MedCrave
Step into the World of Research

Biometrics & Biostatistics International Journal

Review Article

Open Access

CrossMark

Design and analysis of cohort studies: issues and practices

Abstract

Statisticians/Epidemiologists working in collaborative settings face the dilemma of choosing the right design and using appropriate analytical approaches with complex-cohort studies. Breslow and Day (1980, Volume II) in their seminal book address basic issues in detail but with simple applications. However, our challenges have further increased due to complex nature of cohort studies that will be discussed here. In this report we highlight some of the issues that a practitioner will encounter in designing and analyzing complex cohort studies and assess their impact on the findings.

Volume 2 Issue 6 - 2015

Deo Kumar Srivastava,[1] Melissa M Hudson,[2,3] Leslie L Robison,[3] Xiaoyong Wu,[4] Shesh N Rai[4]

[1]Department of Biostatistics, St. Jude Children's Research Hospital, USA

[2,3]Department of Oncology, Department of Epidemiology and Cancer Control, St. Jude Children's Research Hospital, USA

[3]Department of Epidemiology and Cancer Control, St. Jude Children's Research Hospital, USA

[4]Department of Bioinformatics and Biostatistics, University of Louisville, USA

Correspondence: Deo Kumar Srivastava, Department of Biostatistics, St. Jude Children's Research Hospital, Memphis, Texas, USA, Email shesh.rai@louisville.edu

Received: June 26, 2015 | **Published:** July 06, 2015

Abbreviations: SJLIFE, st. jude lifetime cohort study; COG LTFU, children's oncology group long-term follow-up guidelines; QOL, quality of life; AR, at-risk; ALL, acute lymphoblastic leukemia; NR, no-risk; FS, fractional shortening; AF, after load; AAF, abnormal AF; AFS, abnormal FS; CRT, cranial radiation; chest RT, chest radiation; MCAR, missing completely at random; MAR, missing at random; NMAR, not missing at random; MANOVA, multivariate analysis of variance

Introduction

With advances in treatment and supportive care the 5-year survival rates of childhood cancer patients have improved from below 50% in the 1970s to more than 80% today.[2] However, these improved survival rates have come at a cost in terms of late long-term morbidities and mortality associated with the disease or/and its treatment.[3] A comprehensive evaluation of the long-term adverse effects can be achieved by systematically following a cohort of childhood survivors and by the appropriate design and conduct of the analyses using cross-sectional and longitudinal data. To consider the statistical issues involved with follow-up of a large cohort and the studies conducted within that cohort, we will describe the St. Jude Lifetime Cohort Study (SJLIFE). However, it may be noted that issues raised here would apply to similar cohort studies. The objective of SJLIFE is the establishment of a lifetime cohort of survivors treated at St. Jude Children's Research Hospital to facilitate prospective periodic medical assessment of health outcomes among adult survivor of pediatric malignancies. To be eligible for SJLIFE the survivors had to be at least 18 years of age and must have survived for at least 10 years from the date of their cancer diagnosis. The details of the study design of SJLIFE have been previously published.[4] All participants receive a comprehensive risk-based clinical and laboratory assessment consistent with the Children's Oncology Group Long-term Follow-up Guidelines (COG LTFU); see Landier et al.,[5] In addition, survivors complete comprehensive health questionnaires to assess health history and status, social and demographic factors, health behaviors, and psychosocial functioning. As reported in Hudson et al.,[4] a higher prevalence of adverse health outcomes including pulmonary, cardiac, endocrine and neurocognitive abnormalities have been reported. These long-term adverse health outcomes may have significant impact on the quality of life (QOL) of the childhood cancer survivor. It is also seen from several publications that adverse health outcome get progressively worse with longer follow-up and are significantly associated with poor QOL. The key issues in designing and analyzing studies originating within SJLIFE are summarized in the following sections:

i. Selection of cohort
ii. Design
iii. Repeated measures
iv. Competing/confounding risk
v. Missing data
vi. Adjusting for multiplicity

We describe these issues in brief below.

