Category

건강증진 프로그램의 평가 프레임워크 사용: 체계적 문헌고찰 연구의 포괄적 고찰

권준혜[1], 김태웅[1], 이충근[1,2,*]
[1]서울대학교 사범대학 체육교육과
[2]서울대학교 스포츠과학연구소

Use of Evaluation Framework in Health Promotion Program: an Umbrella Review

Junhye Kwon[1], TaeEung Kim[1] and Chung Gun Lee[1,2,*]
[1]Department of Physical Education, College of Education, Seoul National University, Seoul, Korea
[2]Institute of Sport Science, Seoul National University, Seoul, Korea

Abstract

Objectives: The purpose of this study is to synthesize systematic reviews and meta-analyses of the use and reporting of evaluation frameworks for health promotion programs.
Methods: A systematic search was performed using three databases (PubMed, CINAHL, Scopus). Three authors independently reviewed the titles and abstracts of the identified studies, and relevant data were extracted according to the inclusion criteria. The research findings were synthesized descriptively.
Results: Ten studies were included. Seven of these studies reviewed the application of a specific framework for various health promotion programs, while the remaining three analyzed the use of different evaluation frameworks for specific health promotion programs. The RE-AIM framework was most frequently applied in intervention studies, including those related to physical activity, and was predominantly used independently.
Conclusion: Given the continued publication of studies on the issues and directions of applying the RE-AIM framework according to cumulative usage, future research should conduct a systematic evaluation based on such studies.
keywords: *Health Promotion, Program Evaluation, Evaluation Framework*

Introduction

건강 증진 프로그램의 종합적인 평가는 다양한 인구집단 및 상황에도 효과적인 적용이 가능하도록 프로그램 실행 과정에 대한 필수 지식을 제공한다[1, 2]. 평가를 수행하는 궁극적인 목적으로는 개발된 프로그램을 향후 개선, 확장, 혹은 대체할 것인지에 대한 결정[3], 프로그램 목표 실현 정도 공유, 일반화 가능성 파악, 연구 지원금 및 계약 요건의 충족 등이 있다[4, 5]. 연구의 특성에 따라 목적의 우선순위가 달라지기는 하나 건강 증진 프로그램을 효과적으로 실행하기 위해서는 프로그램에 관한 요소들을 체계적으로 평가할 필요가 있다[6]. 그러나 지금까지의 연구들을 살펴보면 건강 증진 분야의 평가는 일관성 없이 적용되어 왔으며[7], 불충분한 평가 계획으로 인해 프로그램 효과를 향상시키는 데 어려움을 겪었다[2, 8]. 프로그램 계획단계에서 평가 부분이 포함되지 않는 것은 전문지식의 부재, 시간 및 자원의 한계, 평가의 이점에 대한 무지, 평가는 전문가에 의해서만 실시되어야 한다는 믿음 혹은 평가자에게 투입되는 인건비 부족 문제가 주요 원인이라고 볼 수 있다[6]. 이러한 문제를 해결하기 위해서는 평가 영역에 대한 지속적인 개발과 관심을 통해 건강 증진 분야의 연구자들이 평가의 효과와 필요성을 자연스럽게 알아갈 수 있도록 해야 한다.

*** Corresponding author:** Gil-Dong Hong (koreajph@gmail.com, 02-123-4567)
Graduate School of Public Health, Seoul National University, 599 Gwanak-ro, Gwanak-gu, Seoul 151-742, Korea.

성공적인 건강 증진 프로그램을 위해서는 개입 요소들이 제대로 실행되었는지 혹은 어느 정도 전달되었는지 파악할 수 있는 과정 평가를 실시해야 한다[9, 10]. 특히, 프로그램 실행 후 대상 건강 행동이 유의미한 변화를 보였을 때 정확히 어떤 구성 요소가 해당 결과에 영향을 미쳤는지 알 수 있어야 한다[11]. 과정 평가는 프로그램이 실제 연구 결과에 미친 영향과 프로그램 실행 시 무슨 일이 일어났는지 자세히 들여다 볼 수 있는, 소위 말해 '블랙박스(black box)' 역할을 한다[12-14]. 그러나 어떤 개입 요소가 더 효과적이었는지 혹은 더 적게 영향을 미쳤는지 알 수 있는 '블랙박스' 내용을 들여다 본 연구는 많지 않다[13, 14]. 또한 건강 증진 개입으로 인해 효과적인 건강 행동 변화가 나타난 경우, 과정 평가는 향후 관련 연구의 개입 실행에 필요한 핵심 자료를 제공하며[15-17], 개입으로 인한 행동 변화에 효과가 없었다면 정확히 어떤 요인 때문에 개입 실행에 차질을 겪은 것인지 밝혀내는 역할을 한다[16, 18, 19]. 대부분의 건강 증진 프로그램은 하나의 개입만 실행하기보다는 다양한 개입 요소들을 활용하기 때문에 복잡한 형태로 개발되는 경우가 많다[20]. 그리고 이러한 복잡성은 프로그램의 특정 구성 요소 및 결과 요인과 관련하여 낮은 수준의 실행력을 야기할 수 있다[21]. 따라서 과정 평가는 추가적인 옵션이 아닌 건강 증진 프로그램 평가 시 반드시 거쳐야 할 절차라고 할 수 있다.

앞서 언급한 바와 같이, 건강 증진 프로그램의 모든 단계를 체계적으로 수행하기 위해서는 평가 프레임워크의 사용이 필수적이다. 그러나 연구자들은 평가 자체를 하지 않거나 과정 평가는 건너뛰고 결과 평가에만 치중하는 경향이 있으며, 기존에 제시된 여러 평가 프레임워크가 존재함에도 불구하고 이를 적절하게 적용하지 못하고 있다. 신체활동과 식습관 변화 개입 연구에서 사용한 평가 프레임워크의 범위 및 특징을 알아보기 위해 주제범위 문헌고찰(scoping review)을 실시한 연구는 총 71개의 평가 프레임워크가 사용되고 있다는 것을 파악하였다[22]. 또한 주제범위 문헌고찰에서 파악된 평가 프레임워크를 바탕으로 체계적 문헌 고찰을 실시한 결과, 신체활동 개입 연구의 평가 도구로는 총 16개의 프레임워크가 사용되고 있다는 점을 파악하였다[23]. 이처럼 건강 증진 프로그램의 평가 프레임워크 사용과 관련하여 체계적 문헌 고찰 연구가 실시되었으나[23-25], 이를 종합하여 평가 프레임워크 사용에 대한 전반적인 내용을 정리한 연구는 없다. 또한 현재까지 실시된 체계적 문헌 고찰 연구들은 한 가지 평가 프레임워크를 선택하여 해당 프레임워크가 특정 건강 증진 프로그램에 제대로 사용되었는지 평가한 연구가 대부분이다. 체계적 문헌 고찰 연구들은 특정 프레임워크 적용에 대한 자세한 정보를 제공하지만 각 프레임워크가 어떻게 적용되고 있는지 어떠한 프로그램에 어떠한 프레임워크가 적용되고 있는지 종합하여 살펴본 연구는 없었다.

체계적 문헌 고찰과 메타 분석은 편향을 최소화하기 위해 체계적인 검색, 자료 추출 및 평가 방법을 사용하여 사전에 지정한 검색 기준에 맞는 경험적 증거를 수집하는 반면[26], 체계적 문헌고찰 연구의 포괄적 고찰(umbrella review)은 체계적 문헌 고찰에서 축적된 증거들을 종합하고 주어진 주제에 대한 포괄적인 검토를 수행하여 보다 넓은 범위의 결론을 제시하는 것이 주된 목적이라고 할 수 있다[27-29]. 체계적 문헌 고찰의 수가 증가함에 따라, 다음 단계로 가장 적절한 연구 방법은 기존 체계적 문헌 고찰 연구들을 종합적으로 분석하는 것이며, 이는 사전에 검토된 결과들의 공통점 파악과 대조를 가능하게 함으로써 의사결정자에게 중요한 정보를 제공할 수 있다[30]. 이에 본 연구는 건강 증진 프로그램의 평가 프레임워크 사용에 대해 알아보고 건강 행동 및 상태별, 프로그램 특성별로 평가 프레임워크 적용에 차이가 있는지 알아보고 이를 종합하는 것이 목적이다.

