Interventions to improve physical activity during pregnancy: A systematic review on issues of internal and external validity using the RE-AIM framework

Author names and affiliations:

Dr Melinda Craike, Ph.D, Institute of Sport, Exercise and Active Living (ISEAL), College of Sport and Exercise Science, Victoria University, Melbourne, Australia; Deakin University, Geelong, Australia, School of Psychology, Centre for Social and Early Emotional Development.

Dr Briony Hill, Ph.D, Deakin University, Geelong, Australia, School of Psychology, Centre for Social and Early Emotional Development.

Dr Cadeyrn J Gaskin, Ph.D, Deakin University, Geelong, Australia, School of Psychology, Centre for Social and Early Emotional Development; Victoria University, Melbourne, Australia.

Professor Helen Skouteris, Ph.D, Deakin University, Geelong, Australia, School of Psychology, Centre for Social and Early Emotional Development.

Corresponding author:

Dr Melinda Craike, College of Sport and Exercise Science, Victoria University, PO Box 14428, Melbourne, Victoria, Australia, 8001., Phone +61 3 9919 5659: Email: Melinda.craike@vu.edu.au
Abstract

Background: Physical activity during pregnancy has significant health benefits for the mother and her child, however, many women reduce their activity levels during pregnancy and most are not sufficiently active. Given the important health benefits of PA during pregnancy, evidence that supports research translation is vital.

Objectives: To determine the extent to which physical activity interventions for pregnant women report on internal and external validity factors using the RE-AIM framework (reach, efficacy/effectiveness, adoption, implementation, and maintenance).

Search Strategy: Ten databases were searched up to 1 June 2015. Eligible published papers and unpublished/grey literature were identified using relevant search terms.

Selection Criteria: Studies had to report on physical activity interventions during pregnancy, including measures of physical activity during pregnancy at baseline and at least one point post intervention. Randomised controlled trials and quasi-experimental studies that had a comparator group were included.

Data Collection and Analysis: Reporting of RE-AIM dimensions were summarised and synthesised across studies.

Main Results: The reach (72.1%) and efficacy/effectiveness (71.8%) dimensions were commonly reported, however, the implementation (28.9%) and adoption (23.2%) dimensions were less commonly reported and no studies reported on maintenance.

Conclusions: This review highlights the under reporting of issues of contextual factors in studies of physical activity during pregnancy. The translation of physical activity interventions during pregnancy could be improved through reporting of representativeness of participants, clearer reporting of outcomes, more detail on the setting and staff who deliver interventions, costing of interventions and the inclusion of process evaluations and qualitative data.
PROSPERO registration number: CRD42015019801

Tweetable Abstract: The systematic review highlights the under reporting of contextual factors in studies of physical activity during pregnancy.

Key words: Physical activity, pregnancy, internal/external validity, translation, intervention, RE-AIM
Introduction

Physical activity (PA) during pregnancy has substantial physical and psychological health benefits for the mother and her child, including reduced risk of developing gestational diabetes, reduced incidence and severity of prenatal and postnatal depressive symptoms and the normalisation of birth weight. Recommended levels of PA during pregnancy are similar to the broader guidelines for PA in healthy adults. Despite the benefits of PA and recommendations that women continue to be active during pregnancy, many women reduce their activity levels during pregnancy and most are not sufficiently active. For example, a study in the US found that less than 20% of pregnant women were meeting the recommended levels of PA. Pregnancy represents an opportunity to promote PA as women receive close medical attention and are often highly motivated to improve their health to benefit their children.

To date, reviews of PA during pregnancy have focused on the efficacy of PA interventions on health and pregnancy outcomes, weight gain during pregnancy, and changing PA and diet to limit gestational weight gain. Only two systematic reviews have specifically examined the outcomes of interventions in terms of increasing PA during pregnancy. Currie et al. reported that PA interventions incorporating behaviour change techniques help reduce the decline in PA throughout pregnancy. Pearce et al. found that few of the PA interventions reviewed improved PA participation during pregnancy. Both of these reviews only included randomised controlled trials (RCTs) and did not provide a thorough assessment of external validity of the included studies.
Assessing external validity when evaluating physical activity interventions in pregnancy is important because it enables conclusions to be reached about the generalizability of study findings to the broader population of pregnant women. Such conclusions can facilitate decision-making about what interventions are likely to be successfully implemented in what settings (e.g., clinical or community settings) and with what sub-populations of pregnant women. If external validity is low or unknown, then changes to the evaluations or the reporting of these evaluations, respectively, may be required.27.

