Memorial Sloan Kettering
Cancer Center

Measuring Tumor Response: Lessons Learned from Clinical Trials

Chaya Moskowitz, Ph.D.
Department of Epidemiology and Biostatistics
Memorial Sloan Kettering Cancer Center

New York: Springtime Snow

New York Times, March 21, 2016

Background

- Information on the change in tumor burden over time is used in many early-phase and some late-phase clinical trials in oncology
- Primary endpoint in Phase II studies
- Phase III studies with progression-free survival or time to progression as a primary endpoint

Measuring Tumor Burden

- Patients imaged pre-treatment and then serial post-treatment imaging
- Radiologists evaluate tumor burden at each time point
- Read criteria provide guidelines to standardize process
  - Suggest imaging modalities, how to measure lesions, how many lesions to measure, how to choose which lesions to measure
  - Vary depending on indication
  - RECIST 1.1 (Response Evaluation Criteria in Solid Tumors)

Endpoints Based on RECIST

- Target lesion: Measure up to 5 lesions, 2 per involved organ
- Measure longest diameter of each target lesion
  - X_l^T = longest diameter of l[th] target lesion at time T , (T = Baseline, visit 1, etc.)
  - $TB^T = \sum_l X_l^T$; tumor burden at time T
  - Use to define relative change in tumor burden:

$$RC = 100 \times \frac{TB^{Follow\text{-}up} - TB^{Baseline}}{TB^{Baseline}}$$

Response rate:

$$RR = \frac{CR + PR}{\#Subjects}$$

- Response criteria used in endpoint definitions (target response):
  - Complete Response (CR): Disappearance of all target lesions
  - Partial Response (PR): *RC* ≤ -30%
  - Progressive Disease (PD): *RC* ≥ 20% or new lesion appears
  - Stable Disease (SD): -30% ≤ RC ≤ 20%

→ **Time to PD or PFS**

A Limitation of Response Criteria

- Do not do a good job of suggesting which therapies will be successful in Phase III trials
  - Varies by indication

- Variability in tumor measurements

Ratain and Eckhardt, 2004; Oxnard et al., 2012; Villaruz and Socinski, 2013

Variability in Tumor Measurements

- Possible contributing factors include:
  - Patient-related sources of variability: e.g. indication, treatment, other biophysiological sources
  - Imaging-related sources of variability: e.g. modality, acquisition techniques, reconstruction parameters)
  - Reader-related sources of variability: e.g. reader expertise, choice of different target lesions, errors in tumor measurements
- Studies in single tumor measurements, *RC*, response criteria
- Most studies have small sample sizes, focus on a single disease site, and are designed experiments using retrospective research reads
- RECIST acknowledges variability and mentions independent central review may be warranted

Thiesse et al., 1997; Erasmus et al., 2003; Sullivan et al., 2013; Yoon et al. 2016

Independent Central Review

- All images collected for the clinical trial are transferred to a central location and reviewed by experts not involved in the study
  - Eligibility and trial endpoints
- Independent central review encouraged by regulatory authorities
- Different review paradigms, e.g:
  - Blinded
  - Two readers and adjudicator; most frequently used for industry-sponsored trials
- Differences between investigators and independent central review
- Lack of studies looking at differences between radiologists participating in an independent central review

Ford et al., 2009

Aim

- **Evaluate factors associated with variability in independent central review response assessment**

Independent Central Review Database

- Commercial Imaging Core Laboratory database capturing data from blinded independent central reviews of industry-sponsored Phase II and Phase III trials
- All trials in the database for which the Imaging Core Laboratory used two reader and adjudicator paradigm
- 79 clinical trials
- 23,476 patients
- Data available aggregated within de-identified clinical trial (no patient-level data)

Trial Characteristics Available

- Indication
- Read criteria
  - Guidelines used for tumor evaluation
- Adjudication variables
  - Variables related to study endpoints used to determine whether adjudication is required during the independent central review
  - Include best response, date of progression, time to progression
- Average number of target lesions
  - Total number of target lesions selected at baseline visit divided by total number of patients
- Average number of time points
  - Total number of time points at which patients were scanned and for which scans were received by the Imaging Core Laboratory divided by the total number of patients
- Average number of exams per time point
  - Total number of imaging exams received divided by the total number of time points