Methods

1. 선정 기준

검색된 문헌들을 최종적으로 검토하기 위한 포함·배제 기준은 PICO-SD(Population/Problem, Intervention, Comparisons, Outcomes, Study Designs)를 사용하여 Table 1에 제시하였다[31]. 먼저 문제(problem)는 다양한 인구집단을 대상으로 한 건강 행동 및 상태 관련 프로그램의 평가 프레임워크 사용으로 설정하였다. 여기서 다양한 인구집단은 대학생을 포함한 일반 성인, 청소년, 직장인, 노인 등을 포함한다. 건강 행동 및 상태 관련 분야가 아닌 다른 주제에 대한 평가 프레임워크 사용을 분석한 경우, 예를 들면 의료 시

스템을 평가하기 위한 도구로 평가 프레임워크가 사용된 경우는 모두 제외하였다. 개입(intervention)의 경우, 의료적 행위가 가해지지 않은 비의료적 건강 상태 및 행동 관련 개입을 포함하였다. 결과 측정(outcomes)도 마찬가지로 비의료적 건강 상태 및 행동으로 신체활동 증진, 건강한 식습관 형성, 그리고 좌식행동 감소 등을 포함하였다. 개입이나 결과 측정 변인에 의료 행위가 포함되었더라도 신체활동, 식습관 개선, 그리고 좌식행동에 초점을 맞춘 논문이 하나라도 포함되어 있으면 선정 기준에 포함하였다. 연구 디자인(study design)으로는 건강 상태 및 행동 개입의 평가 단계를 분석한 체계적 문헌 고찰 및 메타 분석 연구를 포함하였고, 서술적 고찰(narrative review) 및 회색문헌(grey literature)은 제외하였다. 연구기간은 2005년 1월부터 2023년 2월까지 게재된 논문으로 설정하였고 영문이 아닌 다른 언어로 작성된 문헌은 제외하였다.

Table 1. PICO-SD criteria

	Inclusion	Exclusion
Problem	Applying evaluation frameworks for health promotion programs targeting diverse population	Applying evaluation framework for programs not related to health behavior or status (e.g., assessing healthcare system)
Intervention	Non-clinical health behavior or status related intervention (e.g., exercise, education, providing information, etc)	Clinical intervention
Comparison	-	-
Outcomes	Health behavior or status (e.g., physical activity participation, weight, diet, sedentary behavior)	Clinical behavior or status (e.g., cancer screening)
Study Design	Systematic reviews and Meta-analyses	Non-systematic reviews (e.g., grey literature)

2. 검색 용어 및 방법

본 연구는 미국국립의학도서관에서 제시한 COSI(Core, Standard, Ideal) 모델을 사용하였고, 이 중 문헌 검색의 핵심 데이터베이스인 Core와 표준 검색 범위인 Standard에 해당하는 검색원을 포함하였다[31]. Core에 해당하는 PubMed에는 MeSH(Medical Subject Headings) term을, Standard에 해당하는 CINAHL(Cumulative Index to Nursing & Allied Health)에는 Subject Heading을, Scopus에는 수기검색을 실시하였다. 검색어는 연구문제 PICO-SD로부터 도출하였으며, 검색엔진별 검색어는 부록에 기록하였다.

3. 자료 추출

검색된 자료를 추출하기 위해 각 데이터베이스의 서지들을 Endnote(version 20, Clarivate corp., Philadelphia, PA, U.S.) 프로그램으로 반출한 후 중복되는 문헌들은 제외하였다. 중복 문헌 제거 후에는 연구의 제목과 초록을 검토하였고 본 연구 주제와 일치하지 않는 연구는 2차적으로 제외하였다. 제목과 초록만 보고 포함·배제 여부를 판단하기 힘든 경우 원문을 읽고 판단하였다. 문헌 선택을 위해 연구자 3인이 독립적으로 문헌을 고찰하였으며, 문헌 선택 과정에서 불일치한 의견은 연구자 간 논의를 통해 합의를 이루었다. 문헌을 선택하는 모든 과정은 PRISMA 흐름도를 통해 제시하였으며, 최종적으로 선정된 연구는 첫 번째 저자명과 게재연도, 분석 종류, 고찰에 포함된 논문 수, 대상 인구집단, 대상 프레임워크, 프로그램 환경, 대상 건강행동 및 상태의 프로그램 특징으로 분류하였다.

Results

1. 검색결과

동일한 키워드를 바탕으로 각 데이터베이스에

서 추출된 자료는 다음과 같다. 검색을 통해 확인된 문헌 수는 PubMed 357개, CINAHL 414개, Scopus 353개로 총 1,124개의 문헌이 검색되었다. 346개의 중복문헌 제거 후, 논문 제목과 초록 확인을 통해 또 다시 765개를 제거하였다. 원문이 확보된 최종 선정대상 논문 13개 중 프레임워크 사용에 대한 분석이 아닌 과정 평가 변인에만 초점을 맞춘 논문 1편, 의료적 행위를 결과 변인으로 측정한 논문 1편, 의료 체계에 대한 평가 프레임워크 사용을 분석한 논문 1편은 배제하였다. 이에 본 고찰에는 총 10개의 체계적 문헌 고찰 및 메타분석 연구가 포함되었다. 이와 같은 연구 선택 과정은 PRISMA 흐름도를 통해 제시하였다(Figure 1).

2. 선정된 연구 특성

최종적으로 선정된 연구의 특성은 Table 2에 제시하였다. 선정된 10개의 논문 중 체계적 문헌 고찰과 메타분석을 함께 실시한 1개의 논문을 제외한 나머지 9개의 논문은 모두 체계적 문헌 고찰 연구로 확인되었다. 대상 인구집단은 직장인 및 노동자를 대상으로 한 연구 3편, 교회 신도들을 대상으로 한 연구 1편, 다양한 인구 집단을 대상으로 한 연구 6편으로 구성되었으며, 평가 프레임워크의 종류는 여러 개의 평가프레임워크를 혼합한 연구 3편, RE-AIM 3편, Process evaluation for public health 2편, Precede-Proceed model 1편, Intervention Mapping 1편으로 확인되었다. 연구 환경은 다양한 환경에서 실시된 연구들을 포괄적으로 고찰한 연구 5편, 직장 환경에서 수행된 연구 3편, 교회 환경에서 수행된 연구 1편, 특별히 언급하지 않은 연구 1편으로 확인되었다. 대상 건강 행동 및 상태 역시 다양하게 수행되었다. 10개의 논문 중 신체활동과 좌식행동을 각각 고찰한 연구 2편을 제외한 나머지 8개의 연구는 모두 두 개 이상의 건강행동 및 상태를 대상으로 한 건강 증진 프로그램의 평가 프레임워크 사용을 분석하였다.

3. 고찰된 평가 프레임워크 종류

본 고찰에 최종적으로 포함된 10개의 논문 중 7개의 논문은 RE-AIM 3편, Process evaluation for public health 2편, Precede-Proceed 1편, Intervention Mapping 1편으로, 각각의 프레임워크가 다양한 건강 상태 및 행동 분야에 어떻게 사용되고 있는지 고찰한 연구이며, 나머지 3편의 논문은 하나의 건강 행동 분야에 다양한 프레임워크들이 어떻게 적용되어 왔는지 살펴본 연구이다. 앞서 언급한 'RE-AIM'은 대상자 모집 및 선정(Reach): 프로그램에 참여하고자 하는 연구 대상자의 수 또는 비율, 효과(Effectiveness): 경제적 영향 및 부정적 영향을 포함한 개입이 결과 및 삶의 질에 미치는 영향, 수행기관 특성(Adoption): 기관 차원과 프로그램 직원 차원으로 구분, 개입 수행 기관의 수, 비율, 대표성, 수행 인력의 특성 분석, 실행(Implementation): 개입 요소에 대한 실행자의 충실도(개입 제공의 일관성, 시간 및 비용), 지속(Maintenance): 기관 차원과 개인 차원으로 구분, 기관 차원은 '프로그램이 제도화 되거나 조직 및 정책의 일부가 되는 정도', 개인 차원은 '개입으로 인한 변화 지속성 평가'를 의미하며, 크게 다섯 가지 차원으로 구성되어 있다[32-34]. 또한 이 다섯 가지 차원에는 각 평가 차원에서 연구자가 평가해야 할 세부 지표들이 포함되어 있다. Linnan & Steckler의 Process evaluation for public health 프레임워크는 과정 평가 시 필요한 7개의 주요 구성 요소를 의미하며 각 요소는 환경(context): 프로그램에 영향을 미칠 수 있는 경제적 상황 및 사회적 환경, 또는 대조군에게 노출된 환경, 참여(reach): 연구 대상자의 특성 및 각 개입에 참여한 연구 참여자의 수 또는 비율, 전달된 양(dose delivered): 의도한 개입 모두를 제공하기 위한 연구자의 노력 및 각 개입의 유닛이 전달된 양, 받은 양(dose received): 연구 대상자의 특성, 개입 참여 및 상호작용 정도 혹은 제공된 자원 수용 정도, 충실도(fidelity): 개입 제공자의 역할 및 프로그램 개발자가 의도한대로 개입이 전달된 정도, 실행(implementation): 참여, 전달된 양, 받은 양, 충실도 모두를 종합하는 점수, 모집(recruitment): 연구 대상자를 모집하는 과정 및 절차를 의미한다[11]. Precede(Predisposing Reinforcing Enabling Constructs in Educational/environmental Diagnosis and Evaluation)는 대상 인구집단의 목표 행동 변화를 목적으로 한 맞춤형 프로그램 개발의 진단과 계획 과정을 의미하며, 사회적 진단, 역학적 진단, 교육 및 생태학적 진단, 행정 및 정책 진단과 중재 구성까지의 4단계를 의미한다[35, 36]. Proceed(Policy, Regulatory, and Organizational Constructs in

