sensitivity and glycemic responses [58-60]. Similarly, some [61; 62], but not other [54], short-term randomized clinical trials found that ALA or flaxseed oil moderately improved fasting plasma glucose and markers of insulin resistance in humans. Our current results add to these limited but potentially important data that ALA may provide moderate protection against the development of DM. Because plant sources of n-3 PUFA are potentially more widely available on a global basis, our findings highlight the need for further clinical and observational investigation of these effects.

*Strength and limitations*

Our analysis had several strengths. Multiple databases were systematically searched to ensure identification of all relevant published studies, and we also obtained clarification and/or additional data from several authors to minimize misclassification and potential for publication bias. The prospective cohort design of the included studies reduces the possibility of recall and selection bias. The large number of total incident DM cases provided statistical power to detect clinically meaningful associations. We pooled all available information from each study using GLST, which accounts for both categories of exposure and dose-response, rather than simply pooling often incomparable single extreme categories. We evaluated both dietary estimates and objective biomarkers of n-3 PUFA, and consistent findings for each increased confidence in validity of our findings. Studies were identified from wide geographical locations (North America, Europe, Asia, Australia) with varied population characteristics (age, BMI, sex, exposure distributions), increasing generalizability.

Potential limitations should be considered. Although most studies adjusted for major sociodemographic, lifestyle, clinical, and other dietary risk factors for DM, the possibility of residual confounding by unmeasured or imprecisely measured factors remains a possibility. Measurement error in exposures could have attenuated associations towards the null. On the other hand, these same methods have identified significant associations of fish or n-3 PUFA with several other disease outcomes, and we also evaluated both dietary estimates and n-3 PUFA biomarkers that are subject to differing types of error. Most studies evaluated total fish and/or seafood consumption, and we were unable to evaluate specific fish species or preparation methods which might influence effects on DM. As in all meta-analyses, publication bias is possible. However, visual inspection of funnel plots did not suggest publication bias was a substantial problem; our direct contact with authors and experts minimized the possibility of missing unpublished studies; and the overall null pooled findings are typically less subject to influence from publication bias than if we had found significant positive relations.

*Conclusions*

Based on all available evidence from prospective studies, neither EPA+DHA nor fish/seafood intake have significant associations with risk of DM, and plant-derived ALA was also not associated with DM risk.

## Acknowledgements

We thank the following authors for providing clarification on their published articles and/or additional unpublished data: Aaron Folsom, Katie Meyer, Lu Wang, Allison Hodge, Benno Krachler, Nita Forouhi, Pinal Patel, Geertuida van Woudenbergh, Andrew Odegaard, Janine Kröger, Akiko Nanri, and Raquel Villegas. We would also like to thank the NHLBI, NIH and National Heart Foundation of Australia for funding support.

## References


1. Steyn NP, Mann J, Bennett PH *et al.* (2004) Diet, nutrition and the prevention of type 2 diabetes. *Public Health Nutr* **7**, 147-165.

2. Schulze MB & Hu FB (2005) Primary prevention of diabetes: what can be done and how much can be prevented? *Annu Rev Public Health* **26**, 445-467.

3. Banga A, Unal R, Tripathi P *et al.* (2009) Adiponectin translation is increased by the PPARgamma agonists pioglitazone and omega-3 fatty acids. *Am J Physiol Endocrinol Metab* **296**, E480-489.

4. Vemuri M, Kelley DS, Mackey BE *et al.* (2007) Docosahexaenoic Acid (DHA) But Not Eicosapentaenoic Acid (EPA) Prevents Trans-10, Cis-12 Conjugated Linoleic Acid (CLA)-Induced Insulin Resistance in Mice. *Metab Syndr Relat Disord* **5**, 315-322.

5. Itoh M, Suganami T, Satoh N *et al.* (2007) Increased adiponectin secretion by highly purified eicosapentaenoic acid in rodent models of obesity and human obese subjects. *Arterioscler Thromb Vasc Biol* **27**, 1918-1925.

6. Perez-Matute P, Perez-Echarri N, Martinez JA *et al.* (2007) Eicosapentaenoic acid actions on adiposity and insulin resistance in control and high-fat-fed rats: role of apoptosis, adiponectin and tumour necrosis factor-alpha. *Br J Nutr* **97**, 389-398.

7. Oh DY, Talukdar S, Bae EJ *et al.* (2010) GPR120 is an omega-3 fatty acid receptor mediating potent anti-inflammatory and insulin-sensitizing effects. *Cell* **142**, 687-698.

8. Gonzalez-Periz A, Horrillo R, Ferre N *et al.* (2009) Obesity-induced insulin resistance and hepatic steatosis are alleviated by omega-3 fatty acids: a role for resolvins and protectins. *Faseb J* **23**, 1946-1957.

9. Serhan CN, Chiang N & Van Dyke TE (2008) Resolving inflammation: dual anti-inflammatory and pro-resolution lipid mediators. *Nat Rev Immunol* **8**, 349-361.

10. Tanaka N, Zhang X, Sugiyama E *et al.* (2010) Eicosapentaenoic acid improves hepatic steatosis independent of PPARalpha activation through inhibition of SREBP-1 maturation in mice. *Biochem Pharmacol* **80**, 1601-1612.

11. Neschen S, Morino K, Dong J *et al.* (2007) n-3 Fatty acids preserve insulin sensitivity in vivo in a peroxisome proliferator-activated receptor-alpha-dependent manner. *Diabetes* **56**, 1034-1041.

12. Sato A, Kawano H, Notsu T *et al.* (2010) Antiobesity effect of eicosapentaenoic acid in high-fat/high-sucrose diet-induced obesity: importance of hepatic lipogenesis. *Diabetes* **59**, 2495-2504.

13. Jump DB (2011) Fatty acid regulation of hepatic lipid metabolism. *Curr Opin Clin Nutr Metab Care* **14**, 115-120.

14. Friedberg CE, Janssen MJ, Heine RJ *et al.* (1998) Fish oil and glycemic control in diabetes. A meta-analysis. *Diabetes Care* **21**, 494-500.

15. Hartweg J, Perera R, Montori V *et al.* (2008) Omega-3 polyunsaturated fatty acids (PUFA) for type 2 diabetes mellitus. *Cochrane Database Syst Rev*, CD003205.

16. Montori VM, Farmer A, Wollan PC *et al.* (2000) Fish oil supplementation in type 2 diabetes: a quantitative systematic review. *Diabetes Care* **23**, 1407-1415.

17. Fedor D & Kelley DS (2009) Prevention of insulin resistance by n-3 polyunsaturated fatty acids. *Curr Opin Clin Nutr Metab Care* **12**, 138-146.

18. Stroup DF, Berlin JA, Morton SC *et al.* (2000) Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA* **283**, 2008-2012.

19. Salmeron J, Hu FB, Manson JE *et al.* (2001) Dietary fat intake and risk of type 2 diabetes in women. *Am J Clin Nutr* **73**, 1019-1026.

20. Schulze MB, Manson JE, Willett WC *et al.* (2003) Processed meat intake and incidence of Type 2 diabetes in younger and middle-aged women. *Diabetologia* **46**, 1465-1473.

21. Song Y, Manson JE, Buring JE *et al.* (2004) A prospective study of red meat consumption and type 2 diabetes in middle-aged and elderly women: the women's health study. *Diabetes Care* **27**, 2108-2115.

22. Krishnan S, Coogan PF, Boggs DA *et al.* (2010) Consumption of restaurant foods and incidence of type 2 diabetes in African American women. *Am J Clin Nutr* **91**, 465-471.

23. Feskens EJ, Virtanen SM, Rasanen L *et al.* (1995) Dietary factors determining diabetes and impaired glucose tolerance. A 20-year follow-up of the Finnish and Dutch cohorts of the Seven Countries Study. *Diabetes Care* **18**, 1104-1112.

24. Vessby B, Aro A, Skarfors E *et al.* (1994) The risk to develop NIDDM is related to the fatty acid composition of the serum cholesterol esters. *Diabetes* **43**, 1353-1357.

25. Colditz GA, Manson JE, Stampfer MJ *et al.* (1992) Diet and risk of clinical diabetes in women. *Am J Clin Nutr* **55**, 1018-1023.

26. Harding AH, Day NE, Khaw KT *et al.* (2004) Dietary fat and the risk of clinical type 2 diabetes: the European prospective investigation of Cancer-Norfolk study. *Am J Epidemiol* **159**, 73-82.

27. Feskens EJ, Bowles CH & Kromhout D (1991) Inverse association between fish intake and risk of glucose intolerance in normoglycemic elderly men and women. *Diabetes Care* **14**, 935-941.

28. Laaksonen DE, Lakka TA, Lakka HM *et al.* (2002) Serum fatty acid composition predicts development of impaired fasting glycaemia and diabetes in middle-aged men. *Diabet Med* **19**, 456-464.

29. Brostow DP, Odegaard AO, Koh WP *et al.* (2011) Omega-3 fatty acids and incident type 2 diabetes: the Singapore Chinese Health Study. *Am J Clin Nutr* **In press**.

30. Djousse L, Biggs ML, Lemaitre RN *et al.* (2011) Plasma omega-3 fatty acids and incident diabetes in older adults. *Am J Clin Nutr* **94**, 527-533.

31. Djousse L, Gaziano JM, Buring JE *et al.* (2011) Dietary omega-3 fatty acids and fish consumption and risk of type 2 diabetes. *Am J Clin Nutr* **93**, 143-150.

32. Hodge AM, English DR, O'Dea K *et al.* (2007) Plasma phospholipid and dietary fatty acids as predictors of type 2 diabetes: interpreting the role of linoleic acid. *Am J Clin Nutr* **86**, 189-197.

33. Kaushik M, Mozaffarian D, Spiegelman D *et al.* (2009) Long-chain omega-3 fatty acids, fish intake, and the risk of type 2 diabetes mellitus. *Am J Clin Nutr* **90**, 613-620.

34. Krachler B, Norberg M, Eriksson JW *et al.* (2008) Fatty acid profile of the erythrocyte membrane preceding development of Type 2 diabetes mellitus. *Nutr Metab Cardiovasc Dis* **18**, 503-510.

35. Kroger J, Zietemann V, Enzenbach C *et al.* (2011) Erythrocyte membrane phospholipid fatty acids, desaturase activity, and dietary fatty acids in relation to risk of type 2 diabetes in the European Prospective Investigation into Cancer and Nutrition (EPIC)-Potsdam Study. *Am J Clin Nutr* **93**, 127-142.

