Methodological aspects and practical application of a drug quantitative benefit-risk assessment: a case study

PSI conference, 11 June 2025

Hua Liu, Zhaowei Hua (Servier)
Zhaoyang Teng, Sammi (Rui) Tang (Astellas Pharma)
Gaëlle Saint-Hilary (Saryga)

Presented by **Pavel Mozgunov** (MRC Biostatistics Unit, University of Cambridge and Saryga SAS, France)

SERVIER
moved by you

SARYGA
STATISTICS & METHODOLOGY

Introduction: the case-study

- **Ivosidenib (Tibsovo®)**, developed by **Servier**
  - First-in-class, oral, potent, selective inhibitor of mutated isocitrate dehydrogenase 1 (IDH1)
- Status of the development in 2021:
  - Ivosidenib was already indicated for the treatment of acute myeloid leukemia (AML)
  - **Proposed new indication**: Ivosidenib as a treatment for adult patients with previously treated, locally advanced or metastatic **cholangiocarcinoma (CCA)** with an IDH1 mutation
- ClarIDHy Phase 3 study key results:
  - Study met its primary endpoint with statistically significant PFS improvement (HR=0.37, 1-sided $p < 0.0001$)
  - OS improved numerically based on the ITT principle, and was further supported by RPSFT adjustment for the high-rate of crossover* (HR=0.49, 1-sided $p < 0.0001$)

→ A Quantitative Benefit-Risk Assessment (QBRA) was conducted to support future regulatory interaction

*>70% of crossovers

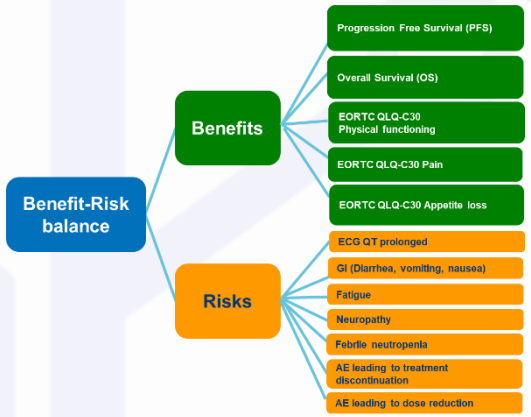
SERVIER moved by you

SARYGA STATISTICS & METHODOLOGY

Quantitative B-R assessment: general principles

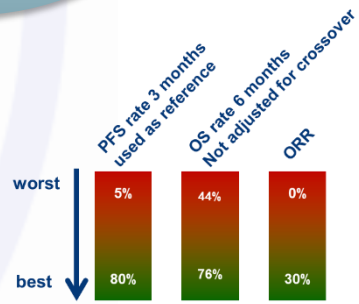
- Define Key Benefits and Risks
- Summarize Performance of treatments on each criterion

Performance of treatments on key B and R



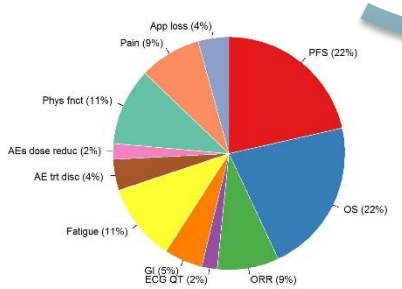
Make B and R variables comparable

- Map the performances on each criterion on a same scale



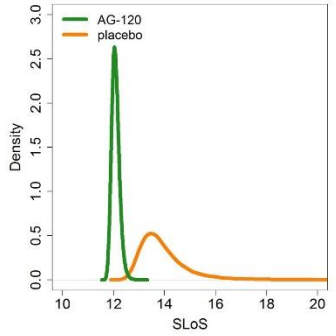
- Reflects the importance of the criteria

Weight elicitation



Benefit-Risk (utility) score

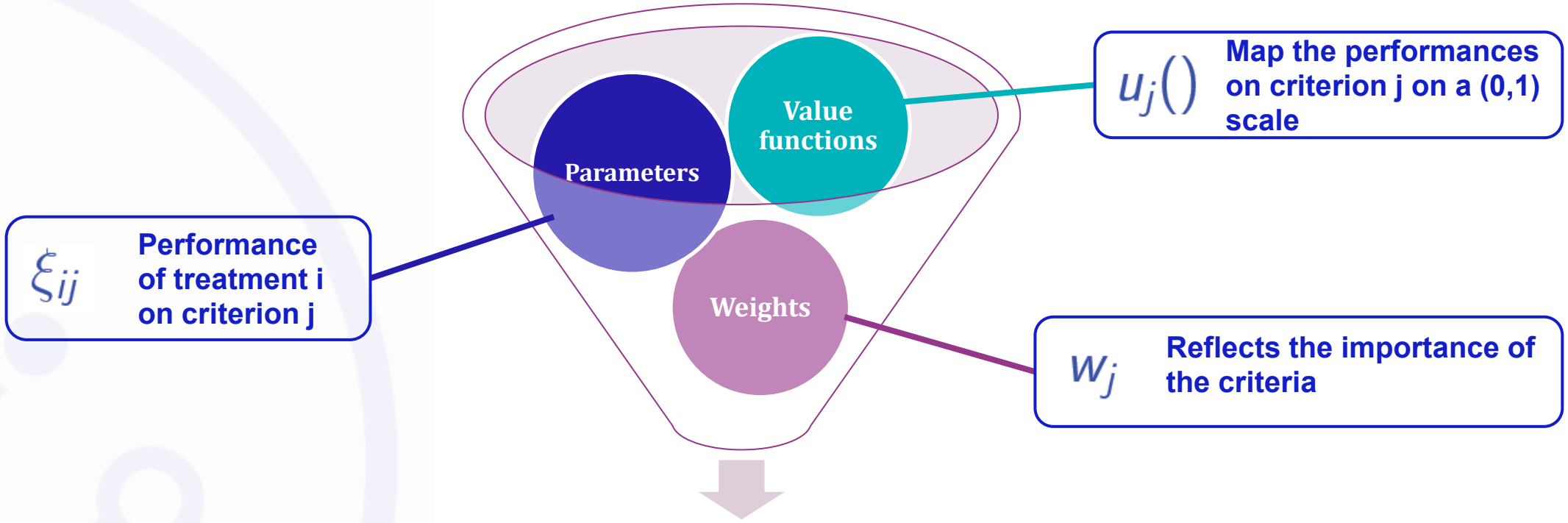
- Linear or non-linear model
- Taking into account uncertainties



SERVIER moved by you

SARYGA STATISTICS & METHODOLOGY

Quantitative B-R assessment: Multi-Criteria Decision Analysis (MCDA)



Example

Scale Loss Score (SLoS, non-linear)

Lower loss score → more preferable B-R balance

$$l(\xi_i, \boldsymbol{w}) := \sum_{j=1}^{n} u_j(\xi_{ij})^{-w_j}$$

Saint-Hilary et al (2019)

