Chapter 2. Methodology

Objectives

The topic of this report was nominated by the National Institutes of Health (NIH) Office of Dietary Supplements (ODS). The three participating Evidence-Based Practice Centers (EPCs) were asked to examine the effects of omega-3 fatty acids, in general, and on the following conditions: Cardiovascular Disease, Transplantation, Immune-Mediated Diseases, Gastrointestinal/Renal Diseases, Cancer, Neurology, Asthma, Child/Maternal Health, Eye Health, and Mental Health. The Southern California EPC (SCEPC) was responsible for examining Immune-Mediated Diseases and Gastrointestinal/Renal Diseases in Year 1 of the project and Cancer and Neurology in Year 2 of the project.

Scope of Work

The methodology that we used for this study included the following:

- Refining the preliminary questions provided by AHRQ,
- Convening a technical expert panel to advise the SCEPC on the study,
- Identifying sources of evidence in the scientific literature,
- Establishing inclusion/exclusion criteria for the articles identified in the scientific literature,
- Identifying potential evidence with attention to controlled clinical trials using omega-3 fatty acids,
- Evaluating potential evidence for methodological quality and relevance,
- Extracting data from studies meeting methodological and clinical criteria,
- Synthesizing the results,
- Performing further statistical analysis on selected studies,
- Performing pooled analyses where appropriate,
- Submitting the results to technical experts for peer review,
- Incorporating reviewers' comments into a final report for submission to AHRQ.

Original Proposed Key Questions

Preliminary questions for the project were developed by ODS in collaboration with the following NIH Institutes: (a) National Cancer Institute (NCI); (b) National Eye Institute (NEI); (c) National Heart, Lung, and Blood Institute (NHLBI); (d) National Institute of Alcohol Abuse and Alcoholism (NIAAA); (e) National Institute of Allergy and Infectious Diseases (NIAID); (f) National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS); (g) National Institute of Child Health and Human Development (NICHD); (h) National Institute of Diabetes

and Digestive and Kidney Diseases (NIDDK); (i) National Institute of Mental Health; and (j) National Institute of Neurological Disorders and Stroke (NINDS).

The general and disease-specific questions that were originally proposed are detailed in Appendix A.1, "Methodologic Approach."

Technical Expert Panel

Each AHRQ evidence report is guided by a Technical Expert Panel (TEP). The TEP advises the SCEPC on refining the preliminary questions, determining the proper inclusion/exclusion criteria for the study and the populations of interest, establishing the proper outcomes measures, and conducting the appropriate analyses.

We convened three TEPs that focused on the following conditions: 1) rheumatoid arthritis (RA), systemic lupus erythematosis (SLE), and bone density/osteoporosis; 2) renal disease and diabetes; and 3) gastrointestinal (GI) diseases. The TEPs were composed of distinguished basic scientists and clinicians, with established expertise in the following areas: omega-3 fatty acids, human nutrition, dietary assessment methods, gastroenterology, nephrology, diabetes, osteoporosis, immunology, and rheumatology. In addition to the experts that we identified, AHRQ and the relevant NIH Institute(s) recommended a number of industry experts. The members of our technical expert panels and a summary of their key comments and recommendations are listed in Appendix A.2.

Key Questions Addressed in this Report

Based on input from our three TEPs, the preliminary disease-specific questions were revised. Additionally, in consultation with the Task Order Officer and the other participating EPCs, we added several questions to our scope of work that had previously been assigned to the NEMC/EPC because they were related to topics we were reviewing. Similarly, a question that had been assigned to the SCEPC for year two was reassigned to the NEMC-EPC. Lastly, one additional question (pertaining to rheumatoid arthritis – number of tender joints) was suggested by a TEP member after reviewing the draft report and was assessed post-hoc. The questions that are addressed in this report are as follows:

Diabetes

What is the evidence in adults or children with a) type II diabetes, or b) insulin resistance/the metabolic syndrome for the efficacy of omega-3 fatty acids in treatment of:

- *total cholesterol*
- *HDL cholesterol*
- *LDL cholesterol*

- *Triglycerides*

What is the evidence in adults and children for an effect of omega-3 fatty acids on insulin sensitivity in a) type II diabetes, or b) the metabolic syndrome?

