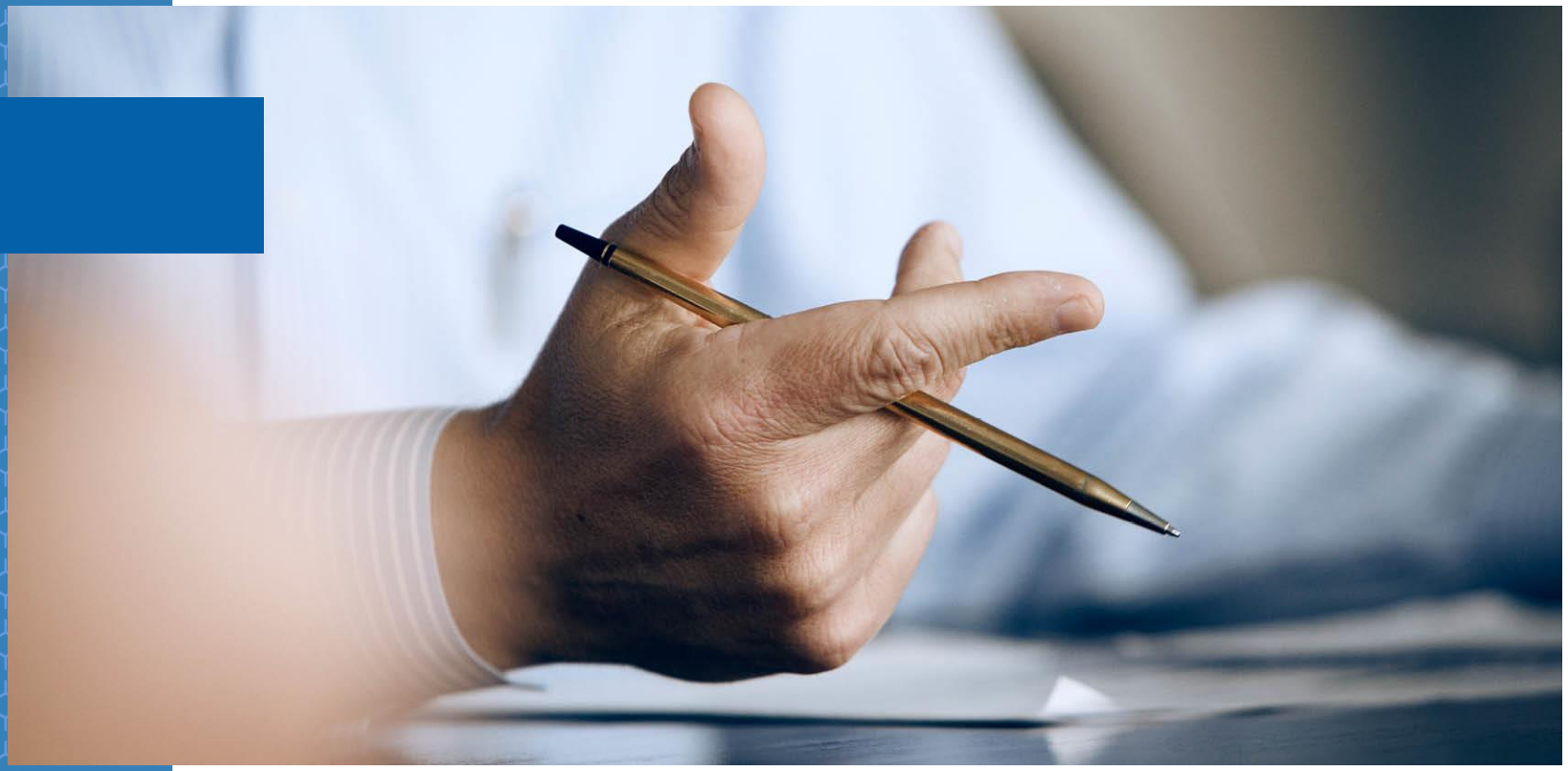


Analysis considerations of phase 3 clinical trials involving CAR-T – Exploring the use of principal stratum strategy