QBRA of Ivosidenib in CCA: panel of external KOLs

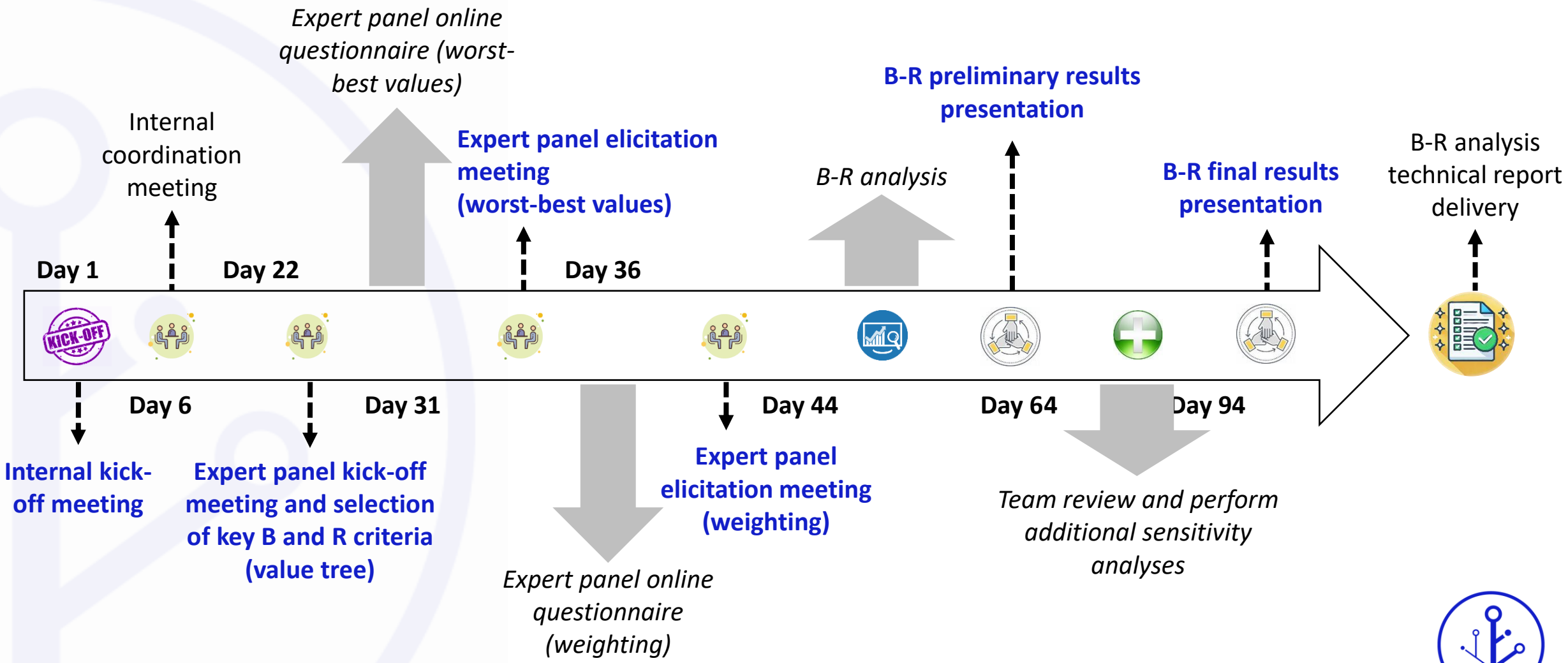
- A panel of 7 KOLs from different regions, **external to the Sponsor**, was identified to select the key benefit and risk criteria and to elicit the preference-dependent parameters of the models
- Highly engaged KOLs – attended 3 meetings and participated 2 online questionnaire in 2 weeks
- **External KOLs voting only**, reduced potential bias for B-R analyses

Faculty	Institution
Ghassan Abou-Alfa, MD	Memorial Sloan Kettering Cancer Center (NY, USA)
Maeve Lowery, MD	Trinity College Dublin (Ireland)
Milind Javle, MD	MD Anderson (TX, USA)
Kate Robin Kelley, MD	University of California San Francisco (CA, USA)
Rachna Shroff, MD	University of Arizona Cancer Center (AZ, USA)
Juan Valle, MD	University of Manchester / The Christie NHS Foundation Trust (UK)
Arndt Vogel, MD	Medizinische Hochschule Hannover (Germany)

SERVIER moved by you

SARYGA STATISTICS & METHODOLOGY

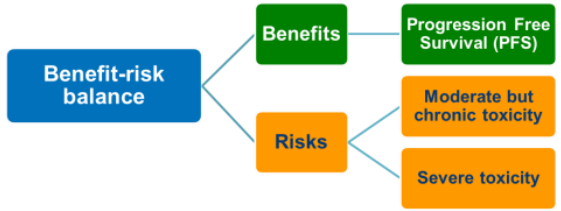
QBRA of Ivosidenib in CCA: Timelines



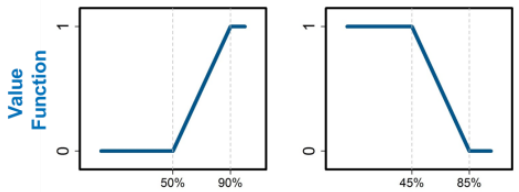
QBRA of Ivosidenib in CCA: Process

Voting by external KOLs only

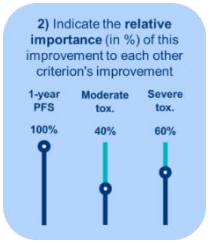
1. Select benefit and risk criteria → **value tree**



2. Elicit **best** and **worst** values



3. Elicit relative **weights**



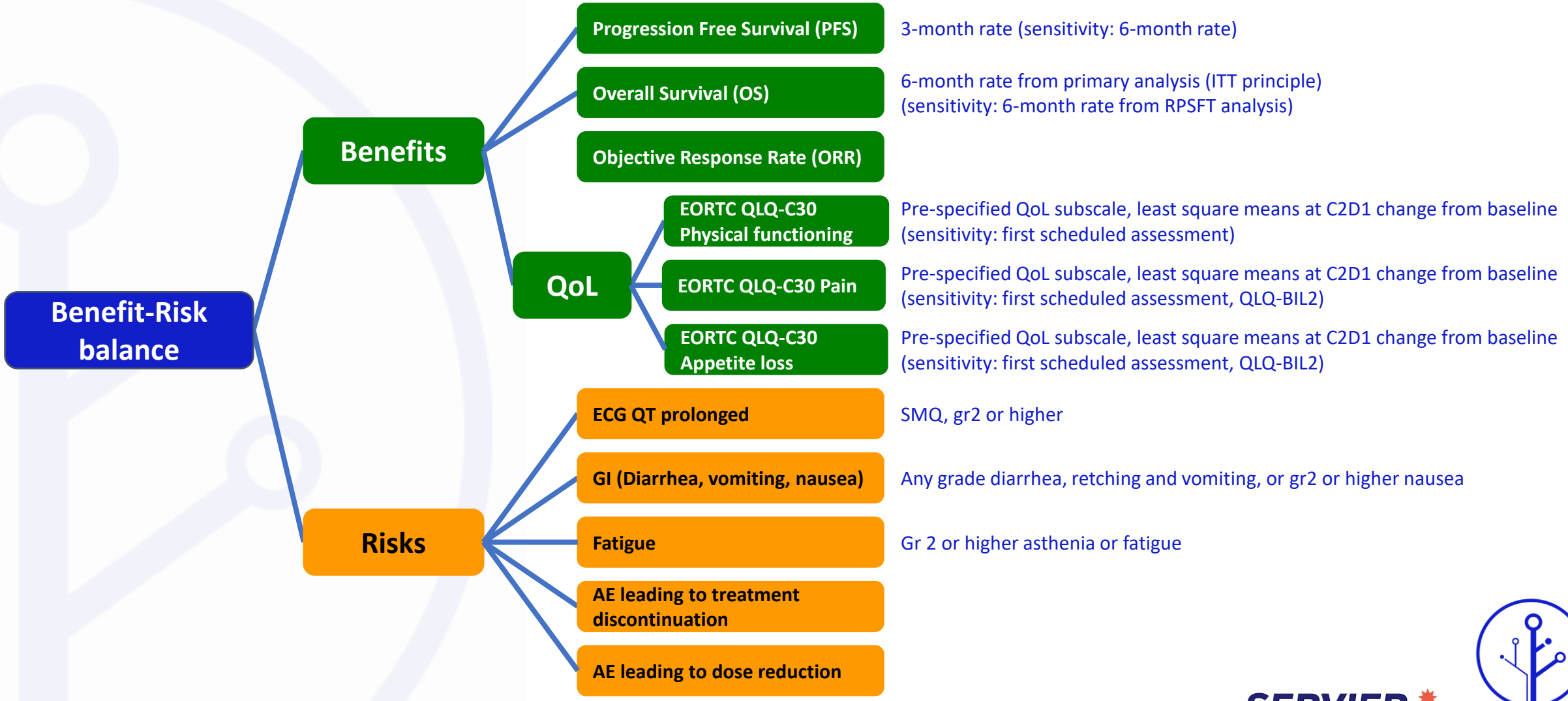
4. Run **quantitative benefit-risk analyses**
   - Main model (SLoS model) and sensitivity analyses (product model, linear model, random weights for all models)

