BBSW 24 May, 2022

BICR - Some Consideration Points in Practice

Disclaimer: *The opinions expressed in this meeting are solely those of the presenter and not necessarily those of Genentech/Roche.*

Topics

- FDA guidance
- Factors for Consideration When Assessing the Need of BICR
- BICR challenges
- BICR Sample Audit - practical challenges
- Food for thoughts

FDA Guidance

FDA released a guidance on “Clinical Trial Imaging Endpoint Process Standards” in April 2018. It says:

“The usefulness of a centralized image interpretation process is determined by the role, variability, and susceptibility to bias of imaging within the trial as well as modality-specific image quality considerations and overall trial design features. Centralized image interpretation is not always critical, even for a phase 3 trial primary endpoint that uses some aspects of quantitative imaging, if the quantitative measures are widely performed and reported in clinical medicine, little imaging acquisition or interpretation variability is anticipated, and potential biases in image interpretation are controlled by the trial design features.”

FDA Guidance for Industry Clinical Trial Imaging Endpoint Process Standards, April 2018

Factors for Consideration When Assessing the Need of BICR

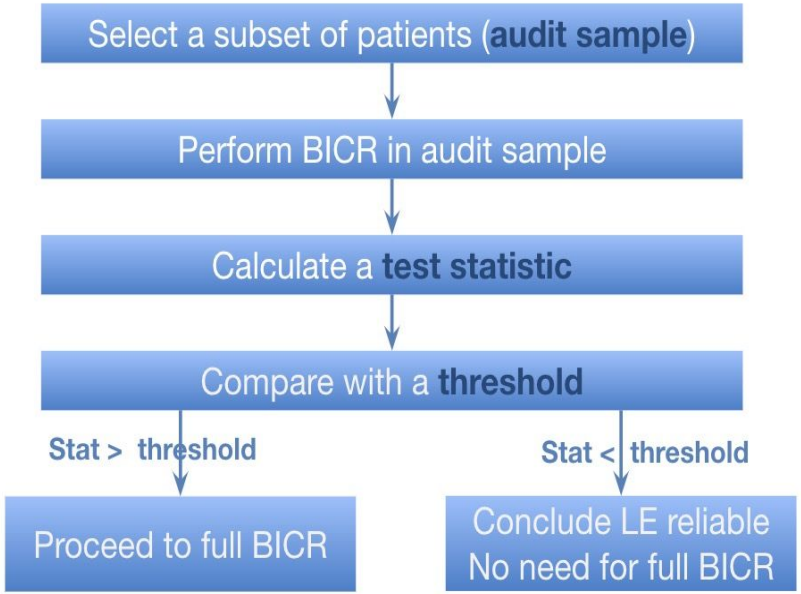
	BICR is less needed	BICR is more needed
Study Design Elements	Double blind	Open label
	Signal seeking	Confirmatory or registrational
	Regional focus	Multicenter international
	Existing indication	Novel indication
	SOC comparator	Novel comparator
	Large anticipated treatment effect	Marginal anticipated treatment effect
Endpoints	Additional EPs (e.g. OS)	Single EP
	Established EP	Not established EP
Assessment Criteria	Single assessment criteria	Multi assessment criteria
	Yes (precedence) (e.g. Lugano/Recist 1.1.)	No (precedence) (e.g. RAPNO, RECIL, Adjuvant)
Other studies	LE only	BICR (as primary or sec. EP)

BICR challenges

- **Potential bias due to informative censoring:** When image collection stops at LE-PD in most studies, LE-PD cases not agreed by BICR are censored for BICR analysis. This informative censoring could lead to larger median-PFS estimate, biased HR estimates and loss of power.
- **Potential bias** could be introduced by the content and the timing of presentation of selected clinical data supporting BICR committee evaluation.
- **Can impact study timelines:** (1) Extra time needed for BICR confirmation of eligibility (2) Difficult event-tracking due to delayed image submission and turnaround time
- **Inherent complexity in BICR procedure** (e.g. vendor selection, training, monitoring)
- **Resource intense:** Additional resources required from sponsors and investigator sites
- **High costs:** ~9% of overall study costs

Quick Review of BICR Sample Audit

The Audit Concept



Current Available BICR Audit Methods

PhRMA (Amit 2011)	NCI (Dodd 2011)	HRR (Stone 2015)
Target: Disagreement in timing of declaring PD	Target: Confirm LE-determined treatment benefit	Target: Bias reflected in treatment effect
Statistic: Differential Discordance; Call PD: Late in treatment / Early in control	Statistic: Estimated upper CI of HR_{BICR}	Statistic: HR_{BICR}/HR_{LE} in audit sample

FDA ODAC (Oncology Drug Advisory Committee) held on 24 July 2012.

