Community Dental Health (2020) **37,** 90–95

doi:10.1922/CDH_SpecialIssueCeleste06

Bias in population oral health research: longitudinal studies

[Special issue of Community Dental Health, to be disseminated at 'Random and Systematic Bias in Population Oral Health Research' IADR symposium, March 2020, Washington DC.]

Roger Keller Celeste

Faculty of Dentistry, Federal University of Rio Grande do Sul, Porto Alegre, Brazil.

***Abstract*:** Bias in longitudinal studies have been well described and the longer the follow-up, the higher the proportion of drop-outs. Here, I present some key issues related to selection bias, time-varying confounders, solutions to bias and challenges in longitudinal studies in dental research. Selection bias creates distortions in measures of disease frequency or association due to losses of follow-up or use of specific population groups. It is shown that even if losses are not associated with baseline values, measures such as odds ratios may be seriously distorted. Such problems can be understood by directed acyclic graphs, identifying the collider bias, or by missing data theory. Time-varying confounding occurs when an exposure varies over time and is affected by past exposure of other time-varying covariates, creating a complex scenario to adjustment in multiple regression. Under some assumptions, missing information may be informed by other variables in the dataset, and techniques such as multiple imputation or inverse probability weighting can be helpful, but the best solution is to prevent losses of follow-up as much as possible. Finally, I present challenges for longitudinal studies that use electronic health records and the need to incorporate area-based contextual measures. The first allows linkage of dental records with other information systems to create longitudinal (big) data. The second allows evaluation longitudinally of the effect of contextual factors, including social and health policies, on oral health.

Keywords*: epidemiology; oral health; dental public health, bias; longitudinal studies.*

Introduction

Longitudinal studies in epidemiology are an important methodological design that establishes temporal sequence between events, reducing recall bias (Hardt and Rutter, 2004), and helping researchers to understand complex relationships among risk factors. Over time, this type of study has become more frequent in dental public health (Celeste *et al*., 2016), indicating availability of such data. Additionally, demographic changes demanded a shift from short term longitudinal studies among children to larger and longer cohorts among adults and older people (Celeste *et al*., 2016).

An important issue when starting a cohort is to decide for how long individuals should be followed. This decision depends on the pathogenesis of the disease of interest and the risk factors we would like to study. The longer the time of follow-up, the more repeated measures will be required and more likely individuals will be lost.

Decisions concerning follow-up times may involve several factors. One aspect is the induction time, that is the time an individual must be exposed to a risk factor up to a level that it irreversibly leads to the disease (Greenland and Lash, 2008; K J Rothman, 1981). The induction time depends on the risk factor and is usually an unknown period. Another aspect is the latency time that can be defined as the time from the beginning of a disease until its clinical diagnosis (Greenland and Lash, 2008; K J Rothman, 1981). This is a disease-specific characteristic that has been described for some oral diseases. For example, clinically detectable dentine carious lesions in children have been estimated to require between four to seven years (Arrow, 2007; Kopycka-Kędzierawski *et al*., 2004). Gingivitis takes around two to three weeks to develop, if oral hygiene is refrained (Löe *et al*., 1965). Evidence of untreated sites of periodontitis among adults suggest that bone loss may progress between 0.3mm to 1.0mm per year (Haas *et al*., 2012; Reddy *et al*., 2000). Oral cancer has a much longer latency, varying from 12 years for the oropharynx to 26 years to tongue tumours (Nadler and Zurbenko, 2014). Having a minimum time of follow-up defined, many operational issues still remain and most will impose difficulties maintaining data quality. After many years of follow-up of a large sample, it is in the interest of all stakeholders and participants to ensure bias in the data is minimal.

Two parameters are usually affected by any bias: the point estimates (i.e. means and proportions) and measures of associations (e.g. attributable or relative risk). In this paper, I will refer mainly to distortions in associations because this is the main objective in aetiological studies. Unbiased point estimates (e.g. prevalence or incidence) are the main objective of epidemiological surveys and therefore, less common in longitudinal studies.

