A novel framework for classification of selection mechanisms in epidemiological research

Jonas Bjork (✉ jonas.bjork@med.lu.se)

Lunds universitet Medicinska fakulteten <https://orcid.org/0000-0003-1883-2000>

Anton Nilsson

Lunds universitet Medicinska fakulteten

Carl Bonander

Goteborgs universitet Sahlgrenska Akademin

Ulf Strömberg

Goteborgs universitet Sahlgrenska Akademin

Research article

Keywords: bias, selection bias, epidemiologic factors, population characteristics, public health

Posted Date: November 5th, 2019

DOI: <https://doi.org/10.21203/rs.2.16849/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at BMC Medical Research Methodology on June 15th, 2020. See the published version at <https://doi.org/10.1186/s12874-020-01015-w>.

Abstract

Background: Non-participation, losses to follow up and other types of study-specific selection mechanisms can be serious concerns in epidemiological studies. There are also selection processes that result in non-random groupings and changes in the composition of populations. These are continuously on-going irrespectively of whether they are subject to sampling in empirical studies. Such population selections are often overlooked, but may lead to lack of comparability of exposed and unexposed populations or decrease study validity in other ways. The overall aim of this study was to develop a simple but general framework for classifying various types of selection mechanisms of relevance for epidemiological research.

Methods: We classify selection mechanisms in three dimensions: i) selection at the population level vs. selection that is study-specific, ii) type of mechanism (selection causing exposure vs. selection in population at risk), iii) timing of the selection (pre-exposure, during exposure or post-outcome).

Results: Examples from the epidemiological literature of selection mechanisms are discussed and classified according to the three dimensions of the proposed framework.

Conclusions: Increased mechanistic understanding of when, how, and why confusion of effects can occur because of selection is an important step towards improved validity of epidemiological research.

Background

Selection bias is a term that still lacks a consistent definition in epidemiology, and it has also varying meaning and usage in other scientific disciplines [1, 2]. This is a source of misunderstandings that may compromise validity, not the least in the interdisciplinary collaborations that are common within empirical research today. Most definitions of selection bias in epidemiology restrict the attention to sample selection in that selection is viewed as a bias-inducing phenomenon occurring at the study level, i.e. during design, enrolment, follow up or analysis [3–5]. The focus on study-related issues when discussing selection bias implies that selection occurring at the population level are often overlooked among applied researchers. As an example of this narrow focus, it is sometimes claimed that selection bias is not an issue in cohort studies involving complete recruitment and follow-up [6].

Another source of confusion in previous work on selection bias is the intermix between selection mechanisms and consequences of the selection. As a result, working definitions of selection bias vary depending on whether the definition includes consequences on internal or external validity [4, 7–9]. Unifactorial selection mechanisms involving exposure or outcome but not both may distort descriptive measures, (e.g. outcome rates, risks and prevalence proportions), but may also hamper external validity of causal effect measures if the effect is heterogeneous across population groups [9]. Bi- or multifactorial selection mechanisms involving both exposure and outcome are generally necessary in order to distort the internal validity of causal effect measures [6, 7].

The overall aim of this study was to develop a simple but general framework for classifying various types of selection mechanisms of relevance for epidemiological research. Examples from the epidemiological literature of selection mechanisms are discussed and classified according to the proposed framework.

Methods

Notation and framework

We define a selection mechanism as a process that results in a non-random split of a population or study cohort into two or more groups, with respect to a particular aspect such as exposure, health status, or some other variable of interest. Causal diagrams (directed acyclic graphs, DAGs) are useful for depicting and defining different types of selection mechanisms formally [7, 10]. Since the timing of a selection mechanism is crucial for judging its effect, it helps if the nodes are ordered chronologically from left to right in a causal diagram so that nodes and arrows placed under each other can be assumed to be acting within the same time window. We use subscripts for the nodes to indicate time, e.g. E_0 represents exposure initiation in time window $t = 0$ and E_1 exposure in some later time window $t = 1$.

In our proposed framework, we classify selection mechanisms in three dimensions. Firstly, these mechanisms can either be present at the population level (and thus occurring independently of study decisions) or at the study level (and thus occurring only because of the study). Secondly, we distinguish between two fundamental types of selection mechanisms: i) selection causing exposure, ii) selection in population at risk. Thirdly, we organize the selection mechanisms with respect to the timing relative the exposure and the outcome, i.e. selection occurring pre-exposure, during exposure or post-outcome. These three dimensions are described in more detail in the following paragraphs.

