

Inter-observer Variability in Radiographic Assessment of Response in Chemotherapy Trials for Pancreatic Cancer

Brian C. Baumann, AB¹, Robert R. Ford, MD¹, Kathy Dasse, PharmD², Kathy Zhou, PhD³, Ruth Feldman, MD¹

¹ RadPharm Imaging Core Lab (Princeton, NJ), ² Pfizer, Inc. (Ann Arbor, MI), ³ Department of Public Health, Weill Cornell Medical College (New York, NY)

Background Information

- In the era of evidence-based medicine, the quality of the evidence is critical
- Previous studies examining inter-observer variability in the radiographic assessment of drug response revealed significant discrepancies
- Response rates reported by investigators (INV) were significantly higher than rates reported by an independent review committee (IRC)
- No studies have examined differences between investigators and independent central review panels for the imaging assessment of drug response in unresectable pancreatic cancer

Summary Table – Previous Studies

Author	Type of Cancer	Total # subjects	# of Responders by INV	Objective response rate (INV or IRC)	# of subjects reviewed by IRC	Response Concordance (IRV=INV)	Change in Response Rate
Gwyther	Ovarian	111	24	25.8% INV 15.2% IRC	21	14/24	42% ↓
Thiesse	Renal	489	86	--	133	52	23% ↓
Miller	Breast	444	N/A	19.1–30.2% INV 9.1–19.8% IRC	400+	N/A	34 – 52% ↓
Rothenberg	Colo-rectal	463	21	13.8% INV 9.9% IRC	463	N/A	28% ↓

Materials and Methods

- 133 patients in two multi-center, randomized Phase II drug trials had their response assessed by measurement of tumor size or new lesions on serial CT scans
- Protocols specified techniques of tumor measurement & defined responses as complete (CR) or partial responses (PR), stable disease (SD) or progressive disease (PD)
- Objective responses were confirmed by repeat imaging evaluations after an interval >4 weeks
- Early withdrawal of PR or SD cases from the study <8 weeks after enrollment were scored as non-evaluable (NE)
- The same images were then blindly reviewed by protocol-trained radiologists at RadPharm, an independent imaging core lab in Princeton, NJ
- Cases of discordance in the interpretation of best response between investigators and reviewers were then blindly re-reviewed by an independent adjudicator

Results

- In 50/133 cases, there was sufficient discrepancy between investigators and reviewers to result in a change in overall response
- In 10 of these cases, investigators reported responses inconsistent with protocol standards for the investigator's reported tumor measurements
- In 40 cases, reviewers disagreed with the radiological findings reported by investigators

Best Response Corrected to Conform to Protocol Criteria

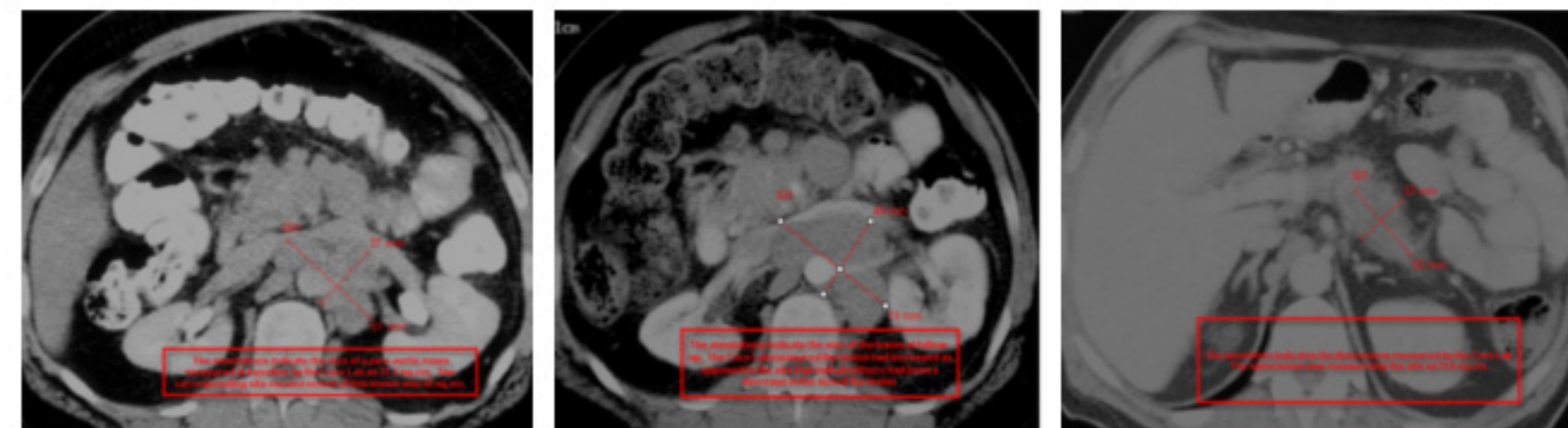
	CR	PR	SD	PD	NE
Investigators	1	20	50	44	18
Reviewers	0	6	67	45	15

- Discrepancies between investigators and reviewers could be classified into the following categories:

- Improperly measured target lesions
- Incorrect selection of target lesions
- Failure to identify new lesions
- Missing data
- Incorrect response form paradigm (disease identified at baseline not subsequently followed, yet response was reported)
- Failure to apply the protocol standards correctly

- In 47 of the 50 cases of discrepancy between investigators and reviewers, the films were available for a second review by the adjudicator
- The adjudicator confirmed the reviewers' assessment of response in 39/47 (83%) of cases
- Discrepancies in response were not randomly distributed between investigators & reviewers
- Investigators reported significantly more objective responses (CR + PR) than reviewers even after investigators' response coding errors were corrected
- Investigators reported favorable responses in 21/133 (16%) of cases compared with 6/133 (5%) responses noted by reviewers, a significant discrepancy (p=.01)

Examples of Discrepancies



Subject 1056 - Baseline

Subject 1056 – Baseline + 1

Subject 1021 – Baseline

Subject 1056

- Investigator Response: SD; Reviewer Response: PD; Adjudicator Response: PD
- Cause of discrepancy: Investigator improperly measured the para-aortic nodal mass target lesion & neglected other target lesions, e.g. nodal mass to left of sma & portacaval node
- At baseline, the investigator obtained a measurement of 63.0 sq cm by combining two adjacent lesions. At baseline + 1, the investigator measured only one of the now clearly distinct lesions, artificially increasing the % drop in bidimensional lesion area which contributed to the investigator's incorrect response of SD.

Subject 1021

- Investigator Response: SD; Reviewer Response: PD; Adjudicator Response: PD
- Cause of discrepancy: Investigator improperly measured the pancreatic mass, including the dilated pancreatic duct proximal to the mass in his measurement
- At baseline, the investigator measured the pancreatic mass as 33.6 sq cm compared to the reviewer who measured 17.5 sq cm. Whereas the reviewer noted an increase in the size of the lesion after baseline, the investigator noted incorrectly that the mass was decreasing.

Discussion & Conclusions

Discussion

- 71% decrease in response rate in this study was even more pronounced than previously reported 23 – 52% decreases in response rates in studies of other primary malignancies
- Possible explanations for the significant discrepancy between investigators and independent reviewers:
 - Investigator bias: Limited alternative treatment options for unresectable pancreas cancer may bias investigators to assess a more favorable response that would ensure patients continued participation in the trial
 - Challenge of accurately measuring pancreatic cancer lesions with existing CT-based imaging technology

Conclusions

- This study supports the use of independent review of trial data when the endpoint of the trial, e.g. response rate, is based on imaging data