Correspondance to: Roger Keller Celeste, Department of Preventive and Social Dentistry, Faculty of Dentistry, Federal University of Rio Grande do Sul, Rua Ramiro Barcelos 2492, 3º floor, Porto Alegre – RS - CEP 90035-003. E-mail: roger.keller@ufrgs.br

The problem of bias in longitudinal studies

Bias is any distortion in estimates of the true population parameters (Greenland and Lash, 2008; Szklo and Nieto, 2007). In epidemiology, selection and measurement (or information) bias, along with issues of confounding are the most commonly described. Although the list of types of bias is long (Delgado-Rodríguez and Llorca, 2004), in longitudinal studies, loss of follow-up is an important selection bias originating mainly from deaths during the study period (differential survival) or attrition (non-participation). While the first is out of a researcher's control, the second is reasonably manageable. Survival bias may distort estimates in cross-sectional studies where prevalent cases represent those who survived prior to the start of the study (Neyman bias) (Banack *et al*., 2019; Delgado-Rodríguez and Llorca, 2004). In this paper, biases that are common to other studies will not be targeted as they have been extensively described elsewhere (Beck *et al*., 1997; Pahel *et al*., 2011; Preisser *et al*., 2017; Slade and Caplan, 1999) nor will issues concerning confounding (Merchant and Pitiphat, 2002).

Selection Bias: understanding the mechanisms

The mechanism for selection bias can be explained by two different approaches. The first, using missing data theories can be helpful, as proposed by Rubin (1976) and extensively described in the statistical literature (Graham, 2009; Schafer and Graham, 2002). This approach describes non-causal mechanisms that may alter probabilities of selection and treats missing data as a source of bias, regardless of the reason for the absence, be it loss of follow-up due to change of address, selective survival, or any other. In the second, using Directed Acyclic Graphs (DAGs) to explain causal relations among variables has gained popularity in epidemiology (Hernán *et al*., 2004). In this approach, a general rule is applied to identify the so called "collider bias", or "confounding by indication".

Rubin defined missing data according to three mechanisms, missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR) (Rubin, 1976; Schafer and Graham, 2002) The first case, MCAR, may be very simple if we understand that a sample selection based on simple randomization means that all individuals not selected (unit non-response) were so because of chance. In other cases, an individual may participate in a given study, but leave some questions blank (item non-response) due to chance. However, MAR happens when the missing values of variable X do not depend on themselves, but are associated with variable Y. Finally, MNAR happens when the missing data depends on its own values. For example, in a longitudinal study, it may occur if individuals that are non-responders in the second wave decided systematically not take part because their baseline clinical exams showed they had no problems (they felt "safe"). Clearly, the mean values will be altered and standard deviation will decrease because of the homogeneity of remaining individuals. Although this situation may not be related to other variables, it will likely affect associations.

As an example, Table 1 shows the proportions of two variables, sex and people with teeth in good conditions,

Table 1. Marginal distribution of sex and teeth conditions of individuals at the baseline in 1968 in the whole sample and among survivors in 2011.

	Baseline		*True Survivors*		*Hypothetical Survivors*	
Sex	*%*	*(n)*	*%*	*(n)*	*%*	*(n)*
Male	49.9	(467)	38.1	(138)	49.9	(181)
Female	50.1	(469)	61.9	(225)	50.1	(182)
Total	100	(936)	100	(363)	100	(363)
Teeth not in good conditions in 1968?						
No	61.2	(573)	54.5	(198)	60.9	(221)
Yes	38.8	(363)	45.5	(165)	39.1	(142)
Total	100	(936)	100	(363)	100	(363)

