

Estimation of Cancer Progression Based Clinical Trial Subgroups

Shankar Srinivasan¹, Lihua Yue² and Weiyuan Chung²

1. Partner, Resource Tepee LLC and Director, Celgene Corporation,

2. Associate Director, Biostatistics, Celgene Corporation.

Abstract: Cancer trials often start investigational therapy at diagnosis or after a selected number of relapses. These are the usual core inclusion criteria in clinical trials. Hence it is helpful when planning a trial to know the likely percentages of patients receiving standard therapy at clinics and hospitals who meet this key inclusion criteria of being newly diagnosed during a period or having just had their first, second or third relapse during an anticipated enrollment time frame. Often regulatory agencies will have approvals tied to the use of a therapy in a relapsed context or in a newly diagnosed context. We provide details on calculations to help those in clinical trial operations make realistic assessments on the number of sites and likely enrollment at clinical trial sites, and the enrollment time frames that might be needed to complete planned total patient enrollment. The estimates complement site feasibility questionnaires which are often sent to gauge patient availability and site interest.

Key words: Clinical trial enrollment, site feasibility, progression/relapse based subgroups.

1. Introduction

The use and approval of a therapy in oncology is often tied to the patient's state in a sequence of transitions from diagnosis through subsequent relapse. See Refs. [1] and [2] for instance. To support the enrollment of patients or the collection of data records constrained to a diagnosis or a relapse based criteria, we estimate the likely proportions of such subsets occurring at medical facilities to complement the use of site feasibility questionnaires [3]. In a number of cancer contexts rough estimates exist for the median across patients of the time period they survived progression free (PFS) and for the time to death (OS) in the population having the disease. Such estimates exist from various starting points such as from diagnosis, from the first relapse, and from second or succeeding relapses [4]. In progression free survival (PFS), the reported median is the time to progression/relapse or death, whichever occurs earlier. The PFS acronym is

often used for hematological cancers. Other terms such as disease free survival (DFS) for the time to disease recurrence or death are common in other oncology contexts.

We first use the median times to these events to compute rough estimates of the proportion of patients who are likely to have their next progression, or die before the end of a specified time-frame, or those who continue through that period without a death or a progression. The latter constitutes those who respond to therapy and would likely stay on therapy, and are not candidates for clinical trials initiating at progressions. We use the exponential distributional assumptions and a uniform incidence of the disease diagnosis over time. The progression, death and ongoing proportion in the enrollment time-frame being considered are used to obtain approximate proportions of trial eligible subsets. Usually additional inclusion/exclusion criteria related to disease, disability and co-morbidity apply as well as competing initiating trials, and the estimates here provide upper bounds and may help personnel in operations make realistic projections about enrollment.

Corresponding author: Shankar Srinivasan, Ph.D., research fields: statistics, epidemiology, clinical trials, and survival analysis.

2. Computed Example

Estimated subsets in a hypothetical context are in Fig. 1 below. Given starting durations going from diagnosis or from some subsequent relapse, the example has median entries under headers "Next PFS" and "Subsequent OS". In the default example we consider a disease condition where the median survival from diagnosis is 4 years and the median time from diagnosis to next (first) progression/death is 1.5 years. Corresponding medians from first relapse are 2 years and 0.75 years (9 months) respectively. In a salvage setting (after second or higher relapse) the medians are 1 year and 0.5 years (6 months) respectively. The period considered is 0.5 years (6 months). Units for time (yrs) are the same for this period and for OS and PFS medians.

The first column estimates the likely breakdown of patients having the disease as a multiple of the number

of newly diagnosed patients. For every newly diagnosed patient during a 6-month enrollment time-frame, 3.85 multiples are ongoing diagnosed patients without a relapse, 0.6 have a first relapse, 0.58 are previously relapsed having a relapse again and 1.60 are previously relapsed ongoing without relapse, all in the same 6-month period. For rapidly progressing disease we also consider unique progressing patients within a period as progressions can occur over more than once within a period for some patients. For every newly diagnosed patient during a 6-month enrollment time-frame above, 0.44 are unique previously relapsed patients having a second or higher relapse again compared to the 0.58 noted when we count relapses instead of the unique patients in the period. This implies a 2.04 multiple unique patients eligible for clinical trials and 5.45 ineligible if trials are to start on relapse or at diagnosis.

