The Estimating effectiveness from efficacy taxonomy (EFFECT): A tool to estimate the real-world impact of health interventions

Karla Galaviz Arredondo, *Emory University*
Mohammed Ali, *Emory University*
Jeehea Haw, *Emory University*
Matthew Magee, *Emory University*
Alysse Kowalski, *Emory University*
Jingkai Wei, *University of North Carolina, Chapel Hill*
Audrey Straus, *CommUnityCare Health Centers*
Mary Weber, *Emory University*
Theo Vos, *University of Washington*
Christopher Murray, *University of Washington*

Only first 10 authors above; see publication for full author list.

Journal Title: DIABETES RESEARCH AND CLINICAL PRACTICE

Volume: Volume 159

Publisher: ELSEVIER IRELAND LTD | 2020-01-01, Pages 107751-107751

Type of Work: Article | Post-print: After Peer Review

Publisher DOI: 10.1016/j.diabres.2019.05.030

Permanent URL: https://pid.emory.edu/ark:/25593/vrp40

Final published version: http://dx.doi.org/10.1016/j.diabres.2019.05.030

HHS Public Access

Author manuscript

Diabetes Res Clin Pract. Author manuscript; available in PMC 2021 January 01.

Published in final edited form as:

Diabetes Res Clin Pract. 2020 January ; 159: 107751. doi:10.1016/j.diabres.2019.05.030.

The Estimating Effectiveness From Efficacy Taxonomy (EFFECT): a tool to estimate the real-world impact of health interventions

Karla I. Galaviz, PhD[a], **Mohammed K. Ali, MD**[a], **Jeehea Sonya Haw, MD**[b], **Matthew James Magee, PhD**[c], **Alysse Kowalski, MPH**[a], **Jingkai Wei, MPH**[d], **Audrey Straus, MSN**[e], **Mary Beth Weber, PhD**[a], **Theo Vos, PhD**[f], **Christopher Murray, PhD**[f], **K.M. Venkat Narayan, MD**[a]

[a]Hubert Department of Global Health, Emory University, Atlanta, GA 30322 USA

[b]Department of Endocrinology and Metabolism, Emory University School of Medicine Atlanta, GA 30307 USA

[c]School of Public Health, Georgia State University, Atlanta GA 30302 USA

[d]Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA

[e]CommUnityCare Health Centers, Austin, TX 78758 USA

[f]Institute for Health Metrics and Evaluation, University of Washington, Seattle WA 98121 USA

Abstract

Aims: To develop and pilot test a taxonomy to empirically estimate health intervention effectiveness from efficacy.

Methods: We developed a taxonomy to score health interventions across 11 items on a scale of 0-100. The taxonomy was pilot-tested in efficacy and effectiveness diabetes prevention studies identified in two separate systematic reviews; here, the face validity, inter-rater reliability and factor structure of the taxonomy were established. Random effects meta-analyses were used to

Corresponding author: Karla I. Galaviz, Hubert Department of Global Health, Rollins School of Public Health, Emory University, #1518 Clifton Rd, Atlanta, GA 30322, CNR Building Room 7040, kgalavi@emory.edu, Tel: 404-202-5802.

AUTHOR CONTRIBUTIONS

KIG conducted the systematic review for diabetes effectiveness studies, helped develop and test the taxonomy, conducted the data analyses, provided interpretation of study findings, drafted the manuscript, and has given final approval of the manuscript. MKA conceptualized and designed the study, helped develop and test the taxonomy, provided guidance with data analyses and the interpretation of study findings, helped drafting the manuscript, and has given final approval of the manuscript. JSH and MJM conducted the systematic review for diabetes efficacy studies, helped develop and test the taxonomy, critically revised the manuscript, and have given final approval of the manuscript. AK and JW helped develop and test the taxonomy, critically revised the manuscript, and have given final approval of the manuscript. AS helped conducting the systematic review for diabetes effectiveness studies, helped test the taxonomy, critically revised the manuscript, and has given final approval of the manuscript. MBW helped develop and test the taxonomy, critically revised the manuscript, and has given final approval of the manuscript. TV and CM helped conceptualize the taxonomy, guided data analyses and interpretation of study findings, critically revised the manuscript, and have given final approval of the manuscript. KMVN helped conceptualize and design the study, helped develop the taxonomy, provided guidance with data analyses and interpretation of study findings, helped drafting the manuscript, and has given final approval of the manuscript.

obtain weight loss and diabetes incidence pooled effects across studies. These effects and taxonomy scores were used to down calibrate efficacy estimates to effectiveness estimates as follows: Efficacy effect*[Efficacy score/highest possible score].

