Finding the Pathway: Mediation Analyses in Randomized Controlled Trials

A. Russell Localio, PhD; Anne R. Meibohm, PhD; and Eliseo Guallar, MD, DrPH

The goal of clinical investigation often lies not only in estimating the effects of treatment or exposure but also in understanding their mechanisms. Identifying the pathways from treatment to outcome and explaining potential causes of the outcome—those are the specific domain of "mediation analysis" (1). In their report in *Annals,* Vallurupalli and colleagues (2) used the Canakinumab Anti-inflammatory Thrombosis Outcomes Study (CANTOS) to ask whether canakinumab reduced the incidence of anemia of chronic inflammation in patients with a history of myocardial infarction and high levels of high-sensitivity C-reactive protein (hsCRP) and whether reductions in hsCRP mediated this effect.

CANTOS participants without anemia at baseline were randomly assigned to receive 1 of 3 dosages of canakinumab (50, 150, or 300 mg; $n = 5796$) or placebo ($n = 2887$). Over a median follow-up of 3.7 years, the incidence rate of anemia was 16% (95% CI, 7% to 23%) lower and the mean time free of anemia 42% (CI, 20% to 67%) longer in the canakinumab versus the placebo group. Using mediation analysis, the authors then investigated whether, and to what extent, the effect of canakinumab on hsCRP levels explained the reduction in anemia rates. We summarize some of the goals, assumptions, and challenges of mediation analysis in medical and public health studies and use the CANTOS report to illustrate key concepts at a time of rapidly developing statistical methodology in this field.

Goals and Models of Mediation Analysis

Mediation analysis seeks to elucidate the pathways through which a treatment or exposure leads to an outcome, to clarify biological mechanisms, to suggest potential ways to intervene to change the relationship of exposure and outcome, and to identify patients who might benefit from early intervention. To this end, mediation analysis follows a counterfactual, or potential outcomes, model (1, 3–5) to estimate the difference in outcomes in 2 situations in which the same patient under the same circumstances either receives or does not receive the exposure or intervention. Only 1 of these situations can be observed—hence the term *counterfactual*. The overall (total) estimated effect of an intervention (denoted by *A* in **Figure 1**) on an outcome (denoted by *Y*) is separated into a direct effect (*A* to *Y*), and 1 or more indirect effects mediated through 1 or more mediators or intervening factors (denoted by *M*). In randomized trials, with proper attention to adherence, dropout, and other postrandomization events, estimating the total effect presents no additional complexities.

In the CANTOS trial, *A* represents randomized assignment to canakinumab or placebo, *Y* the development of incident anemia, and *M* the hsCRP level measured 3 months after randomization. The total effect of canakinumab on anemia was the difference between the rate of anemia if all participants received the drug and the rate if all patients received placebo.

Mediation analysis can then decompose the total effect of exposure into its direct and indirect components. In our example, the "natural direct effect" represents the change in anemia outcome (*Y*), assuming alternatively that all patients were assigned to canakinumab versus all were assigned to placebo, but with the mediator (*M* = hsCRP) held constant for each patient at the level the patient would have if he or she were assigned to the placebo group. The natural direct effect therefore can be interpreted as the part of the total effect that does not operate through the mediator. The CANTOS investigators estimated the natural direct effect (the increase in mean time free of anemia, comparing cankinumab with placebo while holding hsCRP fixed at the level each patient would have if assigned to receive placebo) to be 27% (CI, 7% to 51%) (Table 3 of Vallurupalli and colleagues [2], top section).

Likewise, the "natural indirect effect" estimates the effect of *A* (canakinumab) on *Y* (anemia) through its effect on the level of the mediator *M* (hsCRP). The natural indirect effect represents the change in time free of anemia if a patient were assigned to receive canakinumab and the mediator hsCRP were to change in value from what it would be if the patient were assigned to receive placebo to what it would be if the patient were assigned to receive canakinumab. In CANTOS, the natural indirect effect estimate of 11% (CI, 3% to 20%) (Table 3 of Vallurupalli and colleagues [2], top section) represents the change in outcome resulting from the change in hsCRP. Canakinumab lowers hsCRP, which in turn delays the onset of anemia.