Results

1. Selection at the population vs. study level

Selection mechanisms can either be present at the *population level* (and thus occurring independently of study decisions) or at the *study level* (and thus occurring only because of the study). Epidemiologists are often familiar with non-participation, losses to follow up and other types of study-specific selection mechanisms that can be serious concerns in empirical research. However, there are also selection processes that result in non-random groupings and changes in the composition of the underlying populations. These are continuously on-going irrespectively of whether they are subject to sampling in empirical studies. Population selection occurs both within general populations and specific patient populations. It represents phenomena that only partially overlap with *confounding*, i.e. confusing of effects or lack of exchangeability between exposed and unexposed with respect to background risks for the disease outcome [6, 7]. The confounding resulting from population selection is sometimes apparent, such as differences in health determinants across groups in age, sex or socioeconomic characteristics, and therefore straightforward to adjust for in statistical analysis. However, population selection often

leads to subtle differences across groups that are more difficult to account for, for example occur if personal ambitions make individuals seek higher education, if self-interest drives individuals to choose the occupation that produces the highest utility for them, if physicians select patients into treatments and when health conscious individuals select themselves into preventive screening programs. Additionally, population selection includes phenomena that are distinct from confounding, for instance if an index event must occur in order for someone to enter the population at risk [8]. As examples, disease progression can only be observed among people with the disease, and being unemployed is often a prerequisite for taking part in a job training program. Population selection also act during exposure, for example through survival of the fittest and depletion of susceptibles over time [9]. A common characteristic of population selection effects is that they tend to persist also in “perfect” observational study settings, including register-based studies with complete population samples [10].

2. Type of selection mechanism

Using causal diagrams, we distinguish between two different types of selection mechanisms: i) *selection causing exposure* and ii) *selection in population at risk*. The origin of these mechanisms can either be unifactorial or bi/multifactorial.

Selection causing exposure (unifactorial case)

Each arrow between two nodes in a causal diagram represents a direct causal effect, but also a directed selection mechanism. To see why this the case, consider the direct effect $C_0 \rightarrow E_1$ in Figure 1A, which here means that C-positivity ($C_0 = 1$) increases the likelihood of becoming exposed ($E_1 = 1$). As a consequence, C-positivity will be more common among exposed individuals and less common among unexposed individuals than in the total population. Thus, the direct causal link between C and E is a selection process that leads to two selected groups in the population, exposed and unexposed that have different compositions with respect to C. We will refer to any arrow that ends in exposure as *selection causing exposure*. This selection mechanism does not affect the boundaries of the population as such (i.e. it does not select in or out from the combined population of unexposed and exposed). In occupational epidemiology, *healthy worker hire effect* is a well-known example of a selection mechanism causing exposure, if healthy individuals ($C_0 = 1$) in a population are more likely of becoming employed and thereby occupationally exposed ($E_1 = 1$) [11].

Selection in population at risk (unifactorial case)

The other fundamental selection mechanism, *selection in population at risk*, acts on the population or study cohort, for example by selecting individuals in or out from the population eligible for the exposure (e.g. only survivors until a certain age will be able to retire). It may also select individuals in or out from the population at risk during exposure (e.g. elevated preterm mortality in a drug addict population filters the population at risk of diseases typically occurring at older ages). As a result of selection in population at risk, there will be differences in the composition of selected and filtered subpopulations, e.g. non-

random differences between eligible and non-eligible or between those who remain at risk and those who do not. Selection in population at risk is depicted in Figure 1B, where the directed selection mechanism $C_0 \rightarrow S_1$ here means that C-positivity ($C_0 = 1$) will be more common those who remain in the population at risk ($S_1 = 1$) than among those who do not ($S_1 = 0$). At the population level, the boxed S represents what we refer to as *conditioning by nature*, which means that any causal action in this system downstream from this time window will only act on the selected population ($S_1 = 1$).