Methods

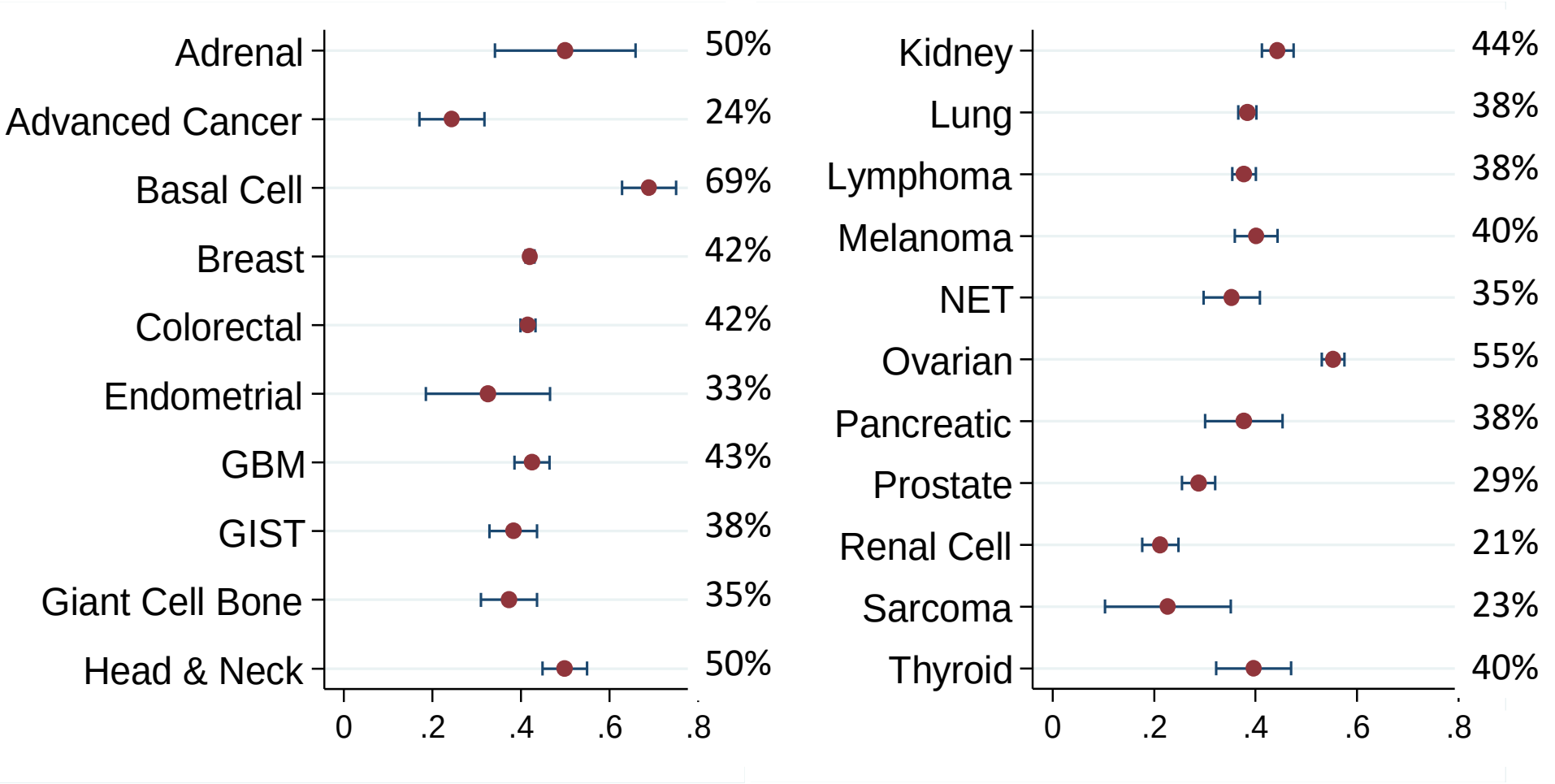
- Excluded information on patients who have no imaging exams after baseline scan
- Estimated and present proportion of cases where the two independent readers disagreed
- Used generalized linear models with weighted least squares and *F* to test for associations

