CLINICAL TRIALS *ARTICLE* *Clinical Trials* 2008; **5**: 107–115

Profile-specific survival estimates: Making reports of clinical trials more patient-relevant

Marilyse Julien[a] *and James A Hanley*[b]

Background When considering treatment options, a physician needs to know the prognosis corresponding to the risk profile of the patient seeking treatment. Reports of clinical trials generally address treatment-specific survival probabilities only in the aggregate, i.e., for the typical patient, and often express the difference in survival as a hazard ratio. Such summaries do not provide treatment-specific survival probabilities (and thus the absolute difference in these probabilities) for patient profiles that are not near the typical of those in the trial. Despite the fact that Cox intended his hazard regression method to be used to produce such profile-specific survival estimates, and even showed how to calculate them, authors are either unaware that this is possible, or else choose not to report them.
Purpose To illustrate how treatment- and profile-specific survival estimates are obtained from the Cox method, and can be displayed in a compact form.
Methods We derive treatment- and profile-specific survival probabilities from the estimated survival function for the 'reference' profile. Data from the Systolic Hypertension in the Elderly Program study serve as an illustration.
Results Two different formats, tabular and nomogram-based, allow the entire set of estimated treatment- and profile-specific survival probabilities to be reported.
Limitations Estimates are limited to the profiles within the covariate-space spanned by the trial, and depend on the correctness of the model.
Conclusion Treatment- and profile-specific survival estimates are practice-relevant, almost never reported, estimable from the Cox model, and easy to report in a compact form. *Clinical Trials* 2008; **5**: 107–115. http://ctj.sagepub.com

Introduction

Physicians rely on reports of both randomized clinical trials and observational studies for evidence on the benefits of various treatments. How informative are the reports of these studies for the physician faced with an individual patient with a specific prognostic profile? We examine two examples.

Case 1. A physician consults the literature to gauge the probability of a survival benefit if a 58-year old man, PSA level 9.1, diagnosed with a 'Gleason 7' prostate cancer, opts for radical rather than conservative treatment. In the only randomized trial of these two options [1], prostate cancer mortality was lower with radical treatment (hazard ratio 0.56). The 10-year 'cumulative incidence' of prostate cancer death was 10% versus 15%. The only profile-specific information was that 'the benefit of radical therapy … differed according to age but not according to the PSA level or the Gleason score.' Two reports of nonrandomized studies contain 'profile-specific' prognoses but are either limited to conservative treatment [2] or have so few patients who took this option that confidence intervals are wide [3]. A third [4] was based on 45,000 men aged 65–80. 'Using propensity scores to adjust for potential confounders,' the authors reported 'a statistically significant survival

[a]Department of Mathematics and Statistics, McGill University Montreal, Quebec, Canada, [b]Department of Epidemiology, Biostatistics and Occupational Health, McGill University Montreal, Quebec, Canada
Author for correspondence: JA Hanley, Department of Epidemiology, Biostatistics and Occupational Health, McGill University, 1020 Pine Avenue West, Montreal, Quebec, Canada. E-mail: james.hanley@mcgill.ca

SAGE Publications, Los Angeles, London, New Delhi and Singapore 10.1177/1740774508089511

advantage' in those who chose radical treatment (hazard ratio, 0.69)''. An absolute 10-year survival difference (in percentage points) was provided for each 'quintile of the propensity score'', but the physician was unable to translate this information into a survival difference for men with his patient's profile.

Case 2. A physician consults the report of a classic randomized trial [5] to assess the 5-year risk of stroke for a 65-year old white woman with a SBP of 160 mmHg, and how much it is lowered if she were to take anti-hypertensive drug treatment. The reported risk difference was $8.2\% - 5.2\% = 3\%$, and the 'favorable effect' of treatment was also found for all age, sex, race, and baseline SBP groups. Again, however, the report did not provide information from which to estimate the risk, and risk difference, for this specific profile.

Are these isolated cases?

Are survival statistics from clinical trials – and nonrandomized studies – limited to the 'average' patient? [6] Is Cox regression used merely to ensure 'fairer comparisons'? How often is it used to provide profile-specific estimates of survival and survival differences? To get a sense of the survival statistics presented, we examined the reports of randomized trials published between January and June 2006 in three widely read general medical journals (NEJM, JAMA, The Lancet). We restricted our attention to the 20 studies in which there was a statistically significant survival difference between the compared treatments with respect to the primary endpoint. We documented how many of these reports presented profile-specific t-year and treatment-specific survival, or its complement, the risk of the event over the t years. Our primary focus was on profile-specific statistics; however, as a side issue, we were also interested in how treatment-specific statistics were reported for the 'aggregate' or the 'typical.' Thus, we documented how often summary statistics were presented in the abstract as (or in a way that allowed the reader to calculate) absolute risk, and risk differences between treatments [7].