Educational and Environmental Development)는 Precede를 사용하여 프로그램을 수행하고 평가하는 단계를 의미하며, 프로그램 실행, 과정 평가, 영향 평가, 결과 평가까지 5-8단계로 구성되어 있다[35, 36]. 한편, Intervention Mapping은 Precede-Proceed 모형과 유사한 프로그램 계획 모형으로 프로그램 설계, 실행, 평가를 위한 문제 파악부터 해결까지 총 6단계로 구성된 프레임워크이며, 각 단계는 (1), 니즈 분석 및 문제 파악, (2) 변화 목표 매트릭스 제시, (3) 이론 기반 개입 선택, (4) 선택한 방법 적용, (5) 프로그램 채택, 실행, 유지를 위한 계획, (6) 평가 계획 생성으로 구성되어 있다[37]. Wierenga et al. [38]의 직장 건강 증진 프로그램의 과정 평가를 고찰한 연구에서는 Process evaluation for public health와 RE-AIM 모델 외에도 Integrative model, Durlak & Dupre의 모델, Baranowski & Stables의 모델, 'Diffusion of Innovations'에 대한 Rogers의 프레임워크를 결합한 모델을 사용하였다. 신체활동 프로그램을 평가한 연구들의 평가 프레임워크 사용에 대해 고찰한 Fynn et al. [23]의 연구에서는 RE-AIM, Process evaluation for public health, Logic Model, Precede-Proceed, Intervention Mapping 프레임워크 외에도 Developing a process evaluation plan, MRC Guidance on evaluation of complex interventions, MRC Guidance on process evaluation, Realist evaluation, Outcome Model, CDC Framework, Evaluation: a Systematic Approach, Model of Implementation, WHO Process Evaluation Workbook, Swiss Model for Outcome Classification, Concepts in process evaluation 프레임워크가 사용되었다. 마지막으로 성인의 좌식 행동 프로그램을 수행한 연구들의 과정 평가 방법에 대해 고찰한 Johansson et al. [39]의 연구에서는 MRC Guidance, RE-AIM, Linnan & Steckler의 Process evaluation framework for public health, Developing a process evaluation plan, Process evaluations of cluster-randomized trials of complex interventions, WHO Process Evaluation Workbook이 사용되었다.

4. 평가 프레임워크 사용 및 프로그램 특성

본 고찰에 포함된 체계적 문헌 고찰 연구는 신체활동에 초점을 맞춘 연구[23], 직장인 대상 스트레스 관리[24], 건강 위험 행동 예방 및 건강 증진 프로그램[38, 40], 좌식행동 프로그램에 초점을 맞춘 연구[39], 다양한 건강증진 프로그램을 포괄적으로 고찰한 연구[33, 34, 41], 교회 기반 건강 관련 프로그램에 초점을 맞춘 연구로 다양한 프로그램에 대한 평가 프레임워크 사용을 분석하였다[25] (Table 2, 3).

Process evaluation for public health 프레임워크 사용을 집중적으로 살펴본 Murta et al. [24]과 Yearly et al. [25]은 7개의 과정 평가 구성 요소 중 건강 증진 프로그램 과정 평가 시 평균 1~3개의 구성 요소가 측정된 것을 확인하였다. 한편, 교회 수준에서 실행된 연구들은 recruitment (88.1%), reach (80.6%), context (34.3%), dose delivered (28.4%), dose received (27.3%), implementation (20.9%), fidelity (9%) 순으로 과정 평가 구성 요소가 측정된 반면, 주로 개인 수준(78.8%)에 초점을 맞춘 Murta et al. [24]의 연구는 recruitment (30%), dose received (22%), attitudes toward intervention (19%), reach (13%), context (9%), fidelity (5%), dose delivered (2%) 순으로 과정 평가 구성 요소가 측정되었다. Linann & Steckler의 과정 평가 프레임워크는 신체활동 분야에서 사용된 16개의 프레임워크 중 세 번째로 빈번하게 사용되었으며 독립적인 프레임워크로 사용되는 것 보다는 다른 평가 프레임워크와 함께 사용되는 경우가 많았다[23].

직장 내 건강 증진 프로그램의 과정 평가를 분석한 Wierenga et al. [38]의 연구에서는 선정된 22편의 논문 중 단 8편의 논문이 과정 평가 프레임워크를 사용하였다. 8편의 논문 중 4편이 Linann & Steckler의 과정 평가 프레임워크를 사용하였고, 나머지 4편이 각각 RE-AIM 모델, Durlak & Dupre의 모델, Baranowski & Stables의 모델, 'Diffusion of Innovations'에 대한 Rogers의 프레임워크를 결합하여 사용하였다. 평균적으로 보고된 평가 요소 8개 중 3.9개의 요소가 보고되었으며, dose received (82%), dose delivered (68%), context (68%), fidelity (41%), satisfaction (41%), reach (32%), recruitment (23%) 순으로 사용되었다.

성인의 좌식 행동을 결과 변인으로 측정한 논문들의 과정 평가 절차를 살펴본 Johansson et al. [39]의 연구에서는 과정 평가를 실시한 17편의 연구 중 평가 프레임워크를 사용한 연구는 13편(76.5%)으로 확인되었다. MRC Guidance(23.5%)와 RE-AIM(23.5%) 프레임워크가 가장 빈번하게 사용되었고, Linnan & Steckler의 과정 프레임워크가 3편(17.6%)의 연구에서 사용되었다. 그 다음으로는 Developing a process evaluation plan, Process

evaluations of cluster-randomized trials of complex interventions, Process Evaluation Workbook by WHO가 각각 1편(5.9%)의 연구에서 사용되었다. 각 프레임워크는 단독으로 사용되기도 했지만 주로 두 개 이상의 프레임워크가 함께 사용된 것으로 확인되었다.

PRECEDE-PROCEED 프레임워크의 사용을 분석한 Kim et al. [41]의 연구는 해당 프레임워크를 세 영역으로 분류하여 건강 증진 프로그램의 평가로 어떤 영역이 포함되었는지 알아보았다. 고찰된 26편의 연구 중 8편(30.8%)이 프로그램 설계 및 계획(planning) 영역에서 사용되었고, 2편(7.7%)의 연구가 실행(implementation)과 평가(evaluation) 영역을 함께 사용하였으며, 16편(61.5%)의 연구가 세 영역을 모두 사용한 것으로 확인되었다. PRECEDE-PROCEED 모델은 신체활동 분야에서 사용된 16개의 프레임워크 중 8번째로 빈번하게 사용되는 것으로 나타났다[23].

직업 안전 및 건강 영역에서의 Intervention mapping(IM) 사용을 분석한 Bakhuys Roozeboom et al. [40]은 핵심 IM 특성(참여, 이론 기반 접근, 생태학적 접근, 실행 계획)과 관련하여 IM 프로토콜 사용의 충실도를 분석하였다. 먼저 참여와 관련하여 12개의 연구 중 오직 2개의 연구가 다양한 이해관계자를 구축하는 과정을 언급하였다. 다른 연구들도 프로그램 관련 이해관계자에 대한 내용을 포함하긴 하였으나 IM에서 제시한 순서에 적합하게 보고하지 않았다. 이론 기반 접근과 관련해서는 모든 연구가 니즈 평가를 실시하였으나 1단계에 해당하는 인과적 경로 언급은 12개의 개입 중 11개의 개입에서 언급되었다. 또한 12개 연구 중 5개의 연구가 2단계에 해당하는 변화 목표 매트릭스를 작성하였고, 모든 연구가 3단계에 해당하는 이론 및 근거 기반 변화 방법 선택에 대해 보고하였다. 10편의 개입 연구가 3단계에 해당하는 변화의 로직 모델을 수행하는 과정에서 행동과 환경 요인을 구분하였고, 모든 연구가 4단계에 해당하는 근로자뿐만 아니라 환경적 맥락(예: 직장)을 대상으로 하는 두 구성 요소를 모두 포함하였다. 마지막으로 단 1편의 연구가 잠재적 프로그램 사용자를 식별하였고, 나머지 연구들은 채택자와 실행자는 식별했으나 유지자를 식별하지 않은 것으로 나타났다. 또한 한 편의 연구가 프로그램 실행을 위한 성과 목표를 수립하였고, 7개의 연구는 실행 동력과 장벽을 식별하였으며, 11개 연구가 프로그램을 실행하기 위한 개입을 설계하였다. Bakhuys Roozeboom et al. [40]의 고찰에 포함된 12개의 연구는 모두 IM 프레임워크의 모든 단계를 평가하지 않은 것으로 나타났으며, 신체활동 연구 분야의 평가 도구로 빈번하게 사용되지 않는 것으로 나타났다[23].