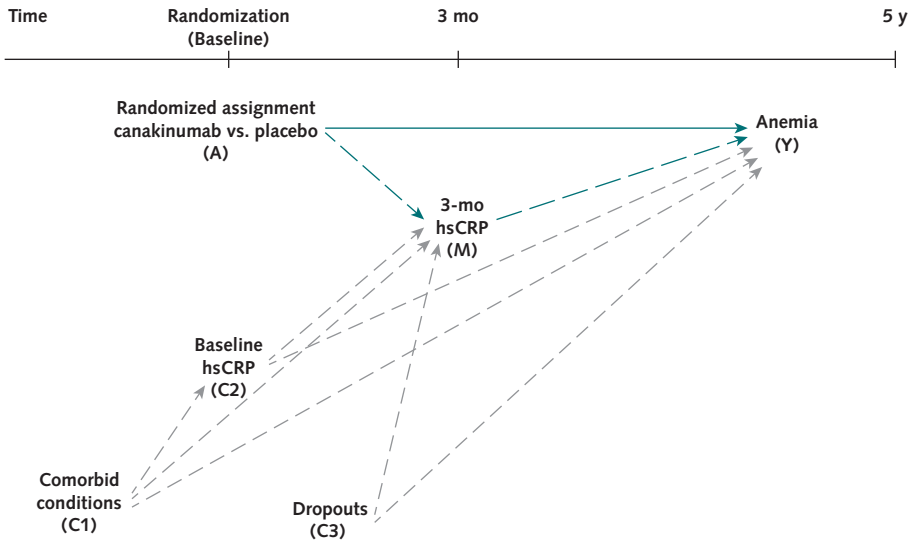
In CANTOS, the total effect of canakinumab on time to anemia (1.42 on the relative scale, or a 42% increase) equates to the product of the 11% natural indirect effect and the 27% natural direct effect ($1.11 \times 1.27 = 1.41$, rounded here and reported as 1.42 by Vallurupalli and colleagues [2] in Table 3). The authors used a multiplicative, time-to-event statistical model, so the total effect was the product rather than the sum of natural direct and indirect effects. If the investigators had implemented an additive model, natural direct plus indirect effect would sum to total effect (6).

See also:

Related article 523

This article was published at Annals.org on 24 March 2020.

Figure 1. Mediation by hsCRP (*M*) from the effect of canakinumab treatment (*A*) on anemia (outcome, *Y*).



The direct effect of treatment is represented by the solid green line and arrow from *A* to *Y* without passing through a mediator (*M*). Indirect effects are represented by the green dashed lines and arrows from *A* through *M* and then to *Y*. The gray dashed lines represent potential confounders (*C*s). They are possible but unlikely for the *A*→*M* relationship in the presence of randomization of *A* and therefore are not drawn. Likewise, baseline hsCRP should not be associated with treatment assignment (*A*) because of randomization at baseline. Baseline hsCRP, however, might be associated with the subsequent hsCRP as well as with the outcome. The potential confounder at baseline (*C1* and *C2*) probably should also be balanced at randomization, and we do not consider them as actual confounders in the *A*→*M* relationship. By contrast, baseline comorbid conditions (*C1*) might confound the *M*→*Y* relationship, as well as baseline hsCRP (*C2*), if the change in hsCRP varies with comorbid conditions, as does outcome (*Y*). *C3* represents another potential confounder that occurs after baseline. Dropout, for example, might vary with hsCRP if a patient withdraws from the study after baseline because of poor treatment response and the withdrawal then leads to a worse outcome. hsCRP = high-sensitivity C-reactive protein.

The CANTOS trial did not report hazard ratios for mediation estimation. Cox models require special methods to decompose total effects into direct and indirect components (1, 7–9). Traditionally, in the social sciences, mediators and outcomes have been continuous variables, often based on test or psychometric scores and modeled with linear regressions. When mediators and outcomes depart from these simple cases, analyses require more assumptions and specialized software. In CANTOS, the randomized exposure (*A*) was an ordinal variable (placebo or canakinumab dosage of 50, 150, or 300 mg), which was treated as binary (placebo vs. canakinumab); the continuous mediator was dichotomized, but also used as continuous (log-transformed); and the outcome was time to event (anemia).

