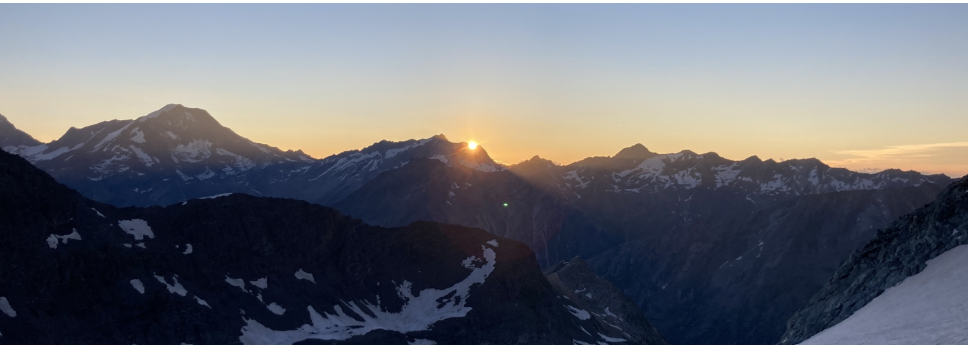

CLL11 – a trial tailored to answer questions from many stakeholders efficiently

Kaspar Rufibach

EFSPi regulatory statistics workshop 2024, Basel



**Every successful drug development program
has many parents.**

**Gabi Bieska, Elina Asikanius,
Uli Burger, Jörg Maurer.**

**I have not been involved in
the design and running of this trial!**

Who can we convince once the data is in?

Regulators are not the only stakeholder.

CLL11 was a platform trial!

Closed testing efficient for multiarm trials.

You need good drug developers!

Impact

Approval and reimbursement of **GAZYVA** in **chronic lymphocytic leukemia** (CLL).

1st Breakthrough Therapy-designated drug to receive FDA approval, [Lee et al. \(2014\)](#).

Clinical publication: [Goede et al. \(2014\)](#).

Statistical publication: [Asikanius et al. \(2016\)](#). Simulation code as supplementary material.

No impact:

**More frequent use of
closed testing in multiarm trials.**

Context at design stage

Context at design stage

CI + Gazyva

2nd generation anti-CD20, experimental

Context at design stage

Chlorambucil

approved standard in Germany (only!)

CI + Gazyva

2nd generation anti-CD20, experimental

Context at design stage

Chlorambucil

approved standard in Germany (only!)

CI + MabThera

1st generation anti-CD20, not approved, off-label use

CI + Gazyva

2nd generation anti-CD20, experimental

2-arm trial G vs. C:

Regulator ☺

Patients ☹

Scientific community / treating physicians ☹

HTA ☹

How to efficiently design a 3-arm trial?

Null hypotheses and type I error protection

Pairwise null hypotheses:

$$H_{0,G \text{ vs. } C} : HR_{G/C} = 1,$$

$$H_{0,R \text{ vs. } C} : HR_{R/C} = 1,$$

$$H_{0,G \text{ vs. } R} : HR_{G/R} = 1.$$

All hypotheses of interest.

Design must **strongly** protect **familywise error rate** (FWER).

Primary endpoint: **progression-free survival**.

Closed testing:

**General principle to construct
testing strategy that protects FWER.**

$$\begin{array}{c}
\boxed{H_{0,G \text{ vs. } C} : \text{HR}_{G/C} = 1} \quad \boxed{H_{0,R \text{ vs. } C} : \text{HR}_{R/C} = 1} \quad \boxed{H_{0,G \text{ vs. } R} : \text{HR}_{G/R} = 1} \\
\uparrow \\
H_{0,G \text{ vs. } C} \cap H_{0,G \text{ vs. } C} \cap H_{0,G \text{ vs. } C} \\
\uparrow \\
\boxed{H_{0,\text{global}} : \text{HR}_{G/C} = \text{HR}_{R/C} = \text{HR}_{R/G} = 1}
\end{array}$$

Reject $H_{0,\text{global}}$ at $\alpha \Rightarrow$ each individual hypothesis **can be tested at α** .