5. Generate final technical report

SERVIER moved by you

SARYGA STATISTICS & METHODOLOGY

QBRA of Ivosidenib in CCA: Value Tree



SERVIER moved by you

SARYGA STATISTICS & METHODOLOGY

QBRA of Ivosidenib in CCA: Best-Worst values

Elicitation process

Slides presented during the elicitation meeting for each criterion

Worst and Best values – Ivosidenib CCA

11

Results of Phase 3 ClarIDHy | Best and Worst values

Criterion	placebo		ivosidenib		overall		Worst	Best
	N	Estimate [95% CI]	N	Estimate [95% CI]	N	Estimate [95% CI]		
Fatigue rate (%)	59	20.3% [11.0% ; 32.8%]	123	21.1% [14.3% ; 29.4%]	182	20.9% [15.2% ; 27.5%]		

Responses from the questionnaire:

Worst	Best
25%	5%
40%	10%
40%	10%
50%	20%
101%	25%

The **lower** the better for the patients

Grade 2 or higher
CI = Confidence Interval

Worst value
- Least desirable value for a treatment
- For benefits, it usually corresponds to what is expected without treatment (lower bound)
- From this value and below, the utility of the treatment in the patient population is at its minimum
- Values below this point may be possible, but are not realistic or are not more unfavorable for the patient population

Best value
- Most desirable value for a treatment
- For risks, it usually corresponds to what is expected without treatment (lower bound)
- From this value and beyond, the utility of the treatment in the patient population is at its maximum
- Values beyond this point may be possible, but are not realistic or do not really have an added value for the patient population

CentEX MVD

Results from the study and 95% CI

Individual responses from KOLs at a questionnaire sent prior to meeting, as a basis for discussion

SERVIER moved by you

SARYGA STATISTICS & METHODOLOGY

QBRA of Ivosidenib in CCA: Best-Worst values

Elicitation process

Slides presented during the elicitation meeting for each criterion

Worst and Best values – Ivosidenib CCA

11

Results of Phase 3 ClarIDHy | Best and Worst values

Criterion	placebo		ivosidenib		overall		Worst	Best
	N	Estimate [95% CI]	N	Estimate [95% CI]	N	Estimate [95% CI]		
Fatigue rate (%)	59	20.3% [11.0% ; 32.8%]	123	21.1% [14.3% ; 29.4%]	182	20.9% [15.2% ; 27.5%]	70%	10%

Consensus reached during the meeting

Responses from the questionnaire:

Worst	Best
25%	5%
40%	10%
40%	10%
50%	20%
101%	25%

The **lower** the better for the patients

Grade 2 or higher
CI = Confidence Interval

Worst value
- Least desirable value for a treatment
- For benefits, it usually corresponds to what is expected without treatment (lower bound)
- From this value and below, the utility of the treatment in the patient population is at its minimum
- Values below this point may be possible, but are not realistic or are not more unfavorable for the patient population

Best value
- Most desirable value for a treatment
- For risks, it usually corresponds to what is expected without treatment (lower bound)
- From this value and beyond, the utility of the treatment in the patient population is at its maximum
- Values beyond this point may be possible, but are not realistic or do not really have an added value for the patient population

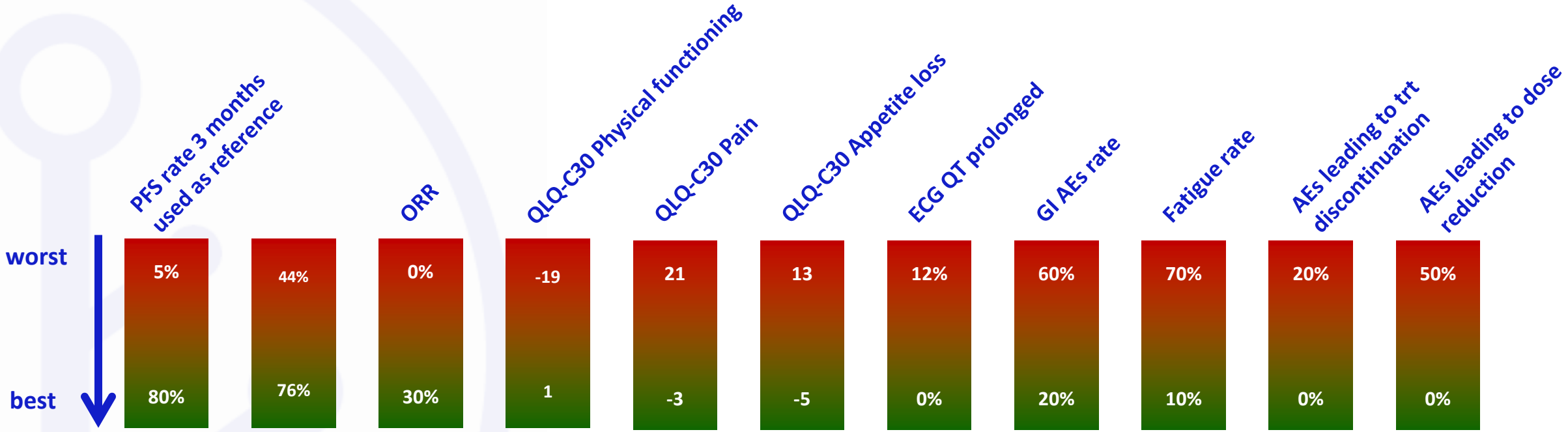
CentEX MVD

SERVIER moved by you

SARYGA STATISTICS & METHODOLOGY

QBRA of Ivosidenib in CCA: Best-Worst values

Finalized by the expert panel via elicitation meeting



And 35% (worst) to 76% (best) for OS rate 6 months adjusted for crossover

Best and Worst values have been rounded

SERVIER
moved by you

SARYGA
STATISTICS & METHODOLOGY

QBRA of Ivosidenib in CCA: Weights

Elicitation process (swing-weighting methodology)

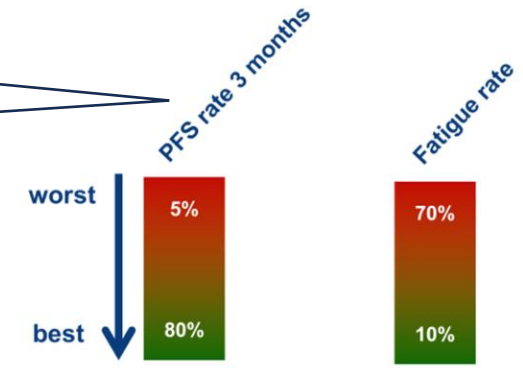
Slides presented during the elicitation meeting for each criterion

Ivosidenib vs placebo, 2nd or 3rd line IDH1m CCA population
Swing-weighting

Consider a treatment with:
- 70% patients with Fatigue (grade 2 or higher)
- 5% PFS rate at 3 months

Quantify the relative importance (in %) of:
- Decreasing the rate of Fatigue from 70% to 10%
- Compared to improving the PFS rate at 3 months from 5% to 80%

Criteria	Relative importance
PFS rate at 3 months	
Fatigue	



Responses from the questionnaire:

5%
25%
50%
50%
60%

Most important (primary) criterion

Individual responses from KOLs at a questionnaire sent prior to meeting, as a basis for discussion

Research and Development
CentEX MVD

If you indicate a relative importance of 50%, it means that improving the rate of Fatigue from 70% to 10% is 50% as important as improving the PFS rate at 3 months from 5% to 80%.

