#### Sensitivity Analysis vs Supportive Analysis under Estimand Framework: A Case Study in Hematological Malignancies

Steven Sun (Johnson & Johnson), Hans-Jochen Weber (Novartis), Marie-Laure Casadebaig (Celgene), Emily Butler (GlaxoSmithKline), Satrajit Roychoudhury (Pfizer), Kaspar Rufibach (Roche), Viktoriya Stalbovskaya (Merus)



#### **Outlines**

- Impact of estimand framework on trial analysis
- Motivating example: GALLIUM study
- Choice of sensitivity analysis and supplementary analysis

#### **Impact of Estimand Framework on Trial Analysis**



# **Sensitivity Analysis**

 Sensitivity analysis: is a series of analyses targeting the same estimand, with differing assumptions to explore the robustness of inferences from the main estimator to <u>deviations from its</u> <u>underlying modelling assumptions</u> and <u>limitations in</u> <u>the data</u>.

### **Supplementary Analysis**

 Supplementary analysis: is a general description for analyses that are conducted in addition to the main and sensitivity analysis to provide additional insights into the understanding of the treatment effect. The term describes a broader class of analyses than sensitivity analyses.

#### **Consequence of Misclassifications of Sensitivity and Supplementary Analysis**

- Often too many sensitivity analyses are planned in the SAP
  - Analysis on different populations (per-protocol population, response-evaluable population, etc)
  - Covariate-adjusted analyses (multivariate analyses)
  - Different censoring schemes
  - Multiple imputations of missing data
- No clear estimand targeted. Some results are likely inconstant with others
  - Interpretation difficult

#### A Case Study --

#### **GALLIUM study**

#### The NEW ENGLAND JOURNAL of MEDICINE

#### ORIGINAL ARTICLE

#### Obinutuzumab for the First-Line Treatment of Follicular Lymphoma

R. Marcus, A. Davies, K. Ando, W. Klapper, S. Opat, C. Owen, E. Phillips, R. Sangha, R. Schlag, J.F. Seymour, W. Townsend, M. Trněný, M. Wenger, G. Fingerle-Rowson, K. Rufibach, T. Moore, M. Herold, and W. Hiddemann

#### ABSTRACT

#### BACKGROUND

Rituximab-based immunochemotherapy has improved outcomes in patients with follicular lymphoma. Obinutuzumab is a glycoengineered type II anti-CD20 monoclonal antibody. We compared rituximab-based chemotherapy with obinutuzumabbased chemotherapy in patients with previously untreated advanced-stage follicular lymphoma.

#### METHODS

We randomly assigned patients to undergo induction treatment with obinutuzumabbased chemotherapy or rituximab-based chemotherapy. Patients with a response received maintenance treatment for up to 2 years with the same antibody that they had received in induction. The primary end point was investigator-assessed progressionfree survival.

#### RESULTS

A total of 1202 patients with follicular lymphoma underwent randomization (601 patients in each group). After a median follow-up of 34.5 months (range, 0 to 54.5), a planned interim analysis showed that obinutuzumab-based chemotherapy resulted in a significantly lower risk of progression, relapse, or death than rituximab-based chemotherapy (estimated 3-year rate of progression-free survival, 80.0% vs. 73.3%; hazard ratio for progression, relapse, or death, 0.66; 95% confidence interval [CI], 0.51 to 0.85; P=0.001). Similar results were seen with regard to independently reviewed progression-free survival and other time-to-event end points. Response rates were similar in the two groups (88.5% in the obinutuzumab group and 86.9% in the rituximab group). Adverse events of grade 3 to 5 were more frequent in the obinutuzumab group than in the rituximab group (74.6% vs. 67.8%), as were serious adverse events (46.1% vs. 39.9%). The rates of adverse events resulting in death were similar in the two groups (4.0% in the obinutuzumab group and 3.4% in the rituximab group). The most common adverse events were infusion-related events that were considered by the investigators to be largely due to obinutuzumab in 353 of 595 patients (59.3%; 95% CI, 55.3 to 63.2) and to rituximab in 292 of 597 patients (48.9%; 95% CI, 44.9 to 52.9; P<0.001). Nausea and neutropenia were common. A total of 35 patients (5.8%) in the obinutuzumab group and 46 (7.7%) in the rituximab group died.

