

Statistical and Trial Design Considerations in Central Nervous System Prophylaxis Studies

Jennifer A. Smith^a and Michael Glantz^b

The prognosis is very poor for patients with cancer who develop central nervous system (CNS) involvement. Prophylactic therapy may be a viable strategy for patients at high risk of CNS metastases. This article explores the rationale, feasibility, and ethics of prophylactic therapy. We discuss various study design considerations for CNS prophylaxis trials, with particular focus on statistical issues, and provide guidance to clinicians trying to decide how to investigate prophylactic therapy. The pool of patients eligible for inclusion in clinical trials is limited. To answer pressing clinical questions, innovative trial designs are needed, along with operational strategies that include risk factor enrichment, target-based end point selection and validation, flexible study accrual and monitoring, and techniques that permit early termination when initial treatment outcomes are poor. Two proposed trials for patients with glioblastoma multiforme and non-Hodgkin lymphoma are presented to highlight design considerations.

Semin Oncol 36 (Suppl 2):S69-S76. © 2009 Published by Elsevier Inc.

Despite progress in controlling systemic cancer, the prognosis for individuals affected by meningeal and other central nervous system (CNS) metastases is poor; affected patients rarely survive for more than a few months. In addition, the spread of cancer cells to the CNS, and cancer cell growth in the CNS, can cause diverse and debilitating neurologic complications, including paralysis, loss of vision, disturbances of gait and speech, altered mental status, and impaired cognitive ability, which severely compromise quality of life (QoL).

A number of factors suggest an increased risk of CNS relapse in patients with non-Hodgkin lymphoma (NHL)¹⁻¹² and a variety of solid tumors.¹³⁻¹⁸ Prophylactic therapy may be a viable treatment alternative for patients at high risk for CNS metastases.¹⁹ However, the role of intra-cerebrospinal fluid (CSF) prophylaxis remains controversial, and questions persist regarding the effectiveness of available prophylactic interventions. This is largely due to the lack of well-powered and properly controlled trials. The small number of

patients available for intra-CSF prophylaxis studies presents a challenge and has important implications for trial design. Moreover, selecting patients who will most likely benefit from prophylactic therapy with the least risk of toxicity may not be straightforward.

This review highlights issues of trial design in patients at risk for developing CNS disease, beginning with end point selection and measures of clinical benefit. Further topics of discussion include patient selection criteria, issues surrounding randomization, blinding, stratification, and sample size. The review concludes with a discussion of several design strategies that may be used in future intra-CSF prophylaxis trials.

END POINTS FOR CNS DISEASE

The primary end points selected to assess clinical outcome should balance the competing goals of patient benefit and scientific rigor sufficient to satisfy the demands of institutional review boards and regulatory agencies.²⁰

Time to Neurologic Progression (clinical)

Time to neurologic progression, which censors patients at the time of death, and progression-free survival (PFS), which includes both progression and death as events, represent end points that attempt to combine time and quality of survival in one measure (Table 1). Traditionally, these outcome measures do not account for treatment-related neurotoxicities, which would be necessary to evaluate the net benefit of treatment. Meticulous prespecification of the clinical definition of

^aGeron Corp, Menlo Park, CA.

^bDepartment of Neurosurgery, Penn State Hershey Medical Center, Hershey, PA.

STATEMENT OF CONFLICT OF INTEREST: M.G. has received honoraria and research support from Enzon Pharmaceuticals. J.A.S. has nothing to disclose.

Address correspondence to Jennifer A. Smith, PhD, Geron Corp, 230 Constitution Dr, Menlo Park, CA 94025. E-mail: jsmith@geron.com 0270-9295/09/\$ - see front matter

© 2009 Published by Elsevier Inc.

doi:10.1053/j.seminoncol.2009.05.008

Table 1. Traditional End Points for Neurooncology Clinical Trials

End Point	Confounders and Concerns
Tumor response	Lack of correlation with QoL
Survival	Test imprecision Variation in subsequent care Death from non-central nervous system cancer
Cause of death	Ignores quality of survival Variation in subsequent care Lack of reliable criteria Poor correlation with other end points
Surrogate markers	Restricted availability Lack of correlation with QoL
Time to tumor progression (radiographic or cytologic)	Lack of correlation with QoL Test imprecision
Time to neurologic progression	Usually unblinded Test imprecision Cumbersome

Abbreviation: QoL, quality of life.