Inflammatory Bowel Disease

What is the evidence for the efficacy of omega-3 fatty acids in treatment of Crohn's disease and ulcerative colitis?

What is the evidence that in adults or children with inflammatory bowel disease, omega-3 fatty acids can replace steroids or other immunosuppressive drugs?

What is the evidence that the benefits of omega-3 fatty acids are influenced by the concomitant administration of various immunosuppressive agents in the treatment of inflammatory bowel disease?

Rheumatoid Arthritis

What is the evidence that in adults or children with rheumatoid arthritis, omega-3 fatty acids affect:

- *pain*
- *number of swollen joints*
- *disease activity*
- *patient's global assessment*
- *joint damage*
- *number of tender joints*

What is the evidence that in adults or children with rheumatoid arthritis, omega-3 fatty acids can replace other more potent anti-inflammatory or immunosuppressive drugs such as steroids and NSAIDs?

What is the evidence that the benefits of omega-3 fatty acids are influenced by the concomitant administration of various immunosuppressive agents in the treatment of rheumatoid arthritis?

Renal Disease

What is the evidence for the efficacy of omega-3 fatty acids in treatment of renal inflammation and glomerulosclerosis?

What is the evidence that in adults or children with immune-mediated renal disease, omega-3 fatty acids can replace steroids or other immunosuppressive drugs?

What is the evidence that the benefits of omega-3 fatty acids are influenced by the concomitant administration of various immunosuppressive agents in the treatment of immune-mediated renal disease?

Systemic Lupus Erythematosus

What is the evidence that in adults or children with systemic lupus erythematosus, omega-3 fatty acids affect disease activity, damage, or patient perceptions of outcomes?

What is the evidence that in adults or children with systemic lupus erythematosus, omega-3 fatty acids can replace steroids or other immunosuppressive drugs?

What is the evidence that the benefits of omega-3 fatty acids are influenced by the concomitant administration of various immunosuppressive agents in the treatment of systemic lupus erythematosus?

Bone Density/Osteoporosis

What is the evidence that omega-3 fatty acids help maintain bone mineral status?

For each of the study questions we also assessed 1) the effect of omega-3 fatty acids on sub-populations, 2) the effects of covariates, dose, source, and exposure duration on the outcomes of interest, and 3) the sustainment of effect.

Assessment of Adverse Events

In addition to assessing the efficacy of omega-3 fatty acids as specified above, we evaluated the data on adverse events that were reported in the studies we reviewed. We recognized, a priori, that adverse events are not reported in a standard way across clinical trials either in terms of the specific adverse events assessed or in the reporting of these adverse events. Hence, the purpose of this analysis was to define in general terms adverse events that occur with omega-3 fatty acids in order to identify specific adverse events that might warrant further investigation.

From each study, we extracted the number of adverse events reported for both intervention and placebo groups. We grouped the adverse events into the following categories:

- Clinical bleeding
- Gastrointestinal complaints or nausea

- Diarrhea
- Headache
- Dermatological
- Withdrawal due to adverse event

We calculated rates of adverse events within the intervention and placebo groups. Adverse event rates were calculated as the percentage of patients pooled across all conditions who had one of the adverse events. Reporting of adverse events varied greatly across studies. Many studies did not report on adverse events. If a study did not report on an adverse event (i.e. missing value), it was not used in that adverse event calculation. If a study reported not having an adverse event (i.e. adverse event rate of 0%), it was used in the calculation. Studies that did not specify the group allocation of the adverse events were excluded. We also excluded studies that reported the number of adverse events but did not report the group sample sizes.

Identification of Literature Sources

Potential evidence for our study came from three sources: on-line library databases, the reference lists of all relevant articles, and industry experts.

Jessie McGowan, Senior Information Scientist, and Nancy Santesso, Knowledge Translation Specialist, at the University of Ottawa were responsible for developing a common search strategy for omega-3 fatty acids for the 3 participating EPCs. Nancy Santesso developed a core omega-3 search strategy in collaboration with project librarians, biochemists, nutritionists, and clinicians, who also provided biochemical names, abbreviations, food sources, and commercial product names for omega-3 fatty acids. The literature search was not restricted by language of publication or by study design, except with the MeSH term, "dietary fats," in order to increase specificity. When possible, the searches were limited to studies involving human subjects. The core search strategy is detailed in Appendix A.3.