#### CONCLUSIONS

Obinutuzumab-based immunochemotherapy and maintenance therapy resulted in longer progression-free survival than rituximab-based therapy. High-grade adverse events were more common with obinutuzumab-based chemotherapy. (Funded by F. Hoffmann–La Roche; GALLIUM ClinicalTrials.gov number, NCT01332968.)

#### **Clinical Trial Design (GALLIUM Study)**



<sup>\$</sup> Chemotherapy choice by site
 <sup>†</sup> Patients in SD enter observation phase for up to 2 years

Objective: To evaluate the efficacy of G-chemo followed by G-maintenance therapy compared with R-chemo followed by R-maintenance therapy in patients with previously untreated advanced follicular lymphoma (FL), as measured by investigator-assessed progression free survival (PFS).

#### **Common features of hematological studies**

- Multiple treatment phases
  - Induction
  - ASCT
  - Consolidation
  - Maintenance

Some patients can be cured

#### Challenges

- Is a common population level summary HR a good measurement for the treatment benefit?
  - Constant proportional hazard at two treatment phases?
  - Patients with stable disease won't get maintenance treatment
- How to isolate the treatment benefit in each phase (FDA's concern)?
  - Overall benefit may be driven by the induction phase only

### **Main Analysis for PFS**

 Stratified analysis (with stratification factors used in randomization) for investigators' assessed PFS without adjustment by other covariates for ITT FL patients

#### **Common Analyses for PFS – Sensitivity or Supplementary?**

- Tumor assessment: by independent review committee (IRC-PFS)
- unstratified analysis
- Stratification per eCRF
- Covariate-adjusted estimator
- Other populations (Per-protocol population, responseevaluable population)
- Different censoring schemes
  - Censoring at subsequent therapy
  - Worst case analysis for loss to follow-up
  - 2 or more consecutive missing assessment
- Maintenance as a time-dependent covariate for PFS Cox regression model

### **Inv-PFS vs IRC-PFS**

 Inv- and IRC-PFS are two estimators of the same estimand → one sensitivity of the other.

Type of potential bias	Inv-PFS	IRC-PFS
Knowledge of treatment assignment	In an open-label study, Investigator knows treatment assignment	Typically performed blinded to treatment assignment
Informative censoring through Inv-PD	Not applicable, i.e. no risk of bias	If PD is called by local assessment prior to IRC-PD, then scan collection is typically stopped, i.e. IRC-PFS will remain censored at date of Inv-PD.

#### **Inv-PFS vs IRC-PFS**

- In Gallium study, INV-PFS was used for the primary analysis, but results included in USPI are based on IRC-PFS
  - The IA boundary is different due to different fraction information
  - IRC-PFS primary endpoint for FDA → the study only had 218 events < then 245 events as pre-specified for IA, and p-value was \*above\* group-sequential boundary (planned boundary is p-value <0.012).</li>

INV-PFS: G-CT vs R-CT: HR = 0.66 , p-value = 0.0012 IRC-PFS: G-CT vs R-CT: HR = 0.71, p-value = 0.0138

### **Stratified vs. unstratified**

- Stratified Cox model: Distinct baseline hazard functions for each stratum, common hazard ratio across strata.
- Unstratified: Identical baseline hazard for each stratum.
- Same baseline hazard = modeling assumption → unstratified sensitivity of primary stratified estimator
  - Consistent with what have been done in the past

Stratified analysis: HR = 0.66, p-value = 0.0012 Unstratified analysis: HR = 0.66, p-value = 0.0013

#### **Stratification per eCRF**

- Stratification per CRF or IWRS should be considered as limitation of data
  - > one a sensitivity analysis of the other
- Discrepancies may reflect different technical assessment methods. And treatment balance within each stratum (per CRF) may no longer hold

> Stratified analysis (per CRF) is a supplementary analysis

# **Covariate-adjusted Analysis**

Estimand	Linear regression	Logistic regression	Cox regression	Aalen additive model
Unadjusted	Marginal	Marginal	Marginal	Marginal
Covariate-adjusted	Effect collapsible, i.e. marginal = conditional	Conditional	Conditional	Effect collapsible, i.e. marginal = conditional

