Chapter 2. Methods

We synthesized evidence from the scientific literature on the effectiveness of behavioral interventions to increase physical activity in the general population, as well as evidence of the effectiveness of physical activity interventions to improve psychosocial and physiologic outcomes for cancer survivors. The methods used for this process were developed by the project team at the Minnesota Evidence-based Practice Center (EPC), in conjunction with representatives from NCI and AHRQ. The Minnesota EPC was established by AHRQ to conduct systematic reviews and technology assessments of all aspects of health care. The Minnesota EPC performs research on improving methods of synthesizing the scientific evidence, developing evidence reports, and conducting technology assessments.

Project staff collaborated with the National Institutes of Health's National Cancer Institute's Division of Cancer Control and Population Sciences, the Task Order Officer at AHRQ, and the Technical Expert Panel for this review on issues related to review topic and methods used.

Scope of Work

The literature review process for this report was divided into two parts. The methodology was similar but not identical for these two parts and will be reported in subsections throughout the remainder of the methodology section.

The literature review process for key question #1, which related to the effectiveness of behavioral interventions to sustainably increase physical activity in the general population was carried out as follows:

- Establish criteria for inclusion of articles in review
- Identify sources of evidence in the scientific literature
- Extract study descriptions, data, and study quality data from studies meeting inclusion criteria
- Attempt to find data that could be synthesized quantitatively
- Summarize findings qualitatively
- Submit results to the technical expert peer reviewers for review
- Incorporate reviewers' comments into a final report for submission to AHRQ

The literature review process for the part of this report on the topic of physical activity in cancer survivors included:

- Establish criteria for inclusion of articles in review
- Identify sources of evidence in the literature
- Extract data from studies meeting the inclusion criteria
- Attempt to find data that could be synthesized quantitatively
- Summarize findings qualitatively
- Submit the results to peer reviewers
- Incorporate reviewers' comments into a final report for submission to AHRQ

Note: Appendixes and Evidence tables cited in this report are provided electronically at http://www.ahrq.gov/clinic/epcindex.htm

Establishing the Technical Expert Panel

A Technical Expert Panel (TEP) was selected to guide the process of refining the key questions and developing the report. Representatives of NCI's Division of Cancer Control and Population Sciences and Minnesota EPC project staff both developed lists of individuals who had content area expertise. There was particular interest in including end users of the evidence report in the TEP. Appendix A lists the individuals who served on the TEP for this report, as well as their areas of expertise.

Developing the Key Questions

The ORIGINAL key questions put forth by AHRQ and NCI were later revised. The original key questions were as follows:

1. What is the evidence that physical activity interventions alone, or combined with diet modification or smoking cessation, are effective in helping individuals in the general population change their aerobic physical activity and maintain an active lifestyle?
    - What settings have been used to deliver behavioral interventions?
    - Are interventions in specific settings more effective than others (e.g., individuals or groups; organizational settings; community settings; public policy)?
    - To what extent have these interventions been delivered to minority or high-risk populations?
    - Is there evidence that effectiveness of interventions varies in minority or high-risk populations?
    - Determine the factors that mediate or moderate the success of these interventions (e.g., gender, race/ethnicity, intervention type, incentives, intervention dose, length of intervention, intervention mode of delivery, exercise setting, physical activity mode, physical activity intensity, research design, other).
2. Are interventions that use behavioral theories more effective in changing aerobic physical activity than those that do not?
    - What theories have been used to design physical activity interventions and to what extent have they been implemented?
    - Are interventions that use particular behavioral theories more effective than others in changing behaviors?
    - Do behavioral interventions have a significant impact on theoretically hypothesized mediators of physical activity?
    - Determine the factors that moderate the success of theoretical interventions (e.g., gender, race/ethnicity, intervention type, incentives, intervention dose, length of intervention, intervention mode of delivery, exercise setting, physical activity mode, physical activity intensity, research design, other).
3. What is the evidence that physical activity interventions, alone or combined with diet modification or smoking cessation, are effective in helping cancer survivors improve their psychosocial outcomes (e.g., anxiety, depression, fatigue, and quality of life) or physiological outcomes (e.g., cardiorespiratory fitness, obesity/total fat/visceral fat, insulin, Insulin-like Growth Factors (IGFs) and IGF binding proteins, and sex hormones steroids and binding proteins)?

