This item is the archived peer-reviewed author-version of:

Short and long-term impairments of cardiopulmonary fitness level in previous childhood cancer cases: a systematic review

Reference:
Kabak Vesile Yildiz, Calders Patrick, Duger Tulin, Mohammed Jibril, van Breda Eric.- Short and long-term impairments of cardiopulmonary fitness level in previous childhood cancer cases: a systematic review
Full text (Publisher's DOI): https://doi.org/10.1007/S00520-018-4483-8
To cite this reference: https://hdl.handle.net/10067/1539630151162165141
Short and Long-Term Impairments of Cardiopulmonary Fitness Level in Previous Childhood Cancer Cases: A Systematic Review

Vesile YILDIZ KABAK\textsuperscript{1,2}, Patrick CALDERS\textsuperscript{2}, Tulin DUGER\textsuperscript{1}, Jibril MOHAMMED\textsuperscript{2,3}, Eric VAN BREDA\textsuperscript{2,4}

\textsuperscript{1}Department of Physiotherapy and Rehabilitation, Hacettepe University, Ankara, Turkey
\textsuperscript{2}Department of Rehabilitation Sciences and Physiotherapy, Ghent University, Ghent, Belgium.
\textsuperscript{3}Department of Physiotherapy, Bayero University, Kano, Nigeria.
\textsuperscript{4}Department of Physiotherapy, Research Group MOVANT, University of Antwerp, Antwerp, Belgium.

Corresponding Author:

Vesile Yildiz Kabak, PT, MSc.
Address: Hacettepe University Faculty of Health Sciences, Department of Physiotherapy and Rehabilitation, 06100, Ankara, Turkey.
E-mail: vesile_vldz@hotmail.com
vesile.yildiz@hacettepe.edu.tr
Telephone: +903122051576
Fax: +903123052012
ORCID iD: 0000-0002-1559-1793
Abstract

Purpose: To describe the impairments in physical fitness in individuals who were previously diagnosed and treated for childhood cancer.

Methods: Using the PRISMA-guidelines, a systematic search was performed in PubMed, Web of Science and Embase using a combination of the following predefined keywords: “exercise capacity” OR “aerobic capacity” OR “fitness” OR “cardiorespiratory fitness” OR “cardiopulmonary fitness” OR “physical fitness” OR “exercise testing” OR “exercise tolerance” OR “exercise” OR “oxygen consumption” AND “leukemia” OR “childhood cancer” OR “childhood cancer survivors (CCS)”. Studies that met our inclusion criteria were reviewed on methodological quality, while the Newcastle-Ottawa Scale was used for evidence synthesis.

Results: A total of 2644 articles were identified from the database search. After screening based on the eligibility (abstracts) and inclusion (full texts) criteria, 49 articles remained. Even though the risk-of-bias scores in the studies were generally low, yet the results from those with high-quality studies revealed that poor fitness levels were prevalent in individuals with acute lymphoblastic leukemia, brain tumor and mixed cancer histories, compared to healthy controls.

Conclusions: A global glance at CCS shows poor levels of fitness that is continuous and life-long even after active cancer treatment has ended. Nevertheless, the results presented in this review were based on a limited number of high-quality studies suggesting the need to for additional clinical trials in the topic area.

Keywords: childhood cancer survivor, survivor, physical fitness, exercise capacity
Introduction

The five-year survival rate has increased from 50% to over 80% during the last decade in childhood cancer [1-3]. This increased survival rates have led to an increased need to rehabilitation of treatment related long-term side effects and comorbidities in this population. Moreover, it is well known that childhood cancer survivors (CCS) represent a clinically heterogeneous group of patients that are dealing with multiple side effects of treatment protocols, lifestyle factors and disease pathophysiology.

Specifically, about three-quarters of CCS develop a chronic illness; cardiovascular disease, pulmonary dysfunction, severe musculoskeletal problems, endocrine abnormalities, and secondary cancers are the most prevalent post-treatment [4,5]. Again, it has been shown that these chronic side-effects continued to increase over time in survivors [5]. As a result, fitness level can be affected negatively in CCS [6]. Poor fitness level leads to further increases in sedentary behavior which, in turn, is partly responsible for chronic fatigue, one of the most reported symptoms in cancer patients. Patients experiencing fatigue tend to be more inactive, and a self-sustaining vicious cycle develops [7]. Consequently, participation to daily activities and overall the quality of life level are affected negatively [8].

Just like in a healthy population, physical activity is an important as lifestyle ingredient in CCS, especially for those having high risk for developing treatment-related chronic health derangements. In a recent Cochrane review, it was reported that within five years after diagnosis, participation in physical activity alone caused a positive impact on cardiopulmonary fitness, body composition, muscle strength, and quality of life [9]. However, this is not surprising because physical inactivity is associated with higher risk for osteoporosis, obesity, and metabolic syndrome [10, 11].

To date, literature regarding fitness levels in CCSs is scarce and inconclusive. The prior attempt at providing evidence by means of two earlier systematic reviews only focused on the maximum oxygen consumption (VO₂ max), as an outcome measure, to assess fitness levels [12, 13]. Moreover, one of these only focused on acute lymphoblastic leukemia (ALL) patients [12].

With the increasing number of CCS patients and also a possible concurrent increase in the rate of cancer related morbidities in these patients, the evidence regarding fitness level in CCS is inadequately reflected in literature. Therefore, the present systematic review was conducted to assess both fitness levels and the different methods used in fitness assessment in all childhood cancers after active cancer treatment.
Methods

Design

This systematic review was registered within the PROSPERO system (registration number: CRD42018082252). And the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) statement checklist was used as a guide in review process [14].

Search Strategy

The PICOS (population, intervention, comparator, outcomes and study design) components were used to establish the research question, and the corresponding eligibility criteria are shown in Table 1. Thereafter, a structured search was conducted in three electronic databases: PubMed, Web of Science, and Embase. The last date of searching databases was on 1st March 2018. The search terms were (“leukemia” OR “childhood cancer” OR “childhood cancer survivors”) AND (“exercise capacity” OR “aerobic capacity” OR “fitness” OR “cardiorespiratory fitness” OR “cardiopulmonary fitness” OR “physical fitness” OR “exercise testing” OR “exercise tolerance” OR “exercise” OR “oxygen consumption”). There were no restrictions regarding publication date. Duplicates were removed manually. In addition, we manually searched for relevant articles based on article citations and reference lists.