Selection of cohort

With the comprehensive follow-up of all the adult survivors within SJLIFE it is clear that many research questions of interest arise. Sometimes the interest could focus on the entire cohort, such as objectives of estimating the prevalence of several adverse outcomes, and as reported in Hudson et al.,[4] identifying the risk factors associated with the outcomes of interest. Often, there is interest in comparing these prevalence estimates to some standard population or to some standardized scores. Population-based sources do exist for selected outcomes (e.g. cancer incidence), but when they are not available it may be necessary to recruit a control cohort to compare the prevalence

estimates. On the other hand, often, the interest could be in evaluating the outcomes in a sub-cohort, e.g. those who are exposed to certain treatment exposures. One such study is CARTOX, described in Hudson et al.,[6] which investigated the association between cancer treatment and subsequent risk of cardiotoxicity. The diagnostic groups potentially at-risk (AR) of cardiotoxicity included survivors of childhood leukemia, lymphoma, sarcoma and embryonal tumors all treated with anthracycline chemotherapy and/or radiation involving the heart. The control group defined at no-risk (NR) was comprised of survivors of acute lymphoblastic leukemia (ALL), Wilms tumor and germ cell tumor who did not receive cardiotoxic treatment. Two outcome measures, fractional shortening (FS) and after load (AF), were used to evaluate cardiotoxicity. Using the reference values for these measures abnormal AF (AAF) and abnormal FS (AFS) were also defined. It may be noted that when the interest is in evaluating specific hypothesis within a sub-cohort and the normative values are not available or the outcome measures of interest have not been collected on all members of the cohort, then it may be necessary to conduct a nested case-control study to economize on cost and time.

Design

When evaluating specific hypothesis, either for the entire cohort or within sub-cohort, it is essential that the study design avoids the potential introduction of selection bias. The cohort reported in Hudson et al.,[6] who participated in the CARTOX study (cross-sectional study) were invited to participate in a follow-up study evaluating declining trends in the cardiovascular health and its impact on functional status (longitudinal study). Similarly, one can evaluate QOL in the entire cohort or in a sub-cohort in a cross-sectional or longitudinal design. These types of designs require that the sample size justification or power estimates are based on appropriate cross-sectional or longitudinal methods (taking into account within subject correlation). Within each type of study there are potentially three types of outcome measures that may be applied:

a. Binary

b. Continuous

c. Time to event

In the context of the CARTOX study, definition of abnormal cardiac function AFS or AAF, based on a cut-off on FS< 0.28 or AF ≥ 74 g/cm2, represents binary data (abnormality: Yes/No). On the other hand if the actual values of FS or AF are utilized then we will be dealing with continuous variables. The decision to use categorical versus continuous data will often be based upon whether well-established clinically meaningful cut-offs exist to identify individuals with the abnormality. However, if the interest was in evaluating the time to development of an outcome (i.e. cardiac abnormality) then we will be utilizing time to event data. Similarly, in the context of the longitudinal study described above one could be interested in:

a. Assessing the increase in abnormal cardiac function over time (binary outcome)

b. Assessing the decline/increase in the actual values of FS/AF (continuous outcome)

c. evaluating the time to developing cardiac abnormality (time to event data)

but applying the methods appropriate for interval censored data. Similarly, one could treat QOL as a continuous outcome (Total Score), binary (abnormal QOL based on a cut-off) or time to developing abnormal QOL. It is important to note that the samples size justification is closely tied to the type of outcome. When a particular hypothesis needs to be evaluated for the entire SJLIFE cohort or for the sub-cohort it is clear that the appropriate design and the plan of analysis will be based on the underlying distribution of the outcome measure. For example if the outcome measure is continuous and if the assumption of normality is appropriate then the normal-theory based methods would be appropriate; otherwise, the use of nonparametric or distribution free methods may be more appropriate. On the other hand, if the outcome measure is binary presence/absence of a condition, then the methods based on evaluating the binary outcome such as logistic or probit regression models may be more appropriate. When the interest is in evaluating time to a specific outcome then methods utilizing survival analysis framework are required.

