Effects of intradialytic cycling exercise on exercise capacity, quality of life, physical function and cardiovascular measures in adult haemodialysis patients: a systematic review and meta-analysis.

Hannah ML Young MSc\(^1\), Daniel S March PhD\(^1\), Matthew PM Graham –Brown MBChB\(^2\), Arwel W Jones PhD\(^3\), Ffion Curtis PhD\(^3\), Charlotte S Grantham MSc\(^1\), Darren R Churchward MRes\(^1\), Patrick Highton\(^1,2\) MSc, Alice C Smith PhD\(^1\), Sally J Singh PhD\(^4\), Chris Bridle PhD\(^3\), James O Burton MD\(^1\)

\(^1\)Dept of Infection, Immunity & Inflammation & John Walls Renal Unit, University of Leicester and University Hospitals of Leicester NHS Trust, UK. \(^2\)National Centre for Sport and Exercise Medicine, Loughborough University, \(^3\)Lincoln Institute for Health, University of Lincoln, UK. \(^4\)Centre for Exercise & Rehabilitation Science, Leicester BRU, Glenfield Hospital, UK

Corresponding author:
Hannah ML Young
Academic Unit
Leicester General Hospital
Gwendolen Road
Leicester
LE5 4PW

Word count: Manuscript body (including abstract) 3499

Short title: Intradialytic aerobic cycling exercise: systematic review.
Abstract

Background. Intradialytic cycling (IDC), delivered during haemodialysis (HD) has the potential to improve many health issues. This systematic review and meta-analysis examines the evidence on the effects of IDC on exercise capacity, quality of life (QOL), physical function & cardiovascular health.

Methods. Twenty-four databases were searched alongside internet, hand searching and consultation with experts. Eligibility criteria were cluster randomised, randomised and quasi-randomised controlled trials (RCTs) of IDC versus usual care in prevalent adult HD patients. Primary outcome measures were exercise capacity (VO$_2$ peak and field tests) and QoL. Secondary measures were cardiac and physical function.

Results. Thirteen RCTs were eligible. Eight provided data for use in meta-analyses, which indicated no significant change in VO$_2$ peak (MD 1.19 ml/kg/min, 95% confidence interval -1.15 to 3.52, p=0.3), physical (MC 1.97, 95% CI –8.27 to 12.22, p=0.7) or mental component (MC 3.37, 95% CI -7.94 to 14.68, p=0.6) summary scores of the Medical Outcomes Short Form 36, pulse wave velocity (MD -0.57m/s, -1.55 to 0.41, p=0.4), systolic (MD -2.28mmHg, -14.46 to 9.90, p=0.7) or diastolic blood pressure (MD 2.25mmHg, -3.01 to 7.50, p=0.4) following IDC. IDC, however, leads to an improvement in performance on the six-minute walk test (MD 87.84m 39.60 to 136.09, p=0.0004). All included studies were considered to have high risk of bias.

Conclusions. There is insufficient evidence demonstrating that cycling exercise during HD improves patient outcomes. High quality, adequately powered, RCTs of IDC are required.

Keywords. ESRD; haemodialysis; exercise; systematic reviews; meta-analysis.
Introduction

The incidence and prevalence of end-stage renal disease (ESRD) requiring dialysis is increasing, with the majority of patients undertaking haemodialysis (HD). These patients have significantly increased morbidity and mortality, with cardiovascular disease the leading cause of death. Skeletal muscle catabolism, malnutrition, anaemia, uraemia, chronic inflammation, co-morbidities, physical inactivity, together with ‘enforced’ sedentary time during HD also contribute to a reduction in exercise and functional capacity which are associated with disability increased healthcare utilisation and reduced quality of life (QoL).

Exercise interventions have the potential to target several of these issues. The majority of previous reviews have examined the effects of exercise in general within the HD population (7-10). ‘Intradialytic exercise’ (an umbrella term covering a range of heterogeneous exercise interventions delivered during HD) is, however, often advocated due to greater adherence rates (11). Recent systematic reviews indicate that intradialytic exercise can significantly improve exercise capacity (12), physical QoL (12, 13) and blood pressure (13), but the interventions used include a range of different components. In choosing to review methods of exercise delivery rather than the specific type of intervention undertaken, clinicians and policy makers lack clear information on which specific modes of IDE are most beneficial, hampering the translation of research evidence into practice.

As intradialytic exercise delivered solely by means of a cycle ergometer (intradialytic cycling, IDC) is most commonly delivered within clinical practice (14) the aim of this review is to provide an up-to-date synthesis of available evidence comparing the effects of IDC.
versus usual care on exercise capacity, QoL, physical function, and cardiovascular health in HD patients.