Source: (Celeste and Fritzell 2018)

based on actual longitudinal data (Celeste and Fritzell, 2018). While the first column presents values at baseline in 1968, the second and third columns show those values considering 61% of losses in 2011. Although, it is a high proportion, selection bias may arise with losses as low as 5% (if all losses concentrate in one cell) and the often-heard "rule-of-thumb", that 20% of losses may be acceptable, has long been refuted (Greenland, 1977). The second column is a case of MAR, as missing data on sex and teeth in good condition were associated with survival in unadjusted analysis. However, the third column shows absence was not associated with hypothetical cases of survival, but it is also a case of MAR, as missing cases of sex (or teeth) were unrelated to their own values. Note that cases of MAR may or may not bias the association between sex and teeth in good condition, as shown in Table 2. What initially appeared as biased data (true survivors) was unbiased based on the odds ratio (full data OR=1.31 versus survivors OR=1.34), while the two scenarios of hypothetical survivors could result in an increase or decrease of the magnitude of association (OR=2.45 or OR=0.60). The scenario of bias in association, when the losses are unrelated to the exposure, as in our hypothetical survivors, was demonstrated by Greenland (1977).

DAGs are useful because they provide an important conceptual aid to anticipate problems in design of longitudinal studies and facilitate selection of appropriate controls in multiple regression analysis. Details about rules and interpretation of DAGs can be seen elsewhere (Glymour, 2006; Merchant and Pitiphat, 2002). The issue about collider bias lays in the D-separation rule, which states that, if two variables are not associated, they will be associated if conditioned – restricted – to a common effect. Here, I illustrate with two hypothetical examples what may happen in dental caries research. In Figure 1a) root caries will not be associated with calcium deficiency in the general population. However, among a specific population (for example institutionalized older people), it is reasonable to presume root caries would be negatively associated. This bias is not trivial, as restricting to institutionalized older people has indirectly created a spurious association (backdoor) between lack of physical mobility (including toothbrushing) and difficulties in food

Table 2. Distribution of teeth conditions by sex in 1968 considering all individuals at baseline in 1968 and only survivors in 2011.

	Teeth not in good conditions in 1968?					
	Yes		*No*		*Total*	
	n	*(%)*	*n*	*(%)*	*n*	*(%)*
	Baseline Odds Ratio=1.31					
Male	196	(42.0%)	271	(58.0%)	467	(100%)
Female	167	(35.6%)	302	(64.4%)	469	(100%)
Total	363	(38.8%)	573	(61.2%)	936	(100%)
	True Survivors Odds Ratio=1.34					
Male	69	(50.0%)	69	(50.0%)	138	(100%)
Female	96	(42.7%)	129	(57.3%)	225	(100%)
Total	165	(45.5%)	198	(54.5%)	363	(100%)
	Hypothetical Survivors 1 Odds Ratio=2.45					
Male	90	(49.7%)	91	(50.3%)	181	(100%)
Female	52	(28.6%)	130	(71.4%)	182	(100%)
Total	142	(39.1%)	221	(60.9%)	363	(100%)
	Hypothetical Survivors 2 Odds Ratio=0.60					
Male	60	(33.1%)	121	(66.9%)	181	(100%)
Female	82	(45.1%)	100	(54.9%)	182	(100%)
Total	142	(39.1%)	221	(60.9%)	363	(100%)

Source: (Celeste and Fritzell 2018)