When designing the CARTOX study, the sample size consisted of the convenience sample of all the research participants who agreed to participate. Out of 1268 available research participants 278 (22%) agreed to participate. Hypothetically, if we had to determine the sample size for comparing the prevalence of cardiotoxicity between the two groups (i.e. AR vs. NR) with sufficient power then the easiest approach would be to use binomial distribution. However, in doing so we would be completely ignoring the length of follow-up, which may be crucial in identifying an appropriate time to intervene if an intervention was possible, and could be possibly more powerful as it would have utilized the information up to the time the event occurred.[7] Sample size estimation for time to event data that account for losses to follow-up, dropouts and noncompliance, under the assumption of exponential distribution, has been proposed by Lachin JM & Foulkes MA.[8] Further, sample size calculation under Cox's proportional hazards model assumption[7] and for comparing two survival curves using log rank test[9,10] without assuming proportional hazards or exponential distribution assumption have been proposed. The approaches based on Cox's proportional hazards model and log-rank test utilize the local alternative framework which assumes that the parameter of interest (e.g. hazard function or survival function) decreases to 0 at the rate of 1/sqrt(n). In the approaches described above it is assumed that the subjects/individuals are followed in real time and the exact time of the event when it happens can be recorded unless they are censored. However, when a cohort is followed intermittently, i.e. every 2 or 3 years, the actual times of the events may not be recorded and the only information available is that the event occurred from last visit to the current visit. Such data are classified as interval censored data. There are no readily available approaches or software, besides simulation methods, to calculate the sample size for interval censored data. Often one will have to use some approximation and make some assumptions in using Cox's proportional hazards model for sample size justification.

In addition to samples size justification, another important aspect of design is the issue of sampling. Once again in the context of CARTOX it is clear that the proportion of survivors in NR groups would be relatively small (20% range) and an effort could be made to recruit all of them. On the other hand, those in AR group constitutes a much larger cohort consisting of four disease categories (leukemia, lymphoma, sarcoma and embryonal tumors). All the survivors within these four disease groups would have received cardiotoxic therapy, but, in general, those with leukemia would have received higher doses of anthracycline, while those with lymphoma might have received higher dose of chest radiation and, thus, it is likely that the risk of developing cardiotoxicity might be different across the four disease groups. The

critical issue then is how to sample from within each disease category and what constitutes the "right" comparative group? The manner in which AR group is constituted will affect the power of comparing the two groups as it will be impacted by the prevalence estimates within each disease category. Assuming, limited heterogeneity among the disease groups that constitute NR group but assuming that the prevalence of cardiotoxicity to be quite heterogeneous across the four disease categories, the constitution of the sample for AR group will have significant impact on the power for comparing the two groups (NR vs. AR). If the research participants are recruited without regard to the disease then the prevalence estimates will be affected by the disease group that is most represented in the entire cohort. On the other hand one could recruit equal number of research participants from each disease group, which will lead to fair representation of all disease groups but may still provide a biased estimate of the prevalence for the entire cohort if the prevalence of cardiotoxicity is different across the four disease categories. However, if the interest is in making inference that would be representative of the cohort from which the sample is drawn then a proportional allocation (proportional to the size of each disease category) or optimum allocation (that takes into account size and variability) should be adopted to get an unbiased estimate of the prevalence for the entire cohort and a fairly accurate estimate of the power.