We found that most abstracts contained information on the risk and risk difference for the 'average' patient. Some articles did provide risk differences or hazard ratios for 'univariate' subgroups (e.g., by age or by sex). However, despite the range of risk profiles in each study, and the common use of Cox regression, none of the 20 reports presented information that would allow the reader to assess treatment-specific risk for a specific profile, e.g., for a specific age-sex combination.

Risk difference: overall and profile-specific

The absence of reported profile-specific risks stems in part from the widespread use of Cox's proportional hazards model [8], with its focus on event rates (i.e., incidence) rather than t-year cumulative incidence or risk (i.e., the complement of t-year survival probability). Since survival differences are expressed primarily as hazard ratios, absolute measures such as risk and risk difference – the recommended, more meaningful, metrics [9,10] – are underused. Similarly, with logistic regression models – where the time element is nonexistent, irrelevant or unimportant – the use of odds ratios has led to a neglect of absolute risk and risk differences [7].

It is becoming more widely understood that if the odds ratio or risk ratio is the same across profiles, then the risk differences can not be the same, i.e., that, patients with different profiles will derive different degrees of absolute benefit; see, for example, Califf *et al*. [11]. However, despite extensive coverage of it in specialized modeling texts [12], there does not seem to be the same widespread appreciation of the implications when a treatment benefit is reported as a hazard ratio. Part of the difficulty lies in the fact that the cumulative incidence, or risk, over the time window $(0, t)$ is a complicated function of the integral of the incidence function over that window. If the integral, and thus the cumulative incidence, is small, the reported hazard ratio can indeed be used as a good approximation to the risk ratio; otherwise, the 'translation' is more complicated. Either way, however, a hazard ratio that is the same across patient profiles, implies that the absolute risk differences can not be the same.

This ambiguity in the interpretation of hazard ratios, and lack of appreciation of the implications for risk differences, are exemplified by the statement, taken from the report of the clinical trial cited in Case 1, that 'the benefit of radical therapy . . . did not differ according to the PSA level or the Gleason score.' A reader might well take this statement to mean that the 10-year risk difference of 5 percentage points – implying a 'number needed to treat' of 20 – was the same for all Gleason scores. In fact, the constancy is with respect to hazard ratios. A $(100 - 56 =)44\%$ reduction in mortality rates, whatever the Gleason score, implies a larger treatment benefit (risk difference) for men with (higher risk) Gleason 7 than with (lower risk) Gleason 5 cancer. The report would have been more helpful if, before addressing subgroup analyses showing no significant treatment-prognostic factor interactions, it focused first on (i) the range of fitted absolute 10-year risks, with conservative treatment, for the various

age-Gleason Score-PSA profiles, and (ii) the absolute risk reductions, if persons with these same profiles underwent radical treatment. The article by Califf *et al*. [11] takes such a counterfactual approach. Since the focus was on 30-day mortality risk, and there were no censored observations, its authors used the logistic regression model.

The prevailing emphasis on ratio measures, and neglect of profile-specific risks, are unfortunate, since Cox's article [8], went well beyond constant-over-time (and nonconstant-over-time) hazard ratios. He entitled his article 'Regression Models and Life Tables', and provided a method to derive profile-specific survival, or its complement, cumulative incidence (risk), from his model. The Life Tables portion has been underused.

Besides low awareness, we suspect two other reasons profile-specific survival estimates are not reported: a lack of understanding as to how they are derived, and the difficulty in presenting them in a compact form.

Thus, our objectives are two-fold, to (i) re-iterate that profile-specific risk estimates are possible, and (ii) show how they can be reported in a compact form. To address (i), we refer to a recent expository article [13]. For (ii), we take advantage of the long-established culture of using scoring systems, to enable end-users, using either a table look-up or a nomogram, to derive profile-specific estimates to quickly and easily from complex regression equations. Probability look-up tables have had a long history in medicine: one of the best known is the one for estimating coronary heart disease risk based on data from the Framingham Heart Study [14]; they are also becoming common in clinical medicine [15,16]. Thanks to publicly available software [17] to create them, nomograms[18] are sometimes used instead of tables [19,20].

Materials and Methods

Data for illustration

The Systolic Hypertension in the Elderly Program (SHEP) study [5] addressed the effectiveness of antihypertensive drug treatment in reducing the risk of stroke in persons with isolated systolic hypertension. We obtained the data, without subject identifications, under an NHLBI program [21]. We analysed 4701 records with complete data on age, sex, race, systolic blood pressure and assigned treatment {active, placebo} (Table 1). In the 20 894 person-years of follow-up, incident stroke was diagnosed in 263 persons.

Risk differences via Cox regression

We took as the 'reference' profile 60 year old white females with a systolic blood pressure of 140 mmHg. We fitted Cox's semi-parametric log-hazard model with the five linear predictors: age (in yrs) minus 60, indicator of male gender and black race, systolic blood pressure minus 140, and an indicator of active treatment. Effect modification was investigated using products of the indicator of active treatment and each of the other four terms. Estimated *t*-year stroke-free survival proportions for persons with the reference profile were obtained using the Breslow estimator [22]; profile-specific risk estimates were then calculated as described in Appendix 1. Risk estimates for two selected profiles, as well as the 'average' cumulative incidence curves in the original article, are shown graphically.