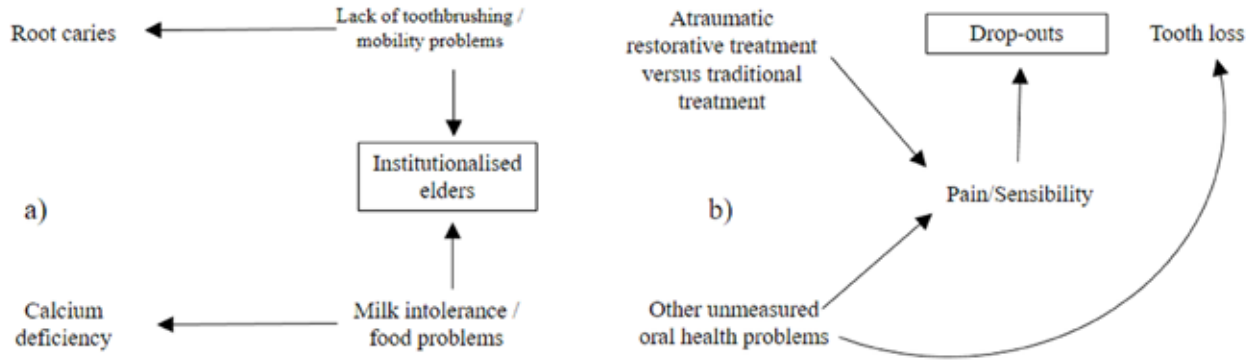

Figure 1. Casual diagram showing hypothetical selection bias in two cases: a) the association between root caries and calcium deficiency due to conditioning on institutional elders. In this example, most elders that have milk intolerance will not have difficulty in toothbrushing; therefore those with calcium deficiency will develop less root caries due to selection bias. b) in a clinical trial two factors can lead to drop-outs, only one is associated with tooth loss; therefore conditioning one drop-outs will create a biased association between atraumatic restorative treatment with tooth loss. In both cases, bias cannot be removed because the causes of conditioning are unmeasured or unknown.

preparation (including milk intolerance) of. Some older people will have both problems, but most will have only one. Those with a milk intolerance may not have problems tooth brushing. In light of this, researchers need to be cautious when developing inclusion and exclusion criteria for sample selection, as this is usually a source of bias. Secondly, the problem in using population subgroups is not an issue of lack of generalizability to other populations, but lack of internal validity, as the association is spurious even among those older people. The degree of bias, and even if there is no bias at all, will vary and we cannot predict with certainty.

In Figure 1b), we also have bias introduced by restricting analysis to those that remained in a clinical trial to investigate if atraumatic restorative treatment (ART) may increase tooth survival. Again, this is an indirect bias, because the collider is not the drop-out variable but its cause. It is plausible that individuals with dental pain or dentin sensitivity may not attend a recall for the trial. If only ART could cause pain and sensibility, no bias would happen, but in fact, other dental and mouth problems can cause pain and sensibility. Therefore, conditioning on those who did not drop-out is equivalent to conducting analysis with those without pain/sensibility, which opens a backdoor for an association between ART and other dental/mouth problem. As other dental problems may lead to tooth loss, then such analysis would show a biased association between ART and tooth loss.

Time-varying covariates

Some exposures vary over time and are affected by past exposure of other time-varying covariates, a situation that often (if not always) occurs in practice (Mansournia *et al*., 2017). In oral epidemiological research, examples of time-varying exposures or time-varying confounders may include smoking status, sugar consumption, income and fluoride exposure. Usually, individual time-varying information is collected only during the period when subjects are observed and is difficult to be addressed in cross-sectional studies.

Figure 2 represents a hypothetical cohort. This figure shows the exposure to sugar consumption at baseline (t0) and at the first follow-up (t1). This time-varying variable is a risk factor for dental caries (t1), but there is reverse causation, as dental caries (t0) affects sugar consumption (t1). Sugar (t0) is also a time-varying confounder between obesity (t1) and dental caries (t1). Finally, because past exposures affect future exposure, sugar (t1) mediates indirect effects of sugar (t0), or obesity (t0), on caries (t1). Controlling for such problems requires longitudinal studies with repeated measures, but conventional methods, as multiple regression, will over-adjust when controlling for mediators that are also confounders and open backdoor associations when controlling for colliders.

Commonly, epidemiologists want to estimate the total (direct) non-confounded effect of a risk factor on a disease. For this, an adequate solution is the use of G-methods (Mansournia *et al*., 2017), although the use of path analysis can also incorporate repeated measures of exposure and outcome, incorporating reverse causation (see (Darin-Mattsson *et al*., 2018)).