36. Meyer KA, Kushi LH, Jacobs DR, Jr. *et al.* (2001) Dietary fat and incidence of type 2 diabetes in older Iowa women. *Diabetes Care* **24**, 1528-1535.

37. Nanri A, Mizoue T, Noda M *et al.* (2011) Fish intake and type 2 diabetes in Japanese men and women: the Japan Public Health Center-based Prospective Study. *Am J Clin Nutr* **94**, 884-891.

38. Patel PS, Sharp SJ, Jansen E *et al.* (2010) Fatty acids measured in plasma and erythrocyte-membrane phospholipids and derived by food-frequency questionnaire and the risk of new-onset type 2 diabetes: a pilot study in the European Prospective Investigation into Cancer and Nutrition (EPIC)-Norfolk cohort. *Am J Clin Nutr* **92**, 1214-1222.

39. Patel PS, Sharp SJ, Luben RN *et al.* (2009) Association between type of dietary fish and seafood intake and the risk of incident type 2 diabetes: the European prospective investigation of cancer (EPIC)-Norfolk cohort study. *Diabetes Care* **32**, 1857-1863.

40. van Dam RM, Willett WC, Rimm EB *et al.* (2002) Dietary fat and meat intake in relation to risk of type 2 diabetes in men. *Diabetes Care* **25**, 417-424.

41. van Woudenbergh GJ, van Ballegooijen AJ, Kuijsten A *et al.* (2009) Eating fish and risk of type 2 diabetes: A population-based, prospective follow-up study. *Diabetes Care* **32**, 2021-2026.

42. Vang A, Singh PN, Lee JW *et al.* (2008) Meats, processed meats, obesity, weight gain and occurrence of diabetes among adults: findings from Adventist Health Studies. *Ann Nutr Metab* **52**, 96-104.

43. Villegas R, Xiang YB, Elasy T *et al.* (2011) Fish, shellfish, and long-chain n-3 fatty acid consumption and risk of incident type 2 diabetes in middle-aged Chinese men and women. *Am J Clin Nutr* **94**, 543-551.

44. Wang L, Folsom AR, Zheng ZJ *et al.* (2003) Plasma fatty acid composition and incidence of diabetes in middle-aged adults: the Atherosclerosis Risk in Communities (ARIC) Study. *Am J Clin Nutr* **78**, 91-98.

45. 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.

46. Greenland S & Longnecker MP (1992) Methods for trend estimation from summarized dose-response data, with applications to meta-analysis. *Am J Epidemiol* **135**, 1301-1309.

47. Orsini N, Bellocco R & Greenland S (2006) Generalized least squares for trend estimation of summarized dose-response data. *Stata Journal* **6**, 40-57.

48. Micha R, Kalantarian S, Wirojratana P *et al.* (2011) Estimating the global and regional burden of suboptimal nutrition on chronic disease: methods and inputs to the analysis. *Eur J Clin Nutr*.

49. Higgins JP, Thompson SG, Deeks JJ *et al.* (2003) Measuring inconsistency in meta-analyses. *BMJ* **327**, 557-560.

50. Begg CB & Mazumdar M (1994) Operating characteristics of a rank correlation test for publication bias. *Biometrics* **50**, 1088-1101.

51. Egger M, Davey Smith G, Schneider M *et al.* (1997) Bias in meta-analysis detected by a simple, graphical test. *BMJ* **315**, 629-634.

52. Griffin MD, Sanders TA, Davies IG *et al.* (2006) Effects of altering the ratio of dietary n-6 to n-3 fatty acids on insulin sensitivity, lipoprotein size, and postprandial lipemia in men and postmenopausal women aged 45-70 y: the OPTILIP Study. *Am J Clin Nutr* **84**, 1290-1298.

53. Giacco R, Cuomo V, Vessby B *et al.* (2007) Fish oil, insulin sensitivity, insulin secretion and glucose tolerance in healthy people: is there any effect of fish oil supplementation in relation to the type of background diet and habitual dietary intake of n-6 and n-3 fatty acids? *Nutr Metab Cardiovasc Dis* **17**, 572-580.

54. Egert S, Fobker M, Andersen G *et al.* (2008) Effects of dietary alpha-linolenic acid, eicosapentaenoic acid or docosahexaenoic acid on parameters of glucose metabolism in healthy volunteers. *Ann Nutr Metab* **53**, 182-187.

55. Chang JW, Chen HL, Su HJ *et al.* (2011) Simultaneous exposure of non-diabetics to high levels of dioxins and mercury increases their risk of insulin resistance. *J Hazard Mater* **185**, 749-755.

56. Rignell-Hydbom A, Lidfeldt J, Kiviranta H *et al.* (2009) Exposure to p,p'-DDE: a risk factor for type 2 diabetes. *PLoS One* **4**, e7503.

57. Grandjean P, Henriksen JE, Choi AL *et al.* (2011) Marine food pollutants as a risk factor for hypoinsulinemia and type 2 diabetes. *Epidemiology* **22**, 410-417.

58. Ghafoorunissa, Ibrahim A & Natarajan S (2005) Substituting dietary linoleic acid with alpha-linolenic acid improves insulin sensitivity in sucrose fed rats. *Biochim Biophys Acta* **1733**, 67-75.

59. Mustad VA, Demichele S, Huang YS *et al.* (2006) Differential effects of n-3 polyunsaturated fatty acids on metabolic control and vascular reactivity in the type 2 diabetic ob/ob mouse. *Metabolism* **55**, 1365-1374.

60. Kelley DS, Vemuri M, Adkins Y *et al.* (2009) Flaxseed oil prevents trans-10, cis-12-conjugated linoleic acid-induced insulin resistance in mice. *Br J Nutr* **101**, 701-708.

61. Wendland E, Farmer A, Glasziou P *et al.* (2006) Effect of alpha linolenic acid on cardiovascular risk markers: a systematic review. *Heart* **92**, 166-169.

62. Bloedon LT, Balikai S, Chittams J *et al.* (2008) Flaxseed and cardiovascular risk factors: results from a double blind, randomized, controlled clinical trial. *J Am Coll Nutr* **27**, 65-74.

63. Huedo-Medina TB, Sanchez-Meca J, Marin-Martinez F *et al.* (2006) Assessing heterogeneity in meta-analysis: Q statistic or I2 index? *Psychol Methods* **11**, 193-206.

64. Harbord RM & Higgins JP (2008) Meta-regression in Stata. *The Stata Journal* **8**, 493-519.

65. Thompson SG & Higgins JP (2002) How should meta-regression analyses be undertaken and interpreted? *Stat Med* **21**, 1559-1573.

66. Sterne JA, Gavaghan D & Egger M (2000) Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. *J Clin Epidemiol* **53**, 1119-1129.

## Figure legend

**Figure 1.** Search, screening, and selection process of prospective cohort studies of dietary n-3 PUFA, fish/seafood, and circulating n-3 PUFA biomarkers and risk of type 2 diabetes.

**Figure 2.** Relative risk of type 2 diabetes according to fish and/or seafood consumption in 13 prospective cohorts including 481,489 participants and 20,830 cases of incident diabetes. Within-study relative risks and 95% CI's were quantified using generalized least squares trend estimation, and study-specific results were pooled using random effects meta-analysis. For 3 cohorts reporting effect estimates for both fish and seafood intake,[29; 43] effects estimates for fish were used in the primary analysis; findings using seafood did not appreciably alter the results (not shown).

**Figure 3.** Relative risk of type 2 diabetes according to estimated dietary EPA+DHA in 16 prospective cohorts including 440,873 participants and 21,512 cases of incident diabetes. Within-study relative risks and 95% CI's were quantified using generalized least squares trend estimation, and study-specific results were pooled using random effects meta-analysis. Exclusion of 4 studies for which the EPA+DHA intake was imputed from dietary fish/seafood [37; 39; 42], did not appreciably alter the results (per 250mg/day, RR=1.06, 95% CI=0.99, 1.14, *P*=0.08).

**Figure 4.** Relative risk of type 2 diabetes according to EPA+DHA biomarker level (as % of total fatty acids) in 5 prospective cohorts including 10382 individuals and 1581 incident DM cases. Within-study relative risks and 95% CI's were quantified using generalized least squares trend estimation, and study-specific results were pooled using random effects meta-analysis. Pooling results using the alternative biomarker measurements (red blood cell phospholipid) in Patel et al [38], did not appreciably alter the results (per 3% of total fatty acids, RR = 0.96, 95% CI = 0.76-1.21). The study by Wang et al [44], was not included due to insufficient information to allow extraction of effect estimates. However, EPA and DHA were not associated with risk of diabetes [44], and inclusion of this study is unlikely to appreciably alter the observed results.

**Figure 5.** Relative risk of type 2 diabetes according to dietary ALA in 7 prospective cohorts including 131940 individuals and 7365 incident DM cases. Within-study relative risks and 95% CI's were quantified using generalized least squares trend estimation, and study-specific results were pooled using random effects meta-analysis.

**Figure 6.** Relative risk of type 2 diabetes according to ALA biomarker (as % of total fatty acids) in 6 prospective cohorts including 13291 individuals and 1833 incident DM cases. Within-study relative risks and 95% CI's were quantified using generalized least squares trend estimation, and study-specific results were pooled using random effects meta-analysis. Pooling results with the alternative compartment of biomarker measurement in Wang et a

[44]l, and Patel et al [38], did not appreciably alter the results (per 0.1% of total fatty acids, RR = 0.93, 95% CI = 0.85-1.02, P = 0.11).

**Supplementary Figure 1.** Relative risk of type 2 diabetes according to fish and/or seafood consumption in 13 prospective cohorts, stratified by mean/median follow-up of included cohorts (10.2years). Mean or median follow-up were available and extracted from all studies. Within-study relative risks and 95% CI's were quantified using generalized least squares trend estimation, and study-specific results were pooled using random effects meta-analysis.

**Supplementary Figure 2.** Relative risk of type 2 diabetes according to estimated dietary EPA+DHA in 16 prospective cohorts, stratified by mean/median follow-up of included cohorts (9.6years). Mean or median follow-up were available for all studies except Meyer et al [36], for which the maximum number of years of follow-up (11y) was used. Within-study relative risks and 95% CI's were quantified using generalized least squares trend estimation, and study-specific results were pooled using random effects meta-analysis. Exclusion of the study by Meyer et al did not appreciably alter the results.