Indication

	# Trials	# Patients
Adrenal	1	38
Advanced Cancer	1	131
Basal Cell	2	222
Breast	18	8,497
Colorectal	7	3,225
Endometrial	1	43
GBM	3	607
GIST	2	311
Giant Cell Bone	2	225
Head & Neck	3	379

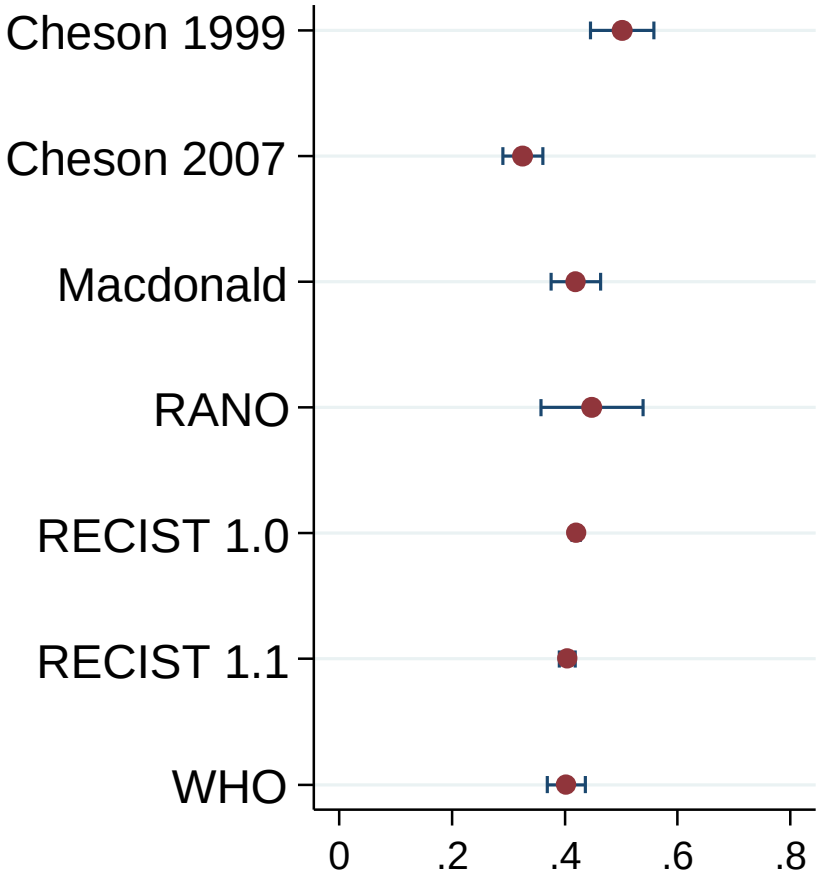
	# Trials	# Patients
Kidney	3	963
Lung	11	2,806
Lymphoma	13	1,692
Melanoma	3	526
NET	1	286
Ovarian	3	1,901
Pancreatic	1	154
Prostate	1	748
Renal Cell	1	509
Sarcoma	1	44
Thyroid	1	169