Long-term survivors report higher prevalence of chronic conditions resulting from their cancer and/or its treatment such as cardiovascular, endocrine and pulmonary, QOL in SJLIFE is evaluated using the SF-36 (http://www.sf-36.org/tools/SF36.shtml), which consists of 36 questions and evaluates physical and mental components of overall health. The physical component consists of four sub-components (Physical Function, Role Physical, Bodily Pain and General Health) and the mental component consists of four sub-components (Mental Health, Role Emotional, Social Function and Vitality). Often the analysis is conducted by taking the total of all eight sub-components or analyzing the physical and the mental component separately. In addition, pre-specified cut-offs are used to categorize survivors/subjects into normal/abnormal category for the two main component of QOL (Physical and Mental). The SF-36 questionnaire consists of 36 likert type questions where each sub-component consists of several items each possibly at different levels. For example, Physical functioning consists of 10 items each at 3 levels with the total score ranging from10-30, whereas the total score for Role physical is based on four items each at two levels (yes/no) and ranges from 4-8. Now if one evaluates one of the main components, i.e. the Physical, then one will have to sum the scores of Physical functioning, Role physical, Bodily Pain and General Health. The critical issue is how one interprets the total score which is obtained by summing the scores that are very different in range that possibly evaluate correlated but slightly different constructs. The interpretation of results has been made easier with the standardization of each of the sub-component measure. This is done by transforming each sub-component score to a mean of 50 and standard deviation of 10 in the general US population. However, one has to wonder how good this standardization is for the sub-components when the range is from 10-30 (possibly normally distributed) vs. when the range is 4-8 (highly discretized).

When evaluating the decline in QOL in survivors of childhood cancer it is imperative that it is compared to a cohort that is very similar to the surviving cohort except for the fact that they never had cancer. In such situations exposed (with cancer) and unexposed (without cancer) groups are matched on certain characteristics such as age, race and sex, and then the two groups are compared and the impact of covariates of interest on QOL measure is evaluated using appropriate statistical approach such as stratified analyses or conditional logistic regression, e.g. see Hosmer DW & Lemeshow S.[11] This study design is different from the traditional retrospective case-control (i.e. nested case-control) design where the cases are identified first and then non-cases are selected from the available pool of participants and matched on key characteristic variables and then data on other covariates of interest are collected and analyzed using conditional logistic regression. Often, in case-control designs, 1:1 or 1:M matching are commonly used. It is well known that 1:1 matching provides the most cost-effective design when exposed and unexposed groups are equally "scarce." However, when the cases are rare and more participants are available in the unexposed group then more than one control is matched to each case. It is seen that the efficiency of the design increases with increasing M, but the gain in efficiency declines rapidly beyond M=4, see Breslow NE & Day NE.[12] It is often seen that many times the investigators design the study as a matched study (matched on key characteristics) but when it is time to conduct the analysis the matching is ignored and the analysis is conducted as if there was no matching, this is inappropriate as it may lead to biased results, Breslow NE et al.,[13] An analysis that appropriately accounts for the matching should be undertaken. Thus, choosing an appropriate outcome measure to evaluate a specific hypothesis and using an appropriate analytical approach are crucial in reporting the results in an unbiased manner.

Repeated measures

In following up a cohort such as SJLIFE it is always an open question as to how often the survivors should be evaluated. Of course, the more often we get this information the more accurately the incidence of adverse health outcomes can be described. However, this has to be balanced with the onerous amount of logistical issues involved in bringing all the survivors for a complete check-up, costs of performing the test procedures, and the burden on the survivors to come to St. Jude Children's Research Hospital and staying for the required duration while all the test are being conducted. If resources permit, there is clear merit in collecting all the medical information and the results of through screening from all the survivors every 2-3 years, irrespective of the treatment exposure. This provides for an internal control group when comparing the prevalence or incidence of certain adverse outcomes in the group of survivors who are exposed vs. those who are not exposed to certain exposures, e.g. certain treatment exposures such as cranial radiation (CRT), chest radiation (ChestRT) or treatment with anthracyclines etc. Another important issue to note and incorporate into the analysis of time to event data (e.g. time to abnormal QOL) is the highly variable length of follow-up, 10 years to 50 years, and its impact on the inference. That is, how does one adjust for the natural decline in QOL as a result of the aging process vs. due to disease and its treatment? Often, the effects of the aging process can be easily accounted for if age specific norms are available. The manner in which the data are collected provide a snap shot of the cohort at pre-specified time points every 2-3 years. Thus, the exact time of the adverse outcomes may not be known exactly but it will be known that it happened anytime from the last visit to the current visit, i.e. what we have is the interval censored data. However, this may be further complicated by the fact that for some adverse events the exact time of onset for some individuals may be known, e.g. by visiting a physician or self-reporting, thus we end up with a data that are a mixture that has time of onset for some survivors as interval censored and for some with known exact time of onset. The