Study selection

The systematic screening of the articles was carried out in two different phases. First, all search results were screened based on title and abstract in a preliminary screening. The first and the forth authors (VYK and JM) independently screened the titles and abstracts of the all articles based on the eligibility criteria. In the second phase, full text articles were evaluated for inclusion criteria. Similarly, VYK and JM independently screened full text of the articles. Thereafter, any disagreements were resolved through discussion to reach a consensus. If consensus could not be reached, a third opinion was given by the last author (EVB). The study selection and screening of articles was performed using the Covidence software (available online at www.covidence.org).

Data Extraction

Two reviewers (VYK and JM) independently performed data extraction by selecting relevant data and integrating it into two separate columns in Covidence. Thereafter, the two columns were compared and integrated into a final
extraction table. Again, disagreements were resolved through discussion between VYK and JM and in case of disagreement EVB was asked to make a final judgement. Specifically, the following data was extracted for each study: name of the first author, year of publication, study design, sample size, study population (demographic characteristics, diagnosis, treatment status, time after diagnosis), control group (demographic characteristics), and outcome measures (used fitness tests, reporting of the results) (Table 2).

Risk of Bias

Two independent reviewers (VYK and JM) independently assessed risk of bias in all the included articles using the Newcastle-Ottawa scale [15]. Specific items checklist from this scale include suitability of randomization and concealment of allocation, blinding of patients, health care providers, data collectors, and outcome assessors and extent of loss to follow-up. This checklist has two different forms for randomized and non-randomized control studies and it is recommended by the Cochrane Collaboration [16]. Maximum scores obtainable from each of the scales are 9 points, which represents the highest methodological quality. The included studies were defined as moderate or high-quality if they scored ≥6 points, whereas studies with <6 were defined as low-quality studies [17].

Results

Study Selection

According to our search terms a total of 3688 articles were identified from the database search and screening of the reference lists yielded 6 eligible studies that were added. After duplicates (n=1048) were removed, the titles and abstracts of totally 2646 articles were screened and 2520 articles were excluded for not meeting the eligibility criteria. The full text articles of the remaining 126 articles were retrieved for screening based on the inclusion criteria. Following the full text screening, 49 articles that fulfilled the included criteria were included in qualitative synthesis (Figure 1).

Study Characteristics

The characteristics of each eligible study were extracted and presented in Table 2. Of the 49 studies included, 19 studies reported on leukemia survivors [18-36], while 15 of them included ALL survivors [21-35]. In 2 studies, physical fitness of lymphoma [37, 38], brain tumor [39, 40], and solid tumor [41, 42] survivors were reported. Twenty four studies reported physical fitness level of individuals with mixed cancer diagnosis [6, 43-65], and 5 of
them included individuals who had undergone hematopoietic stem cell transplantation (HSCT) [61-65]. The study characteristics in the included articles, ordered according to age (child or adult), time since active cancer treatment, and outcome measures are presented in Figure 2.

**Risk of Bias Analysis**

Following the risk of bias analysis of the included studies, we found that methodological quality score of the case-control studies (n=43) ranged from 2/9 to 8/9. While in the cohort studies (n=6), the methodological quality score ranged between 4/9 and 6/9. The risk of bias analysis of the each eligible study is shown in Table 3.

**Synthesis of the Results**

**Leukemia**

Three studies [18-20] with a total of 139 subjects reported physical fitness level of childhood leukemia. Of these, two studies compared the participants’ scores with normative [18, 19], while the last study compared the participants scores with a healthy control group [20]. Although the age of the individual subjects and off-treatment period of 5 years were largely comparable, the mean time period from active treatment was variable. All the included studies assessed physical fitness using the VO\(_2\) max. The results from two studies revealed that fitness level was significantly lower in leukemia survivors (p<0.05) [18, 20]. In the last study, 25% of individuals had poor fitness levels measured at around 10 years after diagnosis [19]. All three studies regarding leukemia are of low-quality and high risk of bias (2 to 4 out of 9).

**Acute Lymphoblastic Leukemia**

For childhood ALL survivors, 16 studies, totaling 1600 participants were included in this review. In 10 studies, the ALL survivors were compared with healthy control group [21, 23, 26, 27, 29-33, 35], while normative data was used as comparator in the remaining 5 studies [22, 24, 25, 28, 34, 36]. To assess physical fitness, VO\(_2\) max was utilized in 10 studies [21, 23, 25-27, 29-31, 33, 36], six minute walk test (6-MWT) was used in 5 studies [22, 24, 28, 34, 35] and shuttle run test was used in one study [32]. In addition, one study utilized VO\(_2\) max as a predictor for physical fitness level using the Duke Activity Status Index (DASI) [25].
Results from the risk of bias indicated that four studies with high-quality study design, investigated fitness level by means of VO$_2$ max [23, 26, 29, 30]. Of these, three studies revealed that ALL survivors had a significantly poorer fitness level compared to controls (p<0.05) [23, 29, 30]. Participants were also compared whether they received cranial radiotherapy (RT) treatment, and results indicated that individuals treated with cranial RT had lower fitness level than individuals treated without cranial RT (p<0.05) [29, 30]. In addition, Ness et al. consisting of 365 ALL survivors who had long period after treatments (≥10 years) indicated that survivors participating in a moderate to vigorous physical activity for a minimum of 30 minutes per day had higher VO$_2$ max than controls [29]. Furthermore, one high-quality study, which assessed fitness levels in mix population (pediatric and adult) reported similar results with controls for the pediatric population (p>0.05), while adult population had lower fitness level compared with the controls (p<0.05) [26]. Similarly, results from two studies with low-quality design that assess VO$_2$ max values were indicated lower fitness in ALL survivors regardless of gender than those of the controls [25, 27].

