

# Research Designs in Mesh and Emtree: A Comparative Study of Coverage

Tanja Bekhuis, PhD, MS, MLIS<sup>1</sup>, Dina Demner-Fushman, MD, PhD<sup>2</sup>,  
Rebecca Crowley, MD, MS<sup>1</sup>

<sup>1</sup>Department of Biomedical Informatics, University of Pittsburgh, PA; <sup>2</sup> Lister Hill National Center for Biomedical Communications, US National Library of Medicine, Bethesda, MD

## Abstract

To aid retrieval of studies common to comparative effectiveness research (CER), we developed a terminology based on the local terminologies of several organizations. We compared coverage of the new terminology in MeSH and Emtree, and developed a crosswalk between the two controlled vocabularies. Patterns of coverage were similar and partial matches predominated. Negated or detailed terms were rarely matched exactly. For unmapped or partially mapped designs, records were retrievable if a substring in a design query matched the language of scientists.

## Introduction

Retrieving CER studies is problematic because the language for common designs is nonstandard and inconsistently indexed<sup>1</sup>. To support expert searchers who want to find such studies, we developed a design terminology; compared term coverage in MeSH and Emtree, the controlled vocabularies for Medline and Embase; developed a crosswalk between MeSH and Emtree to aid retrieval; and considered whether scientists use CER design language.

## Methods

We extracted terms from the local terminologies (LTs) of 5 CER organizations, including the US Agency for Healthcare Research and Quality, the University of Alberta Evidence-based Practice Center (EPC), the Cochrane Collaboration, the Academy of Nutrition and Dietetics, and the RTI-UNC EPC. The new terminology consists of the union of terms. Coverage of designs was defined by the distribution over match type (exact, partial, or no match) in MeSH and Emtree. A crosswalk was developed by recording terms to which designs mapped in both controlled vocabularies. Additionally, queries restricted to titles and abstracts were run in Embase to explore the correspondence between CER terms and scientists' language.

## Results

The union of terms mostly consisted of primary study designs and a few terms useful for evaluating evidence, such as *opinion paper* and *systematic review*. About half the terms appeared in just one LT (47.44%; 37/78); a few terms were common to all (8.97%; 7/78), including *before-after study*, *case-control study*, *case series*, *cross-sectional study*, *prospective cohort study*, *retrospective cohort study*, and *randomized controlled trial*. Patterns of coverage were similar in MeSH and Emtree ( $\gamma=.581$ ,  $P=.002$ ) (Table 1).

**Table 1.** Coverage of CER design terms by controlled vocabulary.

Type of match	MeSH	Emtree	MeSH to Emtree Ratio
No match	14 (17.9%)	7 (9.0%)	2.00 (14/7)
Partial	49 (62.8%)	45 (57.7%)	1.09 (49/45)
Exact	15 (19.2%)	26 (33.3%)	0.58 (15/26)
Total	78 (100%)	78 (100%)	

## Discussion

The new CER design terminology and its crosswalk may be useful for expert searchers. For partially mapped designs, queries could consist of free text for modifiers such as *nonrandomized* or *interrupted* added to broad or related controlled terms (if they exist). Queries for unmapped terms require free text by necessity.

## References

1. Reeves B, Deeks J, Higgins J, Wells G. Including non-randomized studies. In: Higgins J, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions. Chichester (UK): Wiley; 2008.