If you have enough power to test $H_{0,\text{global}}$ – virtually **free lunch!**

Assumptions

Global significance level: $\alpha = 0.05$.

Alternative hypotheses and power for sample size planning:

- **98%** power to detect $HR_{G/C} = 12/27 = \mathbf{0.444}$,
- 80% power to detect $HR_{R/C} = 12/20 = \mathbf{0.600}$,
- 80% power to detect $HR_{G/R} = 20/27 = \mathbf{0.741}$.

Why 98%?

- Futility and efficacy interim for R vs. C at final analysis of G vs. C \Rightarrow 30% adequate information fraction to perform interim at.
- Enough **safety follow up** for C for benefit-risk.
- Randomization to arm G expected to have terminated at G vs. C analysis cutoff.

Four potential strategies

① Three separate trials:

- Each at $\alpha = 0.05$.
- Distribute patients on three trials \Rightarrow use each patient for **one** comparison only.

② One 3-arm trial with Bonferroni correction:

- Each comparison at $\alpha = 0.0167$.
- All patients in same trial \Rightarrow use each patient for **two** comparisons.

③ One 3-arm trial with closed testing, wait until last comparison mature:

- Test $H_{0,\text{global}}$ once targeted number of events for **latest** comparison reached.

④ One 3-arm trial with closed testing, each comparisons analyzed once mature:

- Test $H_{0,\text{global}}$ once targeted number of events for **first** comparison is reached.
- Perform other pairwise comparisons once targeted number of events reached.

Metrics

- **Time to regulatory approval of Gazyva**: determined by first cutoff G vs. C.
- **Time to make patients / scientific community / HTA happy**: determined by cutoff G vs. R.

Consideration:

- Closed test in Strategies 3 and 4 induces **power loss** for each pairwise comparison. Quantify power loss, mainly for G vs. R.

Time to regulatory approval for Gazyva

	G vs. C: 0.444	R vs. C: 0.600	G vs. R: 0.741
Three separate trials	34.4	39.2	hopeless
Bonferroni	21.4	24.3	90.1
Closed testing – last mature	47.4	47.4	47.4
Closed testing – first mature	18.6	19.4	51.0

Time to make patients / HTA happy

	G vs. C: 0.444	R vs. C: 0.600	G vs. R: 0.741
Three separate trials	34.4	39.2	hopeless
Bonferroni	21.4	24.3	90.1
Closed testing – last mature	47.4	47.4	47.4
Closed testing – first mature	18.6	19.4	51.0

Last scenario:

- Slight **power loss (1.7%)** compared to 2-arm trial for G vs. R comparison due to global test.
- Compensate through **17** more events.
- Corresponds to **3.8** months delay compared to 2-arm trial.

G vs. R stopped at interim analysis for efficacy.

**To pull this off you need
good drug developers!**

	A vs. C C: N = 238 / A: N = 118	A vs. B B: N = 233 / A: N = 118	B vs. C ² C: N = 333 / B: N = 330
	Global test of closed testing procedure*		
July 2012	A vs. C primary analysis 105 events (100%) HR = 0.44 Median PFS: 27 vs. 12 months Significance level: 0.05		B vs. C futility / efficacy interim analysis 125 events (31%) futile if HR > 0.88 Non-binding
August 2012		A vs. B primary analysis 145 events (100%) HR = 0.60 Median PFS: 20 vs. 12 months Significance level: 0.05	
May 2013	A vs. C updated analysis ¹	A vs. B updated analysis ¹	B vs. C efficacy interim analysis 300 events (74%) Significance level: 0.019
NA ³			B vs. C final analysis 406 events (100%) Significance level: 0.044

Operational aspects in CLL11

Operational bias:

- Hazard ratios became known over time: $HR_{G/R} = HR_{G/C}/HR_{R/C}$! (under some assumptions).
- Treatment schedule in CLL11 rather fixed once started.
- Define analysis timepoints not only through PFS cutoffs: e.g. all patients needed to be randomized to G prior to cutoff for G vs. C.

