GCIG Harmonization Committee
Statistical Section

4:00pm -- 6.00pm, Thursday, May 29, 2014
Huron Room, Doubletree Hotel, Chicago

MINUTES

Chair: Jim Paul (james.paul@glasgow.ac.uk) – SGCTG (JP)
Co-Chair: Byung Ho Nam (byunghonam@ncc.rc.kr) – KGOG (BHN)

Present:
Andrew Embleton (a.embleton@ctu.mrc.ac.uk) – MRC/NCRI (AE)
Mark Brady (brady@gogstats.org) – NRG (MB)
Val Gebski (Val@ctc.usyd.edu.au) – ANZGOG (VG)
Alexander Reuss (Alexander.reuss@kks.uni-marburg.de) – AGO (AR)
Tetsutaro Hamano (hamato@insti.kitasato-u.ac.jp)– GOTIG (TH)
Wendy Fantl - COGi (WF)

1. **Welcome & Introductions (C.O.I. declaration)**
There were no COI declared.

2. **Statistical Issues in Scanning to Assess Progression (AR/BHN)**

BHN reviewed made a presentation reviewing the guidance produced by the FDA (Guidance for Industry Clinical Trail Endpoints the Approval of Cancer Drugs and Biologics) on assessing progression and AR made a presentation on some issues addressed in the literature on the same topic.

Copies of these presentations are attached.

Key elements are summarized below:-

- Frequency
  - Ideally same in both arms
  - Frequency can be half the median pfs in control arm without significant impact on power
  - Frequency can be different in different countries, as long as it still meets these first two criteria
- Central review
  - Literature/experience suggests this makes little difference
  - Not required for blinded studies our studies with large effects
  -
- Analysis has inherent problem that we don't know exact time of progression

- Analysis further complicated by:-
  - Scan missed because of site/patient lack of compliance
  - Patients being switched to other ant-cancer treatment before progression

        - Patients coming of study therapy early before progression

- Variety of ways of dealing with these (no single correct approach)
- Must ensure how data is handled is addressed in the SAP
    - Have to analyse the data in a number of ways (sensitivity analysis) to ensure conclusions are robust
    - Will produce guidance document on this

3. Finalisation of position paper on study designs for rare tumours (MB)

This is to be finalized for the next meeting.

4. Proposals for discussion topics at future meetings (All)

- The merits of PFS or OS as primary end-point
- Futility boundaries
- Response adaptive designs
- Issues around making data for GCIG trials routinely available for meta-analyses

The "Issues around making data for GCIG trials routinely available for meta-analyses" was selected at the topic for the next meeting.

5. AoB

There was no AOB.

Statistical Issues in Scanning to Access Progression

2104 Spring GCIG Harmonization/STAT

KGOG

Byung-Ho Nam

FDA Guidance

Guidance for Industry
Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)

May 2007
Clinical/Medical

FDA Guidance

- **The methodology for assessing, measuring, and analyzing PFS should be detailed in the protocol and statistical analysis plan (SAP)**
- **It is also important to carefully define tumor progression criteria in the protocol.**

- **There are no standard regulatory criteria for defining progression**
- **Applicants have used a variety of different criteria, including the RECIST criteria**
- **The broad outline presented in most published PFS criteria should be supplemented with additional details in the protocol and SAP**

FDA Guidance

- **Visits and radiological assessments should be symmetric between the two study arms to prevent systematic bias**
- **When possible, studies should be blinded. Blinding is particularly important when patient or investigator assessments are included as components of the progression endpoint**
- **At a minimum, the assessments should be subjected to a blinded independent adjudication team, generally consisting of radiologists and clinicians**

FDA Guidance

- The FDA and the applicant should agree prospectively on the following items:
  - The study design
  - The definition of progression
  - The data to be recorded on the case report form (CRF)
  - The SAP
  - The methodology for handling missing data and censoring methods
  - The operating procedures of an independent endpoint review committee (IRC), if applicable

