Retrospective Evaluation of the Futility Analysis in LUME-Lung 2, a Randomised, Double-blind, Placebo-controlled Phase III Trial of Nintedanib (BIBF 1120) in Combination with Pemetrexed in NSCLC Patients Progressing after One Prior First-line Chemotherapy

#P327

Nasser H. Hanna,[1] Rolf Kaiser,[2] Joo-Hang Kim,[3] Richard N. Sullivan,[4] Osvaldo Aren,[5] Myung-Ju Ahn,[6] Beatrice Tiangco,[7] Isabelle Voccia,[8] José Barrueco,[9] Patricia Glomb;[2] for the LUME-Lung 2 Study Group

[1]*Melvin and Bren Simon Cancer Center, Indiana University, Indianapolis, IN, USA;* [2]*Boehringer Ingelheim GmbH, Biberach, Germany;* [3]*Department of Internal Medicine (Medical Oncology), Yonsei Cancer Research Institute, Yonsei Cancer Center, Seoul, South Korea;* [4]*Department of Oncology, Auckland City Hospital, Auckland, New Zealand;* [5]*Instituto Nacional del Cancer, Santiago, Chile;* [6]*Department of Medicine, Division of Hematology-Oncology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, South Korea;* [7]*Internal Medicine, National Kidney and Transplant Institute, Quezon City, Philippines;* [8]*Boehringer Ingelheim (Canada) Ltd, Burlington, ON, Canada;* [9]*Boehringer Ingelheim GmbH, Ridgefield, CT, USA*

INTRODUCTION

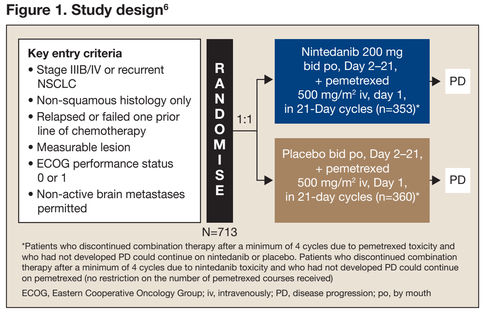
- Nintedanib is a potent, oral, small molecule, triple angiokinase inhibitor of vascular endothelial growth factor receptors 1–3, platelet-derived growth factor receptors α and β, and fibroblast growth factor receptors 1–3, which has demonstrated substantial antitumour and antiangiogenic activity in preclinical experiments[1] and in clinical trials (1199.10, .9, .26, .13)[2–5]
- LUME-Lung 2 (NCT00806819), a Phase III trial, investigated the efficacy and safety of nintedanib in combination with pemetrexed versus placebo plus pemetrexed for the treatment of patients with advanced or recurrent, non-squamous non small-cell lung cancer (NSCLC) who had previously received first-line chemotherapy[6]
- Based on the results of a preplanned futility analysis of investigator-assessed progression-free survival (PFS), conducted by an Independent Data Monitoring Committee (DMC), recruitment was halted early after 713/1300 planned patients had enrolled as the likelihood of meeting the primary endpoint was low, but at the same time there were no safety concerns
- Subsequent analysis showed that the primary endpoint of centrally reviewed PFS was met even though the study was stopped prematurely
  - treatment with nintedanib plus pemetrexed resulted in a statistically significant prolongation of centrally reviewed PFS compared with placebo plus pemetrexed (median, 4.4 vs 3.6 months, respectively; hazard ratio [HR]=0.83; 95% confidence interval [CI]: 0.7–0.99; p=0.04)[6]
- A retrospective analysis of the conditional and predictive power of the futility analysis is presented

METHODS

Study design

- LUME-Lung 2 was a global (Asia, North and South America, Europe) phase III, multicentre, randomised, double-blind, placebo-controlled trial (Figure 1)

Figure 1. Study design[6]



*Patients who discontinued combination therapy after a minimum of 4 cycles due to pemetrexed toxicity and who had not developed PD could continue on nintedanib or placebo. Patients who discontinued combination therapy after a minimum of 4 cycles due to nintedanib toxicity and who had not developed PD could continue on pemetrexed (no restriction on the number of pemetrexed courses received)

ECOG, Eastern Cooperative Oncology Group; iv, intravenously; PD, disease progression; po, by mouth

- Primary endpoint: PFS according to modified Response Criteria in Solid Tumors (RECIST) version 1.0, based on independent central review
- Key secondary endpoints: overall survival (OS); PFS based on investigator assessment; objective tumour response; safety and tolerability