건강 증진 분야의 RE-AIM 프레임워크 사용을 분석한 연구 중 Gaglio et al. [42]은 신체활동 및 비만(36.6%), 질병 자가 관리(29.6%), 흡연 및 약물 남용(9.9%), 정신 건강 및 치매(1.4%), 암 예방(1.4%), 다중 주제(14.1%)를 대상으로 한 건강 증진 프로그램의 RE-AIM 프레임워크 사용을 분석하였다. 해당 분야의 71편의 연구 중 RE-AIM 프레임워크는 대상자 모집 및 선정(91.5%), 실행(90.1%), 효과(77.5%), 수행기관 특성(기관 차원: 75.3%), 수행기관 특성(직원 차원: 74.6%), 지속(조직 차원: 71.8%), 지속(개인 차원: 64.8%) 순으로 각 차원이 보고되었다. 한편, Harden et al. [33]은 Gaglio et al. [42]의 연구에서 RE-AIM의 다섯 가지 차원을 모두 보고한 44개 연구와 이후 연구들을 바탕으로 일반적으로 보고되는 RE-AIM의 각 차원을 평가하였다. 신체활동 참여 및 식습관 개선(32%), 흡연 및 약물 남용(15%), 신체활동 단독(10%) 질병 자가 관리(5%), 식이(5%), 체중(2%) 등의 분야에서 RE-AIM 프레임워크는 대상자 모집 및 선정, 효과, 실행, 수행기관 특성, 지속(기관 차원), 지속(개인 차원) 순으로 보고되었고, 이는 선행 연구인 Gaglio et al. [42]의 연구와 비슷하지만 효과와 실행 차원 보고에 있어 약간의 차이를 보였다. Harden et al. [33]의 연구에서 고찰한 101편의 연구 중 RE-AIM 프레임워크의 다섯 가지 차원이 하나라도 포함된 연구의 개입 수는 82편으로 확인되었다. 그러나 각 차원에 해당하는 지표들의 측정 비율은 저조하였다. 대상자 모집 및 선정의 지표들을 모두 측정한 연구는 17%로 나타났으며, 단 한 편의 연구가 효과에 해당하는 지표들을 모두 측정하였다. 개인 차원의 지속에 해당하는 세 가지 지표를 모두 측정한 연구는 한 편도 없었으며, 한 편의 연구가 조직 차원의 수행기관 특성에 해당하는 여섯 가지 지표를 모두 측정하였고, 마찬가지로 한 편의 연구가 실행에 해당하는 지표들을 모두 측정하였다. 마지막으로 조직 차원의 지속에 해당하는 지표를 모두 측정한 연구는 단 한 편도 없는 것으로 나타났다. 가장

최근 RE-AIM 프레임워크 사용을 분석한 D'Lima et al. [34]은 RE-AIM 프레임워크를 사용한 최신 문헌들을 종합하고 현실 활용 (프레임워크의 부분적 적용) 및 보고 내용을 분석하였다. D'Lima et al. [34]이 고찰한 155편의 연구 중 RE-AIM 프레임워크의 다섯 가지 차원을 모두 보고한 연구는 69%로 나타났으며, 4개 차원을 언급한 연구는 3.5%, 3개는 6.5%, 2개는 5.2%, 1개만 언급한 연구는 5.8%로 나타났다. 다섯 개의 차원 중 대상자 모집 및 선정이 92.9%로 가장 빈번하게 보고되었으며, 다음으로는 실행(90.%), 수행기관 특성(89.7%), 효과(84.7%), 지속(77.4%) 순으로 보고되었다. D'Lima et al. [34]은 이러한 RE-AIM 평가 프레임워크의 보고 수치를 선행 연구인 Gaglio et al. [42]의 연구와 비교하였고 결과적으로 다섯 가지 차원의 보고 비율이 이전보다 증가한 것을 알 수 있었다. D'Lima et al. [34]은 157개의 연구들 중 15편(10%)의 RE-AIM 연구를 무작위로 선정하여 심층 분석한 결과, 9편의 연구가 RE-AIM의 다섯 가지 차원을 모두 평가하였으나 다섯 가지 차원 중 하나라도 평가하지 않은 나머지 6편의 연구 중 4편의 연구가 RE-AIM의 특정 차원을 평가하지 않은 이유를 설명하지 않았다고 보고하였다. 한편, RE-AIM 프레임워크는 69편의 신체활동 증진 연구에서 사용된 16개의 평가 프레임워크 중 가장 활발히 사용되었으며(39.1%), 다른 평가 프레임워크와 함께 사용되지 않고 주로 단독으로 사용되는 것으로 확인되었다[23].

5. 평가 프레임워크 적용 효과(결과 요약)

Linann & Steckler의 과정 평가 프레임워크의 사용을 분석한 Yearly et al. [25]은 과정 평가 구성 요소의 측정 평균이 흡연, 심혈관 건강, 식습관 및 체중 조절, 암, 당뇨, 정신건강, 일반 건강 증진 프로그램 순으로 높게 나타났으나, 프로그램 특성(대상 인구, 대상 건강 상태, 프로그램 목표 등)에 따른 평가 요소의 측정 개수에는 차이가 없다고 보고하였다[25].

직장 수준에서 실행된 22편의 건강 증진 개입 연구 중 7편의 연구는 개입 실행의 dose와 fidelity가 높을수록 주요 결과 변인에 긍정적인 영향을 미친다고 보고하였다. 그러나 7개의 연구 중 단 3편이 평가 프레임워크를 적용한 연구로 확인되었다[38]. Johansson et al. [39]은 MRC Guidance 프레임워크를 바탕으로 성인의 좌식행동 개입 연구들을 평가한 결과, 개입 전달, 상황적 요인, 개인적 요인, 방해 요인, 촉진 요인에 의해 좌식 행동 변화가 나타났다고 보고하였다. 또한 과정 평가를 실행한 17편의 좌식행동 개입 연구 중 3편의 연구가 좌식 행동을 감소시키는데 효과적이었다고 보고하였다. 그러나 3편 모두 평가 프레임워크를 적용하지 않은 것으로 나타났다[39].

Bakhuys Roozeboom et al. [40]은 IM 프로토콜 사용의 충실도를 분석함과 동시에 개입 실행 과정을 평가하기 위해 Linann & Steckler의 과정 평가 요소에 만족(satisfaction) 요소를 추가하여 IM 프레임워크를 사용한 연구들의 평가 지표 사용을 평가하였다. Reach를 측정한 8편의 연구 중, 3편(37.5%)의 연구가 훌륭한 수준으로 reach를 보고하였으며, 5편(62.5%)의 연구가 불만족스러운 수준으로 reach를 보고하였다. dose delivered에 대한 정보를 보고한 8편의 연구 중, 6편(75%)의 연구는 우수, 2편(25%)의 연구는 보통 수준으로 dose delivered를 보고하였다. 한 편의 연구를 제외한 모든 연구에서 dose received를 보고하였고, 이 중 1편(9.1%)의 연구만 우수한 수준으로 dose received를 보고하였으며, 7편(63.6%)의 연구는 보통 수준, 3편(27.3%)의 연구는 불만족스러운 수준의 dose received를 보고하였다. fidelity는 8편의 연구에서 보고되었으며, 이 중 1편(12.5%)의 연구에서만 만족스러운 충실도를 보고하였다. 개입 만족도와 관련된 satisfaction은 10편의 연구에서 보고되었으며, 그 중 1편(10%)은 훌륭한 수준, 8편(80%)은 만족스러운 수준, 나머지 1편(10%)은 보통 수준의 만족도를 보고하였다. 또한 IM 충실도에 따른 실행 과정 및 개입 효과 간의 유의미한 관계성은 없다고 보고하였다.

PRECEDE-PROCEED 프레임워크를 적용한 연구들의 개입 효과는 1차 결과 측정 변인, 즉 지식 관련 변인만 유의미한 것으로 나타났다[41].

RE-AIM의 지표 중 참여자 비율을 정확하게 보고한 연구는 45%로 나타났다. 개별 행동 결과에 대한 효과를 정확하게 보고한 연구들 중 89%는 긍정적 변화를, 11%는 무효 결과를 보고하였다. 프로그램 종료 6개월 후의 개입 효과는 9편(11%)의 연구에서 보고되었으며, 9편 모두 개입 전에 비해 긍정적인 효과가 있는 것으로 나타났다. 평균 조직 차원 및 직원 차원의 수행기관 비율은 각각 75%와 79%로 나타났으며, 조직

차원의 지속 비율은 불충분한 자료로 인해 측정하지 못한 것으로 나타났다[33]. 한편, Gaglio et al. [42]은 RE-AIM의 누락된 평가 기준에 대해 언급하지 않은 반면, D'Lima et al. [34]은 하위 분석을 통해 RE-AIM 프레임워크를 사용한 연구들의 누락된 주요 평가 기준을 파악하였다. (1) 제외된 참여자의 비율 및 이유, (2) 공중보건 목표와 관련된 주요 결과 측정, (3) 수행기관에 대한 직원 제외 비율 및 이유, (4) 충실도를 위한 완벽한 전달 및 비율, 그리고 (5) 하위 그룹의 장기 효과와 관련된 데이터가 몇몇 연구에서 누락된 것으로 확인되었다.