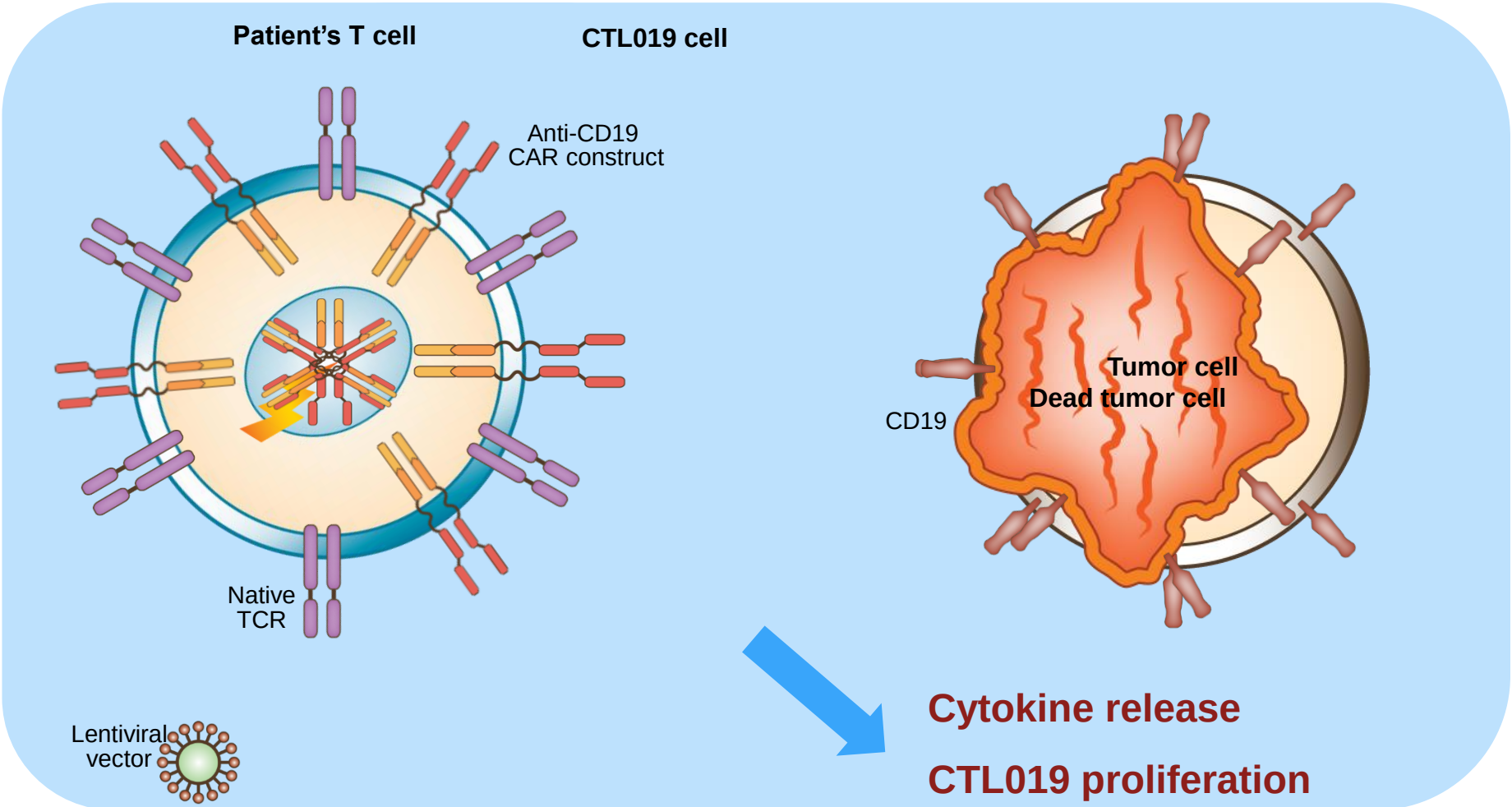
Yuanbo Song, on behalf of Novartis Team

2019 ASA Biopharmaceutical Section
Regulatory-Industry Statistics Workshop
Sep 24, 2019

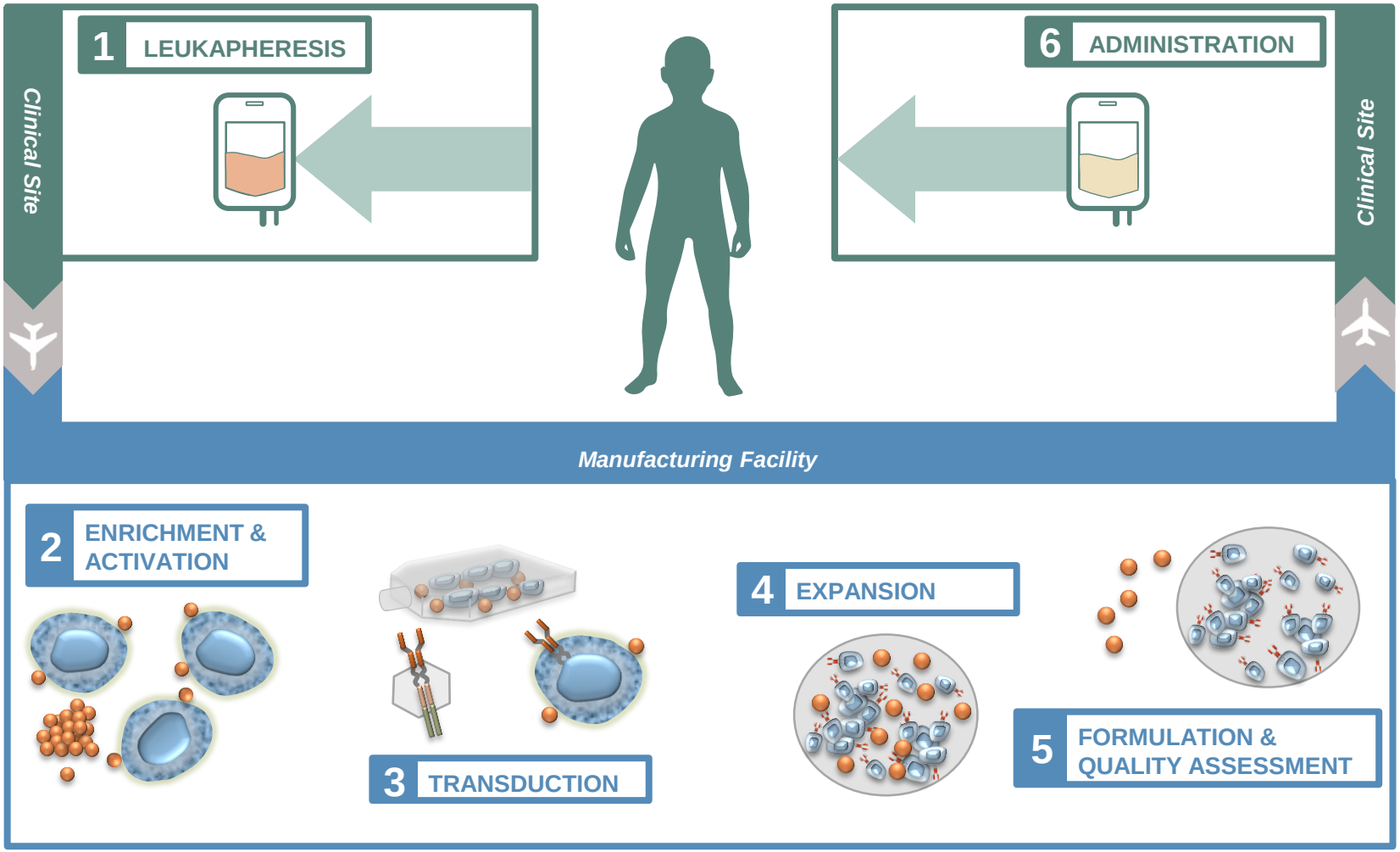
NOVARTIS

Kymriah (CTL019) – Chimeric antigen receptor T cell (CAR-T) therapy

A living drug designed to target CD19+ B cells



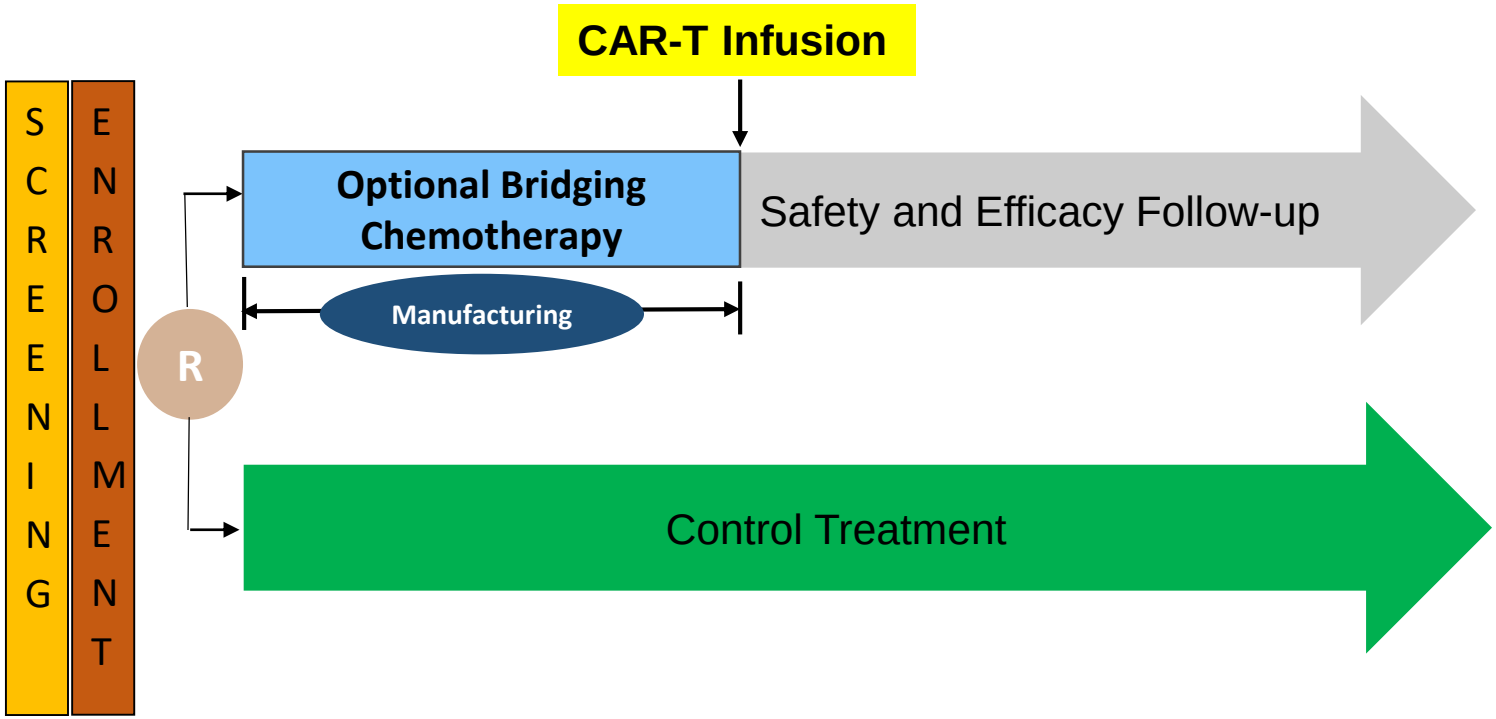
CAR-T cell manufacturing process



NOVARTIS

CAR-T randomized phase III trial

General study design



NOVARTIS

Defining the treatment effect

Challenges and Solutions

- CAR-T treatment not readily available at randomization:
  - Patients in CAR-T arm may take bridging therapy and tumor could respond to the therapy before receiving CAR-T
  - Patients may discontinue from the study prior to CAR-T infusion due to patient's condition or manufacturing related reasons.
- Treatment strategy:
  - CAR-T treatment strategy: BG chemotherapy + LD chemotherapy + CAR-T
  - Control treatment strategy: Standard of Care
- Primary objective
  - Compare the long-term survival of the two treatment strategies regardless of a patient's response to bridging chemotherapy

How to compare treatment effect by response status to bridging therapy?

Question of interest (Secondary):

- What is the long term survival of the CAR-T relative to control in patients who would be in (or not in) remission after bridging if they were given the bridging chemotherapy?