Figure 2. Calculating the proportion (of the effect of exposure on outcome) mediated from the NDE and the NIE on the basis of a multiplicative Weibull accelerated failure time model.

NDE = 1.27

NIE = 1.11

$$\text{PM} = \frac{(\text{NDE} * (\text{NIE} - 1))}{(\text{NDE} * \text{NIE} - 1)}$$

Using the estimates from CANTOS for the mean time free of anemia:

$$\text{PM} = \frac{(1.27 * (1.11 - 1))}{(1.27 * 1.11 - 1)} = 0.34$$

See the main text for details. CANTOS = Canakinumab Anti-inflammatory Thrombosis Outcomes Study; NDE = natural direct effect; NIE = natural indirect effect; PM = proportion mediated.

Proportion Mediated and Eliminated

Having decomposed effects into their direct and indirect components, the CANTOS investigators estimated the proportion of total effect mediated through hsCRP. On the multiplicative scale, this portion is a ratio (Figure 2) equal to 0.34, or 0.35 in Vallurupalli and colleagues' Table 3 (2). Thus, about one third of the effect of the drug operates through changing hsCRP levels at 3 months—in this example, from 2 mg/L or greater to less than 2 mg/L.

Another useful measure in mediation is the "controlled direct effect," the effect of exposure on outcome when the mediator is fixed at the same level in all participants. In CANTOS Table 3, the controlled direct effect—the estimated increase in the mean time free of anemia comparing canakinumab with placebo and setting hsCRP to a given level (such as <2 mg/L) in all study participants—was 29% (CI, 9% to 57%). With this estimate of controlled direct effect, we can calculate another ratio—the proportion of the total effect eliminated—by the change in hsCRP (0.31 in this case [Figure 3]). The proportion eliminated answers the policy question of the potential for reducing the risk for the outcome (anemia) by intervening on the mediator level (hsCRP). As the controlled

direct effect increases, the proportion-eliminated ratio falls.

Thus, a formal decomposition of direct and indirect effects can answer clinically relevant questions about the pathways and potentials for treatments. Nevertheless, as with many ratios, the proportions mediated and eliminated may be unstable and highly variable. Instability becomes especially serious when the natural direct effect and the natural indirect effect operate in opposite directions ("inconsistent mediation"), leading to smaller total effects of treatment on outcome. For that reason, these proportion-mediated and proportion-eliminated ratios will not always be useful in mediation analysis, even if the decomposition of effects is helpful.

ASSUMPTIONS UNDERLYING MEDIATION MODELS

Several assumptions must hold for unbiased estimation of direct and indirect effects (Table). A key assumption, often implicit and sometimes overlooked, is temporality: Treatment (*A*) precedes the mediator (*M*), and both precede the outcome (*Y*). In CANTOS, randomization preceded 3-month hsCRP measurements, and both preceded the development of anemia. In observational studies, however, investigators often measure exposure (*A*) and candidate mediators (*M*) at the same visit. In that case, they must explain and support assumptions about the time and directionality of effects between *A* and *M*.

Mediation analyses also assume the absence of unmeasured confounding in the treatment–outcome relationship as well as in the mediator–outcome relationship. The CANTOS trial satisfied the treatment–outcome assumption, primarily through randomization. In observational studies, or in randomized trials with loss to follow-up, nonadherence, or other postrandomization events, investigators must identify and carefully control for potential confounders. By contrast, randomization of treatment does not protect against confounding in the mediator–outcome association, because the levels of *M* are not randomized. The investigator should always collect and adjust for all potential confounders of the mediator–outcome association. In CANTOS, for example, previous comorbid conditions may increase 3-month hsCRP levels and may also cause anemia, and thus would confound the association of 3-month hsCRP levels with anemia. To address confounding, the investigators adjusted for age, sex, baseline hsCRP (log-transformed), and presence of diabetes, heart failure, and hypertension at baseline.