BICR Sample Audit: Practical Challenges

Technical challenges

- Audit approach applicable only for larger trials (sufficient amount of audit sample data required)
- When determining audit sample size and threshold, trade-off between sensitivity (Sn) and specificity (Sp)
- Imperfect (i.e. <100%) Sn & Sp
- Audit sampling being performed close to database snapshot date, often too late to switch to a full BICR and still deliver data to HA's per original timelines
- Selection of audit method to use (no clear winner)

Operational Challenges

Compared to full BICR, additional uncertainties about impact on resource, cost & timelines:

- Only moderate resource/cost savings
- Need for: an additional "audit plan", a complicated BICR charter, potential additional interactions with the iDMC.
- If audit test fails, full BICR will be conducted in an expedited manner resulting in additional costs for a read and delay delivery of the data, which may postpone submission and approval of new therapies for patients

Food for thoughts

- When is BICR “truly” needed?
- When BICR is needed, will it be a primary EP or a sensitivity analysis?
- Is BICR sample audit applicable/helpful in your study?

Two pieces of work in manuscript preparation (to be submitted in 2022):

- A retrospective meta-analysis and review of all Oncology randomized studies with LE and BICR results during 2006 - 2020 (n=49).
- Using “interim sampling” instead of “simple random sampling” when performing BICR sample audit.

Doing now what patients need next

Review of Past Roche/Genentech Oncology Studies
- Impact of BICR on Study Outcome in 49 Ph3 Studies Investigated

ANALYSIS RESULTS:

- ORR: 1 study
  - LE/BICR results are consistent and LE was used for filing/label
- PFS: 48 studies
  - LE/BICR consistent: 39 (i.e. both or neither crossed pre-specified alpha boundary)
  - LE/BICR discordant: 9
    - 6 studies: filing decision made based on LE results
    - 3 studies: LE (marginal) positive vs BICR negative → totality of data and not BICR itself impacted filing decision

ANALYSIS CONCLUSIONS:

- BICR does not impact filing go/no-go decision and label for 46/49 studies
- Filing go/no-go decision and label are primarily based on population level rather than individual patient level statistics.

Review of Past Roche/Genentech Oncology Studies

- Overview of Characteristics of Investigated Studies

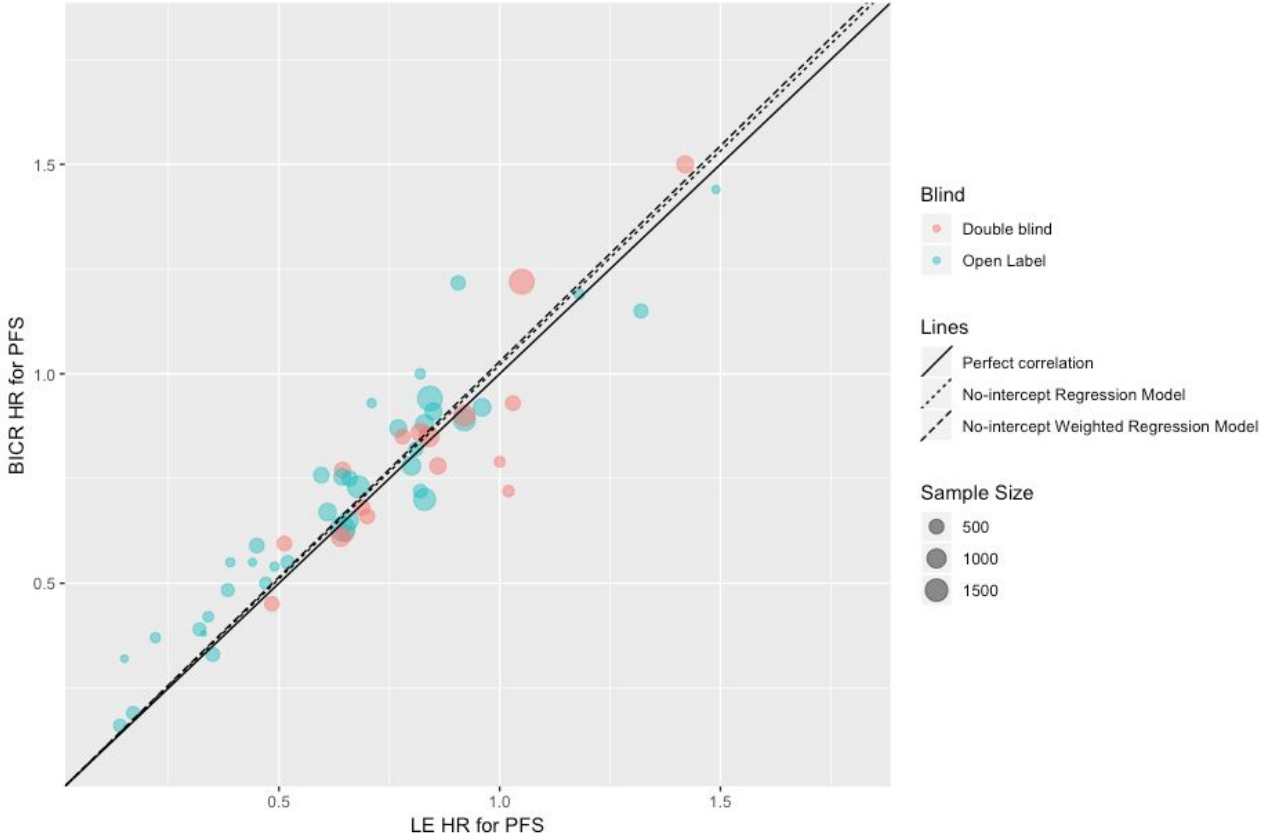
Characteristics		Trials w/ PFS (N=55)	Trials w/ ORR (N=40)
Design	Double Blind	17	7
	Open Label	38	33
Sample Size	Mean	631	625.9
	Min-Max	80-1904	80-1904
Phase	III	48	33
	I/II	7	7
Randomization ratio	1:1	43	34
	2:1	12	6
CCOD	Before 2015	24	18
	After 2015	31	22
Sponsor	Roche	50	38
	Co-op	5	2
Criterion	RECIST/(m)Lugano	43	33
	Other	12	7

