Aim

The aim of this paper is to review the methodological characteristics of mediation analyses performed in observational epidemiologic studies published between 2015 and 2019 and to provide recommendations for the application of mediation analyses in future studies. In this paper we performed a scoping review, as the aim of this paper is relatively broad and concerns the collection of information on a range of methodological characteristics rather than information on a clearly defined substantive question [30]. In the next section, we first provide an overview of traditional and causal mediation analysis methods. Then we describe the methods and results of our scoping review. Finally, we provide recommendations for the application of mediation analysis in future studies.

Traditional mediation analysis

Traditional mediation analysis is based on the estimation of the four pathways shown in Fig. 1 [3, 10]. In Fig. 1A, the c path represents the total exposure-outcome effect. In Fig. 1B, the a path represents the exposure-mediator effect, the b path represents the mediator-outcome effect, and the câ€™ path represents the direct exposure-outcome effect. When the mediator and outcome are both continuous, the paths in Fig. 1 are estimated using the following three linear regression eqs. (9):

$$Y={i}_1+ cX+{d}_1Z+{upvarepsilon}_1$$ (1)

$$M={i}_2+ aX+{d}_2Z+{upvarepsilon}_2$$ (2)

$$Y={i}_3+{c}^{prime }X+ bM+{d}_3Z+{upvarepsilon}_3$$ (3)

where the c coefficient in eq. 1 represents the total exposure-outcome effect. The a coefficient in eq. 2 represents the exposure-mediator effect. The b coefficient in eq. 3 represents the mediator-outcome effect when adjusted for the exposure, and the câ€™ coefficient represents the direct exposure-outcome effect when adjusted for the mediator. The i 1 , i 2 , and i 3 terms represent intercepts and the Îµ 1 , Îµ 2 , and Îµ 3 terms represent residuals. Finally, Z represents a set of confounders. The inclusion of confounders in eqs. 1, 2, and 3 should always be considered when a mediation analysis is performed based on observational data, as the exclusion of confounders will result in biased effect estimates [3].

Fig. 1 Path diagram of a single mediator model. Full size image

Traditional mediation analysis defines the direct, indirect, and total effects in terms of the linear regression coefficients from eqs. 1, 2, and 3 [3, 12]. The total effect is defined and estimated as the c coefficient from eq. 1 and the direct effect is defined and estimated as the câ€™ coefficient from eq. 3. The indirect effect is defined and estimated as the product of the a and b coefficients (ab) and as the difference between the c coefficient and the câ€™

coefficient (c-câ€™). These two indirect effects are mathematically equivalent when the regression coefficients are estimated with linear regression [13]. The relative size of the mediated effect can be assessed using the proportion mediated, which represents the size of the indirect effect estimate relative to the total effect estimate, or by interpreting the standardized indirect effect estimate as a Cohenâ€™s d [3].

Some of the first papers on mediation analysis recommended to assess the statistical significance of the indirect effect estimate with a z-test or a confidence interval based on the multivariate delta standard error [10, 31,32,33]. However, these methods are not recommended, as they assume that the indirect effect estimate follows a normal sampling distribution, which often does not hold [34]. As a result, the z-test and confidence interval based on the multivariate delta standard error have relatively low power to detect a statistically significant indirect effect [35,36,37]. Confidence intervals that do take into account the nonnormal sampling distribution of the indirect effect estimator are therefore preferred, such as the distribution of the product confidence interval, Monte Carlo confidence interval, and bootstrap confidence intervals [34, 36, 38].