The estimation of *natural* direct and indirect effects, used for estimating the proportion mediated, also requires an assumption of the absence of unmeasured confounding for the effect of treatment on the mediator. Randomization, without loss from postrandomization events, again should satisfy this assumption. In addition, the exposure or treatment cannot affect confounders in the mediation–outcome relationship (there are no arrows from *A* to *C1*, *C2*, or *C3* in Figure 1). This assumption may be challenging to meet and, if in question, should be the subject of sensitivity analyses.

Two additional assumptions are noninterference at the individual patient level and at the study site (also known as no spillover effects). Under these assumptions, individual patients should not interact in ways

Figure 3. Calculating the proportion (of the effect of exposure on outcome) eliminated from the CDE and the TE on the basis of a multiplicative Weibull accelerated failure time model.

TE = 1.42

CDE = 1.29

$$\text{PE} = \frac{(\text{TE} - \text{CDE})}{(\text{TE} - 1)}$$

Using the estimates from CANTOS for relative mean time free of anemia:

$$\text{PE} = \frac{(1.42 - 1.29)}{(1.42 - 1)} = 0.31$$

See the main text for details. CANTOS = Canakinumab Anti-inflammatory Thrombosis Outcomes Study; CDE = controlled direct effect; PE = proportion eliminated; TE = total effect.

Table. Assumptions to Satisfy for Use of Mediation Analysis of Treatment (*A*), Mediator (*M*), and Outcome (*Y*)

Assumptions	Approaches to Satisfy Assumptions
Overall assumptions	
Temporality: *A* precedes *M*, and *M* precedes *Y* in time	Capture data at precisely ordered times
No unmeasured confounding in (*A*→*Y*) relationship	Randomization and complete follow-up without losses
No unmeasured confounding in (*M*→*Y*) relationship	Collection of (and control for) covariates
For estimation of natural direct and indirect effects	
No unmeasured confounding in *A*→*M* relationship	Randomization and complete follow-up without losses
No effects of exposure (*A*) on the confounders of the *M*→*Y* relationship	Design; sensitivity analyses
Spillover	
No spillover effects at the individual patient level	Treat and measure patients individually rather than in groups
No spillover effects across study sites	Limit contacts of patients and investigators across sites
Usual design assumptions	
No selection bias	Careful design and selection criteria
No measurement error	Choose measures with regard to precision and reproducibility

that may affect the relationship of the mediator to the outcome, and study sites should not affect the outcome at other sites (10). In CANTOS, both assumptions were probably met, as will be the case with typical randomized controlled trials. In cluster-level randomized trials or group therapy interventions, however, these assumptions are more difficult to satisfy.

Sensitivity Analyses

As with many model-based and assumption-dependent analyses, mediation analyses should demonstrate robustness of findings and conclusions and assess the possible effects of confounding and interference. For the direct intervention effect, one can estimate how influential a confounder would need to be to explain away any observed direct effect. The same task applies to assessing sensitivity to violations of the assumption of no mediator–outcome confounding: how strong an unobserved confounder (or confounders) would need to be to eliminate (reduce to zero) the estimated effects and to render that estimate nonsignificant (1, 5, 11–13). To be able to support such sensitivity analyses, investigators should design studies to collect covariates with no missing values and little or no measurement error. By design, covariates collected and used ideally should also represent unobserved potential confounders.

Specialized Statistical Software

Well-documented, specialized programs are essential for mediation analysis and have become available in the past decade, including the macro used by the CANTOS authors (14), across common software packages, such as SAS (SAS Institute) (15), Stata (StataCorp) (16–18), and R (R Foundation for Statistical Computing) (19–21). Updates and extension will probably occur regularly in this field of ongoing statistical innovation.

Added Complexities and Special Situations

Although CANTOS investigated a single mediation pathway, more complex questions and designs, such as multiple correlated mediators, create methodological challenges (22). Treatment (*A*) and mediator (*M*) may interact, and this interaction may alter, sometimes substantially, the estimates of direct and indirect effects of treatment (1). Mediation and interaction together may decompose into even more refined effects, a complexity beyond the scope of this summary and the CANTOS example (23). Simultaneous effects of interactions of treatments and other factors, or interactions of mediators and other factors, are also details beyond the reach of this summary (1). Likewise, in situations of time-varying exposures, mediators, and confounders, simple concepts of direct and indirect effects do not hold, and analyses require weighted regressions to account for all factors (24).