**Supplementary Figure 3.** Funnel plot of studies of fish/seafood intake (top panel, 13 cohorts) and of EPA+DHA intake (bottom panel, 16 cohorts), with risk of type 2 diabetes. The RR estimate (x-axis) is plotted against the standard error of the log of the RR estimates (y-axis) of each cohort. The vertical line represents the fixed-effect pooled estimate of RRs; and the dashed lines indicates the expected 95% CI for a given standard error under the assumption of no heterogeneity between studies.

**Table 1. Characteristics of the identified 16 studies including 18 prospective cohorts that evaluated dietary n-3 PUFA, fish/seafood consumption, or biomarker n-3 PUFA levels and incidence of type 2 diabetes***

| Author (Year) | Study name (Country) | Total N/number of DM cases | Age, y, mean or range | Mean BMI, (kg/m$^2$) | Men, % | Follow-up years | Exposures† | Diabetes ascertainment | Adjustment‡ | Quality score§ | Additional data requested / provided‖ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Meyer (2001)[36] | Iowa Women's Health Study (USA) | 35,988 / 1,890 | 55-69 | 26.7 | 0 | Max: 11 | Diet<br>EPA+DHA | Self-report. Validated in sub-cohort by physician diagnosis | +++ | 3 | Yes / No |
| van Dam (2002)[40] | Health Professional Follow-up Study (HPFS) (USA) | 42,504 / 1,321 | 40-75 | 25.3 | 100 | Mean: 11 | Diet<br>ALA | Self-report confirmed by supplemental questionnaire | +++ | 5 | No / - |
| Wang (2003)[44] | Atherosclerosis Risk in Communities Study (USA) | 2,909 / 252 | 54 | 26.7 | 46.9 | Mean: 8.1 | Biomarker<br>ALA | Self-reported diabetes medication use or physician diagnosis or fasting or non fasting glucose | ++ | 4 | Yes / Yes |
| Hodge (2007)[32] | Melbourne Collaborative Cohort Study (Australia) | ¶3,737 / 346 | 55 | 27 | 44.1 | Mean: 4.1 | Diet<br>ALA<br>EPA<br>DHA<br>EPA+DHA<br><br>Biomarker<br>ALA<br>EPA<br>DHA<br>EPA+DHA | Self-report confirmed by subject's physician. | ++ | 4 | Yes / Yes |
| Krachler (2008)[34] | Vasterbotten Intervention Program (Sweden) | ¶450 / 159 | 52 | 27 | 58.4 | Mean: 8.8 | Biomarker<br>ALA<br>EPA<br>DHA<br>EPA+DHA | Subjects attended health exam at local primary care center, diabetes confirmed | ++ | 4 | Yes / Yes |

| | | |
|---|---|---|
| | according to WHO 1998 criteria. | |

| Author (Year) | Study name (Country) | Total N/number of DM cases | Age, y, mean or range | Mean BMI, (kg/m$^2$) | Men, % | Follow-up years | Exposures† | Diabetes ascertainment | Adjustment‡ | Quality score§ | Additional data requested / provided‖ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Vang (2008)[42] | Adventist Mortality Study and Adventist Health Study (USA) | 8,401 / 531 | 65 | 24.5 | 61 | Mean: 16 | Diet<br>Seafood | Self-report. | + | 1 | Yes / No |
| Kaushik (2009)[33] | Nurses' Health Study (NHS) (USA) | 61,031 / 4,159 | 56 | 25 | 0 | Mean: 16.7 | Diet<br>Fish<br>EPA+DHA | Self-report confirmed by supplementary questionnaire. | +++ | 5 | No / No |
| Kaushik (2009)[33] | Nurses' Health Study 2(NHS2) (USA) | 91,669 / 2,728 | 36 | 24.5 | 0 | Mean: 13.7 | Diet<br>Fish<br>EPA+DHA | Self-report confirmed by supplementary questionnaire. | +++ | 5 | No / No |
| Kaushik (2009)[33] | Health Professional Follow-up Study (HPFS) (USA) | 42,504 / 2493 | 53 | 25.4 | 100 | Mean: 16.0 | Diet<br>Fish<br>EPA+DHA | Self-report confirmed by supplementary questionnaire. | +++ | 5 | No / No |
| Patel (2009)[39] | EPIC-Norfolk (UK) | 21,984 / 725 | 58 | 26.3 | 44.6 | Median: 10.2 | Diet<br>Seafood | Self-report confirmed by linkage with health registries or cases detected by linkage to health registries alone. | +++ | 5 | Yes / Yes |
| van Woudenbergh (2009)[41] | Rotterdam study (Netherlands) | 4,472 / 463 | 67 | 26.2 | 41 | Median: 12.4 | Diet<br>Fish<br>EPA<br>DHA<br>EPA+DHA | Diabetes status confirmed by general practitioners according to WHO 1999 criteria. | +++ | 4 | Yes / Yes |

| Author (Year) | Study name (Country) | Total N/number of DM cases | Age, y, mean or range | BMI, kg/m$^2$, mean | Men, % | Follow-up years | Exposures† | Diabetes ascertainment | Adjustment‡ | Quality score§ | Additional data requested / provided‖ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Patel (2010)[38] | EPIC-Norfolk (UK) | ¶383 / 199 | 64 | 28.1 | 53.3 | Mean: 10.3 | Diet<br>ALA<br>EPA<br>DHA<br><br>Biomarker<br>ALA<br>EPA<br>DHA<br>EPA+DHA | Self-report confirmed by linkage with health registries. | ++ | 4 | Yes / Yes |
| Brostow (2011)[29] | Singapore Chinese Health Study (Singapore) | 43,176 / 2,252 | 55 | 23 | 42.4 | Mean: 5.7 | Diet<br>Fish<br>Seafood<br>ALA<br>EPA+DHA | Self-report confirmed by linkage with hospital records or supplementary questionnaire | +++ | 5 | Yes / Yes |
| Djousse (2011)[31] | Women's Health Study (WHS) (USA) | 36,328 / 2,370 | 55 | 25.9 | 0 | Mean: 12.4 | Diet<br>Fish<br>ALA<br>EPA<br>DHA<br>EPA+DHA | Self-report confirmed by telephone interview or supplemental questionnaire | +++ | 5 | Yes / Yes |
| Djousse (2011)[30] | Cardiovascular Health Study (CHS) (USA) | 3,088 / 204 | 75 | 26.4 | 38.9 | Mean: 9.6 | Diet<br>Fish<br>ALA<br>EPA+DHA<br><br>Biomarker<br>ALA<br>EPA<br>DHA<br>EPA+DHA | Diabetes medication use, or fasting or nonfasting glucose level | +++ | 4 | Yes / Yes |

| Author (Year) | Study name (Country) | Total N/number of DM cases | Age, y, mean or range | Mean BMI, (kg/m$^2$) | Men, % | Follow-up years | Exposures† | Diabetes ascertainment | Adjustment‡ | Quality score§ | Additional data requested / provided‖ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Kroger (2011)[35] | EPIC-Potsdam (Germany) | 2,724 / 673 | 51 | 26.9 | 38.7 | Mean: 6.3 | Diet<br>ALA<br>EPA<br>DHA<br>EPA+DHA<br><br>Biomarker<br>ALA<br>EPA<br>DHA<br>EPA+DHA | Self-report confirmed by diagnosing physician | +++ | 4 | Yes / Yes |
| Nanri (2011)[37] | JPHCM (Japan) | 22,921 / 572 | 56 | 23.6 | 100 | Mean: 5 | Diet<br>Seafood | Self-report confirmed by medical records | +++ | 5 | No / - |
| Nanri (2011)[37] | JPHCW (Japan) | 29,759 / 399 | 56 | 23.4 | 0 | Mean: 5 | Diet<br>Seafood | Self-report confirmed by medical records | +++ | 5 | No / - |
| Villegas (2011)[43] | Shanghai Men's Health Study (SMHS) (China) | 51,963 / 900 | 54 | 23.6 | 100 | Mean: 4 | Diet<br>Fish<br>Seafood<br>EPA+DHA | Self-report confirmed by diabetes medication use or fasting glucose or OGTT. | +++ | 5 | Yes / Yes |
| Villegas (2011)[43] | Shanghai Women's Health Study (SWHS) (China) | 64,193 / 3,034 | 51 | 23.8 | 0 | Mean: 8.9 | Diet<br>Fish<br>Seafood<br>EPA+DHA | Self-report confirmed by diabetes medication use or fasting glucose or OGTT. | +++ | 5 | Yes / Yes |

EPIC, European Prospective Investigation into Cancer and Nutrition; JPHCM, Japan Public Health Center-based Prospective Study of Men; JPHCW, Japan Public Health Center-based Prospective Study of Women; Fish, refers to finfish; Seafood, refers to finfish and shellfish combined.

* All exposure-type 2 diabetes risk assessments were pre-specified primary analyses, except dietary fish/n-3 PUFA intake in Djousse et al [30], which were secondary analyses.

† Dietary exposures were all assessed using food frequency questionnaires; fatty acid biomarkers were measured in plasma phospholipid [30; 32; 38; 44], plasma cholesterol ester [44]; red blood cell membrane [34]; red blood cell phospholipid [35; 38], by gas or gas-liquid chromatography.

‡ Degree of covariate adjustment indicated by (+): sociodemographics; (++): sociodemographics and other risk factors; (+++):sociodemographics and other risk factors and dietary variables.

§ Quality of each study was assessed by 5 separate criteria on an integer scale (0 or 1, with 1 being better). These included: appropriateness and reporting of inclusion and exclusion criteria, assessment of exposure (1 point if habitual n-3 PUFA/fish/seafood consumption was assessed with a validated diet assessment method, i.e. a validated FFQ or repeated short term measures; for biomarkers, a published laboratory assay), assessment of outcome (1 point if diagnosis of DM was confirmed according to accepted criteria and not based on self report), control of confounding (1 point if adjusted for socio-demographic plus either other risk factors or dietary variables for dietary studies; 1 point if adjusted for socio-demographic plus other risk factors for biomarker studies) and evidence of bias (1 point if no evidence of bias). Scores were summed and studies with scores from 0 to 3 and 4 to 5 were considered lower and higher quality, respectively.

|| Dependent on the study, additional data requested included baseline characteristics (age, BMI, exposure distributions) and exposure category-specific data (number of participants, person-years of follow-up, number of cases, median level of exposure, risk estimates and 95% CI).

¶ Nested selection of cases and controls within the original cohort. The total number of participants recruited in the original cohorts were: Melbourne Collaborative Cohort Study, n=41,528 [32]; Vasterbotton intervention program [34], n=33,336; EPIC-Norfolk, n=25,639 [38].