Some ways to correct and prevent bias in longitudinal studies

Again, bias may lay on measurement or selection problems and so, there are distinct solutions for each case. Most methods to correct bias are based on correcting measurement error. For such cases, when values of specificity and sensitivity are available, and no confounding exists (e.g randomized trials), a correction can be implemented (Antunes, 2019; Greenland and Lash, 2008). Methods to correct selection bias usually rely on external sources, such as population census or baseline data – to reconstruct the proportions by sex, age and possible other important factors (Graham, 2009). While such methods can reconstitute point estimates, it is unclear if they can correct distortions in associations because there is no information on distributions of disease cases by covariates (see examples in Table 2).

As discussed, selection bias can be understood as a case of missing data. For cases of MCAR, the use of complete-case analysis, also called listwise deletion, is satisfactory. Under the MCAR assumption, missing data on prevalence or incidence varies randomly from their true values. However, if we have more missing data among men than among women, overall means and proportions will be biased. If those missing cases are related to a known variable, then it is a case of MAR. In epidemiological surveys, absence is the result of sampling design. This information can be used to produce sampling weights to correct means and standard-deviations. Under MAR assumptions, in longitudinal studies, point estimates can be reconstituted from baseline information. For example, if we have lost half of the men from the first to the second wave, the remaining men may be assign

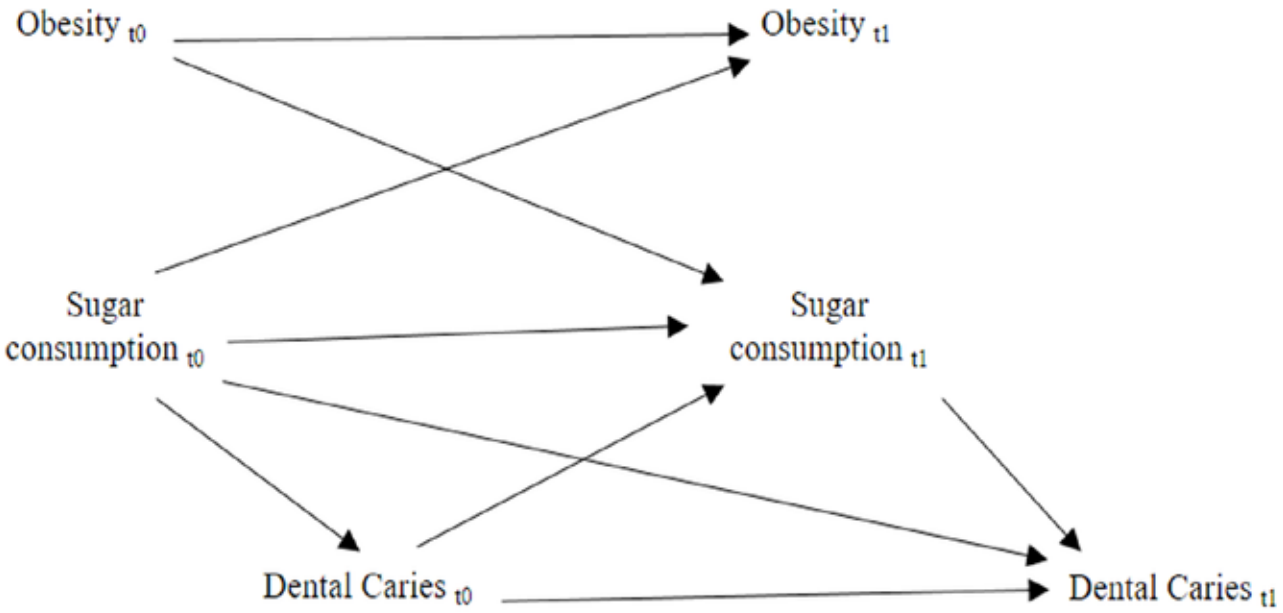

Figure 2. Time varying covariates in longitudinal studies about dental caries with repeated measures: a) $sugar_{t0}$ is confounding factor for associated $obesity_{t1}$ and $caries_{t1}$; b) $sugar_{t1}$ is a mediator for total effect of $sugar_{t0}$ on $caries_{t1}$; c) $sugar_{t1}$ is collider for the association between $obesity_{t0}$ and $sugar_{t0}$; d) there is sample selection due to reverse causation of $caries_{t0}$ on $sugar_{t1}$

a weight of two, and this idea is the basis for inverse probability weighting (IPW). Multiple imputation (MI) is also based on internal information to recover missing data, as long as this is a case of item non-response and good predictors do exit in the data set. Under MNAR, no statistical technique seems good, although under some probability assumptions, it is useful for sensitivity analysis (Graham, 2009), where we can test the effects of different scenarios.