**Materials and methods**

*Protocol registration and eligibility criteria*

A pre-specified protocol was published on PROSPERO ([www.crd.york.ac.uk/PROSPERO](http://www.crd.york.ac.uk/PROSPERO)). Eligibility criteria included cluster randomised, randomised and quasi-randomised controlled trials (RCT) of prevalent HD patients. Trials in which it was possible to elucidate the direct effects of IDC training were included; studies of acute responses to a single bout of IDC were excluded. Control groups were those who received usual care haemodialysis, defined as four hours thrice weekly, not receiving any form of exercise intervention, counselling or education. Primary outcomes measures were: exercise capacity measured by VO\textsubscript{2} peak; treadmill or field tests (e.g. 6MWT or Incremental Shuttle Walk Test (ISWT) and quality of life (QoL) using validated measures. Secondary outcome measures included: cardiac (echocardiogram, blood pressure and pulse wave velocity) and physical function determined using a range of measures (Sit to Stand 5, STS5; Sit to Stand 30, STS30; Sit to Stand 60, STS60; North Staffordshire Royal Infirmary Walk Test, NSRI; Timed Up and Go Test, TUAG).

*Study Identification*

To identify existing relevant trials and systematic reviews, the pre-specified databases, in addition to the National Institute of Health Research (NIHR) Centre for Reviews and Dissemination Database, SciELO (Scientific Electronic Library Online), OAIster (Open
Archives Initiative), SCOPUS, BASE (Bielefeld Academic Search Engine) and Open Grey, were searched from inception to March 2017, supplemented with internet searching until July 4th, 2017, hand searching reference lists and consultation with experts. No restrictions were placed upon publication, language status, or year of publication. Search terms were adapted to database requirements, the full search strategy for the MEDLINE database is shown in supplemental item S1.

Study selection, data extraction and risk of bias assessment

Search results were managed using Refworks (ProQuest, Ann Arbour, MI, USA). Duplicate citations were removed, and the remaining citations screened against eligibility criteria. Titles and abstracts deemed not to meet these criteria were excluded. For the remaining citations, full-text articles were assessed for eligibility by HMLY, DSM and MG. Disagreements were resolved by discussion with recourse to a fourth reviewer if needed.

Data from eligible studies were extracted by one reviewer and checked by another, using a template based upon the Cochrane collaboration tool. Information relating to study design, population, intervention, usual care control, outcomes and adverse events were recorded. Where multiple reports originated from a single study, comparison of the key characteristics were made using tables and were only included, as a single study, if they reported relevant outcomes to avoid overstating results. Where missing data was encountered, original authors were contacted.
Reviewers independently assessed study quality according to the Cochrane Risk of Bias assessment tool. Risk of bias for randomized controlled trials was assessed as high, low, or unclear across five domains (15). The overall risk of bias was determined using the following criteria: 1. Low risk of bias (all criteria graded adequate), 2. Moderate risk of bias (one criterion graded inadequate or two unclear) and 3. High risk of bias (more than one criterion graded inadequate or more than two unclear).

**Statistical analysis**

Descriptive statistics were used for characteristics of included studies. Where means and SD were neither reported nor available from study authors (16, 17), they were estimated from medians and IQRs (18) or standard deviations were calculated from standard error (19, 20). Only data from the first period within cross-over trials was used to reduce the potential influence of carry-over.

All outcomes were treated as continuous data and interpreted as mean differences. Analysis was primarily based upon final values. In circumstances of baseline imbalances between groups, analysis was based upon changes from baseline (15). Statistical heterogeneity was assessed with the $I^2$ test, which describes the percentage of variation across studies above that attributed to chance (21). Heterogeneity was considered unimportant for $I^2$ values up to 40% (15). RevMan 5.3 software was used to undertake meta-analyses. Data pertaining to similar outcome measures were pooled (with participant as the unit of analysis) in a meta-analysis, using a random effects model, which assumes the observed estimates of treatment effect vary across studies due to within and between study variance (22, 23). Pre-planned
subgroup and sensitivity were not possible due to the limited number of studies available for meta-analysis and all studies being classified as high risk of bias respectively.

Results

Search results and study characteristics

Figure 1 provides a flow diagram of included studies. Searching identified 3669 following removal of duplicates. Following screening of titles, abstracts and full text articles. 13 studies were eligible for inclusion. A table detailing the excluded studies can be found in supplemental item S2. Due to inadequate reporting of group sizes, results and wide heterogeneity of measures only eight trials provided information for use in meta-analyses (17, 19, 20, 24-28).