FDA Guidance

Analysis of PFS

- The protocol should define an adequate assessment visit for each patient (i.e., a visit when all scheduled tumor assessments have been done)
- The analysis plan should outline a comparison of the adequacy of follow-up in each treatment arm
- Methodology for analyzing incomplete and/or missing follow-up visits and censoring methods should be specified in the protocol

FDA Guidance

Analysis of PFS

- **The analysis plan should specify the primary analysis and one or more sensitivity analyses to evaluate the robustness of the results**
- **Although any analyses with missing data can be problematic, the results can be strengthened by similar results in both the primary and the sensitivity analyses**
- **The evaluation should include the number of deaths in patients who have been lost to follow-up for a prolonged time period. An imbalance in such deaths could bias the PFS measurement by overestimating PFS in the treatment arm with less follow-up**

FDA Guidance

Analysis of PFS

- Because progression data can be collected from multiple sources (including physical exams at unscheduled visits and radiological scans of various types) and at different times, data collection for each assessment visit should be limited to a specified short time interval around the scheduled visit
- Difficulties can arise in determining the event date and censoring date when data are collected over a prolonged time period

FDA Guidance

Analysis of PFS

- We recommend assigning the progression date to the earliest time when any progression is observed without prior missing assessments and censoring at the date when the last radiological assessment determined a lack of progression
- Plans for PFS data collection and analysis should be discussed with the FDA at end-of-phase 2 meetings and verified in special protocol assessments

FDA Guidance

APPENDIX1
TUMOR MEASUREMENT DATA COLLECTION

- **The CRF and electronic data document the target lesions identified during the baseline visit before treatment. Retrospective identification of such lesions would not be considered reliable.**

- **Tumor lesions be assigned a unique identifying letter or number. This assignment provides differentiation among multiple tumors occurring at one anatomic site and the matching of tumors measured at baseline and tumors measured during follow-up.**

FDA Guidance

APPENDIX1
TUMOR MEASUREMENT DATA COLLECTION

- **A mechanism be in place that ensures complete data collection at critical times during follow-up.**
- **The CRF should ensure that all target lesions are assessed at baseline and that the same imaging or measuring method is used for all tests required at baseline and follow-up.**

- **The CRF contains data fields that indicate whether scans were performed at each visit.**

FDA Guidance

APPENDIX1
TUMOR MEASUREMENT DATA COLLECTION

- **A zero be recorded when a lesion has completely resolved. Otherwise, disappearance of a lesion cannot be differentiated from a missing value**

- **Follow-up tests provide for timely detection of new lesions both at initial and new sites of disease. The occurrence and location of new lesions should be recorded in the CRF and in the submitted electronic data**

FDA Guidance

APPENDIX2
ISSUES TO CONSIDER IN PFS ANALYSIS

- **The protocol and SAP should detail the primary analysis of PFS.**
- **This analysis should include a detailed description of the endpoint, appropriate modalities for evaluating tumors, and procedures for minimizing bias, such as procedures for an IRC**
- **One or two secondary analyses should be specified to evaluate anticipated problems in trial conduct and to assess whether results are robust**

FDA Guidance

- **Definition of progression date**

  - **In PFS analyses, the exact progression date is unknown. The following two methods can be used for defining the *recorded progression date (PDate)* used for PFS analysis**:

FDA Guidance

■ Definition of progression date

1. PDate assigned to the first time at which progression can be declared

− For progression based on a new lesion, *the PDate is the date of the first observation* that the new lesion was detected.

− If multiple assessments based on the sum of target lesion measurements are done at different times, *the PDate is the date of the last observation or radiological assessment of target lesions that shows a predefined increase in the sum of the target lesion measurements*

FDA Guidance

- **Definition of progression date**

2. PDate as the date of the protocol-scheduled clinic visit immediately after all radiological assessments (which collectively document progression) have been done

FDA Guidance

- **Definition of censoring date**

  - **Censoring dates are defined in patients with no documented progression before data cutoff or dropout. In these patients, *the censoring date is often defined as the last date on which progression status was adequately assessed.***
  - **One acceptable approach uses the date of the last assessment performed. However, multiple radiological tests can be evaluated in the determination of progression. *A second acceptable approach uses the date of the clinic visit corresponding to these radiological assessments*.**

FDA Guidance

■ **Definition of an adequate PFS evaluation**

- **In patients with no evidence of progression, censoring for PFS often relies on the date of the last *adequate tumor assessment***
- **A careful definition of what constitutes an adequate tumor assessment includes adequacy of target lesion assessments and adequacy of radiological tests both to evaluate nontarget lesions and to search for new lesions**

FDA Guidance

- **Analysis of partially missing tumor data**

  - **Analysis plans should describe the method for calculating progression status when data are partially missing from *adequate tumor assessment* visits**

FDA Guidance

■ Completely missing tumor data.