There is only one study with high-quality, used 6-MWT as outcome measure and found about 46.5% of ALL survivors had performance limitation (≤1.3 SD from normative) [34]. One study reported that 6-MWT distance covered in feet was similar between ALL subjects with siblings who had controlled medical condition [35]. Whereas in three studies that reported distance following 6-MWT found a comparatively lower score for ALL survivors compared with controls [22, 24, 28]. In these studies participants treated only with chemotherapy (CT) and mean duration after treatment was less than 10 years. All of these four studies using 6-MWT had low-quality level and high risk of bias. The only study with low-quality that reported physical fitness using shuttle run test indicated that ALL survivors treated without HSCT had similar fitness level with controls (p>0.05); on the other hand their fitness level was higher than ALL survivors who treated with HSCT [32].

**Lymphoma**

Only 2 studies comprising 243 investigated fitness levels of individuals diagnosed with lymphoma [37, 38]. One study, that included adults who had experienced childhood cancer, found that 30% of the participants had poor fitness levels (VO$_2$ max<20 ml/kg/m$^2$). Lymphoma patients in this study were previously treated with chest RT and mean duration after treatments was 15 years [37]. The second study by Ehrhardt et al. [38] included 200 adults who were ≥20 years post treatment. Expectedly, results showed that 32% of the participants had fitness impairments,
which is about 22% higher than expected values of the general population [38]. To summarize, both of these two studies, those were of low-quality revealed that around 30% of lymphoma survivors had poor physical fitness even after long post treatment period.

**Brain tumor**

Two studies consisting of 93 participants assessed physical fitness in survivors of childhood brain tumor [39, 40]. Both of these studies used VO$_2$ max as outcome measure. One of the studies, which had a high-quality included adults 10-20 years post-treatment found that there is low level of physical fitness in survivors compared to controls (p<0.001) [39]. The second study included a small sample (n=14) and was of low-quality. The results showed that there is a reduction in physical fitness for pediatric survivors with a posterior fossa tumor who are in their early (<2 years) post treatment phase (p<0.001). Significant reduction was also found when these results are compared to that of children diagnosed with other types of cancer (p<0.001) [40]. Impairments in physical fitness were reported in these two studies both at early and late post treatment period.

**Solid tumors**

Two studies were found concerning solid tumors used 6-MWT as outcome measure, totaling 737 participants [41, 42]. In one study, Fernandez-Pineda et al. included survivors who were diagnosed with bone or soft tissue sarcoma mostly treated with surgery and CT and reported that they had lower fitness levels in comparison to controls (p<0.001) [41]. In the second study, 531 adult survivors who were ≥10 years post diagnosis was included and a few proportion (14%) of the participants who walked less than 500 meters were classified as poor endurance performers [42]. Both of these studies were conducted at least 10 years post treatments yet were of low-quality study design.

**CCS with a mixed histories**

Most of the studies in literature, aimed at assessing fitness of cancer survivors have been performed on a population with a mixed history of cancer. The results from the reviewed studies, presented either as percentiles or numeric values tend to suggest that most of the studies used the VO$_2$ max test [47-52, 54-59]. The 6-MWT was used in 6 studies [6, 44-46, 53, 60], while one study utilized a 12-minute running test [43].
Of these studies mentioned above, four studies were of high-quality and low risk of bias [6, 43, 44, 46]. The 6-MWT used in three of these studies [6, 44, 46] and the last study used 12-minute running test as outcome measure [43]. In this study, participants with brain tumor and hematologic cancer had poor distance covered in 12-minute running test (35% and 18%, respectively), whereas those with a solid tumor had similar fitness level compared to the control subjects during the age of 18 in military service [43]. In another high-quality study, 11.8% of survivors with growth hormone deficiency who treated with cranial RT had poor distance covered in 6-MWT (<400 m), whereas only 5.7% of those without hormone deficiency had poor distance (p<0.05) [44]. Hoffman et al. also reported that all CCS who treated with CT and mean duration after CT was less than 10 years had lower fitness level than controls (p<0.05); according to diagnosis low level of fitness determined in only individuals diagnosed with central nervous system tumor (p<0.001) [6].

Based on the VO$_2$ max in CCS, results in all studies with low-quality design have reported impaired VO$_2$ max for overall CCS population [49-52, 54, 56, 58]. Four of these studies stated that VO$_2$ max was lower among CCS within first ten years after treatment than in controls [49, 50, 52, 58]. In one study, 17% of CCS had a lower fitness level than normative [54], while in the other, the median value was found to be 14% lower than the predicted value [56]. In both studies CCS mostly treated with anthracyclines and mean duration after treatments was around 7 years [54, 56]. Only one study found similar VO$_2$ max level when compared with controls in CCSs mostly treated with anthracyclines who were within the first ten years after treatments [51]. In three other studies, results were presented according to gender, apart from one study it was found that fitness levels were lower in both male and female CCS compared to their controls group counterparts [47, 48, 55]. Only one study reported that there is a similar fitness level between male survivors and controls, while female survivors had lower scores than controls [48]. Similarly, results from the feasibility study, in which steep ramp test and was used, showed that one minute step protocol can be used to assess VO$_2$ max [59].

Furthermore, Smith et al. reported that performance limitation was the most prevalent among bone and central nervous system tumor survivors (15.6 and 9.2%, respectively) [53]. Moreover, performance limitation was most prevalent in CCS older than 50 years (13.8%) and whose BMI lower than 18.5 kg/m$^2$ (12.5%) or higher than 40 kg/m$^2$ (11.2%) [53]. In another study, rate of fitness impairment was 17.6% for all CCSs and participants treated
with chest RT were found to have the highest rate of impairment, which was 27.4% [45]. Both of these studies were of low-quality design and survivors who were in 10 to 50 years after treatments.

**Hematopoietic stem cell transplantation survivors**

Physical fitness levels of CCS who had undergone HSCT were reported in 5 studies [61-65]. In one study that is of high-quality, reported similar results in 6-MWT in CCS mostly undergone allogeneic HSCT and after more than 10 years from transplantation (p>0.05) [63]. Contrastingly, one study by Larsen et al. that used VO$_2$ max as outcome measure found lower fitness levels among HSCT patients compared to controls after less than 10 years from transplantation [62]. Poor fitness levels were also reported in two studies that used shuttle run test as outcome measure [61, 65]. In one study, participants were tested during early (1-2 years) and late (≥4 years) period after HSCT and in both groups lower level of fitness was reported [65]. In the other study, participants were evaluated longitudinally. The VO$_2$ max was lower in HSCT survivors both at the initial and later tests and increased by 4% per year, it was 69.3% of predicted at the end of 5 years. Still, fitness level was lower in comparison with normative (p<0.001) [64].