Results: We scored 82 effectiveness lifestyle modification studies (mean score 49.2), 32 efficacy lifestyle modification studies (mean score 69.8) and 20 efficacy studies testing medications (mean score 77.4). The taxonomy had face validity and good inter-rater reliability (ICC=0.9 [0.87, 0.93]). The between-groups down calibrated weight loss estimate was similar to that observed in the effectiveness meta-analysis (1.7 and 1.8 kg, respectively). The down calibrated diabetes relative risk reduction estimate was also similar to that observed in the effectiveness meta-analysis (30.6% over 2.7 years and 29% over 2 years, respectively).

Conclusions: The taxonomy is a promising tool to estimate the real-world impact of health interventions.

Keywords

effectiveness metrics; implementation science; global health

INTRODUCTION

Gaps in implementation of proven interventions are legion, and result in considerable, potentially avoidable, ill-health [1]. Diabetes and cardiovascular diseases (CVD) for instance, can be prevented or at least delayed for many years [2-5]; yet they are responsible for 1.3 and 17.3 million annual deaths, respectively, globally [6]. Evidence summaries show diabetes and CVD can be prevented through lifestyle modification or medications [2-5, 7, 8], but these research findings often do not translate into real-world practice and therefore fail to achieve the potential population-level health improvements suggested by studies [1, 9].

Most evidence of health intervention effects is derived from efficacy studies that test treatment effects under optimal conditions [10]. These studies favor internal validity (i.e. unbiased causal inference) and often overestimate the effects that can be accrued from application of health interventions under real-world conditions, or effectiveness [9, 10]. Effectiveness studies place more weight on external validity (i.e. generalizability) to better resemble the real-world impact of health interventions but they are more difficult to get funded/conduct and therefore less commonly done. Indeed few health conditions have an effectiveness evidence base to draw conclusions about real-world impact; thus, empirical methods to estimate effectiveness holds great potential for informing health intervention research and decision making.

We developed a method to estimate effectiveness of health interventions that can be derived from down calibrating efficacy estimates. Unlike existing approaches [11-16], our empiric approach optimizes the use of existing efficacy data to estimate potential real-world impact of health interventions. Our method assesses the extent to which a health intervention study is focused on establishing internal or external validity and is designed to use this information to down calibrate efficacy estimates. Herein we describe the development, validation and testing of this tool using diabetes prevention interventions as a case example.

METHODS

We developed the **E**stimating E**F**fectiveness **F**rom **E**ffi**C**acy **T**axonomy (EFFECT) to empirically estimate health intervention effectiveness from efficacy data. The development, testing and validation of the taxonomy comprised the following three steps: 1) development of taxonomy items and scoring system; 2) taxonomy testing among efficacy and effectiveness diabetes prevention studies; and 3) testing and validation of the efficacy-to-effectiveness down calibration approach. Below we provide details about the methods used in each step.

1.1 Taxonomy Development

We developed a taxonomy and associated scoring system that allows us to quantitatively and qualitatively characterize to what degree data in studies reflect efficacy or effectiveness focus. The taxonomy comprises a set of items and domains that were developed iteratively over a 3-month period as follows. A group of three experts in implementation science, epidemiology and health intervention research conducted a literature review to identify intervention characteristics that were linked to internal validity and those linked to external validity; descriptive documents, reviews and seminal manuscripts were reviewed. The team then met to discuss the identified internal/external validity characteristics and selected those deemed most relevant for health interventions. Then, the team held a second meeting and, based on the identified characteristics, developed a set of 11 items and operational definitions.

A final meeting was held to develop a system to score each of the 11 items. Based on expert judgment, each item was assigned a scoring scale ranging from 1 to 10 for items related to study, intervention and population characteristics, 1 to 5 for data analysis and outcomes assessed items, and 4 to 10 for the analytical method used item. Data analysis and outcomes assessed items were given lower scores because they were judged as having lower impact on overall effect estimate than the other items; a scale of 4 to 10 was used for the analytical method item to set the maximum total score at 100. Individual item scores can be summed to obtain composite score ranging from 1 to 100, where higher scores indicate a focus on internal validity (i.e. understanding disease etiology or progression, or proof of principle that an intervention works) and lower scores a focus on external validity (i.e. proof that an intervention works in real-life, is generalizable, acceptable, and can be applied at scale). The taxonomy items and scoring scales are presented in Table 1. Operational definitions for each item and instructions on how to use the taxonomy and scoring system are included in Supplementary File 1.