- Assessment visits where no data are collected are sometimes followed by death or by assessment visits showing progression.
- In other cases, the subsequent assessment shows no progression. In the latter case, it may seem appropriate to continue the treatment and continue monitoring for progression. However, this approach treats missing data differently depending upon subsequent events and can represent informative censoring.
- Another possible approach is to include data from subsequent PFS assessments. This can be appropriate when evaluations are frequent and when only a single follow-up visit is missed.

FDA Guidance

■ Completely missing tumor data.

• **Censoring at the last adequate tumor assessment can be more appropriate when there are two or more missed visits.** The SAP should detail primary and secondary PFS analyses to evaluate the potential effect of missing data.

• Reasons for dropouts should be incorporated into procedures for determining censoring and progression status. For instance, **for the primary analysis, patients going off-study for undocumented clinical progression, change of cancer treatment, or decreasing performance status can be censored at the last adequate tumor assessment.**

• **The secondary sensitivity analysis would include these dropouts as progression events.** Although missed visits for progression can be problematic, all efforts should be made to keep following patients for disease progression irrespective of the number of visits missed.

FDA Guidance

APPENDIX 3:
EXAMPLE TABLES FOR PFS ANALYSIS

- **Sensitivity analyses can be helpful in determining whether the PFS analysis is robust. However, these sensitivity analyses are exploratory and supportive of the results of the primary analysis, and efficacy may not be claimed based on sensitivity analysis alone.**
- **Different sensitivity analyses can be described in tables that specify how dates of progression events and dates for censoring of progression data can be assigned.**

FDA Guidance

Table A represents a sensitivity analysis that only includes well-documented and verifiable progression events. Other data are censored.

In Table A, the progression dates are:

- **Based only on radiological assessments verified by an IRC. *Clinical progression* is not considered a progression endpoint.**
- **Assigned to the first time when tumor progression was noted.**
- **The date of death when the patient is closely followed. However, deaths occurring after two or more missed visits are censored at the last visit.**

FDA Guidance

Table A. PFS 1 (includes documented progression only)

Situation	Date of Progression or Censoring	Outcome
No baseline tumor assessments	Randomization	Censored
Progression documented between scheduled visits	Earliest of: • Date of radiological assessment showing new lesion (if progression is based on new lesion); or • Date of last radiological assessment of measured lesions (if progression is based on increase in sum of measured lesions)	Progressed
No progression	Date of last radiological assessment of measured lesions	Censored
Treatment discontinuation for undocumented progression	Date of last radiological assessment of measured lesions	Censored
Treatment discontinuation for toxicity or other reason	Date of last radiological assessment of measured lesions	Censored
New anticancer treatment started	Date of last radiological assessment of measured lesions	Censored
Death before first PD assessment	Date of death	Progressed
Death between adequate assessment visits	Date of death	Progressed
Death or progression after more than one missed visit	Date of last radiological assessment of measured lesions	Censored

FDA Guidance

The sensitivity analysis in Table B corrects for potential bias in follow-up schedules for tumor assessment *by assigning the dates for censoring and events only at scheduled visit dates.* However, this approach can introduce bias if the progression occurred closer to the last visit, particularly in an open-label study. This approach can be suitable in blinded, randomized studies.