Preventing missing datahas always been associated with good research practices and also quality control (Van Den Broeck *et al*., 2005). In cases where missing data is unavoidable such as deaths in dental geriatric and gerodontology studies, some suggestions may be at hand. First, if we believe that selective survival does not influence the associations of interest, then the major problem is lack of statistical power. For this problem, the use of open cohort may be a solution because individuals may enter the cohort at any time. Although it brings some complexity to analysis, statistical alternatives to cumulative risk have been extensively described (Beck *et al*., 1997; Slade and Caplan, 1999). Second, under the MAR assumption, several variables that may predict loss of follow-up should be collected in the first instance. Techniques such as MI, IPW, propensity scores or instrumental variables may be used to yield unbiased results (Banack *et al*., 2019).

Methodological and Theoretical Challenges for Longitudinal Studies

Methodological challenges for longitudinal studies in dental caries have long been described (Slade and Caplan, 1999). Unfortunately, some are yet to be resolved and some new issues have appeared. In this section, I will briefly consider two contemporary challenges for longitudinal studies: 1) the use of electronic health records, and 2) incorporation of area-based contextual measures.

The use of dental records has been described as secondary data and used as aggregate variables, mostly in ecological studies, with few examples using them at an individual level (Leake and Werneck, 2005; Olsen, 2008). Nevertheless, the advent of electronic records accessed electronically, with linkage to several databases, brought the possibility of using individual level big data in dental research. Current availability of electronic health registers may also allow large (retrospective) cohort studies combining multiple generations with clinical, socioeconomical, pharmacological, behavioural and other data, sometimes representative of the general population if the country has universal dental health coverage.

There are challenges to overcome before widespread use of electronic health records in health research analysis becomes the norm. Firstly, concerns about lack of standardized diagnostic criteria and changes over time affect quality and make longitudinal comparisons almost impossible. If non-differential misclassification happens, then associations may be washed out, but if lack of quality means systematic error, then associations may be seriously distorted. Despite that, there is some evidence of validity of data from dental records, such as tooth loss (Ljung *et al*., 2019). Secondly, selection bias introduced by constraining analysis to healthcare users is also a major concern, which hinders several research questions. For example, while most healthcare users have some health problems that lead them to seek treatment, many health insurance companies deny (or delay) treatment for a long list of pre-existing conditions leading to a healthier population. When data stem from screening programs, such as in school prevention dental programs, then this bias may be minimized. Finally, ethical concerns about privacy are also important (Olsen, 2008). Overall, the more researchers use such data, the faster identified problems will be solved.

Area-based contextual measures have been important determinants in oral health and many measures have been used in social science and policy analysis (Aguiar *et al*., 2017; Locker, 2000). Contextual measures in dentistry include social capital, water fluoridation, access to healthcare, public policies, availability of sugary foods at local shops, neighbourhood infra-structure or contextual deprivation, environmental exposures and any variables that are not a characteristic of an individual.

Most contextual studies use cross-sectional data, and longitudinal studies present some challenges. One of which concerns the geographical mobility of study participants, exposing them to different contexts; therefore, the correct identification of the exposure level is difficult. For example, people commute to work within large areas, and the idea of a small and homogeneous area may not be possible. Another concern is definition of the size and borders of the geographical area that has to match the exposure concept. For example, a study about a local health policy may use the catchment area of healthcare centres, but another study about social capital may need an area with borders socially defined by neighbourhoods. The use of census tracks is the most common solution, because of the availability of valid estimates at a contextual level. Nonetheless, census data may not have the desired theoretical meaning and may mismatch the area borders. A third issue is the unknown induction time, posing difficulties in defining time to follow-up. Finally, even if the sample size of individuals within each area is sufficient, there must be a large number of areas to allow between-area variability, then the final sample size may be large and geographically spread.