SERVIER moved by you

SARYGA STATISTICS & METHODOLOGY

QBRA of Ivosidenib in CCA: Weights

Elicitation process (swing-weighting methodology)

Slides presented during the elicitation meeting for each criterion

Ivosidenib vs placebo, 2nd or 3rd line IDH1m CCA population
Swing-weighting

Consider a treatment with:
- 70% patients with Fatigue (grade 2 or higher)
- 5% PFS rate at 3 months

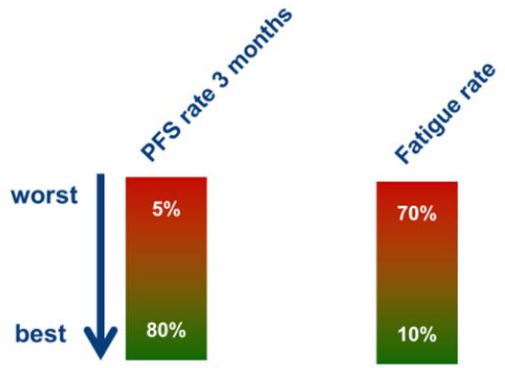
Quantify the relative importance (in %) of:
- Decreasing the rate of Fatigue from 70% to 10%
- Compared to improving the PFS rate at 3 months from 5% to 80%

Criteria	Relative importance
PFS rate at 3 months	100%
Fatigue	50%

Consensus reached during the meeting

Responses from the questionnaire:

5%
25%
50%
50%
60%



Research and Development
CentEX MVD

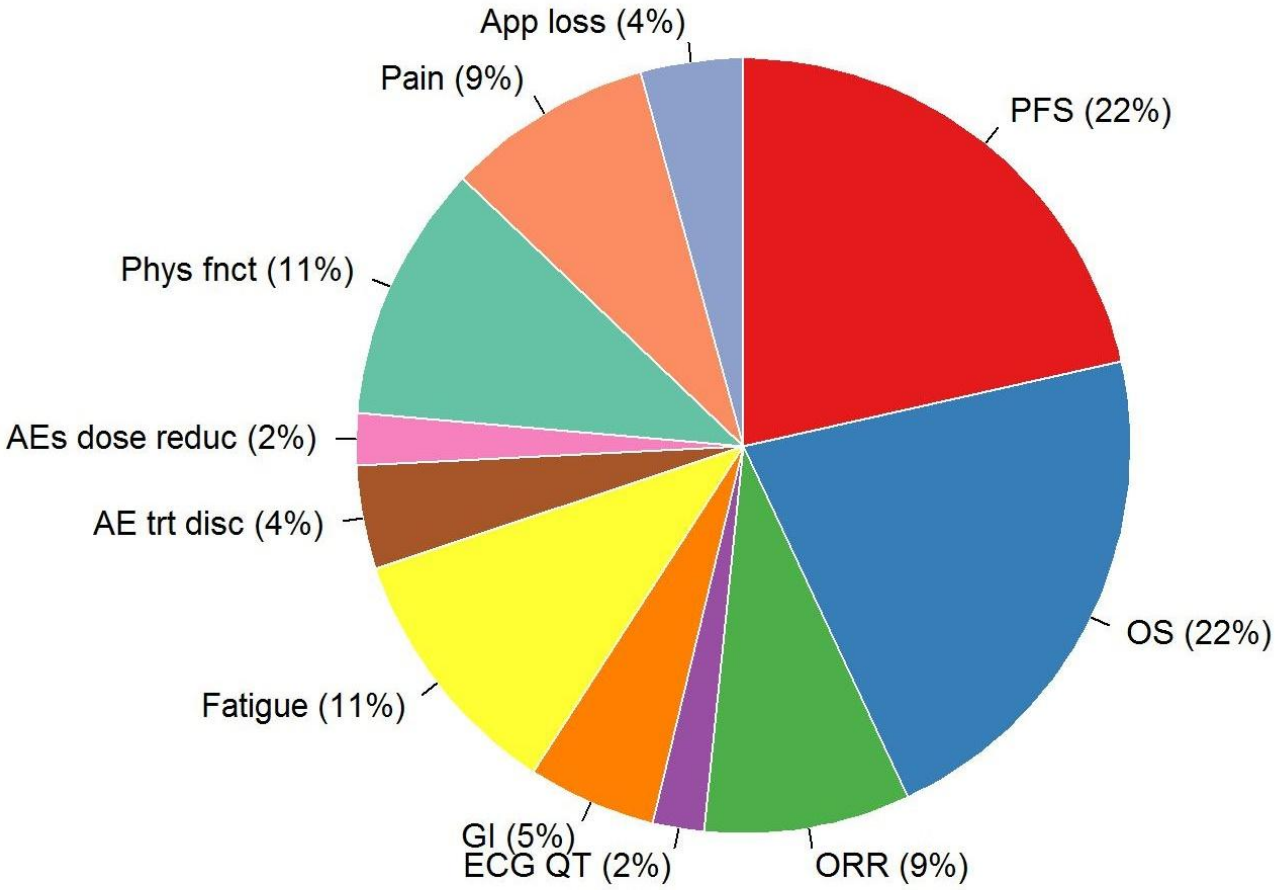
If you indicate a relative importance of 50%, it means that improving the rate of Fatigue from 70% to 10% is 50% as important as improving the PFS rate at 3 months from 5% to 80%.