Denote:

- **S(T) as potential survival outcome for treatment T (T = 0: control; T = 1: CAR-T)**
- **R(1) as potential outcome for response status to bridging chemotherapy in CAR-T arm (R = 0: non-responder; R = 1: responder).**

Interest in contrasting the distribution of:

- **Survival for stratum of patients who did not respond to bridging: {S(T=1)|R(1)=0} vs. {S(T=0)|R(1)=0} or;**
- **Survival for stratum of patients who responded to bridging: {S(T=1)|R(1)=1} vs. {S(T=0)|R(1)=1}**

with hazard ratio as the effect measure

ICH E9 (R1)* – Principal stratum strategy

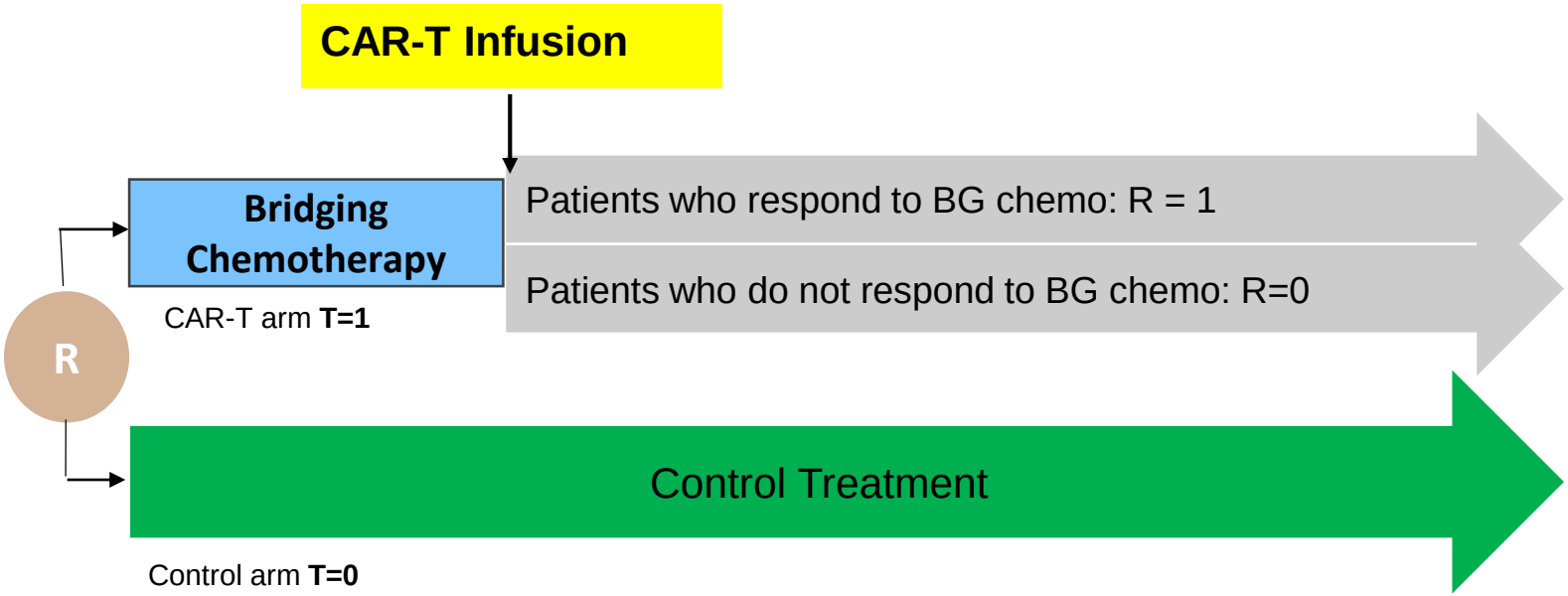
Principal stratum strategy

The target population might be taken to be the principal stratum (see Glossary) in which an intercurrent event would not occur. For example, the target population of interest might be taken to be the stratum of patients in which failure to adhere to treatment would not occur. In other words, a principal stratum is a subset of the broader population who would not experience the intercurrent event. The scientific question of interest relates to the treatment effect only within that stratum.

Effects in principal strata should be clearly distinguished from any type of subgroup or per-protocol analyses where membership is based on the trial data. Principal stratification (see Glossary) is defined by a patient's potential intercurrent events on both treatments: for example, patients who would adhere to either treatment. It is not possible in general to identify these subjects directly, either in advance of the trial since the occurrence of the intercurrent event cannot be predicted, or based on the data from a randomised controlled trial because each patient will be observed on one treatment only.

*Estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials

Response status to bridging therapy

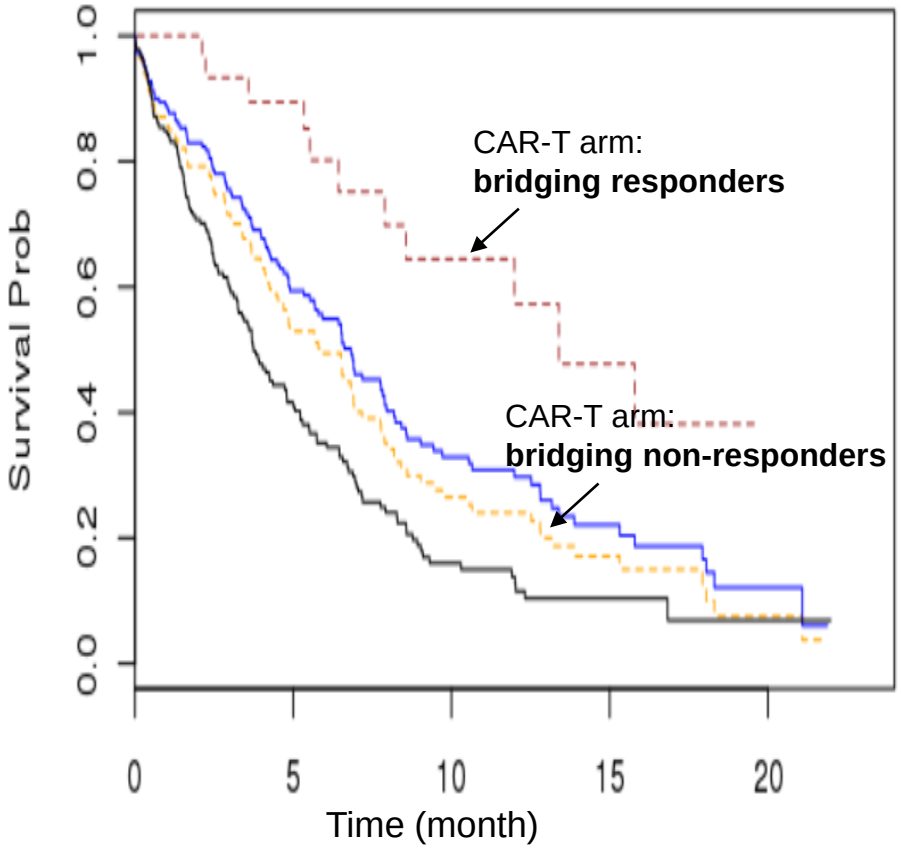


At the end of bridging,

- For CAR-T arm patients, R is observed (either R=1 or R=0).
- For control arm patients, **R is unknown**.

A “naive” comparison

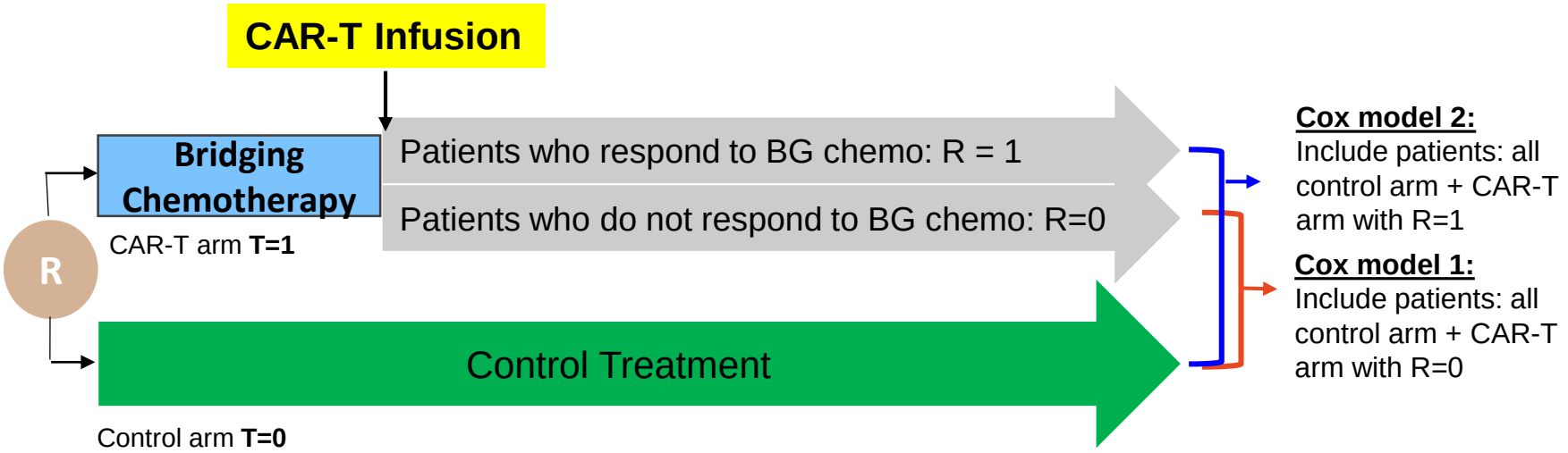
KM curve for stratum of pts responded or not responded to BG chemotherapy in CAR-T arm vs. KM for the control arm



Comparison 1:
Bridging responders from CAR-T arm vs. Control arm (red curve vs. black curve)

Comparison 2:
Bridging non-responders from CAR-T arm vs. Control arm (yellow curve vs. black curve)

Principal stratum 1: the “Independent” approach



- The naive comparisons provide unbiased estimate of treatment effect if **S(T=0) and R are independent**.
  - All patients in control arm share the same survival distribution regardless of their response status