Refining the Key Questions

Examining Past Systematic Reviews

A number of systematic reviews had previously been done examining different aspects of the preliminary key questions. It was the desire of AHRQ, NCI, and the Minnesota EPC that this report make a contribution to the literature. This required an understanding of the focus and conclusions of prior systematic reviews in this topic area. A search was undertaken to identify previous systematic reviews that overlapped with the preliminary key questions and to examine what key questions had been addressed in those reviews. A document outlining key questions addressed by prior reviews was prepared for the Technical Expert Panel for a face-to-face meeting in January 2003.

Meeting of the Technical Expert Panel and Refinement of the Key Questions

The Technical Expert Panel, representatives from NCI, AHRQ, and the Minnesota EPC met face to face on January 29, 2003, to discuss the refinement of the key questions and the development of the report. The group expressed interest in the role of mediators and moderators but did not at that meeting refine the questions further. A series of discussions between AHRQ, NCI, and the Minnesota EPC was held after the expert meeting and additional input was gathered from the Technical Expert Panel where appropriate. The issue was how to refine the key questions so that they would address an area not otherwise addressed in the literature and that would be achievable within the contract. The result of the discussions was that one key factor that had not been as completely addressed in previous reviews was whether interventions to increase physical activity had effects that lasted beyond the end of the intervention period itself. It was decided then that the review would be limited to studies that examined outcomes at least three months after the end of the intervention. At the January meeting, the TEP had suggested excluding studies that were done in the context of acute disease (such as cardiac rehabilitation) and this criterion was added to the exclusions. The revised and final key questions with inclusion and exclusion criteria were:

1. What is the evidence that physical activity interventions alone, or combined with diet modification or smoking cessation, are effective in helping individuals in the general population increase their aerobic physical activity or maintain adequate aerobic physical activity?
    a. Is the effectiveness of theoretically based interventions different?
    b. Do hypothesized moderators affect the results of these interventions?
    c. Do these interventions affect theoretically hypothesized mediators?
    d. In these interventions, is there a relationship between changes in theoretically hypothesized mediators and changes in physical activity?
2. What is the evidence that physical activity interventions, alone or combined with diet modification or smoking cessation, are effective in helping cancer survivors improve their psychosocial outcomes (e.g., anxiety, depression, fatigue, and quality of life) or physiological outcomes (e.g., cardiorespiratory fitness, obesity/total fat/visceral fat, insulin, IGFs and IGF binding proteins, and sex hormones steroids and binding proteins)?

We planned that in answering the first key question, we would also address the following subsidiary questions:

1. What theories have been explicitly used?
2. What theoretical constructs have been implemented within interventions explicitly based on a particular theory or theories?
3. To what extent have mediators been appropriately tested?

To be included, a study must meet the following inclusion criteria:

Key question #1.

- Study must include an intervention designed to increase physical activity
- Study must include a measure of whether physical activity is affected by the intervention. Fitness is an acceptable surrogate measure of physical activity if it was intended for that purpose.
- Study must include a concurrent comparison group (studies that instructed control group participants to avoid exercise were excluded on the basis that those studies were focused on physiologic outcomes, not changes in physical activity behavior)
- Studies with all age groups will be included
- Study must be published in the English language

Key question #2.

- Study must be focused on individuals who have been diagnosed with cancer
- Study must include an intervention designed to increase physical activity
- Study must include a concurrent comparison group
- Study must include adults
- Study must be published in the English language

Exclusion Criteria

Studies with the following characteristics will be excluded from the review:

Key question #1.

- Study has fewer than 75 subjects total between the intervention and comparison group
- Study reports less than three month post-intervention followup data
- Study targets specifically:
    - Individuals with acute disease of any kind
    - Individuals with coronary heart disease, peripheral vascular disease, or cerebrovascular disease
    - Individuals with cystic fibrosis
    - Individuals with osteoarthritis
    - Institutionalized individuals (nursing homes residents or prisoners)
- Studies of cardiac rehabilitation programs
- Studies of rehabilitation/physiotherapy interventions

The Technical Expert Panel discussed at length the advantages and disadvantages of including or excluding from key question 1 any studies that targeted individuals with chronic or acute diseases. After the TEP meeting in January, the Minnesota EPC was guided further by AHRQ and NCI to include studies with diabetic or obese individuals, but not studies with other chronic or acute diseases. The rationale was that the impact of behavioral interventions on individuals with the excluded diseases might differ from the impact on included individuals.

Key question #2.