Further operational aspects:

- Multiple final / interim analyses on different sets of patients.
- iDMC for interim analyses in G vs. R.
- Independent response review: even more important after G vs. C was unblinded.

Who can we convince once the data is in?

Regulators are not the only stakeholder.

CLL11 was a platform trial!

Closed testing efficient for multiarm trials.

You need good drug developers!

Thank you for your attention.

kaspar.rufibach@gmail.com

Slides can be downloaded on

www.kasparrufibach.ch

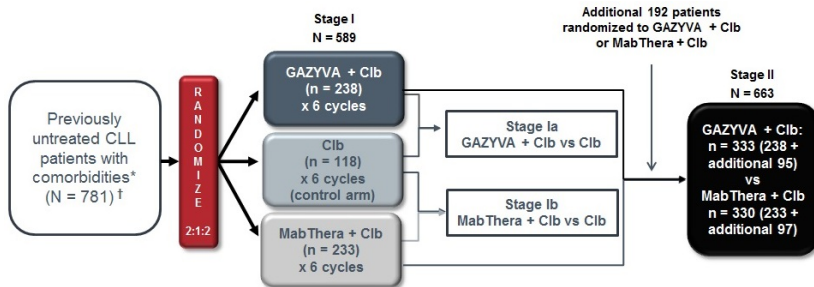
References I

- ▶ Asikanius, E., Rufibach, K., Bahlo, J., Bieska, G. and Burger, H. U. (2016). Comparison of design strategies for a three-arm clinical trial with time-to-event endpoint: Power, time-to-analysis, and operational aspects. *Biometrical journal. Biometrische Zeitschrift* **58** 1295–1310.
- ▶ Goede, V., Fischer, K., Busch, R., Engelke, A., Eichhorst, B., Wendtner, C. M., Chagorova, T., de la Serna, J., Dillhuydy, M. S., Illmer, T., Opat, S., Owen, C. J., Samoylova, O., Kreuzer, K. A., Stilgenbauer, S., Dohner, H., Langerak, A. W., Ritgen, M., Kneba, M., Asikanius, E., Humphrey, K., Wenger, M. and Hallek, M. (2014). Obinutuzumab plus chlorambucil in patients with CLL and coexisting conditions. *N. Engl. J. Med.* **370** 1101–1110.
- ▶ Lee, H. Z., Miller, B. W., Kwitkowski, V. E., Ricci, S., DelValle, P., Saber, H., Grillo, J., Bullock, J., Florian, J., Mehrotra, N., Ko, C. W., Nie, L., Shapiro, M., Tolnay, M., Kane, R. C., Kaminskas, E., Justice, R., Farrell, A. T. and Pazdur, R. (2014). U.s. Food and drug administration approval: obinutuzumab in combination with chlorambucil for the treatment of previously untreated chronic lymphocytic leukemia. *Clin. Cancer Res.* **20** 3902–3907.

Backup

CLL11 design

Primary endpoint: **progression-free survival** (PFS).



Assumptions

Global significance level: $\alpha = 0.05$.

Assumed effect sizes:

- $HR_{G/C} = 12/27 = 0.444$,
- $HR_{R/C} = 12/20 = 0.600$,
- $HR_{G/R} = 20/27 = 0.741$.

Assumptions:

- $n = 640$ patients in each strategy.
- Randomize 1:2:2.
- 20pts/m for 2m, 40pts/m for 15m.

Methods

Strategies 1, 2:

- **Compute** number of necessary events.
- **Compute** cutoffs for analyses based on that.

Strategies 3, 4:

- **Unadjusted** analysis: **Compute** number of necessary events and cutoff.
- **Adjusted** analysis: Global test gates pairwise tests. Increase number of necessary events from unadjusted analysis until **simulations** (10^6 runs) yield targeted power.

