



EDITORIAL

Survivor bias and risk assessment

Dave P. Miller*, Mardi Gomberg-Maitland[#] and Marc Humbert[†]

In the current issue of the *European Respiratory Journal*, LEE *et al.* [1] have done excellent work developing and validating a risk equation for incident pulmonary arterial hypertension (PAH) patients in the UK. In addition, they have very helpfully summarised the differences and similarities between the newly developed Scottish composite score (SCS) and other contemporary PAH survival equations. One of the most important methodological choices was the decision to include only incident cases in the analysis, and this is an issue which deserves further discussion because of the implications for the design of registries and other observational studies in PAH.

Prevalent cohorts are prone to inherent bias. Prevalent cohorts in PAH, and in fact for any disease, are subject to a survivor bias because patients who die soon after disease onset are less likely to be included in a prevalent cohort. For instance, consider two different subgroups, one with a mean survival of 4 yrs and another with a mean survival time of 8 yrs. If the incidence rate is constant over time, patients from the latter cohort, with the better prognosis, will be twice as likely as the former group to appear in a prevalent cohort. The latter group will therefore be overrepresented if one is interested in the distribution of an incident cohort. Survivor bias of this nature impacts the development of survival estimates in two distinct ways. First, an analysis cohort consisting of prevalent patients will have a disproportionately higher number of lower risk patients. In PAH prevalent cohorts, we see more females, fewer functional class (FC) IV patients and more patients with longer 6-minute walk distances (6MWDs). Because there are more of these patients, they receive more weight in the development of equations. Second, the future prognosis of this lower risk cohort will be better, so the aggregate estimates of survival for the overall cohort will be more favourable than if the survival estimate is derived from incident patients.

HUMBERT *et al.* [2] and McGOON and MILLER [3] show that, in aggregate, prevalent patients have a better prognosis than incident or newly diagnosed patients. However, BENZA *et al.* [4] show that the Registry to Evaluate Early and Long-term PAH Disease Management (REVEAL) risk calculator, developed in a predominantly prevalent cohort, is nonetheless effective at predicting risk in newly diagnosed patients. The newly diagnosed cohort in REVEAL has fewer patients in the lowest risk strata and more patients in the highest risk strata, so

survivor bias clearly affects the case mix between the two cohorts, but for risk assessment it appears to be fully accounted for through the risk adjustment. In other words, survivor bias may be an intractable problem when prevalent patients are used to estimate an aggregate curve, but it may not be an important issue in risk assessment. Similarly, in the development of the Pulmonary Hypertension Connection (PHC) equation [5], the baseline right heart catheterisation (RHC) was not always concurrent with diagnosis, but the formula was able to be successfully validated using the RHC date from other studies as time zero [6].

The REVEAL prognostic equation, which is the basis for the risk calculator, was developed using the most recent assessment at any time point prior to enrolment, to assess risk [7]. This reflects a perceived clinical need for ongoing risk assessment throughout the patients' disease course. There is, of course, a very specific need for an accurate assessment of prognosis at the time of diagnosis, as this is the time point when the most potential treatment options may be under consideration.

Like the SCS, the French registry equation was developed [8] using date of diagnosis as the reference point; however, unlike the SCS, prevalent patients diagnosed within 3 yrs of enrolment were included in the analysis. The inclusion of prevalent patients, while using diagnosis as the reference point, creates an issue with left truncated data. Therefore a delayed entry model was used so that patients who were enrolled for example a year after diagnosis were not used in the hazard estimates for the first year of the curve. In other words, the survivors were only utilised to estimate the latter part of the curves, which are estimated exclusively using survivors anyway even if one starts with an incident cohort. McGOON and MILLER [3] demonstrate how the aggregate curve using all patients in a delayed entry model nicely mirrors the aggregate curve explicitly excluding previously diagnosed patients.

It is possible to account for bias with prevalent cohorts utilising statistics that allow for estimation of the risk over time. Both Kaplan–Meier estimates and the Kalbfleisch–Prentice estimator, which is the kernel of the Cox Proportional Hazards model, are products of the instantaneous hazards over time [9]. This means that a purely incident cohort is used to estimate the hazard only in the first moment after the initial diagnosis. But, it is necessary to account for survival after initial diagnosis and to have the hazard estimate change as time passes. In statistical terminology, the hazard estimate for any time point t , greater than zero, is conditional on having survived to at least time t . Thus, a patient who is enrolled in a study at time x and followed through time $x+c$ (c =additional on-study follow-up time) can be used to estimate the hazard at time t as long as $x<t<x+c$. The

*ICON Late Phase and Outcomes Research, San Francisco, CA, and [#]University of Chicago Medical Center, Chicago, IL, USA. [†]Université Paris-Sud, INSERM U999, Hôpital Bicêtre, APHP, Le Kremlin Bicêtre, France.