neurologic progression is essential. Once clearly defined, the assessment of neurologic progression must be blinded, since significant bias is likely when the individual making outcome determinations is not masked to the treatment assignment. Assessments performed by a designated outside physician (someone other than the treating physician) may help overcome this potential problem. Data review and end point determination by an independent expert committee may be another alternative.^{21,22}

The frequency of assessment of neurologic progression, and the interval between repeated tests, are critical as well. Assessment frequency is in large part determined by the size of the expected effect; if treatment is expected to delay progression by only a few weeks, the assessment must be more frequent than if a longer delay is expected. This desire for frequent testing is balanced against the increased cost and feasibility of testing. Both increased frequency of testing and increased sample size increase the power of the study. Generally, the more frequently neurologic progression is assessed, the fewer the number of patients required to complete the trial. To avoid the introduction of bias,

assessment frequency must be equal between comparison groups.

QoL and Neurocognitive End Points

Direct measures of clinical benefit include QoL and neurocognitive end points. These measurements allow the clinician to evaluate the trade-off between clinical benefit and treatment-related toxicity (both neurologic and non-neurologic) and assess the relative impact of treatment on the patient's QoL.

QoL

QoL end points include strategies such as Quality-adjusted Time Without Symptoms of disease or Toxicity of treatment (Q-TWiST)²³ and the Independent Living Score (ILS).²⁴ Q-TWiST provides a single measure that integrates quality and quantity of life. This is a formal mathematical technique that allows the clinician to assess the amount of time a patient experiences treatment-related toxicity, the proportion of the total survival time during which a patient is free of symptoms related to either treatment or disease progression, and the amount of time during which symptoms are present because of disease progression.²⁵ The rationale for the ILS is the understanding that increasing impairment associated with cancer progression decreases independent functioning. Therefore, a simple way to measure QoL may be to quantify the amount of time a patient remains independent in his or her activities. While reductionist, and a simplification of the concept of QoL, this measurement approach makes the reasonable assumption that one critical global variable (independence) measured over time will reflect other aspects of a patient's QoL.²⁴

Neurocognitive Function

Brain tumors may profoundly affect cognitive function (eg, memory, decision-making), as may the treatment. Neurocognitive performance may thus reflect the net benefit of treatment over toxicity. Many instruments measure neurocognitive function (Table 2).²⁶⁻³³ However, given the multifactorial causes of cognitive deficits in patients with CNS metastases, accurate assessment may present a considerable challenge.^{26,34} No single, simple screening tool reliably identifies the degree and course of impairment and also defines the contributions of multiple etiologies. Some tests measure discrete processes such as expressive language function, while others measure more distributed, non-specific, functions throughout the brain.³⁴ Because neurocognitive deficits may limit a patient's ability to complete the tests, the reason for non-completion also must be captured, thereby increasing the power of this end point. The utility of most tests depends on the psychometric properties of the measure, the nature of the populations and treatment under study, and the

Table 2. Instruments Available to Assess Neurocognitive Function

Test	Characteristic Measured	Time to Administer (min)
Hopkins Verbal Learning Test ²⁷	Verbal memory	5
Trail Making Test Part A ²⁸	Visual-motor speed	2
Trail Making Test Part B ²⁸	Executive function	5
Controlled Oral Word Association ³¹	Verbal fluency	5
Digit Span Test ³²	Attention and concentration	10
Folstein MMSE ³³	Global cognitive function	10
Grooved Pegboard ²⁹ (non-dominant and dominant hands)	Motor dexterity and speed	2-3
Categoric Word Fluency ³⁰	Executive function	5

Abbreviation: MMSE, Mini-Mental State Examination.

effect on neurocognitive function of those treatments and of the disease itself.

Few studies have prospectively followed neurocognitive function among patients treated in randomized clinical trials who are suffering from CNS metastases.^{35,36} Corn et al used the Folstein Mini-Mental State Examination (MMSE), a simple and widely used tool to assess dementia.³⁵ Mehta et al used a sophisticated array of neurocognitive tests, including the Hopkins Verbal Learning tests, Controlled Oral Word Association, Grooved Pegboard Test, and TrailMaking tests A and B.³⁶ Performance on neurocognitive testing also depends, in part, on the patient's compliance with treatment and testing. Compliance observed in the Mehta study was high (94%-96%), and almost all patients in the Corn study completed at least one MMSE. These studies support the feasibility of formally evaluating neurocognitive outcome among patients treated for brain metastases. Although many of these tests are relatively simple to administer, a high degree of professional expertise is required for their accurate interpretation.^{26,34}