Concluding remarks

Several types of bias may occur in longitudinal oral health studies, but selection bias is the most common threat to their validity. Researchers need to understand it to prevent and deal adequately. The use of DAGs to identify colliders must be encouraged in early stages of study concept. Nonetheless, missing data have to be dealt with carefully and details of how they were handled must be explicit.

Individual contribution of each author

RKC contributed to development of study design, review of literature, drafting of the manuscript and approval of final version.

Acknowledgments

RKC holds a PQ2-CNPq Fellowship from the Brazilian National Research Council (number 310978/2016-5). I would like to thanks Fernando Neves Hugo and Juliana Balbinot Hilgert for kind comments on a previous version.

References

Aguiar, V. R., Pattussi, M. P. and Celeste, R. K. (2017): The role of municipal public policies in oral health socioeconomic inequalities in Brazil: A multilevel study, *Community Dentistry and Oral Epidemiology.* **46**, 1–6.

Antunes, J. L. F. (2019): How to correct the odds ratio for lack of sensitivity and specificity. *Community Dentistry and Oral Epidemiology*. https://doi.org/10.1111/cdoe.12489

Arrow, P. (2007): Incidence and progression of approximal carious lesions among school children in Western Australia, *Australian Dental Journal.* **52**, 216–226.

Banack, H. R., Kaufman, J. S., Wactawski-Wende, J., Troen, B. R. and Stovitz, S. D. (2019): Investigating and Remediating Selection Bias in Geriatrics Research: The Selection Bias Toolkit. *Journal of the American Geriatrics Society* **67**, 1970–1976.

Beck, J. D., Lawrence, H. P. and Koch, G. G. (1997): Analytic approaches to longitudinal caries data in adults. *Community Dentistry and Oral Epidemiology.* **25**, 42–51.

Celeste, R. K., Broadbent, J. M. and Moyses, S. J. (2016): Half-century of Dental Public Health research: bibliometric analysis of world scientific trends. *Community Dentistry and Oral Epidemiology* **44**, 557–563.

Celeste, R. K. and Fritzell, J. (2018): Do socioeconomic inequalities in pain, psychological distress and oral health increase or decrease over the life course? Evidence from Sweden over 43 years of follow-up. *Journal of Epidemiology and Community Health* **72**, 160–167.

Darin-Mattsson, A., Andel, R., Celeste, R. K. and Kåreholt, I. (2018): Linking financial hardship throughout the life-course with psychological distress in old age: Sensitive period, accumulation of risks, and chain of risks hypotheses. *Social Science & Medicine* **201**, 111–119.

Delgado-Rodríguez, M. and Llorca, J. (2004): Bias. *Journal of Epidemiology and Community Health* **58**, 635–641.

Glymour, M. M. (2006): Using causal diagrams to understand common problems in social epidemiology. 387–422 *in* J. M. Oakes and J. S. Kaufman, eds., *Methods in Social Epidemiology* Jossey-Bass, San Francisco.

Graham, J. W. (2009): Missing Data Analysis: Making It Work in the Real World. *Annual Review of Psychology* **60**, 549–576.

Greenland, S. (1977): Response and Follow-up Bias in Cohort Studies. *American Journal of Epidemiology* **106**, 184–187.

Greenland, S. and Lash, T. L. (2008): Bias Analysis (3rd ed.), 345–380 in Kenneth J Rothman, S. Greenland, and T. L. Lash, eds., *Modern Epidemiology.* Lippincott-Williams, Philadelphia.