1.2 Taxonomy Testing

The taxonomy was tested among diabetes prevention studies for which there is a well-established efficacy and effectiveness evidence base. To identify published diabetes prevention efficacy and effectiveness intervention studies, we conducted two separate systematic reviews – one aimed at identifying effectiveness studies and another aimed at identifying efficacy studies. Details about these systematic reviews are reported elsewhere [2, 17]; there was no overlap between studies included in the efficacy and effectiveness

systematic review. The identified studies were scored using the taxonomy and the face validity, reliability and factor structure of the tool were assessed as described below.

To establish the face validity of the taxonomy, a team of experts (n=7) pilot-tested it on three randomly selected studies and met to discuss operational definitions and scoring instructions. This was aimed at promoting interpretation consistency and at harmonizing the scoring process. Based on team discussions, operational definitions for each taxonomy item were improved (i.e. more detailed item definitions and instructions on how to select a specific score were added) and three additional studies were scored. In a subsequent meeting, the study team compared their scores, discussed whether the taxonomy was accurately measuring internal and external validity characteristics within a study, and concluded that the taxonomy had face validity.

Using the finalized taxonomy, each of the identified published studies was scored twice by random pairs of raters from the study team. Scores from each pair of raters were used to assess the inter-rater reliability of the taxonomy via the intra-class correlation coefficient. Values above 0.75 were deemed as indicative of good reliability [18]. Using the average of the two sets of scores, we computed means and standard deviations to explore the distribution of taxonomy scores. Pearson correlations coefficients among taxonomy items were obtained and used to conduct an exploratory factor analysis. Factors were extracted using the maximum likelihood method and rotated using the direct oblimin rotation [19]. The objective of this factor analysis was to explore the taxonomy factor structure and its influence on each item rather than eliminating items. The factor structure yielding factors with eigenvalues ≥1 and taxonomy items with factor loadings ≥0.4 was considered a parsimonious factor solution. These analyses were conducted using the Statistical Package for Social Sciences version 22 for Windows (IBM, Armonk NY).

1.3 Efficacy to Effectiveness Down Calibration

To down calibrate efficacy estimates to effectiveness estimates, we developed the following formula: Effectiveness estimate = Efficacy effect × (Efficacy score / Highest possible score). The formula was tested among efficacy studies and validated among effectiveness studies.

Random effects meta-analysis weighted by the inverse variance were conducted separately for efficacy and effectiveness lifestyle modification studies. We obtained within-subjects and between-groups pooled effects for two outcomes that represent the two sides of the *Outcomes* item scale in our taxonomy: weight change at 12 months, representing the more proximal/practical outcome, and diabetes incidence representing the more distal/intensive outcome. Using our formula, the efficacy pooled estimate and taxonomy score were then used to down calibrate effect estimates. The obtained down calibrated effect was then compared against that obtained in the effectiveness meta-analysis for each outcome.

To explore the influence of taxonomy score on effect heterogeneity, meta-regressions were conducted including taxonomy score as predictor of weight change and diabetes risk reduction for efficacy and effectiveness studies separately. These analyses were conducted using the Metafor package [20] in R programming language (version 3.2.1).

RESULTS

2.1 Taxonomy Testing

The systematic reviews identified 134 diabetes prevention studies. Of these, 82 were effectiveness studies testing lifestyle modification interventions, 32 efficacy studies testing lifestyle modification interventions and 20 efficacy studies testing medications. The item and total scores for each group of studies are presented in Table 2. In line with the taxonomy scoring system, efficacy studies had scores over 50 and effectiveness studies had scores under 50.

The inter-rater reliability of the taxonomy total score was ICC=0.9 (0.87, 0.93), suggesting good inter-rater reliability. Among individual item correlations, the items “analysis indicators” and “emphasis on adherence” showed the lowest intra-class correlation coefficients (ICC=0.60 and ICC=0.56, respectively; Table 3). Items that were often not reported included emphasis on adherence, intervention delivery setting, and analytical method employed; these were not reported in 31%, 28% and 17% of the articles, respectively. When items were not reported, raters assigned scores based on proxy information reported in the article. For instance, information on intervention provider (e.g. healthcare professional) and treatment characteristics (e.g. individual counseling) was used to determine the intervention delivery setting (e.g. healthcare setting). Similarly, information on study completion (e.g. attrition rates) or treatment adherence (e.g. pill counts, attendance lists, incentives) was used to determine if a study placed high emphasis on adherence (e.g. low attrition rates indicating high emphasis). Finally, the number of participants included in statistical analyses provided information on the analytical method employed (e.g. analysis including non-completers was deemed as intent to treat, while those including completers was deemed as per protocol).