Tumor Response

Direct measures of CNS tumor burden include magnetic resonance imaging (MRI) and CSF cytology. Conventional cytology is not quantitative³⁷ and like conventional histology is subject to large inter- and intra-observer variability. As discussed elsewhere in this supplement, the reliability of CSF cytology is also very dependent on the volume of fluid submitted for analysis, the speed of sample processing, the number of samples obtained, and the site (lumbar *v* ventricular) from which the CSF is sampled. CSF cytology and MRI analysis may be combined; however, end point complexity becomes a serious concern with multimodal combinatorial assessments. Quantitative and semi-quantitative techniques are available for MRI data, but a strong subjective component remains.³⁸ In addition,

the increasingly widespread use of vascular targeted agents such as bevacizumab have rendered traditional measurement strategies, which rely on tumor dimensions on gadolinium-enhanced, T₁-weighted images, obsolete.³⁹⁻⁴¹

Survival

Survival can be used as a measure of clinical benefit. However, the utility of this end point depends on the particular study and the clinical context. For example, survival may not be the most sensitive end point when there are competing risks of death from the systemic and CNS cancers, and the investigational treatment is directed only at the CNS disease component (eg, treatment of neoplastic meningitis or brain metastases in the setting of active systemic cancer). In this case, survival as a measure of treatment effect may be "diluted" by non-CNS-related deaths, and CNS-related mortality may be more useful. Survival or PFS, where both death and progression are considered events, might be more sensitive end points in trials where the mechanism for overall disease progression depends on the status of the CNS component (eg, the glioblastoma multiforme [GBM] study discussed below) or where the CNS acts as a reservoir to "resupply" the brain (eg, GBM) or systemic disease (eg, acute lymphocytic leukemia), or where the treatment itself may hasten death (eg, bone marrow transplant). In these latter situations, all-cause mortality, time to (any) disease progression, or PFS may be valid end points.

TRIAL DESIGN CONSIDERATIONS

An effective trial design calls for adequate numbers of prespecified primary end point events. An adequate number of events can be achieved by enrolling a large enough number of patients, enrolling a smaller number of selected patients more likely to achieve the specified event, liberalizing the definition of the primary out-

come measure, or by extending the duration of follow-up. Each of these strategies has its own benefits and drawbacks. Accruing large numbers of patients and prolonging the length of the trial increase expenses and make administration of the trial more complex. Selecting an at-risk population more likely to reach a study end point is more economical but may slow down enrollment and limit the ability to generalize the results. One important advantage of using an at-risk population in the context of a prophylaxis trial is a reduction in the number of patients who receive treatment unnecessarily, ie, patients who would not have suffered the outcome of interest even without the study intervention. Another way of looking at this trade-off are the concepts of “number needed to treat” (NNT) and “number needed to harm” (NNH), concise expressions of the number of patients who would need to undergo the proposed intervention in order for one patient to benefit from or be harmed by the intervention. An example of these concepts as they apply to a trial of CSF prophylaxis in patients with NHL may be found in the article by Herrlinger et al in this supplement.

In a large-scale retrospective study, Hollender et al sought to determine the incidence and risk factors for CNS relapse in patients with NHL.⁴ A primary objective of this study was to establish risk assessment guidelines to help clinicians identify patient subpopulations that might be ideal targets for CNS prophylaxis. The probability of CNS recurrence within 5 years was calculated as a function of the number of risk factors present. CNS prophylaxis was recommended for patients with high-grade NHL and at least four risk factors (see Table 5 in the article by Herrlinger et al in this supplement).

Risk models such as this may indeed help clinicians identify high-risk patients and decide whether to implement CNS prophylactic therapy. From a clinical trials perspective, the advantage of using such a model to guide patient enrollment is the greater power to detect a treatment effect in a smaller sample size. As noted, however, the ability to generalize study results and a reduced pool of study-eligible patients (and hence slower patient accrual) are potential trade-offs.

Randomizing small patient numbers may lead to other methodologic challenges. For example, the desire to stratify patients (by study site or other variables) using the usual blocked lists for randomization may become too predictable. One possible way to avoid this predictability is to use a centralized adaptive randomization that achieves balance across stratification variables and across sites without using the usual prespecified blocked lists. With this approach, each subsequent patient randomization depends on the characteristics of the preceding randomization decision, and in this way, allocation concealment is preserved.

