



Editorial

Scand J Work Environ Health [2013;39\(2\):121-124](#)

doi:10.5271/sjweh.3343

Understanding mechanisms: opening the “black box” in observational studies

by [Kristensen P](#), [Aalen OO](#)

Affiliation: Department of Occupational Medicine and Epidemiology, National Institute of Occupational Health, PO Box 8149 Dep, N-0033, Oslo, Norway. pk@stami.no

Refers to the following texts of the Journal: [2012;38\(6\):485-488](#)
[2013;39\(2\):134-143](#) [2012;38\(5\):391-392](#) [2012;38\(6\):516-526](#)
[2012;38\(5\):409-417](#)

The following articles refer to this text: [2013;39\(6\):535-549](#);
[2014;40\(5\):473-482](#); [2017;43\(1\):42-49](#)

Key terms: [causality](#); [counterfactual analysis](#); [determinant](#); [disability pension](#); [editorial](#); [mechanism](#); [mediation analysis](#); [mediator](#); [observational study](#); [retirement](#); [sickness absence](#); [socioeconomic status](#); [statistical method](#); [statistics](#); [time-dependent covariate](#); [unemployment](#)

This article in PubMed: www.ncbi.nlm.nih.gov/pubmed/23319154



This work is licensed under a [Creative Commons Attribution 4.0 International License](http://creativecommons.org/licenses/by/4.0/).

Understanding mechanisms: opening the “black box” in observational studies

New statistical methods are increasingly being used in epidemiology. One of the most interesting developments is the new field of causal inference. There is no magic bullet for causality of course, but the new statistical thinking in this area is fruitful in that it clarifies issues. It is important that statistical analyses are transparent, but one should not shy away from the effort to learn about new concepts such as counterfactual or mediation analysis. These ideas are likely to play an increasing role in many fields, including epidemiology. The design and execution of cohort studies are often excellent. The issue then is how to learn as much as possible from the data, and new statistical methods can in some cases yield more than classical ones.

Workers with a low baseline socioeconomic status (SES) were at increased risk of selection out of paid employment in a 10-year follow-up, whereas self-reported ill-health at baseline was associated with an increased risk of selection out of work *as well as* a reduced likelihood of return to work after exit. These are the main results of a Dutch study published in this issue of the *Scandinavian Journal of Work, Environment & Health* (1). The study was based on a rich data set that allowed the separation of different pathways out of work: unemployment, early retirement, disability pension, or becoming economically inactive. Schuring et al recommend that “[P]olicies to improve labor force participation, especially among low socioeconomic level workers, should protect workers with health problems against exclusion from the labor force”.

This appears to be sound advice. The logical prevention would then be targeted at workers with health problems. Such problems could be detected in health examinations at the workplace, or interventions could be targeted at risk groups at an early stage, eg, based on sickness absence registrations.

Baseline SES or health determining the subsequent working career could, in a life course terminology (2), be classified as effects of a critical period. It is however reasonable to consider more complex causal models based on the dynamic relationships between SES, health, and work. The influence of exit from work on subsequent health could be important in this respect (3, 4). Results in some studies suggest that retirement will be beneficial to health (5–7), but there are indications that the picture is more complex (4, 8). Effects on health and well-being are also likely to be dependent on exit pathways, being less favorable after unemployment (9) or involuntary retirement (10). Studies addressing employment often concern different time-dependent states, such as unemployment, sickness absence, disability, and retirement, which add to the analytical complexity (1, 11). In fact, the principal investigator of the Schuring et al study (1) addressed this issue and the call for multistate analytical models in a recent editorial (12).

Given this complexity, the question is how to formulate the proper research questions in order to give a sound basis for prevention (13). One solution would be to ask whether changes in determinants affect changes in outcome entities (14). This requires considering the time dimension not only for the outcome but also for the independent variables. Schuring et al (1) acknowledge the problem of not having measurements of changes in determinants and the potential for such changes influencing employment transitions. Alternative research questions could be: to what extent will changes in SES or changes in health influence changes in employment? How will changes in employment influence SES and health?

This calls for including time-dependent covariates in the analysis. One interesting possibility that has gained increased attention recently is to estimate direct and indirect effects in mediation analysis. Assessing the role of an intermediate between exposure and outcome – a mediator – could be considered as a way to open risk factor epidemiology’s “black box”, thereby improving the mechanistic understanding of an exposure–outcome relationship (15). The mediator of interest could be SES later in life as an

intermediate between childhood SES and chronic disease in adulthood (16) or physical work environment as an intermediate between SES and sickness absence (17). One realistic and interesting possibility is that the mediator might be more accessible for intervention than the exposure and therefore more valuable from a preventive point of view (15). For example, it could be easier to intervene on the work environment than SES level.