**Supplementary Table 1. Summary of number of cohorts available for meta-analysis for each dietary exposure, and the range of intake in the identified studies.**

| | ALA | EPA | DHA | EPA+DHA | Fish/Seafood |
|---|---|---|---|---|---|
| Number of cohorts identified for meta-analysis | n=7 | n=5 | n=5 | n=12 | n=13 |
| | | | **Range of dietary intake** | | |
| **Cohorts** | **(g/day)** | **(mg/day)** | **(mg/day)** | **(mg/day)** | **(g/day)** |
| Meyer (2001)[36]* | - | - | - | 30-390 | - |
| Van Dam (2002)[40]* | 0.32-0.67 | - | - | - | - |
| Hodge (2007)[32]* | 0.78-1.13 | 41-204 | 72-354 | 114-558 | - |
| Vang (2008)[42]† | - | - | - | 0-85 | Seafood: 0-21 |
| Kaushik_NHS (2009)[33] ‡ | - | - | - | 120-320 | Fish: 15-42 |
| Kaushik_NHS2 (2009)[33] ‡ | - | - | - | 110-300 | Fish: 7-30 |
| Kaushik_HPFS (2009)[33] ‡ | - | - | - | 140-420 | Fish: 14-46 |
| Patel (2009)[39] † | - | - | - | 32-194 | Seafood: 8-49 |
| van Woudenbergh (2009)[41] † | - | 4-77 | 19-159 | 24-237 | Seafood: 0-36 |
| Patel (2010)[38]§ | 0.82-1.22 | 50-170 | 50-230 | - | - |
| Brostow (2011)[29]* | 0.28-0.93 | - | - | 130-550 | Fish: 19-86<br>Seafood: 22-94 |
| Djousse_CHS (2011)[30]\|\| | 0.96-1.84 | - | - | 110-730 | Fish: 2-85 |
| Djousse_WHS (2011)[31]* | 0.79-1.59 | 10-120 | 40-270 | 60-390 | Fish: 8-64 |
| Kroger (2011)[35]* | 1-2.3 | 6-222 | 18-334 | 31-555 | - |
| Nanri_JPHCM (2011)[37]\|\| | - | - | - | 382-1793 | Seafood : 37-172 |
| Nanri_JPHCW (2011)[37]\|\| | - | - | - | 369-1703 | Seafood: 35-163 |
| Villegas_SMHS (2011)[43]* | - | - | - | 20-200 | Fish: 10-80<br>Seafood: 14-99 |
| Villegas_SWHS (2011)[43]* | - | - | - | 20-200 | Fish: 10-80<br>Seafood: 14-99 |

ALA, alpha-linolenic acid; EPA, eicosapentaenoic acid; DHA, docosahexaenoic acid; NHS, Nurses' Health Study; NHS2, Nurses' Health Study 2; HPFS, Health Professionals Follow-up Study; CHS, Cardiovascular Health Study; WHS, Women's Health Study; JPHCM, Japan Public Health Center-based Prospective Study of Men; JPHCW, Japan Public Health Center-based Prospective Study of Women; SMHS, Shanghai Men's Health Study; SWHS, Shanghai Women's Health Study.
* Values are median of the 1st quintile to the 5th quintile.
† Values are median of the lowest to the highest category of intake.
‡ Values are inter-quartile range.
§ Values are median of the 1st tertile to the 3rd tertile.
|| Values are median of the 1st quartile to the 4th quartile.

**Supplementary Table 2. Summary of number of cohorts available for meta-analysis for each fatty acid biomarker exposure, and the range of distributions in the identified studies.**

| Study | | ALA | EPA | DHA | EPA+DHA |
|---|---|---|---|---|---|
| Number of cohorts identified for meta-analysis | | n = 6 | n = 5 | n = 5 | n = 5 |
| **Cohorts** | **Measured Compartment** | **Range of distribution, as % of total fatty acids** | | | |
| Wang (2003)[44]* | Plasma_PL | 0.09-0.21 | - | - | - |
| Wang (2003)[44]* | Plasma_CE | 0.29-0.56 | - | - | - |
| Hodge (2007)[32]* | Plasma_PL | 0.09-0.27 | 0.57-1.62 | 2.78-5.42 | 3.57-6.72 |
| Krachler (2008)[34]† | RBC membrane | 0.24-0.48 | 0.89-1.81 | 3.73-5.77 | 4.76-7.44 |
| Patel (2010)[38]‡ | Plasma_PL | 0.15-0.3 | 0.66-1.58 | 3.54-6.07 | 4.3-7.77 |
| Patel (2010)[38]‡ | RBC_PL | 0.1-0.18 | 0.64-1.4 | 4.5-6.47 | 5.16-7.7 |
| Djousse (2011)[30]§ | Plasma_PL | 0.1-0.21 | 0.32-0.87 | 2.04-4.22 | 2.47-5.02 |
| Kroger (2011)[35]* | RBC_PL | 0.1-0.22 | 0.42-1.2 | 2.7-6.2 | 3.2-7.3 |

RBC, red blood cells; PL, phospholipid; CE, cholesterol ester. Note both Wang et al [44], and Patel et al [38] measured fatty acids in 2 different compartments and assessed association of each with risk of type 2 diabetes.

* Values are median of the 1st quintile to the 5th quintile.

† Values are mean-SD to mean+SD.

‡ Values are median of the 1st tertile to the 3rd tertile.

§ Values are median of the 1st quartile to the 4th quartile.

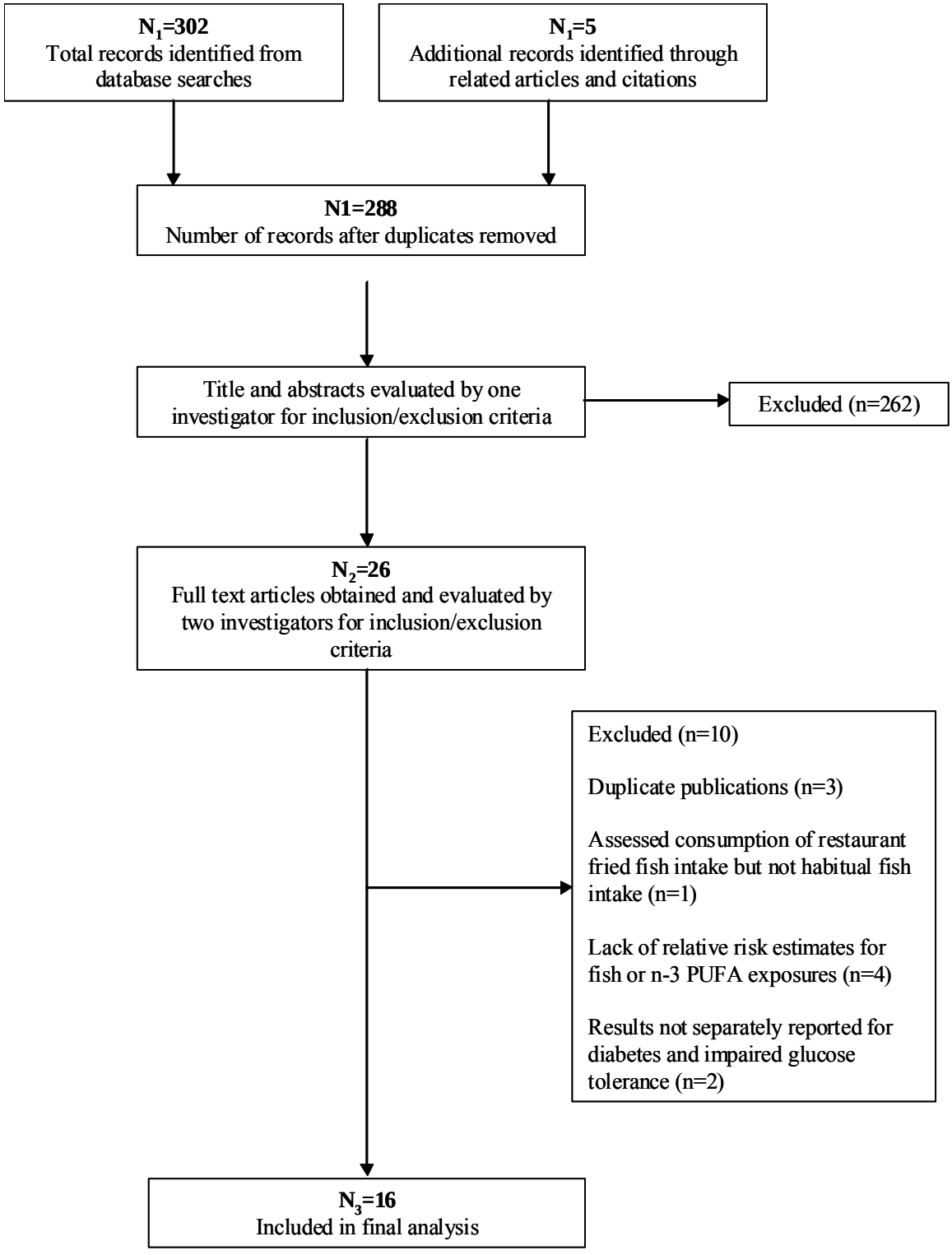

N1=302
Total records identified from database searches
N1=5
Additional records identified through related articles and citations
N1=288
Number of records after duplicates removed
Title and abstracts evaluated by one investigator for inclusion/exclusion criteria
Excluded (n=262)
N2=26
Full text articles obtained and evaluated by two investigators for inclusion/exclusion criteria
Excluded (n=10)
Duplicate publications (n=3)
Assessed consumption of restaurant fried fish intake but not habitual fish intake (n=1)
Lack of relative risk estimates for fish or n-3 PUFA exposures (n=4)
Results not separately reported for diabetes and impaired glucose tolerance (n=2)
N3=16
Included in final analysis