Strong correlations were observed among the items study goal, study design and focus of outcomes (r 0.75 to 0.83), and between the items recruitment setting and intervention delivery setting (r=0.79; Table 3). The factor analysis showed that two main factors explained 64% of the variance across the 11 taxonomy items–study objective (eigenvalue=5.9) and study population (eigenvalu=1.2). Overall, eight items loaded onto study objective (factor loadings ranging from 0.32 to 0.95) and three items onto study population (factor loadings ranging from 0.58 to 0.88; Table 4).

2.2 Down Calibration Validation

Overall, 13 efficacy lifestyle modification studies reported weight loss at 12 months (Table 5). The pooled weight loss within participants receiving an intervention was −3.0 kg (−4.5, −1.4) and the pooled weight loss difference between intervention and control participants was −2.4 kg (−4.3, −0.6). Diabetes incidence was reported in 16 efficacy studies over a mean follow up of 2.7 years; the pooled diabetes relative risk reduction was 44% (RR=0.56 [0.48, 0.65]). To down calibrate these efficacy estimates to effectiveness estimates, we applied our formula as follows:

- Within subjects weight change: 3.0 × (69.5 / 100) = 2.1 kg

- Between subjects weight change: 2.4 × (69.5 / 100) = 1.7 kg
- Diabetes relative risk reduction: 44 × (69.5 / 100) = 30.6%

These down calibrated estimates were compared to those obtained in the random effect meta-analyses of effectiveness studies (Table 5). The within-subjects down calibrated weight loss estimate was lower than the observed effectiveness estimate (−2.1 and −2.5 kg, respectively), while the between-subjects down calibrated estimate was similar to the observed effectiveness estimate (1.7 and 1.8 kg, respectively). The down calibrated relative risk reduction estimate was also similar to that observed in the effectiveness meta-analysis estimate over a mean follow up of 2 years (30.6% and 29%, respectively).

Meta-regressions including the taxonomy total score as predictor of weight change at 12 months within intervention participants were conducted for efficacy and effectiveness studies separately. For efficacy studies, the model explained 20% of the effect heterogeneity in weight change ($p<.0001$), where every increase in total score associated with an additional weight loss of 0.2 kg ($\beta=0.23$). For effectiveness studies, the model did not explain weight loss effect heterogeneity. Meta-regressions exploring total score as predictor of diabetes risk reduction did not explain effect heterogeneity in neither efficacy nor effectiveness studies.

DISCUSSION

Because few health conditions have an effectiveness evidence base to draw conclusions about real-world impact, empirical methods to estimate effectiveness from efficacy are needed. To address this, we developed EFFECT, a taxonomy to estimate the effectiveness of health interventions that can be derived from down calibrating efficacy estimates. EFFECT was found to be a reliable tool for assessing efficacy study characteristics (i.e. internal validity) and effectiveness study characteristics (i.e. external validity). Using taxonomy scores, we down calibrated efficacy intervention effects and found our approach yielded similar estimates to those obtained in meta-analyses of effectiveness studies. Though more testing in other interventions and outcomes is needed, these results show EFFECT is a promising, empirically verified tool to estimate the real-world impact of health interventions.

Our efficacy to effectiveness down calibration approach produced estimates that approached those observed in the effectiveness meta-analysis we conducted. For weight loss at 12 months, the within-subjects down calibrated estimate was lower than the observed effectiveness estimate (−2.1 and −2.5 kg, respectively) but close enough to make a conservative prediction about the potential real-world weight loss intervention participants can achieve. The between-groups down calibrated weight loss estimate was similar to the observed effectiveness estimate (1.7 and 1.8 kg, respectively), accurately reflecting the real-world weight loss participants receiving an intervention can achieve compared to those not receiving one. Our down calibrated weight loss estimates are also similar to within-subjects (2.1 to 2.3 kg) and between-groups (1.6 kg) estimates reported at 12-18 months in previous effectiveness meta-analyses of diabetes prevention interventions [21, 22].

Our down calibrated diabetes relative risk reduction estimate was also similar to that observed in the effectiveness meta-analysis we conducted (30.6% over a mean follow-up of 2.7 years and 29% over 2 years, respectively). This suggests our down calibration approach provides a close estimate of the real-world impact preventive interventions can have on diabetes risk reduction. Though our down calibrated estimate is higher than the 26% estimate observed at 1-1.5 years in a previous meta-analysis of effectiveness diabetes prevention studies [22], it provides a close estimate of the effect that could be achieved from applying diabetes prevention interventions in the real world.