Compact presentation of treatment- and profile-specific estimates

In order to present estimates of the treatment- and profile-specific risk and risk difference for any given profile, we use two different approaches. In the first, we form a scoring system, and couple it with a look-up-table that yields the risk for that profile (i) if untreated and (ii) if treated. The second uses a nomogram: points for each risk factor are read from 'rulers', and the total number of points is converted into risk estimates and a risk difference. Detailed methods and calculations, along

Table 1 For each of the two intervention groups ($Tx = 1$ for Active, $Tx = 0$ for Placebo) in SHEP study, distributions of prognostic indicators; also shown are the respective numbers of subjects and strokes

	Age:			Sex:	Race:	SBP:			No. of	No. of
Tx	Q_{10}	Q_{50}	Q_{90}	% male	% Black	Q_{10}	Q_{50}	Q_{90}	subjects	strokes
0	64	72	81	43	14	161	168	183	2351	158
1	64	72	81	44	14	161	168	185	2350	105

Note: Q_{10}, Q_{50}, and Q_{90} are the 10th, 50th, and 90th centiles. SBP: Systolic Blood Pressure.

with a link to computer code, are presented in Appendices 1 and 2.

Results/Illustration

The estimated 1-, 2-, . . . , 5-year risk of stroke for persons with the reference profile were 0.5, 1, 1.3, 1.8, and 2.2%, respectively. The estimated 5-year risk of stroke for a 65 year old white female with a SBP of 160 mmHg was 3.8% if untreated, and 2.5% if treated. While these imply the risk reduction estimate of 1.3%, the corresponding estimate for an 80 year old black male with a SBP of 180 mmHg (calculations shown in Table A2) was $16.0 - 10.7 = 5.3\%$, both appreciably different from the overall estimate of $7.6 - 4.9 = 2.7\%$. For these low and high-risk profiles, the respective numbers needed to treat (NNT) are $100/1.3 = 77$ and $100/5.3 = 19$, both very different from the NNT estimate of $100/2.7 = 37$ obtained using the risk difference corresponding to the 'typical' or 'average' patient in this trial. This increasing risk difference with increasing risk is also evident in Figure 1.

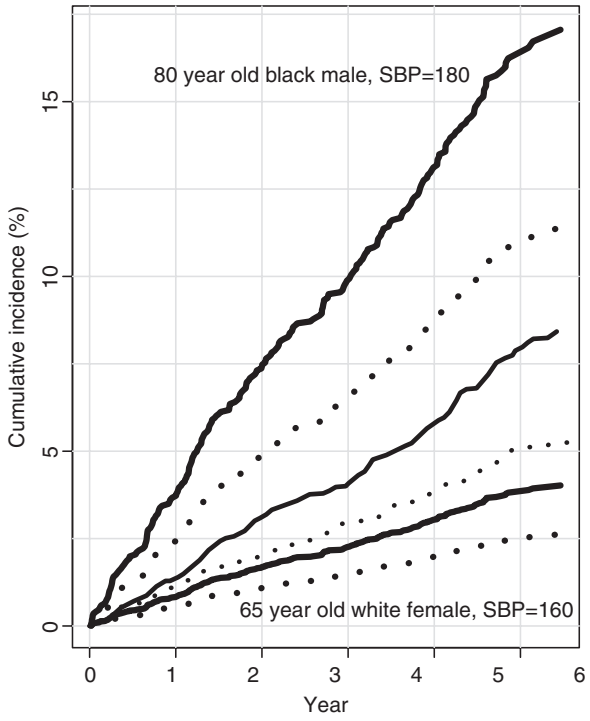

Figure 1 Cumulative incidence (risk) of stroke for patients with higher- and lower-risk profiles, if untreated (solid line) and if treated (dotted line); estimates from Cox regression; data from SHEP study (see text). The middle two curves, extracted from the original report, are the cumulative incidence curves for the 'average' patient, if untreated (solid line) and if treated (dotted line)

The scoring system and the look-up table are shown in Table 2. Alternatively, as shown in Figure 2, the points for each risk factor in the profile can be read from 'rulers.' The scale for each ruler highlights the relative importance of the risk factor.

The 'increasing risk difference with increasing risk' is again evident in Table 2 and Figure 2. Since even the highest risk is below 20%, the (across the spectrum) 35% reduction in incidence translates into approximately an (across the spectrum) 35% reduction in cumulative incidence.

Discussion

Our aim was to highlight four points about profile-specific risk (survival) estimates: they are (i) practice-relevant (ii) almost never reported (iii) directly estimable from the Cox model and (iv) easy to report in a compact form. We end by anticipating some of the issues that our proposal may raise.