**Discussion**

The aim of the present review was to summarize and provide a more detailed insight in the scientific literature concerning the level of physical fitness in CCS after active cancer treatment. To the best of our knowledge this is the first systematic review to analyze and review all available literature on fitness tests during post-treatment period for all type of childhood cancers. Altogether, CCS show poor physical fitness levels in comparison with their healthy counterparts. In addition, impairments in fitness level seem to be present for long even in CCS. Our results also showed that there is a poor level of evidence regarding impairment of physical fitness in leukemia, solid tumor, and lymphoma survivors. Moreover, a significant proportion of the studies reviewed were of low-quality, only 11 (22.4%) of studies had moderate to high-quality. Hence, the need for high-quality studies with better study designs and large sample sizes were sought for the future.

**Summary of evidence**

It is well known that good level of physical fitness in CCS is associated with a higher quality of life and overall survival rates [8]. Physical fitness is one of the most important aspects that are directly related to mortality in
general population. Such an association has been shown among earlier studies conducted in adult cancer patients and has been supported by systematic reviews and meta-analysis [17, 66]. According to these studies in adult cancer survivors who engage in high and moderate levels of physical fitness have a decreased mortality risk in comparison with those who have a low level of fitness (RR: 0.55; 95% CI: 0.47-0.65 and RR: 0.80; 95% CI: 0.67-0.97, respectively). Even though this is not yet proven among children with cancers or CCS, assessment of physical fitness in general, and in CCSs in particular should not be underestimated and should become the standard care in the clinical setting of CCSs. Unfortunately, no such practice is currently performed either during hospitalization or during long-term follow-up is routinely been performed in children with cancer.

*Acute Lymphoblastic Leukemia*

In studies with high-evidence score and based on the extracted information from CCS with a history of ALL, it can be inferred that serious impairments in fitness levels prevalent in this population [23, 29, 30, 34]. In contrast, however, to the latter findings of Bar et al. who reported no difference in cardiopulmonary fitness levels between former ALL patients and age-matched untrained control [26]. A major drawback of the latter study is the small number of included patients (n=19) and the moderate level of evidence (see Table 3) [26]. In addition, when we looked at the different fitness tests, the VO₂ max was the major test utilized in studies with high-quality [23, 26, 29, 30]; only in one study 6-MWT scores were recorded, yet results were presented in percentage; nearly half of the ALL survivors had performance limitation [34]. Nevertheless, the evidence appears to be also scanty. Hence, based on these results, it was concluded that there is limited or conflicting evidence regarding fitness levels in childhood ALL survivors.

*Brain tumor*

Because of the aggressiveness of this type of cancer, it is a big challenge to find sufficient brain tumor survivors that can be incorporated in any high-quality studies. Incidentally, this review found that one high-quality study in brain tumor survivors reported poor level of fitness. We must, however, take into account that fitness level was assessed according to the DASI, a patient reported questionnaire [39]. Therefore, there is a possibility that patient self-reporting proof of fitness is questionable and should be regarded as inadequate. It seems also that for future brain
tumor survivor studies, high-quality design and objective measurement of physical fitness should be emphasized as not only important and valuable, but necessary.

**CCSs with mixed cancer histories**

According to the results of four studies with high-quality, poor fitness level was found in CCS with mixed histories of cancer [6, 43, 44, 46]. Among these studies only 6-MWT was scarcely used results in standardized numeric data (mean±SD) [6]. In the remaining studies, the results were expressed as percentage and the rate of impairment of physical fitness ranged from 5.7% to 35% [43, 44, 46]. Also, there were no high-quality studies that used VO2 max in CCSs. In survivors of HSCT, there was only one study with high-quality and similar fitness level was reported between the groups [63]. There was some statistically differences between the groups in this study such as age, height, and weight. Hence, these results should be interpreted carefully. To conclude more precisely there is great need to further investigate fitness level in HSCT survivors. Furthermore, despite the fact that there were relatively more studies conducted on CCSs diagnosed with different cancer types, quality level of the studies was not satisfactory. In addition, presentation of the results in existing studies is usually not presented in standardized forms, making comparison of results difficult. For these reasons, it is important for future studies to perform studies with higher level of quality with standardized data.

**Outcome Measures for Cardiopulmonary Fitness**

Exercise testing of cardiorespiratory fitness in post-cancer treatment in adults has become an important part of clinical diagnosis in adult cancer survivors [67, 68]. According to the guidelines of the American Thoracic Society, both the 6-MWT and cardiopulmonary exercise testing (CPET) can be used for functional assessment of physical fitness [69]. In this present review, the outcomes of the 6-MWT and CPET, i.e. distance covered and VO2 max, respectively, are being considered as important outcome measures. Nevertheless, since the shuttle run test and the 12-minute running test have been used as well, it is important to suggest a gold standard for adequate classification. Moreover, these tests can be used as a complementary assessment tool, except for the differences in terms of information, which makes comparison among the studies difficult. In addition, there is presently no consensus regarding the best test protocol in assessing physical fitness at present as well. Presently, ergometer or treadmill tests, as in the case of CPET, are mostly used utilized in existing studies; however there is a need to adequately
explore protocols used to determine VO₂ max level in CCS. Moreover, increase in the use of higher number of muscle mass that is involved during treadmill testing may cause a higher values of VO₂ max in comparison to ergometer testing, indicating another limitation in our study and the need for care in interpreting this results [69, 70].

Another different CPET ergometer protocol is the steep ramp test was used in one study and they reported that such a test is feasible and valid in CCSs [59]. In two studies indirect patient-reported inventory, DASI was used to predict VO₂ max. The use of questionnaire that is validated against the cardiopulmonary exercise test in healthy and sick populations; however, because it is a patient-reported inventory, the results could be overestimated [71, 72].

