Spence, R. R. and Heesch, K. C. and Brown, W. J. (2009) A systematic review of the association between physical activity and colorectal cancer risk. *Scandinavian Journal Of Medicine & Science In Sports*, 19(6). p. 764.

TITLE: A systematic review of the association between physical activity and colorectal cancer risk

Authors: R.R. Spence[1], K.C. Heesch[1], W.J. Brown[1]

[1] The University of Queensland, School of Human Movement Studies, Qld 4072, Australia

Running title: Colorectal cancer and physical activity.

Word Count: 4840

Corresponding author:
Rosalind R Spence
Mailing: The University of Queensland, School of Human Movement Studies, Qld 4072, Australia
Tel: +617 3365 6441
Fax: +617 3365 6877
E-mail: rspence@hms.uq.edu.au

Abstract

This review evaluated the strength of the evidence for a causal relationship between physical activity (PA) and colorectal cancer (CRC). A systematic review of databases through February 2008 was conducted to identify studies that assessed the association between total or recreational PA and incidence or mortality of CRC (including CRC, rectal cancer, colon cancer, and proximal or distal colon cancer). Studies were evaluated for significant associations between PA and risk of CRC endpoints and for evidence of dose-response relationships in the highest quality studies. Twenty cohort studies were evaluated; 11 were high quality. Fifty percent of all *studies* and 64% of highest quality studies reported at least one significant association between PA and risk of a CRC endpoint ($p < 0.05$). However, only 28% of all *analyses* (31% of analyses of highest quality studies) were significant ($p < 0.05$). Only forty percent of analyses of highest quality studies resulted in a significant p for trend ($p < 0.05$); however, a non-significant inverse linear association between PA and colon cancer risk was apparent. Heterogeneity in the evidence from all studies and from the highest quality studies was evident. Evidence from cohort studies is not sufficient to claim a *convincing* relationship exists between PA and CRC risk.

Abstract word count: 199 words

Background and Purpose

A growing volume of epidemiological evidence supports the importance of physical activity (PA) for the primary and secondary prevention of chronic diseases, such as coronary heart disease, stroke, hypertension, type 2 diabetes mellitus, osteoporosis, and cancer (American College of Sports Medicine, 1998; Blair & Brodney, 1999; Warburton et al., 2007). However, there is still much that we do not know about these associations, especially in the field of cancer prevention.

Colorectal cancer (CRC) is the third most common cancer worldwide (Parkin et al., 2005). Incidence rates vary 25-fold between countries, with the highest rates observed in Australia, North America, and Western Europe and the lowest rates in Africa and Asia (Parkin et al., 2005). Understanding the risk factors for this disease is integral to the development of effective strategies for the prevention of CRC.

The World Cancer Research Fund/American Institute for Cancer Research's (WCRF/AICR) second expert report on Food, nutrition, physical activity and cancer prevention (World Cancer Research Fund/American Institute for Cancer Research, 2007) provides guidelines for the evaluation of scientific evidence, that can be used to determine whether there is a causal relationship between PA and cancer. The guidelines include six criteria, of which the first three determine whether there are sufficient studies to allow the strength and quality of the evidence to be graded. The first criterion (#1) requires evidence from more than one study type (i.e., from both case-control and cohort studies), and the second (#2) requires evidence from at least two independent cohort studies. The third (#3) requires the presence of 'good quality' studies, defined as those that are free from potential sources of bias or from random or systematic error. Where sufficient evidence is available to meet these three criteria, a rating of 'limited' to 'convincing' is based on three additional criteria that refer to the strength of the evidence. To receive a rating of 'convincing', there must be no substantial unexplained heterogeneity among studies relating to the presence, absence or direction of an association (#4); the presence of a plausible biological gradient ('dose response') in the association (#5); and strong experimental evidence of biological plausibility (#6).

The three most recent comprehensive reviews of the evidence relating inactivity to risk of CRC have used both qualitative and quantitative methods to address the first two WCRF/AICR criteria and to evaluate the relationship between PA and CRC (Friedenreich, 2001; Slattery, 2004; Samad et al., 2005). As most studies conducted prior to these reviews were case-control studies, the three reviews tended to focus on this type of study; they did however include results from the few cohort studies that had been published prior to the reviews.

3

Friedenreich (2001) applied an earlier version of the WCRF/AICR criteria (World Cancer Research Fund/American Institute for Cancer Research, 1997) and concluded that there was convincing evidence of a causal relationship between PA and colon cancer but not for rectal cancer. In her review, Slattery (2004) agreed by concluding that PA reduces the risk of colon cancer, but that the evidence was less clear for rectal cancer. She suggested that 3.5 to 4 hours/week of vigorous PA may be needed to optimise protection. Finally, in a meta-analysis, Samad (2005) also found a protective effect of PA for colon cancer, but not for rectal cancer. As each of these reviews included a number of case-control studies and a sufficient number (i.e. two) of cohort studies, they provide sufficient evidence to meet the first two WCRF/AICR criteria.

The third WCRF/AICR criterion calls for the quality of the evidence to be evaluated. Unfortunately, many of the early case-control studies described in the three reviews had numerous sources of potential error or bias, suggesting these studies did not meet the third criterion. Case-control studies collect data on PA after cancer diagnosis. They are therefore subject to differential recall bias, which occurs when participants' status as cases or controls differently affects their reporting of PA behaviour. Similarly, in case-control studies, participants' status may influence whether or not they agree to participate in the study, which can lead to selection bias. Moreover, to measure PA the early case-control studies tended to rely primarily on job title or a single crude item (e.g. "are you more or less active than the average person your age?"). Such measures do not allow for computation of energy expenditure or 'volume' of PA and create the potential for misclassification of individual PA levels. Additionally, as Slattery (2004) noted, case-control studies often measure PA 1-2 years prior to diagnosis because recall of earlier PA behaviour may be less accurate. That time period, however, may not be the critical one for assessing the association between PA and CRC risk, given that CRC takes 10-15 years to develop from its precursor (the adenoma). These limitations suggest that, even though the first two criteria have been met, the quality of the studies included in previous reviews was not adequate to meet the third WCRF/AICR criterion.

The problem of adequate study quality has been addressed in recent years by the availability of results from several high quality cohort studies, which have been published since the publication of the earlier reviews. These studies are not completely free from potential sources of bias, especially if there is selective reporting of significant findings. However, they do minimise the potential for selection and recall bias because they typically have included large numbers of participants and they have measured PA many years prior to diagnosis. Moreover, their use of multiple items to measure frequency, intensity, duration and type of PA is more robust than the single items that were often used in the early case-control studies. These cohort studies are therefore more likely to meet the third ('good

quality') criterion than the early case control studies that comprised most of the studies included in prior reviews.

Although the previous reviews (Friedenreich, 2001; Slattery, 2004; Samad et al., 2005) included the few early studies that reported dose-response relationships, there was insufficient evidence from high quality studies at the time to examine the WCRF/AICR fourth and fifth criteria. The more recent cohort studies provide additional evidence that can be used to evaluate the homogeneity of study findings (i.e., presence or absence of relationships) and the presence of plausible (dose-response) relationships.

The purpose of this paper is to update the strength of the evidence for an inverse relationship between PA and risk of CRC. The work builds on the earlier reviews, which clearly provide evidence to satisfy the first two WCRF/AICR criteria. This review uses published cohort studies, including the most recently published studies that were not included in previous reviews, to evaluate whether findings from good quality studies (criterion #3) support WCRF/AICR criteria #4 and #5 and thus to judge the overall strength of the evidence.

As cohort studies have limited ability to address the issue of biological plausibility, the final criterion (#6) is not considered in this review. However, it should be noted that the biological mechanisms influencing the relationship between PA and colon cancer may differ from those for PA and rectal cancer. Readers are referred to a review paper by Quadrilatero & Hoffman-Gertz (2003) for a summary of the most widely accepted hypotheses on these issue.