Fitted (predicted) values vs. results from subgroup analyses

We first wish to dispel any suggestion that we are advocating *post-hoc* subgroup analyses – with their risk of false positive findings. Our call for the reporting of profile-specific estimates is not a call for subgroup analysis; it is a call to use Cox regression –- used primarily up to now to redress treatment imbalances – to describe the variation in predicted risk if untreated, and risk difference if treated, across all profiles. Subgroup analyses divide up the data; the profile-specific estimates we describe are 'predicted' values from a proportional hazards model fitted to all of the data, but 'translated' into different risk differences depending on the baseline risk.

Fitted values from statistical models vs. actual data: 'borrowing strength'

Since trials are forced to include patients with a spectrum of prognoses, the result for the treatment comparison within each separate profile has a sizable margin of error. The margin of error for the estimate of profile-specific treatment benefit can be reduced by statistical aggregating/averaging of the estimates across profiles, using (as in meta-analyses) the scale on which the benefits are (or are assumed to be) the most homogeneous. Because of the small sample sizes per profile, we often 'aggregate' using regression models.

Table 2 Risk estimates (%) for stroke in the next 1, . . . , 5 years, if the SBP will not be treated ($Tx = 0$) and if it will ($Tx = 1$), as a function of the four prognostic indicators incorporated in the Total Score [points are proportional to coefficients in Cox model shown in Table 3]

	Total Score	*Tx*	Year 1	2	3	4	5
	200	0	3.4	7.0	9.3	12.3	15.4
		1	2.2	4.6	6.1	8.2	10.3
(No. years beyond 60) × 4 _____	150	0	2.1	4.3	5.7	7.7	9.7
		1	1.4	2.8	3.8	5.0	6.4
Male . . . 25 _____	100	0	1.3	2.6	3.5	4.7	6.0
		1	0.8	1.7	2.3	3.1	3.9
Black . . . 30 _____	50	0	0.8	1.6	2.2	2.9	3.7
		1	0.5	1.0	1.4	1.9	2.4
(Every 10 mm SBP above 140) × 17 _____	0	0	0.5	1.0	1.3	1.8	2.2
		1	0.3	0.6	0.8	1.1	1.5
Total score _____							

Example calculation for 80 year old black male with a SBP of 180 mmHg. Score = $20 \times 4 + 25 + 30 + 4 \times 17 = 203$. Risk estimate for stroke in the next 5 years, if the SBP will not be treated is ≈ 16%, and if it will be treated is ≈ 10%.

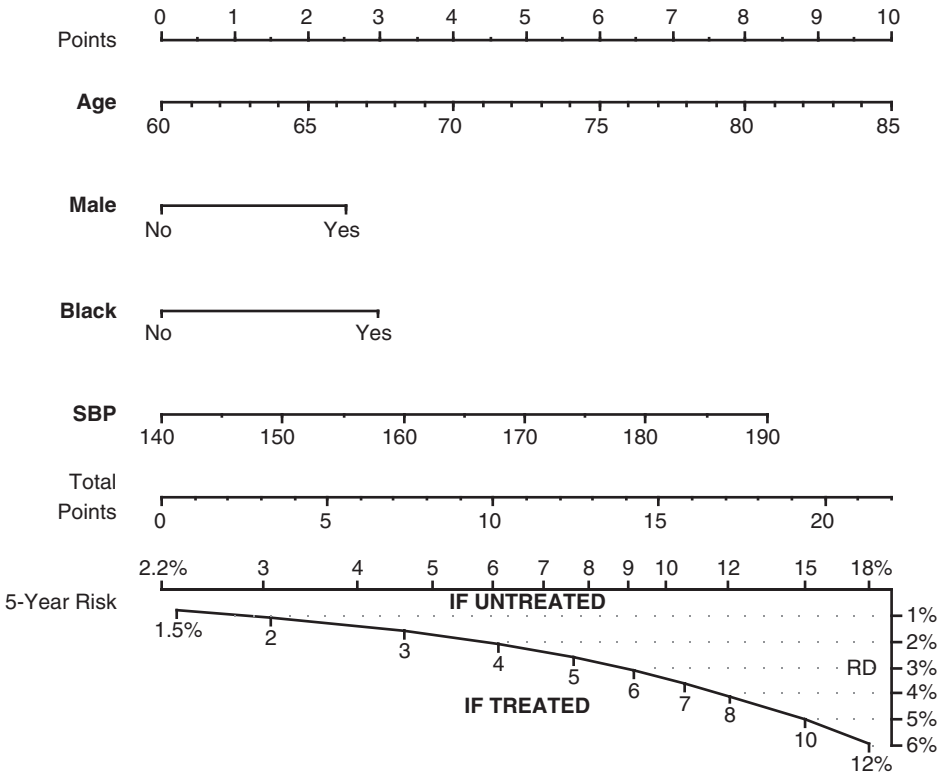

Figure 2 Nomogram to calculate estimated 5-year risk of stroke if untreated, or if treated. Points – proportional to fitted Cox regression coefficients – for the four factors (Age to SBP) are summed and transferred to 'Total Points' scale. The corresponding risks and Risk Difference (RD) are read from the bottom two scales. Data from SHEP study (see text)