For the SCEPC, this core search strategy was incorporated into 6 specific searches that focused on our relevant disease categories: rheumatoid arthritis, bone density, SLE, renal disease, diabetes, and gastrointestinal diseases. The strategies for these searches are detailed in Appendix A.3.

The following databases were searched: Medline (1966-July, 2003), Premedline (July 8, 2003), Embase (1980-Week 27, 2003), Cochrane Central Register of Controlled Trials (2nd Quarter, 2003), CAB Health (1973-June 2003), Dissertation Abstracts (1861-to December 2002). All of these databases were searched using the Ovid interface, except CAB Health, which was searched through SilverPlatter. Any duplicate records were identified and removed within each search question using Reference Manager software, except for the last update, which was imported into EndNote. The citations obtained from these literature searches were sent to the SCEPC via e-mail.

The citations were transferred to a secured Internet-based software system (termed D2D) that enabled us to view article titles and abstracts electronically. Two reviewers, Walter Mojica and James Pencharz, used the computerized software system to independently evaluate the citations and abstracts using the review form in Figure B.1, Appendix B, which was loaded onto the computerized system.

The reviewers flagged article titles that focused on omega-3 fatty acids and any of the following disease conditions: diabetes mellitus, inflammatory bowel disease (ulcerative colitis and Crohn's disease), rheumatoid arthritis, SLE, renal disease, osteoporosis, or bone mineral status. In addition, they flagged article titles that pertained to the disease conditions of the other participating EPCs (i.e., cardiovascular disease or asthma). Language was not a barrier to inclusion. Articles that either reviewer flagged were ordered, as well as those articles in which it was unclear from the title or abstract whether the article was relevant. The articles were ordered from the RAND library, the UCLA library, or Kessler-Hancock, a San Francisco-based literature retrieval firm with contacts around the world. The literature was tracked using ProCite and Access software.

In addition, we sent letters to industry experts recommended by the Office Dietary Supplements to obtain any unpublished data (Figure A.3.1).

Evaluation of Evidence

Two reviewers independently reviewed each article that was ordered to determine whether it should be accepted for further study using a structured screening form (shown in Figure B.2, Appendix B) that included a defined set of inclusive/exclusive criteria (Table A.4.1, Appendix A.4). Walter Mojica reviewed all of the articles; James Pencharz and Jennifer Grossman each reviewed a portion of the articles. The reviewers resolved any disagreements by consensus.

Extraction of Data

For the articles that passed our screening criteria, two reviewers independently abstracted detailed data onto a specialized quality review form (QRF) (Figure B.3, Appendix B). Walter Mojica and Jennifer Grossman reviewed all of the articles except those pertaining to diabetes, which were reviewed by Walter Mojica and Puja Khanna. We consulted with several outside scientists to complete QRFs for foreign-language articles. The reviewers resolved differences through consensus, and a senior physician researcher resolved any disagreements that could not be resolved through this method.

The QRF included questions about the trial design; the outcomes of interest; the quality of the trial; the number and characteristics of the patients; details on the intervention, such as the dose, frequency, and duration; the types of outcome measures; adverse events; and the elapsed time between the intervention and outcome measurements.

Grading Evidence

Methodologic Quality of Randomized Controlled Trials

To evaluate the quality of the design and execution of trials, we also collected information on the QRF about the study design, appropriateness of randomization, blinding, description of withdrawals and dropouts, and concealment of allocation.[6,113] A score for quality was calculated

for each trial using a system developed by Jadad (Appendix A.5, Figure A.5.1). The Jadad score rates studies on a scale of 0 to 5. Empirical evidence has shown that studies scoring 2 or less report exaggerated results compared with studies scoring 3 or more.[114,115] Thus, studies with a Jadad score of 3 or more are referred to as "high quality," and studies scoring 2 or less are referred to as "poor quality." For our purposes, if a trial was associated with more than one study, its quality score was equal to the maximum score calculated across its associated studies. Additionally, a generic summary quality score (A, B, C) was assigned to each study based upon the combination of its Jadad score and reporting of concealment of allocation (Appendix A.5, Table A.5.1).