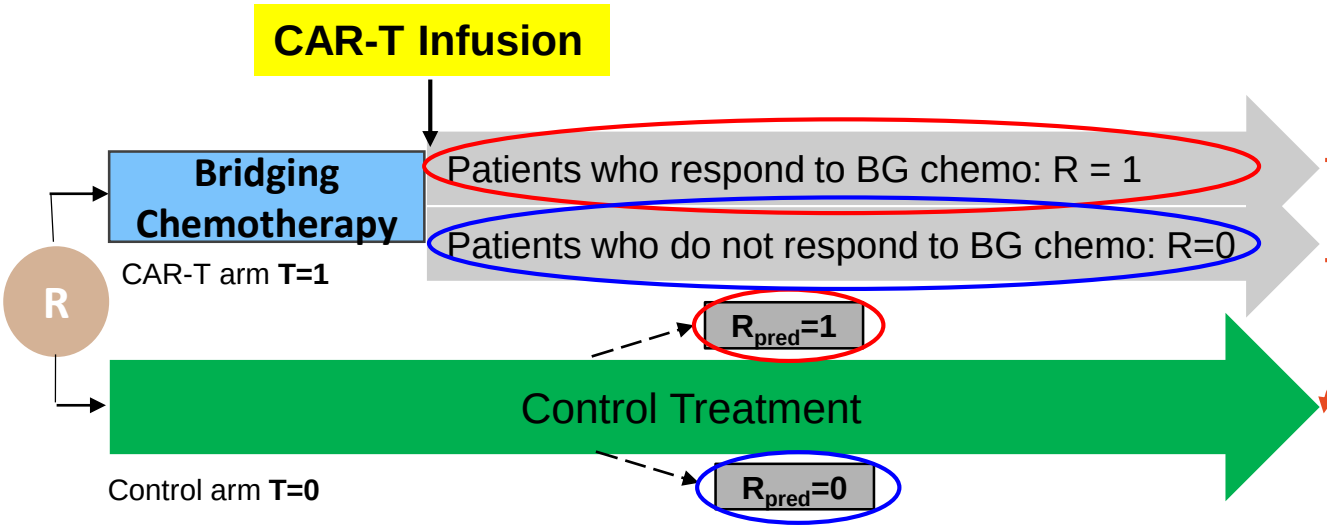
Utilizing covariates (1)

- R and S(0) dependent
  - Populations with R=1 and R=0 differ in their outcome S(0)
  - Assess effect of treatment and population (not a causal effect)

- Assume all covariates X affecting S(0) are known
  - If X is independent of R → R ⊥ S(0) → Can use the naive comparison
  - If all (or some) X are related to R
    → X unbalanced in populations with R=0 and R=1
    → cannot use naive comparison

Principal stratum 2: the "Multiple imputation (MI)" approach

The idea of the MI approach is to **predict a patient's response to bridging** for those randomized to the control arm (as it's not observable)



1. Fit logistic regression model using data from CAR-T arm: R ~ covariates

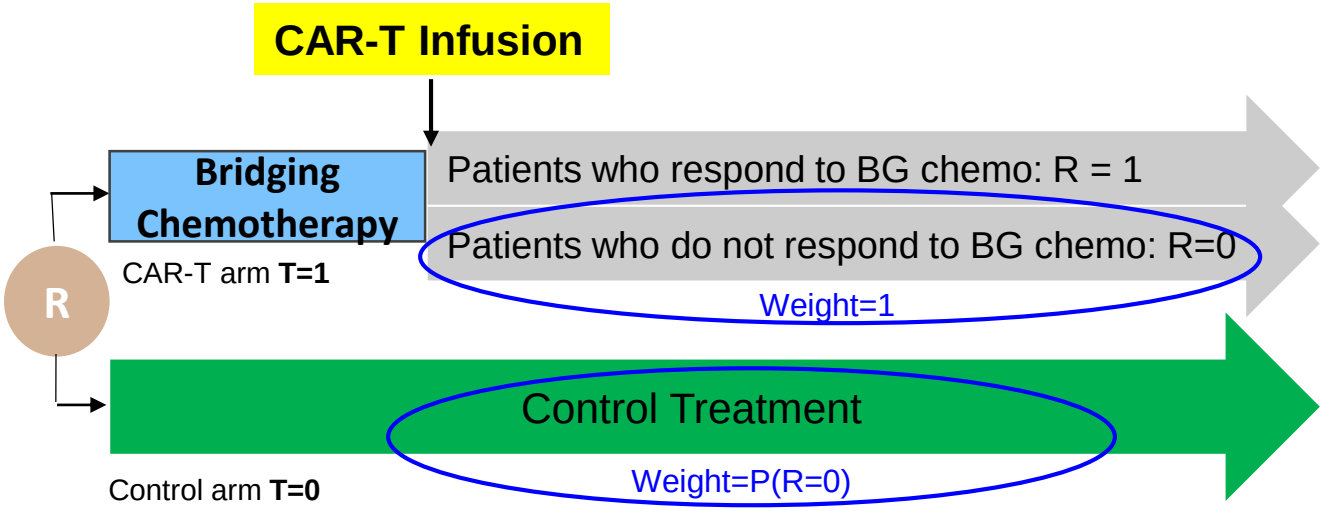
2. Apply the fitted model to control arm and obtain P(R=0|covariates) or P(R=1|covariates)

3. Generate R_{pred} from Bernoulli distribution

4. Fit Cox regression models:
Cox model 1: Estimate HR of CAR-T vs. Control for non-BG responders include patients- CAR-T arm (R=0) + control arm (R_{pred}=0)
Cox model 2: Estimate HR of CAR-T vs. Control for BG responders include patients- CAR-T arm (R=1) + control arm (R_{pred}=1)
5. Combine results with Rubin's rule

Principal stratum 3: the "Weighted" approach

Instead of predicting a patient's response to bridging for those in the control arm, the **probability of being a responder or non-responder to BG** will be used as weight in the **weighted cox regression model**.