Mediation analysis also may apply to studies in which randomization and treatment assignment occur at the cluster level, and in which mediation can occur at the individual as well as at the cluster– or intervention–level (1, 25–27). Finally, mediation often requires special attention to the complexities of variance estimation, measurement error, and missing values. Mediation analyses, as with any modeling, benefit from careful design, including statistical power calculations (28, 29) for estimating both direct and indirect effects. Calculations also should reflect the impact of multiple comparisons.

In conclusion, mediation analysis based on randomized trials can, as in the CANTOS report, help to explain the mechanisms by which treatments might lead to intended outcomes. Careful attention to assumptions and control of confounding are essential. When properly performed, however, a mediation analysis can produce helpful clinical and mechanistic insights.

From University of Pennsylvania, Philadelphia, Pennsylvania (A.R.L.); American College of Physicians, Philadelphia, Pennsylvania (A.R.M.); and Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland (E.G.).

Disclosures: Disclosures can be viewed at www.acponline.org/authors/icmje/ConflictOfInterestForms.do?msNum=M20-0887.

Corresponding Author: A. Russell Localio, PhD, Division of Biostatistics, Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, 617 Blockley Hall, 423 Guardian Drive, Philadelphia, PA 19104; e-mail, rlocalio@pennmedicine.upenn.edu.

Current author addresses and author contributions are available at Annals.org.

Ann Intern Med. 2020;172:553-557. doi:10.7326/M20-0887

References

1. **VanderWeele TJ.** Explanation in Causal Inference. Methods for Mediation and Interaction. Oxford Univ Pr; 2015.
2. **Vallurupalli M, MacFadyen JG, Glynn RJ, et al.** Effects of interleukin-1β inhibition on incident anemia. Exploratory analyses from a randomized trial. Ann Intern Med. 2020;172:523-32. doi:10.7326/M19-2945
3. **Valeri L, Vanderweele TJ.** Mediation analysis allowing for exposure-mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. Psychol Methods. 2013;18:137-50. [PMID: 23379553] doi:10.1037/a0031034
4. **Imai K, Keele L, Tingley D.** A general approach to causal mediation analysis. Psychol Methods. 2010;15:309-34. [PMID: 20954780] doi:10.1037/a0020761
5. **VanderWeele TJ.** Mediation analysis: a practitioner's guide. Annu Rev Public Health. 2016;37:17-32. [PMID: 26653405] doi:10.1146/annurev-publhealth-032315-021402
6. **VanderWeele TJ.** Causal mediation analysis with survival data. Epidemiology. 2011;22:582-5. [PMID: 21642779] doi:10.1097/EDE.0b013e31821db37e
7. **Tchetgen Tchetgen EJ.** On causal mediation analysis with a survival outcome. Int J Biostat. 2011;7:Article 33. [PMID: 22049268] doi:10.2202/1557-4679.1351