Although other approaches have been proposed to assess the quality and real-world impacts of health interventions, none of these has empirically estimated effectiveness from efficacy. For instance, Cohen and colleagues (2015) developed a tool to measure research impact across four domains (corroboration, attribution, reach, and importance), where impact is classified into low, medium and high across scholarly, translational, policy or practice, and long-term population outcomes [16]. An approach based on the RE-AIM framework [23] proposes to multiply the mean of Reach and Effectiveness estimates to obtain a measure of individual-level impact, and to multiply the mean of Adoption and Implementation estimates to obtain a measure of setting-level impact [13]. The use of epidemiological methods has also been proposed to determine population impact based on the reach and risk reduction produced by an intervention [14]. Other approaches have focused on qualitatively evaluating the external validity of interventions [11] or on assessing the extent to which trials are pragmatic or explanatory [12]. In contrast, our empiric approach combines qualitative and quantitative assessments of internal and external validity factors that are used to derive estimates of potential real-world impact of health interventions.

Governments are facing difficulties in determining the real-world benefits of research given that current metrics are insufficient to determine its value [24]. Rather than guessing what real-world impact might be possible, our taxonomy can help optimize the use of existing efficacy data in decision-making where effectiveness data lack. With EFFECT, decision makers can determine, a priori, what effect could be achieved based on different permutations of taxonomy scores and cost inputs they can afford. For instance, variations in intervention intensity (e.g. higher dose versus usual dose), participant characteristics (e.g. high risk versus low risk) and outcomes assessed (e.g. behavior versus disease incidence) can be explored to obtain effect estimates. Such assessment can aid decision making in planning health interventions around cost and expected effects.

The present study represents the first step towards developing a practical decision-making tool that can be used to estimate the real-world impact of health interventions. We used diabetes prevention interventions to test the taxonomy given the robust efficacy and effectiveness evidence base available for this condition; however, the goal is for the taxonomy to be used across a range of health conditions and interventions. To achieve this, future work will focus on the external and longitudinal validation of the taxonomy for different health conditions, interventions and outcomes.

The present findings should be interpreted in light of the study limitations. Because the context and characteristics of populations receiving interventions may influence intervention

implementation and impact, a taxonomy-derived effect estimate may not be perfectly precise for different contexts or populations. Another limitation is that the taxonomy was developed and tested based on published data and is beholden to the level of detail provided by the authors of the original articles. Indeed some taxonomy items were poorly reported; still, we used other information reported in the articles to complete the scoring. Despite operational definitions and detailed scoring instructions were used, subjectivity and assumptions of the reader may have influenced scoring, which may bias taxonomy scores and ultimately effect estimates. However, we are unsure if the direction (over vs. under estimation) of this bias was the same in every case; assuming variation was random, our models could account for some over and under-estimation by providing uncertainties around intervention effect estimates. If a study obtains a score of 100, our down calibration formula would not work; other strategies should be explored in such case. Finally, we did not explore the impact of risk of bias in effect estimates; future applications of the taxonomy will explore this.

CONCLUSION

Because few health interventions have an effectiveness evidence base from where to draw conclusions about real-world impact, we developed a taxonomy to estimate effectiveness from efficacy data. The taxonomy was found to be a reliable tool for assessing efficacy study characteristics (i.e. internal validity) and effectiveness study characteristics (i.e. external validity). We developed an approach that uses taxonomy scores to down calibrate efficacy to effectiveness intervention effects; our approach yielded similar estimates to those obtained in meta-analyses of effectiveness studies. Though more testing in other interventions and outcomes is needed, these results show EFFECT is a promising, empirically verified tool to estimate the potential real-world impact of health interventions. Once externally validated, EFFECT could be used to guide health-related spending and to estimate the potential impacts of implementing interventions for common conditions more widely.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

FUNDING AND COMPETING INTERESTS

This project was funded by the Disease Control Priorities Network grant to the Institute for Health Metrics and Evaluation from the Bill & Melinda Gates Foundation (UWSC7007). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Authors KIG, MKA, KVN and MBW were partially supported by the Georgia Center for Diabetes Translation Research (P30DK111024). Authors declare no competing interests.