- Marginal effect:
  - Average effect of **moving entire population from untreated to treated**.
  - Unadjusted estimate in RCT
- Conditional effect:
  - Average effect of treatment on individual, i.e. of moving a subject from untreated to treated.
  - Estimated from regression coefficient for treatment assignment indicator variable in multiple regression model.
- Do not routinely run adjusted and unadjusted analysis → they may target different estimand! One supplementary of the other for PFS analysis based on Cox regression model

#### **Per-protocol population vs ITT**

- ITT vs PP <u>analysis data set</u>
  - It depends on the definition of PP analysis dataset
  - Are both PP and ITT analysis datasets represent the target population of interest?
    - Usually PP analysis datasets include patients who meet all eligibility criteria, in this case, they are random samples of target population → sensitivity analysis
    - If PP analysis datasets include patients with certain conditions (e.g., receive at least 6 cycles of study drug), then they are not considered as random samples fof target population → supplementary analysis
- Recommendation: PP population is not useful for superiority study.
  - There was no analysis based on PP analysis set in Gallium study

# **ITT vs response evaluable population**

- Response evaluable population usually includes patients meeting certain criteria, which may depend on the outcome of treatment
  - They do not represent the target population defined in the study  $\rightarrow$  supplementary analysis

### **Different censoring schemes**

- Patients may cross-over or receive subsequent anticancer therapy before PD
  - FDA guideline recommends censoring patients at the last adequate disease assessment before subsequent therapy
    - Hypothetical strategy
  - EMA prefers using all data available regardless of subsequent therapies
    - Treatment policy strategy
  - Treating subsequent therapy as an event
    - Composite strategy
- Different strategies correspond to 'different estimand'
  → supplementary analysis is more appropriate

#### **Different censoring schemes**

- Worst case analysis for lost to follow-up
  - Treat loss to follow-up as an event for patients in treatment arm and censor it for patients in the control arm
- Is lost to follow-up an intercurrent event?
  - If so, then the the worst case analysis coresponds to composite strategy for patients in treatment arm and hypotheitical strategy in the control → target on different estimand: supplementary analysis!!
- Is Lost to follow-up considered as missing data (ICH guideline hinted so)
  - If so, then worst case analysis could be viewed as a sensitivity analysis
  - BUT, is the assumption logical??? Unlikely

#### **Different censoring schemes**

- Two or more consecutive missing assessment
  - Censor at the last adequate disease assessment prior to missing assessment
  - Limitation of data  $\rightarrow$  sensitivity analysis

### **Analyses to Address the Confounding Issue by Maintenance Therapy**

- During the filing of Gallium study, the question came up about the contribution of maintenance to the treatment effect
  - Is there additional benefit with G maintenance? If yes, What is the effect of G maintenance?
  - Is the benfit of G is same in both induction and maintennace?
  - Is the overall benfit driven by the mantenance only
- Supplementary analyses are needed to addess these questions
  - Targetting on different estimands

- Is there additional benefit with G maintenance? If yes, What is the effect of G maintenance?
  - To provide an unbiased estimate of maintenance effect size, a 2<sup>nd</sup> randomization at the time of maintenance is needed
  - With the current design, below analyses can indirectly check the benefit with G maintenance
    - PFS analysis by censoring patients at the time of maintenance
    - Analysis on time from maintenance start to PD for those who got maintenance therapy

- Is the benfit of G is same in both induction and maintenance?
  - Model diagnostics for constant hazard ratio (is this enough? Patients with SD assessment at the end of treatment won't receive maintenance therapy)

- Is the overall benefit driven by the maintenance only
  - Other endpoints can better characterize the benefit of G in the induction phase (PFS rate at the time of maintenance, ORR or CR rate in induction phase)
  - Proportion of patients received maintenance

- Maintenance as a time-dependent covariate for PFS Cox regression model
  - What is the corresponding estimand?
    - 4 attributes of an estimand are implicitly for a fixed treatment strategy
    - Treating maintenance as an confounding factor implies maintenance is not considered as part of treatment strategy

### **Discussions & Conclusions**

- Many common analyses performed should not be treated as sensitivity analyses in the Estimand framework
  - Reduce overall number of (unfocused) analyses.
- Supplementary analyses should be carefully selected to address the scientific questions to be answered after study completion
- Multiple estimands may be needed to align with a study objective