Whereas regression models are often used to 'adjust' imbalanced comparisons, the fitted values from these models can also be used as more reliable estimates for profiles with few subjects. These estimates 'borrow' from the remainder of the data. To convince colleagues that they 'smooth/borrow' instinctively, we asked them to estimate, from Table 1 in [3,23], the prognosis for a man diagnosed at age 58, with a Gleason 5 prostate cancer (there were only eight such men in the data set)

or – worse – a similarly aged man with a Gleason 8 cancer (there were only two such men in this cell). We observed them as they examined the experience in adjacent cells, and based their estimate for the cell of interest on an informal weighted average of the data in it and in the adjacent cells. The metaphor of 'borrowing strength', which was used to justify the regression models in [23], was a favourite of the eminent statisticians Frederick Mosteller and his colleague, John Tukey, who used it several decades ago when making USA election-night projections from partial returns. The advice of statistician George Box is also relevant here: 'all models are wrong, but some are more useful than others.' Forced to estimate the prognosis for men in a 'statistical cell' for which we have little data, we prefer a more precise but possibly slightly off-target 'synthetic' estimate, i.e. one based on an '*inter*polation' model and substantial surrounding data, than a less precise 'standalone' one. We do not advocate *extra*polation.

More complex, and 'time-irrelevant', risk models

The data in our worked example were adequately fitted by hazard ratios for treatment that were constant over time and over covariate-patterns. However, our call for the production of profile-specific estimates does not preclude pre-planned subgroup analyses (with varying hazard ratios), or more complex hazard models. Moreover, although we focused on the Cox model and cumulative incidence, our plea for 'individualization' is also relevant for data-analyses carried out under other statistical models, such as logistic regression, where the time element is nonexistent, irrelevant, or unimportant.

Alternative presentations of profile-specific estimates

When – as in the 45 000 patient study cited earlier – we have a large amount of information concerning the profile of a particular patient, we depend less on data from adjacent profiles and statistical models: the data in that cell are sufficient. Since a printed article cannot include estimates for each of the profiles, they would need to be provided as supplementary information on the journal website. Alternatively, since the pattern of the profile-specific statistics can usually be adequately summarized (and 'held together') by a statistical model, the fitted values for each profile might be compactly presented within the printed report, using a scoring system and a table/nomogram.

One other way to present information on treatment benefit is via a 'prognostic stratification' system[24,25], in which several prognostic variables are mapped into discrete categories within which risks are more homogeneous. An example is the use of the TNM items to define cancer stages I, II, III, and IV. These systems are used to guide treatment choices, to achieve better balance of prognostic factors in clinical trials, and to come closer to 'comparing like with like' in nonrandomized comparisons of treatments, especially over time or between institutions.

If such a sufficiently fine prognostic stratification system already exists, it can be used to add specificity in the presentation of treatment results. For example, one might, as in[3], use D'Amico's risk groupings for clinically localized prostate cancer to report the statistics with different treatment choices, separately for men with high, intermediate and low-risk prostate cancers.

The 'prognostic stratification' alternative will appeal to those who are concerned about the number and nature of the mathematical assumptions inherent in the regression-based scoring systems, with numbers of points (coefficients) derived from statistical models fitted to the entire dataset. Such end-users may prefer to rely on estimates based on the 'strictly local' (but more limited) data and a much simpler – or no – statistical model, rather than on estimates derived from the global data tied together by a model. However, this local approach will produce wide confidence intervals. A more sensible approach [26] is to 'derive a continuous risk score from a model in which all relevant covariates are kept continuous, and then to apply categorization at the final step, so that profiles are divided into several groups for clinical application by applying cutpoints to the risk score.'

Beyond the one-number summary of treatment benefit

Before presenting the illustration, we asked clinicians to recall the treatment benefit reported in the SHEP study. They remembered 'a 30% reduction' – but not the average risk difference of 3% at five years. Relying on a one-number estimate of percentage risk reduction, rather than a one-number average risk difference, may not be entirely inappropriate: indeed a case could be made that if one could only remember one 'treatment benefit' number, it should be the ratio measure. The absolute risk difference observed in a trial is a function both of the follow-up time, and the average risk profile of those enrolled – possibly different from that of the clinician's own patients.

For a specific profile, the clinician informally estimates the level of absolute risk (if untreated) [27] and applies the percentage reduction to it, knowing instinctively that the risk difference will likely be larger if the patient is higher risk. However, without a good sense of the absolute risk to which this percentage reduction applies, it is not possible to accurately weigh the probability of benefit against the costs and the probability of harm.