Mediation analysis is based on the assumption of temporal precedence of the exposure, mediator, and outcome, which means that changes in the exposure are assumed to precede changes in the mediator, and that changes in the mediator are assumed to precede changes in the outcome [3, 39]. Furthermore, traditional mediation analysis is based on parametric regression assumptions. In other words, the residuals of the linear regression models are assumed to be normally distributed and homoscedastic across values of the independent variables in the model, the a, b, c, and câ€™ coefficients are assumed to represent their correct functional form (e.g., linear or quadratic), the observations are assumed to be independent, and it is assumed that there are no effect modifiers or omitted confounders of the estimated effects [3, 40]. Effect modifiers can be taken into account by including interaction terms (i.e., exposure-by-covariate or mediator-by-covariate) in the models and by subsequently estimating the direct and indirect effects for different values of the effect modifier. This can, for example, be done by estimating the effects for specific categories of a categorical effect modifier or by centering a continuous effect modifier at a clinically relevant value [3, 11]. The effect estimates can be adjusted for measured confounders by adding the confounder variables to all estimated regression equations.

Ambiguities arise when traditional mediation analysis is used to estimate the effects for mediation models with

non-continuous mediator and outcome variables [12, 41, 42]. For example, the product-of coefficients and difference-in-coefficients methods provide different indirect effect estimates when based on the coefficients from non-linear regression models, such as logistic regression or Cox proportional-hazards regression [12, 41, 43]. Furthermore, although it has been recommended to assess the presence of exposure-mediator interactions in the traditional mediation analysis literature, guidance is scarce on the estimation and interpretation of effects for mediation models with an exposure-mediator interaction [3, 9]. Recent papers have shown that group-mean centering of the continuous mediator variable in traditional mediation analysis yields effect estimates similar to the effect estimates from causal mediation analysis for mediation models with a continuous outcome and an exposure-mediator interaction [16], but not necessarily for mediation models with a binary outcome and an exposure-mediator interaction [18].

Causal mediation analysis

Causal mediation analysis clarifies the ambiguities that arise in traditional mediation analysis [16, 18, 44]. Causal mediation analysis is based on the counterfactual framework [4, 14, 15], and distinguishes causal effect definitions from causal effect estimation [45]. A strength of the causal effect definitions is that they are non-parametric and therefore can be applied to any type of mediation model to derive the causal effect estimates. This includes models with an exposure-mediator interaction and models with non-continuous mediator variables or non-continuous outcome variables [46].

Causal effect definitions

Causal mediation analysis defines causal effects as the difference between two counterfactual outcomes [47, 48]. A counterfactual outcome is an individualâ€™s outcome value that would be observed when exposed to a certain exposure value. In the remainder of this section we denote the outcome as Y, and the exposure values of interest as x and x*. In theory, two counterfactual outcomes can be observed for one individual over the same time period, one based on exposure value x and one based on exposure value x* [47, 48]. The individualâ€™s counterfactual outcome under exposure value x is denoted as Y i (x), and the individualâ€™s counterfactual outcome under exposure value x* is denoted as Y i (x*). The causal exposure effect is defined as the difference between these two counterfactual outcomes observed for the same individual over the same time period, i.e., Y i (x) â€“ Y i (xâˆ—).

The counterfactual outcomes in a mediation model are not only dependent on exposure values, but also on mediator values [4]. We denote the mediator as M and the mediator values as m. The counterfactual notation for the outcome can be extended by including this mediator value. An individualâ€™s counterfactual outcome under exposure value x and mediator value m is denoted as Y i (x, m), and the same individualâ€™s counterfactual outcome under exposure value x* and mediator value m as Y i (x*, m). The difference between these two counterfactual outcomes observed for the same individual over the same time period is the controlled direct effect (CDE), i.e., Y i (x, m) â€“ Y i (xâˆ—, m). The CDE is the direct effect of changing an individualâ€™s exposure value from x to x*, while holding the mediator value constant at m [4]. The mediator value m is determined by the researcher and reflects a value of clinical or policy relevance [4].