References

[1]. Glasgow RE and Emmons KM. How can we increase translation of research into practice? Types of evidence needed. Annu Rev Public Health. 2007; 28: 413–33. [PubMed: 17150029]

[2]. Haw JS, Galaviz KI, Straus AN, Kowalski AJ, Magee MJ, Weber MB, et al. Long-term sustainability of diabetes prevention approaches: A systematic review and meta-analysis of randomized clinical trials. JAMA Intern Med. 2017; 177(12): 1808–1817. [PubMed: 29114778]

[3]. Ettehad D, Emdin CA, Kiran A, Anderson SG, Callender T, Emberson J, et al. Blood pressure lowering for prevention of cardiovascular disease and death: A systematic review and meta-analysis. The Lancet. 2016; 387(10022): 957–967.

[4]. Fretheim A, Odgaard-Jensen J, Brors O, Madsen S, Njolstad I, Norheim O, et al. Comparative effectiveness of antihypertensive medication for primary prevention of cardiovascular disease: Systematic review and multiple treatments meta-analysis. BMC Medicine. 2012; 10(1): 33. [PubMed: 22480336]

[5]. Gillies CL, Abrams KR, Lambert PC, Cooper NJ, Sutton AJ, Hsu RT, et al. Pharmacological and lifestyle interventions to prevent or delay type 2 diabetes in people with impaired glucose tolerance: Systematic review and meta-analysis. BMJ. 2007; 334(7588): 299. [PubMed: 17237299]

[6]. Global Burden of Disease Collaborators. Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: A systematic analysis for the global burden of disease study 2013. The Lancet. 2015; 385(9963): 117–171.

[7]. Balk EM, Earley A, Raman G, Avendano EA, Pittas AG, and Remington PL. Combined diet and physical activity promotion programs to prevent type 2 diabetes among persons at increased risk: A systematic review for the community preventive services task force. Ann Intern Med. 2015; 163(6): 437–51. [PubMed: 26167912]

[8]. Merlotti C, Morabito A, and Pontiroli AE. Prevention of type 2 diabetes; a systematic review and meta-analysis of different intervention strategies. Diabetes Obes Metab. 2014; 16(8): 719–27. [PubMed: 24476122]

[9]. Glasgow RE, Lichtenstein E, and Marcus AC. Why don't we see more translation of health promotion research to practice? Rethinking the efficacy-to-effectiveness transition. American Journal of Public Health. 2003; 93(8): 1261–1267. [PubMed: 12893608]

[10]. Flay BR. Efficacy and effectiveness trials (and other phases of research) in the development of health promotion programs. Preventive Medicine. 1986; 15(5): 451–474. [PubMed: 3534875]

[11]. Green LW and Glasgow RE. Evaluating the relevance, generalization, and applicability of research: Issues in external validation and translation methodology. Evaluation & the Health Professions. 2006; 29(1): 126–153. [PubMed: 16510882]

[12]. Thorpe KE, Zwarenstein M, Oxman AD, Treweek S, Furberg CD, Altman DG, et al. A pragmatic–explanatory continuum indicator summary (precis): A tool to help trial designers. CMAJ : Canadian Medical Association Journal. 2009; 180(10): E47–E57. [PubMed: 19372436]

[13]. Glasgow RE, Klesges LM, Dzewaltowski DA, Estabrooks PA, and Vogt TM. Evaluating the impact of health promotion programs: Using the re-aim framework to form summary measures for decision making involving complex issues. Health Education Research. 2006; 21(5): 688–694. [PubMed: 16945984]

[14]. Koepsell TD, Zatzick DF, and Rivara FP. Estimating the population impact of preventive interventions from randomized trials. Am J Prev Med. 2011; 40(2): 191–8. [PubMed: 21238868]

[15]. Banzi R, Moja L, Pistotti V, Facchini A, and Liberati A. Conceptual frameworks and empirical approaches used to assess the impact of health research: An overview of reviews. Health Research Policy and Systems. 2011; 9(1): 26. [PubMed: 21702930]

[16]. Cohen G, Schroeder J, Newson R, King L, Rychetnik L, Milat A, et al. Does health intervention research have real world policy and practice impacts: Testing a new impact assessment tool. Health Research Policy and Systems. 2015; 13(1): 3. [PubMed: 25552272]

[17]. Galaviz KI, Weber MB, Straus A, Haw JS, Narayan KMV, and Ali MK. Global diabetes prevention interventions: A systematic review and network meta-analysis of the real-world impact on incidence, weight, and glucose. Diabetes Care. 2018; 41(7): 1526–1534. [PubMed: 29934481]

[18]. Koo TK and Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. Journal of Chiropractic Medicine. 2016; 15(2): 155–163. [PubMed: 27330520]

[19]. Costello AB and Osborne JW. Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis (2005). Practical Assessment, Research & Evaluation. 2005; 10(7).