The primary aim of this systematic review was to determine the extent to which PA interventions for pregnant women report on internal and external validity factors using the RE-AIM framework. A secondary aim was to examine whether there were differences in reporting of the RE-AIM dimensions and individual indicators by study design (RCTs compared to quasi-experimental designs).

Methods

Literature search

This review follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. The systematic review protocol was registered with the International Prospective Register of Systematic Reviews (PROSPERO) on 14 May 2015 and updated on 3 November 2015.

Both published papers and unpublished/grey literature (abstracts were excluded) were eligible for inclusion in this review. We searched electronic databases, scanned the reference lists of included studies and relevant reviews, and wrote to the authors of included papers to
determine if they had conducted any additional research that we had not identified or had any further information relating to their published studies.

Individualised search strategies for each database were developed in collaboration with a professional librarian. Searches were performed up to 1 June 2015. We included ten databases in our main search strategy; Health Policy Reference Center via EBSCO; PsycINFO via EBSCO; Medline Complete via EBSCO; CINAHL Complete via EBSCO; Informit: Health Subset; Embase; Scopus; SportDiscus with full text via EBSCO; Global Health via EBSCO; Academic Search Complete via EBSCO. An example search strategy is included in Supplementary Table 1 (Table S1). We also conducted supplementary searches (e.g., Google advanced search, Greylit, The Grey Literature Report) to identify additional literature, using similar search terms to our main search.

MC and BH conducted the screening and study selection process. Initially, MC and BH independently screened 605 records against the inclusion criteria (Cohen's kappa = 0.875; very good inter-rater agreement). BH and MC then screened the remaining entries. We obtained full reports for all titles that appeared to meet the inclusion criteria or where there was any uncertainty. MC and BH, independently of each other, conducted a pilot full text screen of 50 reports using the inclusion criteria (Cohen's kappa = 0.901; very good inter-rater agreement). MC and BH then screened the remaining full-text of reports. For the intervention studies that met the inclusion criteria for our review, qualitative studies, process evaluations, cost or other information relating to the intervention were included as companion papers.
**Inclusion criteria**

Studies were selected according to the following inclusion criteria: **Participants**: Women during pregnancy; **Intervention**: PA interventions, including where PA was the sole focus or part of interventions with multiple lifestyle factors. We excluded studies that included only pelvic floor exercises; **Comparators**: Comparison of alternative interventions, usual care, or no intervention; **Outcomes**: PA measured during pregnancy at baseline and at least one point post intervention initiation (but still during pregnancy); **Study Designs**: Randomised controlled trials and quasi-experimental that had a comparator including alternative interventions or usual care.

**Data Extraction**

**Study Characteristics**

We extracted the following data from each study: country, study design, control condition, number of participants and demographic characteristics, PA intervention description, measure of PA and timing of assessments, study results, other outcomes assessed.

**RE-AIM evaluation**

Incorporating aspects of both internal and external validity, the RE-AIM framework (reach, efficacy/effectiveness, adoption, implementation and maintenance) was designed to focus attention on aspects of interventions that can improve the translation of strategies into practice \(^{28,29}\). The RE-AIM framework has been applied to reviews of PA interventions for a range of population groups, including breast cancer survivors \(^{30}\), Latin Americans \(^{31}\), and people with type 2 diabetes \(^{32}\). The RE-AIM framework as also been applied to health interventions during pregnancy, including an evaluation of the Alcohol and Pregnancy Project\(^{35}\) and a comparison of two approaches to promoting smoking abstinence in pregnant and postpartum women\(^{36}\).
RE-AIM dimensions include reach, efficacy/effectiveness, adoption, implementation and maintenance. Reach is a measure of individual-level participation, including the proportion of the population targeted that are affected by the intervention, as well as the representativeness of the participants to the target population. Efficacy/effectiveness measures include the effectiveness of the intervention, positive or negative consequences of the intervention, as well as behavioural, quality of life, and participant satisfaction outcomes. Adoption is concerned staff and settings, the proportion of existing or available settings that offer the intervention and how representative these settings are of the community as a whole. Implementation refers to the degree to which the intervention is delivered as intended, and evaluated based on the faithfulness of the program administrators to the design of the intervention. Finally, maintenance includes an individual and an institutional level component about the sustainability of the intervention.