- Studies with no intervention designed to increase physical activity
- Studies with no concurrent comparison group
- Studies conducted in children only
- Studies published in languages other than English

Definition of followup. Studies that reported less than three months of post-intervention followup data were excluded in the review for key question #1 only. The definition of this time interval is illustrated in Figure 2. The followup period was defined as starting when contact from the investigators intended to affect the physical activity of the subjects concluded. Contact for measurement was allowed. It should be noted that individual investigators may not have defined followup this way, so the length of followup reported by the investigators within the publications reviewed may not be the same as the length of followup within this report. For example, a physical activity intervention may be faded over time, such that an intensive intervention may last for six months, followed by a minimal maintenance intervention for 18 months. The investigators may consider the 18 month period after the end of the intensive intervention to be a followup period. By contrast, in this report, followup would not start until the end of the 18 months of minimal maintenance intervention.

Literature Search Design

Identification of Literature Sources

Potential evidence for the report came from online databases, reference lists of all relevant articles and reviews, and files of project staff and TEP members with specific expertise in behavioral physical activity interventions and/or research on physical activity in cancer survivors. MEDLINE® was used as the only online database. In the process of peer review, only one paper from the cancer survivor literature and no papers from the general population literature were identified as missing. This indicates that although the search was not repeated in additional online databases, the existing literature for the key questions to be addressed was comprehensively identified.

Search for key question #1. A MEDLINE® search was performed to identify trials of physical activity interventions. The specific search strategy is shown in Appendix B. The titles and, if necessary, the abstracts of the results of this search were reviewed by an expert in physical activity interventions to identify references that required full review. We also identified previous systematic reviews of physical activity interventions in the literature.[16, 19-25, 27-34, 36, 37, 44, 45] The titles and, if necessary, the abstracts of all of the references in those reviews were also reviewed by an expert in physical activity interventions to identify references that required full review. These two lists were combined and all references were reviewed to determine whether the studies met inclusion criteria.

Additional references were identified in two ways. The list of references meeting inclusion criteria was shared with the Technical Expert Panel. TEP members could then suggest references missed that should be included.

The second manner in which additional references were added was in the process of reviewing the papers for whether they met inclusion criteria. During this review, the abstractors

identified other references that should be reviewed. Abstractors primarily identified other references related to the study in the reference under review.

Search for key question #2. For the part of the review on physical activity interventions in cancer survivors, we conducted two MEDLINE® searches with separate search strategies to identify possible papers for inclusion in the review. The strategies for these searches are included in Appendix B. We limited our search to English language papers. The first of these searches, conducted on July 6, 2003, yielded 16 papers. In July we also reviewed the references of a recent review on the topic of exercise in cancer survivors[7] and identified 39 additional papers. The second MEDLINE® search was conducted on September 17, 2003, and yielded 73 papers. These lists were combined using the bibliographic software EndNote® and duplicates were deleted, yielding a total of 128 titles. These titles were reviewed by a representative at NCI, a member of the TEP with special expertise in physical activity and cancer research, and a member of the project team to see if there were any papers obviously missing. Several additional titles were suggested by the TEP member, and several were deleted as well, resulting in a total of 128 titles to be reviewed for inclusion. Two additional articles were identified during the process of reviewing the papers included in the review. In the process of peer review, 14 additional references were identified.

Evaluation of Evidence

Data Collection

Question 1: General population. A data collection form was developed using the Guide to Community Preventive Services data abstraction instrument as a template.[46] All information necessary to the review was collected on this form except for the specific outcome information (see below). This instrument was reviewed by the Technical Expert Panel and relevant changes were made. The instrument was then computerized as a Filemaker Pro database to allow computerized data entry. This instrument was then pilot tested by the data abstractors and a member of the project staff with expertise in physical activity interventions. Revisions were made to the Filemaker Pro data entry screens until entry could be efficiently and accurately accomplished.

One member of a team of four data abstractors reviewed each reference identified for full review. The data abstractors all had expertise in the area of physical activity research. Any reference that was felt to not meet inclusion criteria was reviewed by another member of the team and if there was disagreement, the reference was brought to the full group. Each included reference underwent a second full review by a senior member of the team and any questions were discussed with the full team.

A second data abstraction form was developed for the abstractions of the specific outcome data from the included studies. This abstraction was done in an Excel spreadsheet. The outcomes of the included studies were abstracted by one of two members of the team with significant experience in abstracting outcomes for systematic reviews. The specific outcomes to be abstracted from each reference were reviewed by the entire team. The outcomes were re-abstracted as a check by a senior member of the research team and where there were questions, discussed with the full team.