Supplementary item S3 provides a summary the characteristics of included trials. Included trials were published between 1995 and 2016 in English. Twelve were individually randomised (17, 19, 20, 24-33), and one cluster randomised (16). Twelve used a parallel group design and one a crossover design (17). All trials measured outcomes at baseline and at the end of the intervention (16, 17, 19, 20, 24-33), two also measured outcomes at an interim time-point (27, 33). Only one study included longer follow up, one month post intervention (28). In total, 369 patients were randomised to receive IDC (n=190) or usual care (n=179) with sample sizes ranging from 18-49.

Characteristics of IDC

Full reporting of the IDC intervention was lacking in several studies, and the characteristics of IDC interventions ranged widely between studies. Five studies did not provide information
regarding the mode of IDC training used (17, 24, 26, 29, 30, 32), four provided continuous training (19, 20, 27, 31), two interval training (16, 33) and two a combination (25, 28). Mean duration of exercise training was 14 weeks (range 6-26). In all but one study, where patients exercised twice a week (16), patients were encouraged to exercise at each HD session. The mean duration of each planned bout of training was 31 minutes (range 20-45) and the majority required patients to exercise at a moderate intensity (19, 20, 24-27, 29, 33). Actual levels of concordance were difficult to ascertain, as only four studies reported adherence level (17, 25, 26, 28, 29) and only three summarised the amount of exercise achieved (17, 26, 29, 33). Where reported, mean adherence rates were 81%. Only five trials reported adverse events (16, 17, 25-27, 29) of which four reported no events (16, 17, 25, 27), and one no ‘significant complications’ (26, 29). Exercise training was provided by researchers in three studies (20, 25, 27), “physical activity experts” in one study (19) and HD staff in another (17). One study reported the intervention being provided by medical and nursing staff, but it was unclear whether these were attached to the trial or the HD unit (31). Seven did not report who was delivering the intervention (16, 24, 26, 28-30, 32, 33).

Risk of bias assessment

Risk of bias summaries for all included studies are provided in table 1. All studies were rated as high risk of bias, primarily due to insufficient reporting. Only one study provided the information required to assess bias across all domains and was judged to be of high risk in all (25).

Effect of IDC

Exercise capacity
Nine studies reported outcomes related to exercise capacity. Six studies measured VO$_2$ peak (16, 19, 26-28), however it was not possible to include VO$_2$ data from the cluster RCT in meta-analyses due to inadequate reporting (16). One study used maximum work capacity, measured in watts (33). Three studies used the 6MWT (19, 25, 31) and one the ISWT (20), both field tests of exercise capacity. Four studies including 85 participants provided VO$_2$ peak data appropriate for meta-analysis. A non-significant improvement of 1.19 ml/kg/min (95% confidence interval -1.15 to 3.52; p=0.3) figure 2A) was observed immediately following IDC compared to usual care (19, 26-28). Statistical heterogeneity was low $I^2$=10%. The study that reported maximal work capacity also reported no significant difference between IDC and usual care (33).

Two of the three studies that assessed 6MWT provided sufficient information to pool results (19, 25). These trials included 48 participants and demonstrated a statistically significant improvement of 87.84m (39.60 to 136.09, p=0.0004, figure 2B) in favour of IDC. There was no evidence of statistical heterogeneity ($I^2$=0%). The study excluded from synthesis also suggested a significant improvement in the IDC group (p<0.05) (31). Only one included study assessed ISWT (20) whereby a statistically significant 15% (p=0.03) improvement in distance walked was observed in the IDC group.

Quality of life (QoL)

The SF-36 tool reports QoL as physical (PCS) and mental component scores (MCS). (34) Two studies with a total of 52 participants reported the effect of IDC on PCS and MCS. Due to baseline imbalances between groups for one of the two included studies (25) both mental and physical component summary scores for the SF-36 were assessed as change from
baseline scores. Of these studies, only one reported mean change and standard deviation. For the remaining study (25) we used a conservative estimated correlation coefficient value of 0.5(35) to calculate SD of change scores. No statistical differences were observed between IDC and usual care for the PCS (1.97, –8.27 to 12.22, p=0.7, figure 3A) or MCS scores (3.37, -7.94 to 14.68, p=0.6, figure 3B) of the SF-36. Statistical heterogeneity was not evident for the MCS ($\gamma^2=0\%$), and was negligible for the PCS ($\gamma^2=18\%$).

Cardiac outcomes

Five studies reported cardiovascular outcomes, of these three measured resting systolic (SBP) and diastolic blood pressure (DBP)(17, 20, 25), two measured pulse wave velocity(17, 25). Three studies recorded findings from echocardiogram (20, 29, 30).