Specifically, this paper aims to answer two research questions:

- Is there homogeneity between study findings in the reporting of the presence or absence of associations between PA and CRC (i.e., does the evidence support WCRF/AICR criterion #4)?
- Do the findings from the highest-quality studies provide consistent evidence of a plausible gradient (dose-response) in the relationship between PA and CRC (i.e., does the evidence support WCRF/AICR criterion #5)?

Methods

Selection of studies for the review

PubMed and Embase databases were searched using a comprehensive list of search terms (Table 1). The criteria for inclusion were that the studies: (1) were prospective cohort or nested case-control studies; (2) were published in English prior to February 1, 2008; (3) evaluated the relationship between PA and CRC risk (endpoints included CRC and colon, rectal, distal or proximal colon cancers); (4) included a cohort of men, women, or both men and women, who did not have CRC at baseline; and (5) provided risk estimates and corresponding 95% confidence intervals for the association between PA and CRC. Studies

that included cohorts of specific groups whose risk of CRC were likely to be different from that of the general population, such as elite athletes or smokers, were excluded.

INSERT TABLE 1 ABOUT HERE

Titles and abstracts were reviewed to exclude any studies that clearly did not meet the inclusion criteria. Full texts of the remaining studies were then reviewed to ensure they met the inclusion criteria, and reference lists of relevant articles and past reviews were searched for studies not identified in the initial search.

Data extraction methods

The Cochrane Data-Extraction Guidelines for non-randomised trials (Cochrane Non-randomised Studies Methods Group, 2008) were used to extract and summarise the list of studies. Reports of preliminary or early results that were updated in subsequent manuscripts were identified during data extraction and excluded from the analyses.

As there was considerable heterogeneity in study design and in measurement of PA, no attempt was made to estimate overall quantitative effect sizes. Instead, the strength of the evidence was rated using criteria #4 and #5 of the WCRF/AICR guidelines, which relate to the homogeneity of the evidence and the presence of a biologically-plausible dose-response.

Analysis of the homogeneity of the evidence

To examine the homogeneity of the evidence (i.e., criterion #4 and research question 1), the proportion and number of analyses that resulted in statistically significant findings ($p<0.05$) were calculated and compared with the total proportion and number of associations examined in these studies. To determine whether the number of significant associations that was reported in previous reviews reflects a true association between PA and the endpoints, or whether associations were due to type 1 error as a result of selective discussion of only the significant relationships within any study, these calculations were conducted separately for each endpoint reported (e.g. colorectal, colon, rectal, proximal colon and distal colon cancer) as well as for all endpoints combined. The number of studies that reported more than one significant association was also calculated, in order to assess whether the overall strength of the evidence could be influenced by a small number of studies that found significant associations across multiple endpoints.

Differences in endpoints between individuals in the highest and lowest PA categories were compared, following the method used by the WCRF/AICR panel to examine associations between exposures and outcomes (World Cancer Research Fund/American Institute for Cancer Research, 2007). As multiple measures of PA were included in some

6

studies, the analysis that included the most comprehensive PA measure was selected for inclusion. For example, if a study included separate analyses of occupational, recreational and total PA, only the analysis of total PA was selected. If a study did not have a measure of total PA, the recreational PA measure providing the most detailed information on frequency, duration and intensity was included, as this was deemed the most comprehensive measure provided. In some studies PA was measured for a number of time periods across the lifespan. For these analyses, the time period selected for inclusion was the period closest in time to the assessment (e.g. PA over the previous 3 years rather than PA over the previous decade). This allowed for the greatest possibility of homogeneity in the time period used to measure PA across studies, given that most studies assessed PA over the year prior to the assessment. This inclusion criterion also allowed for the least possibility of recall bias. Furthermore, if only occupational PA was measured, the study was not included. This decision was made for a number of reasons. First, occupational PA was generally measured in a crude manner, usually by job title. Moreover, in the earlier studies, many of the female participants were not employed full-time, and, therefore, assessment of only occupational PA could not accurately reflect their PA levels.

Analysis of the homogeneity of the evidence in the highest quality studies

To address research question 2 (and criterion #5), only the highest quality studies were included. These were studies that measured volume of PA (defined here as intensity and duration) and were appropriate for evaluating dose-response (i.e., PA was categorised into more than two categories). Again, the most comprehensive PA measure was used, and each endpoint was analysed separately: first by comparing the highest and lowest PA volume categories and then by identifying the 'p for trend' across the PA categories for each study. The proportion of significant 'p for trend' ($p < 0.05$) values was determined for each endpoint.

Results

Selection of studies for the review

We identified 1068 potentially relevant articles on PA and CRC risk from PubMed and 1749 from Embase (Figure 1). After removing duplicates (n=342) and articles that did not meet the inclusion criteria based on title and abstract (n=2375), the full text of 100 articles was assessed for inclusion. Eighty-two articles were excluded because they used a cross-sectional or retrospective study design (n=48); either participants or exposure/outcome measures did not meet inclusion criteria (n=27); insufficient information was provided to describe the PA measure (n=1); or the article was part of multiple publications on the same source population (n=6). Two articles that were identified from the reference list of other articles were added to the list of studies to be considered. Twenty studies (Wu et al., 1987; Gerhardsson et al., 1988;

Severson et al., 1989; Ballard-Barbash et al., 1990; Thun et al., 1992; Lee & Paffenbarger, 1994; Giovannucci et al., 1995; Thune & Lund, 1996; Lee et al., 1997; Davey Smith et al., 2000; Chao et al., 2004; Schnohr et al., 2005; Calton et al., 2006; Friedenreich et al., 2006; Johnsen et al., 2006; Larsson et al., 2006; Lee et al., 2007; Mai et al., 2007; Wolin et al., 2007; Nilsen et al., 2008) were identified that met the inclusion criteria and were therefore included in this review.

INSERT FIGURE 1 ABOUT HERE

Characteristics of studies included in the review

Characteristics of the included studies are summarized in Table 2, and the relative risk of each cancer endpoint reported in each study is summarised in Table 3. The studies were published between 1987 and 2008. The cohorts ranged in size from 4,214 (Ballard-Barbash et al., 1990) to 413,044 (Friedenreich et al., 2006). Five studies included cohorts of men only (Severson et al., 1989; Lee & Paffenbarger, 1994; Giovannucci et al., 1995; Lee et al., 1997; Davey Smith et al., 2000); three included cohorts of women only (Calton et al., 2006; Mai et al., 2007; Wolin et al., 2007); and the remaining 12 included mixed cohorts (Wu et al., 1987; Gerhardsson et al., 1988; Ballard-Barbash et al., 1990; Thun et al., 1992; Thune & Lund, 1996; Chao et al., 2004; Schnohr et al., 2005; Friedenreich et al., 2006; Johnsen et al., 2006; Larsson et al., 2006; Lee et al., 2007; Nilsen et al., 2008). Most studies were conducted in the United States (n=11) or Europe (n=8). One was conducted in Japan. There was considerable variation among the studies in the PA measure used, the covariates included, and the duration of follow-up. Five studies assessed total activity only. Nine others measured recreational PA only; and an additional six assessed a mixture of occupational, recreational and/or domestic PA. PA measures included daily or weekly time spent in specific activities or activities of a specified intensity; total energy expenditure; or summary scores based on frequency, intensity and duration of PA. The cancer outcome was mortality in two studies (Thun et al., 1992; Davey Smith et al., 2000) and both incidence and mortality in the other 18 studies (Wu et al., 1987; Gerhardsson et al., 1988; Severson et al., 1989; Ballard-Barbash et al., 1990; Lee & Paffenbarger, 1994; Giovannucci et al., 1995; Thune & Lund, 1996; Lee et al., 1997; Chao et al., 2004; Schnohr et al., 2005; Calton et al., 2006; Friedenreich et al., 2006; Johnsen et al., 2006; Larsson et al., 2006; Lee et al., 2007; Mai et al., 2007; Wolin et al., 2007; Nilsen et al., 2008). Most studies (13 of the 20) reported associations for more than one endpoint, with a maximum of five endpoints (Larsson et al., 2006; Lee et al., 2007) and a minimum of one (Gerhardsson et al., 1988; Ballard-Barbash et al., 1990; Thun et al., 1992; Lee et al., 1997; Davey Smith et al., 2000; Schnohr et al., 2005; Johnsen et al., 2006). Endpoints included CRC in five studies (Wu et al., 1987; Ballard-Barbash et al., 1990; Davey Smith et al., 2000;