Applicability

In this report, the focus is on the U.S. population. To capture the potential applicability of studies to the different populations of interest as defined in the scope of work (namely Americans with inflammatory bowel disease, rheumatoid arthritis, renal disease, systemic lupus erythematosus or osteoporosis), we categorized the populations in the studies we reviewed in terms of 1) applicability to the U.S. population and 2) health state (Appendix A.5, Table A.5.2). In the summary tables, each study receives a combined applicability grade consisting of the applicability and health state.

Data Synthesis

We performed both a qualitative and quantitative synthesis of the evidence. We performed a meta-analysis for those studies that sufficiently assessed interventions, populations, and outcomes to justify pooling. For the remaining studies, we performed a qualitative analysis.

Meta-Analysis

Our meta-analytic methods are sufficiently comparable across conditions and outcomes that we describe them in general in this section. Individual approaches and decisions are discussed as necessary and appropriate in the discussion of results for particular conditions and outcomes.

Selection of Trials for Descriptive Analysis or Meta-Analysis

For each condition, we identified a set of relevant outcomes, e.g., cholesterol outcomes for the condition of diabetes, based on input from our TEP. Trials were considered for further analysis if they contained information on a chosen outcome collected within a follow-up interval for which measures were considered clinically comparable.

For some trials, several publications presented the same outcome data. In these cases, we picked the most informative of the duplicates; for example, if one publication was a conference abstract with preliminary data and the second was a full journal article, we chose the latter. The publications dropped for duplicate data do not appear in the evidence table but are noted in the

results text. We note that multiple citations of the same article were removed at the title screening stage of the project.

In order for a trial to be included in further analysis, the associated publication(s) had to report on the outcome, and contain sufficient statistical information for the calculation of a summary statistic. A trial also had to provide data prior to the crossover point if the trial was a crossover design to mirror the data available from a non-crossover trial, i.e., to enable the inclusion of a treatment effect uncontaminated by other treatments. Had data been included after the cross-over, the uncontaminated placebo or control group outcome would not have been available for example.

Trial Summary Statistics

Each trial contained one control or placebo group. Some trials contained more than one treatment (omega-3) group. In order not to double-count patients, we chose the most clinically relevant treatment group to enter our analysis, or in some cases combined treatment groups. For those outcomes that were dichotomous, the summary statistic was a risk ratio, that is, the risk of the outcome in the treatment (omega-3) group divided by the risk of the outcome in the control or placebo group. A risk ratio greater than one indicates that the risk of the outcome in the treatment group is larger than that in the control or usual care arm. For example, if the risk ratio is 1.10, then patients in the treatment group are 1.10 times as likely to have the outcome as those in the control or placebo group.

For each study, we estimated the log risk ratio and its standard deviation. We conducted the analysis on the logarithmic scale for variance-stabilization reasons.[7] We then back-transformed to the risk ratio scale for interpretability.

For those outcomes that were continuous, we extracted the follow-up means and standard deviations for the treatment and control or placebo groups respectively. If a study did not report a follow-up mean, or a follow-up mean could not be calculated from the given data, the study was excluded from analysis. For studies that did not report a standard deviation or for which a standard deviation could not be calculated from the given data, we imputed the standard deviation by using those studies and groups that did report a standard deviation and weighting all groups equally, or we assumed that the standard deviation was 0.25 of the theoretical range for the specific measure in the study. For example, if a study measured pain on a 0-100 scale, we assumed the standard deviation was 25.

If all studies measured the outcome on the same scale or the measures could all be converted to the same scale, e.g., cholesterol measurements measured in mg/dL or mmol/L which could all be converted to mg/dL, the summary statistic was the *mean difference* (MD) between the treatment group follow-up mean and the control or placebo group follow-up mean:

Mean difference = treatment follow-up mean – control follow-up mean

We estimated the standard deviation for that mean difference.[8] If the studies used different measurements of the same outcome and we could not convert them all to the same scale, the summary statistic was an effect size. The effect size is the mean difference at follow-up divided by the pooled standard deviation. This summary statistic is unitless and indicates the number of standard deviations by which the treatment and control or placebo group means differ. We estimated an unbiased estimate[9] of Hedges' g effect size[10] and its standard deviation. A negative

mean difference or effect size indicates that the treatment is associated with a decrease in the outcome at follow-up as compared with the control or usual care group.