Table B. PFS 2 (uniform progression and assessment dates)

Situation	Date of Progression or Censoring	Outcome
No baseline tumor assessments	Randomization	Censored
Progression documented between scheduled visits	Date of next scheduled visit	Progressed
No progression	Date of last visit with adequate assessment	Censored
Treatment discontinuation for undocumented progression	Date of last visit with adequate assessment	Censored
Treatment discontinuation for toxicity or other reason	Date of last visit with adequate assessment	Censored
New anticancer treatment started	Date of last visit with adequate assessment	Censored
Death before first PD assessment	Date of death	Progressed
Death between adequate assessment visits	Date of death	Progressed
Death or progression after more than one missed visit	Date of last visit with adequate assessment	Censored

FDA Guidance

***The sensitivity analysis in Table C evaluates PFS according to the investigator's assessment*. However, this approach can introduce bias if the progression occurred closer to the last visit, particularly in an open-label study. This approach can be suitable in blinded, randomized studies.**

Table C. PFS 3 (includes investigator claims)

Situation	Date of Progression or Censoring	Outcome
No baseline assessment	Randomization	Censored
Progression documented between scheduled visits	Next scheduled visit	Progressed
No progression	Date of last visit with adequate assessment	Censored
Investigator claim of clinical progression	Scheduled visit (or next scheduled visit if between visits)	Progressed
Treatment discontinuation for toxicity or other reason	Date of last visit with adequate assessment	Censored
New anticancer treatment started with no claim of progression	Date of last visit with adequate assessment	Censored
Death before first PD assessment	Date of death	Progressed
Death between adequate assessment visits or after patient misses one assessment visit	Date of death	Progressed
Death after an extended lost-to-follow-up time (two or more missed assessments)	Last visit with adequate assessment	Censored

Statistical issues in scanning to assess progression – some views from the literature

GCIG Harmonization Committee

Statistical Section

Alexander Reuss

Chicago, 29 May 2014

Statistical issues in scanning to assess progression

Questions from our last meeting:

- Must it *[scanning frequency]* be the same across the whole study?
- How frequent does scanning have to be?
  - How does this impact on power?
  - How does it impact on the estimate of the HR?
- What if scans don't happen at scheduled times? Adjustment?
- Assessment for bias in treatment comparisons from differences in scan timing
- Need for central review?

Must it be the same across the whole study?

- Not quite sure how the question was meant.
- Answering the question: Can follow-up for progression be stopped before observation of PD (e.g. if treatment is discontinued due to tox or at start of a new anticancer treatment)?

Can follow-up stop before observation of PD?

MISSING DATA AND CENSORING IN THE ANALYSIS OF PROGRESSION-FREE SURVIVAL IN ONCOLOGY CLINICAL TRIALS

J. S. Denne[1], A. M. Stone[2], R. Bailey-Iacona[3], and T.-T. Chen[4]

[1]*Eli Lilly & Co., Indianapolis, Indiana, USA*
[2]*AstraZeneca, Macclesfield, United Kingdom*
[3]*AstraZeneca, Wilmington, Delaware, USA*
[4]*Bristol Myers Squibb, Wallingford, Connecticut, and Columbia University, New York, New York, USA*

Progression-free survival (PFS) is increasingly used as a primary endpoint in oncology clinical trials. However, trial conduct is often such that PFS data on some patients may be partially missing either due to incomplete follow-up for progression, or due to data that may be collected but confounded by patients stopping randomized therapy or starting alternative therapy prior to progression. Regulatory guidance on how to handle these patients in the analysis and whether to censor these patients differs between agencies. We present results of a reanalysis of 28 Phase III trials from 12 companies or institutions performed by the Pharmaceutical Research and Manufacturers Association-sponsored PFS Expert Team. We show that analyses not adhering to the intention-to-treat principle tend to give hazard ratio estimates further from unity and describe several factors associated with this shift. We present illustrative simulations to support these findings and provide recommendations for the analysis of PFS.

Journal of Biopharmaceutical Statistics, 23: 951–970, 2013

Can follow-up stop before observation of PD?

