

Mathematical and Statistical Models for Research

Samuel Litwin, Ph.D., Member

Eric Ross, Ph.D., Associate Member, Senior Director of Academic Informatics, Director, Biostatistics and Populations Studies Facilities

Eugene Toll, M.S., Volunteer



Interpretation of clinical or biological data often requires a mathematical or statistical model. Such models can increase the power of experiments to decide between clinical treatments or to evaluate hypotheses. An example of our work in the design of a phase II clinical trial is given.

Early stopping designs based on survival: A new paradigm. Litwin in collaboration with Wong,[§] Hudes[§]

New cancer treatments are first tested in clinical trials to determine the maximum tolerated dose of a particular regimen. Since these clinical trials are often small and consist of patients with a variety of advanced cancers, they provide little information about a regimen's activity in any particular malignancy. Promising agents are then tested in larger, uncontrolled phase II studies of specific malignancies to determine their anticancer activity in a particular disease setting.

Many phase II studies are based on the Simon 'minimax' design. In this design an initial cohort of patients is accrued and evaluated for futility. A second cohort is accrued only if the treatment has an encouraging response rate (R. Simon, *Control Clin. Trials* 10:1, 1989). For studies of traditional cytotoxic chemotherapy agents that cause tumor shrinkage, this is an attractive design because it permits prompt evaluation of response while also minimizing patient exposure to drugs with minimal activity.

However, the introduction of novel anti-cancer agents in the last decade has prompted investigators to consider alternative endpoints for clinical trial designs (Korn, et al., *J. Clin. Oncol.* 19:265, 2001). Several of these 'targeted therapies' have been found to cause prolonged tumor stabilization, rather than tumor shrinkage. Therefore, in clinical trials of these agents, progression free survival (PFS) may be an appropriate primary endpoint, since the response rate does not reflect the potential benefit of disease stabilization. Making cancer a chronic disease (by tumor stabilization) would be a major accomplishment.

Designing such studies can be challenging. Some clinical trials use PFS as the primary endpoint in the traditional 'minimax' design. This requires accrual to be halted between stages while PFS is being evaluated, and this can prolong the study's length. Another option is to use PFS as the primary endpoint with the response rate as the criterion for early stopping. While this may prevent the delays inherent in the use of PFS as the early stopping criterion, it causes trials to be closed early for lack of response, even though their main effect is prolongation of PFS rather than tumor shrinkage.

We describe an early stopping design where 'early' means: 1) patients are evaluated for survival at a time much earlier than the target endpoint; and 2) only an initial cohort, usually about half, of the targeted sample size, is evaluated.

We make no assumptions regarding the distributions of survival time or time of patient arrival. This contrasts with the usual practice in survival. We describe an early stopping design based only on stated probabilities of survival at early and final evaluation times. Early stopping is the result of patients not progression-free (PF) to the early time point. Initial cohort patients who are PF past the early time remain on study until final evaluation. Only sums and products of the binomial distribution are needed for an exact determination of power, chances of incorrect conclusions and early stopping probabilities.

The phase II design requires that we place n_1 patients on study and evaluate each patient for PFS at time t_1 after his or her arrival. The study will be stopped for lack of efficacy if too few of them are PF at t_1 . If enough patients are PF at t_1 then accrual is continued until a total of n_2 can be evaluated. Patients in the original cohort (n_1) who are PF at t_1 continue on study. This aspect of the design differs from the usual early stopping phase II study where all n_1 patients are fully evaluated before continuing. Patients who continue on study from the first group plus the $n_2 - n_1$ new patients are evaluated after being on study for time t_2 .

We test the composite null hypothesis that the chance of PFS at time t_1 does not exceed p_{01} and the chance of PFS at time t_2 , conditional on being progression free at t_1 does not exceed p_{02} . The alternative of interest is that both probabilities are greater, namely: $p_{11} > p_{01}$ and $p_{12} > p_{02}$. In that case it is also true that $p_{11} p_{12} > p_{01} p_{02}$. This product is the chance of PFS at time t_2 .