Citation: Srivastava DK, Hudson MM, Robison LL, et al. Design and analysis of cohort studies: issues and practices. *Biom Biostat Int J.* 2015;2(6):161–166. DOI: 10.15406/bbij.2015.02.00044

standard statistical approaches may not be applicable in such setting and alternative approaches need to be explored or developed.

Competing/confounding risk

Competing risk arises in the context of survival analysis framework where a research participant could be at risk of developing several adverse outcomes (competing adverse outcomes) that are mutually exclusive and occurrence of any one of them precludes one from observing the times to developing other adverse outcomes. In the context of the CARTOX study, death without cardiac abnormality can be considered a competing risk. When evaluating a complex phenotype such as QOL it is easy to comprehend that QOL may be impacted by a particular diagnosis and/or by its treatment. Also, it can be further hypothesized that the mechanisms through which the QOL may be affected is through chronic health conditions. That is, a survivor would be in a healthy state (no chronic condition and normal QOL) when newly diagnosed. However, as a result of the disease or its treatment a survivor may develop one of the many chronic conditions such as cardiac, pulmonary, endocrine or musculoskeletal and connective tissue disorder which, in turn, may be responsible for poor QOL. On the other hand a survivor may have poor QOL without having and chronic health condition. This can be formulated in framework of illness-death model, see Rai SN et al.,[14] and the analysis conducted within this framework can take into account the competing risk and provide better estimate of the incidence rates. In the above context one may focus on one chronic condition but, quite often, the interest could be on evaluating the effect of all chronic conditions on QOL simultaneously. Also, the chronic conditions may be correlated and, thus, in such situations multivariate modeling will have to be undertaken and new approaches and methods need to be developed.

Confounding commonly occurs when a particular factor is related to the exposure and outcome but does not lie on the causal pathway between the exposure and outcome. When the confounding factor is not appropriately accounted for in the analysis then the association between the exposure and the outcome could be over or under estimated. The potential confounding variables need to be identified at the design stage so that valid information on them can be collected and appropriately incorporated into the analysis. In the context of the CARTOX trial it is likely that insulin resistance may be related to the exposure and the outcome (FS or AF). However, if we do not collect the information on insulin resistance or if we don't take into account appropriately when analyzing the data we could get biased estimates of the association between exposure (AR vs. NR) and the outcome (FS or AF). There are essentially two approaches to account for confounding variable: a)Stratify the analysis by confounding variables and use methods that appropriately pool information across the strata to provide an estimate or test procedure for the entire cohort and b)Utilize regression approach that adjusts for all the confounders simultaneously in the model as covariates. However, the limitation of approach in (a) is having too many confounding factors that may lead to many strata resulting in unreliable estimates within each stratum. A disadvantage of the approach in (b) is that one may lose sight of the specific details of the data and may not provide the best possible understanding of the relationships between the confounding factors and the outcome measure.