We used the 21-item data collection tool that has been used in several previous systematic reviews which have reported the RE-AIM dimensions. This tool was the basis for calculating percentages of studies meeting criteria for the five RE-AIM dimensions (reach, efficacy/effectiveness, adoption, implementation, and maintenance). Of note, one of the maintenance indicators relates to maintenance of individual behaviour 6 months post intervention completion; given that we were interested in PA during pregnancy, this indicator was deemed not applicable and was not included in our reporting. In addition to the 21-items, we also included assessments of eight additional indicators that were applied by Galaviz et al. to provide a more comprehensive assessment of the RE-AIM framework.

We developed a coding manual, based on the coding manual used by Blackman et al. (personal communication, 2015), and MC and BH piloted this on eight studies. We discussed
any areas where discrepancies in consistency of data extraction arose. MC and BH coded the remaining studies and if there were any uncertainties, these were discussed; resolution was by consensus with direct reference to the research article.

We summarised RE-AIM criteria using means and frequencies. First, the average proportion of indicators reported within each RE-AIM dimension was computed (i.e., number of indicators reported for a given dimension divided by the total number of possible indicators within the dimension). Second, the proportion of studies that reported specific indicators within each RE-AIM dimension were computed (i.e., number of studies that reported divided by total number of studies).

Results

Figure 1 presents the flow of studies included in this review; 52 documents (representing 38 studies) met inclusion criteria. Note that Dodd et al.\textsuperscript{39} included a nested trial; this was assessed as a separate study and therefore the paper by Dodd et al.\textsuperscript{39} was counted as two studies for the purpose of this review. To simplify reporting and distinguish between these studies, we used the citation for the thesis on which the nested study was based.\textsuperscript{40} Ten companion documents (which did not meet the inclusion criteria) were also identified and provided additional information for the data extraction.

Study characteristics

Study characteristics are summarised in Supplementary Table 2 (Table S2). For ease of reading throughout the Results section, for studies with more than one reference we have referenced each study using the first reference in Table S2. Of the 52 documents retrieved, 42
were peer-reviewed publications and 10 were theses. Of the 38 studies identified, 33 were RCTs and the remaining five were quasi-experimental trials. Each of the studies included in this review reported several outcomes in addition to PA. The most common outcomes were behaviour change relating to diet, gestational weight gain, improving outcomes for women who were obese/or had gestational diabetes mellitus (GDM), or the prevention of GDM. Most interventions were unsupervised. Six studies reported both objective and self-report measures of PA, two studies used only objective measures of PA, and the remaining 30 studies used self-report only. A number of different self-report measures were used; the most commonly used was the Pregnancy Physical Activity Questionnaire (PPAQ, n = 10).

**RE-AIM evaluation**

On average, reach (72.1%) and efficacy/effectiveness (71.8%) were the most highly reported RE-AIM dimensions, fewer reported implementation (28.9%) or adoption (23.2%) and no studies reported maintenance indicators. Table 1 shows a detailed breakdown of the individual RE-AIM indicators, total and by study design.

Table 1 here

**Reach**

On average, 3.6 (72.1%) of the 5 reach indicators were reported across the 38 studies; these indicators were more likely to be reported in RCTs than quasi-experimental studies (73.9% compared to 60.0%). Method to identify target population (n = 36, 94.7%), inclusion criteria (n = 35, 92.1%), exclusion criteria (n = 32, 84.2%), and participation rate (n = 25, 65.8%)
were all highly reported, however the characteristics of participants and non-participants (or other indicator of representativeness) was not highly reported (n = 9, 23.7%). The reporting of these indicators were similar across RCTs and quasi-experimental studies, except for inclusion criteria, which was more likely to be reported in RCTs (100% compared to 40%).