Because the power of a study (for time to event end points) is driven by the number of events, not the

entire sample size, the usual strategy for powering a trial assumes a targeted event rate and calculates the number of patients to achieve the desired number of events. There are four basic methods for preserving power in a given trial: (1) increase the number of observed events either by extending the follow-up period, increasing the frequency of testing, or redefining the end point; (2) increase the hypothesized effect size (a strategy that is only acceptable if the investigators believe the revised effect size is valid); (3) increase alpha; or (4) use a one- versus two-tailed test. (In some contexts, a two-tailed test is not necessary; for example, the traditional Simon Two-Stage Design is a one-tailed test.)

Commonly, enrollment is based on a fixed sample size that was generated by assuming a targeted event rate. Because the power of a trial is critically dependent on this potentially uncertain rate, a safer approach may be to base enrollment on the estimated number of events, not on sample size (ie, continue enrollment until a predefined number of events is achieved). This approach maintains statistical power but comes with the trade-off of an uncertain sample size that cannot be specified at study outset. Furthermore, this approach requires a relatively short time between the occurrence of the event and collection of pertinent data by the study personnel.

Simon Two-Stage Designs

This methodology describes a one-armed phase II design that requires a relatively small sample size.⁴² Two-stage designs are widely used in oncology trials to reduce the number of patients placed on therapies that might be ineffective. This design provides the option to stop the trial early when clear-cut termination rules are met. In the conventional two-stage design, the number of patients treated in the two stages is fixed. Once the target number of patients is enrolled in stage one, response is assessed using an interim analysis, and the study proceeds to the second stage, enrolling the remainder of patients if a specified number of responses are achieved in the first stage. In the event that the first stage fails to achieve the prespecified number of responses, the trial is terminated.

A large number of two-stage designs (with varying sizes of stages one and two) can be generated that correspond to specified values of the null response rate (treatment has no effect), specific alternative response rate of interest (corresponding to a treatment effect), type I error rate, and power. Then a particular design can be selected from the large pool based on additional specified criteria. For example, Simon originally proposed selection criteria that generated what he termed the optimal and the minimax designs. The optimal design is associated with the smallest expected sample size assuming the treatment has no effect. The minimax

Table 3. Glioblastoma Phase III Study Design (survival end point, 4-year accrual, 18-month follow-up; progression-free survival end point, 4-year accrual, 12-month follow-up)*

14.6 Versus 19.5 Months Median Survival (HR, 0.75)

	Sample Size Per Arm	
	85% Power	90% Power
Alpha = 0.05	275	321
Alpha = 0.10	220	262

6.9 Versus 9.2 Months Median PFS (HR, 0.75)

	Sample Size Per Arm	
	85% Power	90% Power
Alpha = 0.05	236	276
Alpha = 0.10	189	225
Alpha = 0.20	142	173

Abbreviations: PFS, progression-free survival; HR, hazard ratio.
*Assumes log-rank tests for all comparisons.

design is associated with the smallest overall sample size. The total N for this design can be significantly smaller than the total N for the optimal design; however, there is generally a larger stage one sample size. This larger stage one sample size might be preferable for heterogeneous populations, as a very small stage one sample size may not be representative of the entire eligible population.

There are several potential problems with any two-stage design. Investigators may fail to identify a subgroup that could potentially benefit from the intervention if a trial with small numbers is prematurely terminated. In addition, the purpose of the trial might be to gather relevant safety (or other end point) information, and this might not be sufficiently accomplished if a trial is stopped early on the basis of response rate data. Similarly, investigators may wish to continue the trial even though the treatment does not meet the prespecified response criteria, because they believe even the more modest observed effect is clinically important.

PROPOSED GBM PHASE III STUDY

The phase III GBM trial proposed elsewhere in this supplement considers two possible end points: survival (4-year accrual, 18-month follow-up) and PFS (4-year accrual, 12-month follow-up) (see the article by Glantz et al in this supplement). This trial assumes a median survival of 14.6 versus 19.5 months and a median PFS of 6.9 versus 9.2 months in the standard versus investigational treatment arms, both with a hazard ratio (HR) of

0.75. Survival and PFS estimates for the standard treatment arm are based on published results from a well-designed phase III trial.⁴³ Survival and PFS targets for the investigational arm are based on the investigators' judgment about what would constitute a clinically meaningful response advantage. Table 3 provides the number of patients required to detect these differences for various values of alpha and power levels. Of note, although the medians differ for the two end points (survival *v* PFS), the proposed treatment effect (HR) is the same. The follow-up time and the number of patients needed to achieve this effect size are smaller when PFS is chosen as the end point, because the PFS end point results in an earlier and a greater number of events.