Selection causing exposure (bifactorial case)

Exposures are in reality often caused by several selection mechanisms, either acting independently or in concert. Importantly, the origin of some these selections can be unknown. In Figure 1C, a bifactorial selection mechanism with two independent, directed selections into a common effect E is depicted, one originating from a known source C and one from an unknown source U. This combined selection mechanism will induce inverse associations between C and U within strata of E that are different from those of the underlying population. As an example, suppose heavy alcohol consumption (E) is caused by low socioeconomic status (C) but also by a genetically determined tolerance for alcohol that allows escalation of drinking, here assumed to be unknown (U). The induced inverse association within strata of alcohol consumption will imply that individuals with low socioeconomic status will have a worse tolerance on average than individuals with high socioeconomic status within the same stratum.

Selection in population at risk (bifactorial case)

Bi- and multifactorial selection mechanisms also frequently operate on the population at risk. An example is depicted in Figure 1D, where a disease D has an exposure E as a known origin but also an unknown origin U. Any subsequent causal action on disease progression in D can only be observed among people with the disease ($D = 1$), which in the causal diagram is represented by D. Such conditioning by nature on an index event necessary for entering the population at risk creates inverse associations between the underlying determinants E and U. This may lead to bias commonly referred to as *index event bias* (or *collider stratification bias*) in studies where the collider D (i.e. the event where the two directed selection mechanisms collide) constitutes the population at risk [12, 13]. As an example, smoking is a well-established risk factor for the development of rheumatoid arthritis (RA) [13]. Nevertheless, null or even inverse associations between smoking and disease progression among patients with RA have been observed. The antiinflammatory role of nicotine has been put forward as a possible explanation of the lower systemic inflammation and structural disease progression in current smokers with RA [14]. However, index event bias stemming from selection mechanisms of the type depicted in Figure 1D is a compelling alternative explanation, as the necessary conditioning on the index event D (incident RA) induces spurious inverse associations between the exposure E (smoking) and other risk factors U that may also cause disease progression [13]. Another example of such paradoxical results that could be explained by index event bias is the apparent protective effect of obesity on mortality among end-stage renal disease patients [15].

3. Timing of the selection - pre-exposure, during exposure or post-outcome

Figure 1A–1D all represent selection mechanisms occurring prior to any exposure effect in the downstream population. At the study-level, selection occurring pre-exposure corresponds to non-random sample selection at study entry in for example surveys or cohort studies. Selection mechanisms may also affect exposure continuation, even when exposure initiation was not subject to selection. As an example, there was no apparent socioeconomic gradient (C) in the propensity to start smoking (E_0) among young adults in the 1950's [16]. However, the propensity to continue smoking (represented by E_1 in Figure 2A) despite growing evidence of serious health hazards exhibited a clear inverse socioeconomic gradient [17].

Additional examples

In Table 1, we show how selection mechanisms occurring at the population level can be classified according to our suggested framework. Some of these examples are described further below. Examples of selection that might occur at the study level are classified in Table 2.

Selection in exposure caused by side effects or subclinical symptoms

The selection from exposure initiation E_0 to exposure continuation E_2 is in many situations mediated by a side effect or an adverse event A_1 caused by E_0 (Figure 2B). This situation could for example occur if initial smokers ($E_0 = 1$) who experience bad cough ($A_1 = 1$) are more likely to quit smoking than others. Because of this selection, continuing smokers ($E_2 = 1$) would then have experienced bad cough to a lesser degree than those who have given up smoking ($E_2 = 0$). *Healthy worker survivor effect* is a similar selection mechanism within occupational epidemiology, implying that employees who can tolerate the physical or psychosocial working conditions are more prone to stay in the workplace and thereby remain occupationally exposed [11]. Selection may in some situations also cause changes in exposure post-outcome, for example if a certain exposure may relieve subclinical symptoms. As an example, it has been suggested that smoking in schizophrenia may begin in the prodromal phase of the disorder [18].

Competing event - avoidance necessary for staying in the population at risk

Exposure may also cause changes in the population at risk. Figure 2C represents a situation where an exposure E causes an adverse event D (with death as an extreme example) that precludes continued stay in the population at risk with respect to a different outcome Y . Such mutually exclusive outcomes are commonly referred to as *competing event (competing risk)* [19]. Continued causal actions beyond a certain time window can only occur if D has not occurred until then. This will imply that the exposure in the population at risk will gradually become more and more inversely related to other determinants of the adverse event (represented by U in Figure 2C). As an example, suppose we want to study the effect of sedentary lifestyle (E) on the risk for dementia (Y) at older ages. Thus, people will have to have survived

to older ages in order to become part of the exposed population at risk. This necessary conditioning on survival ($D = 0$) implies that surviving individuals with sedentary lifestyle can be expected to have a more favorable risk profile with respect to other determinants (U) of survival than surviving individuals with a physically active lifestyle. Bias would occur if these determinants are also related to the risk of dementia and are not accounted for.