Larsson et al., 2006; Lee et al., 2007)), colon cancer in 17 studies (Gerhardsson et al., 1988; Severson et al., 1989; Thun et al., 1992; Lee & Paffenbarger, 1994; Giovannucci et al., 1995; Thune & Lund, 1996; Lee et al., 1997; Chao et al., 2004; Schnohr et al., 2005; Calton et al., 2006; Friedenreich et al., 2006; Johnsen et al., 2006; Larsson et al., 2006; Lee et al., 2007; Mai et al., 2007; Wolin et al., 2007; Nilsen et al., 2008)), rectal cancer in nine studies (Severson et al., 1989; Lee & Paffenbarger, 1994; Giovannucci et al., 1995; Thune & Lund, 1996; Chao et al., 2004; Friedenreich et al., 2006; Larsson et al., 2006; Lee et al., 2007; Nilsen et al., 2008)) and proximal or distal colon cancer in 11 studies (Wu et al., 1987; Giovannucci et al., 1995; Thune & Lund, 1996; Chao et al., 2004; Calton et al., 2006; Friedenreich et al., 2006; Larsson et al., 2006; Lee et al., 2007; Mai et al., 2007; Wolin et al., 2007; Nilsen et al., 2008)). Covariates included in the multivariable analyses ranged from age and BMI only, to complex lists of family and medical histories and lifestyle and demographic characteristics. Follow-up duration ranged from 5 years (Wu et al., 1987) to up to 28 years (Ballard-Barbash et al., 1990), with a mean follow-up of 12.4 years.

Eight studies (Calton et al., 2006; Friedenreich et al., 2006; Johnsen et al., 2006; Larsson et al., 2006; Lee et al., 2007; Mai et al., 2007; Wolin et al., 2007; Nilsen et al., 2008) were published for the first time or had published updated results after the publication of the last comprehensive review (Samad et al., 2005) of PA and CRC risk.

INSERT TABLES 2 & 3 ABOUT HERE

Homogeneity of the evidence

Fifty percent of studies (10 of 20 studies) reported at least one statistically significant association, and four of these reported statistically significant associations for more than one endpoint. Of these four, two reported statistically significant associations for the endpoints of colon and distal colon cancer (Giovannucci 1995, Nilsen 2008), one for the endpoints of colon and proximal colon cancer (Chao 2004) and one for the endpoints of CRC, colon and proximal colon cancer (Lee 2007).

However, when individual associations across all studies were examined, only 28% (15 of 53 analyses) of all reported associations between PA and an endpoint (i.e., CRC, colon, rectal or proximal or distal colon cancer) were statistically significant. All were inverse associations, showing a reduction in cancer risk for the highest PA category when compared with the lowest PA category (Figure 2).

The highest proportion of statistically significant associations for any endpoint was 60% (3 of 5 analyses) for CRC. In these analyses (Wu et al., 1987; Ballard-Barbash et al., 1990; Lee et al., 2007) the associations were statistically significant for men but not women. The proportion of statistically significant associations for the other endpoints ranged from 0% (of 9 analyses) for rectal cancer to 35% (6 of 17 analyses) for colon cancer. For colon cancer,

three of six associations that reached statistical significance were only significant for men (Severson et al., 1989; Giovannucci et al., 1995; Lee et al., 2007); the others found significant associations for both sexes. Only 27% (3 of 11 analyses) of the examined associations for both proximal and distal colon cancer endpoints reached statistical significance. Two of the three proximal colon cancer associations (Chao et al., 2004; Friedenreich et al., 2006) were statistically significant for both sexes, but one (Lee et al., 2007) was significant only for men. Of the three statistically significant associations found for distal colon cancer, one was for men (Giovannucci et al., 1995), one was for women (Wolin et al., 2007) and one was for men and women (Nilsen et al., 2008).

Because no studies reported a statistically significant association between PA and rectal cancer, the proportion of all reported associations was recalculated without the rectal cancer endpoint. With rectal cancer excluded, the proportion of statistically significant associations increased from 28% (15 of 53 analyses) to 34% (15 of 44 analyses).

INSERT FIGURE 2 ABOUT HERE

Homogeneity of the evidence in the highest quality studies

Eleven studies met the criteria for inclusion in the analysis of the highest quality studies (Figure 2) (Severson et al., 1989; Lee & Paffenbarger, 1994; Giovannucci et al., 1995; Chao et al., 2004; Calton et al., 2006; Friedenreich et al., 2006; Larsson et al., 2006; Lee et al., 2007; Mai et al., 2007; Wolin et al., 2007; Nilsen et al., 2008). The characteristics of these studies did not differ greatly from those of the other nine studies included in the general analyses, other than the method of PA measurement.

> **Comment [RS1]:**
> References have been added.

When this restricted group of studies was considered, 31% (12 of 39 analyses) of all associations were significant, which is similar to the 28% for all studies included in this review (Figure 2). However, compared with all studies (50%, 10 of 20 studies), a greater proportion of the highest quality studies (64%, 7 of 11 studies) had at least one statistically significant result. When rectal cancer was removed from the analysis, the proportion of all reported associations that were statistically significant increased from 31% to 39% (12 of 31 analyses).

Fifty percent of associations between PA and CRC were statistically significant among studies in the restricted sample (compared with 60% of associations for all studies). However, as only two studies reported CRC as an endpoint, this is not a meaningful proportion, and both studies reported associations for men only. The proportion of statistically significant associations for rectal cancer remained at 0% (of 8 analyses). For colon cancer and proximal and distal colon cancers, higher proportions of the examined associations reached statistical significance when the analysis was limited to the highest quality studies than when

all studies were included. For colon cancer, 45% of associations in the restricted sample (5 of 11 analyses) were statistically significant, compared with 35% in the total sample. Three of the studies in the restricted sample found statistical significance for men only (Severson et al., 1989; Giovannucci et al., 1995; Lee et al., 2007). For both proximal and distal colon cancers, 33% of findings (3 of 9 analyses) were statistically significant, compared with 27% for the total sample. One of the three statistically significant findings for proximal colon cancer was significant for men only (Chao et al., 2004), and for distal colon cancer one was significant for women only (Wolin et al., 2007) and another for men only (Giovannucci et al., 1995).

Evidence of dose-response relationship in the highest quality studies

With only two high-quality studies reporting the endpoint of CRC, no further analysis was conducted on this sub-sample. Moreover, because cases of proximal or distal colon cancer were always included in a study's reported number of colon cancer cases, these endpoints were not used in the final analysis to prevent duplication of findings.

Figure 3 shows the relative risk of colon cancer and 95% confidence intervals for all PA categories as reported for the 11 studies in the restricted sample. Both studies that included both men and women showed a statistically significant 'p for trend' (p<0.05) whereas only 43% (n=3) of the seven studies of men and 17% (n=1) of the six studies of women found a statistically significant 'p for trend' (p<0.05). The studies that found a significant 'p for trend' did not differ in characteristics from those that did not.

Although the proportion of studies with a significant 'p for trend' is not high, there is a clear trend across studies towards an inverse linear association between PA and colon cancer risk. When looking at the graphical representation of the relative risks across PA categories, it appears that while not statistically significant, the relative risk in most studies decreases as PA increases. Additionally, for most of the studies the lowest risk was for the highest PA category (Figure 3).