As clinicians increasingly go online or use mobile devices for decision support in the course of patient care, their ability to use risk functions to estimate profile-specific risks also increases. No longer constrained by how many treatment-benefit numbers they can store in their memory, they can now have electronic 'access' to the two relevant risks, for whichever risk profile. Indeed, they already use support systems to compute risks for myocardial infarction [28], and diagnostic probabilities [29]. In order to provide doctors with more patient– and practice-relevant information, reports of clinical trials should routinely provide the predicted (i.e., fitted) survival function for the reference profile, together with the regression coefficients (or hazard ratios) required to convert it into profile-specific survival functions. Or, they can present the risk estimates using risk scores, which are then converted into risks via either a nomogram or a table. Moreover, if a sufficiently fine prognostic stratification system already exists, reports should also provide the *t*-year risk, and risk difference, specific to each stage or risk group.

Acknowledgments

This work was supported by grants from The Natural Sciences and Engineering Research Council of Canada and Le Fonds Québécois de la recherche sur la nature et les technologies. We thank the large number of people who provided comments on earlier versions of this manuscript.

References

1. **Bill-Axelson A, Holmberg L, Ruutu M,** *et al.* Radical prostatectomy versus watchful waiting in early prostate cancer. *N Engl J Medicine* 2005; **352**: 1977–84.
2. **Albertsen PC, Hanley JA, Fine J.** 20-year outcomes following conservative management of clinically localized prostate cancer. *Journal of American Medical Association* 2005; **293**: 2095–101.
3. **Albertsen PC, Hanley JA, Penson DF, Barrows G, Fine J.** 13-year outcomes following treatment for clinically localized prostate cancer in a population based cohort. *J Urology*. 2007; **177**: 932–36.
4. **Wong YN, Mitra N, Hudes G,** *et al.* Survival associated with treatment vs observation of localized prostate cancer in elderly men. *Journal of American Medical Association*. 2006; **296**: 2683–93.
5. **SHEP Cooperative Research Group.** Prevention of stroke by antihypertensive drug treatment in older persons with isolated systolic hypertension. Final results of the Systolic Hypertension in the Elderly Program (SHEP). *Journal of American Medical Association* 1991: **265**: 3255–64.
6. **Feinstein AR, Horwitz RI.** Problems in the ''evidence'' of ''evidence-based medicine''. *American J Medicine* 1997; **103**(6): 529–35.
7. **Schwartz LM, Woloshin S, Dvorin EL, Welch HG.** Ratio measures in leading medical journals: structured review of accessibility of underlying absolute risks. *BMJ* 2006; **333**: 1248. (epub 23 Oct 2006).
8. **Cox DR.** Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society*, Series B 1972; **34**: 187–220.
9. **The CONSORT Group.** Consort statement. www.consort-statement.org/Statement/revisedstatement.htm (accessed 21 May 2007).
10. **STROBE statement.** checklist of essential items. www.strobe-statement.org/PDF/STROBE-Checklist-Version3.pdf (accessed 21 May 2007).
11. **Califf RM, Woodlief LH, Harrell Jr FE,** *et al.* Selection of thrombolytic therapy for individual patients: development of a clinical model. *American Heart Journal* 1997; **133**(6): 630–39.
12. **Harrell FE.** Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis. Section 17.2.3. Springer, New York, 2001.
13. **Hanley JA.** The Breslow estimator of the non-parametric baseline survivor function in Cox's regression model: some heuristics. *Epidemiology* 2008; **19**(1): 101–102.
14. **Estimating Coronary Heart Disease (CHD)**: Risk Using Framingham Heart Study Prediction Score Sheets http://www.nhlbi.nih.gov/about/framingham/riskabs.htm (accessed 01 December 2007).
15. **Kannel WB, D'Agostino RB, Silbershatz H,** *et al.* Profile for estimating risk of heart failure. *Archives of Internal Medicine* 1999; **159**(11): 1197–204.
16. **Steyerberg EW, Vergouwe Y, Keizer HJ. Habbema JDF.** Residual mass histology in testicular cancer: development and validation of a clinical prediction rule. *Statistics in Medicine* 2001; **20**: 3847–59.
17. **Harrell FE.** The Design package for R. http://cran.r-project.org/web/packages/Design/index.html (accessed 17 Feb 2008).
18. **Wikipedia. Nomogram.** http://en.wikipedia.org/wiki/Nomogram (accessed 01 December 2007).
19. **Tisman G.** Using Nomograms to Predict Prostate Cancer Treatment Outcomes. http://www.prostate-cancer.org/resource/pdf/Is8-4.pdf (accessed 01 December 2007).
20. **Bevilacqua JL, Kattan MW, Fey JV,** *et al.* Doctor, what are my chances of having a positive sentinel node? A validated nomogram for risk estimation. *Journal of Clinical Oncology* 2007 Aug 20; **25**(24): 3670–9. (epub 30 July 2007).
21. **National Heart Lung and Blood Institute (USA).** The NHLBI Limited Access Dataset Program //www.nhlbi.nih.gov/resources/deca/directry.htm) (accessed 06 March 2007).
22. **Breslow NE.** Contribution to the discussion of paper by D.R. Cox. *Journal of the Royal Statistical Society*, Series B. 1972; **34**: 216–17.
23. **Albertsen PC, Hanley JA, Gleason DF, Barry MJ.** Competing risk analysis of men aged 55 to 74 years at diagnosis managed conservatively for clinically localized