This database of past studies has a good coverage of studies with various designs: double blind and open label, a wide range of sample size, various phases, different rand. ratios, old and recent, Roche sponsored or run by co-op group, with or without standardized criterion for image reading.

As expected, BICR was implemented in more studies with PFS as a primary endpoint than studies with ORR as a primary endpoint.

Review of Past Roche/Genentech Oncology Studies

- BICR versus LE PFS-HRs



- The BICR and INV PFS-HRs are highly correlated.
- The high concordance is observed regardless of sample size
- The high concordance is observed regardless of double-blind or open-label.

Review of Past Roche/Genentech Oncology Studies
- Concordance between BICR and LE PFS-HRs

		PFS HR		
	# of studies	r (95% CI) [1]	R^2 (95% CI) [2]	HRR [3]
Overall	55	0.946 (0.909,0.968)	0.893 (0.802,0.942)	1.044 (1.009,1.081)
Double Blinded	17	0.882 (0.697,0.957)	0.858 (0.643,0.949)	1.014 (0.958,1.073)
Open Label	38	0.956 (0.916,0.977)	0.904 (0.803,0.956)	1.062 (1.016,1.110)

[1] r: Pearson Correlation on log(HR)
[2] R^2: coefficient of determination from weighted linear regression model on log(HR), weighted by sample size
[3] HRR: HRR = (BICR PFS HR)/(LE PFS HR), estimated from linear mixed effect model, weighted by sample size

Review of Past Roche/Genentech Oncology Studies
- Impact of BICR on Conclusion/Decision based on PFS

Among the total 55 studies investigated:

- 36.4% had HRR <= 1, i.e. BICR estimated stronger treatment benefit than LE did. BICR didn't capture additional LE bias favoring treatment arm.
- 32.7% had HRR between 1 and 1.15, i.e. BICR & LE assessed PFS HRs are close. There is no change of study outcomes^ except for 2 out of the 18 studies. (*slide 25 for details*)
- 30.9% had HRR >1.15
  - Majority (10/17, 58.8%) are due to small HRs, i.e., both BICR HR & LE HR are <0.6. The differences did not change the study conclusion.
  - Other reasons include: Ph2 study w/ small sample size; PFS is not a primary endpoint.
  - There is no change of study outcomes^ except for 1 out of the 17 studies. (*slide 26 for details*)

BICR is better than LE

		HRR* HRR= (BICR PFS HR)/(LE PFS HR)			
Characteristics		<=0.85	(0.85,1]	(1,1.15]	>1.15
Overall		3 (5.5%)	17 (30.9%)	18 (32.7%)	17 (30.9%)
Design	Double Blind	2	8	4	3
	Open Label	1	9	14	14
Sample Size	Mean	625.7	776.9	691	422.2
	Min-Max	222-1400	121-1873	152-1873	80-1904
Rand. ratio	1:1	1	14	17	11
	2:1	2	3	1	6
Time	Before 2015	1	8	8	7
	After 2015	2	9	10	10
Sponsor	Roche	3	16	16	15
	Co-op	0	1	2	2