Discussion

본 연구는 체계적 문헌고찰 연구의 포괄적 고찰을 통해 여러 개의 평가 프레임워크가 다양한 수준에서 건강 증진 프로그램을 평가하는데 적용되고 있다는 점을 파악하였다. 그 중에서도 7편의 체계적 문헌 고찰 연구는 건강 증진 분야의 특정 주제에 대한 RE-AIM[32], Process evaluation for public health[11], Intervention Mapping[43], Precede-Proceed Model[36] 이 네 가지 프레임워크 사용을 독립적으로 분석하여 각 프레임워크가 어떻게 적용되고 있는지에 대한 자세한 내용을 파악할 수 있었다. 신체활동 분야의 경우, Developing a process evaluation 프레임워크가 RE-AIM 모델 다음으로 빈번하게 사용되는 것으로 나타났으나 종종 Linann & Steckler의 프레임워크와 함께 쓰이거나[23], 신체활동 분야 외에는 빈번하게 쓰이지 않는 것으로 확인되었다. 이는 아마도 Developing a process evaluation 과정 평가 모델이 Linann & Steckler의 주요 과정 평가 요소를 채택하여 발전된 것이기 때문에 선행 연구들이 Linann & Steckler의 프레임워크를 인용하여 프로그램 평가를 진행했을 가능성이 높다.

RE-AIM 프레임워크 사용을 분석한 체계적 문헌 고찰은 총 세 편으로, 세 편 모두 RE-AIM 프레임워크의 다섯 가지 차원의 평가 비율 및 각 차원에 해당하는 34가지 지표들의 보고 비율을 분석하였다[33, 34, 42]. RE-AIM의 다섯 가지 차원의 보고 비율은 다른 프레임워크에 비해 높은 편이라는 것을 알 수 있었다. 그러나 각 차원에 해당하는 지표들을 포괄적으로 보고한 연구는 많지 않았다. RE-AIM의 다섯 가지 차원 중 대상자 모집 및 선정이 가장 빈번하게 보고된 반면, 지속은 가장 적게 보고되었다. 이와 관련하여 Harden et al. [33]은 포함된 연구들의 자료 부족으로 인해 기관 차원의 지속 효과를 측정하기 어렵다고 보고하였으며, 실제로 기관 차원의 지속에 해당하는 지표들을 모두 평가한 연구는 없었다[42]. 이에 대해 Holtrop et al. [44]은 기존 연구들이 지속 효과를 측정했더라도 초기 RE-AIM 연구의"개입 종료 후 6개월 이후 시점의 추적조사"라는 정의 때문에 해당 시점에 비해 단기 혹은 장기 효과를 측정했을 때 이를 보고하기 어려웠을 수 있다고 설명하였다. 한편, RE-AIM을 사용한 연구들은 종종 대상자 모집 및 선정과 수행기관 특성을 혼동하는 경향을 보였다[45]. 이는 RE-AIM 사용을 고찰한 체계적 문헌 고찰 연구 세 편에서도 공통적으로 지적한 부분으로 RE-AIM 사용 시 각별히 주의해야 하는 부분이다. Holtrop et al. [44]은 이 문제에 대해 RE-AIM 사용 전 프레임워크 적용 및 측정에 관한 안내와 교육을 참고할 필요가 있다고 제언하였다. 최근 20년 간 RE-AIM 프레임워크의 적용 및 발전 과정을 살펴본 Glasgow et al. [45]은 RE-AIM 사용 시 지속적으로 언급되는 문제를 해결하기 위해서는 RE-AIM 프레임워크에 대한 여러 가지 자원 및 가이드라인을 제시하는 www.re-aim.org 웹사이트 이용의 필요성을 강조하였다. RE-AIM 프레임워크의 사용이 증가함에 따라 현재 사용 현황 및 문제점, 그리고 앞으로의 방향성에 대한 연구가 지속적으로 게재되고 있기 때문에 향후 RE-AIM 프레임워크를 사용하고자 하는 연구자들은 해당 자료들을 통해 RE-AIM의 정확한 측정 방법을 확인할 수 있을 것으로 판단된다. 한편, RE-AIM의 각 다섯 가지 차원 보고 비율은 1999년부터 2010년까지의 문헌들을 고찰한 Gaglio et al. [42]의 연구에 비해 2011년부터 2017년까지의 문헌들을 고찰한 D'Lima et al. [34]의 연구에서 증가하였다. 이는 건강 증진 분야의 개입 연구들이 RE-AIM의 지표들을 전보다 더 포괄적으로 적용하여 프로그램을 평가하고 있다고 볼 수 있다.

Linann & Steckler의 프레임워크는 신체활동을 포함한 다양한 건강 증진 프로그램의 평가 프레임워크로 사용되었고, RE-AIM 모델 다음으로 빈번하게 사용되는 것으로 나타났다. Linann & Steckler의 프레임워크는 단독으로 사용되기도 했지만 다른 평가 프레임워크와 함께 적용되는 경우가 많았다[23]. 이는 아마도 Linann &

Steckler의 프레임워크의 안면 타당도가 과정 평가에 속하는 것뿐만 아니라 대부분이 프로그램 실행 과정을 평가하는데 초점을 맞추었기 때문일 수 있다. 실제로 신체활동 개입 연구를 실시한 연구들 중 개입 실행 과정을 평가한 연구들은 과정 평가 프레임워크를 사용하는 경향을 보였다[23]. Murta et al. [24]은 Linann & Steckler의 프레임워크가 유용한 과정 평가 도구임에도 불구하고 해당 프레임워크를 사용한 직장 기반 연구들의 과정 평가는 미흡한 수준에서 수행되었다는 점을 지적하였다. 반면, Wierenga et al. [38]은 해당 프레임워크가 프로그램 실행 과정을 촉진시키거나 방해할 수 있는 상황적 요인에 대한 부분을 간과한다는 점을 지적하면서 다른 프레임워크를 기반으로 한 체계적이고 포괄적인 과정 평가 수행을 제안하였다. Yearly et al. [25]와 Murta et al. [24]은 Linann & Steckler의 프레임워크를 사용하여 과정 평가를 실시한 연구들이 7개의 구성 요소 중 3개의 요소만을 측정했다는 것을 공통적으로 확인하였다. 그리고 이러한 불완전한 보고는 개입 실행 및 결과에 영향을 미친 구성요소가 무엇인지 파악하는데 어려움을 주었다고 보고하였다[38]. 고찰 결과, Linann & Steckler의 프레임워크를 사용한 연구들은 대부분 해당 프레임워크를 부분적으로 적용하는 경향을 보였으며, 어떠한 구성요소가 실제 개입 결과에 가장 큰 영향을 미치는지는 정확히 알 수 없었다.

한편, Bakhuys Roozeboom et al. [40]은 IM 프레임워크의 4가지(참여, 이론 기반 접근, 생태학적 접근, 실행 계획) 특성 중 이론 기반 접근은 개입 실행 과정(만족도 및 개입 수신)과 긍정적인 연관성이 있었으나, IM 프로토콜의 충실도는 개입 실행 및 결과에 영향을 미치지 않았다고 보고하였다. Bakhuys Roozeboom et al. [40]은 IM 프레임워크를 사용한 연구가 많지 않아 IM 프레임워크의 충실도가 개입 실행 과정과 결과에 미치는 영향을 양적으로 분석하는 것이 쉽지 않았다고 보고하였다. 본 고찰에는 포함되지 않았으나 직업 관련 장애 예방 개입 연구의 IM 프레임워크 사용을 분석한 Fassier et al. [46]의 체계적 문헌 고찰 연구 역시 8편의 연구만이 최종 고찰에 포함되었다. IM 프레임워크는 질병 예방[47] 및 암[48] 관련 개입 연구에서 비교적 활발하게 사용되고 있는 것으로 나타났다. 두 연구 모두 IM 프레임워크를 적용한 연구들은 질병 예방 개입 이행 수준을 유의미하게 증가시켰으며[47], 암 검진 비율 증가 및 사회적 불공평을 감소시키는데 효과적이었다고 보고하였다[48]. 한편, 신체활동 분야에서는 평가를 수행한 69편의 연구 중 단 2편의 연구만이 IM 프레임워크를 사용한 것으로 확인되었다[23]. 프레임워크가 적게 사용된다고 해서 해당 프레임워크가 평가도구로 부적절한 것은 아니다. 그러나 IM 프레임워크는 절차가 다소 복잡하고 지나치게 세부적이어서 모든 단계를 수행하기에는 오랜 시간이 소요된다는 단점이 있다[37]. 직업 안전 및 건강 증진 연구들도 이러한 맥락에서 참여적 접근, 변화 목표 매트릭스 실행, 그리고 개입 실행 계획 절차를 따르는데 어려움을 겪었을 수 있다.