[20]. Viechtbauer W. Conducting meta-analyses in r with the metafor package. Journal of Statistical Software. 2010; 36(3): 1–48.

[21]. Dunkley AJ, Bodicoat DH, Greaves CJ, Russell C, Yates T, Davies MJ, et al. Diabetes prevention in the real world: Effectiveness of pragmatic lifestyle interventions for the prevention of type 2 diabetes and of the impact of adherence to guideline recommendations: A systematic review and meta-analysis. Diabetes Care. 2014; 37(4): 922–33. [PubMed: 24652723]

[22]. Ashra NB, Spong R, Carter P, Davies MJ, Dunkley A, Gillies C, et al. A systematic review and meta-analysis assessing the effectiveness of pragmatic lifestyle interventions for the prevention of type 2 diabetes mellitus in routine practice [article online], 2015 Available from https://www.Gov.Uk/government/uploads/system/uploads/attachment_data/file/456147/phe_evidence_review_of_diabetes_prevention_programmes-_final.Pdf. Accessed march 13 2017. 2015.

[23]. Glasgow RE, Vogt TM, and 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. [PubMed: 10474547]

[24]. Bornmann L. Measuring the societal impact of research: Research is less and less assessed on scientific impact alone--we should aim to quantify the increasingly important contributions of science to society. EMBO Rep. 2012; 13(8): 673–6. [PubMed: 22777497]

Highlights

- We developed the Estimating EFfectiveness From EffiCacy Taxonomy (EFFECT).
- Pilot-testing shows EFFECT reliably identifies effectiveness and efficacy studies.
- EFFECT yielded down calibrated estimates that were similar to empirical estimates.
- EFFECT is a promising tool to estimate the real-world impact of interventions.

Table 1.

Taxonomy items and scoring scales.

1. Study stated goal	1	2	3	4	5	6	7	8	9	10
	Focused on generalizability (external validity)									Focused on causality (internal validity)
2. Study design employed	1	2	3	4	5	6	7	8	9	10
	Low causal inference (e.g. Pre/Post)									High causal inference (e.g. RCT)
3. Population inclusion criteria	1	2	3	4	5	6	7	8	9	10
	General public					High-risk populations				
	Less selective				*Highly selective*	*Less selective*				*Highly selective*
4. Population recruitment setting	1	2	3	4	5	6	7	8	9	10
	Community settings					Clinical settings				
	Heterogeneous				*Homogenous*	*Heterogeneous*				*Homogeneous*
5. Intervention delivery setting	1	2	3	4	5	6	7	8	9	10
	Communities; workplaces				Primary care		Secondary Hospital		Tertiary Hospital	
6. Intervention intensity	1	2	3	4	5	6	7	8	9	10
	Intervention similar to general care					Research (costly, intense) intervention				
	Commonly used; low threshold for side-effects					*High-dose or intensity; high threshold for side effects*				
7. Emphasis on intervention adherence	1	2	3	4	5	6	7	8	9	10
	No or low effort to ensure adherence					Focus on achieving high adherence				
	Nothing				*Reminders*	*Documenting adherence*				*Blood levels*
8. Data analysis indicators	1		2		3		4		5	
	Process-focused									Precision-focused
9. Outcomes assessed	1		2		3		4		5	
	Proxy or process					Intermediate				Events/Death
	e.g. self-reported behaviors		*e.g. measured weight*		*e.g. single biochemical tests*		*e.g. repeated/ sophisticated testing*			*e.g. strokes*
10. Focus of outcomes	1	2	3	4	5	6	7	8	9	10
	Applicability and scalability					Understand etiology & mechanisms				
	Pathways to implementation explored					*Pathways to outcomes explored*				
11. Analytical method used	4		5		6		7	8	9	10
	Intention to treat		Both Intention to treat and per-protocol					Per-protocol (*adherent*)		
TOTAL SCORE	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**
	Effectiveness						**Efficacy**			

Table 2.

Item and total scores of the included studies (N=134) by study focus and treatment type.