Depletion of susceptibles

This situation, which is depicted in Figure 2D, is similar to the previous situation with a competing event in that the exposure E causes selection gradually in the population at risk. However, here the outcome at later time points (represented by Y_2 in Figure 2D) does not get competition from other outcomes but from the same outcome at earlier time points (Y_1). This phenomenon is commonly referred to as *depletion of susceptibles* [20], as the early events of interest caused by the exposure may lead to a depletion of susceptibles in the population at risk over time. As an example, an effect of smoking on a particular disease outcome may seemingly decrease over time, or even change direction, as early disease events “eat” of the causal components required for the disease to manifest at later time points. It has been suggested that the hazardous effect of smoking on mortality may disappear for ages 85 and above [21]. However, an alternative explanation is that the harmful effect of smoking is disguised by depletion of susceptibles, i.e. smokers among the oldest are likely to be a more selected group of survivors than non-smokers at the same age with respect to other determinants of survival. Unless accounted for, bias will occur in the causal inference regarding smoking effects.

Berkson's fallacy

Selection mechanisms altering the population at risk may also occur post-outcome. A well-known example is *Berkson's fallacy*, where the population of patients who come to the hospital is structurally different from patients with the same disease who for various reasons do not come [22]. The selected population can be structurally different from the filtered population with respect to a single determinant of the selection (Figure 1B with S as outcome) but also with respect to associations between different determinants (Figure 1D with D as outcome). Selection post outcome of the Berkson type may also affect validity in studies of malformations in live births [23], as malformations often increase the risk of miscarriages and these are often impossible to observe.

Discussion

The important distinction between population and study selection is presently lacking in most working definitions of selection bias. Early work by Kleinbaum et al. used the term selection bias to denote a distortion in the estimate of effect resulting from the manner in which subjects are selected into the study population [3]. Similarly, Delgado-Rodriguez & Llorca defined selection bias as the error introduced when the study population does not represent the target population [19]. However, these restrictive definitions to a large extent exclude selection occurring at the population level, which may therefore fly under the radar

in the mind of applied researchers when interpreting their results. Furthermore, in the commonly used working definition of confounding (confusion of effects [24]), the reason why exposed and unexposed individuals differ with respect to a third variable is not specified. Knowledge about the underlying selection mechanisms are essential for the choice of adequate study design, appropriate statistical methods and correct analytical decisions [8]. We therefore find it hazardous to have restrictive or unspecific definitions of central validity concepts in epidemiology that do not naturally lead investigators to think carefully about potential reasons why the populations of exposed and unexposed individuals may lack comparability, either already at baseline or evolving gradually during follow up.

Our proposed framework may bridge some of the gaps between previous work on selection in different scientific disciplines that have either viewed selection as a phenomenon that occurs prior to exposure [25, 26], or as a phenomenon that occurs downstream the exposure from selecting on a collider [7, 10]. The Cochrane handbook for systematic reviews of interventions defines selection bias in clinical observational studies as “systematic differences between baseline characteristics of the groups that arise from self-selection of treatments, physician-directed selection of treatments, or association of treatment assignments with demographic, clinical, or social characteristics” [26]. Self-selection is thus a keyword for this definition, i.e. the selection occurs into exposure or treatment. The baseline characteristics for which the differences are observed are not necessarily the true underlying factors responsible for the selection, but are likely to be associated with them. A related issue is therefore to what extent the resulting selection effects can be accounted for by the observed differences in baseline characteristics.

Selection into social programs was the specific selection mechanism considered in a seminal paper by Heckman et al. [25]. The authors made a decomposition of the resulting selection effect into three components: 1) systematic differences in covariate patterns among treated and untreated, 2) lack of sufficient support (overlap) for certain covariate patterns, 3) selection bias, precisely defined. With this decomposition, confounding in the sense that the treatment effect is confused with that of the covariates would be due to differences and lack of support in covariate pattern (component 1 and 2), whereas any differences not attributable to the observed covariates would be the remaining selection effect (component 3).