Alternatively, dramatically smaller sample sizes and time required for patient accrual could be achieved with a single-arm trial using a PFS end point. However, the lack of blinding in these types of single-arm trials has the potential to introduce substantial observer bias, and the lack of randomization requires faith that the historical control population is very comparable to the study population and that the historical outcome data are reliable.

PROPOSED NHL PROPHYLAXIS STUDY

The phase III NHL prophylaxis study proposed by Herrlinger et al elsewhere in this supplement also would consider two end points: time to progression (2 years accrual, 2 or 3 years follow-up) and PFS (2 years

Table 4. Non-Hodgkin Lymphoma Phase III Prophylaxis Study Design (time to progression end point, 2-year accrual, alpha = 0.05; progression-free survival end point, 2-year accrual, alpha = 0.05)*

TTP: Two or More Risk Factors: 2-Year Event Rate 10% *v* 4%

	Sample Size Per Arm	
	85% Power	90% Power
2-Year follow-up	236	276
3-Year follow-up	180	210

PFS: 60% *v* 68.2% Event-Free at 2 Years (HR, 0.75)

	Sample Size Per Arm	
	85% Power	90% Power
2-Year follow-up	460	538
3-Year follow-up	379	443

Abbreviations: TTP, time to progression; PFS, profession-free survival; HR, hazard ratio.
*Assumes log rank tests for all comparisons.

Table 5. Sample Size Calculations for a Proposed non-Hodgkin Lymphoma Prophylaxis Study Using Various Designs (primary outcome variable is defined as relapse in the central nervous system)

Randomized Controlled Phase III Design (end point: relapse rate)*				
Relapse Rates		Sample Size (per arm)		
Treatment 1 (L-AraC)	Treatment 2 (MTX)	80% Power	90% Power	
0.01	0.07	180	225	
0.02	0.06	400	520	
0.03	0.05	1,600	2,100	
Single-Arm Phase II Model†				
Relapse Rates		Sample Size		
Treatment (L-AraC or MTX)	Historical High-Risk Population	80% Power	90% Power	
0.01	0.15	35	35	
0.02	0.15	35	46	
0.03	0.15	46	56	
0.05	0.15	75	102	
Randomized Phase II "Pick the Winner" Design‡				
Relapse Rates		Sample Size (per arm)		
Treatment (L-AraC or MTX)	Historical High-Risk Population	80% Power	90% Power	
0.01	0.15	35	35	
0.02	0.15	35	46	
0.03	0.15	46	56	
0.05	0.15	75	102	
0.06	0.15	102	127	
0.07	0.15	127	160	

NOTE. All calculations assume a two-sided alpha of 0.05.

Abbreviations: L-AraC, liposomal cytarabine; MTX, methotrexate; NHL, non-Hodgkin lymphoma.

*Experimental treatment reduces the relapse rate to 1% to 3%, while the relapse rate varied between 5% and 7% in patients treated with MTX. Test: Fisher's exact test.

†Test: one-sample exact test of a binomial proportion.

‡Test: Two separate one-sample exact tests of a binomial proportion. Power is for each test independently.

accrual, 2 or 3 years follow-up) (Table 4). Required risk factors for study eligibility would be elevated lactate dehydrogenase and either involvement of more than one extranodal site or a specific site of disease associated with a high risk of CNS relapse. The sample size calculations would assume a two-sided alpha of 0.05, both risk factors present, and a time to progression event rate of 10% versus 4% and PFS of 60% versus 68.2% at 2 years (Table 4). The authors also would consider the benefit-to-risk trade-off by comparing the NNT to prevent one case of CNS relapse with the NNH. Assuming a 10% CNS relapse rate, the NNT would be 13.6 and the NNH 3.5, with an odds ratio in favor of prophylaxis of 3.9 (95% confidence interval, 1.4-11.0). Using the risk factors and assumptions outlined in their proposal, these ranges fall within what is usually considered acceptable in other disease models.