The number of participants in the studies ranged from 1542 to 221239 (median = 151); the median participation rate was 58.8%. An indication of the representativeness of the sample was reported by nine studies (23.7%). Three interventions presented characteristics of non-participants (i.e., those who refused participation)51, 53, 60, four compared study participants to broader populations45, 50, 66, 67, one compared the sample with a large cohort study74, and one study compared the sample with other research and the broader population72. Other indicators assessed included per cent of participants who were excluded (e.g., were ineligible; n = 21, 55.3% reported). No studies provided information on the cost of recruitment.

**Efficacy/effectiveness**

On average, 2.9 out of 4 (71.8%) efficacy/effectiveness indicators were reported; these were more likely to be reported in RCTs than quasi-experimental trials (73.5% compared to 60%). Each of the included studies had measures of PA at baseline and at least one follow-up during pregnancy (100%). Data on attrition (specifically in relation to PA measures) were reported in 26 studies (68.4%), fewer measured quality of life or unintended consequences (n = 23, 60.5%) or stated that they used intention-to-treat analysis (n = 22, 57.9%). RCTs were more likely than quasi-experimental studies to use intention-to-treat (n = 21, 63.6% compared to n = 1, 20%) and report on quality of life measures or unintended consequences (n = 22, 66.7% compared to n = 1, 20%); however, quasi-experimental studies were more likely to report on attrition (100% compared to 63.6%).
An improvement in PA was reported in 19 studies\(^41, 42, 46, 48, 51, 55, 58-60, 63, 65, 66, 68-72, 74, 75\). Attrition rates were examined specifically in relation to PA assessment; median attrition rates were high but similar across the intervention (21.4%) and control conditions (23.4%). Other indicators assessed included imputation procedures, which were specified in 10 studies (26.3%), and measure of PA relative to public health goal, which was reported in six studies (15.8%).

Adoption (setting and staff)

One average, 1.4 of the 6 indicators for adoption were reported (23.2%); quasi-experimental studies were slightly more likely to report these indicators than RCTs (33.3% compared to 21.7%). Level of expertise of staff (delivery agent) were reported in almost all of the studies (n = 35, 92.1%) and 12 (31.6%) provided an explicit description of the characteristics of the intervention location. However, all other aspects were reported in less than 6% of studies: description of staff who delivered the intervention (i.e., explicit behavioural/demographic characteristics of staff; n = 2, 5.3%), rate of adoption at the setting or delivery agent level (n = 2, 5.3%), inclusion/exclusion criteria of delivery agent or setting (n = 2, 5.3%), method to identify staff who delivered the intervention (0%). Each of these indicators were more commonly reported in quasi-experimental studies than RCT studies, except description of staff who delivered the intervention.

Interventions were delivered by a range of staff, including researchers and practitioners, and several interventions were delivered by several different types of staff. Common examples of staff included midwife or nurse (including research midwife/nurse, obstetric nurse, student nurse)\(^41, 43-45, 51, 64, 75\), physiotherapist\(^43, 57, 60, 67, 68, 74\) and professional exercise trainer or
exercise physiologist. Only two studies included explicit descriptions of study staff; one study described the staff as parish nurses fluent in Spanish and a second reported that staff were bilingual and bicultural health educators.

Most of the included studies named the intervention location, which included health care and community settings, however fewer (n = 12, 31.6%) provided an explicit statement of characteristics of the location of the intervention. The rate of adoption of delivery settings was reported by two studies and the inclusion criteria for the delivery agent for setting was reported in two studies.

Implementation

On average, 1 of the 3 (mean = 0.9, 28.9%) of the implementation indicators were reported; these indicators were slightly more likely to be reported in quasi-experimental studies than RCTs (33.3% compared to 28.3%). Intervention intensity (including all three elements of timing, duration and intensity) were described in 27 (71.1%) of the included studies. Fewer reported the extent to which the protocol was delivered as intended (n = 4, 10.5%) or measures of the cost of implementation (we included cost effectiveness evaluations; n = 2, 5.3%). Intervention intensity and extent the study protocol was delivered as intended was more commonly reported in quasi-experimental studies, however measures of cost were only reported in RCTs. Reporting on the extent that the protocol was delivered as intended was reported in a number of ways including participant reporting of interactions with staff, an adherence scale and staff audio diaries. Additional indicators that we assessed included participant adherence to the intervention (n = 28, 73.7%), consistency of implementation across settings and delivery agents (n = 3, 7.9%), and use of qualitative methods to understand implementation (n = 2, 5.3%).
Maintenance

Maintenance of the program was not reported in any of the studies and no studies reported on the cost of maintenance. An additional indicator was included, which was the use of qualitative data to understand setting level institutionalization; this was not reported in any study.