SERVIER moved by you

SARYGA STATISTICS & METHODOLOGY

QBRA of Ivosidenib in CCA: Weights

Finalized by the expert panel via elicitation meeting



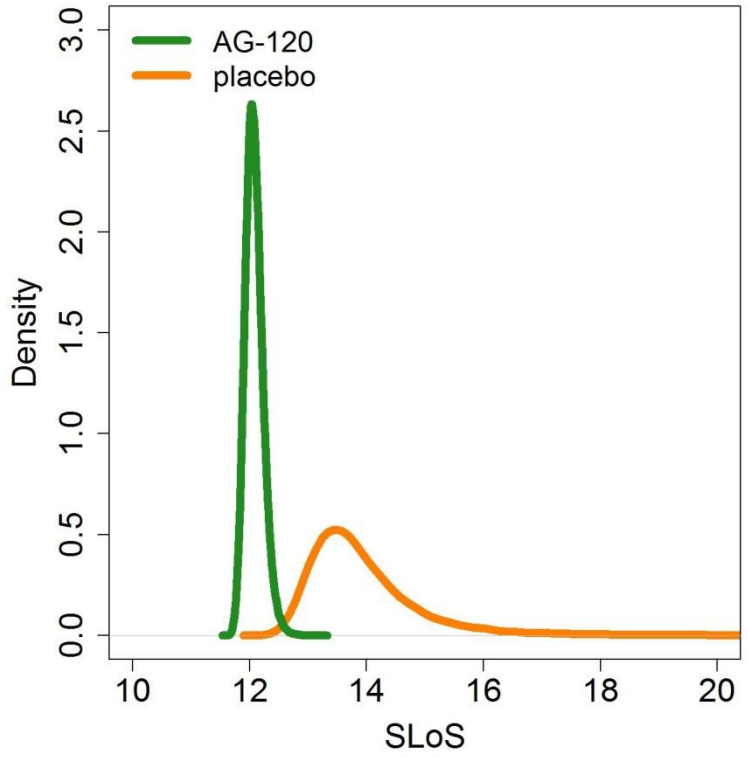
SERVIER
moved by you

SARYGA
STATISTICS & METHODOLOGY

QBRA of Ivosidenib in CCA: Results

Main analysis: SLoS model

Distribution of the B-R loss scores



The **lower** the better

Probability treatment is better than placebo

Probability Treatment > Placebo
95.24%

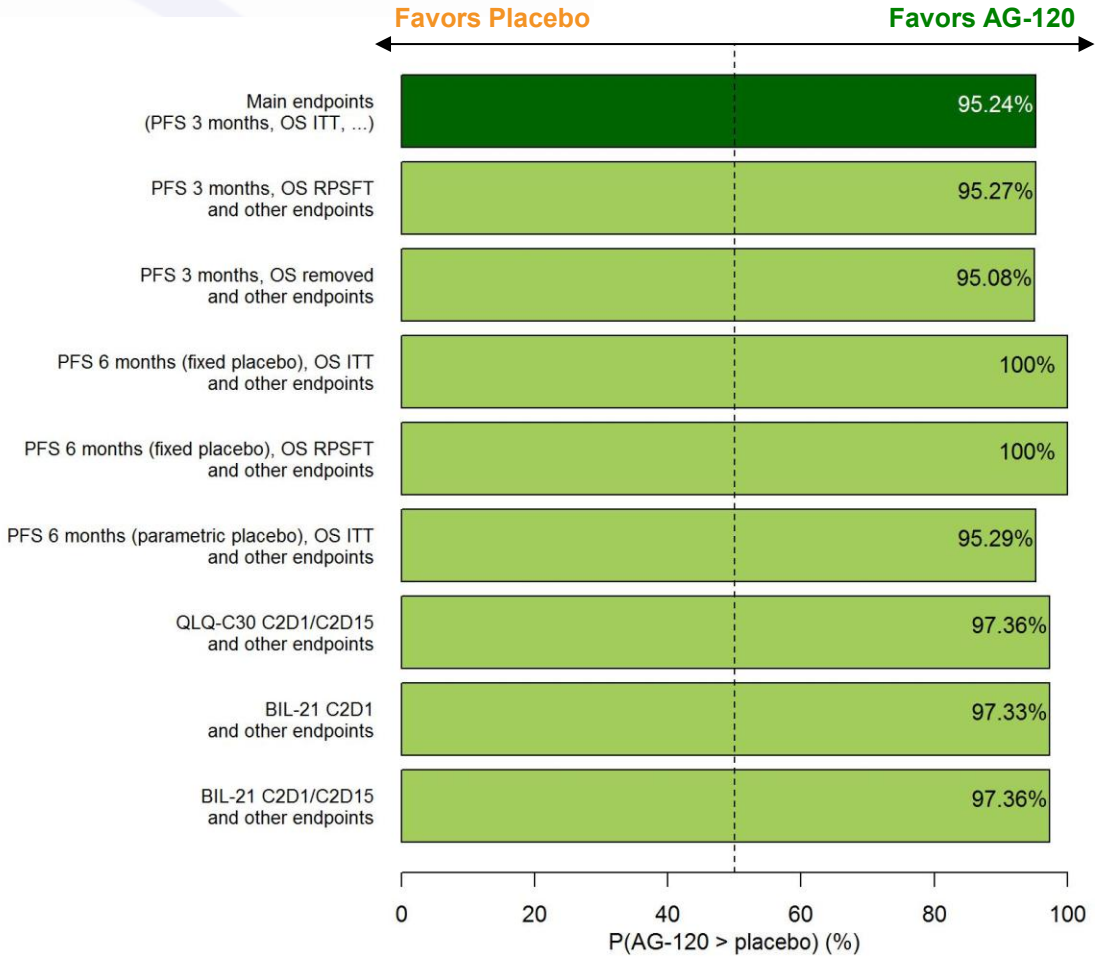
Saint-Hilary G, Robert V, Gasparini M, Jaki T, Mozgunov P. A novel measure of drug benefit-risk assessment based on Scale Loss Score. *Stat Methods Med Res*. 2019; 28(9):2738-2753.

SERVIER moved by you

SARYGA STATISTICS & METHODOLOGY

QBRA of Ivosidenib in CCA: Results

Sensitivity analyses: SLoS model



Sensitivity analyses

- **OS RPSFT**: Replace OS rate at 6 months *not adjusted* for crossover by OS rate at 6 months *adjusted* for crossover, and all other main endpoints
- **OS removed:** Given >70% patients in the placebo arm switched to treatment, OS ITT may not be considered a reliable endpoint and is removed from the model
- **PFS 6 months (fixed placebo)**: Replace PFS rate at 3 months by PFS rate at 6 months (deterministic for the placebo arm), and all other main endpoints
- **PFS 6 months (parametric placebo)**: Replace PFS rate at 3 months by PFS rate at 6 months (log-logistic survival model for the placebo arm), and all other main endpoints
- **QoL QLQ-C30 C2D1/C2D15:** Replace QLQ-C30 assessments at C2D1 by first scheduled assessment (C2D1 or C2D15), and all other main endpoints
- **QoL BIL-21 C2D1:** Replace QLQ-C30 Pain and Appetite Loss by BIL-21 Pain and Eating, at C2D1, and all other main endpoints
- **QoL BIL-21 C2D1/C2D15 :** Replace QLQ-C30 Pain and Appetite Loss by BIL-21 Pain and Eating, at first scheduled assessment (C2D1 or C2D15) , and all other main endpoints

Saint-Hilary G, Robert V, Gasparini M, Jaki T, Mozgunov P. A novel measure of drug benefit-risk assessment based on Scale Loss Score. *Stat Methods Med Res*. 2019; 28(9):2738-2753.

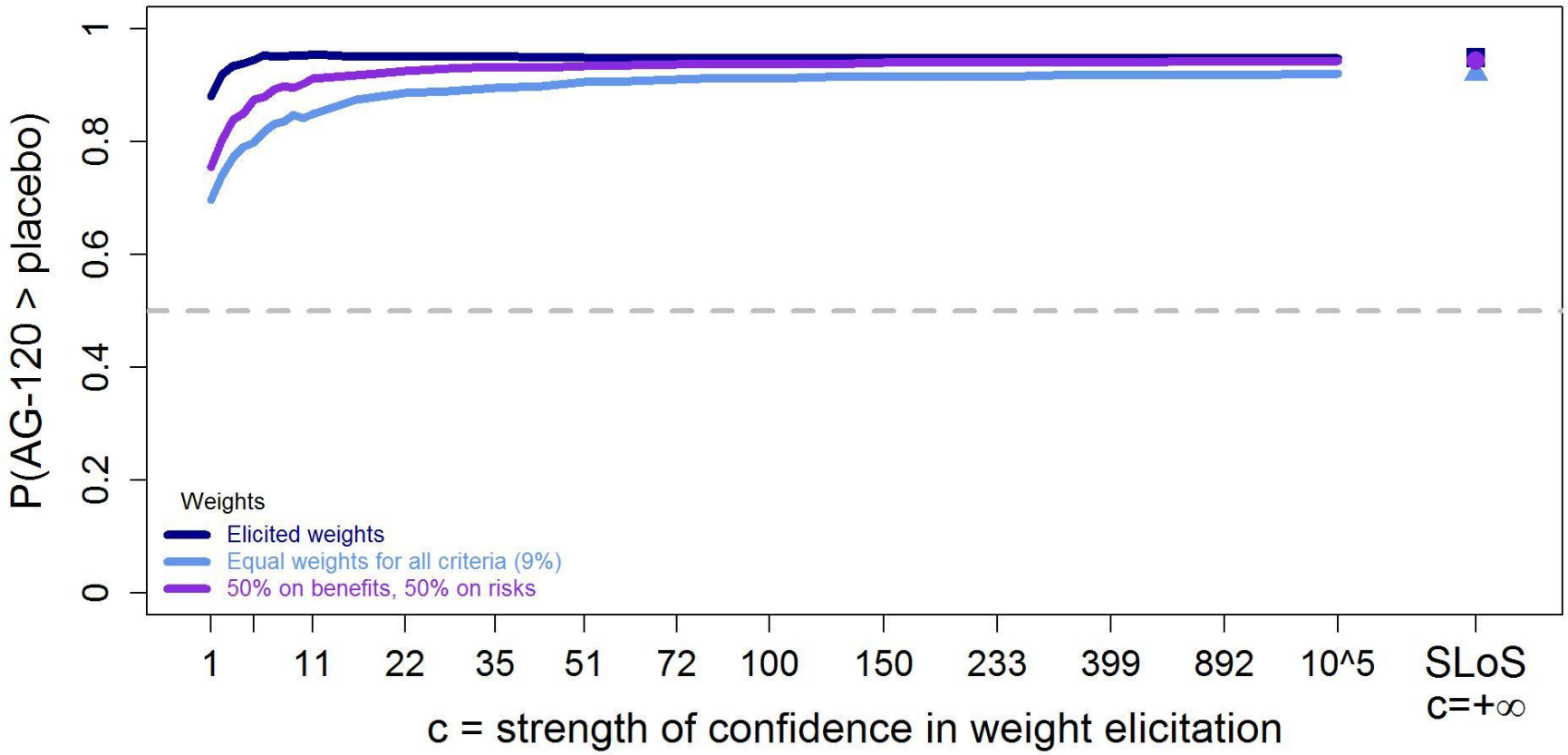
SERVIER moved by you

SARYGA STATISTICS & METHODOLOGY

QBRA of Ivosidenib in CCA: Results

Sensitivity analyses: Dirichlet SLoS model

- What if KOLs were uncertain in their weight elicitation?
- What if we used other weights for the criteria?



