ELSEVIER

Journal of Clinical Epidemiology ■ (2015) ■

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

A newly developed tool for classifying study designs in systematic reviews of interventions and exposures showed substantial reliability and validity

Hyun-Ju Seo[a], Soo Young Kim[b,*], Yoon Jae Lee[c], Bo-Hyoung Jang[d], Ji-Eun Park[e], Seung-Soo Sheen[f], Seo Kyung Hahn[g]

[a]*Department of Nursing, College of Medicine, Chosun University, Gwangju, Republic of Korea*
[b]*Department of Family Medicine, College of Medicine, Hallym University, Seoul, Republic of Korea*
[c]*Department of Oriental Gynecology, CHA Bundang Medical Center, CHA University, Seongnam, Republic of Korea*
[d]*Department of Preventive Medicine, College of Korean Medicine, Kyung Hee University, Seoul, Republic of Korea*
[e]*Department of Health Care Management and Policy, School of Public Health, Seoul National University, Seoul, Republic of Korea*
[f]*Department of Pulmonology, Ajou University Hospital, Suwon, Republic of Korea*
[g]*College of Medicine, Seoul National University, Seoul, Republic of Korea*

Accepted 16 September 2015; Published online xxxx

Abstract

Objective: To develop a study Design Algorithm for Medical Literature on Intervention (DAMI) and test its interrater reliability, construct validity, and ease of use.

Study Design and Setting: We developed and then revised the DAMI to include detailed instructions. To test the DAMI's reliability, we used a purposive sample of 134 primary, mainly nonrandomized studies. We then compared the study designs as classified by the original authors and through the DAMI. Unweighted kappa statistics were computed to test interrater reliability and construct validity based on the level of agreement between the original and DAMI classifications. Assessment time was also recorded to evaluate ease of use.

Results: The DAMI includes 13 study designs, including experimental and observational studies of interventions and exposure. Both the interrater reliability (unweighted kappa = 0.67; 95% CI [0.64−0.75]) and construct validity (unweighted kappa = 0.63, 95% CI [0.52−0.67]) were substantial. Mean classification time using the DAMI was 4.08 ± 2.44 minutes (range, 0.51−10.92).

Conclusions: The DAMI showed substantial interrater reliability and construct validity. Furthermore, given its ease of use, it could be used to accurately classify medical literature for systematic reviews of interventions although minimizing disagreement between authors of such reviews.

Keywords: Research design; Intervention studies; Reliability; Validity; Systematic review; Algorithm

1. Introduction

Numerous systematic reviews of health-related interventions have been carried out in the last two decades. Systematic reviews are the cornerstone of evidence-based health care, either on their own or through their incorporation into clinical practice guidelines or evidence briefings for evidence-based practice and policy making, respectively [1].

When conducting systematic reviews, researchers must appropriately categorize studies to effectively make decisions regarding study eligibility and exclusion criteria, assessment of risk of bias, consolidation of study results through qualitative synthesis or quantitative pooling, interpretation of findings, and grading the quality of evidence [2]. Studies are typically categorized according to complex characteristics, such as populations, eligibility criteria, study processes, and analyses [3]. The Cochrane Collaboration review groups mainly include randomized controlled trials aimed at evaluating the treatment effects of interventions [4]; therefore, they find it relatively unimportant to

Conflicts of interest: None.

Funding: This study was supported by a grant from the Health Insurance Review and Assessment Service (HIRA), Republic of Korea.

* Corresponding author. Kangdong Sacred Heart Hospital, 445 Gil-Dong, Gangdong-Gu, Seoul 134-814, Republic of Korea. Tel.: +82-2-2224-2406; fax: +82-2-2224-2409.

E-mail address: hallymfm@gmail.com (S.Y. Kim).

http://dx.doi.org/10.1016/j.jclinepi.2015.09.013
0895-4356/

What is new?

Key finding

- This study demonstrates the reliability and validity of the newly developed Design Algorithm for Medical Literature on Intervention (DAMI), which provides clear instructions for classifying study designs during systematic reviews on the effects of health interventions.
- The results showed that the DAMI has high interrater agreement and construct validity.

What this adds to what was known?

- Because the DAMI has good interrater reliability and construct validity, authors of systematic reviews of interventions and exposures will be able to use it to correctly classify various study designs.

What is the implication and what should change now?

- The DAMI could be an effective means of classifying medical literature for systematic reviews of interventions because it provides detailed instructions and minimizes disagreement between authors of such reviews.

classify studies by their designs. However, current comparative effectiveness research has begun synthesizing the use of different research methodologies to compare health interventions and identify the most effective medical interventions [5]. Alongside this advancement in comparative effectiveness reviews, the inclusion of nonrandomized studies in evidence synthesis has increased [6]. Therefore, it is becoming increasingly more important to be able to accurately classify various study designs and sort research evidence [7].

Although tools exist to help researchers and health care professionals classify studies [8−11], they can be difficult to apply to systematic reviews because they lack detailed guidance and are often insufficiently validated. The recent typology developed by Hartling et al. [3] was found to have moderate reliability and low accuracy, rendering it difficult to use for classifying study designs in the literature. Therefore, there is a need for a user-friendly typology to enable appropriate classification of study designs during systematic reviews on the effects of health interventions.

This study aims to (1) develop a new classification algorithm with clear instructions and that considers both experimental and observational study designs; (2) determine the algorithm's interrater reliability and validity; and (3) assess the algorithm's ease of use.

2. Methods

2.1. Development of the study Design Algorithm for Medical Literature on Intervention

2.1.1. Developing the draft algorithm

The main principles of development were as follows: (1) establish the ontology for classifying intervention studies; (2) create a general tool to use across different study designs, including experimental and observational studies; (3) develop an algorithm that is easy to understand; (4) develop a tool for categorizing studies wherein variables are measured at the individual level; and (5) include detailed instructions to reduce variation between reviewers. To develop this algorithm, we reviewed previous typologies by Hartling et al. [3], Zaza et al. [9], and West et al. [10], and tools used in some health technology assessment agencies, such as the National Institute for Health and Clinical Excellence (NICE) [11], US Preventive Services Task Force [12], and the Cochrane Collaboration [13]. Furthermore, we reviewed several epidemiology textbooks [14−16] and then drafted an algorithm with detailed instructions on how to reduce variance among reviewers. Three consultative meetings with researchers of epidemiology, statistics, and systematic review methodology were held to discuss the first draft algorithm. Each item's contents were examined and revised on the basis of the researchers' consensus.

2.1.2. Collecting opinions regarding the unpublished draft algorithm

Ten Korean researchers gathered data on users' opinions and experiences regarding the problems and weaknesses of the unpublished draft algorithm after it had been used in several health technology assessment institutions, including the National Evidence-based Healthcare Collaborating Agency and the Health Insurance Review and Assessment Service in South Korea. The first version of the Design Algorithm for Medical Literature on Intervention (DAMI) was also presented and discussed at an evidence-based health care forum, wherein experts in evidence-based health care address and discuss current methodological issues in medical research in South Korea.

2.1.3. Creating the final version of the DAMI

After revising the first draft algorithm on the basis of the aforementioned feedback, we obtained comments from two international clinical epidemiology experts. Users' experiences and the advice of local and international experts on evidence-based health care were incorporated into the final version of the DAMI.

2.2. Validation

We chose 134 studies to cover all relevant study designs that can be identified using the DAMI from among nine systematic reviews and health technology assessment