Calculator Inputs			
Start of Event Duration	Median Next PFS	Median Subsequent OS	Period Length
Diagnosis	1.5	4	0.5
First Relapse	0.75	2	0.5
Second or Higher to Next Event	0.5	1	0.5
-			
	Per 1	Subgoup	Trial
Subgroup	ND	Proportions	Eligible
Newly Diagnosed (ND)	1	13.36%	49.09%
Ongoing without any relapse	3.85	51.41%	
First Relapse This Period	0.60	7.99%	29.34%
Any Relapse After First in This Period	0.58		
Second Relapse	0.34		
Third Relapse	0.14		
Fourth or Beyond	0.10		
Any Relapse After First in This Period (Unique			
Patients without Prior Relapse in Period)	0.44	5.87%	21.57%
Second Relapse (Unique Patient)	0.27	3.62%	13.28%
Third Relapse (Unique Patient)	0.10	1.32%	4.86%
Fourth or Beyond (Unique Patient)	0.07	0.94%	3.43%
Ongoing Relapsed Patients	1.60	21.37%	
Ongoing after First Relapse	1.02	13.60%	
Ongoing after Second Relapse	0.34	4.55%	
Ongoing after Third Relapse	0.14	1.89%	
Ongoing after Fourth or Higher Relapse	0.10	1.33%	
Trial Eligible Total (Diagnosed or Relapsing			
this Period)	2.04	27.22%	100.00%
Ineligible Total (Ongoing through Period)	5.45	72.78%	
Total	7.48	100.00%	

Fig. 1 Progression based cancer clinical trial subset estimation.

2.1 Interpretation

These are converted to percentages in the second column. In addition to some bias when the assumptions noted earlier do not hold, these percentages are associated with error. This error works out to about $\pm 5\%$ (95% confidence intervals) when considering about 100 patients and it is generally proportional to the inverse of the square root of the number of patients. For every 100 patients in the sites being considered, about 73 are likely to do reasonably well through the 6-month period and are likely to come in for scheduled visits and possibly just pick their prescriptions or undergo any therapies at the site. These patients would not typically volunteer or be considered for clinical trials. Further for those 100 patients about 13 patients would be newly diagnosed during the period, about 8 would have had their first relapse and another 6 would have had a subsequent relapse, leading to only about 27 eligible for a clinical trial. The last column computes these trial eligible proportions as a percent of the total eligible.

Often regulatory agencies request separate study of therapies in the newly diagnosed and the relapsed settings and licensing is restricted to the setting studied. For the hypothetical example considered, only about 13 in 100 patients at clinics and hospitals would be qualified for a trial by a sponsor pursuing a newly diagnosed indication for their therapy. Often sponsors consider approvals for patients difficult to treat such as those with 1 or more relapses. For the default example in this calculator, this would be about 14 patients among 100 visiting clinics and hospitals over a 6-month period. As noted earlier there would be additional attrition in the available patient pools as one applies inclusion and exclusion criteria. For instance, a requirement of refractoriness to a class of medications in addition to being relapsed would further restrict enrollment.

The calculator allows the entry of a fraction of a year or whole years for an enrollment period of interest such as 0.5, 1, and 1.5 years, etc. The reader may use the interactive calculator in Ref. [5] to enter a period of 1.5 years (18 months) instead of the 0.5 years (6 months) used in Fig. 1. An 18-month enrollment is closer to norm in cancer clinical trials than 6 months, though 6 months may sometimes be possible in the context of a disease with high incidence and poor prognosis together with aggressive enrollment. For an 18-month enrollment, with the same medians to progression and death, for 100 patients, 42 may not be eligible for a trial, and 58 eligible for some trial. About 34 and 24 of the 58 eligible will be newly diagnosed and relapsed during the period respectively. Enrollment prospects are a lot better. The next section provides details on the calculations. These are also available in the technical report [6], though this article adds a section on unique relapsing patients during enrollment time frames.