How should direct and indirect effects be unraveled? The traditional way has been to compare exposure effects in regression models with and without the mediator (18). This is a strategy that will often produce flawed results, one of the problems being confounding from time-dependent covariates (16, 19). Analysis of time-to-event data in Cox regression and comparing models with and without the mediator involves additional problems because the important assumption of proportional hazards can never be satisfied for both models. One alternative to the traditional regression methods (eg, logistic or Cox regression) is to apply a counterfactual framework (20). The concept behind this is, to a large degree, based on ideas of intervention described in Judea Pearl's famous book "Causality" (21). The aim is to analyze observational data in a way that mimics randomized experiments. Among the options for making counterfactual deductions from observational data, marginal structural model (MSM) with inverse probability weighting (IPW) and the use of causal diagrams (directed acyclic graphs or DAG) is the most common (16, 20, 22). This approach is well suited for causal mediation analysis (16). Dynamic path analysis is an alternative for estimating direct and indirect effects (17, 20, 23). This method is particularly useful when time-to-event is the outcome, ie, in survival analysis (17, 20). Lange & Hansen (17) estimated the mediating role of the physical work environment on the relationship between SES and sickness absence. Two methods were compared: one traditional where Cox models with and without the covariate (work environment) were compared, and the other in a counterfactual analysis where exposure (SES) and covariates were time-dependent, modeling the rate using the the Aalen additive hazard model (17). The traditional way was performed in accord with an earlier study applying the same data, where the degree of reduction in rate ratios in models including the covariate were interpreted as a corresponding degree of mediation through that factor (24). The results of the counterfactual analysis were qualitatively in accordance with the mediating effect in ordinary Cox regression analysis, but the quantitative effect was stronger (17).

High quality event data (1, 11, 24) from registers and other sources have become increasingly available. Counterfactual frameworks could be a valuable basis for analytical progress and increased causal understanding of social and biologic processes and, thereby, for correct interventions. These causal methods have limitations [eg, dependent on the assumption of no unmeasured confounding of exposure or mediator (19, 20)]. This problem should be minimized by the careful collection of covariates that could act as confounders of the exposure–or mediator–outcome relationships (19) as well as sensitivity analysis assessments (25).

Emerging tools and techniques open new avenues for identification and estimation of causal effects in observational studies. This is one reason why they have become increasingly more popular. One example is the 2012 awarding of the Rothman Epidemiology Prize to first author Theis Lange for his article "Direct and indirect effects in a survival context" (26).

Counterfactual causal analysis requires methods that could be challenging to the social or occupational epidemiologist. Does this mean that the floor will be completely taken over by biostatisticians? There is no reason to suspect such a scenario: the study of social or biological processes will depend on asking the right questions (13). Howards et al (25) provide an excellent example of the need for biologic knowledge and reasoning, pursuing the right model for analyzing the relationship between smoking and miscarriage. In this sense, the driving force behind causal inference is still biologic knowledge and hypotheses. Rather than fearing undue dominance by biostatisticians, there should be a call for a more intimate collaboration in all stages of studies applying the new causal methods.