**Figure 1**

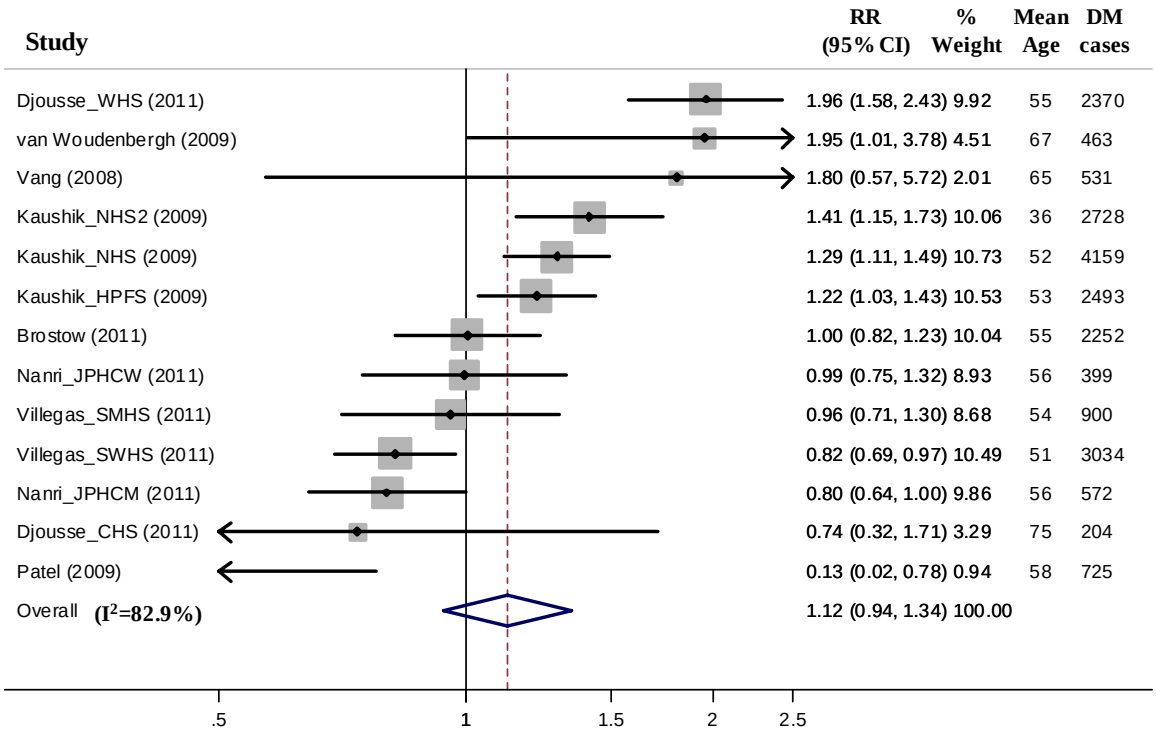



**Figure 2. Relative risk of incident type 2 diabetes per 100g/day of fish/seafood intake**

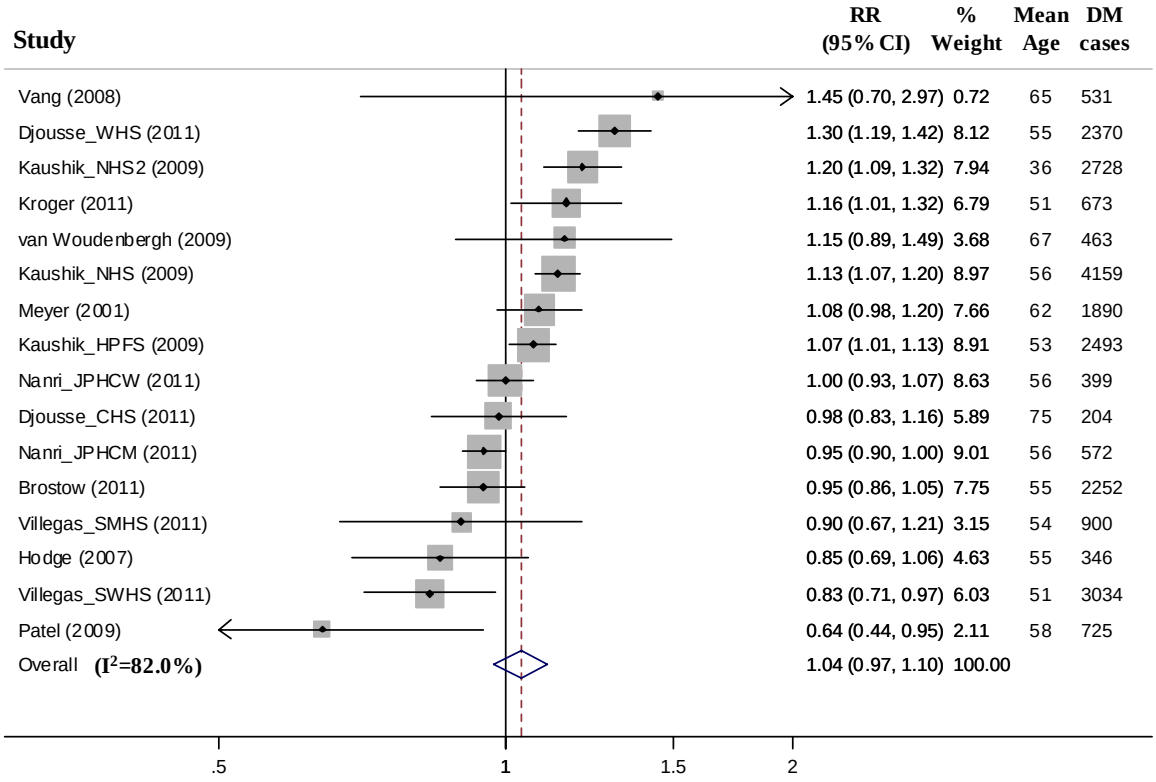



**Figure 3. Relative risk of incident type 2 diabetes per 250mg/day of EPA+DHA intake**

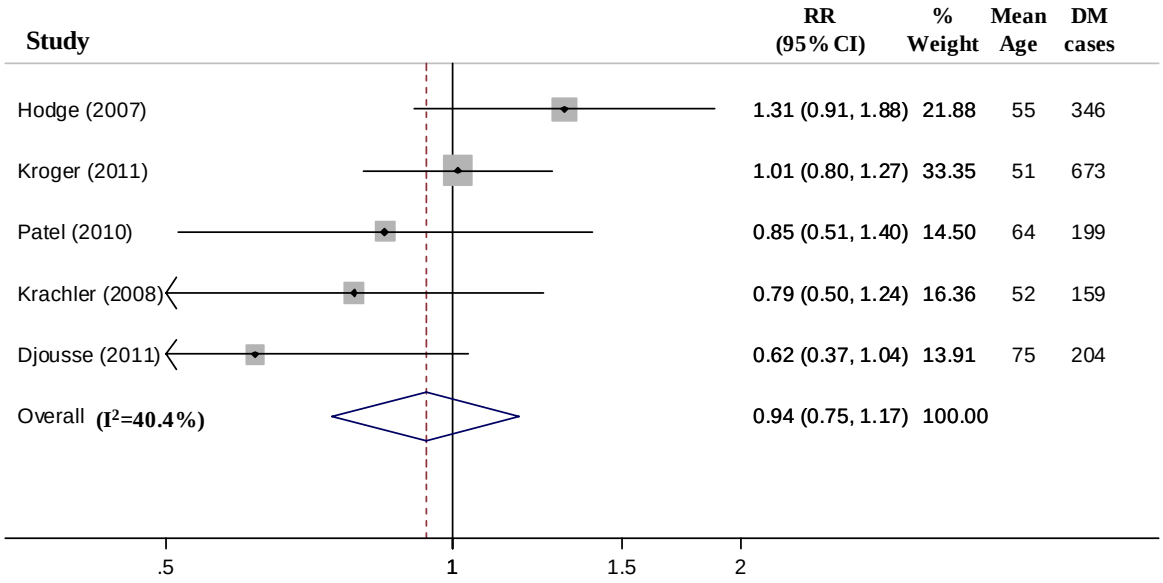