INSERT FIGURE 3 ABOUT HERE

Discussion

The purpose of this paper was to evaluate the strength of the evidence for an inverse association between PA and risk of CRC. The guidelines created by the WCRF/AICR panel for judging the evidence for associations between risk factors and cancer were used as a framework for this review (World Cancer Research Fund/American Institute for Cancer Research, 2007). Previous reviews (Friedenreich, 2001; Slattery, 2004; Samad et al., 2005), which summarised findings from both case-control studies and early cohort studies, showed that there was sufficient support for an inverse relationship between PA and CRC risk to meet

the first two WCRF/AICR criteria (evidence from more than one study type [#1] and from at least 2 cohort studies [#2]). However, due to measurement and study design limitations of the studies available for inclusion in the earlier reviews, those reviews were not able to address the subsequent WCRF/AICR criteria.

Since those earlier reviews, a number of prospective cohort studies have been conducted. These studies have overcome enough of the limitations of the earlier studies to allow for a critical evaluation of 'good quality' studies, as required by the third WCRF/AICR criterion (#3). The current review, therefore, included only prospective cohort studies. A systematic search of the literature yielded 20 studies for inclusion in this review, eight of which had been published since the publication of the previous reviews, or had been updated to include results from further follow-up of cohorts previously studied (Calton et al., 2006; Friedenreich et al., 2006; Johnsen et al., 2006; Larsson et al., 2006; Lee et al., 2007; Mai et al., 2007; Wolin et al., 2007; Nilsen et al., 2008). With the inclusion of only 'good quality' studies, this review expanded on past reviews to address WCRF/AICR criteria #4 and #5. Criterion #4 requires homogeneity among study findings, and criterion #5 calls for consistent support among high quality studies for a dose-response relationship between PA and CRC.

With respect to WCRF/AICR criterion #4, our results suggest much heterogeneity among study findings. The proportion of studies that reported a statistically significant inverse association between PA and at least one CRC endpoint (i.e., CRC, colon cancer, rectal cancer, proximal or distal colon cancer) was 50%. This percentage increased to 64% when only the highest quality studies with the best measures of PA were analysed. However, when the proportion of reported associations for any specific endpoint was considered (for the total sample or among only the highest quality studies), less than 35% of all the associations reported were statistically significant. As previous evidence supports different biological mechanisms for PA and rectal cancer than for PA and colon cancer, and because there were no significant outcomes for the relationship between PA and rectal cancer in the reviewed literature, the analysis was re-run without the rectal cancer endpoint. Removing rectal cancer from the analysis increased the proportion of all reported associations to 34% of all studies and 39% of the highest quality studies. Although these findings are encouraging, particularly for colon cancer, which has the most homogeneity of evidence, they do not suggest that the evidence is consistent, or 'convincing' as defined by WCRF/AICR (2007), for an association between PA and CRC. Additionally, only 20% (4/20) of all the studies and 36% (4/11) of the highest quality studies reported statistically significant associations for more than one end point. Overall these findings suggest that there is considerable heterogeneity among study findings for each endpoint.

This heterogeneity may be explained in part by the variability in PA measures and the type of PA measured, even among the highest quality studies. For example, participants in the

Wolin et al. (2007) study were asked to report the average time per week spent in each of eight leisure-time activities: walking or hiking outdoors; jogging; running; bicycling; lap swimming; playing tennis; playing squash or racquetball; and participating in calisthenics, aerobics, aerobic dance or use of a rowing machine. Individuals also reported their usual walking pace and number of flights of stairs climbed daily. These data were used to derive a weekly physical activity score (MET hours per week) that was divided into quintiles for analysis. This measure was much more comprehensive than the one used in the study by Gerhardsson et al. (1988) , in which participants were categorized into a 'high' or 'low' physical activity category. Highly active participants were those who reported having a physically demanding job or reported engaging in regular and/or hard exercise during recreational hours. Thus, Gerhardsson et al. used a cruder PA measure than did Wolin et al., but one that include occupational PA as well as leisure-time PA.

The heterogeneity could also be due in part to the variation among studies in the selection of covariates for inclusion in each analysis, the duration of follow-up, and the life stage at which PA was assessed. For example, Wu et al. (1987) adjusted only for age, whereas Giovannucci (1995) adjusted for 12 covariates: age, BMI, parental history of CRC, history of endoscopic screening or polyp diagnosis, smoking, aspirin use, intake of folate, methione, alcohol, dietary fibre, red meat and total energy. Moreover, the follow-up duration varied greatly: from 5 years (Wu et al., 1987) to 28 years (Ballard-Barbash et al., 1990). The life stages for which PA was assessed included: 'current' PA at baseline (Calton et al., 2006); average PA over the past 3 years (Mai et al., 2007); PA when aged 25-50 years (Gerhardsson et al., 1988); and 'current' PA reported at baseline and 3 years later, which were then averaged (Lee et al., 1997). The variation in follow-up duration and life stage for which PA is recorded is particularly important, because it is not known whether PA at a particular age is more protective against CRC, or whether PA at a particular life stage may reduce the risk of cancer, 10-15 years from that exposure (the latency period for the development of CRC).

If more of the studies had used comparable methodologies, a systematic review could have examined subgroups of studies defined by methodologies, such as by the PA measure used or by the covariates included in analyses. At this time, however, too few studies have used similar methodologies to allow for analysis of sub-groups of homogeneous studies. Therefore, important factors associated with CRC, such as BMI and age of diagnosis, cannot be examined separately. The one factor that can be preliminarily examined is sex, because some studies have examined the association between PA and CRC separately by sex. In this review, more positive associations between PA and CRC endpoints were found for men than for women although the number of studies available for analysis was too few to draw meaningful conclusions about sex differences.

In summary, there is a lack of homogeneity among prospective cohort studies in the reporting of significant associations between PA and CRC, and the heterogeneity remains even after restricting the analysis to studies with high quality PA measures. It should be noted, however, that 75% (40/53) of analyses evaluated showed either significant or non-significant reductions in risk of CRC with increasing PA (Figure 2). Thus, the totality of the evidence, which includes both significant and non-significant inverse associations between PA and CRC risk, indicates a trend towards an inverse association between PA and CRC risk. It remains unclear as to whether this association is more likely to be due to a decreased risk of proximal or distal colon cancer or both. Thus, such relationships warrant further investigation, especially given that at least some of the heterogeneity could be explained by variations in PA measurement and study design. Additionally, as concluded in previous reviews and suggested by discussions of potential biological mechanisms, this review found convincing evidence of a lack of association between PA and rectal cancer (Friedenreich, 2001; Giovannucci, 2002; Quadrilatero & Hoffman-Goetz, 2003; Slattery, 2004; Samad et al., 2005).

With respect to WCRF/AICR criterion #5, support for the presence of a dose-response relationship between PA and specifically the colon cancer endpoint was encouraging but inconclusive (due to the evidence of a lack of association between PA and rectal cancer from both this study and previous studies, this final criterion was addressed using the evidence for colon cancer only). The two studies that included both men and women reported a statistically significant linear trend in reduced risk of cancer with increased PA. In contrast, of the seven studies that included only men and the six that included only women, four showed a statistically significant linear trend. When all 11 studies included in the restricted sample were evaluated together, a linear dose-response pattern was evident, but the wide confidence intervals precluded the finding of statistically significant relationships. It is possible that the number and location of PA category boundaries obscured the true relationship between PA and CRC and that non-linear effects of PA were missed due to the use of insufficient categories (World Cancer Research Fund/American Institute for Cancer Research, 2007). Additionally, until the optimal method of measuring PA and the life stages at which PA is most important for protecting against cancer risk are known, it is likely to remain difficult to collect more convincing evidence of a dose-response relationship between PA and colon cancer.

Limitations

One limitation was the criteria used to define highest quality studies (i.e., studies that measured PA intensity and duration and categorized PA into sufficient levels to examine dose-response associations). It is possible that defining 'highest quality' more strictly with additional criteria would have resulted in different results for this sub-sample. However, to

date there are too few published studies to have stricter inclusion criteria. With few studies, it was not possible to examine associations within subgroups, such as those defined by the type of PA reported, or the timing of the PA across the lifespan, or by other CRC risk factors, such as age at diagnosis or BMI. It may be that a stronger relationship between PA and CRC risk would be evident in certain subgroups. The decisions to exclude studies that measured only occupational PA, as done in a previous review of PA and breast cancer (Monninkhof et al., 2007), and to include only the PA measure hypothesized to have the least recall bias, may have caused us to miss the type of PA or timing of PA that had the greatest impact on CRC risk. Even so, no evidence is available to support a specific time period as the optimal one for engaging in PA to reduce CRC risk.