Instead of holding the mediator constant at a predetermined value, we can also let the mediator take on the value that would naturally be observed under exposure values x and x* [4]. Two counterfactual mediator values can be observed for an individual under the two exposure values x and x*: the counterfactual mediator value under exposure value x, i.e., M i (x), and the counterfactual mediator value under exposure value x*, i.e., M i (x*). We can now replace mediator value m with these two counterfactual mediator values, resulting in four nested counterfactual outcome values: Y i (x, M i (x)), Y i (x, M i (xâˆ—)), Y i (xâˆ—, M i (x)), and Y i (xâˆ—, M i (xâˆ—)) [4, 49]. These four counterfactual outcomes are referred to as nested counterfactual outcomes, because the counterfactual mediator values are nested within the counterfactual outcomes values [4].

Five causal effects are defined based on the differences between these nested counterfactual outcomes: the pure natural direct effect (PNDE), the total natural direct effect (TNDE), the pure natural indirect effect (PNIE), the total natural indirect effect (TNIE), and the total effect (TE) [4, 15]. Table 1 provides an overview of these causal effects and their respective interpretations. For the natural direct effects we block the effect through the mediator by holding each individualâ€™s mediator constant at either M i (x) or M i (x*), while for the natural indirect effects we block the effect through the exposure by holding the exposure constant at either x and x* [1, 50]. For the TE, we allow information to flow through both the exposure and mediator, varying both the exposure value and the counterfactual mediator value.

Table 1 Overview of the definitions and interpretations of the causal mediation effects Full size table

The causal effects are defined at the individual level, but in practice we are unable to observe multiple

counterfactual outcomes for the same individual over the same time period [47, 48]. Therefore, we are unable to estimate individual-level causal effects. This has been referred to as the fundamental problem of causal inference [47]. Instead, we can estimate the population-average causal effects based on the expected difference between two population-average (nested) counterfactual outcomes [4, 14, 47]. To ensure that the PNDE, TNDE, PNIE, and TNIE have a causal interpretation at the population-average level, the following four assumptions need to hold [4, 46]:

1. no unmeasured confounding of the exposure-outcome effect; 2. no unmeasured confounding of the mediator-outcome effect; 3. no unmeasured confounding of the exposure-mediator effect; 4. no confounders of the mediator-outcome effect that are affected by the exposure.

Assumption 4 is also known as the cross-world independence assumption. In practice this is often a strong assumption [51], for example because often there will be multiple mediators of the exposure-outcome effect. For the CDE only assumptions 1 and 2 have to hold, and for the TE only assumption 1 has to hold. Finally, consistency is assumed, which means that the observed mediator and outcome values would also have been observed had the individual randomly been assigned the observed exposure and mediator values [46, 52].

Causal effect estimation

Various estimation approaches have been developed to estimate the causal direct, indirect, and total effects at the population-average level, including simulations, numerical integration, multiple regression analysis, and natural effect models [19, 23, 53,54,55]. Most of these methods use eq. 2 and/or eq. 3 as input. Provided that the relevant parametric assumptions hold, the regression coefficients from eqs. 2 and 3 can be used to compute the causal mediation effects. To accommodate the estimation of pure and total natural direct and indirect effects, eq. 3 is typically extended with an exposure-mediator interaction term.

The simulation-based approach can be applied based on both parametric and non-parametric models [25, 53]. The parametric simulation-based approach uses the sampling distributions of the estimated parameters from eqs. 2 and 3 to simulate the potential mediator and outcome values for each subject. Based on the simulated potential outcomes, the causal effects are computed for each subject. Subsequently, the causal effects are averaged to arrive at the population-average causal effects. The non-parametric simulation-based approach estimates possibly non-parametric models for the mediator and outcome variables within a prespecified number of bootstrap resamples. Based on these

models the potential mediator and outcome values are simulated for each subject. Then based on these simulated potential outcomes, the causal effects are estimated and averaged to get the population-average causal effects.