Analysis cutoffs

			G vs. C	R vs. C	G vs. R
Hazard ratio			0.444	0.600	0.741
Strategy 1: Three separate trials		computed #required events computed cutoff (months)	111 34.4	136 39.2	349 –
Strategy 2: 3-arm with Bonferroni		computed #required events computed cutoff (months)	136 21.4	181 24.3	465 90.1
Strategy 3: 3-arm with closed testing	unadj.	computed #required events computed cutoff (months)	275 47.2	303 47.2	349 47.2
	adj.	ass. (G vs. R)/resulting (R/G vs. C) #events cutoff (months) corresponding to #events	276 47.4	303 47.4	350 47.4
Strategy 4: 3-arm with closed testing	unadj.	computed #required events computed cutoff (months)	111 18.6	136 19.4	349 47.2
		assumed #required events cutoff (months) corresponding to #events	111 18.6	136 19.4	366 51.0
	power	simulated power corresponding to #events	0.974	0.807	0.800
		simulated unadj. power corresp. to #events	0.988	0.809	0.817

Patients for each comparison:

- Strategy 1: 64/128; 64/128; 128/128.
- Strategies 2-4: 128/256; 128/256; 256/256.

Results - power loss

Detailed results in backups.

			G vs. C	R vs. C	G vs. R
Hazard ratio			0.444	0.600	0.741
Strategy 1: Three separate trials		computed #required events computed cutoff (months)	111 34.4	136 39.2	349 –
Strategy 2: 3-arm with Bonferroni		computed #required events computed cutoff (months)	136 21.4	181 24.3	465 90.1
Strategy 3: 3-arm with closed testing	unadj.	computed #required events computed cutoff (months)	275 47.2	303 47.2	349 47.2
	adj.	ass. (G vs. R)/resulting (R/G vs. C) #events cutoff (months) corresponding to #events	276 47.4	303 47.4	350 47.4
Strategy 4: 3-arm with closed testing	unadj.	computed #required events computed cutoff (months)	111 18.6	136 19.4	349 47.2
	adj.	assumed #required events cutoff (months) corresponding to #events	111 18.6	136 19.4	366 51.0
	power	simulated power corresponding to #events simulated unadj. power corresp. to #events	0.974 0.988	0.807 0.809	0.800 0.817

Patients for each comparison:

- Strategy 1: 64/128; 64/128; 128/128.
- Strategies 2-4: 128/256; 128/256; 256/256.

Results

Results: with CLL11 strategy,

- save between $\sim 3\text{m}$ and $\sim 29\text{m}$ to first cutoff,
- $\sim 2\%$ power loss for G vs. R, corresponding to 17 events or $\sim 4\text{m}$.

Explore strategy based on closed testing in multi-arm trials.

Paper compares strategies with respect to

- operational complexity,
- operational bias,
- difficulty of inference in pairwise comparisons,
- type I error protection for secondary endpoints.
- Sensitivity analysis: CLL11 assumed quite large effect sizes. Strategy also feasible for smaller effect sizes?

Operational aspects in CLL11

Operational bias: Information from ongoing CT causes changes to participant pool, investigator or patient behavior, or other clinical aspects that affect conduct such that conclusions about efficacy or safety are impacted by differences in data collected post public availability of interim results.

CLL11:

- G vs. C became available quickly.
- Treatment schedule in CLL11 rather fixed once started.
- Define analysis timepoints not only through PFS cutoffs: e.g. all patients needed to be randomized to G prior to cutoff for G vs. C.

Further operational aspects:

- Multiple final / interim analyses on different sets of patients.
- iDMC for interim analyses in G vs. R.
- Independent response review: even more important after G vs. C was unblinded.

R version and packages used to generate these slides:

R version: R version 4.4.1 (2024-06-14 ucrt)

Base packages: stats / graphics / grDevices / utils / datasets / methods / base

Other packages: reporttools / xtable

This document was generated on 2024-09-05 at 09:37:30.