A further limitation was the inclusion of only the lowest and highest levels of PA in examining homogeneity across studies and analyses. This approach did not always provide the most pronounced risk estimates. For example, in the study by Severson et al. (1989) the greatest risk reduction was for middle PA category. As this study suggests, the association between PA and CRC risk may not be linear. However, the WCRF/AICR guidelines (World Cancer Research Fund/American Institute for Cancer Research, 2007) and prior reviews (Monninkhof et al., 2007) have recommended this technique as a valid method for establishing the presence or absence of a relationship. Also, as the analysis of the high quality studies shows (Figure 3), for colon cancer most studies found the highest PA level to be associated with the lowest risk reduction. Finally, the review is not free from publication bias because only published studies were included. The bias would likely be the exclusion of studies that found no association between PA and the CRC endpoints, which would further weaken the homogeneity of the evidence for an association between PA and CRC risk. Notwithstanding, as discussed by other reviewers who included only published studies in their review (Monninkhof et al., 2007), the inclusion of only published studies probably resulted in the inclusion of the most rigorous studies, which were likely to have their findings reported no matter the outcome.

Perspectives

Previous reviews have concluded that there is an inverse association between PA and CRC risk. Given that few cohort studies had reported on this association prior to these reviews, the reviews relied largely on findings from case-control studies. Since the reviews were published, however, a number of large population-based cohort studies, often with detailed measures of PA, have been published. As a result, there is now sufficient evidence to evaluate the consistency of findings from cohort studies. The current paper reviewed cohort studies to evaluate the strength of the evidence supporting an association between PA and CRC risk. The main finding was that most analyses of the associations between PA and CRC endpoints

15

did not result in statistically significant findings. In fact, the percentages of statistically significant findings that supported an inverse association between PA and any CRC endpoint or a dose-response relationship between PA and colon cancer was low. These findings indicate that the evidence for an association between PA and CRC risk is not homogeneous. Based on WCRF/AICR guidelines (World Cancer Research Fund/American Institute for Cancer Research, 2007), the strength of the evidence for this association is insufficient to be given a grading of 'conclusive'. However, as a high percentage of studies reported some risk reductions for the highest levels of PA and the presence of non-significant linear trends, the totality of the evidence for an association is promising, in particular for the relationship between PA and colon cancer. It is likely that the heterogeneity is attributable to study characteristics and methodological limitations, primarily an inability to accurately measure PA; a lack of knowledge as to the relevant time periods to measure PA; and an inability to adjust for other factors within the analyses, as few studies measured the same factors. Future research should thus use comprehensive measures of PA; examine the association between PA across the lifespan and risk of colon cancer; and investigate the influence of risk factors for colon cancer, such as age at diagnosis, sex, and BMI, on the association between PA and colon cancer. Addressing the methodological limitations of previous studies will prepare for the next generation of large-scale prospective cohort studies that examine the association between PA and colon cancer risk.

Key Words:

Colorectal Neoplasm, Primary Prevention, Exercise, Health Promotion.

References

American College of Sports Medicine. American College of Sports Medicine position stand. Exercise and physical activity for older adults. *Med Sci Sports Exerc* 1998: 30: (6): 992-1008.

Ballard-Barbash R, Schatzkin A, Albanes D, Schiffman MH, Kreger BE, Kannel WB, Anderson KM, Helsel WE. Physical activity and risk of large bowel cancer in Framingham Study. *Cancer Research* 1990: 50: (12): 3610-3613.

Blair SN, Brodney S. Effects of physical inactivity and obesity on morbidity and mortality: current evidence and research issues. *Med Sci Sports Exerc* 1999: 31: (11 Suppl): S646-S662.

Calton BA, Lacey JV, Jr, Schatzkin A, Schairer C, Colbert LH, Albanes D, Leitzmann MF. Physical activity and the risk of colon cancer among women: a prospective cohort study (United States). *Int J Cancer* 2006: 119: (2): 385-391.

Chao A, Connell CJ, Jacobs EJ, McCullough ML, Patel AV, Calle EE, Cokkinides VE, Thun MJ. Amount, type, and timing of recreational physical activity in relation to colon and rectal cancer in older adults: the Cancer Prevention Study II Nutrition Cohort. *Cancer Epidemiol Biomarkers Prev* 2004: 13: (12): 2187-2195.

Cochrane Non-randomised Studies Methods Group. Guidelines for using non-randomised studies in Cochrane reviews (Draft) (on-line). Available: http://www.cochrane.dk/nrsmg/guidelines.htm. Accessed May 13, 2008.

Davey Smith G, Shipley MJ, Batty GD, Morris JN, Marmot M. Physical activity and cause-specific mortality in the Whitehall Study. *Public Health* 2000: 114: (5): 308-315.

Friedenreich C, Norat T, Steindorf K, Boutron-Ruault MC, Pischon T, Mazuir M, Clavel-Chapelon F, Linseisen J, Boeing H, Bergman M, Johnsen NF, Tjonneland A, Overvad K, Mendez M, Quiros JR, Martinez C, Dorronsoro M, Navarro C, Gurrea AB, Bingham S, Khaw KT, Allen N, Key T, Trichopoulou A, Trichopoulos D, Orfanou N, Krogh V, Palli D, Tumino R, Panico S, Vineis P, Bueno-de-Mesquita HB, Peeters PH, Monninkhof E, Berglund G, Manjer J, Ferrari P, Slimani N, Kaaks R, Riboli E. Physical activity and risk of colon and rectal cancers: the European prospective investigation into cancer and nutrition. *Cancer Epidemiol Biomarkers Prev* 2006: 15: (12): 2398-2407.

Friedenreich CM. Physical activity and cancer prevention: from observational to intervention research. *Cancer Epidemiol Biomarkers Prev* 2001: 10: (4): 287-301.

Gerhardsson M, Floderus B, Norell SE. Physical activity and colon cancer risk. *Int J Epidemiol* 1988: 17: (4): 743-746.

Giovannucci E. Modifiable risk factors for colon cancer. *Gastroenterol Clin North Am* 2002: 31: (4): 925-943.

Giovannucci E, Ascherio A, Rimm EB, Colditz GA, Stampfer MJ, Willett WC. Physical activity, obesity, and risk for colon cancer and adenoma in men. *Ann Intern Med* 1995: 122: (5): 327-334.

Johnsen NF, Christensen J, Thomsen BL, Olsen A, Loft S, Overvad K, Tjonneland A. Physical activity and risk of colon cancer in a cohort of Danish middle-aged men and women. *Eur J Epidemiol* 2006: 21: (12): 877-884.

Larsson SC, Rutegard J, Bergkvist L, Wolk A. Physical activity, obesity, and risk of colon and rectal cancer in a cohort of Swedish men. *Eur J Cancer* 2006: 42: (15): 2590-2597.

Lee IM, Manson JE, Ajani U, Paffenbarger RS, Jr, Hennekens CH, Buring JE. Physical activity and risk of colon cancer: the Physicians' Health Study (United States). *Cancer Causes Control* 1997: 8: (4): 568-574.

Lee IM, Paffenbarger RS, Jr. Physical activity and its relation to cancer risk: a prospective study of college alumni. *Med Sci Sports Exerc* 1994: 26: (7): 831-837.

Lee KJ, Inoue M, Otani T, Iwasaki M, Sasazuki S, Tsugane S. Physical activity and risk of colorectal cancer in Japanese men and women: the Japan Public Health Center-Based Prospective Study. *Cancer Causes Control* 2007: 18: (2): 199-209.