3. Details on Calculations

Consider a newly diagnosed patient at some time S_0 and at every time point obtained by adding or subtracting all whole multiples of a period L. A uniform incidence of the disease can then be obtained by integrating over a uniform measure from S_0 to (S_0 + L). We will see shortly that the subset fractions we obtain are independent of S₀ and hence computed fractions using sequences of patients diagnosed at time-points separated by a period L, hold for a uniform incidence of the disease over time. For each newly diagnosed patient at the start S₀ of a given period, let P_{pfs} and P_{d} be the probabilities of having their first progression or death and death alone in the succeeding period of length L. These can be computed using the median PFS and Median OS, denoted as M_{pfs} and M_d using exponential assumptions (see Ref. [7] for instance) through the hazards h_{pfs} and h_d as:

and

 $P_{d} = \{1 - exp[-h_{d} \times L]\} = \{1 - exp[(ln(0.5) / M_{d}) \times L]\}$

 $\mathbf{P}_{pfs} = \{1 - exp[-\mathbf{h}_{pfs} \times \mathbf{L}]\} =$

 $\{1 - \exp[(\ln(0.5) / M_{pfs}) \times L]\}$

Note that these probabilities are independent of S_0

and depend only on the medians and the period length L. All computations that follow use these or other probabilities that depend only on the medians and the length of the period. Hence, all computations will hold under the uniform incidence rates of the disease over time. The probabilities P_{p} and P_{o} for those with progression (first) alone, and those ongoing through the period without a progression, can be obtained as $P_p =$ $P_{pfs} - P_d$, and $P_o = 1 - P_p - P_d$. Further note that for every newly diagnosed patient from two periods back, $P_p \times P_o$ is the probability of first progression in this period, and for every newly diagnosed patient from three periods back, $P_p \times P_o \times P_o$ is the probability of having their first progression in this period and so on, yielding the expected total of first progressions in a given period T_p, corresponding to the sequence consisting of a newly diagnosed patient at the start of every period, as an infinite geometric series (see Ref. [8] for instance) which was evaluated as follows:

 $T_p = P_p (1 + P_o + P_o^2 + ...) = P_p / (1 - P_o)$

Similarly the expected total ongoing after diagnosis in a period, without having a progression, is given by:

 $T_o = P_o (1 + P_o + P_o ^2 + ...) = P_o / (1 - P_o).$

Proceeding in a similar manner and using the medians after first, second and higher relapses, we can obtain for every patient with the $(k-1)^{\text{th}}$ relapse in the previous period, the probability P_{pk} of having the k^{th} relapse, and the expected fraction P_{ok} ongoing without relapse, in a succeeding period. Using these we can obtain the expected total of second and third progressions in a given period, T_{p2} and T_{p3} respectively, corresponding to the sequence of one newly diagnosed patient per period start, as:

 $T_{p2} = T_p \times P_{p2} (1 + P_{o1} + P_{o1}^2 + ...) =$ $T_p \times P_{p2} / (1 - P_{o1})$

and

$$T_{p3} = T_{p2} \times P_{p3} (1 + P_{o2} + P_{o2}^{2} + ...) =$$

$$T_{p2} \times P_{p3} / (1 - P_{o2})$$

and the expected total ongoing after the first, second and third progressions, T_{o1} , T_{o2} , and T_{o3} respectively, in

a given period, corresponding to one newly diagnosed patient per period, as:

$$\begin{split} T_{o1} &= T_p \times P_{o1} \left(1 + P_{o1} + P_{o1}^2 + ... \right) = \\ T_p \times P_{o1} / \left(1 - P_{o1} \right), \\ T_{o2} &= T_{p2} \times P_{o2} \left(1 + P_{o2} + P_{o2}^2 + ... \right) = \\ T_{p2} \times P_{o2} / \left(1 - P_{o2} \right) \end{split}$$

and

$$T_{o3} = T_{p3} \times P_{o3} (1 + P_{o3} + P_{o3}^{2} + ...) =$$

$$T_{p3} \times P_{o3} / (1 - P_{o3})$$

3.1 Computing the Fourth or Higher Progression Patients in a Period

Using similar derivations $T_{p4} = T_{p3} \times P_{p4}/(1 - P_{o3})$ for the newly diagnosed patient sequence, the calculator presumes the same next PFS and subsequent OS medians in the salvage setting (after second relapse). This leads to $P_{pk} = P_{p4}$ for $k \ge 4$ and $P_{om} = P_{o3}$ for $m \ge 3$. Using this we have:

$$\begin{split} T_{p5} &= T_{p4} \times P_{p5} / (1 - P_{o4}) = T_{p4} \times P_{p4} / (1 - P_{o3}) = \\ T_{p3} \times \{P_{p4} / (1 - P_{o3})\} \times \{P_{p4} / (1 - P_{o3})\} = T_{p3} \times (P_{pc})^{A}2, \\ \text{with } P_{pc} &= \{P_{p4} / (1 - P_{o3})\}. \text{ Similarly} \end{split}$$