Robust results: the similar trend is observed for all sets of weights, and all converge quickly toward the main analysis results

Saint-Hilary, G., Cadour, S., Robert, V. and Gasparini, M. (2017). A simple way to unify multicriteria decision analysis (MCDA) and stochastic multicriteria acceptability analysis (SMAA) using a Dirichlet distribution in benefit–risk assessment. *Biometrical Journal* **59**(3): 567-578.

SERVIER moved by you

SARYGA STATISTICS & METHODOLOGY

QBRA of Ivosidenib in CCA


EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

23 February 2023
EMA/173654/2023
Committee for Medicinal Products for Human Use (CHMP)

Assessment report

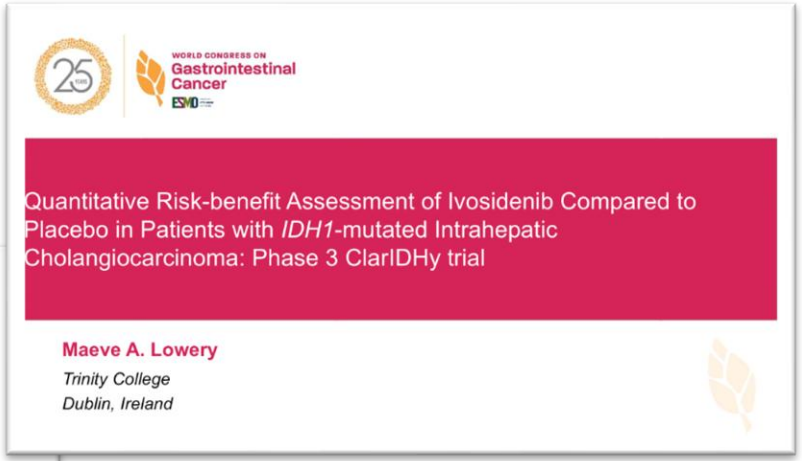
Tibsovo

International non-proprietary name: ivosidenib

- Submitted as part of the submission package to EMA in 2023

  *"(…) **The MCDA framework**, together with its elicitation process and analysis, **can be helpful in transforming multiple aspects of the data into a loss or utility score**. **Several models were investigated to assess the robustness of the main analysis results** (…). The elicited criteria and corresponding weights and values are dependent on a panel of 7 KOLs, and different panels may have provided different recommendations, leading to some natural variability in the selections (…)."*

- Presented at ESMO 2023
- Published in ESMO Gastrointestinal Oncology 2025


WORLD CONGRESS ON Gastrointestinal Cancer
ESMO

Quantitative Risk-benefit Assessment of Ivosidenib Compared to Placebo in Patients with *IDH1*-mutated Intrahepatic Cholangiocarcinoma: Phase 3 ClarIDHy trial

Maeve A. Lowery
Trinity College
Dublin, Ireland


ESMO GOOD SCIENCE BETTER MEDICINE BEST PRACTICE

ESMO GASTROINTESTINAL ONCOLOGY

ORIGINAL ARTICLE

Quantitative benefit–risk assessment of data from the phase III ClarIDHy study of ivosidenib versus placebo in patients with m*IDH1* cholangiocarcinoma

J. W. Valle[1,2*†], G. K. Abou-Alfa[3,4,5], R. K. Kelley[6], M. A. Lowery[5], R. T. Shroff[7], Y. Bian[8], G. Saint-Hilary[9], H. Liu[8], Z. Teng[8], Z. Hua[8], C. Gliser[8], A. Vogel[10,11,12] & M. M. Javle[13]

SERVIER
moved by you

SARYGA
STATISTICS & METHODOLOGY

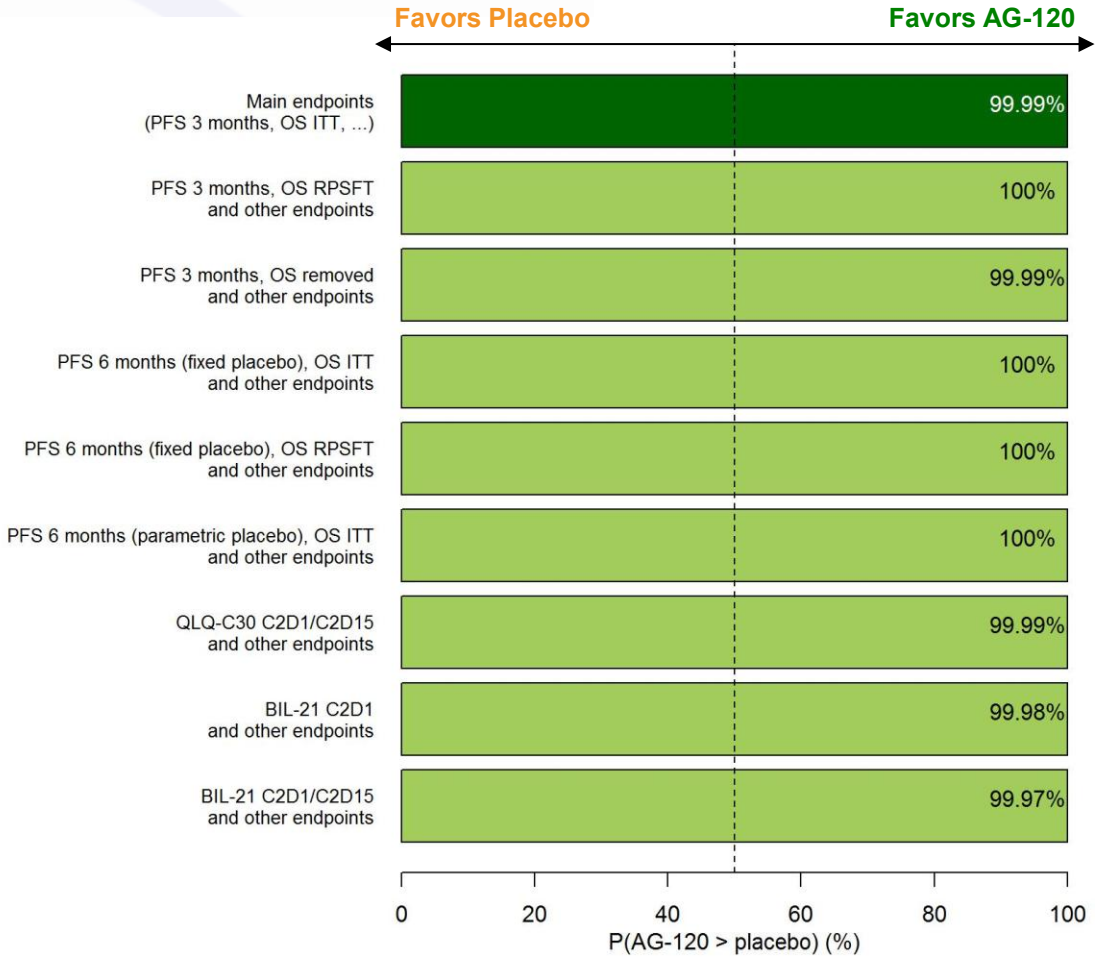
Thank you!