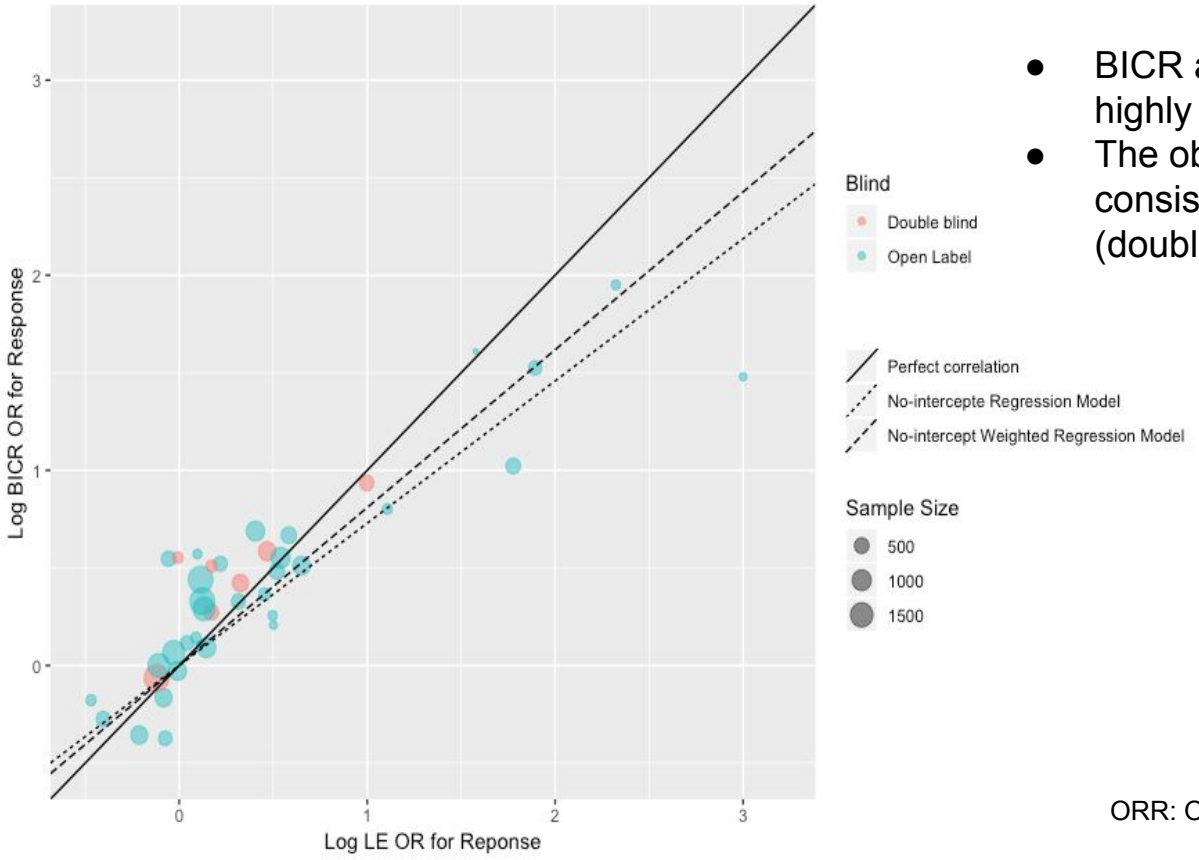
^ Study outcomes refer to filing go/no-go decision or impact on product label.

Review of Past Roche/Genentech Oncology Studies
- Summary of BICR Impact on PFS Analysis Outcomes

- Highly correlated LE and BICR PFS outcomes have been observed.
    - High degree of association between BICR and LE estimates of the PFS treatment effect was observed with Pearson correlation r = 0.946 [95%CI: 0.909,0.968]
    - Based on weighted linear regression model, 89.3% [95%CI: 80.2%,94,2%] of the variability in the BICR PFS can be explained by the LE PFS.
    - The overall HRR (of BICR HR over LE HR) from the random effects model was 1.044 [95%CI: 1.009,1.081], indicating only a 4.4% difference between the two evaluations.
- The results are consistent regardless of design (double blind or open label).

Review of Past Roche/Genentech Oncology Studies

- BICR versus LE for ORR



- BICR and LE log(OR) for ORR are highly correlated.
- The observed correlation is generally consistent regardless of design (double blind or open label).

ORR: Overall Objective Response; OR: Odd Ratio

Review of Past Roche/Genentech Oncology Studies
- Concordance between BICR and LE ORR

		Response Odds Ratio (OR)		
	# of studies	r (95% CI) [1]	R^2 (95% CI) [2]	Ratio of OR of Response[3]
Overall	40	0.890 (0.801,0.941)	0.767 (0.631,0.874)	1.065 (0.983,1.154)
Double Blinded	7	0.826 (0.194,0.974)	0.868 (0.098,0.997)	1.108 (0.978,1.257)
Open Label	33	0.900 (0.806,0.950)	0.760 (0.603,0.880)	1.055 (0.959,1.162)

ORR: Overall Response Rate; OR: Odd Ratio
[1] r: Pearson Correlation on log(OR)
[2] R^2: coefficient of determination from weighted linear regression on log(OR), weighted by the sample size
[3] Ratio of OR of Response = (BICR OR for response)/(LE OR for response), estimated from linear mixed effect model weighted by sample size

Review of Past Roche/Genentech Oncology Studies
- Summary of BICR Impact on ORR Analysis Outcomes

- Highly correlated LE and BICR ORR outcomes have been observed.
    - Degree of association between BICR and LE estimates for the overall response was numerically lower than for PFS, but still relatively high with Pearson Correlation r=0.890 [95%CI: 0.801,0.941]
    - Based on the weighted linear regression model, 76.7% [95%CI: 63.1%, 87.4%] of the variability in the BICR log(OR) can be explained by the LE log(OR).
    - The overall ratio of ORs (BICR vs LE) from the random effects model was 1.065 [95%CI: 0.983,1.154], indicating only a 6.5% difference between the two evaluation.
- The results are generally consistent regardless of design (double blind or open label).