$$T_{p6} = T_{p3} \times (P_{pc})^{A3}$$

and the pattern continues leading to fourth or higher progression rates T_{p4+} corresponding to the newly diagnosed patient sequence as:

$$T_{p4+} = T_{p3} \times P_{pc} \times ((1 + P_{pc} + P_{pc}^{2} + ...)) =$$
$$T_{p3} \times P_{pc} / (1 - P_{pc})$$

3.2 Unique Relapsing Patients

The expressions for progressions in a period derived earlier in Section 3 and in Section 3.1 look at the number of progressions in a period and not necessarily the number of unique patients having progressions in a period. For long enrollment periods the same patient may have one or more progressions in the same period and the results that follow adjust for these. For the k^{th} progression, the probability P_{k+1} that the next progression occurs in the same period is given by:

$$P_{k+1} = (1/L) J(exp[-h_{d(k+1)} \times a])$$

$\exp[-h_{pfs(k+1)} \times a])da$

where the integral operates over values of going from 0

to L and the hazard subscripts (k+1) indicate hazards for the PFS and OS events after the k^{th} progression. The integral evaluates to the following manner:

$$P_{k+1} = (1/L) \{ \exp[-h_{pfs(k+1)} \times a] / h_{pfs(k+1)} \\ \exp[-h_{d(k+1)} \times a] / h_{d(k+1)} \}$$

When the expression inside the braces, evaluating at a = 0 is subtracted from that evaluated at a = L, we get:

$$= (1/L)\{(1 - \exp[-h_{d(k+1)} \times L]) / h_{d(k+1)} -$$

$$(1 - \exp[-h_{pfs(k+1)} \times L]) / h_{pfs(k+1)}$$

Using the relationship between the hazard h and the Median M for the exponential distribution of $h = -\ln(0.5)/M$, the expression further evaluates to:

$$= (1/L) \{ -(M_{d(k+1)} \times P_{d(k+1)}) / \ln(0.5) + (M_{pfs(k+1)} \times P_{pfs(k+1)}) / \ln(0.5) \}$$

The median times to death and to subsequent relapse or death, and the probabilities of events have subscripts (k+1) indicative of times and probabilities after the k^{th} progression. Note that the number of $(k+1)^{\text{th}}$ relapses in a period is given by $T_{p(k+1)}$ unconditional on the occurrence of the k^{th} relapse. Further conditional on the occurrence of the k^{th} relapse, the number of $(k+1)^{\text{th}}$ relapses in a period is given by $T_{pk} \times P_{k+1}$. Hence the number of $(k+1)^{\text{th}}$ relapses in a period that is conditional on the non-occurrence of the k^{th} relapse in the same period is given by $T_{p(k+1)} - T_{pk} \times P_{k+1}$. This is the number of unique $(k+1)^{\text{th}}$ relapses $U_{p(k+1)}$ in the period. No prior relapse occurred in the period. Hence for k > 1 we have:

$$U_{pk} = T_{pk} - T_{p(k-1)} \times P_k$$

Using the assumption of a constant hazard of a subsequent progression in the salvage set for the third and subsequent relapse, it follows that $P_k = P_3$ for all $k \ge 4$. Hence the fourth or higher unique progression rate U_{p4+} corresponding to the newly diagnosed patient sequence is given as the sum for $k \ge 4$, by:

$$\begin{split} U_{p4+} &= \sum [T_{pk} \text{ - } T_{p(k-1)} \times P_k] \\ &= T_{p4+} \text{ - } \sum \left[P_3 \times T_{p(k-1)} \right] = T_{p4+} \text{ - } P_3 \times (T_{p3+} T_{p4+}) \end{split}$$

3.3 Patients Ongoing after Fourth or Higher Progression in a Period

We note here that on proceeding in an analogous

manner $T_{o4} = T_{p4} \times P_{oc}$ with $P_{oc} = \{P_{o3} / (1 - P_{o3})\}$ and $P_{ok} = P_{o3}$ for $k \ge 3$ per the salvage setting constant hazard assumption. Further,