PRECEDE-PROCEED 모델을 적용한 건강 증진 개입 연구들은 계획, 실행, 평가 단계를 모두 적용한 연구도 많았지만 평가 단계를 제외한 계획 및 실행 단계만 적용한 연구가 삼분의 일 이상을 차지했다[41]. 이는 PRECEDE-PROCEED 모델이 IM 프레임워크처럼 프로그램 평가를 용이하게 하지만 계획 및 개발에 대한 지침을 주로 제공하기 때문이라고 볼 수 있다[23]. Kim et al. [41]의 메타 분석 결과 PRECEDE-PROCEED 모델을 적용한 연구는 건강 행동 및 상태와 관련한 지식 변화에 긍정적 영향을 미쳤다. 본 연구에는 포함되지 않았지만 암 검진 및 교육 프로그램과 관련한 연구들을 고찰한 Saulle et al. [49]의 연구에서도 PRECEDE-PROCEED 모델을 적용한 연구는 건강 관련 이해와 지식을 향상시키는데 효과적이라고 보고하였다. 이는 PRECEDE-PROCEED 모델의 교육 및 생태학적 평가 단계에서 선행 요인(predisposing factor) 지식을 긍정적으로 변화시켰다는 것을 의미한다. 그러나 Kim et al. [41]은 고찰한 논문들의 연구 디자인과 결과 측정 변인이 지나치게 다양하다는 이유로 건강 증진 및 상태 변화에 대한 효과 크기는 측정하지 못했다. PRECEDE-PROCEED 모델은 건강 증진 프로그램을 설계하는데 유용하고 프레임워크 특성에 따라 건강 관련 지식을 향상시키는데 효과적이지만 여전히 프로그램 계획 단계에서만 적용되는 경우가 많았다.

앞서 논의한 네 가지 프레임워크 외에도 신체활동 개입 연구에서 빈번하게 사용되는 프레임워크로 MRC Guidance가 있었다[23]. 여러 개의 평가 프레임워크 사용을 살펴본 Johansson et al. [39]의 연구에서는 4편의 좌식행동 개입 연구가

과정 평가를 위해 MRC 프레임워크를 인용한 것을 확인하였다. 그러나 네 편의 연구 중 단 한 편의 연구만이 MRC 프레임워크를 적용하여 과정 평가를 실시하였다. Johansson et al. [39]은 해당 연구가 전체적으로 고품질 연구 중 하나였으며, 전체 과정 평가 수행을 위한 프레임워크 적용은 연구의 질을 향상시킬 수 있다고 제언하였다. 그러나 한 편의 연구만으로는 MRC 프레임워크가 실제 과정 평가 도구로 효과적인지 알 수 없다. 게다가 Fynn et al. [23]의 연구에서 MRC 프레임워크를 사용한 5편의 연구도 해당 프레임워크 가이드라인에 의해 평가를 수행했다는 내용을 명확하게 제시하지 않았기 때문에 추가적인 고찰이 필요할 것으로 사료된다.

본 고찰에 포함된 연구들을 종합한 결과, 건강 행동 및 상태 혹은 프로그램 특성별로 비교적 많이 사용되는 프레임워크는 확인할 수 있었으나 각 특성에 따라 적용되는 프레임워크의 차이는 명확하게 알 수 없었다. 고찰 결과, RE-AIM 프레임워크가 건강 증진 분야의 평가 프레임워크로 가장 빈번하게 사용되고 있다는 점을 파악할 수 있었다. 이는 RE-AIM 프레임워크가 과정 및 결과 평가 요소에 대한 포괄적인 가이드라인을 제시하기 때문으로 판단된다. RE-AIM 프레임워크는 신체활동 개입 연구에서도 가장 빈번하게 사용되었고, 주로 단독으로 사용되었다.

본 연구에는 두 가지 한계점이 있다. 첫째, 2005년부터 최근까지의 연구를 검색하였지만 영문으로 작성된 연구와 개입 및 결과 측정 변인이 비의료적 형태인 연구만을 포함하였기 때문에 연구의 누락이 발생했을 수 있다. 둘째, 빈번하게 사용되지 않는 평가 프레임워크 적용에 대한 체계적 문헌 고찰 연구를 포함하지 못했기 때문에 해석 시 주의가 필요하다. 이러한 제한점에도 불구하고 본 고찰은 하나의 평가 프레임워크 사용에 대한 고찰이 아닌 건강 증진 분야의 여러 평가 프레임워크 적용 내용을 종합하여 연구자들에게 해당 분야에 대한 주요 정보를 제공한 것에 의미가 있다고 사료된다.

Conclusion

본 연구는 건강증진 프로그램의 평가 프레임워크 사용을 분석한 체계적 문헌고찰 및 메타분석 연구들을 종합적으로 고찰하였다. 고찰 결과, RE-AIM 프레임워크가 신체활동을 포함한 건강증진 프로그램의 평가 도구로 가장 빈번하게 사용되고 있다는 점을 파악하였다. 누적 사용량에 따라 RE-AIM 프레임워크 적용의 문제점 및 방향성에 대한 연구들이 지속적으로 게재되고 있으므로 향후 연구에서는 이러한 내용을 참고하여 체계적인 프로그램 평가를 실시할 필요가 있다. 본 연구는 기존에 수행되지 않은 평가 프레임워크 사용에 대한 포괄적 고찰을 수행함으로써 건강증진 분야의 평가 프레임워크 사용에 대한 종합 내용을 제시했다는데 의의가 있다.

References

1. Oldenburg B, Absetz P. Lost in translation: overcoming the barriers to global implementation and exchange of behavioral medicine evidence. Translational Behavioral Medicine. 2011;1(2):252–5.
2. Kozica, SL, Lombard CB, Hider K, Harrison CL, Teede HJ. Developing comprehensive health promotion evaluations: a methodological review. MOJ Public Health. 2014;1(1):39-48.
3. Poland B, Frohlich KL, Cargo M. Context as a fundamental dimension of health promotion program evaluation. Health Promotion Evaluation Practices in the Americas: Values and Research. 2009;299-317.
4. McKenzie JF, Neiger BL, Thackeray R. Planning, implementing and evaluating health promotion programs. Jones & Bartlett Learning. 2022.
5. Windsor R, Clark N, Boyd NR, Goodman RM. Evaluation of health promotion, health education, and diverse prevention programs. Mayfield Publishing. 2004.
6. O'Connor-Fleming ML, Parker E, Higgins H, Gould T. A framework for evaluating health promotion programs. Health Promotion Journal of Australia. 2006;17(1):61-66.
7. Swinburn B, Bell C, King L, Magarey A, O'Brien K, Waters E. Obesity prevention programs demand high-quality evaluations. Australian and New Zealand Journal of Public Health. 2007;31(4):305-307.
8. Fleming ML, Parker E. Health promotion: Principles and practice in the Australian context. Allen & Urwin. 2007.
9. Baranowski T, Stables G. Process evaluations of the 5-a-day projects. Health Education & Behavior. 2000;27(2):157-166.

10. Cook TD, Campbell DT, Day A. Quasi-experimentation: Design & analysis issues for field settings (Vol. 351). Boston: Houghton Mifflin. 1979.
11. Linnan L, Steckler A. Process Evaluation for Public Health Interventions and Research: An Overview. Jossey-Bass. 2002.
12. Bouffard JA, Taxman FS, Silverman R. Improving process evaluations of correctional programs by using a comprehensive evaluation methodology. Evaluation and Program Planning. 2003;26(2):149-161.
13. Harachi TW, Abbott RD, Catalano RF, Haggerty KP, Fleming CB. Opening the black box: using process evaluation measures to assess implementation and theory building. American Journal of Community Psychology. 1999;27(5):711-731.
14. McCreary LL, Kaponda CP, Kafulafula UK, Ngalande RC, Kumbani LC, Jere DL, et al. Process evaluation of HIV prevention peer groups in Malawi: a look inside the black box. Health education research. 2010;25(6):965-978.
15. Israel BA, Cummings KM, Dignan MB, Heaney CA, Perales DP, Simons-Morton BG, et al. Evaluation of health education programs: Current assessment and future directions. Health Education Quarterly. 1995;22(3):364-389.
16. Spillane V, Byrne MC, Byrne M, Leathem CS, O'Malley M, Cupples ME. Monitoring treatment fidelity in a randomized controlled trial of a complex intervention. Journal of Advanced Nursing. 2007;60(3):343-352.
17. Wight D, Obasi A. Unpacking the black box: the importance of process data to explain outcomes. In: Stephenson JM, Imrie J, Bonell C (Eds.). Effective Sexual Health Interventions: Issues in Experimental Evaluation (pp. 151-66). Oxford University Press. 2003.
18. Flowers P, Hart GJ, Williamson LM, Frankis JS, Der GJ. Does bar-based, peer-led sexual health promotion have a community-level effect amongst gay men in Scotland?. International Journal of STD & AIDS. 2002;13(2):102-108.
19. Plummer ML, Wight D, Obasi AIN, Wamoyi J, Mshana G, Todd J, et al. A process evaluation of a school-based adolescent sexual health intervention in rural Tanzania: the MEMA kwa Vijana programme. Health Education Research. 2007;22(4):500-512.
20. Robbins LB, Pfeiffer KA, Wesolek SM, Lo YJ. Process evaluation for a school-based physical activity intervention for 6th-and 7th-grade boys: Reach, dose, and fidelity. Evaluation and Program Planning. 2014;42:21-31.
21. Young DR, Steckler A, Cohen S, Pratt C, Felton G, Moe SG, et al. Process evaluation results from a school-and community-linked intervention: the Trial of Activity for Adolescent Girls (TAAG). Health Education Research. 2008;23(6):976-986.
22. Fynn JF, Hardeman W, Milton K, Jones AP. A scoping review of evaluation frameworks and their applicability to real-world physical activity and dietary change programme evaluation. BMC Public Health. 2000;20(1):1-16.
23. Fynn JF, Hardeman W, Milton K, Murphy J, Jones A. A systematic review of the use and reporting of evaluation frameworks within evaluations of physical activity interventions. International Journal of Behavioral Nutrition and Physical Activity. 2000;17(1):1-17.
24. Murta SG, Sanderson K, Oldenburg B. Process evaluation in occupational stress management programs: a systematic review. American Journal of Health Promotion. 2007;21(4):248-254.
25. Yearly KHCK, Klos LA, Linnan L. The examination of process evaluation use in church-based health interventions: a systematic review. Health Promotion Practice. 2012;13(4):524-534.
26. Higgins J, Green S. Cochrane Handbook For Systematic Reviews Of Interventions. Wiley. 2008.
27. Becker L, Oxman A. Overview of reviews. In: Higgins JP, Green S (Eds.). Cochrane handbook for systematic reviews of interventions: Cochrane book series (pp. 607-631). John Wiley & Sons. 2008.
28. Grant MJ, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. Health Information & Libraries journal. 2009;26(2):91-108.
29. Hartling L, Chisholm A, Thomson D, Dryden DM. A descriptive analysis of overviews of reviews published between 2000 and 2011. PloS One. 2012;7(11):e49667.
30. Aromataris E, Fernandez R, Godfrey CM, Holly C, Khalil H, Tungpunkom P. Summarizing systematic reviews: methodological development,