Mai PL, Sullivan-Halley J, Ursin G, Stram DO, Deapen D, Villaluna D, Horn-Ross PL, Clarke CA, Reynolds P, Ross RK, West DW, Anton-Culver H, Ziogas A, Bernstein L. Physical activity and colon cancer risk among women in the California Teachers Study. *Cancer Epidemiol Biomarkers Prev* 2007: 16: (3): 517-525.

Monninkhof EM, Elias SG, Vlems FA, van der Tweel I, Schuit AJ, Voskuil DW, van Leeuwen FE. Physical activity and breast cancer: a systematic review. *Epidemiology* 2007: 18: (1): 137-157.

Nilsen TI, Romundstad PR, Petersen H, Gunnell D, Vatten LJ. Recreational Physical Activity and Cancer Risk in Subsites of the Colon (the Nord-Trondelag Health Study). *Cancer Epidemiol Biomarkers Prev* 2008: 17: (1): 183-188.

Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. *CA Cancer J Clin* 2005: 55: (2): 74-108.

Quadrilatero J, Hoffman-Goetz L. Physical activity and colon cancer. A systematic review of potential mechanisms. *J Sports Med Phys Fitness* 2003: 43: (2): 121-138.

Samad AK, Taylor RS, Marshall T, Chapman MA. A meta-analysis of the association of physical activity with reduced risk of colorectal cancer. *Colorectal Dis* 2005: 7: (3): 204-213.

Schnohr P, Gronbaek M, Petersen L, Hein HO, Sorensen TI. Physical activity in leisure-time and risk of cancer: 14-year follow-up of 28,000 Danish men and women. *Scand J Public Health* 2005: 33: (4): 244-249.

Severson RK, Nomura AM, Grove JS, Stemmermann GN. A prospective analysis of physical activity and cancer. *Am J Epidemiol* 1989: 130: (3): 522-529.

Slattery ML. Physical activity and colorectal cancer. *Sports Med* 2004: 34: (4): 239-252.

Thun MJ, Calle EE, Namboodiri MM, Flanders WD, Coates RJ, Byers T, Boffetta P, Garfinkel L, Heath CW, Jr. Risk factors for fatal colon cancer in a large prospective study. *J Natl Cancer Inst* 1992: 84: (19): 1491-1500.

Thune I, Lund E. Physical activity and risk of colorectal cancer in men and women. *Br J Cancer* 1996: 73: (9): 1134-1140.

Warburton DE, Nicol CW, Gatto SN, Bredin SS. Cardiovascular disease and osteoporosis: balancing risk management. *Vasc Health Risk Manag* 2007: 3: (5): 673-689.

Wolin KY, Lee IM, Colditz GA, Glynn RJ, Fuchs C, Giovannucci E. Leisure-time physical activity patterns and risk of colon cancer in women. *Int J Cancer* 2007: 121: (12): 2776-2781.

World Cancer Research Fund/American Institute for Cancer Research. Food, nutrition and the prevention of cancer: a global perspective. AICR, Washington, DC, 1997.

World Cancer Research Fund/American Institute for Cancer Research. Food, nutrition, and physical activity, and the prevention of cancer: a global perspective. AICR, Washington, DC, AICR, 2007.

Wu AH, Paganini-Hill A, Ross RK, Henderson BE. Alcohol, physical activity and other risk factors for colorectal cancer: a prospective study. *Br J Cancer* 1987: 55: (6): 687-694.

Table 1. Search strategy used in PubMed and Embase databases to identify studies for
inclusion

PubMed	Embase
Keywords relating to physical activity [exposure] (physical activity, physical fitness, inactivity, exercise, exertion, energy expenditure, walking, exercise therapy, physical education, physical training)	Keywords relating to physical activity [exposure] (physical activity, exercise training, physical education, physical inactivity, exertion, physical fitness, exercise therapy)
Key words relating to colorectal cancer [outcome] (colorectal cancer, colon cancer, rectal cancer, colonic neoplasm) as Medical Subject Headings (MeSH), and keywords in the title/abstract or text and as exploded variants.	Keywords relating to colorectal cancer [outcome] (colorectal cancer, bowel cancer, colon cancer, rectal cancer, colorectal neoplasm, colon neoplasm, rectal neoplasm). All terms were mapped to preferred terminology and searched also for synonyms and explosions:

Table 2. Summary of included studies

Citation	Sample[a]	PA measure	Case ascertainment	Covariates	Follow-up[b]
Wu., 1987	Residents of a Los Angeles retirement community (USA) N=11,644 Age not given 126 CRC cases (58 men, 68 women)	Hours per day in recreational PA (provided examples of swimming, biking, dancing). Categories: (1) <1 hr/day (2) 1-2 hrs/day (3) > 2 hrs/day	Self-report and hospital records from 5 hospitals (incidence); health department, next of kin, and obituaries (mortality).	Age	5 yrs
Gerhardsson et al., 1988	Twin Registry (Sweden) N=16,477 Age 42-82 yrs Number of cases not provided	Occupational PA, rated as sedentary, moderately active or physically demanding; recreational PA when aged 25-50 yrs, rated as 'hardly any', 'light exercise', 'regular exercise' or 'hard exercise'. Categories: (1) sedentary occupation AND hardly any or light recreational PA (2) physically demanding occupation AND regular or 'hard' intensity recreational PA	Linkage with Swedish cancer registry and cause-of death registry	Age, gender, geographic region, degree of urbanization, meat consumption, coffee consumption	14 yrs
[c]Severson et al., 1989	Japan-Hawaii Cancer Study: Japanese men living on Hawaiian island Oahu (USA) N=7,925 men Age 46-68 yrs 192 CC cases 95 RC cases	Weighted sum of the usual time spent per 24 hrs sleeping, sedentary, and in light, moderate or heavy physical activity. Categories (Tertiles, PA index[d]): (1) 0-30.1 (2) 30.2-34.2 (3) ≥34.3	Linkage with Hawaii tumour registry	Age, BMI	8-11 yrs

Ballard-Barbash et al., 1990	The Framington Study: Population-based cohort (USA) N= 4214 Age 30-62 yrs 152 CRC cases (73 men, 79 women)	Weighted sum of usual time per 24 hrs in basal, light, moderate and heavy activities. Categories (Tertiles, total PA index) Men: (1) 25-29 (2) 30-33 (3) 34-83 * highest category is referent Women: (1) 25-29 (2) 30-31 (3) 32-55 * highest category is referent	Review of records from biennial physician-administered examinations	Age, BMI, height, education, cholesterol, alcohol intake, smoking, and, for women, parity and menopause status	up to 28 yrs
Thun et al., 1992	Cancer Prevention Study II (CPS II) Mortality Cohort: National population-based cohort (USA)[e] N cohort= 763,343 n cases= 1,150 n control= 5,746 Mean age= 65 yrs 1150 CC deaths (611 men, 539 women)	'Physical activity in work or play', rated as none, slight, moderate or heavy. Categories: (1) 'None' (2) 'Slight' (3) 'Moderate' (4) 'Heavy'	Self-reported or next of kin report and death certificates	Age, BMI, family history of CC, vegetable and total fat intake, aspirin use	6 yrs
[c]Lee et al., 1994	Harvard Alumni Health Study: male graduates of Harvard University (USA) N= 17,607 men	Kilocalories expended per week in stair climbing, walking city blocks, and sports or recreational activities Categories: (1) <1000 kcal/wk at baseline and 10 yrs	Self-reported incidence, death certificate verification of mortality.	Age, BMI, parental history of cancer.	24 yrs