One trial was excluded from meta-analysis for systolic blood pressure (SBP) due to a large baseline difference of 21.4mmHg between the usual care and IDC groups (20). This study did not report any significant effect on BP results. Synthesised data from the remaining two trials (17, 25), including 50 participants revealed a non-significant reduction in SBP of 2.28mmHg with IDC (-14.46 to 9.90, p=0.7, figure 4A). Synthesised data from three trials including 67 participants demonstrated a non-significant increase of 2.25mmHg in diastolic blood pressure with IDC (-3.01 to 7.50, p=0.4, figure 4B). There was no evidence of statistical heterogeneity for the SBP ($\gamma^2=0\%$) and negligible heterogeneity for the DBP ($\gamma^2=2\%$).

Synthesised data from two trials measuring pulse wave velocity (17, 17, 25) including 50 participants demonstrated a non-significant improvement of -0.57 m/s (-1.55 to 0.41, p=0.4,
figure 4C in the IDC group. There was no evidence of statistical heterogeneity ($I^2=0\%$). Of
the three trials reporting echocardiography measures, two trials, including 37 participants
provided results for left ventricular mass index (LVMI). However, we deemed meta-analysis
inappropriate due to a large, clinically relevant baseline difference between the control and
IDC groups of 28.3g/m$^2$ in one study (20) and 13.9g/m$^2$ in the other (29). Neither reported
any significant difference in LVMI following IDC.

Two studies including 62 patients measured left ventricular ejection fraction (LVEF). These
were not included in a meta-analysis due to inadequate reporting of group size in one study
(30). This study saw a significant improvement in LVEF in the IDC group (p=0.004)(30)
whilst the other reported no significant change (29).

**Physical function**

Three studies reported physical function outcomes. The clinical diversity in outcome
measures prevented pooling of results, but inspection of individual studies revealed limited
impact of IDC across a range of measures. A single study of 29 participants revealed no
statistically significant change in TUAG scores (-0.3 seconds, -1.39 to 0.77, p=0.9) between
IDC and usual care (25). Similarly, a single trial of 22 participants that reported a number of
function measures observed non-statistically significant improvements in STS5 (-0.57
seconds, -1.85 to 0.71, p=0.4), STS30 (0.59 stands, -2.76 to 3.94, p=0.7), STS60 (0.50 stands,
-5.67 to 6.67, p=0.9), normal gait speed (0.43 seconds, -0.67 to 1.53, p=0.5) and fast gait
speed (0.08 seconds, -0.41 to 0.57, p=0.8) (24). A baseline difference of 31.98 seconds
between the IDC and control group was noted for the NSRI walk test, but analysis of mean
change scores revealed a statistically significant improvement in the NSRI walk test (-10.2
seconds, -17.6 to -2.8, p=0.007) following IDC (24). Finally, a single trial of 21 participants reported a significant improvement in the number of steps completed in six minutes in the IDC group compared with usual care (32), however a clinically relevant baseline imbalance of 15 steps between groups may have influenced these results.

Discussion

Main Findings

The results of this systematic review and meta-analysis suggest that IDC can lead to statistically and clinically significantly improvements in exercise capacity measured via field testing, but current evidence demonstrates no statistically significant effect upon VO$_2$ peak, QoL, blood pressure or arterial stiffness. Due to issues with trial reporting, substantial baseline imbalances in outcome measures between intervention groups and inconsistency in selection of outcome measures across studies, it was not possible to synthesise the effect of IDC on cardiac or physical function. The quality of the included evidence was considered to be at high risk of bias. Based upon this review, there is currently insufficient evidence to support the use of IDC in clinical practice to improve exercise capacity, QoL, cardiac or physical function.

Comparison to previous reviews

Our review, which focussed specifically upon intradialytic aerobic exercise, delivered using a static cycle ergometer, stands predominantly in contrast to previous publications. These reviews have shown significant improvements in exercise capacity (12, 13) physical QoL (12) and blood pressure (13). These differences may be explained by the inclusion of additional evidence in our data synthesis, but also by the decision of previous reviews to include
intradialytic exercise programmes consisting of a range of types of exercise within their
meta-analyses and interventions that were not exclusively aerobic training(36, 37) or
delivered to HD patients (38) within subgroup analyses investigating types of intradialytic
exercise modality (13).