**Figure 4. Relative risk of incident type 2 diabetes per 3% (as % of total fatty acids) of EPA+DHA biomarker**

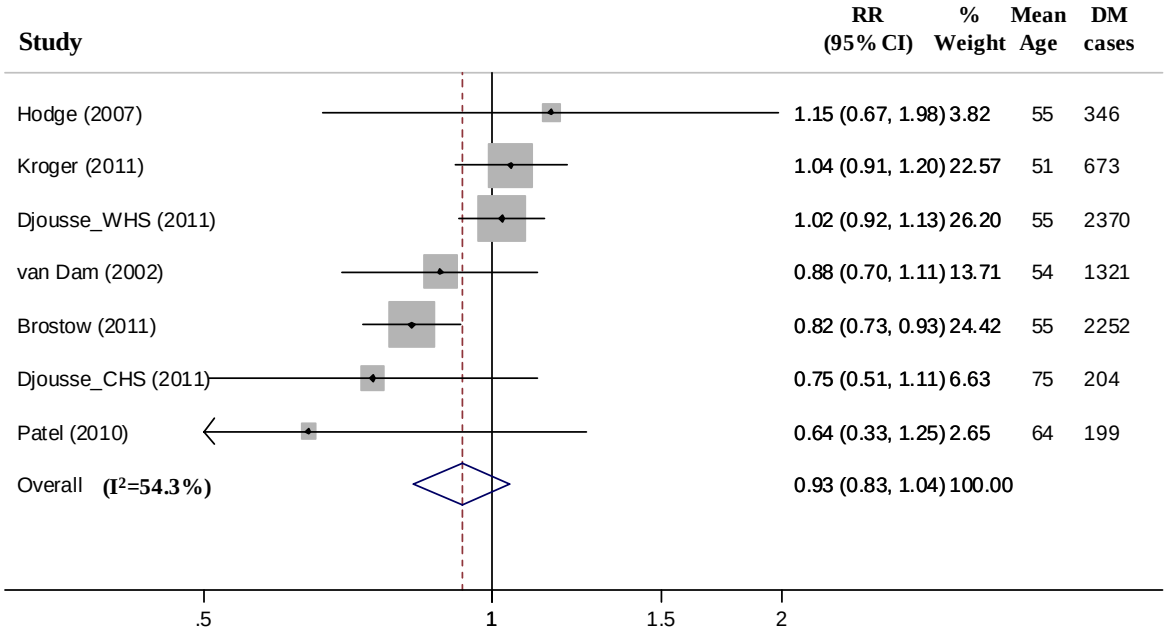



**Figure 5. Relative risk of incident type 2 diabetes per 0.5g/day of ALA intake**

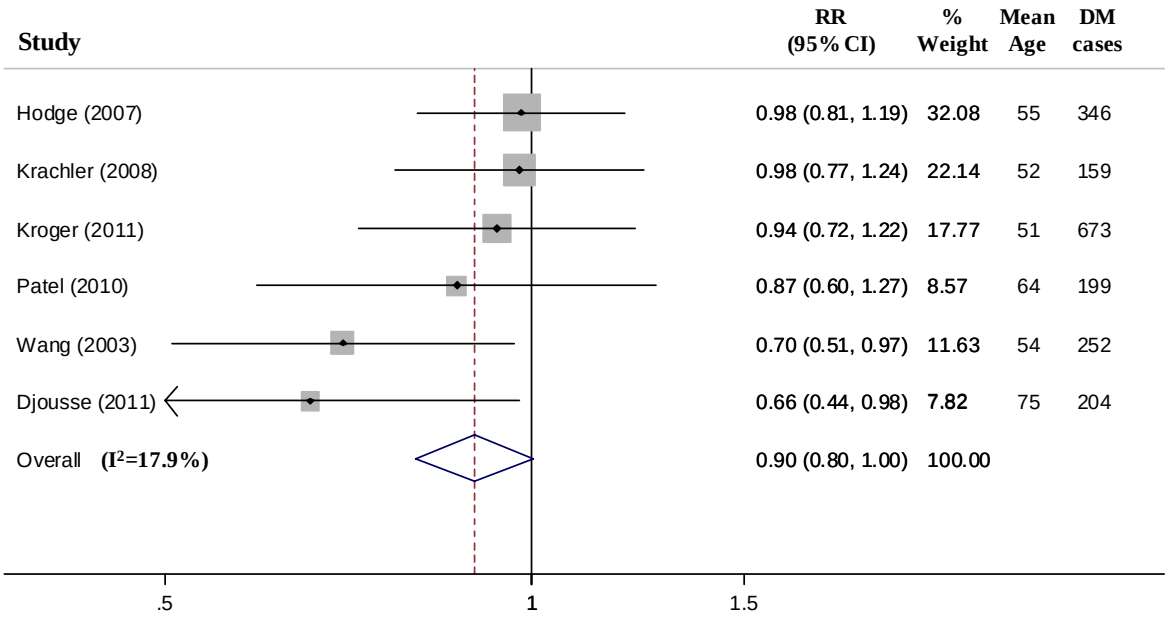
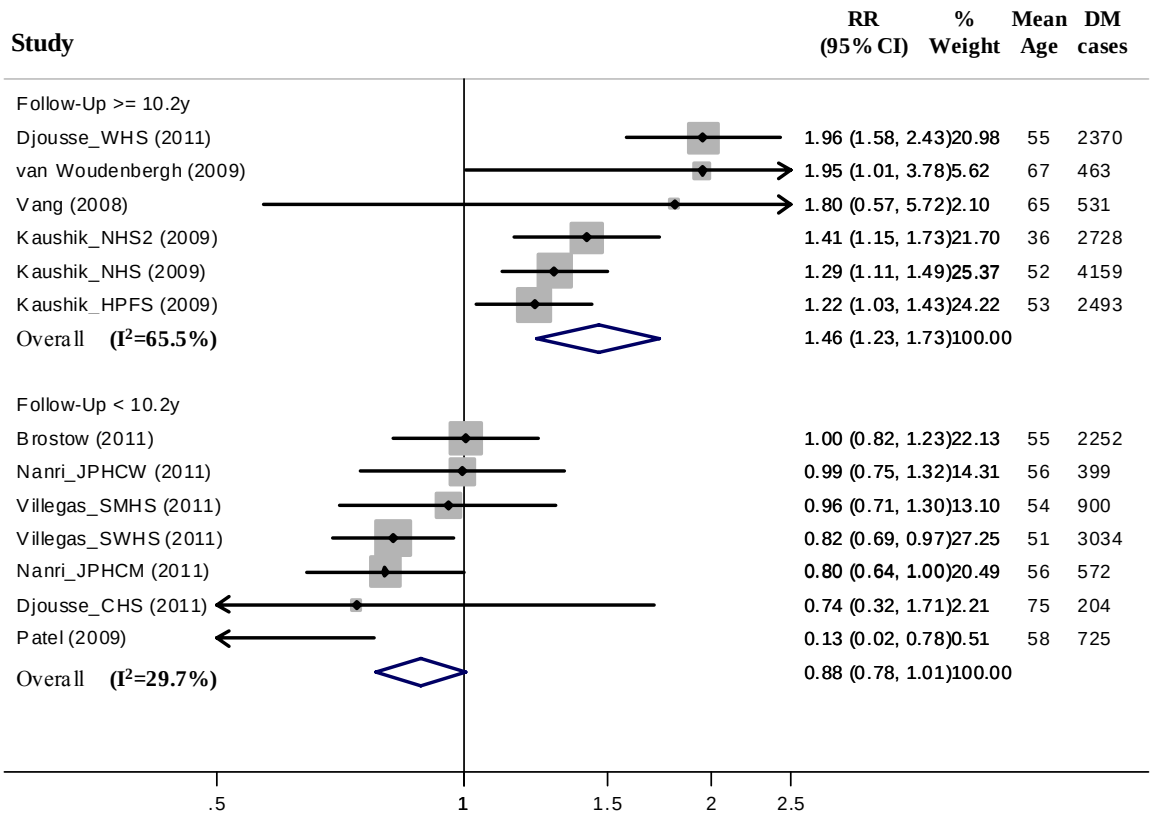
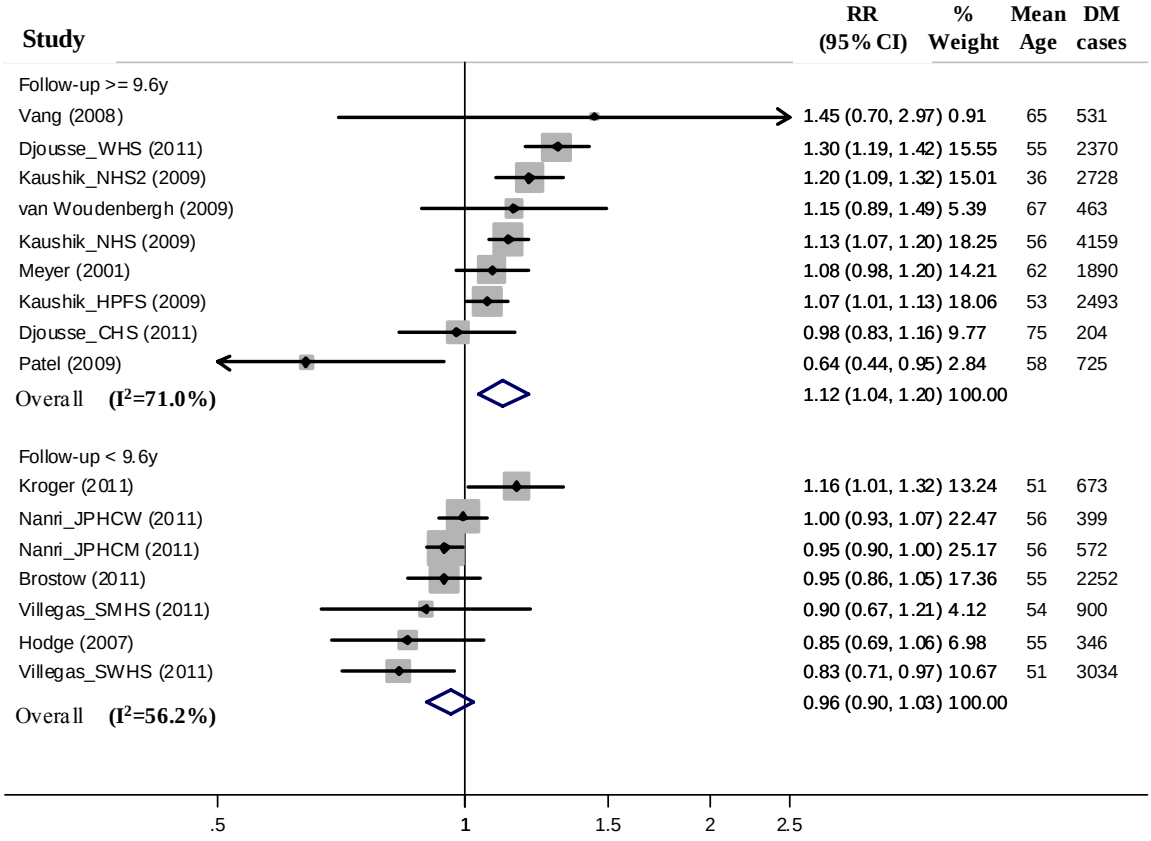