Around one-third of all studies in the present review used the 6-MWT to measure physical fitness level in CSSs. The major advantage of this test is inexpensive and can be performed easily in clinical setting [73, 74]. This test has also been accepted as an important prognostic factor for survival in adult cancer patients [75, 76]. However, there are inconclusive results in the eligible studies. For instance, it was discovered that different cut-off points were used to determine poor fitness level in different studies. Therefore, we suggest that there is a need for more studies using the 6-MWT to adequately and conclusively determine presence of impairment in physical fitness in CCS.

Finally, among the studies that were included in the present study, shuttle run test and 12-minute running test were used as well. In the literature, there is no consensus on the data concerning feasibility and validity of these two test protocols in cancer patients in general and in CCS in particular. Taken together, these results in the current literature suggest that there is discrepancy in the test methods for physical fitness in childhood cancer thereby warranting more studies in CCS. Also, the development of consensus and recommendations on exercise testing and results analyses for CCSs is important in future research.

In conclusion, physical fitness in CCSs is hampered by either tumor growth, medication of long-term side-effects that can occur during and post cancer treatment. According to results of the high-quality studies, poor fitness level is described in ALL, brain tumors, and individuals with mixed cancer histories. However, these results should be interpreted with care since of the small number of studies with high-evidence. Because of the latter phenomenon, there is great need to study physical fitness level within a better designed study in larger populations. In addition, there is a need for more information about implementetation of fitness tests and interpretation of the results of these tests for CCS. To date, it is difficult to interpret and compare results according to diagnosis. Although a reason is not easy to provide, a lack of consensus should be put aside and focus should be directed towards the beneficial effects
for future studies. To uncover impairments of physical fitness in CCSs will, in the future, contribute to treatment protocols and to design exercise programs for this population. As a result, physical fitness level of previous childhood cancer cases should be carefully examined even after long period from treatment.

**Compliance with Ethical Standards**

**Funding:** Vesile Yildiz Kabak was financially supported by the Erasmus Program of Turkish National Agency during her study period at Ghent University.

**Ethical approval:** This article does not contain any studies with human participants or animals performed by any of the authors.

**Conflict of interest:** The authors indicated no potential conflicts of interest.
References


Records identified through database searching (n = 3688)
  PubMed: 836 (1 March 2018)
  Web of Science: 977 (28 February 2018)
  Embase: 1875 (27 February 2018)

Records identified through reference lists (n = 6)

Records after duplicates removed (n = 2646)

Records screened (n = 2646)

Records excluded (n = 2520)

Full-text articles assessed for eligibility (n = 126)

Full-text articles excluded, with reasons (n = 77)
  28: conference paper
  23: wrong outcomes
  9: articles not in English
  7: patient population
  7: study design
  3: not an original research

Studies included in qualitative synthesis (n = 49)

Figure 1. Study Selection Process
Table 1 PICOS based eligibility criteria

<table>
<thead>
<tr>
<th>Inclusion criteria</th>
<th>Exclusion criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>P</strong></td>
<td></td>
</tr>
<tr>
<td>Children</td>
<td>Adults (&gt;18 years)</td>
</tr>
<tr>
<td>Adolescents</td>
<td>cancer patients</td>
</tr>
<tr>
<td>Adults with completed</td>
<td>Children with cancer (&lt; 18 years) during</td>
</tr>
<tr>
<td>childhood cancer treatment</td>
<td>cancer treatment</td>
</tr>
<tr>
<td>(surgery, chemotherapy, radiotherapy</td>
<td></td>
</tr>
<tr>
<td>or hematopoietic stem cell</td>
<td></td>
</tr>
<tr>
<td>transplantation)</td>
<td></td>
</tr>
<tr>
<td><strong>I</strong></td>
<td></td>
</tr>
<tr>
<td>Fitness level based</td>
<td></td>
</tr>
<tr>
<td>Tests for submaximal or maximal</td>
<td>Intervention study</td>
</tr>
<tr>
<td>exercise capacity in numbers or</td>
<td></td>
</tr>
<tr>
<td>percentages</td>
<td></td>
</tr>
<tr>
<td><strong>C</strong></td>
<td></td>
</tr>
<tr>
<td>-</td>
<td></td>
</tr>
<tr>
<td><strong>O</strong></td>
<td></td>
</tr>
<tr>
<td>$\text{VO}_2\text{max}$</td>
<td></td>
</tr>
<tr>
<td>6-min walking test</td>
<td></td>
</tr>
<tr>
<td>Shuttle run</td>
<td></td>
</tr>
<tr>
<td>12-minute running test</td>
<td></td>
</tr>
<tr>
<td><strong>S</strong></td>
<td></td>
</tr>
<tr>
<td>Case Control Study</td>
<td>Non-English-, Dutch-, French-, or German,</td>
</tr>
<tr>
<td>Cohort Study</td>
<td>Systematic Review, Meta-analysis</td>
</tr>
<tr>
<td>Cross-sectional study</td>
<td></td>
</tr>
<tr>
<td>Author, Year</td>
<td>Study design</td>
</tr>
<tr>
<td>-------------</td>
<td>-------------</td>
</tr>
<tr>
<td><strong>Participants Diagnosed with Leukemia</strong></td>
<td></td>
</tr>
<tr>
<td>Black P et al, 1998 [19]</td>
<td>CC</td>
</tr>
<tr>
<td>Jenney ME et al, 1995 [20]</td>
<td>CC</td>
</tr>
<tr>
<td>van Brussel M et al, 2006 [18]</td>
<td>CC</td>
</tr>
<tr>
<td><strong>Participants Diagnosed with ALL</strong></td>
<td></td>
</tr>
<tr>
<td>Hartman A, et al, 2013 [22]</td>
<td>CC</td>
</tr>
<tr>
<td>Hung SH et al, 2017 [28]</td>
<td>CC</td>
</tr>
<tr>
<td>Ness KK et al, 2012 [34]</td>
<td>CC</td>
</tr>
<tr>
<td>Ruble K et al, 2015 [35]</td>
<td>CC</td>
</tr>
<tr>
<td>Shimomura Y et al, 2011 [24]</td>
<td>CC</td>
</tr>
</tbody>
</table>
Bell W et al, 2006 [27]  
Healthy siblings  
M: 12.4±3.4/12.8 ±2.7  
F:11.8±3.7/12.2±2.8  
AC+Cranial RT  
VO₂max, motorized treadmill  
ALL<C (M: 39.9 vs 47.6 ml/kg/min, p<0.05; F: 30.5 vs 41.3 ml/kg/min, p<0.05)