**Implications for practice**

Current evidence suggests patients, on average, observe a clinically important improvement
of 87m in the 6MWT after a programme of IDC, although the strength of this conclusion is
limited by the high risk of bias within the two included studies (39). Theoretically an
improvement in field tests of exercise capacity might reasonably be expected to be reflected
in improvement in VO$_2$ peak, given that aerobic capacity contributes towards overall exercise
capacity. The contrasting results seen in these two outcomes may be due to the inclusion of
greater numbers of different trials within the meta-analysis for VO$_2$ peak in comparison to
that for the 6MWT, other factors (e.g. muscular fatigue, anaerobic threshold, motivation)
known to influence field test performance(5) and the use of different protocols to measure
exercise capacity, which can influence the measurement of VO$_2$ peak, and its sensitivity to
change following a programme of exercise(40). The high prevalence of autonomic
dysfunction in many HD patients may also be another reason for the discrepancy between the
results of the two outcomes (41). Autonomic dysfunction can lead to cardiac
unresponsiveness which is in turn associated with poor physical performance. Whilst
speculative, it is possible that this may play a more important role during a higher intensity
exercise test and therefore, patients may achieve greater improvements in field tests such as
the 6MWT, when compared with VO$_2$ peak. Although, reduced exercise capacity and
sedentary behaviour are powerful predictors of mortality in end-stage renal disease (4, 42,
these results should not be extrapolated, as a causal link between increased exercise capacity through exercise training and decreased mortality has yet to be established. IDC appeared also to have little influence upon cardiovascular outcomes within the current review, although this may reflect a lack of high quality RCTs rather than evidence of ineffectiveness.

Previous reviews have reported no significant adverse events (9, 12, 13) due to exercise training, citing these as evidence of safety. In the current review, eight trials failed to report information on adverse events and those which did only provided limited information. Therefore, it is difficult to make a clear judgement about the safety of IDC. A recent study of IDC noted a significant drop in blood pressure one hour post exercise, with no reported adverse events (44). Given the association between asymptomatic intradialytic hypotension and adverse outcomes (45, 46), further research is needed to confirm that IDC is not associated with subclinical adverse events or to provide clinicians with information about groups of patients for whom IDE is not appropriate.

Exercise is beneficial across a spectrum of chronic diseases(47) and current Kidney Disease Outcomes Quality Initiative guidance recommends patients with chronic kidney disease CKD undertake “an exercise program compatible with cardiovascular health and tolerance, aiming for at least 30 minutes 5 times per week” (48). Expert statements provide more specific guidance (49, 50) and systematic reviews have extolled the benefits of exercise for HD patients (7-10, 12, 13). These have all led to increasing calls for exercise, particularly intradialytic, to become a routine part of the care of HD patients (13, 51, 52). The results of this review suggest, however, that the specific effects of IDC are currently unknown,
primarily due to the methodological shortcomings of existing trials. Further high quality, well-reported, RCTs are required. Two such trials, PEDAL (NCT02222402) and CYCLE-HD (ISRCTN11299707), will imminently provide evidence of the ability of intradialytic exercise to influence cardiac function and quality of life, which may further inform best practice.

Implications for research

Table 2 outlines several implications for future research which have been highlighted by this review. Unclear reporting was a feature of most IDC trials, and the primary reason for high risk of bias. Evidence presented by unclear trials rarely leads to change in practice or advances in research, because clinicians and policy makers lack confidence in the validity of findings and interventions can rarely be replicated (53, 54). Whilst we acknowledge that several IDC trials included within this review precede the publication of this guidance, future trials should adhere to both CONSORT guidance for the reporting of non-pharmacologic trials and use the Template for Intervention Description and Replication (TIDieR) checklist for intervention reporting (54, 55).

Most trials were small, and not powered to detect a true effect (56). Studies reporting echocardiography and harms are particularly vulnerable to lack of statistical power (57-60). Lack of blinding, which can influence participant, intervention provider and outcome assessor behaviour, leading to over-estimation of effect(56, 61, 62) particularly within non-pharmacologic trials (63) and ‘per protocol’ analysis, which may bias the estimated effects, were also common (61, 64, 65). Future trials should endeavour to adequately power studies, and aim to report blinding methods and fidelity explicitly (56).
Inconsistent use of a wide range of outcome measures limited meta-analyses, highlighting the need for a core outcome set to be measured and reported in all trials, alongside outcomes relevant to the individual study. Standardised Outcomes in Nephrology (SONG) aims to develop a core set of validated outcomes that reflect the main concerns of key stakeholders, including patients (66). Thirty-four priorities areas have been identified, approximately 21 of which may potentially be influenced by exercise interventions. Once standardised outcome measures have been established, better synthesis of studies will be possible, allowing comparisons of different interventions across outcomes (66, 67).

Limitations

Due to inadequate reporting two potentially relevant studies were excluded (38, 68). Additionally, six included trials presented outcome data incompletely (16, 20, 27, 31-33). Attempts to obtain these data were unsuccessful. The inclusion of these may have provided additional information for meta-analyses, potentially providing larger sample sizes and greater statistical power. The limited number of eligible studies meant it was not possible to assess publication bias or conduct subgroup analyses. Further analyses of specific characteristics of the IDC intervention (e.g. duration of programme, adherence, intensity of exercise) would have provided information on potential reasons for differences in outcomes. Despite the lack of funnel plots, risk of publication bias is minimal, as many included studies reported statistically non-significant results. Unpublished studies with statistically non-significant results may exist, but their addition is not likely to change our conclusions.