SERVIER
moved by you

SARYGA
STATISTICS & METHODOLOGY

QBRA of Ivosidenib in CCA: Results

Sensitivity analyses: Linear model



Sensitivity analyses

- **OS RPSFT**: Replace OS rate at 6 months *not adjusted* for crossover by OS rate at 6 months *adjusted* for crossover, and all other main endpoints
- **OS removed:** Given >70% patients in the placebo arm switched to treatment, OS ITT may not be considered a reliable endpoint and is removed from the model
- **PFS 6 months (fixed placebo)**: Replace PFS rate at 3 months by PFS rate at 6 months (deterministic for the placebo arm), and all other main endpoints
- **PFS 6 months (parametric placebo)**: Replace PFS rate at 3 months by PFS rate at 6 months (log-logistic survival model for the placebo arm), and all other main endpoints
- **QoL QLQ-C30 C2D1/C2D15:** Replace QLQ-C30 assessments at C2D1 by first scheduled assessment (C2D1 or C2D15), and all other main endpoints
- **QoL BIL-21 C2D1:** Replace QLQ-C30 Pain and Appetite Loss by BIL-21 Pain and Eating, at C2D1, and all other main endpoints
- **QoL BIL-21 C2D1/C2D15 :** Replace QLQ-C30 Pain and Appetite Loss by BIL-21 Pain and Eating, at first scheduled assessment (C2D1 or C2D15) , and all other main endpoints

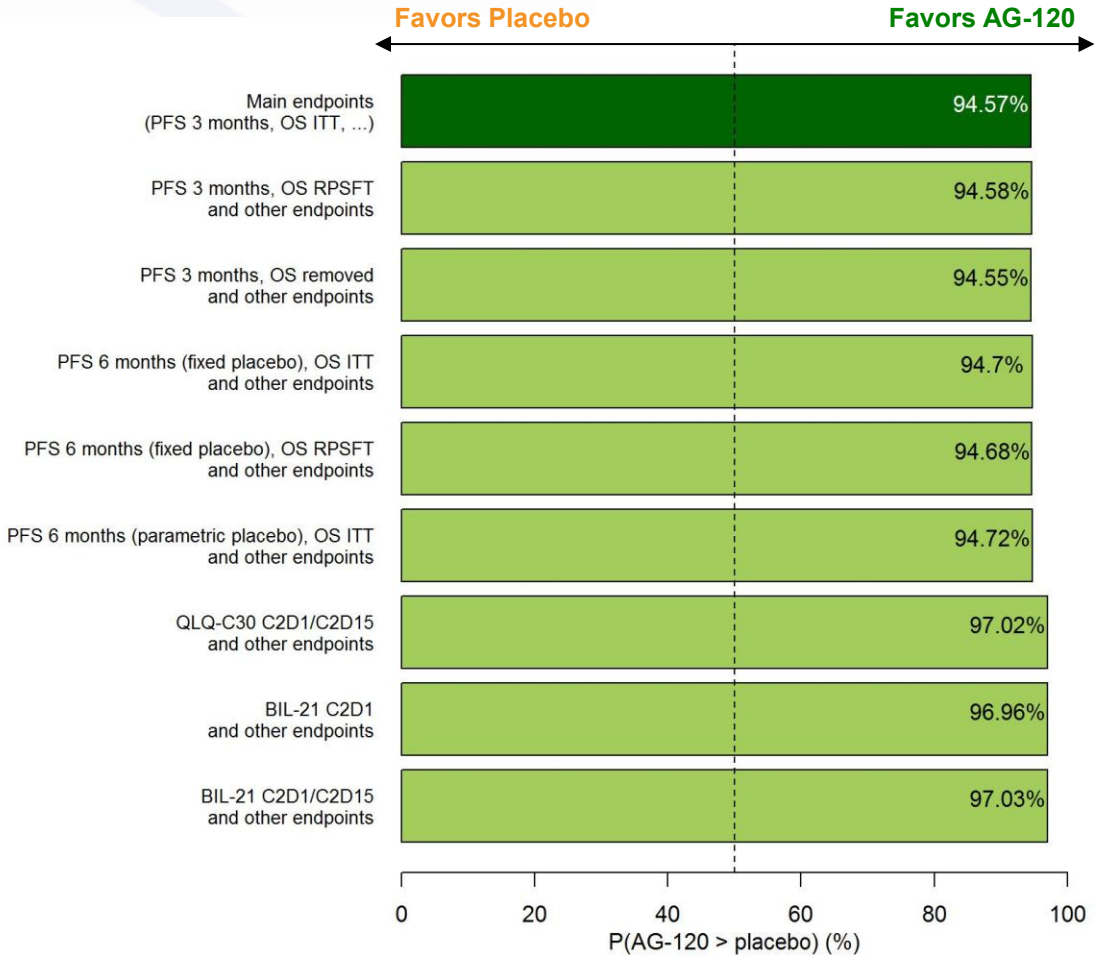
Saint-Hilary G, Robert V, Gasparini M, Jaki T, Mozgunov P. A novel measure of drug benefit-risk assessment based on Scale Loss Score. *Stat Methods Med Res*. 2019; 28(9):2738-2753.

SERVIER moved by you

SARYGA STATISTICS & METHODOLOGY

QBRA of Ivosidenib in CCA: Results

Sensitivity analyses: Product model



Sensitivity analyses

- **OS RPSFT**: Replace OS rate at 6 months *not adjusted* for crossover by OS rate at 6 months *adjusted* for crossover, and all other main endpoints
- **OS removed:** Given >70% patients in the placebo arm switched to treatment, OS ITT may not be considered a reliable endpoint and is removed from the model
- **PFS 6 months (fixed placebo)**: Replace PFS rate at 3 months by PFS rate at 6 months (deterministic for the placebo arm), and all other main endpoints
- **PFS 6 months (parametric placebo)**: Replace PFS rate at 3 months by PFS rate at 6 months (log-logistic survival model for the placebo arm), and all other main endpoints
- **QoL QLQ-C30 C2D1/C2D15:** Replace QLQ-C30 assessments at C2D1 by first scheduled assessment (C2D1 or C2D15), and all other main endpoints
- **QoL BIL-21 C2D1:** Replace QLQ-C30 Pain and Appetite Loss by BIL-21 Pain and Eating, at C2D1, and all other main endpoints
- **QoL BIL-21 C2D1/C2D15 :** Replace QLQ-C30 Pain and Appetite Loss by BIL-21 Pain and Eating, at first scheduled assessment (C2D1 or C2D15) , and all other main endpoints

Saint-Hilary G, Robert V, Gasparini M, Jaki T, Mozgunov P. A novel measure of drug benefit-risk assessment based on Scale Loss Score. *Stat Methods Med Res*. 2019; 28(9):2738-2753.