Christiansen JR et al, 2015 [21]  
Age, body weight, SBP-matched HC  
28.6[18.6-46.5]/30.2[19.4-45.2]  
Age at diagnosis: 5.3 [0.3-16]  
CT:AC(78%), MTX (95%), VCR(100%)  
RT: 15%  
HSCT: 2%  
VO₂max, electrically braked CE  
47% had poor fitness  
No AC < LD and M-HD AC (17% vs 48% and 67%, respectively p<0.001)

Hauser M et al, 2001 [33]  
Age, body surface area matched HC  
69.3±27.1/72.4±22.3 months  
At least 6 months  
CT:AC+CP  
VO₂max, Bruce protocol, treadmill  
ALL=C (38.2±13.3 vs 42.9±9.3 ml/kg/min, p<0.05)

Hrstkova H et al, 2006 [31]  
Norm  
28.5[18.6-46.5]  
Age at diagnosis: 5.4[0.3-16]  
CT: VCR (100%)  
MTX (95%)  
AC (77%)  
VO₂max, CE  
42% of had poor fitness  
Mean: 35±8.5 ml/kg/min

Jarvela LS et al, 2010 [23]  
Non-athletic HC  
22.07[16.7-30.3]/NR  
28.6±18.6/28.9±7.5  
21.9[11.0-30.7]  
CT: DR  
VO₂max, electronically braked BE  
ALL<C (34.8±9.3 vs 40.5±8.8; −5.7 ml/kg/min, p<0.001)

Mrydal OH et al, 2018 [36]  
Norm  
24.3±4.9/ NR  
25-38y: 36.5%/37.7%  
≥15 y: 65.2%  
CT: AC (72.2%)  
CP (40.9%)  
RT: cranial (33.9%)  
VO₂max, graded maximal exercise test, treadmill  
ALL<C (30.7 vs 39.9 ml/kg/min, p<0.001); treated with cranial RT< without RT (22.1±5.5 vs 25.5±6.2 ml/kg/min, p<0.001)

Ness KK et al, 2007 [25]  
Norm  
30.2±7.1  
24.6±4.8  
CT: MTX and VCR(100%)  
RT: 69.3%  
VO₂ max, Duke Activity Status Index  
ALL<C (M: 32.4±4.68 vs 41±5.68; F: 29.18±7.12 vs 33.8±4.89 ml/kg/min; p<0.001)

Ness K et al, 2015 [29]  
Age, sex, race matched C  
28.6±5.9/28.9±7.5  
21.9[11.0-30.7]  
CT: MTX+VCR(100%)  
L-Asp (98.9%)  
AC (40.8%)  
VO₂max, submaximal CPET  
ALL<C, treated with cranial RT< without RT (22.1±5.5 vs 25.5±6.2 ml/kg/min, p<0.001)

Age, sex, race matched C  
24.3±4.9/ NR  
18-24y: 63.5%/62.3%  
25-38y: 36.5%/37.7%  
4-9 y: 10.4%  
10-14 y: 24.4%  
≥15 y: 65.2%  
CT: AC (72.2%)  
CP (40.9%)  
RT: cranial (33.9%)  
VO₂max, graded maximal exercise test, treadmill  
ALL<C (30.7 vs 39.9 ml/kg/min, p<0.001); treated with cranial RT<without (F:23.5 vs 27.3 ml/kg/min, p=0.01; M:31.6 vs 38.9 ml/kg/min, p=0.01)