Haas, A. N., Gaio, E. J., Oppermann, R. V., Rösing, C. K., Albandar, J. M. and Susin, C. (2012): Pattern and rate of progression of periodontal attachment loss in an urban population of South Brazil: a 5-years population-based prospective study. *Journal of Clinical Periodontology*, **39**, 1–9.

Hardt, J. and Rutter, M. (2004): Validity of adult retrospective reports of adverse childhood experiences: Review of the evidence. *Journal of Child Psychology and Psychiatry and Allied Disciplines*. https://doi.org/10.1111/j.1469-7610.2004.00218.x

Hernán, M. A., Hernández-Díaz, S. and Robins, J. M. (2004): A structural approach to selection bias. *Epidemiology* **15**, 615–625.

Kopycka-Kedzierawski, D. T. and Billings, R. J. (2004): A longitudinal study of caries onset in initially caries-free children and baseline salivary mutans streptococci levels: A Kaplan-Meier survival analysis. *Community Dentistry and Oral Epidemiology* **32**, 201–209.

Leake, J. L. and Werneck, R. I. (2005): The use of administrative databases to assess oral health care. *Journal of Public Health Dentistry*. **65**, 21-35.

Ljung, R., Lundgren, F., Appelquist, M. and Cederlund, A. (2019): The Swedish dental health register - Validation study of remaining and intact teeth. *BMC Oral Health* **19**, 116.

Locker, D. (2000): Deprivation and oral health: a review. *Community Dentistry Oral Epidemiology* **28**, 161–169.

Löe, H., Theilade, E. and Jensen, S. B. (1965): Experimental Gingivitis in Man. *Journal of Periodontology* **36**, 177–187.

Mansournia, M. A., Etminan, M., Danaei, G., Kaufman, J. S. and Collins, G. (2017): Handling time varying confounding in observational research. *BMJ Online.* **359**, j4587.

Merchant, A. T. and Pitiphat, W. (2002): Directed acyclic graphs (DAGs): an aid to assess confounding in dental research. *Community Dentistry and Oral Epidemiology* **30**, 399–404.

Nadler, D. L. and Zurbenko, I. G. (2014): Estimating Cancer Latency Times Using a Weibull Model. *Advances in Epidemiology* **2014**, 1–8.

Olsen, J. (2008): Using Secondary Data in Kenneth J. Rothman, S. Greenland, and T. L. Lash, eds., *Modern Epidemiology* (3rd ed.), 481–91. Lippincott William & Wilkins, Philadelphia.

Pahel, B. T., Preisser, J. S., Stearns, S. C. and Rozier, R. G. (2011): Multiple imputation of dental caries data using a zero-inflated Poisson regression model. *Journal of Public Health Dentistry* **71**, 71–78.

Preisser, J. S., Long, D. L. and Stamm, J. W. (2017): Matching the Statistical Model to the Research Question for Dental Caries Indices with Many Zero Counts. *Caries Research* **51**, 198–208.

Reddy, M. S., Geurs, N. C., Jeffcoat, R. L., Proskin, H. and Jeffcoat, M. K. (2000): Periodontal disease progression. *Journal of Periodontology.* **71**, 1583–1590.

Rothman, K J (1981): Induction and latent periods. *Am J Epidemiol* **114**, 253–259.

Rubin, D. B. (1976): Inference and missing data. *Biometrika* **63**, 581–592.

Schafer, J. L. and Graham, J. W. (2002): Missing data: Our view of the state of the art. *Psychological Methods* **7**, 147–177.

Slade, G. D. and Caplan, D. J. (1999): Methodological issues in longitudinal epidemiologic studies of dental caries. *Community Dentistry and Oral Epidemiology.* **27**, 236–248.

Szklo, M. and Nieto, F. J. (2007): *Epidemiology: Beyond the Basics* (2nd ed.), Aspen, Sudbury, 489.

Van Den Broeck, J., Cunningham, S. A., Eeckels, R. and Herbst, K. (2005): Data cleaning: Detecting, diagnosing, and editing data abnormalities. *PLoS Med* **2**, e267.