Once again, the length of time patients are followed (2 v 3 years) has a significant impact on the number of

patients required to complete the study, since a longer follow-up period translates into a higher number of events. The desired power (85% v 90%) also influences study size and reliability. A lower power conveys a higher probability of a type II error. For example, an 80% power incurs a 20% chance of a type II error. In earlier drug development settings (phase II), adequate control of the type II (false-negative) error may be extremely important, because a false-negative result at this stage could prematurely terminate the development of a promising agent. In late stages of development, the burden shifts to avoiding a false-positive result (a type I error) to avoid the widespread use of a drug that is ineffective.⁴⁴

Herrlinger et al also offer a randomized phase II "pick the winner" trial design.⁴⁵ In this design, the two treatment arms would each be compared independently to a prespecified historical control but not to each other. This design may be practically superior,

since study size would be unreasonably large if a phase III trial design was used, and successful accrual to a phase III trial might be very difficult to achieve in this patient population. Table 5 provides sample size estimates for various response assumptions and power choices for this randomized phase II “pick the winner” design, a randomized phase III design with the end point of relapse rate, and a single-arm phase II design with an historical control comparison. These calculations illustrate the substantial variation in sample size dictated by choice of study design and by other factors such as estimates of effect size and choice of power.

CONCLUSION

Definitive evidence for the effectiveness of CNS prophylaxis is lacking except in small cell lung cancer,^{46,47} acute leukemia, and very aggressive lymphomas (ie, Burkitt lymphoma and lymphoblastic lymphoma). In other patient cohorts at high risk for CNS relapse, prophylactic therapy may represent an important treatment option; however, the patient pool for clinical trial participation is relatively small. A small number of patients magnifies the challenges of effective trial design. Researchers involved in trial design should be particularly mindful of the following issues that could be detrimental to a successful study. Do not manipulate the effect size simply to achieve a manageable estimate of sample size. This will likely result in too few patients to arrive at a meaningful conclusion at the end of the study. Consider centralized randomization to avoid predictable treatment allocation that could bias trial results. Consider both the risks (slower accrual, possibly decreased generalizability) and the benefits (smaller study size, shorter duration, fewer patients exposed to the intervention in prophylactic trials) of limiting the study to an at-risk population. Avoiding type II error is perhaps the most important goal of early-phase trials; therefore, consider a one-tailed test or a two-stage test that does not have higher-than-desired probability of early termination. Do not modify an end point simply to increase the number of events (or power) at the expense of using the most clinically relevant or treatment-responsive end point. Blinded outcome assessment, randomization, and stratification for prognostically important variables are also critically important considerations. More research directed at optimal trial design and outcome measures is needed for CNS prophylaxis trials.