Taskinen MH et al, 2013 [32]  
Retro HC  
13.3[9.2-20.1]/12.0 (9.0-30.0)/12  
6.8[2.8-13.4]/5.2[3.5-11.6]  
CT: L-Asp, DR, Dexa, VCR, MTX  
10×5 m shuttle run  
Patients without HSCT=C (p>0.05), patients without HSCT>with HSCT (SDS:−0.5±1.9 vs -1.3±1.8, p<0.001)
<table>
<thead>
<tr>
<th>Authors</th>
<th>Year</th>
<th>Cohort</th>
<th>Age, Sex, or Other Matched C</th>
<th>Age, Sex, or Other Matched HC</th>
<th>Bone, Soft Tissue</th>
<th>Bone, Soft Tissue</th>
<th>CT, RT</th>
<th>CT, RT</th>
<th>VO₂ max, Duke Activity Status Index</th>
<th>Percentage Rank</th>
<th>6 MWT</th>
<th>BT/C BT/C+chest RT</th>
<th>VO₂ max</th>
<th>VO₂ max</th>
<th>Percentage Rank</th>
<th>6 MWT</th>
<th>BT/C BT/C+chest RT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adams MJ et al, 2004</td>
<td></td>
<td>43</td>
<td>NA</td>
<td>NA</td>
<td>31.9 [18.7-49.5]</td>
<td>14.3 [5.9-27.5]</td>
<td>CT: 48.8%</td>
<td>RT: 100%</td>
<td>VO₂ max (VO₂ max↓; ≤20 ml/kg/m²)</td>
<td>6 MWT (Impairment: &lt;10th percentile)</td>
<td>32% had fitness impairment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ehrhardt MJ et al, 2016</td>
<td></td>
<td>200</td>
<td>Norm</td>
<td>10 [1-19]</td>
<td>34 [20-58]</td>
<td>CT: Anti-metabolite (91.5%), Alkylating agent (90.5%), RT: 44%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ness KK et al, 2010</td>
<td></td>
<td>78/78</td>
<td>Age, sex, zip code matched C</td>
<td>22 [18.4-58.3]/25[18-54]</td>
<td>5.9 y; 15.4%</td>
<td>10-14 y: 38.5%</td>
<td>CT: any (30.8%)</td>
<td>RT: CNS (96.2%)</td>
<td>S: 87.2%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wolfe KR et al, 2012</td>
<td></td>
<td>14</td>
<td>Chronic illnesses and HC</td>
<td>14.4±1.86/NR</td>
<td>Age at diagnosis:</td>
<td>5.59±2.89</td>
<td>Adjuvant CT (78.5%)</td>
<td>RT+S: 6%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fernandez-Pineda et al, 2017</td>
<td></td>
<td>206/206</td>
<td>Age, sex, race matched C</td>
<td>Bone: 38.0±9.7, soft tissue: 34.7±9.1/33.1±7.2</td>
<td>Age at diagnosis:</td>
<td>13.1±4.2, soft tissue: 11.4±5.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ness KK et al, 2013</td>
<td></td>
<td>531</td>
<td>NA</td>
<td>31.6 [18.7-63.8]</td>
<td>25.2 [10.7-48.2]</td>
<td>CT: Vinca-alkoloid (54%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Armstrong GT et al, 2015</td>
<td></td>
<td>1807</td>
<td>NA</td>
<td>31 [18-65]</td>
<td>22.6 [10.4-48.3]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chemaitilly W et al, 2015</td>
<td></td>
<td>748</td>
<td>NA</td>
<td>34.2 [19.4-59.6]</td>
<td>27.3 [10.8-47.7]</td>
<td>Cranial RT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beulertz J et al, 2016</td>
<td></td>
<td>13/13</td>
<td>Age, sex matched HC</td>
<td>11.1±3.53/11.29±3.42</td>
<td>2.13±1.18</td>
<td>CT(n=11) RT(n=4) S (n=9)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hartman A et al, 2018</td>
<td></td>
<td>71/75</td>
<td>Age matched C</td>
<td>AML-NBL-WT:31.5[22.5-62.6]-29.1[20.4-43.3]-28.2[18.8-47.9]/26.9[17.9-61.7]</td>
<td>Age at diagnosis:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Participants Diagnosed with Brain Tumor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants Diagnosed with Solid Tumor</td>
</tr>
<tr>
<td>Participants with Mix Childhood Cancer Histories</td>
</tr>
<tr>
<td>Authors</td>
</tr>
<tr>
<td>----------------------</td>
</tr>
<tr>
<td>Hoffman MC et al.</td>
</tr>
<tr>
<td>Smith WA et al.</td>
</tr>
<tr>
<td>Ahomäki R et al.</td>
</tr>
<tr>
<td>Smith WA et al.</td>
</tr>
<tr>
<td>Braam KI et al.</td>
</tr>
<tr>
<td>De Caro E et al.</td>
</tr>
<tr>
<td>De Caro E et al.</td>
</tr>
<tr>
<td>Johnson D et al.</td>
</tr>
<tr>
<td>Matthys D et al.</td>
</tr>
<tr>
<td>Miller AM et al.</td>
</tr>
</tbody>
</table>
Pihkala J et al., 1995 [56]  
CC 30  Norm 17[8-25] 7[2-13]  CT: AC(93.3%) RT: 100% HSCT:16.6% VO2max, CE  
Exercise tolerance:[(35.4±9.7 ml/kg/min, median 14% lower than norm)

Sato T et al, 2001 [58]  
CC 29/41  C referred to hospital without heart disease 12.3±3.0/13.1±2.9 24±10 months  CT: AC VO2max, CE CC<C (22.0±3.7 vs 28.5±7.1 ml/kg/min, p< 0.01

Tham EB et al, 2013 [54]  
CC 30/30  Age and gender matched HC 15.2±2.7/13.8±3.4 7.6±4.5  CT: AC(100%) RT: chest(17%) VO2max, CE  
35±10 ml/kg/min (17%↓ than norm)

Warner JT et al, 1997 [47]  
CC 56/32  Healthy siblings 12.4±3.4/12.8±2.7 11.8±3.7/12.2±2.8  VO2max, motorised treadmill  
CC˂C (22.0±3.7 vs 28.5±7.1 ml/kg/min, p< 0.01

Braam KI et al, 2015 [59]  
Feasibility 61  NA 12.9±3  VO2max, steep ramp test and CPET  
VO2max according to steep ramp test/CPET: 26.6±22.2-34/29.8±24.2-36.4 ml/kg/min

Participants Treated with HSCT

Hogarty AN et al, 2000 [64]  
CC/ follow-up 33  Age, sex matched norm 17.4(7.8-33.9) 1.6[0.3-9] test: Allo: 57.5% Auto: 42.4% VO2max, ramp cycle protocol, CE  
Initial test: HSCT<norm (24.7±5.1 ml/kg/min, 61.6% of norm, p<0.05). VO2max↑ at the end of 5 y: HSCT<norm (69.3% of norm, p<0.001)

Larsen RL et al, 1992 [62]  
CC 31/70  Nonathletic HC 15.8±7.5/12.2±3.2 3.9±3.3  Allo: 61.2%, Auto: 35.4%, Syn: 3% VO2max, incremental protocol, CE  
HSCT survivors=C (24±6 vs 34±7 ml/kg/min, p<0.05)

Slater ME et al, 2015 [63]  
CC 119/66  Age, sex matched siblings 27.4±0.7/25.0±1.0 12.7±0.6  Allo: 73.1% Auto: 26.9% 6 MWT  
HSCT survivors=C (583.1±29.8 vs 591.9±39.1 m, p>0.05)

Bianco et al, 2014 [61]  
CC 18/40  Non-athletic HC 7.55±2.43/7.92±1.78 10-24 months  4×10 m shuttle run test  
HSCT<C (16.04±2.20 sec vs 14.28±1.5 sec, p<0.05)

Hovi L et al, 2010 [65]  
CC 94/522  Age, sex matched C 11(6-20) Late test: 13(7-30)  Allo 10x5 m shuttle run test  
HSCT<C both at early and late tests (SDS: -1.2(1.8) and -1.2(1.9), respectively, p<0.001)