CORRESPONDENCE: D.P. Miller, ICON Late Phase and Outcomes Research, 188 The Embarcadero, Suite 200, San Francisco, CA 94105, USA. E-mail: dave.miller@iconplc.com

inclusion of prevalent patients in this manner avoids survivor bias, and, equally importantly, avoids immortal time bias. Immortal time bias is different from survival bias in that it occurs when patients are analysed as though they are at risk of an event, when in fact they are not. Examples of immortal time are the time between diagnosis and enrolment or, in flawed comparative effectiveness studies, the time between enrolment and treatment initiation [10].

It should be noted that neither a delayed entry model nor a restrictive inclusion criteria address the survivor bias associated with patients not surviving to diagnosis. GRIMES and SCHULZ [11] note that Neyman bias, a form of incident-prevalent bias, is particularly common in “diseases that are quickly fatal... or subclinical”. Because PAH is a rare disease with a frequently long delay between nonspecific symptom onset and diagnosis [12], it is unknown how many patients die without ever having a RHC and thus without any attribution of the death to undiagnosed PAH. Furthermore, due to differences in healthcare systems, including access to specialists and the different diagnostic requirements associated with receiving advanced treatments, the pre-diagnosis survivor bias may differ between countries. Therefore, incident cohorts from one country may not be easily generalisable to another country. Generalisability is an important benefit of observational study analyses, largely because of the good external validity that observational studies provide. External validity is defined simply as “does it hold up to be true in other cohorts”. This is complementary to the weaker generalisability of randomised clinical trials, which are designed with a greater focus on internal validity [13]. Randomised trials use strict inclusion criteria to create a well-defined patient cohort and use highly structured follow-up schedules to optimise evaluation of treatment effects. There are significant costs, as well as potential ethical issues, associated with randomising patients in whom a treatment effect is less likely to be observed. The balance of internal and external validity can apply to observational studies as well. Studies with a narrow focus and a single objective may benefit from tighter inclusion criteria, but the inclusion of prevalent patients in a cohort expands the number of questions that may be answered.

More work needs to be done to understand if the risk factors for prevalent patients differ from the risk factors for incident patients. It is plausible that some modifiable risk factors are less important for incident patients, assuming that they receive immediate and effective treatment, while those same risk factors may portend worse outcomes in prevalent patients who are already on therapy. This is a testable hypothesis if a cohort includes both incident and prevalent patients.

In studies which enrol incident and prevalent patients, interactions between risk profile and prognosis could be evaluated by the formal inclusion of an interaction term in the models. On first glance this is a simple solution, one that is frequently used in general modelling techniques. However, in studies which include only incident cases, a longer and more comprehensive post-enrolment follow-up would be necessary to test the interaction. Interactions are difficult to test because they generally require very large sample sizes which in a rare disease is not easy to obtain. Also, it can be difficult to select a manageable number of interactions to assess, with too large a pool assessed leading to spurious results. Thus, hypothesising

and testing interactions may be one of the areas where there is the most to be gained by the collective wisdom associated with multiple large observational studies in many countries, ideally through the use of meta-analysis. While a single study may be underpowered to detect an interaction, a consistent trend across multiple studies would be a more reliable signal. Additionally, one study might prove to be hypothesis generating where the other studies can be used to replicate and validate the results to guard against spurious findings.

In fact, much has been learned already through the collective experience of the different registries. Prognosis today is clearly better than it was at the time of the National Institutes of Health's first major PAH registry [5, 8, 14]. The paradox that PAH is a predominantly female diagnosis, but that men with the disease do worse, is now firmly established. Worse prognosis for systemic sclerosis associated PAH is evident, while mixed signals still exist for portopulmonary hypertension and familial PAH. The 6MWD and haemodynamics are the most consistent predictors across studies, but the particular haemodynamic parameters of greatest importance remains debatable. Other variables, such as brain natriuretic peptide and diffusing capacity of the lung for carbon monoxide, were identified as univariate predictors in the development of the SCS, but did not reach statistical significance for inclusion in the final score. These variables may be candidates to consider for a possible interaction. That is, they could be stronger markers for a better or worse prognosis later in a patients' disease course.