(Supp. Info) Details of BICR Impact on Study Outcomes with PFS Endpoints

Among the 18 studies whose HRRs were between 1 and 1.15, the study conclusion (i.e. stat significant treatment benefit shown or not) were the same based on BICR and LE for 13 studies except for the following 5 studies:

- BO22589 (MARIANNE): open label ph3; 3-arm, BICR PFS is the primary for two sets of testing; for (T-DM1vs T+P) test, the LE and BICR outcomes are consistent, both were negative; for the other test (T-DM1+P vs T+P), BICR negative, LE positive, filing no-go based on totality of data including BICR. BICR is not a main factor for the decision making. BICR has implication for filing decision and label.
- WO30070 (IMvigor130): double blind ph3; LE PFS is the primary; LE positive, BICR negative, waiting for additional OS data; filing no-go based on totality of data including BICR. This is a PMR study aiming to convert AA label into full label, and ODAC held on Apr 28,2021. BICR is not a main factor for the decision making. BICR has implication for filing decision and label.
- WO29637 IMmotion151: open label; LE is the primary; LE positive, BICR negative; filing go based on LE, however, submission withdrawn following negative OS data. BICR no impact on filing, no impact on label.
- CO39262 (IMspire150): double blind ph3; LE PFS is the primary; LE positive, BICR negative, US filing-go and USPI based on LE; EU filing pending more mature OS data awaited. BICR no impact on filing, no impact on label.
- BO21223 (GALLIUM): open label ph3; FL population & ITT population, LE PFS is the primary; LE positive, BICR negative; filing go based on LE data; EU SmPC based on LE data, US PI includes both LE and BICR stating consistent. BICR no impact on filing, no impact on (EU) label.

In summary, there is no change of study outcomes due to BICR except for 2 (i.e. BO22589 and WO300070) out of the 18 studies

(Supp. Info) Details of BICR Impact on Study Outcomes with PFS Endpoints

Among the 17 studies whose HRRs were >1.15,

- 10 studies are due to small HRs, i.e., both BICR HR & LE HR are <0.6. The differences did not change the study conclusion.
- 4 studies (BO20924,BO29554[phII, n=154, negative for both],GO29537,BO20094) had no impact on study conclusion (both IRC and INV or neither of them crossed the pre-specified alpha boundary).
- WO29074(IMmotion150): ph2 with small sample size (n=202), no filing implication.
- GO29438 (IMpower132): open label ph3; LE PFS is the primary, LE positive but not clinical meaningful, BICR negative; filing no-go based on totality of data including BICR and competitive landscape change. BICR is not a main factor for the decision making. BICR has implication for filing decision and label.
- NO16966: ph3; 2x2 design, two co-primary endpoints: (1) open label non-inferiority portion (XELOX Versus FOLFOX-4) HRR > 1.15, LE was better than BICR, filing go based on LE, EU label based on LE; BICR no impact on filing, no impact on label. (2) double blind superiority portion for Avastin in colorectal cancer, HRR < 1, LE and BICR are consistent, BICR no impact on filing, no impact on label.

In summary, there is no change of study outcomes due to BICR except for 1 (i.e. GO2948) out of the 17 studies.

(Supp. Info) Details of BICR Impact on Study Outcome for the 49 Ph3 Studies Investigated

Among the 49 Ph3 studies (in PFS or ORR analyses), BICR and LE are consistent (i.e. both or neither of them crossed the pre-specified boundary), except for the following studies:

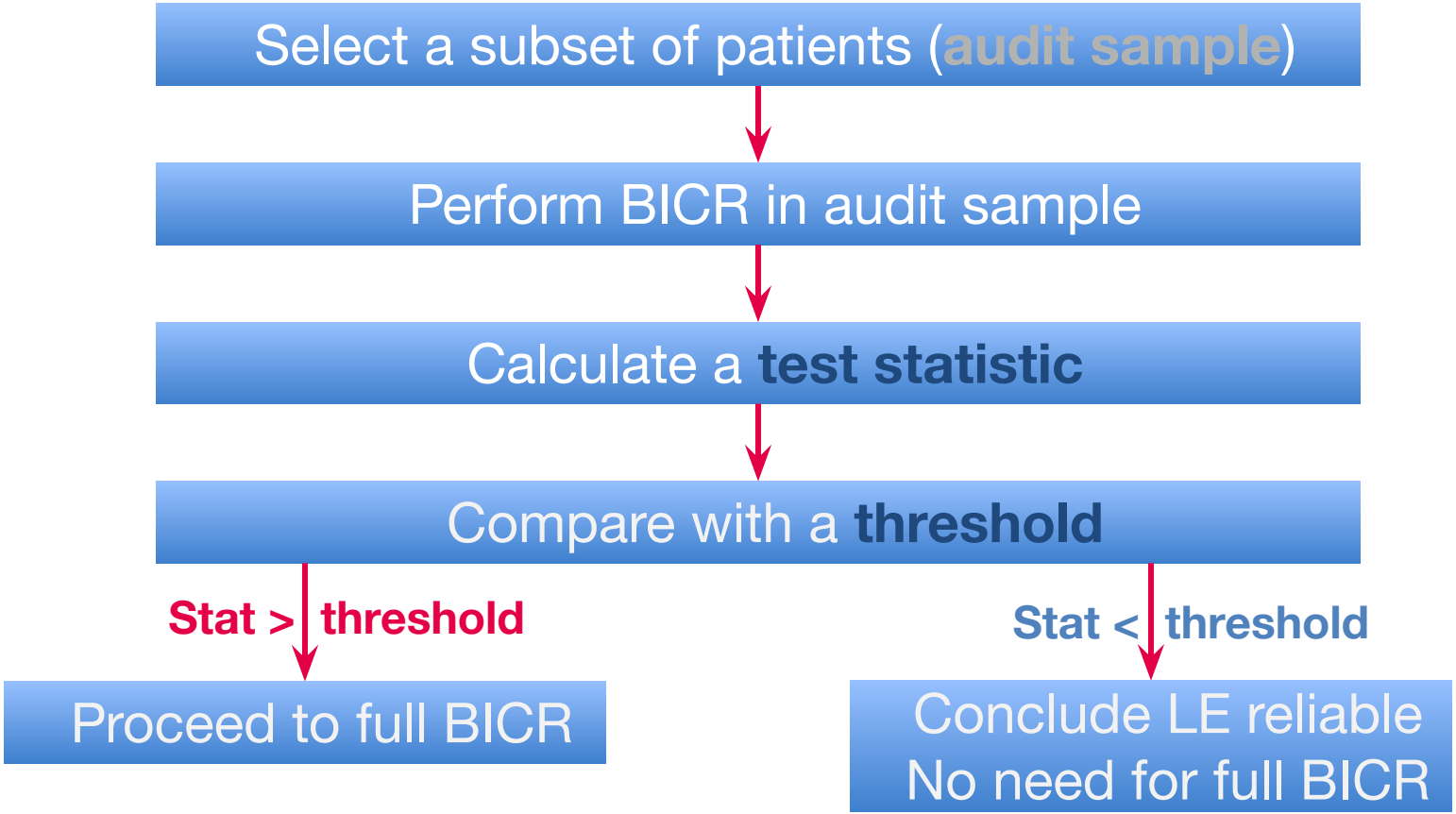
1. MO39196 IMpassion131: double blind; LE is the primary; LE negative; BICR positive; filing no-go based on LE. BICR no impact on filing, no impact on label.
2. BO20231 AVEREL: open label; LE is the primary, LE negative; BICR positive; filing no-go based on LE. BICR no impact on filing, no impact on label.
3. TDM4788g/BO22589 MARIANNE (see slide 30): filing no-go based on totality of data including BICR. BICR has implication for filing decision and label.
4. WO30070 IMvigor130 (see slide 30): filing no-go based on totality of data including BICR. BICR has implication for filing decision and label.
5. WO29637 IMmotion151 (see slide 30): filing go based on LE, later submission withdrawn following negative OS data. BICR no impact on filing, no impact on label.
6. CO39262 IMspire150 (see slide 30): filing go based on LE. BICR no impact on filing, no impact on label.
7. BO21223 GALLIUM (see slide 30): filing go based on LE. BICR no impact on filing, no impact on label.
8. GO29438 IMpower132 (see slide 31): filing no-go based on totality of data including BICR and competitive landscape change. BICR has implication for filing decision and label.
9. NO16966 (see slide 31): filing go based on LE. BICR no impact on filing, no impact on label.

In total, BICR didn't change the filing go/no-go decision and has no impact on label for 46 out of 49 Ph3 studies investigated.