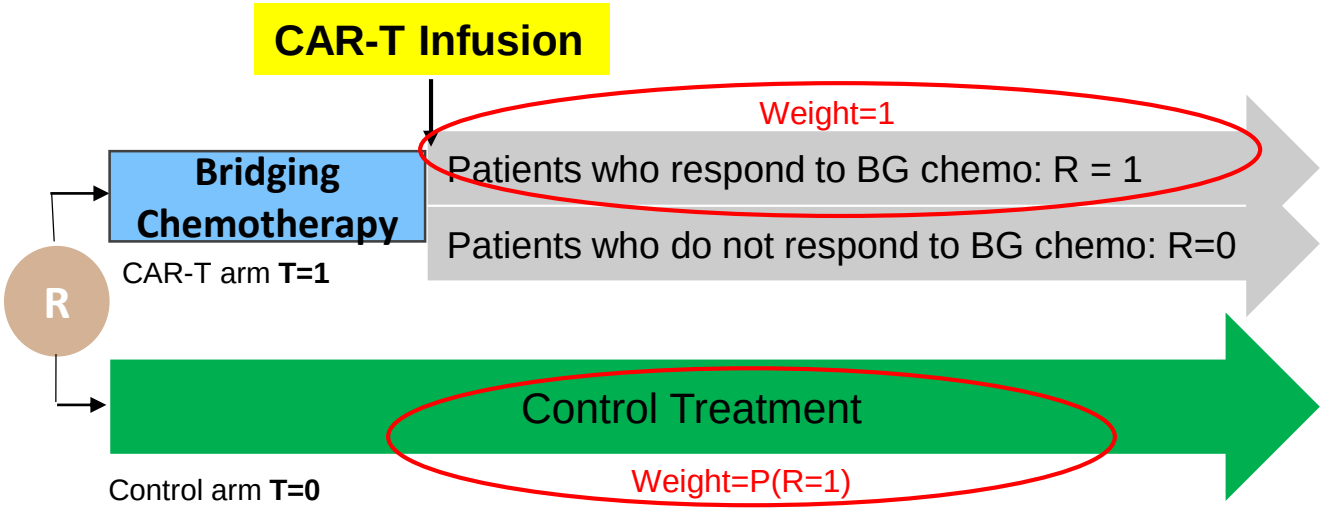
1. Fit logistic regression model using data from CAR-T arm:
R (binary 1 vs. 0) ~ covariates

2. Apply the fitted model to control arm and predict a patient's probability of responding/not responding to BG: P(R=1) (or P(R=0))

3. Fit weighted Cox regression models:
Cox model 1: Estimate HR of CAR-T vs. Control for non-BG responders
include patients- CAR-T arm (R=0, weight=1) + all control arm (weight=P(R=0))

Principal stratum 3: the "Weighted" approach

Instead of predicting a patient's response to bridging for those in the control arm, the **probability of being a responder or non-responder to BG** will be used as weight in the **weighted cox regression model**.



1. Fit logistic regression model using data from CTL arm:
R (binary 1 vs. 0) ~ covariates

2. Apply the fitted model to control arm and predict a patient's probability of responding/not responding to BG: P(R=1) (or P(R=0))

3. Fit weighted Cox regression models:
Cox model 1: Estimate HR of CAR-T vs. Control for non-BG responders
include patients- CAR-T arm (R=0, weight=1) + all control arm (weight=P(R=0))
Cox model 2: Estimate HR of CTL vs. SOC for BG responders
include patients- CTL arm (R=1, weight=1) + all control arm (weight=P(R=1))

Utilizing covariates (2)

- Achieve balance of X by
  - Building a model to predict R using X in CAR-T arm. This can be used to predict R for patients in control arm
  - The distribution of X for bridging responders (R=1) in CAR-T arm is then similar to the distribution of X in the weighted (by P(R=1)) population on control
    - Or equivalently the distribution of X for imputed (R=1) responders in the control arm
- Notes
  - In theory, only variables X affecting S(0) and R need to be utilized. In practice, predictors of R might not be well known
    - Utilize covariates X that affect S(0) and as can plausibly also affect R.
  - Formal assumption is that $S(0) \perp R \mid X$ (conditional independence)

Simulation

Scenario 1 - setting

- Simulate R for **CAR-T arm** and **Control arm** using logistic model: BG response (R) ~ baseline tumor burden (X)
  - X ~ lognormal (μ, σ^2), where μ and σ were from historical study
  - R ~ Bernoulli (P(R=1|X)): R = 0 (70%) vs. R = 1 (30%)
  - AUC under ROC curve: 0.60, 0.72, 0.90
- Simulate survival (S) for the two arms using baseline tumor burden (X)

➔ **As R and S are both generated as a function of X, they are conditionally independent.**

- Other parameters
  - N: Total 300 patients with 1:1 randomization
  - Recruitment: staggered enrollment with 15 pts/months after 3 months.
  - Drop-out: 2-year drop-out rate ~15%

Simulation

Scenario 1 - results

* results averaged across 500 simulations

	AUC under ROC curve	Bias in the estimated HR from observed **BT-responders**			Bias in the estimated HR from observed **Non BT-responders**		
		IND	MI	Weighted	**IND**	MI	Weighted
Sim 1	*0.60*	**-0.040 (SE=0.215)**	-0.007 (SE=0.285)	-0.006 (SE=0.248)	**0.029 (SE=0.170)**	<0.001 (SE=0.205)	0.001 (SE=0.184)
Sim 2	*0.72*	**-0.084 (SE=0.218)**	-0.005 (SE=0.279)	-0.005 (SE=0.250)	**0.065 (SE=0.169)**	<0.001 (SE=0.208)	0.001 (SE=0.184)
Sim 3	*0.90*	**-0.132 (SE=0.215)**	-0.001 (SE=0.258)	-0.002 (SE=0.244)	**0.138 (SE=0.170)**	-0.001 (SE=0.213)	<0.001 (SE=0.189)

- IND approach: Bias decreases as model becomes less predictive
  - Less predictive → Less dependent between R and S
- MI/Weighted approach: Bias could be ignored regardless of model's predictability
  - Good prediction of R is not required, as long as the conditional independence holds for given covariates.