Table 1 Analysis censoring rules

Analysis		Description
1	ITT	All PFS events included regardless of stopping randomized therapy or subsequent therapy. Date of objective progression or if not available, death is used in analysis. If no progression or death, then censored at last objective progression-free disease assessment.
2	PDT	Censor patients who receive subsequent anticancer therapy prior to progression at latest prior visit.
3	DISC	Censor patients who prematurely discontinue randomized therapy due to toxicity or other, non-progression-related reasons at latest prior visit
4	MV	Censor patients who progress, or die (in the absence of progression), after two or more missed visits at latest prior visit
5	ALL	Censor patients who are censored in either PDT, DISC, or MV at earliest censoring time

Table 3 Summary of the extent of censoring across studies

	Percentage of ITT events censored, median (range)			Percentage reduction in ITT follow-up, median (range)		
Censor	Control	Experimental	Experimental minus control	Control	Experimental	Experimental minus control
PDT	9 (0, 32)	8 (0, 32)	0 (−15, 15)	7 (0, 39)	7 (0, 26)	0 (−30, 13)
DISC	18 (2, 58)	17 (1, 52)	−1 (−15, 18)	16 (1, 45)	12 (0, 45)	−1 (−31, 12)
MV	5 (0, 21)	5 (0, 18)	0 (−7, 11)	5 (0, 31)	6 (0, 22)	0 (−19, 9)
ALL	25 (6, 59)	23 (3, 57)	0 (−14, 16)	24 (3, 51)	20 (3, 50)	0 (−28, 9)

Can follow-up stop before observation of PD?

Table 4 Simple measure of informative censoring

	Censoring event rate ratio (CERR), median (5th, 95th percentiles)			
Censor	Control ($n = 28$)	Experimental ($n = 28$)	Overall ($n = 28$)	Ratio of experimental to control ($n = 28$)
PDT	1.4 (0.5, 3.1)	1.3 (0.6, 6.6)	1.2 (0.6, 5.2)	1.0 (0.5, 2.6)
DISC	1.3 (0.5, 3.0)	1.6 (0.5, 4.7)	1.4 (0.6, 3.1)	1.1 (0.5, 2.8)
MV	0.7 (0.3, 1.5)	0.9 (0.2, 2.1)	0.7 (0.3, 1.7)	1.1 (0.5, 2.5)
ALL	1.3 (0.5, 2.5)	1.3 (0.6, 3.7)	1.3 (0.5, 2.7)	1.0 (0.5, 2.4)

CERR = (hazard post censoring)/(hazard pre censoring)

(assumption of exponential distribution)

CERR > 1 means censored patients at greater risk, i.e. informative censoring

How does frequency of scanning impact on power?

Table 2
Simulated power for a trial designed to have 80% power to detect a hazard ratio of 0.67 at a one-sided significance level of 20%, for various scanning frequencies

Visit frequency	Cox proportional hazards		
	Power[a]	Hazard ratio[b]	Follow-up[c] (weeks)
a) Comparison of treatments with medians of 4 and 6 months			
Constant	79.8%	0.67	50
2 weeks	79.6%	0.67	51
1 month	79.0%	0.67	52
2 months	78.1%	0.67	54
4 months	74.5%	0.69	59
b) Comparison of treatments with medians of 8 and 12 months			
Constant	79.6%	0.67	86
2 weeks	79.6%	0.67	87
1 month	79.2%	0.67	88
2 months	78.8%	0.67	91
4 months	77.5%	0.68	95
6 months	76.0%	0.69	100
8 months	73.4%	0.70	104

Each row is the result of 5000 simulations, each with 50 observations per treatment arm and waiting for 69 events to occur, assuming an exponential distribution with the stated medians.
NA, not applicable.
[a] Proportion of simulations with a one-sided p-value <0.2 in favor of the more effective therapy.
[b] Calculated as the geometric mean of the hazard ratios estimated from each dataset.
[c] Calculated as the time from start of recruitment to the 69th event observed, patients recruited over 26 weeks.

"For the Cox proportional hazards model, there is only a marginal loss of power (<3%) when assessments are made at a frequency that is half the control median."

How does it impact on the estimate of the HR?