Missing data

It is unfortunate but is typical when evaluating a cohort in a cross-sectional manner or following it in a longitudinal manner that there will be missing data due to subjects dying, withdrawing from the study or becoming lost to follow-up. In such situations it is imperative to conduct a thorough evaluation of how the missing values affect the results. The treatment of missing data will depend on the type of study (cross-sectional or longitudinal) and whether the missingness is in the outcome measure, explanatory variables or correlated variables (particularly in longitudinal evaluations). However, before conducting any analysis with missing data one must understand the missing-data mechanism. The missing data mechanism describes the probability of a response being observed or missing and Rubin DB;[15] see also Little RJA & Rubin DB[16] and Fitzmaurice G et al.,[17] who characterized three types of missing mechanisms

i. Missing completely at random (MCAR)

ii. Missing at random (MAR)

iii. Not missing at random (NMAR).

Under MCAR, it is assumed that the missing data process does not depend on the observed outcomes nor the observed covariates. MAR implies that the missing data process only depends on the observed covariates and the outcome measure. However, NMAR implies that the missing data mechanism depends on the values of the missing covariates or the outcome measure.

Missing data in cross-sectional studies

In general one talks about missing data in the context of the outcome variable (Y), but it is very likely that the missingness could be in the explanatory variables. For examples, in the CARTOX study, which is a cross-sectional study, the data could be missing for the outcome measure (missing AF or FS values), covariates (Years since diagnosis or BMI), or both outcome measure and covariates. When the data are MCAR the "complete case" analysis (subjects with no missing data) will provide valid inferences as with no missing data.

When one encounters missing data two questions immediately come to mind: (a) is the missingness significant and (b) are the characteristics of subjects who provide complete information (completers) different from those who don't (non-completers)? In practice when the missingness is low, less than 10% ("ad hoc") of the total sample size, and the characteristics of the completers and non-completers don't differ significantly then one tends to ignore it assuming it will not have significant impact on the results. However, when the missingness is high (more than 10% of the total sample size) and when the characteristics between the completers and non-completers differ then rigorous approaches, such as weighting or imputing discussed below, need to be adopted.

When the data are MAR then some "ad hoc" methods for imputing the data, such as mean value of the available observations, may provide biased estimates. However, there are essentially two scientifically rigorous approaches that can be adopted to handle missing data;

i. Weighting methods to appropriately weight the observed data in some manner

ii. Use imputation approaches to fill the missing values and then conduct the analysis on the complete data set.

It may be noted that the imputation approaches can be broadly classified further into two groups one that is based on likelihood or quasi-likelihood (GEE) approach and the other that is essentially based on regression methods; Zhao LP et al.,[18] Molenberghs G & Fitzmaurice G.[19] The methods based on likelihood or generalized estimating equations (GEE) approaches provide valid inferences.

PROC MI and PROC MIANALYZE in SAS can be used to do multiple imputations using regression methods and conduct the analysis. However, when the data are NMAR the missingness is considered to be informative and all standard methods of analysis are invalid. However, when the missingness is NMAR, one is encouraged to conduct sensitivity analysis and assess the impact of various assumptions on the missingness process on the findings.

Missing data in longitudinal studies

In longitudinal studies the measurements for an outcome measure are obtained by following a research participant over time and recording the values of the outcome measure at regular intervals. In longitudinal studies, in addition to all the issues highlighted above for the cross-sectional studies, there is an added problem of accounting for correlation among repeated observations. In longitudinal studies the missing pattern can be "monotonic" or "non-monotonic." If in a cohort subjects are followed and evaluated at K time points then monotonic missing means that if a subject has a missing value at j-th time point (j=1,2,…,K) then this subject will have missing values for all subsequent time points beyond (j+1), i.e. the subject drops out of the study and never comes back. On the other hand non-monotonic missing pattern means that there is no pattern and the observation can be missing at any time point, i.e. a subject can go in and out of the study at any time point. Once again the approaches described for the cross-sectional studies can be adopted for longitudinal studies as well with the caveat of appropriately accounting for the correlation among the repeated observations within an individual and the type of missing pattern. A comprehensive account of the analysis of longitudinal data in presence of missing data can be found in Little RJA & Rubin DB,[16] and Fitzmaurice G et al.,[17] Molenberghs et al.,[19] Although, these references outline general guidelines for analyzing longitudinal studies, a careful consideration is required for its application to specific examples. For example, when comparing group effects (treatment vs. control), multiple imputation in the context of mixed effect model may lead to overestimated variances and may provide heavily biased estimates. However, under ignorable or non-ignorable missing assumption, a mixed model approach based on generalized estimating equations can yield unbiased estimates; see Fong DYT, et al.,[20]