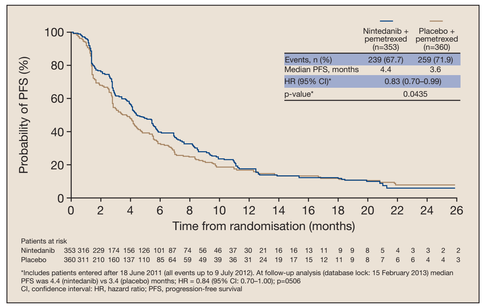
Futility analysis

- A DMC charter was developed prospectively to continually assess trial data to ensure overall safety in the patients treated, monitor the efficacy and quality and provide advice about the conduct of the trial and the integrity of the data.
- A preplanned futility analysis was to be performed by the DMC after 50% of the events for the primary PFS analysis had been observed (~356 events)
- The interim futility analysis included a number of statistical factors, including conditional and predictive power derived from investigator-assessed PFS results
  - ***Conditional power*** *– data from an interim examination used to calculate the probability of obtaining a statistically significant result at the final analysis. Where the conditional power falls below a predefined value, the trial is declared futile*
  - ***Predictive power*** *– averages the conditional power with previous data or current knowledge about the unknown parameters*
- The threshold for futility was set in the DMC charter at a conditional power of 20%
- Predictive (18.5%) and conditional (10.3%) power were calculated based on the data snapshot available at the time of the futility analysis; the database in this retrospective analysis was more mature due to new data entries following the DMC snapshot
- Analysis of the primary endpoint, independently assessed PFS, was conducted after 498 (database lock: 9 July 2012) PFS events (Figure 2)
- Retrospective analyses of the conditional and predictive power, and corresponding HR estimations, were undertaken on investigator-assessed and centrally reviewed PFS data (data snapshot from 9 July 2012) following the approach of Jennison et al.[7]
  - conditional and predictive power, and HRs for PFS, were investigated over time after approximately 10%, 20%, 30%, 40%, 50%, 60% and 70% of events for the primary PFS analysis (713 PFS events) had occurred
- Conditional and predictive powers for the trial were also recalculated at the time of the futility analysis

RESULTS

Progression-free survival

Figure 2. Primary endpoint of PFS by central independent review after 498 events[3]



*Includes patients entered after 18 June 2011 (all events up to 9 July 2012). At follow-up analysis (database lock: 15 February 2013) median PFS was 4.4 (nintedanib) vs 3.4 (placebo) months; HR = 0.84 (95% CI: 0.70–1.00); p=0506

CI, confidence interval: HR, hazard ratio; PFS, progression-free survival

Retrospective analysis of the conditional and predictive power

- As shown in Figure 3 and Table 1, the conditional and predictive power based on investigator-assessed PFS fluctuated over time, but remained on average ~40% above the predefined futility criteria
- Based on investigator-assessed PFS, very low predictive and conditional power were only observed at the time point of the futility analysis (21.6% and 13.3% power, respectively, Figure 3)
  - at that point in time (14 March 2011), 345 PFS events had occurred representing 48% of the events required for the primary PFS analysis
- The HR for investigator-assessed PFS was 0.92 (95% CI: 0.74–1.14) confirming the DMC's recommendation to stop the trial for futility
- Remarkably, the only time when the power dropped by a notable amount was at the time of the futility analysis (Figure 3)
- Retrospective calculation of the predictive and conditional power over time based on centrally reviewed PFS data showed a similar pattern, although conditional and predictive powers were higher compared with investigator-assessed PFS (Figure 4)

Table 1. Conditional and predictive power, and HRs for investigator-assessed PFS over time.

Number of PFS events			Fraction of events for primary PFS analysis (%)	Date of last event	Conditional power (%)	Predictive power (%)	HR	95% CI
Placebo	Nintedanib	Total						
41	32	73	10	23 December 2009	100.0	89.0	0.63	0.37–1.05
69	75	144	20	29 April 2010	37.7	42.8	0.88	0.62–1.25
112	103	215	30	27 August 2010	59.0	53.4	0.85	0.64–1.12
149	137	286	40	6 December 2010	35.4	39.5	0.88	0.69–1.12
176	169	345	48	14 March 2011	13.3	21.6	0.92	0.74–1.14
184	174	358	50	31 March 2011	21.0	27.9	0.90	0.73–1.11
221	207	428	60	15 June 2011	40.5	41.9	0.87	0.72–1.06
259	241	500	70	28 September 2011	31.8	34.1	0.88	0.74–1.05

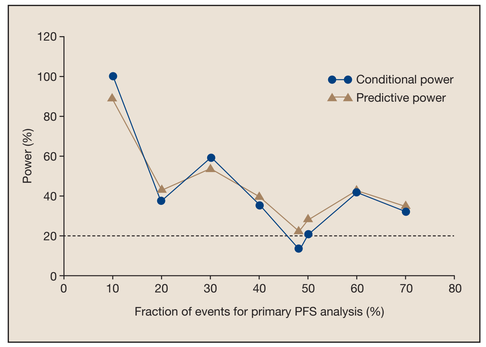
CI, confidence interval; HR, hazard ratio; PFS, progression-free survival.