It is interesting that FC was retained in the multivariable model only in the REVEAL risk calculator, although it was a strong univariate predictor in the development of the SCS and the French registry equation. LEE *et al.* [1] are correct to note the variability of FC assessments [15] as a potential culprit explaining the absence of FC in the final equation. Given the large number of sites participating in REVEAL, it may be that variability exists for other tests as well and that the consistent strength of the 6MWD as a predictor is related to the ability to obtain consistent results across sites without a core laboratory.

Finally, the PAH research community is rapidly improving its understanding of immortal time bias and survival bias. Accurately reporting the at-risk period during which patients were on study is one of the most critical steps for avoiding immortal time bias. Accurately reporting time zero and clarifying the target population for generalisability is one of the most important steps in avoiding survivor bias. If a target cohort for generalisability is all PAH patients under care at a practice, the target cohort is comprised of survivors and the term survivor bias is not meaningful. If a target cohort is incident patients, prevalent patients should only be utilised to estimate the later years of the survival curve. The PH community will need to work together to further our understanding of the disease. Indeed, this will be one of the questions explicitly addressed by the task force on epidemiology and registries at the upcoming fifth World Symposium (www.wsph2013.com) which will take place next year in Nice, France (February 26 to March 1, 2013) [16].

STATEMENT OF INTEREST

Statements of interest for all authors of this manuscript can be found at www.erj.ersjournals.com/site/misc/statements.xhtml

REFERENCES

- 1 Lee W-TN, Ling Y, Sheares K, *et al.* Predicting survival in pulmonary arterial hypertension in the UK. *Eur Respir J* 2012; 40: 604–611.
- 2 Humbert M, Sitbon O, Yaïci A, *et al.* Survival in incident and prevalent cohorts of patients with pulmonary arterial hypertension. *Eur Respir J* 2010; 36: 549–555.
- 3 McGoon MD, Miller DP. REVEAL: a contemporary US pulmonary arterial hypertension registry. *Eur Respir Rev* 2012; 21: 8–18.
- 4 Benza RL, Gomberg-Maitland M, Miller DP, *et al.* The REVEAL registry risk score calculator in patients newly diagnosed with pulmonary arterial hypertension. *Chest* 2011; 141: 354–362.
- 5 Thenappan T, Shah SJ, Rich S, *et al.* Survival in pulmonary arterial hypertension: a reappraisal of the NIH risk stratification equation. *Eur Respir J* 2010; 35: 1079–1087.
- 6 Thenappan T, Glassner C, Gomberg-Maitland M. Validation of the pulmonary hypertension connection equation for survival prediction in pulmonary arterial hypertension. *Chest* 2012; 141: 642–650.
- 7 Benza RL, Miller DP, Gomberg-Maitland M, *et al.* Predicting survival in pulmonary arterial hypertension: insights from the Registry to Evaluate Early and Long-Term Pulmonary Arterial Hypertension Disease Management (REVEAL). *Circulation* 2010; 122: 164–172.
- 8 Humbert M, Sitbon O, Chaouat A, *et al.* Survival in patients with idiopathic, familial, and anorexigen-associated pulmonary arterial hypertension in the modern management era. *Circulation* 2010; 122: 156–163.
- 9 Kalbfleisch JD, Prentice RL. *The Statistical Analysis of Failure Time Data*. New York, Wiley and Sons, Inc, 1980.
- 10 Suissa S. Immortal time bias in pharmaco-epidemiology. *Am J Epidemiol* 2008; 167: 492–499.
- 11 Grimes DA, Schulz KF. Bias and causal associations in observational research. *The Lancet* 2002; 359: 248–252.
- 12 Brown LM, Chen H, Halpern S. Delay in recognition of pulmonary arterial hypertension: factors identified from the REVEAL registry. *Chest* 2012; 140: 19–26.
- 13 Gliklich RE, Dreyer NA, eds. *Registries for Evaluating Patient Outcomes: A User's Guide*. 2nd ed. AHRQ Publication No.10-EHC049. Rockville, Agency for Healthcare Research and Quality, 2010.
- 14 Benza RL, Miller DP, Barst RA. An Evaluation of long-term survival from time of diagnosis in pulmonary arterial hypertension from REVEAL. *Chest* 2012; [Epub ahead of print DOI: 10.1378/chest.11-1460].
- 15 Taichman DB, McGoon MD, Harhay MO, *et al.* Wide variation in clinicians' assessment of New York Heart Association/World Health Organization functional class in patients with pulmonary arterial hypertension. *Mayo Clin Proc* 2009; 84: 586–592.
- 16 Humbert M. The fifth world symposium on pulmonary hypertension will REVEAL the impact of registries. *Eur Respir Rev* 2010; 21: 4–5.