Hernan et al. [7] have proposed a now widespread distinction between confounding and selection bias based on causal diagrams: confounding occurs when there exist common causes of exposure and outcome, while selection bias results from conditioning on common effects of exposure and outcome. The underlying mechanism in the Hernan selection is, as noted above, different from the selection bias defined by Heckman [25], where selection is into exposure rather than as a consequence of the exposure. In our proposed framework, we argue that selection mechanisms are ubiquitous, i.e. present in all the arrows of a causal diagram, and we therefore include both Hernan and Heckman selection in our proposed framework.

Elwert et al.[2] have used a similar organization of selection mechanisms affecting the study population according to the timing of the collider relative to the exposure and the outcome. Our framework allows for

distinction between selection mechanisms and the potential consequences of selection, which are closely linked to the context and study objectives. As an example, suppose there is a selection mechanism affecting a particular population at risk at exposure initiation in a unifactorial manner (e.g. related to health status only). An investigator whose main interest is to report a valid estimate of the relative exposure effect may be less concerned about this selection than another investigator whose objective is to report risk differences (or number needed to harm), assess the etiologic fraction, or generalize the findings to the filtered population.

Conclusions

Based on chronologically ordered causal diagrams, we have proposed a framework for classification of selection mechanisms occurring at the population or study level, useful for epidemiology as well as for other areas of population research. We distinguished between selection mechanisms that cause exposure initiation or continuation and selection mechanisms that cause changes in the population at risk. We further organized the selection mechanisms with respect to the timing relative the exposure and the outcome.

Our proposed framework can be used to identify and classify selection mechanisms that may lead to lack of comparability of exposed and unexposed populations or decrease study validity in other ways, either already at exposure initiation or gradually evolving during follow up. We expect the use of this framework to increase awareness among applied researchers of how selection mechanisms at the population level may influence, or even jeopardize, validity of epidemiological research.

Abbreviations

DAGDirected Acyclic Graph

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Available of data and materials

Not applicable.

Competing interests

None of the authors has any competing interest to declare.

Funding

This work was funded by the Swedish Research Council for Health, Working Life and Welfare (FORTE, grant no. 2017–00414 and 2018–01596).

Authors' contributions

All authors conceived the study and JB developed it further. JB drafted the manuscript; AN, CB and US revised it for important intellectual content. All authors read and approved the final version of the manuscript.

Acknowledgements

Not applicable.

References

1. Infante-Rivard C, Cusson A: *Reflection on modern methods: selection bias-a review of recent developments. International journal of epidemiology* 2018, *47*(5):1714–1722.
2. Elwert F, Winship C: *Endogenous Selection Bias: The Problem of Conditioning on a Collider Variable. Annual review of sociology* 2014, *40*:31–53.
3. Kleinbaum DG, Morgenstern H, Kupper LL: *Selection bias in epidemiologic studies. American journal of epidemiology* 1981, *113*(4):452–463.
4. Porta M (ed.): *A dictionary of epidemiology*. New York: Oxford University Press Inc; 2014.
5. Chang CH: *Applications of the propensity score weighting method in psychogeriatric research: correcting selection bias and adjusting for confounders. International psychogeriatrics* 2017, *29*(5):703–706.
6. Pearce N, Checkoway H, Kriebel D: *Bias in occupational epidemiology studies. Occupational and environmental medicine* 2007, *64*(8):562–568.
7. Hernan MA, Hernandez-Diaz S, Robins JM: *A structural approach to selection bias. Epidemiology (Cambridge, Mass)* 2004, *15*(5):615–625.