REFERENCES

- Boehme V, Zeynalova S, Kloess M, Loeffler M, Kaiser U, Pfreundschuh M, et al. Incidence and risk factors of central nervous system recurrence in aggressive lymphoma—a survey of 1693 patients treated in protocols of the German High-Grade Non-Hodgkin's Lymphoma Study Group (DSHNHL). *Ann Oncol*. 2007;18:149-57.
- Sørli T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A*. 2001;98:10869-74.
- Haioun C, Besson C, Lepage E, Thieblemont C, Simon D, Rose C, et al. Incidence and risk factors of central nervous system relapse in histologically aggressive non-Hodgkin's lymphoma uniformly treated and receiving intrathecal central nervous system prophylaxis: a GELA study on 974 patients. *Groupe d'Etudes des Lymphomes de l'Adulte. Ann Oncol*. 2000;11:685-90.
- Hollender A, Kvaloy S, Nome O, Skovlund E, Lote K, Holte H. Central nervous system involvement following diagnosis of non-Hodgkin's lymphoma: a risk model. *Ann Oncol*. 2002;13:1099-107.
- van Besien K, Ha CS, Murphy S, McLaughlin P, Rodriguez A, Amin K, et al. Risk factors, treatment, and outcome of central nervous system recurrence in adults with intermediate-grade and immunoblastic lymphoma. *Blood*. 1998;91:1178-84.
- Zinzani PL, Magagnoli M, Frezza G, Prologo G, Gherlinzoni F, Bendandi M, et al. Isolated central nervous system relapse in aggressive non-Hodgkin's lymphoma: the Bologna experience. *Leuk Lymphoma*. 1999;32:571-6.
- Tomita N, Kodama F, Sakai R, Koharasawa H, Hattori M, Taguchi J, et al. Predictive factors for central nervous system involvement in non-Hodgkin's lymphoma: significance of very high serum LDH concentrations. *Leuk Lymphoma*. 2000;38:335-43.
- Herman TS, Hammond N, Jones SE, Butler JJ, Byrne GE Jr, McKelvey EM. Involvement of the central nervous system by non-Hodgkin's lymphoma. The Southwest Oncology Group experience. *Cancer*. 1979;43:390-7.
- Liang R, Chiu E, Loke SL. Secondary central nervous system involvement by non-Hodgkin's lymphoma: the risk factors. *Hematol Oncol*. 1990;8:141-5.
- Keldsen N, Michalski W, Bentzen SM, Hansen KB, Thorling K. Risk factors for central nervous system involvement in non-Hodgkin's lymphoma—a multivariate analysis. *Acta Oncol*. 1996;35:703-8.
- Litam JP, Cabanillas F, Smith TL, Bodey GP, Freireich EJ. Central nervous system relapse in malignant lymphomas: risk factors and implications for prophylaxis. *Blood*. 1979;54:1249-57.
- Ersboll J, Schultz HB, Thomsen BL, Keiding N, Nissen NI. Meningeal involvement in non-Hodgkin's lymphoma: symptoms, incidence, risk factors and treatment. *Scand J Haematol*. 1985;35:487-96.
- Kesari S, Batchelor TT. Leptomeningeal metastases. *Neurol Clin*. 2003;21:25-66.
- Bromberg JE, Breems DA, Kraan J, Bikker G, van der Holt B, Smitt PS, et al. CSF flow cytometry greatly improves diagnostic accuracy in CNS hematologic malignancies. *Neurology*. 2007;68:1674-9.
- Benesch M, Urban C. Liposomal cytarabine for leukemic and lymphomatous meningitis: recent developments. *Expert Opin Pharmacother* 2008;9:301-9.
- Ceresoli GL, Reni M, Chiesa G, Carretta A, Schipani S, Passoni P, et al. Brain metastases in locally advanced nonsmall cell lung carcinoma after multimodality treatment: risk factors analysis. *Cancer*. 2002;95:605-12.
- Evans AJ, James JJ, Cornford EJ, Chan SY, Burrell HC,