Discussion

Main Findings

We conducted a systematic review of PA interventions during pregnancy to identify reporting of elements relevant to internal and external validity that may inform the translation of interventions. We found that reporting was higher for aspects of internal validity, such as explicit inclusion and exclusion criteria, than issues of external validity, such as staff a description of staff who delivered the intervention or the method to identify staff who delivered intervention. The findings of our review also revealed several differences in reporting of RE-AIM dimensions between RCTs and quasi-experimental studies, however these were not substantial.

Individual level indicators, such as inclusion and exclusion criteria and participation rate, were well reported, however, the representativeness of participants was not. Other reviews of PA interventions have found low reporting of the representativeness of participants. Given that, on average, 40% of women refused participation in the included studies, examination of the representativeness of participants is important. Among the studies that reported on aspects of generalisability, some reported that characteristics of participants and non-participants were similar or representative of their study population (e.g.,); in contrast, others found demographic or behavioural differences (e.g.,). Knowing who declined to
participate in studies, and their reasons for doing so, may help in the development of targeted
and accessible interventions for these populations. Examples from this review include the
comparison of study participants to the host maternity hospital population characteristics \(^50\),
women giving birth in the district and state \(^45\), and the national pregnant population \(^67\).

Measures of PA at follow-up and attrition rates were highly reported, but intention-to-treat
analysis and measures of quality of life or negative consequences were less well reported.
Although there is strong evidence that PA is beneficial during pregnancy \(^3, 6\), reporting of
negative outcomes and adverse events is important because pregnancy is a time of
physiological changes for women \(^79\). For example, blood volume and cardiac output increase
during pregnancy, and other metabolic functions are altered to provide for the demands of the
fetus \(^79\). Nineteen studies reported positive outcomes, in terms of higher levels of PA in the
intervention compared to the control condition. Comparison of effectiveness across
interventions is difficult, however, due to high levels of attrition (on average, over 20%), and
heterogeneity in the detail of reporting of findings. Furthermore, only 15.8% reported PA
participation relative to public health recommendations during pregnancy. Reporting of the
effect of PA programs in a meaningful way, where comparisons can be made across
interventions, allows decision-makers to assess the relative effectiveness of interventions.

The cost of delivery of interventions is a key factor in determining the translation of research
findings into practice. The need to make the best use of limited resources at all levels, from
the national health service level to the local level, is imperative \(^80, 81\). Two studies included in
our review reported on the cost effectiveness of the intervention; one showed the intervention
was not cost effective \(^82\) and a second showed it was cost neutral \(^83\). Although several other
studies commented that the intervention was not resource intensive (e.g., \(^56, 61\)), they did not
report actual costs. The emphasis of research on PA has been on achieving significant outcomes, which often produce interventions that are intensive, expensive, and demanding. Low-intensity interventions that are less efficacious, but have the potential to be delivered to large numbers of women, may have a more pervasive impact and be more cost-effective than high intensity interventions that are delivered to fewer women. However, the state of the evidence for PA interventions in pregnancy does not yet allow us to draw this conclusion. The inclusion of indicators of the cost of interventions to promote PA during pregnancy are a priority.

The reporting of staff and setting level indicators was low and similar to previous reviews of PA interventions in other populations. Details of the settings where interventions are delivered and staff who deliver interventions allow an assessment of whether an intervention produces a generalised effect or whether implementation varies according to local conditions. The staff involved in research studies often have high levels of training, expertise, or supervision, or they are employed solely to deliver the intervention being evaluated rather than having multiple competing responsibilities. It is important to document the extent to which staff are willing to be involved in a study, their characteristics, and the level of training or skill required to implement the intervention.