Adjustment for multiplicity

Just as QOL is assessed using the SF-36, which has two components (Physical Health and Mental Health) and each component consists of four sub-scales there are many questionnaires administered during SJLIFE evaluation where a subset of questions form a construct and then several correlated constructs are evaluated and compared either to the normative data or to the control group. However, often, when the interest is in evaluating several constructs emerging from a questionnaire then the correlation among the constructs is ignored and the analysis is conducted independently for each construct. This approach will often lead to false discoveries as the type I error rates would be inflated. One approach to controlling the overall type I error rate is to use Bonferroni correction and conduct the test for each hypothesis under consideration at level (α / M) where M is the total number of hypothesis to be tested. However, this approach is subject to criticism as being too conservative and alternative approaches that model the constructs jointly should be adopted. This can be achieved by using multivariate methods such as multivariate analysis of variance (MANOVA) or linear mixed effect models or GEE approach that appropriately model the outcomes jointly and take care of the correlation between the outcomes within an individual. Since the number of hypotheses are not large compared to high-throughput data analysis, methods appropriate for controlling false discovery rate, Benjamini Y & Hochberg Y [21], may not be required. A simple approach such as Holm's adjusted p-values can be used; see Holm SA[22] and Westfall PH et al.,[23]

Early methods for design and analyses of cohort studies are provided in Breslow NE & Day NE.[1] Many of those issues have become more complex, which have been raised in this note from an application perspective. Using SJLIFE as the model, we have highlighted some of the issues that are encountered when evaluating specific hypothesis or designing studies to evaluate specific hypothesis related to sub-cohorts based on exposure or some other characteristics within a cohort that is being followed longitudinally.[24,25] The purpose of our discussion is to raise the awareness of the issues involved with cohort studies while realizing that each of the issues raised here merit a discussion of their own. We intent to follow-up some of these issues, systematically evaluate their impact on the results and report our findings in subsequent manuscripts. However, it is clear that, prior to conducting studies within a cohort that is followed longitudinally, each of the issues raised here need to be carefully evaluated in the context of the studies to be undertaken.

Acknowledgement

Dr. Rai is grateful to generous support from Dr. DM Miller, Director James Graham Brown Cancer Center and Wendell Cherry Chair in Clinical Trial Research. The research work of Deo Kumar Srivastava, Leslie Robison and Melissa Hudson was in part supported by the Cancer Center Support (CORE) grant CA 21765 and by the American Lebanese Syrian Associated Charities (ALSAC).

Conflict of interest

The authors declare that they have no financial or non-financial competing interests.

References

1. Breslow NE, Day NE. *Statistical Methods in Cancer Research, Volume II – The Analysis of Cohort Studies, International Agency for Research on Cancer*. IARC Scientific Publications, Lyon, France. 1980;281 p.
2. Howlader N, Noone AM, Krapcho M, et al. *SEER Cancer Statistics Review, 1975–2012*. National Cancer Institute.
3. Robison LL, Hudson MM. Survivors of childhood and adolescent cancer: life–long risks and responsibilities. *Nat Rev Cancer*. 2014;14(1):60–69.
4. Hudson MM, Ness KK, Gurney JG, et al. Clinical ascertainment of health outcomes among adults treated for childhood cancer. *JAMA*. 2013;309(22):2371–2381.
5. Landier W, Bhatia S, Eshelman DA, et al. Development of risk–based guidelines for pediatric cancer survivors: the Children's Oncology Group Long–Term Follow–Up Guidelines from the Children's Oncology Group Late Effects Committee and Nursing Discipline. *J Clin Oncol*. 2004;22(24):4979–4990.
6. Hudson MM, Rai SN, Nunez C, et al. Noninvasive evaluation of late anthracycline cardiac toxicity in childhood cancer survivors. *J Clin Oncol*. 2007;25(24):3635–3643.
7. Chow SC, Shao J, Wang H. Sample Size Calculations in Clinical Research. *Chapman & Hall/CRC Biostatistics Series, 2nd ed*. Taylor and Francis Group, Boba Raton, USA. 2008.