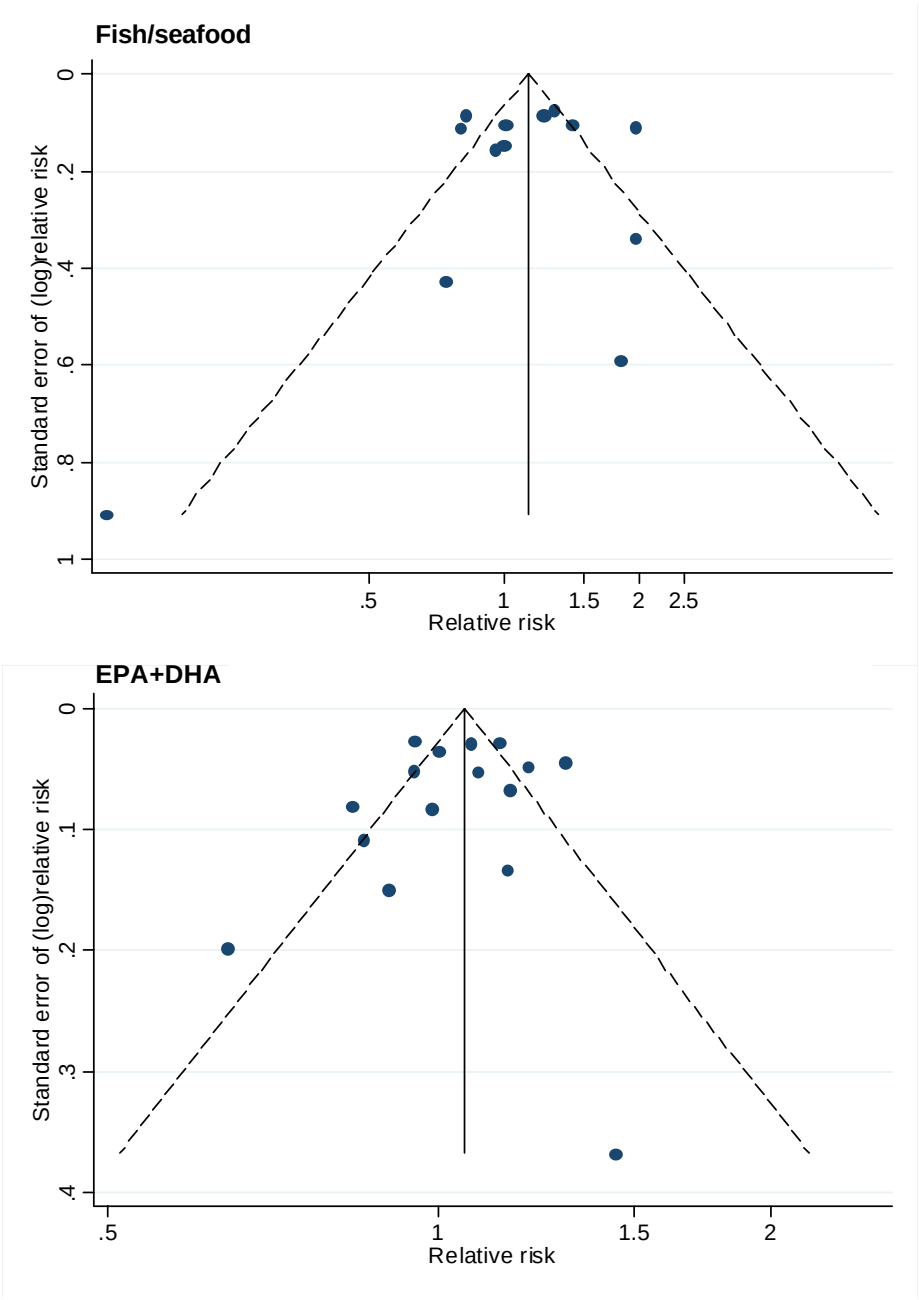

**Figure 6. Relative risk of incident type 2 diabetes per 0.1% (as % of total fatty acids) of ALA biomarker**




**Supplementary Figure 1. Relative risk of incident type 2 diabetes per 100g/day of fish/seafood intake, stratified by the duration of study follow-up**




**Supplementary Figure 2. Relative risk of incident type 2 diabetes per 250mg/day of EPA+DHA intake, stratified by the duration of study follow-up**




**Supplementary Figure 3. Funnel plots for studies that assessed the association of dietary fish/seafood (top panel), and of dietary EPA+DHA (bottom panel), with risk of incident type 2 diabetes**

## Appendix A: generalized least squares for trend estimation

As frequently observed in nutritional epidemiology, all except one of the identified studies categorized the exposures of interest, and obtained RR in the higher categories compared to the lowest (referent) category. A common approach adopted for meta-analysis is to pool the RR associated with the top category from each study. However, this has the disadvantage of pooling categories which are incomparable across different studies, and also not fully utilizing the available data (the middle categories are discarded). An important methodology utilized in this project is the generalized least squares trend (GLST) modeling [46; 47]. For GLST, we extracted the following basic information from each study: person-years of follow-up, number of DM cases, the median level of exposure in each quantile category, relative risk and 95% CI, which is entered into Stata (a Stata database example is shown below for 2 studies):

| n | cases | dose | rr | lci | uci | id |
|---|---|---|---|---|---|---|
| 3445 | 159 | 31.3 | 1 | 1 | 1 | 1 |
| 3310 | 94 | 136.3 | 1.01 | 0.71 | 1.46 | 1 |
| 3373 | 97 | 188.4 | 0.82 | 0.58 | 1.18 | 1 |
| 3329 | 122 | 270.8 | 0.97 | 0.69 | 1.36 | 1 |
| 3613 | 198 | 555.2 | 1.29 | 0.95 | 1.75 | 1 |
| 95041.08 | 458 | 60 | 1 | 1 | 1 | 2 |
| 92723.51 | 477 | 110 | 1.15 | 1.01 | 1.3 | 2 |
| 79403.28 | 410 | 160 | 1.18 | 1.03 | 1.35 | 2 |
| 94708.96 | 521 | 240 | 1.39 | 1.22 | 1.58 | 2 |
| 90200.72 | 504 | 390 | 1.45 | 1.27 | 1.65 | 2 |

Based on this information we calculated log of the relative risk as well as its standard error. The first stage of GLST analysis than obtains the dose-response relationship within each study by linear regression of ln(RR) on the exposure level. Importantly, because there is likely correlation between the ln(RR) estimates (due to the common reference category), the GLST method developed by Greenland and Longnecker (and implemented in Stata) takes this correlation into account [46; 47]. This has the effect of altering the standard error (SE) of the dose-response beta-coefficient estimate (generally increase the SE), To illustrate this, we look at an example using the study by Djousse et al [30], which investigated the association of EPA+DHA intake and DM risk.

We first run the linear regression model without taking into account the correlation between the ln(RR) estimates (by using the vwls option in the GLST command):

```
. glst logrr doser if id==3, se(se) cov(n cases) ir vwls

Variance-weighted least-squares regression          Number of obs   =       3
Goodness-of-fit chi2(2)    =    1.42                Model chi2(1)   =    0.06
Prob > chi2                =  0.4905                Prob > chi2     =  0.8141
------------------------------------------------------------------------------
       logrr |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
```

```
-------------+------------------------------------------------------------
-
       doser |  -.0000739   .0003145    -0.24   0.814    -.0006903    .0005424
------------------------------------------------------------------------------
```

We then repeat the analysis but taking into account the correlation between the ln(RR) estimates:

```
. glst logrr doser if id==3, se(se) cov(n cases) ir

Generalized least-squares regression                Number of obs   =       3
Goodness-of-fit chi2(2)    =    2.73                Model chi2(1)   =    0.05
Prob > chi2                =  0.2549                Prob > chi2     =  0.8312
------------------------------------------------------------------------------
       logrr |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       doser |  -.0000709   .0003328    -0.21   0.831    -.0007231    .0005812
------------------------------------------------------------------------------
```

As expected, the standard error of the beta-coefficient estimate increased, and this will have the effect of altering the weight associated with this coefficient in the second stage meta-analysis where studies are pooled.

In the second stage, the study specific dose-response beta-coefficents are pooled using either fixed (inverse variance weighted) or random effects (Darsimonian and Larid) meta-analysis (see below for explanation of these principles and the decision to choose between the methods).

A current limitation of the GLST analyses implemented in Stata is that it is unable to generate a forest plot, which is useful in to visualize the study specific dose-response risk estimates and the extent of heterogeneity. Therefore, I had to manually took the output from the first stage GLST analysis and enter this into a new database and used the *metan* command in Stata to plot the forest plot and calculated $I^2$ for assessing study heterogeneity.

## Appendix B: $I^2$ statistic for heterogeneity, fixed and random effects meta-analysis

In order to make conclusions about the pooled results from a meta-analysis, it is important to know how consistent the results are between the different studies. Highly consistent results would increase our confidence in the pooled result, whereas large heterogeneity raises important questions such as what could be the underlying cause of the heterogeneity, and also affect the choice of statistical model used to pool the results. Heterogeneity between studies are to be expected due to differences in a multitude of factors such as study quality, participant characteristics, duration of follow-up / treatment, and also variation in RR estimates due to chance alone.

A commonly used test statistic to assess heterogeneity is the Cochran's Q. This is calculated by summing the square deviation of each study's effect estimate from the pooled effect estimate. The contribution of each study is weighted by the inverse of its variance. The Q statistic follows a $\chi^2$ distribution with k (number of studies) – 1 degrees of freedom under the null hypothesis of homogeneity [(63)]. There are several problems with the Q statistic. It has been shown to have low power at detecting true heterogeneity between studies, especially for meta-analyses that include small number of studies (for example – our analyses of EPA+DHA biomarkers and DM risk) [(49)]. Furthermore, while the Q statistic assists in testing of the potential presence of heterogeneity, it does not help to quantify the magnitude of this effect.

The $I^2$ was developed and proposed by Higgins et al [(49)] as an alternative way to assess heterogeneity. It is calculated as $I^2$ = 100% x (Cochran's Q statistics –degrees of freedom). The $I^2$ statistic takes on values between 0 to 100%, and indicates the proportion of total variation across studies that is due to heterogeneity rather than chance. For the current work, we used the $I^2$ as a tool to decide whether to use fixed effects or random effects model to pool the study results (see below). We chose 35% which is between 'low' (<25%) and 'moderate heterogeneity'(50%) as proposed by Higgins et al [(49)]. This was an arbitrary decision and obviously it is still crucial to visually inspect the Forest plot for each analysis which further aids in the interpretation of the data. For example, in the analysis of ALA biomarkers we can see the limited heterogeneity between studies as the effect size estimates were in a reasonably narrow band (and all <1) with large overlaps in their confidence intervals. It was also interesting to observe that in the subgroup analyses for both fish/seafood (30 and 66%) and EPA+DHA intake (56 and 71%) stratified by duration of follow-up, the $I^2$ was lower compared to the overall meta-analysis (>80%). This indirectly supports the meta-regression result that this factor was a source of heterogeneity between studies.

Fixed effect meta-analysis assumes that all studies are estimating the same underlying risk estimate, i.e. variations between studies is only due to chance created by sampling different patients. On the other hand, random effects modeling assume there are real differences in the risk estimate between studies, on top of the sampling variability. In the procedure implemented in this study (DerSimonian and Laird), a between study variance (normally called $\tau^2$) is added to the standard error of the study specific estimates. Therefore, whenever $I^2$ is >0, use of random effects model will result in a pooled estimate with wider 95% CI then the fixed effect model. For example, the association of ALA biomarker with risk of DM was 0.9 (95% CI, 0.8-1) and 0.9 (95% CI, 0.78-1.01) for fixed effect and random effects meta-analysis, respectively. It is inappropriate to utilize fixed effects modeling when there is substantial heterogeneity present, as this estimates a 'common' risk estimate among all studies which is not supported by the data.

## Appendix C: Meta-regression analysis

We explored potential sources of heterogeneity using random effects meta-regression [64]. This model assumes that a study *i* of a total of *n* studies provide an estimate, $y_i$, of the effect of interest. The standard error of $y_i$ is $\sigma_i$, which are also estimated from each study. The model allows the true effects $\theta_i$, to vary between studies by assuming they have a normal distribution around a mean effect, $\theta$; i.e.