Table 7 Simulation study (informative dropout/censoring—constant hazard)

PFS distribution			Probability of being lost to follow-up at assessment visit prior progression		Treatment effect		
Median		True	Control	Experimental	OHR	OHR	OHR
C	E	OHR	p_C	p_E	1M	3M	6M
8	8	1.00	0.00	0.15	0.87 (0.65, 1.16)	0.88 (0.66, 1.18)	0.89 (0.66, 1.19)
			0.00	0.30	0.71 (0.52, 0.97)	0.73 (0.54, 1.00)	0.76 (0.56, 1.03)
			0.15	0.15	1.01 (0.75, 1.37)	1.02 (0.75, 1.38)	1.01 (0.74, 1.37)
			0.15	0.30	0.85 (0.62, 1.17)	0.85 (0.62, 1.17)	0.86 (0.63, 1.19)
			0.30	0.30	1.01 (0.72, 1.42)	1.01 (0.72, 1.41)	1.01 (0.72, 1.41)
			0.15	0.00	1.19 (0.89, 1.60)	1.16 (0.87, 1.56)	1.15 (0.86, 1.55)
			0.30	0.00	1.43 (1.05, 1.95)	1.40 (1.03, 1.91)	1.35 (0.99, 1.84)
8	10	0.8	0.00	0.15	0.69 (0.51, 0.92)	0.70 (0.52, 0.94)	0.71 (0.53, 0.95)
			0.00	0.30	0.57 (0.42, 0.77)	0.59 (0.43, 0.80)	0.60 (0.44, 0.82)
			0.15	0.15	0.80 (0.59, 1.09)	0.80 (0.59, 1.09)	0.81 (0.59, 1.10)
			0.15	0.30	0.67 (0.48, 0.92)	0.67 (0.49, 0.93)	0.68 (0.49, 0.94)
			0.30	0.30	0.81 (0.58, 1.14)	0.81 (0.58, 1.14)	0.80 (0.57, 1.13)
			0.15	0.00	0.94 (0.70, 1.26)	0.93 (0.70, 1.25)	0.92 (0.68, 1.23)
			0.30	0.00	1.14 (0.84, 1.56)	1.12 (0.82, 1.53)	1.09 (0.80, 1.49)
8	12	0.67	0.00	0.15	0.57 (0.42, 0.77)	0.58 (0.43, 0.78)	0.59 (0.44, 0.79)
			0.00	0.30	0.47 (0.35, 0.65)	0.48 (0.35, 0.66)	0.50 (0.36, 0.68)
			0.15	0.15	0.67 (0.50, 0.92)	0.67 (0.49, 0.92)	0.67 (0.49, 0.91)
			0.15	0.30	0.56 (0.40, 0.77)	0.56 (0.41, 0.78)	0.56 (0.40, 0.78)
			0.30	0.30	0.67 (0.47, 0.94)	0.67 (0.47, 0.94)	0.66 (0.47, 0.93)
			0.15	0.00	0.79 (0.59, 1.06)	0.77 (0.57, 1.04)	0.77 (0.57, 1.04)
			0.30	0.00	0.95 (0.69, 1.30)	0.93 (0.68, 1.27)	0.91 (0.66, 1.24)

Note. If a subject is considered to be missing, the subject is censored at the last scheduled assessment prior to the simulated event time. Event times follow an exponential distribution. Censoring mechanism: Each subject is censored at the last scheduled assessment prior to the simulated event time with probability p_C and p_E, respectively, depending on their treatment arm. Maximum follow-up time was fixed at 40 months. Abbreviations: C, control; E, experimental; OHR, observed hazard ratio and 95% CI; and 1M, 3M, and 6M, assessments every 1, 3, or 6 months, respectively.

Frequency of assesments has no big impact on HR estimates as long as rates of informative censoring do not differ markedly between treatment groups.

Denne et al, Journal of Biopharmaceutical Statistics, 23: 951–970, 2013

Deviation from scheduled times? Adjustment?