conduct and reporting of an umbrella review approach. JBI Evidence Implementation. 2015;13(3):132-140.
31. 김수영, 박지은, 서현주, 이윤재, 손희정, 장보형, 외. NECA 체계적 문헌고찰 매뉴얼. 한국보건의료연구원. 2011. (Korean)
32. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. American Journal of Public Health. 1999;89(9):1322-1327.
33. Harden SM, Gaglio B, Shoup JA, Kinney KA, Johnson SB, Brito F, et al. Fidelity to and comparative results across behavioral interventions evaluated through the RE-AIM framework: a systematic review. Systematic reviews. 2015;4:1-13.
34. D'Lima D, Soukup T, Hull L. Evaluating the application of the RE-AIM planning and evaluation framework: An updated systematic review and exploration of pragmatic application. Frontiers in Public Health. 2022;9:755738.
35. Crosby R, Noar SM. What is a planning model? An introduction to PRECEDE-PROCEED. Journal of Public Health Dentistry. 2011;71:S7-S15.
36. Green LW, Kreuter MW. Health Promotion Planning : An Educational and Environmental Approach. (2nd ed). Mayfield Publishing Co. 1991.
37. Kok G, Mesters I. Getting inside the black box of health promotion programmes using intervention Mapping. Chronic Illness. 2011;7(3):176-180.
38. Wierenga D, Engbers LH, Van Empelen P, Duijts S, Hildebrandt VH, Van Mechelen W. What is actually measured in process evaluations for worksite health promotion programs: a systematic review. BMC Public Health. 2013;13(1):1-16.
39. Johansson JF, Lam N, Ozer S, Hall J, Morton S, English C, et al. Systematic review of process evaluations of interventions in trials investigating sedentary behaviour in adults. BMJ open. 2022;12(1):e053945.
40. Bakhuys Roozeboom MC, Wiezer NM, Boot CR, Bongers PM, Schelvis RM. Use of intervention mapping for occupational risk prevention and health promotion: a systematic review of literature. International Journal of Environmental Research and Public Health. 2021;18(4):1775.
41. Kim J, Jang J, Kim B, Lee KH. Effect of the PRECEDE-PROCEED model on health programs: a systematic review and meta-analysis. Systematic Reviews. 2022;11(1):213.
42. Gaglio, B., Shoup, J. A., & Glasgow, R. E. (2013). The RE-AIM framework: a systematic review of use over time. American journal of public health, 103(6), e38-e46.
43. Bartholomew LK, Parcel GS, Kok G, Gottlieb NH. (2006) Planning health promotion programs: an intervention mapping approach. John Wiley & Sons.
44. Holtrop JS, Estabrooks PA, Gaglio B, Harden SM, Kessler RS, King DK, et al. Understanding and applying the RE-AIM framework: Clarifications and resources. Journal of Clinical and Translational Science. 2021;5(1):e126.
45. Glasgow RE, Harden SM, Gaglio B, Rabin B, Smith ML, Porter GC, et al. RE-AIM planning and evaluation framework: adapting to new science and practice with a 20-year review. Frontiers in public health. 2019;7:64.
46. Fassier, J. B., Sarnin, P., Rouat, S., Peron, J., Kok, G., Letrilliart, L., & Lamort-Bouche, M. (2019). Interventions developed with the intervention mapping protocol in work disability prevention: a systematic review of the literature. Journal of occupational rehabilitation, 29, 11-24.
47. Garba, R. M., & Gadanya, M. A. (2017). The role of intervention mapping in designing disease prevention interventions: A systematic review of the literature. PloS one, 12(3), e0174438.
48. Lamort-Bouché M, Sarnin P, Kok G, Rouat S, Peron J, Letrilliart L, et al. Interventions developed with the Intervention Mapping protocol in the field of cancer: a systematic review. Psycho-oncology. 2018;27(4):1138-1149.
49. Saulle R, Sinopoli A, Baer ADP, Mannocci A, Marino M, De Belvis AG, et al. The PRECEDE–PROCEED model as a tool in Public Health screening: a systematic review. La Clinica Terapeutica. 2020;171(2):e167-e177.

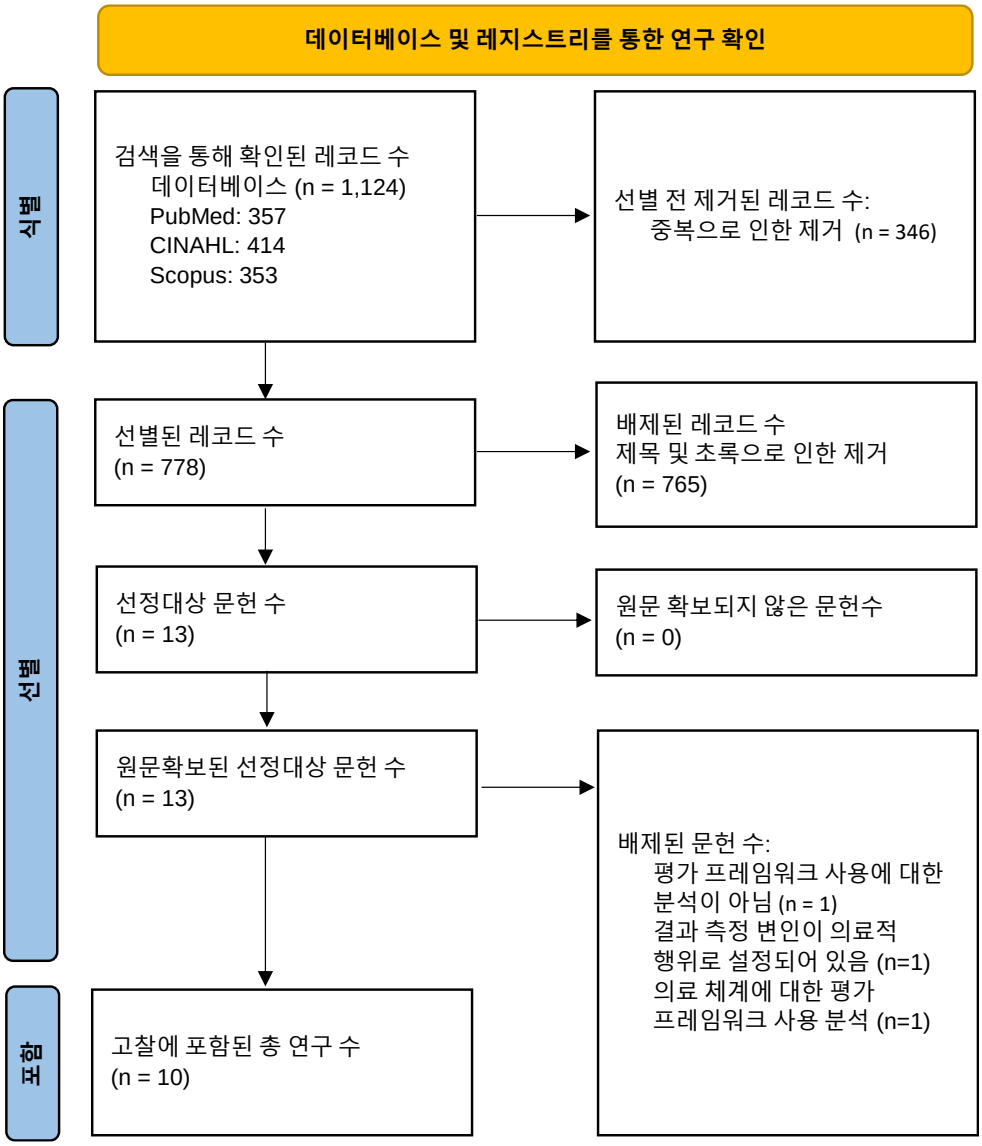

Figure 1. Study diagram

Table 2. Characteristics of selected studies

Author, year	Review type	Studies included in the systematic review	Target population	Target framework	Program settings	Target behavior/status
D'Lima et al., 2022	Systematic review	Articles included (n=157), in-depth analysis (n=15)	N/A	RE-AIM	N/A	multiple behavioral outcomes (physical activity and fruit & vegetable consumption) (n=1), physical activity (n=5), cardiovascular disease (n=2), smoking cessation (n=1), palliative care (n=1), prostate cancer (n=1), healthy eating (n=2), tobacco smoke exposure in children (n=1), diabetes (n=1)
Fynn et al., 2020	Systematic review	69	children or young adults (54%), adults (35%), older people (7%), multiple population groups (4%)	Mixed	schools (40%), health care (20%), workplace (6%), community (35%)	physical activity