	Mean age across PA categories = 47-48 yrs (SD= 10-11 yrs) 280 CC cases 53 RC cases.	post-baseline (2) 1000-2499 kcal/wk at baseline and 10 yrs post-baseline (3) >=2500(kcal/wk at baseline and 10 yrs post-baseline			
[c]Giovannucci et al., 1995	Health Professionals' follow-up study (HPFS): male dentists, optometrists, osteopaths, podiatrists, pharmacists and veterinarians (USA) N= 47,723 men Age 40-75 yrs 249 CRC cases 203 CC cases 46 RC cases	MET.hrs/wk in recreational PA, from weighted sum of reported number of hours/wk spent in each of eight moderate and vigorous activities (walking/hiking, jogging, running, bicycling, lap swimming, tennis/squash/racquetball and callisthenics or rowing). Categories for total recreational PA (Quintiles, median MET.hrs/wk): (1) 0.9 (2) 4.8 (3) 11.3 (4) 22.6 (5) 46.8	Self-reported or next-of-kin report, verified by hospital records and pathology reports	Age, BMI, parental history of CRC, history of endoscopic screening or polyp diagnosis, smoking, aspirin use, intake of folate, methione, alcohol, dietary fibre, red meat, total energy.	6 yrs
Thune et al., 1996	Population-based cohort (Norway) N= 81,516 Age 20-69 yrs 236 CC cases 170 RC cases	Recreational PA, rated as sedentary; PA ≥4 hrs/wk; exercise to keep fit ≥4 hrs/wk; or regular hard training. Occupational PA, rated as mostly sedentary; work with much walking; work with much lifting and walking; or heavy manual work. Total PA scores computed as 'sedentary' if in low categories in occupational and recreational PA; 'moderate' if low category in either occupation or recreation; 'active' if moderate to high in both occupation and recreation.	Linkage with cancer registry in Norway	Age, BMI, geographic region	mean= 16.3 yrs (men), 15.5 yrs (women)

		Categories for occupational and recreational PA combined (1) Sedentary (2) Moderate (3) Active			
Lee et al., 1997	Physicians' Health Study (PHS): male physicians (USA) N=21,807 men Age 40-84 yrs 217 CC cases	Frequency of exercising vigorously enough to work up a sweat, responses of 1-3/ mth, 1/wk, 2-4/wk, 5-6/ wk, daily. Categories (times/wk at both baseline and 3 yrs post-baseline): (1) <1 (2) 1 (3) 2-4 (4) 5+	Self-reported and verified by medical records (incidence). Family members or postal authorities notification (mortality).	Obesity, alcohol consumption, group allocation in RCT.	mean= 10.9 yrs
Davey Smith et al., 2000	Whitehall study: male civil servants (UK) N= 6,702 men. Age 40-64 yrs 89 CRC <u>deaths</u>	Participation in 'hobbies or sport', rated as 'inactive' (no participation), 'moderately active' (active hobbies), 'active' (vigorous sports). Categories: (1) Inactive (2) Moderately active (3) Active	National Health Service Central Registry and Death certificates	Age, BMI, civil service employment grade, smoking, forced air expiratory volume in 1 second.	25 yrs
[c]Chao et al., 2004	Cancer Prevention Study II (CPS II) Nutrition Cohort: a subset of the CPS II mortality study (USA) N= 151,174 Median Age = 63 yrs	MET.hrs/wk in recreational PA, from weighted sum of reported number of hours/wk spent in seven recreational activities (walking/hiking, jogging/running, bicycling, lap swimming, tennis/racquetball, aerobics/callisthenics or dancing). Categories for recreational PA (MET.hrs/wk):	Self-reported, verified by medical record abstraction, linkage with state cancer registry and national death index.	Age, sex, education, past exercise levels, smoking, alcohol use, red meat, folate, fibre, multivitamin use, hormone therapy use.	6-7 yrs

24

		(1) No activity (2) <7 (3) 7-13 (4) 14-23 (5) 24-29 (6) ≥30			
940 CC cases (536 men and 404 women) 390 RC cases (247 men and 143 women)					
Schonhr et al., 2005	Copenhagen Centre for Prospective Population Studies: 2 population-based studies and 1 study of male workers (Denmark) N= 28,259 Age 20-93 395 CC cases (180 men, 215 women)	Recreational PA rated as 'low' (almost entirely sedentary), 'moderate' (light physical activity 2-4 hrs/wk) or 'vigorous' (light physical activity >4hrs/wk or highly vigorous physical activity >4 hrs/wk). Categories for recreational PA: (1) Low Level (2) Moderate Level (3) Vigorous Level	Linkage with Danish cancer registry	Age, BMI, birth cohort, cohort membership, education, occupational PA, smoking, alcohol use consumption	14 yrs
[c]Calton et al., 2006	Breast Cancer Detection Demonstration Project (BCDDP) : women attending breast cancer screening centres (USA) N= 31,783 women Mean age across PA categories = 61-62 yrs 243 CC cases	MET.hrs/day in all activities, from weighted sum of reported usual hours per day spent sleeping and engaging in light, moderate and vigorous activities (examples provided) on both weekdays and weekends. Categories for total PA (Quintiles of MET.hrs/day): (1) 34.00-48.50 (2) 48.51-54.30 (3) 54.31-59.00 (4) 59.10-64.90 (5) 65.00-98.10	Self-report, pathology report and/or state cancer registries (incidence) and National Death Index (mortality).	Age, BMI, education, family history of CRC, smoking status, menopausal hormone use, aspirin use, alcohol consumption and energy-adjusted intakes of total calcium and red meat.	6-11 yrs
[c]Friedenreich	European Prospective	MET.hrs/wk in recreational PA, from	Population-based	Height, weight,	6-7 yrs

et al., 2006	Investigation into Cancer and Nutrition (EPIC): international population-based and defined population sample. N= 413,044 Age 35-70 yrs 1693 CRC cases 1094 CC cases 599 RC cases.	weighted sum of reported weekly frequency and duration of household and recreational PA. Occupational PA, rated as nonworker, sedentary, standing, manual, heavy labour or unknown. Recreational and occupational PA were 'cross-classified' to derive the total score. Categories for total PA index (Quartiles): (1) Inactive (2) Moderately Inactive (3) Moderately Active (4) Active	cancer registries (except France, Germany and Greece: health insurance records, cancer and pathology registries, and active follow-up through study subjects and their next of kin)	education, other types of PA, smoking status, current alcohol intake, energy intake, fibre intake. For rectal cancer only, also fish intake	
Johnsen et al., 2006	Diet, Cancer and Health Study: Population-based cohort (Denmark) N= 57,053 Age 50-64 yrs 297 CC cases (157 males, 140 women)	MET.min/wk in recreational PA, from weighted sum of reported time spent in six activities (sports, cycling, walking, gardening, housework, do-it-yourself work). PA measure: Incidence rate ratios for colon cancer 'per 10 units' of MET.min-score.	Link to Danish cancer registry	Other activity, BMI, education, NSAIDs, present hormone replacement therapy use, smoking, intake of total energy, fat, dietary fibre, red meat and alcohol.	mean = 7.6 yrs
[c]Larsson et al., 2006	Cohort of Swedish Men (COSM): Population-based cohort in Vastmanland and Orebro counties (Sweden) N= 45,906 Age 45-79 yrs	MET.hrs/day (24-hr) in total PA, from weighted sum of reported time spent in PA at work (6 categories, from 'mostly sitting down' to 'heavy manual labour'), exercising and in inactive activities/ sleeping. Categories for total activity (Quintiles of MET.hrs/day): (1) <37.9 (2) 37.9-40.7	Linkage to regional Swedish cancer registries	BMI , education, family history of CRC, history of diabetes, smoking, aspirin use.	mean= 7.1 yrs