Question 2: Cancer survivors. Detailed information about all but the outcomes from each of the 24 included trials was collected on a specialized data collection instrument (the Cancer Abstraction Form) designed for this purpose. This Cancer Abstraction form (Appendix C) was developed in consultation with a representative of NCI (Louise Masse) and a member of the TEP with special expertise in exercise and cancer survivors (Kerry Courneya). We included questions about trial design, study quality, the number and characteristics of participants, participant recruitment information, and details on the intervention (such as dose of exercise and non-exercise components). The outcome data were initially abstracted by a member of the project staff in Excel, just to list outcomes. This listing was checked by a second member of the project staff. Then tables of study descriptions and outcomes were developed. These tables were reviewed and checked by a second project staff member as well. All abstraction was checked by a second project staff member, though independent double abstraction was not conducted.

Two project staff members, both trained in the critical analysis of scientific literature, independently reviewed each of the identified articles to determine eligibility. The data abstraction was first conducted by a research assistant who had been trained in data abstraction procedures; then each abstraction was checked by a PhD trained member of the project staff with content area expertise in physical activity and cancer research. From the 53 articles initially reviewed for eligibility, 29, representing 25 trials, were accepted for further study. During the process of peer review, an additional 14 papers were identified. One of the 25 studies initially included was deemed unacceptable by peer reviewers, since it focused on physiotherapeutic exercise to increase shoulder range of motion after mastectomy.[47] One additional paper that reported outcomes for the remaining 24 studies was identified during peer review. The remaining 13 papers identified during peer review were excluded and have been added to the final list of excluded papers.

All outcomes were acceptable for abstraction for this part of the report, as one of the goals was to assess what outcomes have been included in this literature. The 29 articles presented data on 24 trials. In five cases, there were two articles that presented information for a single trial. To be clear, a 'trial' refers to a controlled clinical trial; an 'article' refers to a published document. An article may present more than one trial, or a trial may be described in more than one article. Trial is the unit for summarizing the results of the review.

To evaluate the quality of the study design and execution of trials, we collected data in a format that was based on the abstraction form developed by the Guide to Community Preventive Services[46] and shared with this project staff by a TEP member who had worked on the exercise/physical activity portion of that task force. For each trial, 11 questions were answered in four categories related to description of the study and participants, study measurement quality, analytic approach, and interpretation of results.

Data Synthesis: General population

Developing a common metric. The original methodologic plan was to attempt to pool main effects across included studies as well as compare the effects of subgroups such as populations studied or intervention type. Such data pooling requires a common metric that can be applied to each study. Because the goal of each of the included studies is to increase physical activity, and physical activity requires energy expenditure, the hope was that the outcome of a significant portion of the studies could be expressed as energy expenditure. A subset of the studies reported

energy expenditure and an additional group of studies reported sufficient data to calculate energy expenditure (e.g. time spent exercising and exercise intensity).

We were able to calculate energy expenditure for less than 12 of the 47 included studies on the general population. In those studies where energy expenditure was given or could be calculated before and after intervention in both control and intervention groups, we sought to compute a common intervention effect estimate (IEE). The IEE we sought to calculate for studies that reported energy expenditure (or enough data to calculate energy expenditure) is more commonly termed the 'raw mean difference.'[48] IEE is calculated as follows:

IEE = (PostT - PreT) - (PostC - PreC), where T indicates treatment group and C indicates control group.

The variance of this measure is calculated as:

Var(IEE) = Var(PostT - PreT) + Var(PostC - PreC) =
Var(PostT) + Var(PreT) - 2*SQRT[Var(PostT)Var(PreT)]Cor(PreT,PostT) + Var(PostC) + Var(PreC) - 2*SQRT[Var(PostC)Var(PreC)]Cor(PreC, PostC).

Calculation of the above variance clearly depends on the pre-post correlation. This correlation was not routinely available in the articles in question, nor was a measure of the standard deviation of the difference in means. Therefore it was not felt to be possible to derive one common metric from which to calculate the IEEs.