(Supp. Info) 17 Studies with HRR > 1.15

	Study.Number.Name	HRR	INV.PFS	IRC.PFS	INV.p.value	IRC.p.value	Consistent	Blind	randomazation.ratio	Phase	N.PFS.IRC
1	AVF3694g / BO20094 Ribbon 1	1.195652	0.644	0.770	<0.0001	0.04	1	Double blind	2	III	622
2	WO29074 IMmotion150	1.219512	0.820	1.000	0.2541	0.9819	NA	Open Label	1	II	202
3	NO16966	1.161905	1.050	1.220	<NA>	<NA>	0	Double blind	1	III	1904
4	MO22224 AURELIA	1.260417	0.384	0.484	<0.0001	<0.0001	1	Open Label	1	III	361
5	BO20924 BERNIE	1.309859	0.710	0.930	0.0977	0.7189	1	Open Label	1	II	154
6	GO28141	1.162109	0.512	0.595	<0.0001	0.0003	1	Double blind	1	III	495
7	YO40245 IMbrave150	1.311111	0.450	0.590	<0.0001	<0.0001	1	Open Label	2	III	501
8	GO30140	1.250000	0.440	0.550	0.0004	0.0108	1	Open Label	1	Ib	119
9	GO29537 Impower130	1.172628	0.643	0.754	<0.0001	0.0026	1	Open Label	2	III	679
10	GO29438 IMpower 132	1.271812	0.596	0.758	<0.0001	0.0055	0	Open Label	1	III	578
11	YO25121	1.235294	0.340	0.420	<0.0001	0.0001	1	Open Label	1	III	217
12	ML20650 EURTAC	1.410256	0.390	0.550	<0.0001	0.003	1	Open Label	1	III	153
13	BO29554 BFAST	1.343267	0.906	1.217	0.3458	0.064	1	Open Label	1	II/III	471
14	YO29449 Alesia	1.681818	0.220	0.370	<0.0001	<0.0001	1	Open Label	2	III	187
15	MO29750 ALUR	2.133333	0.150	0.320	<0.001	<0.001	1	Open Label	2	III	107
16	GO29365	1.151515	0.330	0.380	<0.0001	0.0002	NA	Open Label	1	Ib/II	80
17	BO21004	1.218750	0.320	0.390	<0.0001	<0.0001	1	Open Label	2	III	351

The Audit Concept



Current Available Statistics

PhRMA (Amit 2011)

Target

Disagreement in timing of declaring PD

Statistic

Differential Discordance

Call PD	Late in treatment
	Early in control

NCI (Dodd 2011)

Target

Confirm LE-determined treatment benefit

Statistic

Estimated upper CI of HR_{BICR}

HRR (Stone 2015)