Process evaluations, including the use of qualitative research, can assist in understanding participant level and setting and staff level indicators. Several reported participant compliance to the intervention of less than 50% and half of the intervention included in the review had no impact on PA behaviour. Poston et al. conducted a comprehensive process evaluation following Steckler and Linnan’s framework and provided important insights into intervention delivery. Such evaluations are recommended in future studies.
The findings of our systematic review revealed differences between RCTs and quasi-experimental designs in the reporting of RE-AIM dimensions but these were not substantial. Studies using RCT designs tended to more highly report individual level factors whereas studies with quasi-experimental designs tended to report contextual factors more often. There have been criticisms of RCTs for their focus on internal validity at the expense of external validity and not providing information on how results can be implemented in practice; we found that the quasi-experimental studies followed a similar pattern. Our comparison of RCTs and quasi-experimental studies must be interpreted with caution, however, due to the small number of quasi-experimental studies that we identified (n=5).

**Strengths and Limitations**

Our systematic review was novel in that it was first to assess issues of internal and external validity of physical activity interventions during pregnancy. Our review was comprehensive; we identified a greater number of studies than previous reviews (1424 and 925 studies). The review has some limitations, however. We only included studies with baseline and post intervention assessments of PA. Thus, studies that targeted gestational weight gain but did not assess PA behaviour did not meet the inclusion criteria for this review (e.g.,90). Second, our review included studies that targeted health outcomes as well as those specifically focused on PA behaviour change and therefore there was a level of heterogeneity in aims and type of intervention delivery in the included studies. Finally, the researchers of the reviewed studies may have collected some of the information required to complete a RE-AIM evaluation, but did not report this in the articles and their intention may be to publish this information in the future.
Interpretation

Our findings showed that although researchers frequently report on the internal validity of studies of PA during pregnancy, they do not report external validity as extensively \(^{28,91}\). The translation of research to enhance PA during pregnancy could be improved through the reporting of information relating to the representativeness of study populations, clearer reporting of the effectiveness of interventions, more detail of the setting and staff who deliver interventions, costing of interventions, and the inclusion of process evaluations and qualitative data.

Conclusions

Reporting of issues of external validity needs to be improved so that physical activity interventions during pregnancy can be translated into practice. The onus should fall on funding bodies, researchers, journals, and policy makers to ensure that this detail becomes standard practice when designing, conducting, and reporting findings of interventions \(^{92}\). Given the important health benefits of PA during pregnancy, evidence that supports research translation is vital.

Acknowledgements: Nil

Disclosure of Interests: Nil

Contribution to Authorship:

MC was responsible for the study design, literature search, data extraction and analysis, and drafting the article. BH had significant input in to the study design, data extraction and provided critical reviews of the content of the article. CJG and HS had input in to the study design and interpretation of data and provided critical reviews of the content of the article. All authors have approved the final version of the article to be published and all authors agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.
Details of Ethics Approval: No ethics approval was necessary for this review, as only published data were used.

Funding: Nil


35. Payne JM, France KE, Henley N, D'Antoine HA, Bartu AE, O'Leary CM, et al. RE-AIM evaluation of the Alcohol and Pregnancy Project: Educational resources to inform
health professionals about prenatal alcohol exposure and fetal alcohol spectrum disorder.


40. Sui Z. Physical Activity During Pregnancy Among Women who are Overweight or Obese. [Ph.D]: The University of Adelaide; 2013.


59. Smith KM. The Blossom Project Online: Use of a behaviorally-based website to promote physical activity and prevent excessive gestational weight gain in previously sedentary pregnant women [Ph.D.]. Ann Arbor: Iowa State University; 2014.


67. Oostdam N, van Poppel MN, Wouters MG, Eekhoff EM, Bekedam DJ, Kuchenbecker WKH, et al. No effect of the FitFor2 exercise programme on blood glucose, insulin
sensitivity, and birthweight in pregnant women who were overweight and at risk for gestational diabetes: results of a randomised controlled trial. BJOG. 2012;119(9):1098-107.


73. Nodine PM. The Contribution of Physical Activity to Sleep Parameters in Pregnancy within the Context of Prepregnant Body Mass Index [Ph.D.]: University of Colorado Health Sciences Center; 2011.


102. Avery MD. Exercise as a Therapeutic Intervention in Gestational Diabetes [Ph.D.]: University of Minnesota; 1993.


Figure 1. Flow diagram of studies included in the review