Conclusions
The renal community remains in a position of equipoise regarding IDC, because of a lack of high quality RCT data. This review highlights the need for adequately powered trials that adhere to published reporting guidance and, as far as possible, take steps to remedy the methodological limitations of trials that have gone before.

Acknowledgments

We thank Keith Nockles (University of Leicester Library) for his assistance with the development of search strategies. Finally, we wish to thank Dr Joao Viana (University Institute of Maia - ISMAI, Portugal) and Yan Song (University of Leicester) for their assistance with the translation of non-English language abstracts and full-texts.

Conflicts of Interest Statement

All authors declare that they have no conflicts of interest. The results presented in this paper have not been published previously in whole or part, except in abstract format.

Contributions: Conception; HMLY, SJS, ACS, JOB. Designed the protocol; HMLY. Assisted with protocol design; DSM, MGB, AJ, FC, SJS, CB. Co-ordinated the review process; HMLY. Searched for trials; HMLY, DSM, MGB, DC, CG, PH. Screened search results; HMLY, DSM, MGB, DC, CG, PH, JOB. Assessed the trials for quality; HMLY, DSM, MGB, SJS, JOB. Extracted data; HMLY, DSM, MGB. Analysed data; HMLY, AJ, FC, CB. Manuscript preparation; HMLY. Reviewed final manuscript; all.
Each author contributed important intellectual content during manuscript drafting or revision and accepts accountability for the overall work ensuring that questions pertaining to the accuracy or integrity of any portion of the work are appropriately investigated and resolved.

**Funding**

This report is independent research supported by the National Institute for Health Research Leicester Biomedical Research Centre. The views expressed are those of the author(s) and not necessarily those of the NHS, the National Institute for Health Research Leicester BRC or the Department of Health. This research was part funded by the Stoneygate Trust. Hannah Young and Dr James Burton are supported by grants from the National Institute of Health Research (NIHR; Grant award number: DRF-2016-09-015 and CS-2013-13-014).

The funders had no role in study design; collection, analysis and interpretation of data; writing the report; and the decision to submit the report for publication.

**References**


10. Cochrane handbook for systematic reviews of interventions version 5.1.0. [Internet]; 2011 []. Available from: http://handbook-5-1.cochrane.org/.


52. Bennett PN, Capdarest-Arest N, Parker K. The physical deterioration of dialysis patients—Ignored, ill-reported, and ill-treated. Seminars in Dialysis; Wiley Online Library; 2017.


Comparison of interstudy reproducibility of cardiovascular magnetic resonance with two-  
dimensional echocardiography in normal subjects and in patients with heart failure or left  

mass in hemodialysis patients relative to magnetic resonance imaging.  

580 59. Ioannidis JP. Adverse events in randomized trials: neglected, restricted, distorted, and  

582 60. Tsang R, Colley L, Lynd LD. Inadequate statistical power to detect clinically significant  

585 61. Friedman LM, Furberg C, DeMets DL, Reboussin DM, Granger CB. Fundamentals of  

587 62. Nüesch E, Reichenbach S, Trelle S, et al. The importance of allocation concealment and  
patient blinding in osteoarthritis trials: A meta-epidemiologic study. Arth Care Res.  

590 63. Hróbjartsson A, Emanuelsson F, Skou Thomsen AS, Hilden J, Brorson S. Bias due to  
lack of patient blinding in clinical trials. A systematic review of trials randomizing patients to  

593 64. Nuesch E, Trelle S, Reichenbach S, et al. The effects of excluding patients from the  

595 65. Abraha I, Cherubini A, Cozzolino F, et al. Deviation from intention to treat analysis in  
randomised trials and treatment effect estimates: Meta-epidemiological study. BMJ.  
2015;350(h2445).

598 66. Standardised Outcomes in Nephrology - Haemodialysis.  