References

1. Schuring M, Robroek SJW, Otten Ferdy WJ, Arts CH, Burdorf A. The effect of ill health and socioeconomic status on labor force exit and re-employment: a prospective study with ten years follow-up in the Netherlands. *Scand J Work Environ Health*. 2013;39(2):134–143. <http://dx.doi.org/10.5271/sjweh.3321>.
2. Kuh, D, Ben-Shlomo Y, Lynch J, Hallqvist J, Power C. Life course epidemiology. *J Epidemiol Community Health*. 2003;57:778–783. <http://dx.doi.org/10.1136/jech.57.10.778>.
3. Oksanen T, Virtanen M. Health and retirement: a complex relationship. *Eur J Ageing* 2012; 9: 221–225. <http://dx.doi.org/10.1007/s10433-012-0243-7>.
4. Härmä M. Is retirement beneficial or harmful to mental health? *Scand J Work Environ Health* 2012;38:391–392. <http://dx.doi.org/10.5271/sjweh.3315>.
5. Oksanen T, Vahtera J, Westerlund H, Pentti J, Sjösten N, Virtanen M, et al. Is retirement beneficial for mental health? Antidepressant use before and after retirement. *Epidemiology*. 2011;22:553–559. <http://dx.doi.org/10.1097/EDE.0b013e31821c41bd>.
6. Westerlund H, Kivimäki M, Singh-Manoux A, Melchior M, Ferrie JE, Pentti J, et al. Self-rated health before and after retirement in France (Gazel): a cohort study. *Lancet*. 2009;374:1889–1896. [http://dx.doi.org/10.1016/S0140-6736\(09\)61570-1](http://dx.doi.org/10.1016/S0140-6736(09)61570-1).
7. Westerlund H, Vahtera J, Ferrie JE, Singh-Manoux A, Pentti J, Melchior M, et al. Effect of retirement on major chronic conditions and fatigue: French Gazel occupational cohort study. *BMJ*. 2010;341:c6149. <http://dx.doi.org/10.1136/bmj/c6149>.
8. Laaksonen M, Metsä-Simola N, Martikainen P, Pietiläinen O, Rahkonen O, Gould R et al. Trajectories of mental health before and after old-age and disability retirement: a register-based study on purchases of psychotropic drugs. *Scand J Work Environ Health*. 2012;38:409–419. <http://dx.doi.org/10.5271/sjweh.3290>.
9. Bartley M, Ferrie J, Montgomery SM. Health and labour market disadvantage: unemployment, non-employment, and job insecurity. In: Marmot M, Wilkinson RG, editors. *Social determinants of health*. Oxford: Oxford University Press; 2006. p78–96.
10. Clarke P, Marshall VW, Weir D. Unexpected retirement from full time work after age 62: consequences for life satisfaction in older Americans. *Eur J Ageing*. 2012;9:207–219. <http://dx.doi.org/10.1007/s10433-012-0229-5>.
11. Pedersen J, Bjorner JB, Burr H, Christensen KB. Transitions between sickness absence, work, unemployment, and disability in Denmark 2004–2008. *Scand J Work Environ Health*. 2012;38:516–526. <http://dx.doi.org/10.5271/sjweh.3293>.
12. Burdorf A. The need for novel strategies to analyze the dynamic pattern of worker’s health over time and the consequences for sustained employability. *Scand J Work Environ Health*. 2012;38:485–488. <http://dx.doi.org/10.5271/sjweh.3325>.
13. Kaufman JS, Hernán MA. Epidemiologic methods are useless: they can only give you answers. *Epidemiology*. 2012;23:785–786. <http://dx.doi.org/10.1097/EDE.0b013e31826c30e6>.
14. Harper S, Strumpf EC. Social epidemiology: questionable answers and answerable questions. *Epidemiology*. 2012;23:795–798. <http://dx.doi.org/10.1097/EDE.0b013e31826d078d>.
15. Hafeman DM, Schwartz S. Opening the Black Box: a motivation for the assessment of mediation. *Int J Epidemiol*. 2009;38:838–845. <http://dx.doi.org/10.1093/ije/dyn372>.
16. De Stavola BL, Daniel RM. Marginal structural models: the way forward for life course epidemiology? *Epidemiology*. 2012;23:233–237. <http://dx.doi.org/10.1097/EDE.0b013e318245847e>.
17. Lange T, Hansen JV. Direct and indirect effects in a survival context. *Epidemiology*. 2011;22:575–581. <http://dx.doi.org/10.1097/EDE.0b013e31821c680c>.
18. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J Personality Soc Psychol*. 1986;51:1173–1182. <http://dx.doi.org/10.1037/0022-3514.51.6.1173>.
19. Cole SR, Hernán MA. Fallibility in estimating direct effects. *Int J Epidemiol*. 2002;31:163–165. <http://dx.doi.org/10.1093/ije/31.1.163>.
20. Aalen OO, Røysland K, Gran JM, Ledergerber B. Causality, mediation and time: a dynamic viewpoint. *J Royal Stat Soc Series A*. 2012;175, part 4:831–861. <http://dx.doi.org/10.1111/j.1467-985X.2011.01030.x>.
21. Pearl, J. *Causality: models, reasoning and inference*. 2nd edition. Cambridge: Cambridge University Press; 2009.
22. Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*. 2000;11:550–560. <http://dx.doi.org/10.1097/00001648-200009000-00011>.
23. Gamborg M, Jensen GB, Sørensen TIA, Andersen PK. Dynamic path analysis in life-course epidemiology. *Am J Epidemiol*. 2011;173:1131–1139. <http://dx.doi.org/10.1093/aje/kwq502>.
24. Christensen KB, Labriola M, Lund T, Kivimäki M. Explaining the social gradient in long-term sickness absence: a prospective study of Danish employees. *J Epidemiol Community Health* 2008; 62: 181–183. <http://dx.doi.org/10.1136/jech.2006.056135>.
25. Howards PP, Schisterman EF, Poole C, Kaufman JS, Weinberg CR. “Toward a clearer definition of confounding” revisited with directed acyclic graphs. *Am J Epidemiol*. 2012;176:506–511. <http://dx.doi.org/10.1093/aje/kws127>.
26. Anonymous. Theis Lange, winner of the Kenneth Rothman Epidemiology Prize, 2012. *Epidemiology*. 2012;23(5):655. <http://dx.doi.org/10.1097/EDE.0b013e318261c76c>

Petter Kristensen
Department of Occupational Medicine and Epidemiology
National Institute of Occupational Health
PO Box 8149 Dep, N-0033
Oslo Norway
[e-mail: pkr@stami.no]

Odd O Aalen
Department of Biostatistics
Institute of Basic Medical Sciences
University of Oslo
PO Box 1122 Blindern, 0317
Oslo, Norway
[e-mail: o.o.aalen@medisin.uio.no]