 $T_{o5} = T_{p4} \times P_{pc} \times P_{oc}, \text{ and } T_{o6} = T_{p4} \times (P_{pc})^{^{\wedge 2}} \times P_{oc}$

This pattern continues leading to ongoing rates after fourth or higher progression T_{04+} , corresponding to the newly diagnosed patient sequence, given by:

$$T_{o4+} = T_{p4} \times [1 + P_{pc} + (P_{pc})^{2} + (P_{pc})^{3} + \dots] \times P_{oc} = T_{p4} \times P_{oc} / (1 - P_{pc})$$

3.4 Calculations of Subgroup Proportions

The calculations above provide the occurrence rates within relapse and ongoing groups corresponding to one newly diagnosed patient at the start of each period. This provides the total occurrence rate as a multiple of the unit newly diagnosed rate. This total is then used to provide rough estimates of the proportion of patients in each subgroup on setting the total to a 100%. Note that through arguments made earlier and the uniform disease incidence assumption, newly diagnosed patients will accrue uniformly through the enrollment time-period considered and the multiples and the subgroup proportions will be invariant to the starting time of the intervals. The last column in the calculator estimates the trial eligible fractions based on the total of the newly diagnosed and the relapse groups in each period. For disease registries which take all comers, irrespective of therapy use and stage of the disease, and for electronic health record datasets, the subgroup proportions provide likely subgroup sizes for analyses carve-outs for exploratory analyses given the total number of subjects.

4. Discussion

This calculator uses the median times to next PFS and subsequent OS measured from diagnosis and after relapses/progressions under exponential distribution assumptions. These medians correspond to patients on a mixed set of standard of therapy options and should typically be assessed by observational studies allowing for a latitude of patient consented clinician choices of therapy. The disease prognosis is presumed stable, with the assessed medians holding true over a medium time frame. This would not hold if there is a breakthrough therapy in that time frame with marked efficacy compared to previous standards, having considerable market penetration. Further the calculations assume a uniform incidence rate of the disease over time. The expressions provided allow analysts to rework the computations for the Time to Progression endpoint (TTP) instead of the PFS endpoint, and to relax our assumption of a constant high hazard of PFS and OS events in the salvage setting. Usually additional inclusion/exclusion criteria such as co-morbidity exclusions, requirement of certain biomarkers, and specified risk profiles apply, as well as competing initiating trials, and the starting estimates computed here would need to be further fractionated by the subgroup proportions meeting these additional constraints.

References

 Castellino, A., Chiappella, A., LaPlant, B. R., et al. 2018.
"Lenalidomide plus R-CHOP21 in Newly Diagnosed Diffuse Large B-Cell Lymphoma (DLBCL): Long-Term Follow-Up Results from a Combined Analysis from Two Phase 2 Trials." *Blood Cancer J.* 8 (11): 108. doi: 10.1038/s41408-018-0145-9D.

- [2] O'Brien, S., Furman, R. R., Coutre, S., et al. 2018. "Single-Agent Ibrutinib in Treatment-Naive and Relapsed/Refractory Chronic Lymphocytic Leukemia: A 5-Year Experience." *Blood* 131: 1910-9.
- [3] Burgess, L. J., Burgess, L. J., and Sulzer, N. U. 2011. "Examining the Clinical Trial Feasibility Process and Its Implications for a Trial Site." *Journal of Clinical Trials* 3: 51-4. doi:10.2147/OAJCT.S23631.
- [4] Jagannath, S., Rifkin, R. M., Gasparetto, C., et al. 2018. "Development of a Predictive Model of Multiple Myeloma (MM) Patient Outcomes Based on Treatment Sequencing Using Data from The Connect® MM Patient Registry." *Blood* 132: 3232.
- [5] Calculator at https://resourcetepee.com/free-statistical-calculators/trialenrollment/estimating-clinical-trial-enrollment-eligible-su bsets-in-oncology/.
- [6] Srinivasan, S. S. "Details on the Estimation of Cancer Progression Based Enrollment Subgroups." *Technical Report on Research Gate*, doi: 10.13140/RG.2.2.33118.48963.
- [7] Collett, D. 2003. Modelling Survival Data in Medical Research, 2nd ed. Boca Raton, Florida: Chapman and Hall/CRC.
- [8] Kline, M. 1976. *CALCULUS: An Intuitive and Physical Approach*. Mineola, New York: Dover Publications.