Numerical integration uses eqs. 2 and 3 as input [4, 23]. Based on these equations, average expected outcome values are estimated conditional on the two exposure levels of interest, i.e., x and x*, and all mediator values. These expected outcome values are weighted by the mediator distributions observed under x and x* to estimate the population-average nested potential outcomes, which are subsequently subtracted to get the population-average causal effect estimates.

The regression-based method estimates the average potential outcomes based on the regression coefficients in eqs. 2 and 3 [19, 46, 56]. These estimated potential outcomes are subsequently subtracted to estimate the population-average causal mediation effects. The regression-based effects for mediation models with a binary or time-to-event outcome were originally derived on the risk-ratio scale, therefore this method poses an additional rare outcome assumption when the causal effects are estimated on the odds-ratio scale or hazard-ratio scale [56, 57]. This assumption requires the outcome prevalence to be low across all strata of the exposure and mediator variable [58]. When this assumption is violated, the effect estimates on the odds-ratio scale or hazard-ratio scale can still be used to assess the presence of a mediated effect, but they do not have a causal interpretation [56]. To ensure a causal interpretation, the effects can alternatively be estimated on the risk-ratio scale using log-linear regression or on the survival-time ratio scale using accelerated failure time models [28, 57].

In natural effect models the natural direct effect and natural indirect effect are each represented by a single regression coefficient [25]. In contrast with the other estimation methods, natural effect models require the estimation of only one of the aforementioned regression equations, i.e., eqs. 2 and 3, in addition to the natural effect model [59]. Natural effect models are estimated using a weighting-based approach or a imputation-based approach. The weighting-based approach creates an expanded dataset with weights for each subject based on eq. 2 [54, 60]. The natural effects model is subsequently estimated by regressing the outcome on the two exposure values of interested, i.e., x and x*, and the covariates, while weighting each observation based on the computed weights. The imputation-based approach creates an expanded dataset in which the missing potential outcome values are imputed based on information from eq. 3 [55]. Based on this complete dataset, a natural effects model is estimated.

Traditional mediation analysis versus causal mediation analysis

For certain mediation models, traditional mediation analysis provides the same effect estimates as causal mediation analysis. Traditional mediation analysis provides the same effect estimates as causal mediation analysis for single mediator models with a continuous mediator and a continuous outcome [16, 17, 45]. This also means that traditional mediation analysis fails to provide causal effect estimates when the four no (unmeasured) confounding assumptions are violated. For mediation models with a binary or time-to-event outcome, traditional and causal mediation analysis do not necessarily provide the same effect estimates [16, 18]. For these models, the effect estimation in traditional mediation analysis is most closely related to the regression-based estimation approach in causal mediation analysis, which also estimates the indirect effect using the product-of-coefficients method in the absence of exposure-mediator interaction. However, an important difference is the rare outcome assumption posed by causal mediation analysis for mediation models with a binary or time-to-event outcome. This rare outcome assumption clarifies that the traditional effect estimates based on logistic regression and Cox proportional hazards regression only have a causal interpretation when the outcome is rare.

When there are multiple mediators of the exposure-outcome effect, it is important to take into account all these mediators, because they may be correlated or they may influence one another violating the fourth no confounding assumption, i.e., no confounders of the mediator-outcome effect that are affected by the exposure. Causal mediation analysis clarifies the necessary additional causal assumptions for models with multiple mediators and various methods have been developed for the estimation of causal effects for multiple mediator models [25, 61,62,63].

In recent years, various causal mediation software packages have been developed that enable researchers to apply causal mediation analysis based on only a few lines of code [21,22,23,24,25,26,27, 64]. However, it remains unclear whether the availability of these causal mediation programs has increased the uptake of causal mediation analysis in practice. In the next section we describe the set-up of our scoping review in which we collected information on the methodological characteristics of mediation analyses in published observational studies, with a special focus on the mediation analysis method used.

Study design

This scoping review is reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [65] and the PRISMA-ScR extension [66]. The PRISMA-ScR checklist can be found in supplementary appendix 1. The protocol for this scoping review was not registered in the international register of

systematic reviews, because we did not extract data on clinical outcomes [67].