Table 2. Conditional and predictive power, and HRs for centrally reviewed PFS over time

Number of PFS events			Fraction of events for primary PFS analysis (%)	Date of last event	Conditional power (%)	Predictive power (%)	HR	95% CI
Placebo	Nintedanib	Total						
40	34	74	10	21 December 2009	99.6	77.5	0.71	0.43–1.17
71	72	143	20	8 May 2010	76.4	60.2	0.82	0.58–1.16
115	99	214	30	3 September 2010	98.2	85.8	0.75	0.57–1.00
146	140	286	40	7 January 2011	23.0	31.0	0.90	0.71–1.15
185	172	357	50	1 May 2011	82.5	73.0	0.82	0.66–1.02
223	205	428	60	26 July 2011	76.5	70.1	0.83	0.68–1.01

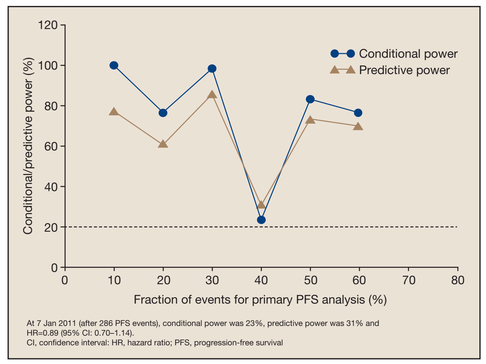
CI, confidence interval; HR, hazard ratio; PFS, progression-free survival

Figure 3. Retrospective conditional and predictive powers of investigator-assessed PFS over time



- Despite alternating increases and decreases over time, conditional power was generally >70% and predictive power was typically >60% (Table 2)
- Similar to that seen with investigator-assessed PFS, there was a decrease in conditional (23.0%) and predictive (31.0%) power after 40% of PFS events had occurred; this was followed by an immediate increase to 82.5% and 73.0%, respectively, after 50% of PFS events had occurred (Figure 4 and Table 3).

Figure 4. Retrospective conditional and predictive powers of centrally reviewed PFS over time



At 7 Jan 2011 (after 286 PFS events), conditional power was 23%, predictive power was 31% and HR=0.89 (95% CI: 0.70–1.14).

CI, confidence interval: HR, hazard ratio; PFS, progression-free survival

CONCLUSIONS

- The primary endpoint of LUME-Lung 2 was met even though the study was stopped prematurely following a preplanned futility analysis of investigator-assessed PFS by the DMC
- Retrospective investigations indicate that at the very time of the futility analyses, conditional and predictive power had dropped below the threshold predefined by the DMC charter. Had DMC analysis been performed at another timepoint, or had centrally reviewed data been used, the outcome of the futility analysis may have been different
- The DMC concluded retrospectively that the futility calculation upon which it based its decision to stop the study was an inadequate estimate of the true futility
- Learnings from this retrospective evaluation of the futility analysis of LUME-Lung 2 show that futility analyses should be repeated at different timepoints and potential discordances between investigator-assessed and centrally reviewed PFS should be considered

REFERENCES

1. Hilberg F, et al. Cancer Res 2008;68:4774–82
2. Reck M, et al. Ann Oncol 2011;22:1374–81
3. Ledermann JA, et al. J Clin Oncol 2011;29:3798–804
4. Eisen T, et al. J Clin Oncol 2013;31:18s (abstract 4506)
5. Reck M, et al. J Clin Oncol 2013;31:18s (abstract LBA8011)
6. Hanna N, et al. J Clin Oncol 2013;31:18s (abstract 8034)
7. Jennison C, Turnbull BW. Group sequential methods with applications to clinical trials. Boca Raton: Chapman & Hall, 2000

ACKNOWLEDGEMENTS

The authors were fully responsible for all content and editorial decisions, were involved at all stages of poster development, and have approved the final version. Medical writing assistance, supported financially by Boehringer Ingelheim, was provided by Duncan Campbell of GeoMed

To download a copy of the poster, scan the QR code or type the following URL into your browser: http://bit.ly/1c81bn6

The European Cancer Congress, Amsterdam, The Netherlands, 27 September–1 October, 2013