Stratification of Trials

For each condition, we performed, as permissible given available data, stratified analyses on subgroups of studies defined by patient population, type of omega-3, and dose of omega-3. We will discuss the particular strata definitions for each condition in the relevant results sections in Chapter 3. In general, a paucity of available data precluded us from pooling data separately in most strata. However, we do discuss the results qualitatively in each stratum when possible.

Performance of Meta-Analysis

In some cases, the trials were judged too clinically heterogeneous to combine. Furthermore, for each outcome, condition, and trial stratum combination, we required that at least three trials be available for pooling. In heterogeneous settings and those with insufficient data, we conduct only a descriptive analysis and present the study-level summary statistics but do not estimate a pooled effect.

For those conditions for which trials were determined to be clinically comparable and for which there were at least three trials, we estimated a pooled random-effects estimate[11] by combining summary statistics across trials. We also report the chi-squared test of heterogeneity p-value.[9]

Forest plots were constructed for each setting. Each individual trial summary statistic is shown as a box whose area is inversely proportional to the estimated variance of the summary statistic in that trial. The trial's confidence interval is shown as a horizontal line through the box. The pooled estimate and its confidence interval are shown as a diamond at the bottom of the plot with a dotted vertical line indicating the pooled estimate value. A vertical solid line at one for dichotomous outcomes or at zero for continuous outcomes indicates no treatment effect.

Sensitivity Analyses

We conducted post hoc sensitivity analysis for meta-analyses that exhibited significant ($p<0.05$) heterogeneity based on the chi-squared test of heterogeneity. In these sensitivity analyses, we removed the most outlying study chosen based on a visual inspection of the forest plot of the original meta-analysis, and estimated a new pooled estimate. We compared this pooled estimate to the original result as well as observed whether significant heterogeneity still remained.

Publication Bias

We assessed the possibility of publication bias by evaluating a funnel plot of summary statistics for asymmetry, which can result from the nonpublication of small trials with negative results. These funnel plots include a horizontal line at the fixed-effects pooled estimate and pseudo–95% confidence limits.[9] If bias due to nonpublication exists, the distribution is asymmetric or skewed. Because graphical evaluation can be subjective, we also conducted an adjusted rank correlation test[10] and a regression asymmetry test[9] as formal statistical tests for publication bias. The correlation approach tests whether the correlation between the effect sizes and their variances is significant, and the regression approach tests whether the intercept of a regression of the effects sizes on their precision differs from zero; that is, both formally test for asymmetry in the funnel plot. We acknowledge that other factors, such as differences in trial quality or true study heterogeneity, could produce asymmetry in funnel plots.

Interpretation of the Results

The mean difference pooled results are readily interpretable as they are measured in a clinically interpretable metric. To aid in interpreting the pooled effect size and risk ratio, whenever possible we back-transformed each pooled estimate to a specific metric. In order to do this, we multiplied each pooled effect size estimate by the average standard deviation of the most clinically relevant outcome measured across the trials, e.g., pain on the VAS scale, included in the pooled estimate. For each pooled risk ratio, we estimated a number needed to treat (NNT) or number needed to harm (NNH) depending on whether the risk ratio was less than or greater to one, by assuming that the population outcome risk was equal to the average control group risk observed across the trials. By average in either calculation, we mean a simple average across relevant placebo/control and/or treatment groups in the relevant studies. We note these back-transformations require assuming a particular underlying standard deviation or outcome risk. Readers may wish to apply their own standard deviation or underlying risk, based on the particular patient population to which they wish to apply the results. We conducted all analyses and drew all graphs using the statistical package Stata.[11]

Peer Review

This draft report was sent for review to a select group of experts in omega-3 fatty acids, epidemiology, nutrition, rheumatoid arthritis, SLE, IBD, nephrology, osteoporosis, and diabetes. The names, expertise, and affiliations of the peer reviewers are listed in Table A.6.1, Appendix A. Additionally, the report was sent to the members of the TEP for review. We entered all comments that we received into a database and collated those pertaining to similar sections of the report. For each comment or group of related comments, we prepared a response detailing how we changed the report or why we did not believe a change was justified. The complete list of peer reviewed comments and our responses are included in Appendix D. Service as a peer reviewer or as a technical expert panelist does not imply agreement or endorsement of the findings of this report.