8. **Lange T, Hansen JV.** Direct and indirect effects in a survival context. Epidemiology. 2011;22:575-81. [PMID: 21552129] doi:10.1097/EDE.0b013e31821c680c
9. **Wang W, Albert JM.** Causal mediation analysis for the Cox proportional hazards model with a smooth baseline hazard estimator. J R Stat Soc Ser C Appl Stat. 2017;66:741-757. [PMID: 28943662] doi:10.1111/rssc.12188
10. **Vanderweele TJ, Hong G, Jones SM, et al.** Mediation and spillover effects in group-randomized trials: a case study of the 4Rs educational intervention. J Am Stat Assoc. 2013;108:469-482. [PMID: 23997375]
11. **Imai K, Keele L, Yamamoto T.** Identification, inference and sensitivity analysis for causal mediation effects. Stat Sci. 2010;25:51-71.
12. **Ding P, Vanderweele TJ.** Sharp sensitivity bounds for mediation under unmeasured mediator-outcome confounding. Biometrika. 2016;103:483-490. [PMID: 27279672]
13. **Ding P, VanderWeele TJ.** Sensitivity analysis without assumptions. Epidemiology. 2016;27:368-77. [PMID: 26841057] doi:10.1097/EDE.0000000000000457
14. **Valeri L.** Computational tools. Accessed at www.hsph.harvard.edu/linda-valeri/computational-tools on 20 February 2020.
15. **SAS Institute.** CAUSALMED. SAS/STAT® 15.1 User's Guide. SAS Institute; 2018.
16. **Hicks R, Tingley D.** Causal mediation analysis. Stata J. 2011;11:605-619.
17. **Emsley R, Liu H.** PARAMED: Stata module to perform causal mediation analysis using parametric regression models. 2013. Accessed at https://EconPapers.repec.org/RePEc:boc:bocode:s457581 on 26 January 2020.
18. **Daniel RM, De Stavola BL, Cousens SN.** Gformula: estimating causal effects in the presence of time-varying confounding or mediation using the g-computation formula. Stata J. 2011;11:479-517.
19. **Tingley D, Yamamoto T, Hirose K, et al.** mediation: R package for causal mediation analysis. J Stat Softw. 2014;59:1-38.
20. **Tingley D, Yamamoto J, Hirose K, et al.** mediation: R package for causal mediation analysis, version 4.4.5. 2015. Accessed at https://cran.r-project.org/web/packages/mediation/vignettes/mediation.pdf on 30 January 2020.
21. **Steen J, Loeys T, Moerkerke B, et al.** Flexible mediation analysis using natural effect models. Accessed at https://cran.r-project.org/web/packages/medflex/vignettes/medflex.pdf on 9 March 2020.
22. **VanderWeele TJ, Vansteelandt S.** Mediation analysis with multiple mediators. Epidemiol Methods. 2014;2:95-115. [PMID: 25580377]
23. **VanderWeele TJ.** A unification of mediation and interaction: a 4-way decomposition. Epidemiology. 2014;25:749-61. [PMID: 25000145] doi:10.1097/EDE.0000000000000121
24. **VanderWeele TJ, Tchetgen Tchetgen EJ.** Mediation analysis with time varying exposures and mediators. J R Stat Soc Series B Stat Methodol. 2017;79:917-938. [PMID: 28824285] doi:10.1111/rssb.12194
25. **Talloen W, Moerkerke B, Loeys T, et al.** Estimation of indirect effects in the presence of unmeasured confounding for the mediator-outcome relationship in a multilevel 2-1-1 mediation model. J Educ Behav Stat. 2016;41:359-391.
26. **Stapleton LM, Pituch KA, Dion E.** Standardized effect size measures for mediation analysis in cluster-randomized trials. J Exp Educ. 2015;83:547-582.
27. **Ten Have TR, Joffe MM, Lynch KG, et al.** Causal mediation analyses with rank preserving models. Biometrics. 2007;63:926-934.
28. **Kenny DA.** PowMedR. Accessed at www.davidakenny.net/progs/PowMedR.txt on 24 July 2016.
29. **Vittinghoff E, Neilands TB.** Sample size for joint testing of indirect effects. Prev Sci. 2015;16:1128-35. [PMID: 25418811] doi:10.1007/s11121-014-0528-5

Current Author Addresses: Dr. Localio: Division of Biostatistics, Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, 617 Blockley Hall, 423 Guardian Drive, Philadelphia, PA 19104.
Dr. Meibohm: American College of Physicians, 190 N. Independence Mall West, Philadelphia, PA 19106.
Dr. Guallar: Departments of Epidemiology and Medicine, Johns Hopkins Bloomberg School of Public Health, 2024 East Monument Street, Room 2-645, Baltimore, MD 21205.

Author Contributions: Conception and design: A.R. Localio.
Analysis and interpretation of the data: A.R. Localio.
Drafting of the article: A.R. Localio, A.R. Meibohm.
Critical revision for important intellectual content: A.R. Localio, E. Guallar.
Final approval of the article: A.R. Localio, A.R. Meibohm, E. Guallar.
Statistical expertise: A.R. Localio, E. Guallar.
Administrative, technical, or logistic support: A.R. Localio.