Our search strategy is based on the MEDLINE search performed by Vo and colleagues [29] who conducted a review aimed to assess the methodological characteristics of mediation analyses conducted in randomized controlled trials between 2017 and 2018. We adapted the search conducted by Vo and colleagues [29] in four ways. First, we searched both the MEDLINE and EMBASE, as EMBASE has been shown to contain many unique references compared to MEDLINE when performing medically-oriented searches [68]. Second, we extended the search period to 5 years, including papers published between January 1st 2015 and December 31st 2019, as estimation methods for causal mediation analysis have been implemented in all major software packages since 2015 [21,22,23,24,25,26,27,28]. Third, in addition to the keywords â€œmediation analysisâ€•, mediation, and mediator used by Vo and colleagues [29], we also included the following keywords to increase the chances of finding papers that conducted a mediation analysis: â€œmediation analys*â€•, mediators, â€œindirect effectâ€•, â€œindirect effectsâ€•, â€œcausal stepsâ€•, â€œproduct-of-coefficientsâ€•, and â€œdifference-in-coefficientsâ€•. Fourth, we searched for observational studies only, as the earlier study performed by Vo and colleagues [29] examined the methodological characteristics of mediation analyses conducted in randomized controlled trials. The MEDLINE (accessed through PubMed) and EMBASE (accessed through embase.com) searches were performed on May 20th 2020. The complete MEDLINE and EMBASE search strategies can be found in supplementary appendix 2.

After removing duplicate records, two authors (JJMR and SJL) independently screened the titles and abstracts of the identified records for eligibility using Rayyan software [69]. Records were eligible for inclusion when published between 2015 and 2019, written in English, based on observational human subjects data, and the title or abstract indicated that it concerned an original research paper in which mediation analysis was performed. Full texts of the eligible records were obtained. When full texts were not available, full texts were requested from the corresponding author by email. Two authors (JJMR and SJL) independently screened the full texts for eligibility. Full texts in which mediation analysis was not performed as one of the primary analysis methods and conference abstracts were excluded, as we expected that these records did not contain a sufficient amount of details on the performed mediation analyses. Disagreements at any stage of the screening process were resolved by a third author (MJV).

A data extraction form was developed and pilot tested by one author, who subsequently extracted data from all eligible papers (JJMR). To ensure the quality of the extracted data, two authors (MJV and SJL) each independently extracted data from a random subsample of 12.5% of the eligible papers, i.e., 25% of the papers in total. Disagreements were

resolved through discussion. The data extraction included the mediation analysis method used, publication year, study design, sample size, software used, the number of exposure, mediator, and outcome variables, each variableâ€™s measurement level, use of a path diagram, use of repeated measurements, single or multiple mediator model, the types of estimated regression models, the type of confidence interval for the indirect effect estimates, the reporting of standard errors and p-values for the indirect effect estimates, use of effect size measures, inclusion of confounders in the analyses, use of sensitivity analyses for unmeasured confounders, assessment of exposure-mediator interaction, assessment of effect modifiers (i.e., exposure-by-covariate or mediator-by-covariate), and the discussion of the rare outcome assumption for mediation models with a binary or time-to-event outcome estimated based on traditional mediation analysis or regression-based causal mediation analysis. For papers based on longitudinal data we extracted the number of measurement waves included in the analyses and the type of longitudinal mediation model estimated. For multiple mediator models we extracted the type of multiple mediator model and the assessment of mediator-by-mediator interactions. The extracted data were summarized using descriptive statistics stratified by the mediation analysis method used. Categorical variables were summarized using frequencies and percentages, and continuous variables were summarized using medians and interquartile ranges.

Reference

[Vermiculture Technology: Earthworms, Organic Wastes, and Environmental Management]

[The Software Engineering Manager Interview Guide]