- Pinder SE, et al. Brain metastases from breast cancer: identification of a high-risk group. *Clin Oncol (R Coll Radiol)*. 2004;16:345-9.
18. Melisko ME, Glantz M, Rugo HS. New challenges and opportunities in the management of brain metastases in patients with ErbB2-positive metastatic breast cancer. *Nat Clin Pract Oncol*. 2009;6:25-33.
 19. Mazhar D, Stebbing J, Bower M. Non-Hodgkin's lymphoma and the CNS: prophylaxis and therapy in immunocompetent and HIV-positive individuals. *Expert Rev Anticancer Ther*. 2006;6:335-41.
 20. Thall PF. A review of phase 2-3 clinical trial designs. *Lifetime Data Anal*. 2008;14:37-53.
 21. Mehta MP, Shapiro WR, Glantz MJ, Patchell RA, Weitzner MA, Meyers CA, et al. Lead-in phase to randomized trial of motexafin gadolinium and whole-brain radiation for patients with brain metastases: centralized assessment of magnetic resonance imaging, neurocognitive, and neurologic end points. *J Clin Oncol*. 2002;20:3445-53.
 22. Meyers CA, Smith JA, Bezjak A, Mehta MP, Liebmann J, Illidge T, et al. Neurocognitive function and progression in patients with brain metastases treated with whole-brain radiation and motexafin gadolinium: results of a randomized phase III trial. *J Clin Oncol*. 2004;2:157-65.
 23. Gelber RD, Cole BF, Gelber S, Goldhirsch A. Comparing treatments using quality-adjusted survival: the Q-TwiST method. *Am Stat*. 1995;49:161-9.
 24. Recht L, Glantz M, Chamberlain M, Hsieh CC. Qualitative measurement of quality outcome in malignant glioma patients using an independent living score (ILS). *J Neurooncol*. 2003;61:127-36.
 25. Cole BF, Glantz MJ, Jaeckle K. Quality-of-life-adjusted survival comparison of sustained release cytosine arabinoside versus intrathecal methotrexate for treatment of solid tumor neoplastic meningitis. *Cancer*. 2003;97:3053-60.
 26. Glantz M, Conlee J. High grade gliomas. In: Meyers C, Perry A, editors. *Cognition and cancer*. Cambridge, UK: Cambridge University Press; 2008. p. 156-69.
 27. Benedict RHB, Schretlen D, Groninger L, Brandt J. Hopkins Verbal Learning Test-Revised: normative data and analysis of inter-form and test-retest reliability. *Clin Neuropsychol*. 1998;12:43-55.
 28. Lezak MD, Howleson DB, Loring DW. *Neuropsychological assessment*. New York: Oxford University Press; 1994.
 29. Lezak MD. *Neuropsychological assessment* (ed 3). New York: Academic Press; 1995.
 30. Benton AD. Differential behavioral effects in frontal lobe disease. *Neuropsychologia*. 1968;6:53-60.
 31. Benton AL, Hamsher K de S. *Multilingual Aphasia Examination*. Iowa City, IA: AJA Associates; 1989.
 32. Wechsler D. *Wechsler Adult Intelligence Scale-Revised*. San Antonio, TX: The Psychological Corp; 1981.
 33. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res*. 1975;12:189-98.
 34. Smith J, Wefel J. Neurocognitive testing in clinical trials. In: Meyers C, Perry A, editors. *Cognition and cancer*. Cambridge, UK: Cambridge University Press; 2008. p. 320-8.
 35. Corn BW, Moughan J, Knisely JPS, Fox SW, Chakravarti A, Yung WK, et al. Prospective evaluation of quality of life and neurocognitive effects in patients with multiple brain metastases receiving whole-brain radiotherapy with or without thalidomide on Radiation Therapy Oncology Group (RTOG) trial 0118. *Int J Radiat Oncol Biol Phys*. 2008;71:71-8.
 36. Mehta MP, Rodrigus P, Terhaard CHJ, Rao A, Suh J, Roa W, et al. Survival and neurologic outcomes in a randomized trial of motexafin gadolinium and whole-brain radiation therapy in brain metastases. *J Clin Oncol*. 2003;21:2529-36.
 37. Melisko ME, Kunwar S, Prados M, Berger MS, Park JW. Brain metastases of breast cancer. *Expert Rev Anticancer Ther*. 2005;5:253-68.
 38. Shah BD, Kesari S, Xu R, Batchelor TT, O'Neill AM, Hochberg FH, et al. Comparison of linear and volumetric criteria in assessing tumor response in adult high-grade gliomas. *Neurooncology*. 2006;8:38-46.
 39. Ananthnarayan S, Bahng J, Roring J, Nghiemphu P, Lai A, Cloughesy T, et al. Time course of imaging changes of GBM during extended bevacizumab treatment. *J Neurooncology*. 2008;88:339-47.
 40. Norden AD, Young GS, Setayesh K, Muzikansky A, Klufas R, Ross GL, et al. Bevacizumab for recurrent malignant gliomas: efficacy, toxicity, and patterns of recurrence. *Neurology*. 2008;70:779-87.
 41. de Groot JF, Yung WK. Bevacizumab and irinotecan in the treatment of recurrent malignant gliomas. *Cancer J*. 2008;14:279-85.
 42. Simon R. Optimal two-stage designs for phase II clinical trials. *Control Clin Trials*. 1989;10:1-10.
 43. Stupp R, Mason WP, van den Bent MJ, Weller M, Fisher B, Taphoorn MJ, et al. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N Engl J Med*. 2005;352:987-96.
 44. Lee JJ, Feng L. Randomized phase II designs in cancer clinical trials: current status and future directions. *J Clin Oncol*. 2005;23:4450-7.
 45. Simon R, Wittes RE, Ellenberg SS. Randomized phase II clinical trials. *Cancer Treat Rep*. 1985;69:1375-81.
 46. Aupérin A, Arriagada R, Pignon JP, Le Pécoux C, Gregor A, Stephens RJ, et al. Prophylactic cranial irradiation for patients with small-cell lung cancer in complete remission. *Prophylactic Cranial Irradiation Overview Collaborative Group*. *N Engl J Med*. 1999;341:476-84.
 47. Lee JJ, Bekele BN, Zhou X, Cantor SB, Komaki R, Lee JS. Decision analysis for prophylactic cranial irradiation for patients with small-cell lung cancer. *J Clin Oncol*. 2006;24:3597-603.