Calculating effect size from all outcomes. Because the diversity of outcomes prevented derivation of a common measure of physical activity for all studies, we elected instead to calculate effect size (e.g. standardized mean difference) from the outcomes represented across the studies. We did this to aid the interpretation, as it may be easier to compare studies using a single outcome measure, effect size, than the diversity of outcomes reported in the included studies. However, the results still reflect different outcomes and different underlying measurement domains; therefore, it is not necessarily reasonable to directly compare the results of two individual studies without examining the outcome measure underlying the effect size. This issue, as it relates specifically to this literature, is discussed in more detail in the results section of this report.

The effect sizes were calculated using the software program ES.[11] Effect sizes (e.g. standardized mean differences between the treatment and control group(s)) were calculated from all outcomes where one of the following combinations of data was available. (Note: We quote here from the ES software manual[11]):

- "Raw score means, standard deviations, and sample sizes OR
- Dichotomous outcomes in two by two tables with cell frequencies OR
- Dichotomous outcomes in two by two tables with chi-square and total sample size OR
- Between-groups t-test on raw post test scores OR
- Raw means and sample sizes on three or more groups, with a t-statistic comparing one group to a combination of other groups OR
- T-test for two matched groups, sample sizes, correlation between groups OR
- Between-groups t-test on change scores with intraclass correlation OR

- Change-score means and change-score standard deviations with intraclass correlation OR
- Two-group between-groups oneway F-statistic on change scores with intraclass correlation OR
- Change score means and sample sizes on three or more groups, t-statistic comparing one group to a combination of the other groups OR
- Two-group between-groups F-statistic on raw posttest scores OR
- Probability level and sample size for groups OR
- Coding results described only as significant if sample size for groups is known."[11]

Assumptions were made regarding missing information where it was felt it could be reasonably assumed. One example is assuming sample size when an enrollment number was available and there were sufficient clues as to the number analyzed at followup even if it was not stated. When an exact p-value was not given for a statistically significant study, it was assumed to be 0.05. This may systematically bias the effect size downward. If the p-value were actually smaller, the effect size would be greater. Where the within-person repeated measures correlation coefficients (intraclass correlation coefficients) for the outcome variable were missing for studies that reported change scores, it was assumed to be 0.6. It should be noted that for some studies it was not possible to incorporate baseline values into the effect size calculation because of inadequate information regarding the correlation of repeated measures. If the intervention and control groups were different at baseline, this difference could bias the post effect size. Given the important issues with the calculation of the effect size, the reader should understand that what is reported gives a reasonable approximation of the effect of the studies but is inexact.

We elected not to perform any mathematical pooling of the results for the general population. The studies differed in terms of intervention type, study duration, patient populations, outcome measures, and clinical outcomes. Although it would have been mathematically feasible to perform a quantitative meta-analysis, it was not clear that the numbers obtained would have any meaning. We elected instead to present the effect size results themselves so that the reader could understand the distribution of effect sizes within the diverse populations rather than reducing that distribution to a point estimate of questionable validity. The other metric examined is whether a study had statistically significant positive results. This criterion likely underestimates the results of the studies but is able to provide an additional level of understanding to the report of the effect sizes alone.

Data Synthesis: Cancer Survivors

For the portion of the review on the topic of physical activity in cancer survivors, effect sizes were also calculated using the software program ES.[11] Effect sizes (e.g. standardized mean differences between the treatment and control group(s)) were calculated from all outcomes where one of the following combinations of data was available. (Note: We quote here from the ES software manual[11]):

- "Raw score means, standard deviations, and sample sizes at post intervention
- Between-groups t-test on raw post-test scores."[11]

No change score effect sizes were possible in this section of the report given lack of information regarding the correlation of pre- and post-intervention values for the wide variety of

outcomes assessed. For studies in which there were no between group differences at baseline, this post-test effect size is a more acceptable measure of the impact of the intervention on the outcome. However, unlike the general population section, effect size calculations were made regardless of whether there were between group differences at baseline. Comments on interpretation of effect sizes in the cancer survivor literature are provided in the results section.

Publication Bias

The great variations in populations, interventions, and outcomes make the usual techniques for detecting publication bias both impractical and unreliable. It would be difficult to conclude that variations in outcome seen with varying trial size was related to publication bias and not confounded by any of the many other ways that the trials differed from each other. We do present the effect sizes and statistical significance of studies by study size, which provides some information about the possibility of publication bias. Yet, this is limited by possible confounding by differences in the studies as well as the fact that negative studies may be more likely to not allow a calculation of effect size (as they are less likely to present variance estimates or exact p-values for non-statistically significant outcomes).