Gaglio et al., 2013	Systematic review	71	variety of populations	RE-AIM	community or policy settings (25.4%), primary care (22.5%), health care or hospitals (5.6%), schools (2.8%), other or non- specified settings (43.7%)	physical activity & obesity (36.6%), disease self-management (29.6%) tobacco or substance abuse (9.9%), mental health & dementia (1.4%), cancer prevention (1.4%), other or multiple topics (14.1%)
Harden et al., 2015	Systematic review	101	variety of populations	RE-AIM	individual-level (57%), setting-level (14%), both (26%), individuals clustered within a setting (i.e., athletes on a team and church members within a congregation) (2%)	multiple behavioral outcomes (32%), smoking/substance abuse (15%), physical activity (10%), disease self-management (5%), diet (5%), weight (2%), other (12%) or no targeted behavioral outcome (19%)

					cluster unit (n=5), community (n=7), primary care (n=2), Internet (n=1), general practice (n=1), antenatal clinic (n=1)	
Johansson et al., 2022	Systematic review	17	clinical or non-clinical adults	Mixed		sedentary behavior
Kim et al., 2022	Systematic review & Meta-analysis	SR=26, MA=11	patients (n =11), parents of asthmatic children (n = 1), students (n = 6), health care providers (n =3), community-dwelling participants (n = 5)	Precede-Proceed model	hospitals (n=12),communities-schools or universities (n=6), health centers (n=4), community pharmacies (n=1), military installation (n=1), district (n=1), vocational program site (n=1)	symptom or disease management (n=10), health behavior promotion (n=10), psychological health (n=3), improvement in quality of life (n=2), dental examination techniques (n=1)

Murta et al., 2007	Systematic review	84	workers	Process evaluation for public health by Linann and Steckler	health care (33%), educational (17%), industrial (17%), communication organizations (11%), public agency (6%), finance (6%), transport (4%), army (2%), not mentioned (4%)	individual interventions (78.8%), occupational stress management program (interface stress management interventions) (17.3%), organization-focused interventions (3.8%), combined (7.7%)
Roozeboom et al., 2021	Systematic review	38 articles on 12 interventions	workers	Intervention Mapping	occupational settings	Occupational risk prevention and health promotion (weight gain prevention and / or physical activity promotion in workplace, influenza vaccination of workers, worker's safety, reduction of quartz exposure, mental health-related outcomes, recovery and relaxation, work engagement and mental health)

Wierenga et al., 2013	Systematic review	22	employees aged 18-65 years	Mixed	worksite-settings	physical activity (63%), healty nutrition (55%), smoking and tobacoo use (23%), relaxation (9%)
Yearly et al., 2012	Systematic review	67	congregation members (71.6%) / community members (17.9%)	Process evaluation for public health by Linann and Steckler	church-based settings	general health (11%), cardiovascular health (19%), cancers (25%), mental health (8%), nutrition/weight control (19%), smoking (3%), diabetes (3%), other/not specified (12%)

Table 3. Use of framework and summary and effectiveness of study result

Author, year	Target Framework	Use of Framework	Summary/Effectiveness
D'Lima et al., 2022	RE-AIM	Studies applied all five RE-AIM dimensions (69.0%), four (13.5%), three (6.5%), two (5.2%), one (5.8%) Reach (92.9%), Implementation (90.3%), Adoption (89.7%), Effectiveness (84.5%), Maintenance (77.4%) were used respectively	not reported criteria of RE-AIM: reach – exclusion criteria effectiveness – measures of a primary outcome in relation to a public health goal adoption – exclusion criteria of delivery staff implementation – percentage of delivery (fidelity) maintenance – robust data related to sub-group effects over the long-term
Fynn et al., 2020	Mixed	RE-AIM (n=27), Developing a process evaluation plan (n=12), Process evaluation for public health (n=10), MRC Guidance on evaluation of complex interventions (n=8), MRC Guidance on process evaluation (n=8), Logic Model (n=7), Realist Evaluation (n=4), Precede-Proceed (n=3), Intervention Mapping (n=2), Outcome Model (n=2), CDC Framework (n=1), Evaluation: a Systematic Approach (n=1), Model of Implementation (n=1), WHO Process	51/69 studies explicitly used stated evaluation framework (74%), 26/69 studies provided detailed descriptions (38%)

		Evaluation Workbook (n=1), Swiss Model for Outcome Classification (n=1), Concepts in process evaluation (n=1)	
Gaglio et al., 2013	RE-AIM	reach (91.5%), implementation (90.1%), effectiveness (77.5%), adoption at the setting-level (75.3%), staff-level adoption (74.6%), setting-level maintenance (71.8%), individual-level maintenance (64.8%)	none of the studies included in the review addressed all 34 criteria of 5 dimensions
Harden et al., 2015	RE-AIM	reach: all four indicators (17%) effectiveness: all five indicators (1.2%) individual-level maintenance: all three indicators (0%) setting-level adoption: all six indicators (1.2%) implementation: all three indicators (1.2%) organizational-level maintenance: all three indicators (0%)	reach: average participation rate (45%) effectiveness: positive changes in the primary outcome (89%), reported broader outcomes (13.4%) individual-level maintenance: effects of ≥6 months post-program (11%) adoption: average setting and staff adoption rates (75% and 79% respectively) implementation: interventions delivered as intended (82%) intervention reported adaptations to delivery (22%)

Johansson et al., 2022	Mixed	RE-AIM (n = 4), MRC Guidance (n =4), Process evaluation framework by Steckler & Linnan (2002) (n = 3), Saunders et al. (2005) (n = 1), Grant et al. (2013) (n = 1), WHO (2001) (n = 1), not specified (n = 9)	review and synthesis by MRC Guidance: barriers, facilitators, intervention delivery, contextual factors, and individual factors were the ones that affect changes in sedentary behavior 3/17 studies (not applied theoretical framework) reported a significant reduction in sedentary behavior (17.6%)
Kim et al., 2022	PRECEDE-PROCEED	planning (30.8%), planning and implementation (7.7%), planning, implementation, evaluation (61.5%)	significant effect on changes in primary variable (knowledge)
Murta et al., 2007	Process evaluation for public health by Linann & Steckler	recruitment (n = 27), dose received (n = 20), attitudes toward intervention (n = 17), reach (n = 12), context (n = 8), fidelity (n =5), dose delivered (n = 2)	not systematically addressed the relationship between process evaluation and outcome evaluation
Roozeboom et al., 2021	Intervention Mapping	step1 - planning group (n=2), needs assessment (n=12), causal pathways to describe the logic model of the problem (n=1), participative approach (n=12) step2 - constructing matrices of change objectives (n=5), participative approach (n=12), differentiation between behavioral and environmental factors (n=12)	process evaluation components assessed: (1) reach: excellent (37.5%), unsatisfactory (62.5%) (2) dose delivered: excellent (75%), moderate (25%), (3) dose received: excellent (9.1%), moderate (63.6%), unsatisfactory (27.3%) (4) fidelity: satisfactory (12.5%), moderate (87.5%)

		step3 - choose theory and evidence-based change methods (n=12), participative approach (n=10), step4 - participative approach (n=10), worker and workplace component of intervention (n=12), step5 - identify potential program users (n=12), state outcomes and performance objectives for program use(implementation planning) (n=9), design implementation interventions (n=12), participative approach (n=9)	(5) satisfaction: excellent (10%), satisfactory (80%), moderate (10%) No association was found between either the IM fidelity and the intervention implementation or the IM fidelity and the intervention effects
Wierenga et al., 2013	Mixed	dose received (82%), dose delivered (68%), context (68%), fidelity (41%), satisfaction (41%), reach (32%), recruitment (23%), maintenance (9%)	high dose and fidelity was positively associated with a change in primary outcomes
Yearly et al., 2012	Process evaluation for public health by Linann and Steckler	recruitment (88.1%), reach (80.6%), context (34.3%), dose delivered (28.4%), dose received (27.3%), implementation (20.9%), fidelity (9%)	process evaluation components assessed: smoking (4.0 ± 1.4), cardiovascular health (3.5 ± 1.3), weight control/nutrition (3.2 ± 1.0), cancer (3.1 ± 0.9), diabetes (2.5 ± 0.7), mental health (2.2 ± 1.3), general health (2.1 ± 1.3) number of process evaluation components evaluated in the selected studies did not vary by program features