603 68. Akiba T, Matsui N, Shinhoara S, Fujiwara H, Nomura T, Marumo F. Effects of  
recombinant human erythropoietin and exercise training on exercise capacity in hemodialysis  

reporting of harms in randomized trials: an extension of the CONSORT statement. Ann  
<table>
<thead>
<tr>
<th>Study</th>
<th>Random sequence generation</th>
<th>Allocation concealment</th>
<th>Blinding of participants</th>
<th>Blinding of personnel</th>
<th>Blinding of outcome assessment</th>
<th>Incomplete outcome data</th>
<th>Reporting bias</th>
<th>Other bias</th>
<th>Overall risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carmack 1995(28)</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Low risk</td>
<td>High risk</td>
</tr>
<tr>
<td>DeLima 2013(32)</td>
<td>High risk</td>
<td>Low risk</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Low risk</td>
<td>High risk</td>
<td>High risk</td>
<td>High risk</td>
<td>High risk</td>
</tr>
<tr>
<td>Giannaki 2013(16, 16, 24)</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
<td>High risk</td>
<td>Low risk</td>
<td>Unclear</td>
<td>Low risk</td>
<td>High risk</td>
<td>High risk</td>
</tr>
<tr>
<td>Groussard 2015(19)</td>
<td>Unclear</td>
<td>Unclear</td>
<td>High risk</td>
<td>High risk</td>
<td>High risk</td>
<td>Low risk</td>
<td>Unclear</td>
<td>Low risk</td>
<td>High risk</td>
</tr>
<tr>
<td>Koh 2010(25)</td>
<td>Low risk</td>
<td>Low risk</td>
<td>High risk</td>
<td>High risk</td>
<td>High risk</td>
<td>High risk</td>
<td>High risk</td>
<td>High risk</td>
<td>High Risk</td>
</tr>
<tr>
<td>Liao 2016(31)</td>
<td>Unclear</td>
<td>Unclear</td>
<td>High risk</td>
<td>High risk</td>
<td>Low risk</td>
<td>High risk</td>
<td>Low risk</td>
<td>High risk</td>
<td>High risk</td>
</tr>
<tr>
<td>Momeni 2014(30)</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Low risk</td>
<td>High risk</td>
<td>Low risk</td>
<td>High risk</td>
<td>High risk</td>
</tr>
<tr>
<td>Moug 2004(16)</td>
<td>Unclear</td>
<td>High risk</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Low risk</td>
<td>High risk</td>
<td>High risk</td>
<td>High risk</td>
</tr>
<tr>
<td>Painter 2002(27)</td>
<td>Low risk</td>
<td>Unclear</td>
<td>Unclear</td>
<td>High risk</td>
<td>Unclear</td>
<td>High risk</td>
<td>Low risk</td>
<td>High risk</td>
<td>High risk</td>
</tr>
<tr>
<td>Parsons 2004(33)</td>
<td>Low risk</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
<td>High risk</td>
<td>Unclear</td>
<td>High risk</td>
<td>High risk</td>
<td>High risk</td>
</tr>
<tr>
<td>Reboredo 2010 and 2011(26, 29, 29)</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
<td>High risk</td>
<td>Unclear</td>
<td>High risk</td>
<td>High risk</td>
<td>High risk</td>
</tr>
<tr>
<td>Toussaint 2008(17)</td>
<td>Unclear</td>
<td>Low risk</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Low risk</td>
<td>Unclear</td>
<td>Low risk</td>
<td>Low risk</td>
<td>High risk</td>
</tr>
<tr>
<td>Wilund 2010(20)</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Low risk</td>
<td>High risk</td>
<td>High risk</td>
<td>High risk</td>
<td>High risk</td>
</tr>
</tbody>
</table>

Table 1. Risk of bias for included studies.
<table>
<thead>
<tr>
<th>Aim</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improved reporting of trials</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Methodological limitation</th>
<th>Rationale for inclusion</th>
<th>Potential strategies to address limitation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poor reporting of delivery of intervention</td>
<td>Seven studies did not report intervention providers. Four studies reported adherence levels and three summarised amount of exercise data. Five studies did not provide any information regarding the type of IDC training used</td>
<td>- Adhere to CONSORT guidance for the reporting of non-pharmacologic trials (54, 55); - Use Template for Intervention Description and Replication (TIDieR) checklist for intervention reporting.</td>
</tr>
<tr>
<td>Poor reporting of study design and trial procedures</td>
<td>All studies were rated as high risk of bias, primarily due to lack of sufficient reporting. Nine trials provided insufficient details and allocation concealment.</td>
<td>- Adhere to CONSORT guidance for the reporting of NPT (53)</td>
</tr>
<tr>
<td>Poor reporting of testing procedures in relation to field testing and patient reported outcomes</td>
<td>Two studies provided no information on the procedures used for testing of field tests. Of those reporting, use of a familiarisation test was not reported in one study, and another did not report if tests were conducted on non-HD or HD days or were standardised for follow-up.</td>
<td>Report: - Timing of testing in relation to HD treatment; - Familiarisation procedures; - Procedures for collecting patient reported measures; - Explicit reporting of standardisation of these procedures for follow-up.</td>
</tr>
<tr>
<td>Poor reporting of adverse events</td>
<td>Five trials reported whether any adverse events had occurred. Statements about adverse events lacked detail.</td>
<td>- Adhere to CONSORT reporting guidance, including specific guidance for reporting harms (69); - Use of standardized and validated measurement instruments for adverse events, where possible.</td>
</tr>
<tr>
<td>Reporting bias in relation to study results</td>
<td>Selective outcome reporting in five of studies due to incomplete reporting of outcomes or time-points, reporting of only statistically significant outcomes. Only one study had published a protocol paper.</td>
<td>- Registration of trials; - Publication of protocol paper; - Presentation of all data in numerical form, not solely figures; - Adherence to Consort for NPT (54); - Raw data freely available in repository.</td>
</tr>
<tr>
<td>Attrition bias</td>
<td>Of the ten studies providing information about</td>
<td>Reduce dropout:</td>
</tr>
</tbody>
</table>
attrition, a mean of 23% (range 6-56%) of participants were lost to follow-up in the IDC group, and 14%, (range 0-56%) in usual care. Only one study reported the use of an intention to treat analysis for their primary outcome measure.