When You Look Matters: The Effect of Assessment Schedule on Progression-Free Survival

Katherine S. Panageas, Leah Ben-Porat, Maura N. Dickler, Paul B. Chapman, Deborah Schrag

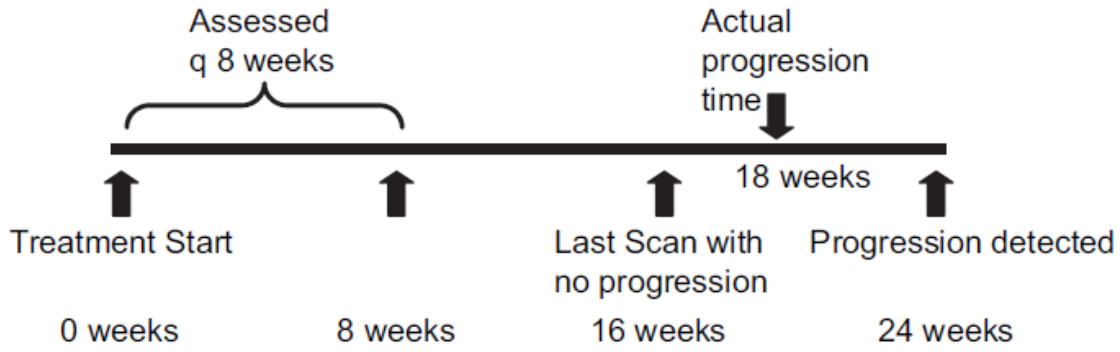

Fig. 1. Assessment intervals every 8 weeks from start of treatment. Although actual progression occurs at 18 weeks, it would not be detected until the assessment at 24 weeks.

Ignoring interval censoring leads to overestimation of median PFS => possibly false claim of clinical significance

"Furthermore, if surveillance intervals are heterogenous within a disease group, comparisons of median PFS across studies may not be meaningful."

Same problem can result from differing assesment intervals between treatment groups in one RCT.

Panageas et al, J Natl Cancer Inst 2007;99: 428 – 32

Assessment for bias from differences in scan timing

Table 1. Sensitivity Analysis for Assessment Bias: Summary of Time From Treatment Start to the First, Second, and Third Lesion Assessments, GM301 Trial (oblimersen sodium for metastatic melanoma)[7]

	No. of Patients		Median Time From Treatment Start (days)				
			G3139 + Dacarbazine		Dacarbazine		
Assessment	G3139 + Dacarbazine	Dacarbazine	No.	95% CI	No.	95% CI	*P**
First	321	311	41	41 to 42	40	40 to 41	< .0001
Second	135	106	88	84 to 91	83.5	82 to 84	.0009
Third	75	67	131	127 to 138	126	124 to 130	.0065

**P* value from log-rank test comparing time to assessment between treatment groups. Nominal *P* values are reported here.

Table 2. Simulation Results Under Equal PFS Distributions of Different Tumor Assessment Schedules, GM301 Trial (oblimersen sodium for metastatic melanoma)[7]

Configuration	Control Group	Experimental Group	Sample Size Per Treatment Group	Probability of False Inference*	Log-Rank *P*†
1A	Days 42, 84, 126, 168, 210, 252	(Delayed by 2 days) Days 44, 86, 128, 170, 212, 254	100	0.65	.100
			300	0.98	.004
1B	Days 42, 84, 126, 168, 210, 252	(Assessment time interval 2 days longer) Days 44, 88, 132, 176, 220, 264	100	0.60	.114
			300	0.97	.007

Abbreviation: PFS, progression-free survival.

*The probability of false inference was estimated by the proportion of the 5,000 replications for which the null hypothesis was rejected. This represented the probability of falsely inferring a difference in PFS between the two treatment groups.

†The average of 5,000 *P* values. Each simulation produced a *P* value. These *P* values were from a two-sided log-rank test comparing PFS between treatment groups.

Bhattacharya et al, J Clin Oncol 27:5958-5964, 2009

Need for central review?