Data are expressed as mean ± standard deviation, mean(range), median(range), or number (n or %).
Figure 2 Flow chart of streaming of included studies by diagnosis, study population, time since active cancer treatment, and outcome variables. ALL: Acute lymphoblastic leukemia, HSCT: Hematopoietic stem cell transplantation, 6 MWT: Six minute walk test
<table>
<thead>
<tr>
<th>Author, Year</th>
<th>Case-Control Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>1 Black P et al, 1998 [19]</td>
<td>+</td>
</tr>
<tr>
<td>2 Jenney MEM et al, 1995 [20]</td>
<td>+</td>
</tr>
<tr>
<td>3 van Brussel M et al, 2006 [18]</td>
<td>+</td>
</tr>
<tr>
<td>4 Hartman A et al, 2013 [22]</td>
<td>+</td>
</tr>
<tr>
<td>5 Hung SH et al, 2017 [28]</td>
<td>+</td>
</tr>
<tr>
<td>6 Ness KK et al, 2012 [34]</td>
<td>+</td>
</tr>
<tr>
<td>7 Ruble K et al, 2015 [35]</td>
<td>+</td>
</tr>
<tr>
<td>8 Shimomura Y et al, 2011 [24]</td>
<td>+</td>
</tr>
<tr>
<td>9 Bar G et al, 2007 [26]</td>
<td>+</td>
</tr>
<tr>
<td>10 Bell W et al, 2006 [27]</td>
<td>+</td>
</tr>
<tr>
<td>12 Hauser M et al, 2001 [33]</td>
<td>+</td>
</tr>
<tr>
<td>13 Hristkova H et al, 2006 [31]</td>
<td>4</td>
</tr>
<tr>
<td>14 Jarvela LS et al, 2010 [23]</td>
<td>+</td>
</tr>
<tr>
<td>15 Myrdal OH et al, 2018 [36]</td>
<td>+</td>
</tr>
<tr>
<td>17 Ness K et al, 2015 [29]</td>
<td>+</td>
</tr>
<tr>
<td>18 Taskinen MH et al, 2013 [32]</td>
<td>+</td>
</tr>
<tr>
<td>21 Wolfe KL et al, 2012 [40]</td>
<td>+</td>
</tr>
<tr>
<td>22 Fernandez-Pineda et al, 2017 [41]</td>
<td>+</td>
</tr>
<tr>
<td>23 Beulertz J et al, 2016 [46]</td>
<td>+</td>
</tr>
<tr>
<td>24 Hartman A et al, 2018 [60]</td>
<td>+</td>
</tr>
<tr>
<td>26 Smith WA et al, 2014 [53]</td>
<td>+</td>
</tr>
<tr>
<td>27 Ahomaki R et al, 2017 [43]</td>
<td>+</td>
</tr>
<tr>
<td>28 Braam KI et al, 2016 [49]</td>
<td>+</td>
</tr>
<tr>
<td>29 De Caro E et al, 2006 [51]</td>
<td>+</td>
</tr>
<tr>
<td>30 De Caro E et al, 2011 [52]</td>
<td>+</td>
</tr>
<tr>
<td>31 Johnson D et al, 1997 [50]</td>
<td>+</td>
</tr>
<tr>
<td>32 Kaneko S et al, 2016 [57]</td>
<td>+</td>
</tr>
<tr>
<td>33 Matthys D et al, 1993 [48]</td>
<td>+</td>
</tr>
<tr>
<td>34 Miller AM et al, 2013 [55]</td>
<td>+</td>
</tr>
<tr>
<td>35 Pihkala J et al, 1995 [56]</td>
<td>+</td>
</tr>
<tr>
<td>36 Sato et al, 2001 [58]</td>
<td>+</td>
</tr>
<tr>
<td>37 Tham EB et al, 2013 [54]</td>
<td>+</td>
</tr>
<tr>
<td>38 Warner JT et al, 1997 [47]</td>
<td>+</td>
</tr>
<tr>
<td>39 Bianco et al, 2014 [61]</td>
<td>+</td>
</tr>
<tr>
<td>40 Hogarty AN et al, 2000 [64]</td>
<td>+</td>
</tr>
<tr>
<td>41 Hovi L et al, 2010 [65]</td>
<td>+</td>
</tr>
<tr>
<td>42 Larsen KL et al, 1992 [62]</td>
<td>+</td>
</tr>
<tr>
<td>43 Slater ME et al, 2015 [63]</td>
<td>+</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cohort Studies</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>NOS for Cohort Studies Total Score (0-9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>44 Adams MJ et al, 2004 [37]</td>
<td>+</td>
<td>NA</td>
<td>+</td>
<td>NA</td>
<td>+</td>
<td>NA</td>
<td>NA</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>45 Armstrong GT et al, 2015 [45]</td>
<td>-</td>
<td>NA</td>
<td>+</td>
<td>NA</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>46 Chemaitilly W et al, 2015 [44]</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>47 Ehrhardt MJ et al, 2017 [38]</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>NA</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>48 Ness KK et al, 2013 [42]</td>
<td>+</td>
<td>NA</td>
<td>-</td>
<td>NA</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>49 Braam KI et al, 2015 [59]</td>
<td>+</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>-</td>
<td>NA</td>
<td>NA</td>
<td>Feasibility study*</td>
<td></td>
</tr>
</tbody>
</table>

Table 3 Risk of bias analysis of the included studies
+ = score fulfilled; - = score not fulfilled, NOS: Newcastle-Ottawa Quality Assessment Scale, NA: not applicable.

Newcastle-Ottawa Quality Assessment Scale for Case-Control Studies; 1-Case Definition: Is the case definition adequate? (Independent validation, record linkage or self-reported), 2-Case Description: Representativeness of cases (Random sample: description of area, hospital and clinic), 3-Selection of Controls: Selection of controls (Community controls with no history of disease), 4-Control Definition: Controls with no history of disease (endpoint), 5-Comparability: Controlled for the most important confounders (age or other factors), 6-Blindness: Researchers were blinded for participant’s status, 7-Same method used for controls and cases, 8: Non-Response Rate: same rate for both groups.

Newcastle-Ottawa Quality Assessment Scale for Cohort Studies; 1-Representativeness of exposed cohort (truly representative average in the community), 2-Selection of the non-exposed cohorts: Drawn from the same community, 3-Ascertainment of exposure: Independent validation or self-reported; 4-Demonstration that outcome of interest was not present at start of study; 5-Comparability of cohorts on the basis of the design or analysis (age or other factors); 6-Assessment of outcome: Independent or blind assessment or record linkage, 7-Enough follow-up time for disease 8-Adequacy of follow up of cohort.

* As our knowledge there is no bias analysis method for feasibility studies. Because of this, the risk of bias analysis of this study could not be performed comprehensively.