| Small sample sizes and underpowered studies | Only three studies based their sample sizes on an a priori power calculation for their primary outcome. One was unable to recruit the required number of patients. Trials of exercise and rehabilitation can be challenging to deliver and the numbers of patients required to adequately power some outcomes may be large depending on the chosen primary outcome measure. | - Minimise number of visits or integrate into usual care;  
- Allow step-wise withdrawal if unavoidable;  
- Engage a patient involvement group to provide patient perspective on trial procedures;  
- Judicious selection of outcome measures;  
- Feasibility studies prior to full scale trial. Undertake an intention to treat analysis. Explicitly state the methods used to address missing data.  
- A priori power calculation for primary outcome measure for RCTs or report as a feasibility or pilot study, with appropriate outcomes, limitations and conclusions;  
- Collaborative working to increase sample sizes and facilitate multi-centre working (key for some outcomes e.g., echocardiogram). |
| Lack of blinding of participants and intervention providers | Three studies reported that they were unable to blind participants and personnel, others do not report blinding. | - Report blinding explicitly including methods and the fidelity;  
- Where blinding not possible, strategies to reduce bias should be reported; |
| Lack of blinding of outcome assessors | Blinding of outcome assessment was reported incompletely in only five studies, all of which were high risk due to explicit lack of blinding or researchers assisting participants to complete QoL questionnaires. | - Aim to use blinded outcome assessors where possible;  
- Report blinding explicitly including methods and the fidelity;  
- Where blinding not possible, strategies to reduce bias should be reported. |
| Baseline imbalances | Five studies demonstrated large, potentially | - Adequate sample sizes and randomisation |
clinically significant differences between control and IDC groups for baseline LVMI, SBP, the SF-36 and some measures of physical function, although all authors state that these differences were not statistically significant.

| Enhanced synthesis of trial results | Wide heterogeneity of outcome measures | Limited number of outcomes appropriate to combine in meta-analyses. | - Use of core outcomes as identified by SONG-HD in addition to outcomes pertinent to the aims of the specific trial (66);
- Outcomes selected are appropriate to trial design;
- Validation/ reliability studies of these functional measures for the population. |
| --- | --- | --- | --- |
| Discuss clinical significance of results. | Individual trials report statistically significant results that seem unlikely to translate to a meaningful benefit for patients. | - Report confidence intervals and effect sizes;
- Comment on the clinical relevance of findings within the results alongside statistical significance;
- Use MCIDs from other chronic disease populations where ones for HD do not exist. |

Table 2: Recommendations for enhancing the reporting, quality and synthesis of clinical trials of IDC. (CONSORT, CONsolidated Standards of Reporting Trials; HD, haemodialysis; IDC, intradialytic cycling; LVMI, left ventricular mass index; MCID, minimum clinically important difference; NPT, non-pharmacological trials; QoL, quality of life; RCT, randomised controlled trial; SBP, systolic blood pressure; SF-36, Medical Outcomes Short Form 36; SONG-HD, Standardised Outcomes in Nephrology- haemodialysis; TIDieR, Template for intervention description and replication.).
Legends to figures

Figure 1: Flow diagram of study selection. *: some studies excluded for multiple reasons. (HD, haemodialysis; IDC, intradialytic cycling, RCT, randomised controlled trial).

Figure 2: Forest plot comparing IDC with usual care on VO2 peak (A) and the six-minute walk test (B).

Figure 3: Forest plots comparing change in physical (A) and mental component summary scores of the SF-36 (B) in IDC and usual care.

Figure 4: Forest plot comparing IDC with usual care on systolic (A), diastolic blood pressure (B) and pulse wave velocity (C).