Blinded independent central review of progression in cancer clinical trials: Results from a meta-analysis ☆

O. Amit [a,*], F. Mannino [a], A.M. Stone [b], W. Bushnell [a], J. Denne [c], J. Helterbrand [d], H.U. Burger [e]

Conclusion: The meta-analysis demonstrates that **local evaluation (LE) provides a reliable estimate of the treatment effect** with little evidence for systematic evaluation bias. **Therefore, when a trial is blinded or a large effect on PFS is observed a BICR may not be warranted.** When a BICR is warranted, a sample-based BICR may provide a reduction in operational complexity without compromising the credibility of trial results. While **for large trials that are not adequately blinded a sample-based BICR may be recommended**. A full BICR should be considered in the case of smaller trials or in situations in which there is a particular need to increase the confidence in the LE results.

Fig. 1 – BICR versus LE HR by blinding status of trial.

Amit et al, EJC 47 p1772-1778, 2011

Recommendations

- Bhattacharya et al 2009:

Table 5. Potential Sources of Bias When Evaluating the PFS Endpoint

Potential Source of Bias	Examples	Sensitivity Analyses
Assessment time	Deviations from scheduled assessment times	Simulation studies: Make hypothetical true PFS times and vary the distribution of observed progression times
	Adverse events causing delays in one arm	Backdating: Adjusting progression times by moving assessments back to scheduled times
	More frequent assessments in one arm because of worsening symptoms	
	Early evaluation of control arm in open-label studies because of a concern about lack of efficacy	
	Delay in assessments because patients are doing well	
	Difficulty in aligning assessment times when treatments have different cycle lengths	
Symptomatic/nonradiologic progression events	These types of events may be declared earlier in one arm	Remove clinical progression events: Consider only radiologic progression and death as PFS events
	Inclusion of objective progression events without documentation of lesion measurements	Consider only objective progression with documentation and deaths as PFS events; backdate objective progression events to the previous complete assessment in the event of missing or incomplete assessments
Missing data	Imbalances between treatment arms in number of missing tumor measurements	Treat two arms identically and apply conservative assumptions to missing data (eg, backdate the PD date)
	Patients lost to follow-up or rate of censoring is imbalanced between treatment arms	Make increasingly conservative assumptions about experimental arm, while making more liberal assumptions in the control arm (ie, consider events in the active arm and censored in the control)

Abbreviations: PFS, progression-free survival; PD, progressive disease.

Recommendations

- Stone et al 2007 (mainly regarding phase II designs):
  - progression endpoints that utilize all available progression data rather than early fixed timepoint analyses
  - little gain from assessing PD more frequently in routine clinical practice
- Panageas et al 2007 (mainly regarding phase II designs):
  - design trials to limit the influence of interval censored data
  - consider using aqequate IC analysis methods
  - using both the lower (the assessment before the detection occurred) and upper endpoints (as usually done) of the assessment intervals. “This approach will mimic the extreme scenarios and will bracket the true distribution.”

Recommendations

- Bushnell, Stone 2013:
  - in the presence of inadvertent unequal visit spacing, IC methods are substantially more robust to bias compared to conventional methods
- Stone et al 2011:
  - Patients should be assessed at the same frequency in each treatment arm and interval censoring methods should be included as a sensitivity analysis. Once validated software is widely available, consideration should be given to the use of ICA methods to replace the log-rank test and Cox regression as the primary tool for analysing PFS data.
- Sun et al 2013:
  - review IC analysis methods and implications for trial design (size, group sequential designs)
  - recommend Turnbull, Finkelstein and own methods
  - software mentioned: %EMICM, proc logistic (SAS); SAND, Icens (R);

Recommendations

- Amit et al 2011:
  - no BICR necessary, if blinded trial or large PFS effect observed
  - in large unblinded trial maybe sample-based BICR

References

- Denne et al, Journal of Biopharmaceutical Statistics, 23: 951–970, 2013
- Stone et al, Contemporary Clinical Trials 28 (2007) 146–152
- Panageas et al, J Natl Cancer Inst 2007;99: 428 – 32
- Amit et al, EJC 47 p1772-1778, 2011
- Bhattacharya et al, J Clin Oncol 27:5958-5964, 2009
- Bushnell, Stone, Journal of Biopharmaceutical Statistics, 23: 949–950,al 2013
- Sun et al, Journal of Biopharmaceutical Statistics, 23: 986–1003, 2013
- Stone et al, EJC 47 p1763-1771, 2011