- 8.Haneuse S: *Distinguishing Selection Bias and Confounding Bias in Comparative Effectiveness Research. Medical care* 2016, *54*(4):e23–29.
- 9.Hernan MA: *Invited Commentary: Selection Bias Without Colliders. American journal of epidemiology* 2017, *185*(11):1048–1050.
- 10.Pearl J: *Causal diagrams for empirical research. Biometrika* 1995, *82*:669–710.
- 11.McNamee R: *Confounding and confounders. Occupational and environmental medicine* 2003, *60*(3):227–234; quiz 164, 234.
- 12.Dahabreh IJ, Kent DM: *Index event bias as an explanation for the paradoxes of recurrence risk research. Jama* 2011, *305*(8):822–823.
- 13.Choi HK, Nguyen US, Niu J, Danaei G, Zhang Y: *Selection bias in rheumatic disease research. Nature reviews Rheumatology* 2014, *10*(7):403–412.
- 14.Vesperini V, Lukas C, Fautrel B, Le Loet X, Rincheval N, Combe B: *Association of tobacco exposure and reduction of radiographic progression in early rheumatoid arthritis: results from a French multicenter cohort. Arthritis care & research* 2013, *65*(12):1899–1906.
- 15.Flanders WD, Eldridge RC, McClellan W: *A nearly unavoidable mechanism for collider bias with index-event studies. Epidemiology (Cambridge, Mass)* 2014, *25*(5):762–764.
- 16.Wald N, Kiryluk S, Darby S, Doll R, Pike M, Peto R: *UK Smoking Statistics*. Oxford: Oxford University Press; 1988.
- 17.Jarvis MJ: *A profile of tobacco smoking. Addiction (Abingdon, England)* 1994, *89*(11):1371–1376.
- 18.Leonard S, Mexal S, Freedman R: *Smoking, Genetics and Schizophrenia: Evidence for Self Medication. Journal of dual diagnosis* 2007, *3*(3–4):43–59.
- 19.Delgado-Rodriguez M, Llorca J: *Bias. Journal of epidemiology and community health* 2004, *58*(8):635–641.
- 20.Stovitz SD, Banack HR, Kaufman JS: *'Depletion of the susceptibles' taught through a story, a table and basic arithmetic. BMJ evidence-based medicine* 2018, *23*(5):199.
- 21.Cohen-Mansfield J: *Smoking and mortality among persons aged 75–94. Preventive medicine* 2013, *56*(3–4):185–189.
- 22.Berkson J: *Limitations of the application of fourfold table analysis to hospital data. International journal of epidemiology* 2014, *43*(2):511–515.

23. Leck I: *McKeown, Record, and the epidemiology of malformations. Paediatric and perinatal epidemiology* 1996, *10*(1):2–16.
24. Rothman KJ: *Epidemiology - an introduction*, 2nd edn. New York: Oxford University Press Inc.; 2012.
25. Heckman JJ, Ichimura H, Smith J, Todd P: *Sources of selection bias in evaluating social programs: an interpretation of conventional measures and evidence on the effectiveness of matching as a program evaluation method. Proceedings of the National Academy of Sciences of the United States of America* 1996, *93*(23):13416–13420.
26. Higgins JPT, Green S (eds.): *Cochrane handbook for systematic reviews of interventions (version 5.1.0, updated March 2011)*: The Cochrane Collaboration; 2011.

Tables

Table 1. Commonly described selection mechanisms at the population-level in the epidemiological literature classified according to the proposed framework.

	Selection mechanism	
Timing of the exposure	Selection causing exposure	Selection in population at risk
Pre-exposure	Self-selection Confounding by indication Healthy worker hire effect	Index event
During exposure	Healthy worker survivor effect	Competing event Depletion of susceptibles
Post-outcome	Subclinical symptoms relieved by exposure	Berkson's fallacy

Table 2. Commonly described study-specific selection mechanisms in the epidemiological literature classified according to the proposed framework.

	Selection mechanism	
Timing of the exposure	<i>Selection causing exposure</i>	<i>Selection in population at risk</i>
<i>Pre-exposure</i>	Selective participation (RCT ^a)	Study base definition Selective participation (cohort study)
<i>During exposure</i>	Non-compliance (RCT ^a)	Loss to follow-up (cohort study) Collider stratification
<i>Post-outcome</i>	Matching (case-control studies)	Selective participation (case-control study)

^a Randomized controlled trial

Figures

Figure 1

Fundamental selection mechanisms illustrated by causal diagrams with subscripts of the nodes indicating time. a. Unifactorial selection causing exposure (E1) b. Unifactorial selection in population at risk (boxed S1) c. Bifactorial selection causing exposure (E1) d. Bifactorial selection in population at risk (boxed D1, induced inverse association is marked with a dashed line).

Figure 2

Selection during exposure illustrated by causal diagrams with subscripts of the nodes indicating time. a. A known determinant (C) does not influence exposure initiation (E_0) but exposure continuation (E_1) b. E_0 causes a side effect (A_1) that influences E_2 . c. E_0 causes an index or competing event (boxed D_1) that precludes the disease outcome (Y_2) d. E_0 causes an early disease event (boxed Y_1) that leads to depletion of susceptibles that precludes later disease events (Y_2) to occur. Induced inverse associations are marked with dashed lines.