prostate cancer. *Journal of American Medical Association* 1998; **280**: 975–80.
24. **Feinstein AR.** Clinical biostatistics. XIV. The purposes of prognostic stratification. *Clin Pharmacol Therapeutics* 1972; **13**(2): 285–97.
25. **Feinstein AR.** Clinical biostatistics. XVI. The process of prognostic stratification. *Clin Pharmacol Therapeutics* 1972; **13**(4): 609–24.
26. **Royston P, Altman DG, Sauerbrei W.** Dichotomizing continuous predictors in multiple regression: a bad idea. *Statistics in Medicine* 2006; **25**: 127–41.
27. **McAlister FA, Straus SE, Guyatt GH, Haynes RB.** Users' guides to the medical literature: XX. Integrating research evidence with the care of the individual patient. Evidence-Based Medicine Working Group. *Journal of American Medical Association* 2000; **283**(21): 2829–36.
28. **National Heart Lung and Blood Institute (USA).** Information for health professionals: Interactive Tools and Resources (http://www.nhlbi.nih.gov/health/prof/other/index.htm) (accessed 02 May 2007).
29. **Partin tables for the probability of pathological stage of prostate cancer.** http://urology.jhu.edu/prostate/partintables.php (accessed 06 March 2007).
30. **Pocock SJ, Clayton TC, Altman DG.** Survival plots of time-to-event outcomes in clinical trials: good practice and pitfalls. *The Lancet*. 2002; **359**(9318): 1686–89.

Appendix 1: Obtaining profile-specific risks from Cox model

As with any regression equation, the point of departure is the estimate for a 'reference' or 'starting' point, or category, before proceeding in increments from there. A useful example is sex- and height-specific estimates of ideal weight (the health risks of being below this weight balance those of being above it). A convenient starting point is the one for women, five feet tall. Before recent research revised the estimates upwards, the equation used to be 100 lbs (reference value), plus 5 lbs for every inch above the reference height, plus '10 additional lbs and 1 additional lb/inch' if male.

Similarly, the point of departure for profile-specific survival curves is to fit a 'reference' survival curve $S_{\text{reference}}(t)$ – the curve corresponding to the profile in which the value of each prognostic variable is in/at the reference category/value (we will explain in the next paragraph how this curve is estimated). Then, since the log hazard functions in Cox's regression model are held together by a linear relationship, each profile-specific $\log\{S(t)\}$ curve is obtained from the reference $\log\{S(t)\}$ curve using the relationship

$$\log\{S_{\text{profile}}(t)\} = \log\{S_{\text{reference}}(t)\} + LP_{\text{profile}}$$

where the 'linear predictor' LP_{profile} for a profile is the score formed by multiplying each fitted regression coefficient by the amount the variable exceeds the reference point, and summing these products. When transformed to the survival scale, this relationship becomes

$$S_{\text{profile}}(t) = \{S_{\text{reference}}(t)\}^{\exp(LP_{\text{profile}})}.$$

There are a number of estimators of the 'reference' curve $S_{\text{reference}}(t)$, including the one suggested by Cox. Breslow's estimator [22] has an intuitive form: it is *the Kaplan–Meier curve obtained if, based on the fitted regression coefficients, each actual subject in the study was converted into a number of hypothetical 'reference-profile' subjects who collectively carry the same short-term risk as the actual one*. [13] For example, a subject whose profile implies a short-term risk 2.5 times that of subjects with the 'reference' profile contributes 2.5 'reference-equivalent' observations to each denominator (number at risk). The curve $S_{\text{reference}}(t)$ is then estimated by applying the Kaplan–Meier estimator to this 'effective number' of 'reference-equivalent' observations. Since it is estimated using all of the observations (with each given a weight), the reference curve is a step-function containing as many jumps as there are events in the entire data set (e.g., each curve shown in Figure 1 is based on 263 strokes). The use of all observations, 'homogenized' by a statistical model, makes the estimated reference curve a more stable starting point (or 'intercept') for the estimated survival curves for other profiles. The fitted $S_{\text{profile}}(t)$ is, of course, invariant to the choice of reference profile.

Each of the mainstream software packages for survival analysis under the Cox model includes the option of obtaining survival curves for specified profiles. The steps are listed in Table A1. The complements of these (descending) survival curves represent cumulative incidence. Clinical trial authorities[30] consider these ascending risk functions more informative.