	496 CRC cases 309 CC cases	(3) 40.8-44.8 (4) ≥44.9			
[c]Lee et al., 2007	Japan Public Health Centre-based Prospective Study: Japanese residents with registered addresses at health centres (Japan) N= 65,022 Age 40-69 yrs 486 CRC cases	MET.hrs/day in recreational PA, from weighted sum of average reported time per day spent in heavy physical work/strenuous exercise, sedentary activity, and walking or standing. Categories (Quartiles, median MET.hrs/day) Men: (1) 28.25 (2) 33.25 (3) 35.25 (4) 43.75 Women: (1) 28.50 (2) 33.25 (3) 35.25 (4) 43.75	Active patient notification from major hospitals and linkage to population-based cancer registries (incidence); death certificates (mortality).	Age, BMI, study area, family history of CRC, smoking status, alcohol intake, intake of red meat, dietary fibre and folate.	6 yrs
[c]Mai et al., 2007	California Teachers Study: current, recent and retired female public school teachers and administrators in California (USA) N= 210,147 women Age 22-84 yrs 395 CC cases	Hrs/wk over the past 3 yrs in moderate or strenuous recreational PA, from average reported number of hrs/wk and mth/yr in moderate activities (examples included brisk walking, recreational tennis, cycling on level street) and strenuous activities (examples included swimming laps, aerobics, running, cycling on hills or racquetball). Categories (hrs/wk) (1) 0-0.51 (2) 0.51-3.99 (3) ≥4.00	Annual linkage with Californian Cancer registry	BMI, menopause status/ hormone therapy use, NSAID, smoking, total caloric intake, total folate intake, total calcium intake, total dietary fibre intake.	mean = 6.6 yrs

[c]Wolin et al., 2007	Nurses Health Study: female registered nurses across USA (USA) N= 79,295 women Age 30-55yrs 547 CC cases	MET.hrs/wk, from weighted sum of reported time spent in eight leisure-time activities (walking/hiking, jogging, bicycling, swimming, tennis, squash/racquetball, callisthenics/aerobics/aerobic dance/ using a rowing machine) Categories (Quintiles of MET.hrs/wk score) (1) <2 (2) 2.1-4.5 (3) 4.6-10.3 (4) 10.4-21.4 (5) ≥21.5	Self-report verified with medical records and pathology reports	Age, BMI, family history of CC, endoscopic history, previous CRC polyp, smoking, multivitamin use, aspirin use, red meat intake, processed meat intake, vitamin D intake, calcium intake, alcohol consumption	16 yrs
[c]Nilsen et al., 2008	Nord-Trondelag Health Survey: County-based cohort (Nord-Trondelag, Norway) N= 75,043 Age 20-101 yrs 736 CC cases (346 men, 390 women) 294 RC cases (170 men, 124 women)	Summary score of recreational PA, from reported average frequency (0, <1, 1, 2-3 or ≥4), duration (<15, 15-30, 31-60 and >60 min) and intensity (light, moderate or vigorous) of recreational PA (categorised as walking, skiing, swimming or other sports) in a usual week. Categories: (1) No activity (2) Low (score < median score) (3) High (score ≥ median score)	Linkage with Cancer Registry of Norway (incidence) and Cause of Death Registry at Statistics Norway (mortality)	Age, BMI, Smoking status, use of alcohol, marital status and education.	17 yrs

[a] All ages reported in this table are age at baseline.
[b] Follow-up is reported as years since baseline
[c] Used in restricted analysis of highest quality studies including the more comprehensive PA measures.
[d] PA index: value calculated by study to categorise individuals. Values have no meaning outside of studies.
[e] Nested Case-control Study.

PA, physical activity; hr/s, hour/s; yrs, years; CRC, colorectal cancer; CC, colon cancer; RC, rectal cancer; BMI, body mass index; MET, metabolic equivalent; wk, week; min, minutes; mth, month; RCT, randomised controlled trial.

Table 3. Relative risk of colorectal cancer endpoints for the highest versus lowest level of physical activity examined

Study	Colorectal Cancer	Colon Cancer	Rectal Cancer	Proximal colon	Distal colon
Wu, 1987	**M: 0.40 (0.2-0.8)** [a] W: 0.89 (0.5-1.6)			M: 0.50 (0.2-1.3) W: 1.16 (0.4-2.5)	M: 0.36 (0.1-1.1) W: 0.68 (0.3-1.5)
Gerhardsson et al., 1988		**M&W: 3.6 (1.3-9.8)** [b]			
[c]Severson et al., 1989		**M: 0.71 (0.51-0.99)**	M: 1.41 (0.84-2.36)		
Ballard-Barbash et al., 1990	**M: 1.8 (1.0-3.2)** W: 1.1 (0.6-1.8)				
Thun et al., 1992		M: 0.60 (0.28-1.27) W: 0.90 (0.41-1.96)			
[c]Lee et al., 1994		M: 1.08 (0.81-1.46)	M: 1.71 (0.88-3.31)		
[c]Giovannucci et al., 1995		**M: 0.53 (0.32-0.88)**	M: 1.83 (0.83–3.84)	M: 0.75 (0.36-1.55)	**M: 0.50 (0.25-1.00)**
Thune et al., 1996		M: 0.97 (0.63-1.50) W: 0.63 (0.39-1.04)	M: 1.20 (0.72-2.02) W: 1.27 (0.59-2.72)	M: 0.96 (0.47-1.93) W: 0.62 (0.31-1.28)	M: 0.99 (0.55-1.80) W: 0.61 (0.30-1.23)
Lee et al., 1997		M: 1.3 (0.9-2.0)			
Davey Smith et al., 2000	M: 1.52 (0.8-2.7)				
[c]Chao et al., 2004		**M: 0.60 (0.41-0.87)** W: 0.77 (0.48-1.24) **M&W: 0.65 (0.49-0.87)**	M&W: 0.83 (0.59-1.16)	**M&W: 0.63 (0.45-0.88)**	M&W: 0.82 (0.55-1.24)
Schonhr et al., 2005		M: 0.72 (0.47-1.11)			

	W: 0.90 (0.56-1.46)			
[c]Calton et al., 2006	W: 1.15 (0.76-1.75)		W: 0.87 (0.46-1.62)	(tertiles instead of quintiles) W: 1.36 (0.75-2.46)
[c]Friedenreich et al., 2006	M&W: 0.78 (0.59-1.03)	M&W: 1.02 (0.73-1.44	(right) **M&W: 0.65 (0.43-1.00)**	(left) M&W: 0.96 (0.64-1.45)
Johnsen et al., 2006	[d]M: 0.97 (0.93-1.01) [d]W: 1.00 (0.96-1.04)			
[c]Larsson et al., 2006 M: 0.82 (0.60-1.10)	M: 0.79 (0.53-1.17)	M: 0.86 (0.53-1.37)	M: 0.71 (0.39-1.29)	M: 0.70 (0.38-1.27)
[c]Lee et al., 2007 **M: 0.69 (0.49-0.97)**	**M: 0.58 (0.39-0.87)**	M: 1.06 (0.56-2.00)	**M: 0.29 (0.14-0.60)**	M: 0.89 (0.53-1.51)
W: 1.16 (0.76-1.77)	W: 0.89 (0.54-1.49)	W: 2.23 (0.99-5.01)	W: 0.55 (0.24-1.26)	W: 1.37 (0.66-2.85)
[c]Mai et al., 2007	W: 0.81 (0.63-1.05)		W: 0.81 (0.59-1.11)	W: 0.78 (0.48-1.25)
[c]Wolin et al., 2007	W: 0.77 (0.58-1.01)		W: 0.97 (0.68-1.38)	**W: 0.54 (0.34-0.84)**
[c]Nilsen et al., 2008	**M: 0.69 (0.48-0.98)**	M: 1.12 (0.65-1.96)	M&W: 0.81 (0.59-1.10)	**M&W: 0.56 (0.37-0.83)**
	W: 0.72 (0.53-0.98)	W: 1.01 (0.58-1.75)		

[a] Statistically significant associations at p< 0.05 are in bold.

[b] Highest physical activity category was referent.

[c] The study is included in the restricted sample of highest quality studies.

[d] Incidence rate ratios, MET.minute-score (per 10 units).

M, men; W, women.

Figure Captions

Figure 1: Flow chart of search strategy.

Figure 2: Number of reported associations between PA and CRC.
Legend: All, All endpoints; CC, colon cancer; RC, rectal cancer; CRC, colorectal cancer.

Figure 3: Dose-response for colon cancer risk across categories of physical activity.
Legend: 95% CI, 95% Confidence Interval.

Figure II