Target

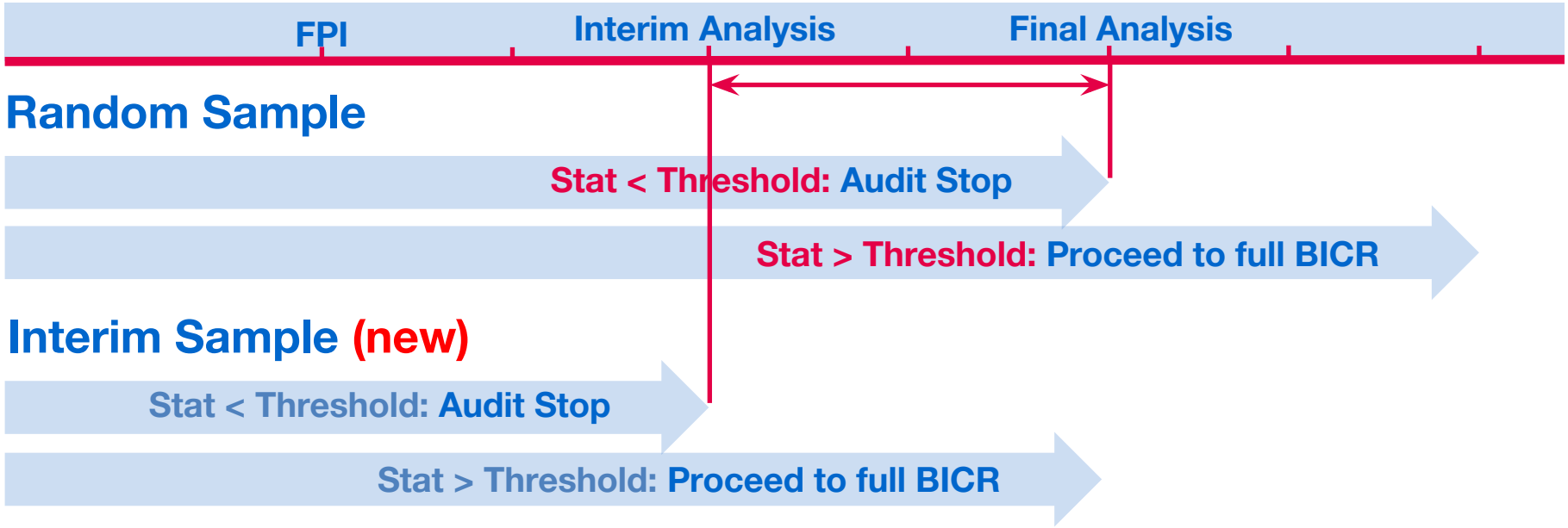
Bias reflected in treatment effect

Statistic

HR_{BICR}/HR_{LE} in audit sample

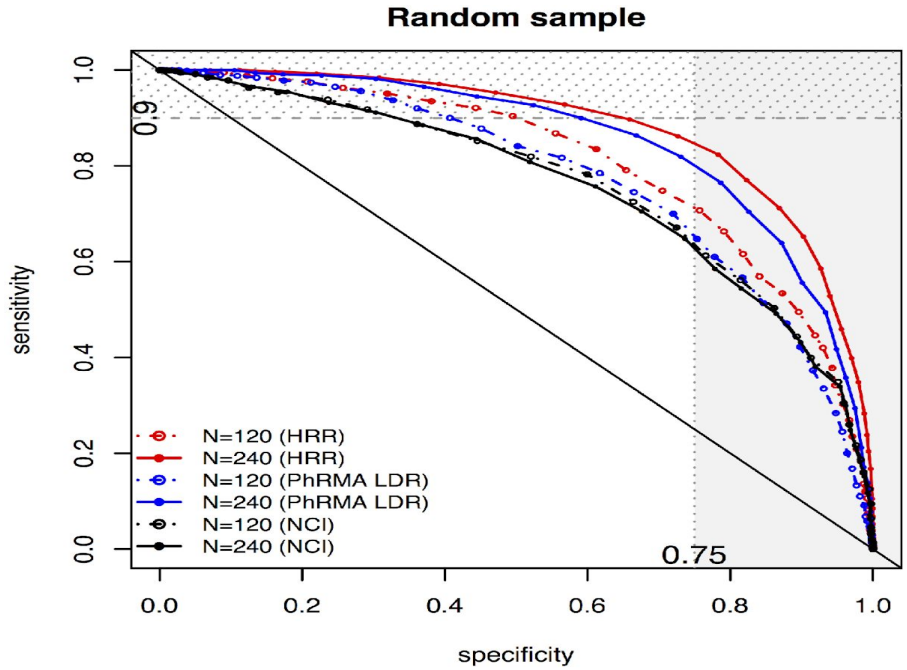
Study Timeline

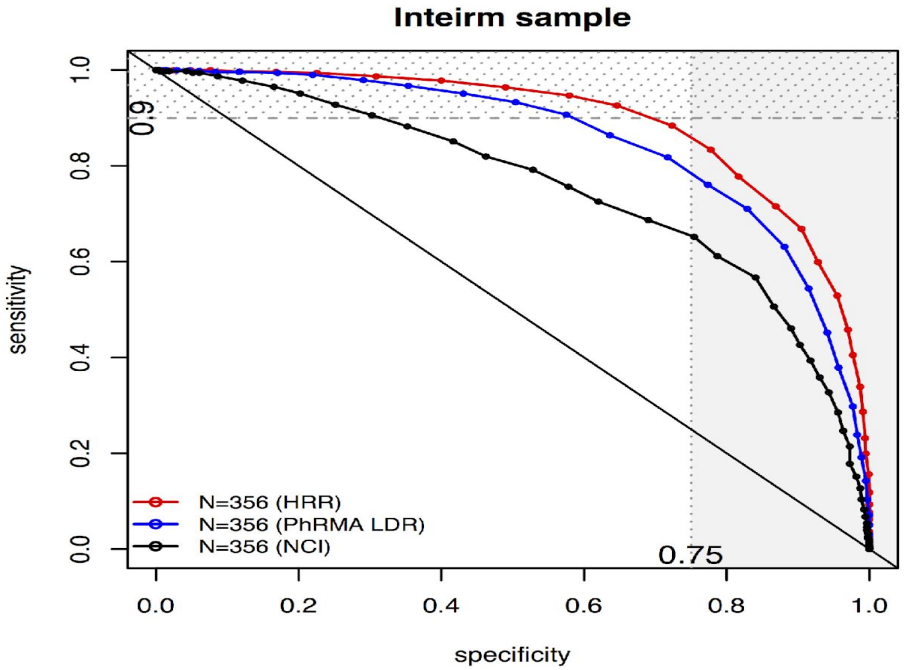
Sampling Strategies



- "Random Sample" (i.e. all tumor images for the selected patients through the whole study) may cause delayed timeline and increased cost if audit checking failed to pass the consistency check for LE and BICR.
- "Interim sample" (i.e. all tumor images for all enrolled patients up to a certain time point) may potentially save cost with no delayed timeline.
- "Interim sample" gives similar performance compared to using "random sample" based on simulation (results shown on next slide)

Simulation Evaluation





- HRR (ratio of HR) outperforms PhRMA and NCI methods
  - Capture what we care about: bias reflected in treatment effect
- Interim sample performs as well as random sample

Above statements hold under various simulation settings.

Summary and Future Steps

- **Recommendation**
  - Interim sample based PFS audit using HRR as test statistics

- **Shiny app (simulation tool and real data application)**
  https://shiny.roche.com/public/shix19/shiny/PFS_Audit/

- **Future steps:**
  - Collect opinion on if this is a reasonable and feasible approach.
  - Share the findings with our pharmaceutical community by publication and/or conference presentation.
  - Seek FDA feedback on this proposed approach.
  - Get management advice on if/when/how we can start making this suggestion to project teams to consider this PFS strategy.

Reference:

- Dodd, Lori E., et al. "An Audit Strategy for Progression-Free Survival." Biometrics 67.3 (2011): 1092-1099.
- Amit, O., et al. "Blinded independent central review of progression in cancer clinical trials: results from a meta-analysis." European Journal of Cancer 47.12 (2011): 1772-1778.
- Stone, Andrew, et al. "Model free audit methodology for bias evaluation of tumour progression in oncology." Pharmaceutical statistics 14.6 (2015): 455-463.