Taxonomy item (scale)	Effectiveness LSM studies (n=82) (mean ± SD)	Efficacy LSM studies (n=32) (mean ± SD)	Efficacy medication studies (n=20) (mean ± SD)
Study Goal (1-10)	4.0 ± 1.2	7.9 ± 1.2	8.4 ± 1.1
Study Design (1-10)	3.7 ± 1.7	7.5 ± 1.2	9.0 ± 0.6
Population Inclusion Criteria (1-10)	6.1 ± 1.8	7.4 ± 1.3	8.6 ± 1.0
Population Recruitment (1-10)	4.3 ± 1.6	6.1 ± 1.2	7.2 ± 1.6
Intervention Delivery Setting (1-10)	3.7 ± 1.4	6.2 ± 1.3	7.2 ± 0.9
Intervention Intensity (1-10)	5.3 ± 1.9	6.9 ± 1.7	7.4 ± 1.3
Emphasis on Adherence (1-10)	4.5 ± 1.6	6.1 ± 1.6	7.1 ± 1.3
Data Analysis Indicators (1-5)	4.0 ±0.6	4.8 ± 0.3	5.0 ± 0.2
Outcomes Assessed (1-5)	2.9 ± 0.8	4.0 ± 0.3	4.2 ± 0.5
Focus of Outcomes (1-10)	4.5 ± 1.3	7.3 ± 1.1	8.5 ± 0.6
Analytical Method (4-10)	6.8 ± 1.6	5.6 ± 1.8	4.9 ± 1.4
Total score (1-100)	49.2 ± 7.5	69.8 ± 6.0	77.4 ± 4.6

SD=standard deviation

LSM = lifestyle modification

Lower scores indicate external validity focus; higher scores indicate internal validity focus.

Table 3.

Taxonomy item correlations computed using the average of the two sets of scores among 134 articles.

	1.GOAL	2.DESIGN	3.INCLUSION	4.RECRUIT	5.SETTING	6.INTENSITY	7. ADHERENCE	8. INDICATORS	9.OUTCOME	10.FOCUS	11.METHOD
1	1.000										
2	.800	1.000									
3	.404	.546	1.000								
4	.569	.595	.495	1.000							
5	.675	.688	.530	.789	1.000						
6	.327	.317	.427	.195	.363	1.000					
7	.427	.505	.454	.298	.427	.530	1.000				
8	.581	.648	.535	.505	.575	.329	.337	1.000			
9	.625	.644	.324	.451	.479	.240	.294	.675	1.000		
10	.832	.748	.306	.525	.635	.377	.485	.498	.615	1.000	
11	−.387	−.492	−.264	−.280	−.304	−.268	−.388	−.231	−.329	−.461	1.000

** All significant at $p < 0.01$

Table 4.

Factor structure, factor loadings and intra-class correlation coefficients for the taxonomy items.

Taxonomy items	Objective Factor	Population Factor	ICC (95% CI)
Focus of Outcomes	0.95	−0.08	0.72 (0.60 to 0.80)
Study Goal	0.88	0.04	0.84 (0.77 to 0.88)
Study Design	0.74	0.19	0.93 (0.89 to 0.95)
Outcomes	0.69	0.20	0.82 (0.74 to 0.87)
Analytical Method	−0.56	0.09	0.71 (0.59 to 0.80)
Emphasis on Adherence	0.46	0.09	0.56 (0.37 to 0.69)
Intervention Intensity	0.40	0.34	0.67 (0.54 to 0.77)
Data Analysis Indicators	0.32	0.12	0.60 (0.43 to 0.71)
Population Recruitment	−0.05	0.89	0.75 (0.65 to 0.83)
Intervention Delivery Setting	0.12	0.82	0.81 (0.73 to 0.86)
Population Inclusion Criteria	0.04	0.58	0.65 (0.51 to 0.75)

ICC= Intra class correlation coefficient to test inter-rater reliability.

Table 5.

Effect estimates obtained in random effects meta-analyses for efficacy and effectiveness studies.

	Efficacy LSM studies (N=16)	Effectiveness LSM studies (N=24)
Taxonomy Score (Mean (SD))	69.5 (5.5)	50.1 (6.3)
Follow up years (Mean (SD))	2.7 (1.5)	1.2 (0.6)
Within subjects weight loss at 12 months (kg)	n=13 −3.0 [−4.5, −1.4]	n=24 −2.5 [−3.3, −1.6]
Between Subjects weight loss at 12 months (kg)	n=10 −2.4 [−4.3, −0.6]	n=9 −1.8 [−3.4, −0.2]
Diabetes relative risk *	n=16 0.56 [0.48, 0.65]	n=8 ** 0.71 [0.59, 0.85]

LSM = lifestyle modification.

* These are controlled studies only.

** Score for these 8 studies is 56.1 (SD=6.7) and mean follow up is 2.0 years (SD=1.3). n=number of studies included in each analysis.