Appendix 2: Application to SHEP data

We fitted Cox's semi-parametric log-hazard model with 'linear predictor' $\Sigma\beta_k X_k$, where $X_1 = \text{Age (in yrs)} - 60$, $X_2 =$ Indicator of male gender, $X_3 =$ Indicator of Black race, $X_4 = \text{Systolic BP (in mmHg)} - 140$ and $X_5 =$ Indicator of active treatment. The fitted values for the parameters are given in Table A2. The estimated 1-, 2-, . . . , 5-year stroke-free survival proportions for persons with the reference profile are $S_{\text{ref.}}[1] = 0.995$, $S_{\text{ref.}}[2] = 0.990$, $S_{\text{ref.}}[3] = 0.987$, $S_{\text{ref.}}[4] = 0.982$, and $S_{\text{ref.}}[5] = 0.978$, respectively.

Although, one would normally use the software to perform it, we illustrate the arithmetic involved by calculating the estimated 5-year risk

Table A1 How to obtain profile-specific survival curves under the Cox model in each of the major statistical packages

Package	Procedure	Statement	Steps
SAS	PHREG	BASELINE	1. Form separate dataset containing profiles of interest. 2. In BASELINE statement, point to this dataset.
Stata[a]	stcox	basesurv( )	1. Use basesurv( ) to store estimated curve for baseline profile. 2. For each profile of interest, calculate corresponding hazard ratio HR from regression coefficients. 3. Insert HR in '$S_{\text{reference}}(t)$ to $S_{\text{profile}}(t)$' relationship.
R/S-Plus/survival	coxph	survfit	Specify profiles via newdata option
R/S-Plus/Design	cph	survest	See also survplot and nomogram
SPSS	COXREG	/PATTERN	Specify profiles in /PATTERN statement.

Note: [a]See http://www.ats.ucla.edu/STAT/stata/seminars/stata_survival/default.htm.

of stroke for a 65 year old white female with a SBP of 160 mmHg. Relative to persons in the reference profile, the hazard for white females like her, 5 years and 20 mmHg above the reference values, is $\exp(0.041 \times 5 + 0.259 \times 0 + 0.303 \times 0 + 0.017 \times 20 - 0.435 \times 0) = \exp(0.545) = 1.72$ if untreated, and $\exp(0.545 - 0.435) = 1.12$ if treated. Thus, the estimates of stroke-free survival at 5 years are $0.978^{1.72} = 0.962$, and $0.978^{1.12} = 0.975$, respectively, so that the cumulative incidence (CI) estimates are $1 - 0.962 = 0.038$ and $1 - 0.975 = 0.025$ respectively.

The quantity 0.545 in the worked example was obtained – from what statisticians call the 'linear predictor' – by summing 'points': 0.041 for every year of age above the reference age, 0.017 for every mmHg above the reference SBP, etc. The resulting 'score' is exponentiated and the profile-specific survival estimate obtained as $\{S_{\text{ref.}}[t]\}^{\exp(\text{score})}$. We can generalize this approach, as has been done with Framingham data, to form a scoring system, and couple it with a look-up-table that yields the cumulative incidence (risk) for any given profile. To avoid decimals, the regression coefficients (i.e., the 'points per unit value above the reference value') used in the scoring are scaled up, and then rounded to integers. The scoring system and the look-up table are shown in Table 2. If required, a confidence interval for the estimated risk, or risk difference, can be derived from the variance-covariance matrix of the estimated regression coefficients by the parametric bootstrap.

Table A2 Fitted values for the coefficients of Cox regression model: data from SHEP study

Age-60	I_{Male}	I_{Black}	SBP − 140	$I_{\text{ActiveTreatment}}$
0.041	0.259	0.303	0.017	−0.435[a]

Note: [a]Hazard ratio = exp(−0.435) = 0.65 (35% reduction).

Alternatively, as shown in Figure 2, the points for each risk factor in the profile can be read from 'rulers.' Figures such as this can be formed using the nomogram function in the 'Design' package [17] in R. Computer code is provided in http://www.epi.mcgill.ca/hanley/software.

Since even the highest risk in Figure 2 is below 20%, the (across the spectrum) 35% reduction in incidence translates into approximately an (across the spectrum) 35% reduction in cumulative incidence. If cumulative incidence (CI) is low, say less than 20%, the exact relationship between $CI[t]$ and the hazard function $h[u]$, namely $CI_{0\ to\ t} = 1 - \exp[-\int_0^t h[u]\mathrm{d}u]$, can be approximated by $CI_{0\ to\ t} \approx \int_0^t h[u]\mathrm{d}u$. In such cases, where the event rate ratio in the treated versus untreated is $HR:1$, then for any specific profile, $CI_{0\ to\ t,\ \text{if treated}} \approx HR \times CI_{0\ to\ t,\ \text{if not treated}}$.