$y_i \mid \theta i \sim N(\theta i, \sigma_i^2)$, where $\theta i \sim N(\theta, \tau^2)$

so, $y_i \sim N(\theta, \sigma_i^2 + \tau^2)$

or equivalently, $y_i = \theta + u_i + \varepsilon_i$, where $u_i \sim N(0, \tau^2)$ and $\varepsilon_i \sim N(0, \sigma_i^2)$

The random effects meta-regression assumes that the true effects follow a normal distribution around a linear predictor (the covariate whose effect we're interested in), i.e.:

$y_i \mid \theta i \sim N(\theta i, \sigma_i^2)$, where $\theta i \sim N(x_i\beta, \tau^2)$

so, $y_i \sim N(x_i\beta, \sigma_i^2 + \tau^2)$

or equivalently, $y_i = x_i\beta + u_i + \varepsilon_i$, where $u_i \sim N(0, \tau^2)$ and $\varepsilon_i \sim N(0, \sigma_i^2)$

In meta-analysis where there are substantial heterogeneity, it is nor reasonable to assume that all of the heterogeneity is explained by the covariate(s) in the model. The random-effects meta-regression analysis is therefore appropriate as it retains the estimate $\tau^2$, which acknowledges this possibility of residual heterogeneity not explained by the covariates in the model, and will yield wider confidence intervals for the regression coefficients than a fixed effect analysis [65]. Stata carries out the meta-regression by first estimating the between-study variance, $\tau^2$, and then estimate the coefficients, $\beta$, by weighted least squares by using the weights $1/(\sigma_i^2 + \tau^2)$[64] The weighting allows the more precise studies to have more influence in the analysis. The meta-regression of fish intake and the covariates BMI and duration of study follow up is shown below to illustrate meta-regression analysis.

**Meta-regression of log RR of fish intake and risk of DM, with median duration of follow-up in each cohort as the covariate**

```
Meta-regression                                     Number of obs  =      13
REML estimate of between-study variance             tau2           =       0
% residual variation due to heterogeneity           I-squared_res  =   0.00%
Proportion of between-study variance explained      Adj R-squared  = 100.00%
With Knapp-Hartung modification
------------------------------------------------------------------------------
       logrr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  _Ifu_med_1 |   .3941229   .0623378     6.32   0.000     .2569183    .5313275
       _cons |   -.137315   .0385368    -3.56   0.004     -.222134    -.052496
------------------------------------------------------------------------------
```

The analysis suggests that the duration of follow-up is significantly associated with log RR. Studies with duration of follow-up > the median of all studies were associated with higher logRR (β=0.39, 95% CI=0.26-0.53, P<0.001). This factor accounted for most of the heterogeneity between the studies (0% of the residual variation is due to heterogeneity, no remaining between study variance, i.e. $\tau^2$ = 0, and proportion of between study variance explained by this covariate, $R^2$ = 100%).

Similar results were observed with median BMI as the covariate (not shown) so we carried out multivariate meta-regression to see possible independent effects of these 2 factors:

```
Meta-regression                                         Number of obs  =      13
REML estimate of between-study variance                 tau2           =       0
% residual variation due to heterogeneity               I-squared_res  =   0.00%
Proportion of between-study variance explained          Adj R-squared  = 100.00%
Joint test for all covariates                           Model F(2,10)  =   20.90
With Knapp-Hartung modification                         Prob > F       =  0.0003
------------------------------------------------------------------------------
       logrr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 _Ibmi_med_1 |  -.4404343   .3250415    -1.36   0.205    -1.164672    .2838032
  _Ifu_med_1 |   .8271465   .3255956     2.54   0.029     .1016743    1.552619
       _cons |  -.1329545    .038671    -3.44   0.006    -.2191188   -.0467903
------------------------------------------------------------------------------
```

The analysis suggests that only the duration of follow-up is significantly associated with log RR, and the effects of BMI may have been confounded by duration of follow-up.

These analyses additionally highlight many of the interesting theoretical and practical aspects of meta-regression [65]. For example, we were careful to pre-specify a set of *a priori* selected potential factors that may have contributed to the heterogeneous results. This was important to avoid 'data-dredging' (i.e. multiple post-hoc analyses) which may lead to many false-positive conclusions. We also encountered one of the common problems of meta-regression which was lack of trial level summary data for the covariates, e.g. BMI and median years of follow-up in each cohort, which necessitated direct author contact. We were able to obtain the great majority of the covariate values this way, but it is easy to envision that in a meta-analysis involving large number of studies this may be difficult to achieve (the meta-regression would then suffer from potential bias by excluding studies that did not provide the covariate values). Meta-regression was also not possible (or at least not meaningful to carry out) when the number of identified studies were small (e.g. for EPA+DHA biomarkers where only 5 studies were included). While our analysis indicated that the duration of follow-up could be an important source of heterogeneity, it is vital to keep in mind this result is observational in nature and subjected to possible ecological bias, i.e. there may be confounders of this relationship which could not be assessed. The finding therefore should be thought of as hypothesis generating and interpreted with caution.

## Appendix D: Assessment of publication bias

We used both graphical (funnel plot) and the Begg's as well as the Egger's formal statistical tests to assess the potential for publication bias in our analyses. These two approaches are complimentary and should be considered in conjunction to help ascertain potential publication bias. The funnel plot is based on visualizing the relationship between study sample size (or the precision of the effect estimate, which will increase as the sample size of the study increases) and effect size estimate. Therefore, it is expected that larger studies will have narrower 'spread' around the pooled effect estimate, while smaller studies should scatter widely. In the absence of bias, this should resemble a symmetrical inverted funnel. The Begg's test is a rank correlation test for the association between effect size and its variance [50]. The test is based on empirical grounds, i.e. based on the fact that if there was publication bias, this would lead to skewness in the funnel graph and we would expect to see correlation between the effect size and its variance [50]. In the first step of the test the effect sizes are standardized (by subtracting each effect size $t_i$ from the overall effect estimate, $t_{overall}$, then divided by the standard deviation of ($t_i$-$t_{overall}$). The test than enumerate the number of pairs of studies that are ranked in the same order with respect to the standardized effect size and its variance (call these x) and also those ranked in the opposite order (call these y), then the normalized test statistic is

$$Z = (x-y) / [k(k-1)(2k+5)/18]^{1/2}$$

Where k is the number of studies in the meta-anlaysis.

The Egger's test is carried out by linear regression of the standard normal deviate (SND) of the effect size, against the precision of the effect estimate [51]. The SND is calculated by dividing the effect estimate (RR) by its standard error, whereas the precision is defined as the inverse of the standard error. The regression equation is therefore: SND = $a + b$ x precision. We expect small studies to be close to zero on the axis, because precision is largely dependent on sample size. While small studies may also produce RR estimates that differ from unity, their standard error is likely large and therefore the SND will also be close to zero. Therefore, the small studies should be close to the origin. On the other hand, large studies should have more precise estimates, and if they showed an effect according to the exposure, will also produce large SND. Therefore, for meta-analysis that are not substantially influenced by publication bias, one would expect the Egger's regression to run through the origin (i.e. $a = 0$) with the slope $b$ indicating the size and direction of effect associated with the exposure. As en example we carried out these statistical tests examining potential publication bias in the meta-analysis of dietary fish intake and risk of DM:

**Begg's Test**

```
  adj. Kendall's Score (P-Q) =      -4
```

```
          Std. Dev. of Score =   16.39
           Number of Studies =      13
                           z  =   -0.24
                     Pr > |z| =   0.807
                           z  =    0.18 (continuity corrected)
                     Pr > |z| =   0.855 (continuity corrected)
```

**Egger's test**

| Std_Eff | Coef. | Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| slope | .0906547 | .1391525 | 0.65 | 0.528 | -.2156179 | .3969272 |
| **bias** | -.7960065 | 1.274108 | -0.62 | **0.545** | -3.6003 | 2.008287 |

We actually used both approaches (but reported one for simplicity in the study manuscript) and the results were consistent. In this example – Begg's test indicated no substantial evidence of publication bias (P=0.86) and similarly Egger's test (the intercept *a* is labeled as 'bias by Stata) supports this conclusion (P=0.55). Both of the tests have somewhat limited statistical power (with the Egger's test performing somewhat better than the Begg's) especially when the number of studies in the meta-analysis is low (which applies to most of our pooled results with ≤16 studies) [66]. However, visual inspection of the funnel plots also supported lack of substantial publication bias in our meta-analyses. It is also important to note that we carried out systematic searching of multiple databases as well as checking for additional relevant papers (e.g. in references of included papers and links of 20 most relevant papers via Pubmed) to ensure identification of all relevant published studies, which would have also reduced the potential for missing publications and causing bias.

## Appendix E: Stata codes

Stata codes are shown below for the assessment of EPA+DHA dietary intake and risk of DM, with all other analyses using similar codes.

```
**********calculate variables needed for GLST***********

gen logrr= log(rr)
gen loglci= log(lci)
gen loguci= log(uci)
gen double se = (loguci - loglci)/ (2*invnorm(.975))


********** GLST*****************

xi: glst  logrr dose, se(se) pfirst(id study) cov(n cases) ts(f) eform



*******Secondary analyses by pooling the top vs bottom category risk estimate
from the studies********

metan  logrr loglci loguci if top_q==1, randomi eform texts(130)
label(namevar=aut_name, yearvar=year)



*******study specific log(RR) and 95% CI were manually entered into a second
data base to allow assessment of forest plots, I^2 values, meta-regression and
potential publication bias**********



*********overall meta-analysis and forest plot*************
gsort -rr

metan  logrr loglci loguci, randomi eform texts(150) boxsca(100)
label(namevar=aut_name, yearvar=year) nowarning xlabel(0.5, 1, 1.5, 2)force
nohet graphregion(color(white)) rcols(age cases)



*********meta-regression if keep all imputed studies but keep Meyer**********


xi: metareg logrr i.age_med ,wsse(selogrr)
xi: metareg logrr i.bmi_med ,wsse(selogrr)
xi: metareg logrr i.fu_med ,wsse(selogrr)

xi: metareg logrr age , wsse(selogrr)
xi: metareg logrr bmi , wsse(selogrr)
xi: metareg logrr fu , wsse(selogrr)
```

```
*********Meta-analysis and forest plot stratified by median follow-up
years*************

metan  logrr loglci loguci, by(fu_med) randomi eform texts(120)
label(namevar=aut_name, yearvar=year) nowarning xlabel(0.5, 1, 1.5, 2,
2.5)force nohet graphregion(color(white)) rcols(age cases) nooverall



******publication bias and funnel plot**********

metabias logrr selogrr, graph(begg)
```