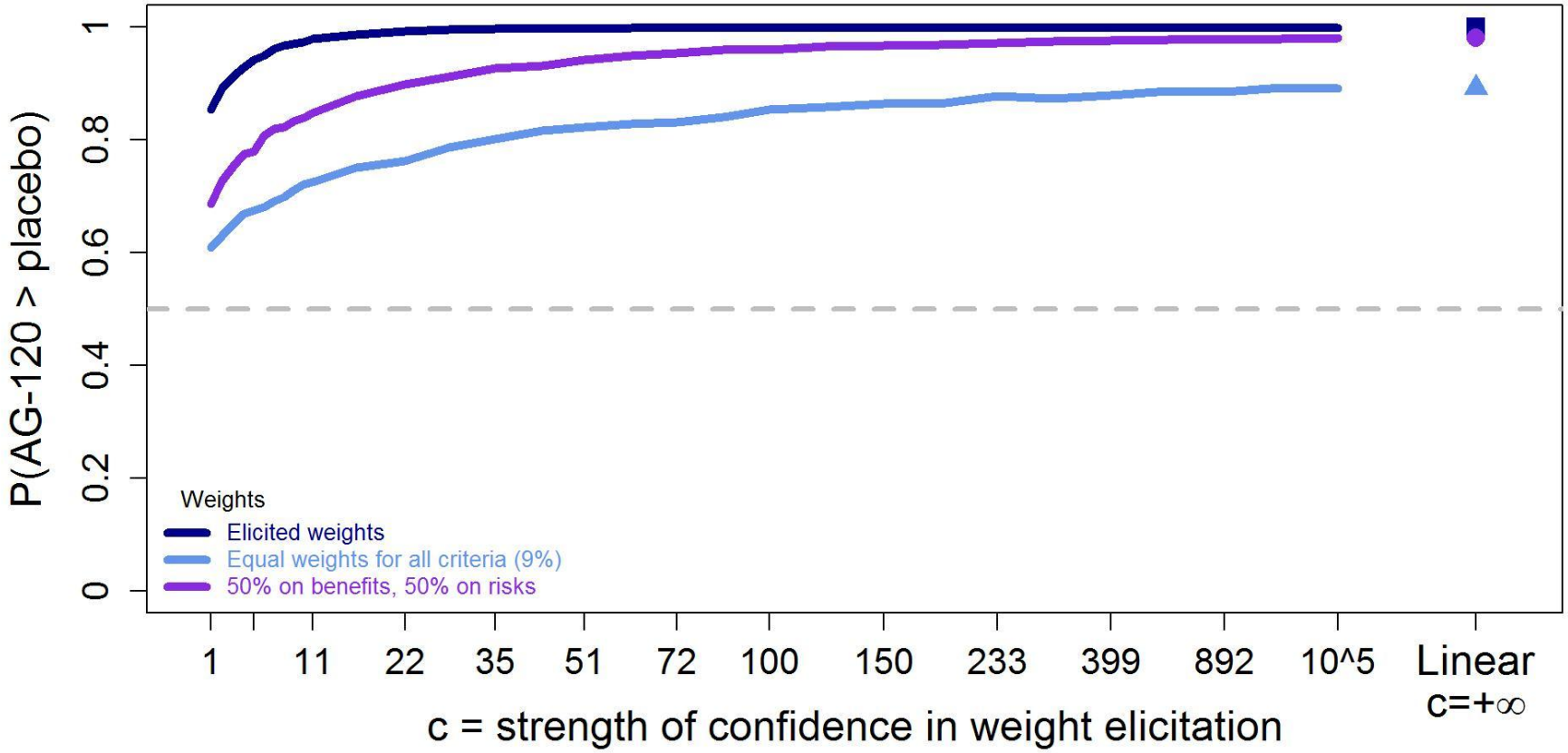
SERVIER moved by you

SARYGA STATISTICS & METHODOLOGY

QBRA of Ivosidenib in CCA: Results

Sensitivity analyses: Dirichlet Linear model

- What if KOLs were uncertain in their weight elicitation?
- What if we used other weights for the criteria?



Robust results: the similar trend is observed for all sets of weights, and all converge quickly toward the main analysis results

Saint-Hilary, G., Cadour, S., Robert, V. and Gasparini, M. (2017). A simple way to unify multicriteria decision analysis (MCDA) and stochastic multicriteria acceptability analysis (SMAA) using a Dirichlet distribution in benefit–risk assessment. *Biometrical Journal* **59**(3): 567-578.

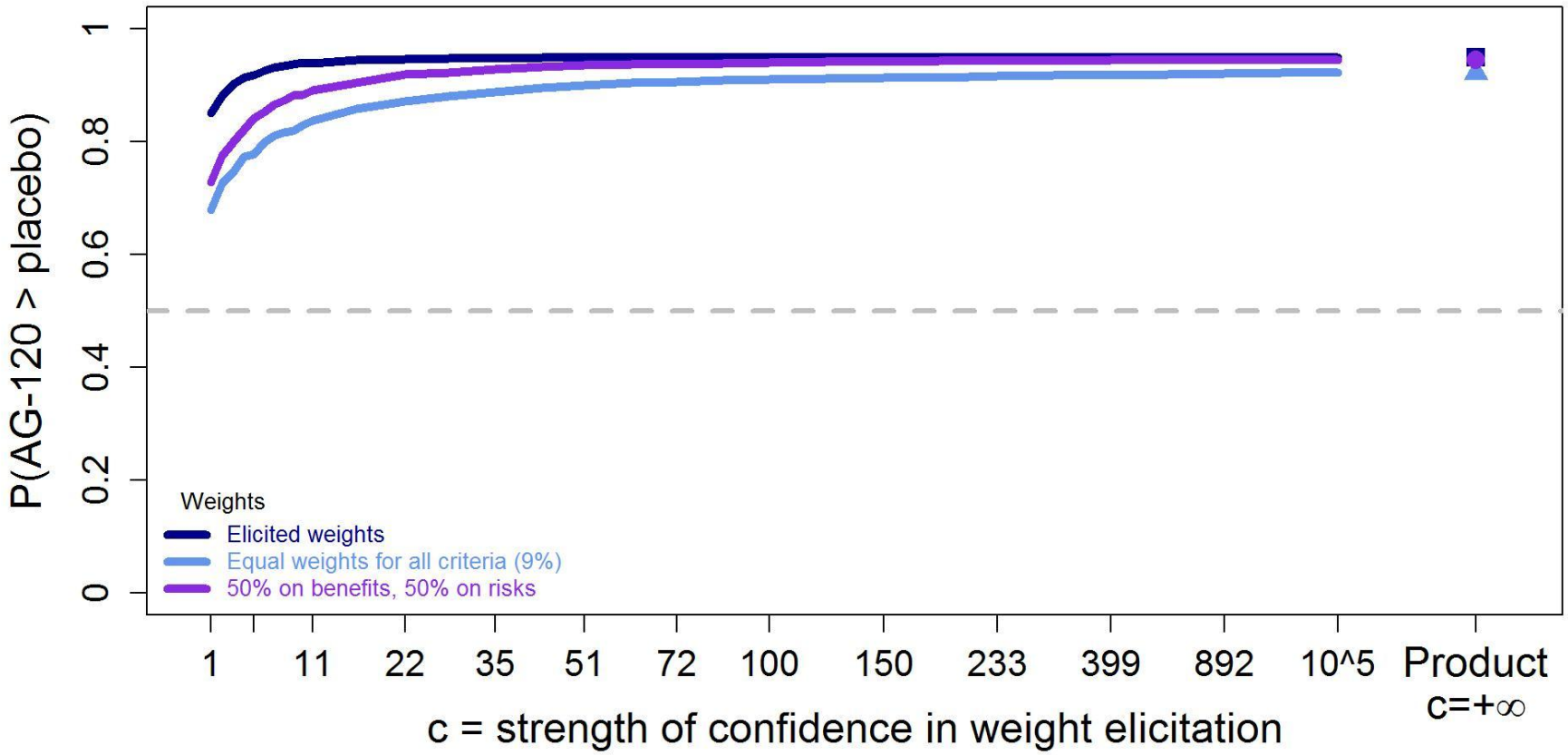
SERVIER moved by you

SARYGA STATISTICS & METHODOLOGY

QBRA of Ivosidenib in CCA: Results

Sensitivity analyses: Dirichlet Product model

- What if KOLs were uncertain in their weight elicitation?
- What if we used other weights for the criteria?



Robust results: the similar trend is observed for all sets of weights, and all converge quickly toward the main analysis results

Saint-Hilary, G., Cadour, S., Robert, V. and Gasparini, M. (2017). A simple way to unify multicriteria decision analysis (MCDA) and stochastic multicriteria acceptability analysis (SMAA) using a Dirichlet distribution in benefit–risk assessment. *Biometrical Journal* **59**(3): 567-578.

SERVIER moved by you

SARYGA STATISTICS & METHODOLOGY