Simulation

Scenario 1 - results

* results averaged across 500 simulations

	AUC under ROC curve	Bias in the estimated HR from observed **BT-responders**			Bias in the estimated HR from observed **Non BT-responders**		
		IND	**MI**	**Weighted**	IND	**MI**	**Weighted**
Sim 1	*0.60*	-0.040 (SE=0.215)	**-0.007 (SE=0.285)**	**-0.006 (SE=0.248)**	0.029 (SE=0.170)	**<0.001 (SE=0.205)**	**0.001 (SE=0.184)**
Sim 2	*0.72*	-0.084 (SE=0.218)	**-0.005 (SE=0.279)**	**-0.005 (SE=0.250)**	0.065 (SE=0.169)	**<0.001 (SE=0.208)**	**0.001 (SE=0.184)**
Sim 3	*0.90*	-0.132 (SE=0.215)	**-0.001 (SE=0.258)**	**-0.002 (SE=0.244)**	0.138 (SE=0.170)	**-0.001 (SE=0.213)**	**<0.001 (SE=0.189)**

- IND approach: Bias decreases as model becomes less predictive
  - Less predictive → Less dependent between R and S
- MI/Weighted approach: Bias could be ignored regardless of model's predictability
  - Good prediction of R is not required, as long as the conditional independence holds for given covariates.

Simulation

Scenario 2 - Setting

- Simulate R for CAR-T arm and Control arm using logistic model: BG response (R) ~ baseline tumor burden (X)
  - X ~ lognormal (μ, σ^2), where μ and σ were from historical study
  - R ~ Bernoulli (P(R=1|X)): R = 0 (70%) vs. R = 1 (30%)
  - AUC under ROC curve: 0.60, 0.72, 0.90
- **Simulate survival (S) for the two arms based on patients' response status (R) to bridging**
- Other parameters
  - N: Total 300 patients with 1:1 randomization
  - Recruitment: staggered enrollment with 15 pts/months after 3 months.
  - Drop-out: 2-year drop-out rate ~15%

Simulation

Scenario 2 - results

* results averaged across 500 simulations

	AUC under ROC curve	Bias in the estimated HR from observed **BT-responders**			Bias in the estimated HR from observed **Non BT-responders**		
		IND	MI	Weighted	**IND**	MI	Weighted
Sim 1	*0.60*	**-0.059 (SE=0.203)**	-0.059 (SE=0.269)	-0.059 (SE=0.236)	**0.067 (SE=149)**	0.065 (SE=0.179)	0.065 (SE=0.164)
Sim 2	*0.72*	**-0.059 (SE=0.201)**	-0.052 (SE=0.261)	-0.052 (SE=0.233)	**0.069 (SE=0.150)**	0.060 (SE=0.180)	0.059 (SE=0.166)
Sim 3	*0.90*	**-0.053 (SE=0.191)**	-0.024 (0.234)	-0.024 (SE=0.220)	**0.076 (SE=0.153)**	0.042 (SE=0.182)	0.042 (SE=0.172)

- IND approach: Biases are large yet similar across all three models
  - Predictability has little impact on bias as S is fully dependent on R.
- MI/Weighted approach: HR estimates become closer to the truth, as model becomes more predictive
  - Non-negligible bias still exists even when AUC=0.90

Simulation

Scenario 2 - results

* results averaged across 500 simulations

	AUC under ROC curve	Bias in the estimated HR from observed **BT-responders**			Bias in the estimated HR from observed **Non BT-responders**		
		IND	**MI**	**Weighted**	IND	**MI**	**Weighted**
Sim 1	*0.60*	-0.059 (SE=0.203)	**-0.059 (SE=0.269)**	**-0.059 (SE=0.236)**	0.067 (SE=149)	**0.065 (SE=0.179)**	**0.065 (SE=0.164)**
Sim 2	*0.72*	-0.059 (SE=0.201)	**-0.052 (SE=0.261)**	**-0.052 (SE=0.233)**	0.069 (SE=0.150)	**0.060 (SE=0.180)**	**0.059 (SE=0.166)**
Sim 3	*0.90*	-0.053 (SE=0.191)	**-0.024 (0.234)**	**-0.024 (SE=0.220)**	0.079 (SE=0.153)	**0.042 (SE=0.182)**	**0.042 (SE=0.172)**

- IND approach: Biases are large yet similar across all three models
  - Predictability has little impact on bias as S is fully dependent on R.
- MI/Weighted approach: HR estimates become closer to the truth, as model becomes more predictive
  - Non-negligible bias still exists even when AUC=0.90

Summary

- Patients in CAR-T arm may take and respond to BG chemotherapy:
  - Primary: compare the long term survival of the two treatment strategies
  - Secondary: treatment effect in patients who respond/not respond to BG
- Challenge of principal stratum strategy: identify all X that relate to S and/or R. A pragmatic strategy for selection of X
  - Include all important covariates X predictive of S
    - If none of X relates to R → Independent Approach
    - If all (or some) of X relates to R → MI/Weighted approach
  - If no predictors of S are known or conditional independence not hold
    - Perfect prediction of R is needed to get unbiased estimates
- Implementation of principal stratum strategy
  - Not common for primary but useful for secondary or exploratory analysis

Acknowledgment

- Aby Buchbinder
- Aiesha Zia
- Amy Racine
- Antonella Maniero
- Bjoern Bornkamp
- Ekkehard Glimm
- Emmanuel Zuber
- Eric Bleickardt
- Evgeny Degtyarev
- Frank Bretz
- Jessie Gu
- Kapildeb Sen
- Oezlem Anak
- Yiyun Zhang

Thank you

NOVARTIS