Let $S_1(t_1)$ be the number of patients in the first group who are PF at t_1 , with $S_1(t_2)$ and $S_2(t_2)$ similarly referring to the time, t_2 , of final evaluation. If $S_1(t_1)$ is not greater than a_1 the study will be stopped after n_1 patients are evaluable at t_1 . If either $S_1(t_1)$ is not greater than a_1 or $S_1(t_2) + S_2(t_2)$ is not greater than a_2 the null hypothesis is accepted. The chance that the null hypothesis is rejected, $p(R)$, is:

$$p(R) = p[S_1(t_1) > a_1, S_1(t_2) + S_2(t_2) > a_2]$$

$$p(R) = \sum_B p[S_1(t_1) = i, S_1(t_2) = j, S_2(t_2) = k]$$

$$p(R) = \sum_B b(i, n_1, p_1) b(j, i, p_2) b(k, n_2 - n_1, p_1 p_2)$$

where $b(k, n, p)$ is the binomial distribution, p_1 is the chance of PFS at time t_1 , p_2 is the chance of PFS at time t_2 , given PFS at t_1 and

$$B = \{(i, j, k) \mid a_1 + 1 \leq i \leq n_1, 0 \leq j \leq i, \max(0, a_2 - j + 1) \leq k \leq n_2 - n_1\}.$$

If, for example: $n_1 = 17$, $n_2 = 34$, $a_1 = 10$, $a_2 = 11$, and $p_1 = 0.5$, $p_2 = 0.6$ then $p(R) = 0.0969$. These values of p_1 and p_2 constitute our null hypothesis, so, in this case, $p(R)$ is the chance of concluding that the treatment is efficacious, when it is not. Their product, $p_1 p_2 = 0.3$, is the chance of survival to time t_2 under the null hypothesis. If $p_1 = 0.8$ and $p_2 = 0.625$, ($p_1 p_2 = 0.5$), then $p(R) = 0.9381$ is the power to reject the null. This design makes no assumptions as to the distributions of survival time or accrual time. Therefore, it can be used for rigorous power analysis.

Publications

Cooper, H.S., Chang, W.C., Coudry, R., Gary, M.A., Everley, L., Spittle, C.S., Wang, H., Litwin, S., Clapper, M.L. Generation of a unique strain of multiple intestinal neoplasia (Apc(+)/Min-FCCC) mice with significantly increased numbers of colorectal adenomas. *Mol. Carcinog.* **44**:31-41, 2005.

Henry, L.R., Sigurdson, E., Ross, E., Hoffman, J. Hydronephrosis does not preclude curative resection of pelvic recurrences after colorectal surgery. *Ann. Surg. Oncol.* **12**:786-792, 2005.

Lewis, J.S., Meeke, K., Osipo, C., Ross, E.A., Kidawi, N., Li, T., Bell, E., Chandel, N.S., Jordan, V.C. Intrinsic mechanism of estradiol-induced apoptosis in breast cancer cells resistant to estrogen deprivation. *J. Natl. Cancer Inst.* **97**:1746-1759, 2005.

Litwin, S., Toll, E., Jilbert, A.R., Mason, W.S. The competing roles of virus replication and hepatocyte death rates in the emergence of drug-resistant mutants: Theoretical considerations. *J. Clin. Virol.* **35**:S96-S107, 2005.

Schnoll, R.A., Wang, H., Miller, S.M., Babb, J.S., Cornfeld, M.J., Higman-Tofani, S., Henigan-Peel, T., Balslem, A., Slater, E., Ross, E., Boyd, C.S., Engstrom, P.F. Change in worksite smoking behavior following cancer risk feedback intervention: A pilot study. *Am. J. Health Behav.* **29**:215-227, 2005.

§ Fox Chase researcher