8. Lachin JM, Foulkes MA. Evaluation of sample size and power for analysis of survival with allowance for nonuniform patient entry, losses to follow–up, noncompliance, and stratification. *Biometrics*. 1986;42(3):507–519.
9. Lakatos E. Sample size determination in clinical trials with time dependent rates of losses and noncompliance. *Control Clin Trials*. 1986;7(3):189–199.
10. Lakatos E. Sample sizes based on the log–rank statistic in complex clinical trials. *Biometrics*. 1988;44(1):229–241.
11. Hosmer DW, Lemeshow S. *Applied Logistic Regression, 2nd ed*. John Wiley & Sons Inc., New York, USA. 2000;375 p.
12. Breslow NE, Day NE. *Statistical Methods in Cancer Research, Volume I – The Analysis of Case–Control Studies*. International Agency for Research on Cancer, Lyon, France. 1980;281 p.
13. Breslow NE, Day NE, Halvorsen KT, et al. Estimation of multiple relative risk functions in matched case–control studies. *Am J Epidemiol.* 1978;108(4):299–307.
14. Rai SN, Pan J, Yuan X, et al. Estimating incidence rate on current status data with application to a phase IV cancer trial. *Communications in Statistics – Theory and Methods*. 2013;42(17):3117–3135.
15. Rubin DB. Inference and missing data. *Biometrika*. 1976;63(3):581–592.
16. Little RJA, Rubin DB. *Statistical Analysis with Missing Data, 2nd ed*. John Wiley & Sons, Inc. Publication, Hoboken, New Jersey, USA. 2002;371 p.
17. Fitzmaurice G, Davidian M, Verbeke G, et al. *Handbook of Modern Statistical Methods – Longitudinal Data Analysis*. Chapman & Hall/CRC, New York, USA. 2009.
18. Zhao LP, Lipsitz S, Lew D. Regression analysis with missing covariate data using estimating equations. *Biometrics*. 1996;52(4):1165–1182.
19. Molenberghs G, Fitzmaurice G. *Incomplete data: Introduction and Overview. In Handbook of Modern Statistical Methods, Longitudinal Data Analysis. (Chapter 17)*. In: Fitzmaurice G, editors. Chapman & Hall/CRC, New York, USA. 2009;395–408 p.
20. Fong DY, Rai SN, Lam KS. Estimating the effect of multiple imputation on incomplete longitudinal data with applications to a randomized clinical study. *J Biopharm Stat*. 2013;23(5):1004–1022.
21. Benjamini Y, and Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*. 1995;57:289–300.
22. Holm S. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*. 1979;6(2):65–70.
23. Westfall PH, Tobias RD, Wolfinger RD. *Multiple Comparisons and Multiple Tests Using SAS, 2nd ed*. SAS institute Inc, Cary, North Carolina, USA. 2011.
24. Rai SN. *Analysis of Occult Tumor Studies*. In: Tan WY, Hanin L, editors. *Handbook of Cancer Models with Applications*. World Scientific Press, Singapore. 2008.
25. Diggle PJ, Heagerty P, Liang KY, et al. *Analysis of Longitudinal Data*. Oxford Statistical Science Series, New York, USA. 2005.

Citation: Srivastava DK, Hudson MM, Robison LL, et al. Design and analysis of cohort studies: issues and practices. *Biom Biostat Int J.* 2015;2(6):161–166. DOI: 10.15406/bbij.2015.02.00044