Proportion Discordant by Indication



% of patients for whom adjudication is required with 95% confidence intervals

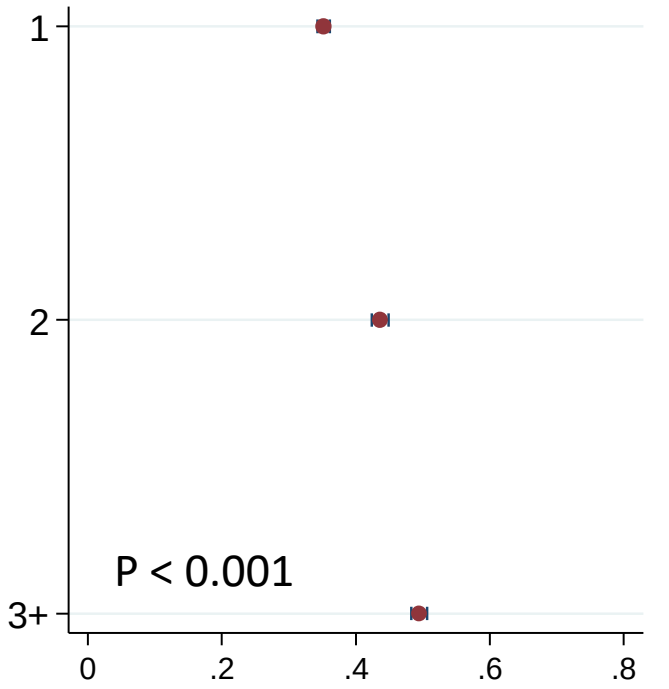
Read Criteria



	# Trials	# Patients	Proportion Discordant
Cheson 1999	4	303	50%
Cheson 2007	6	667	33%
Macdonald	2	491	42%
RANO	1	116	45%
RECIST 1.0	44	16,384	42%
RECIST 1.1	18	4,681	40%
WHO	4	834	40%

% of patients for whom adjudication is required with 95% confidence intervals

Number of Adjudication Variables



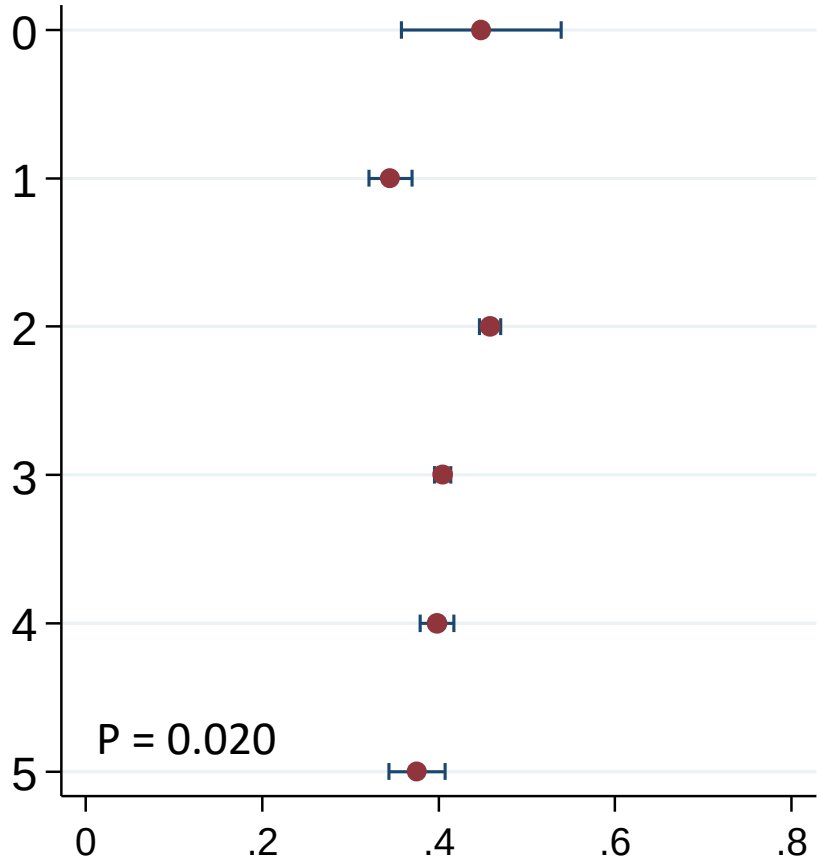
	# Trials	# Patients	Proportion Discordant
	36	10,521	35%
	20	6,285	44%
	23	6,670	50%

Among trials using a single adjudication variable

Date of progression	26	9,130	35%
Best response	5	701	28%

% of patients for whom adjudication is required with 95% confidence intervals

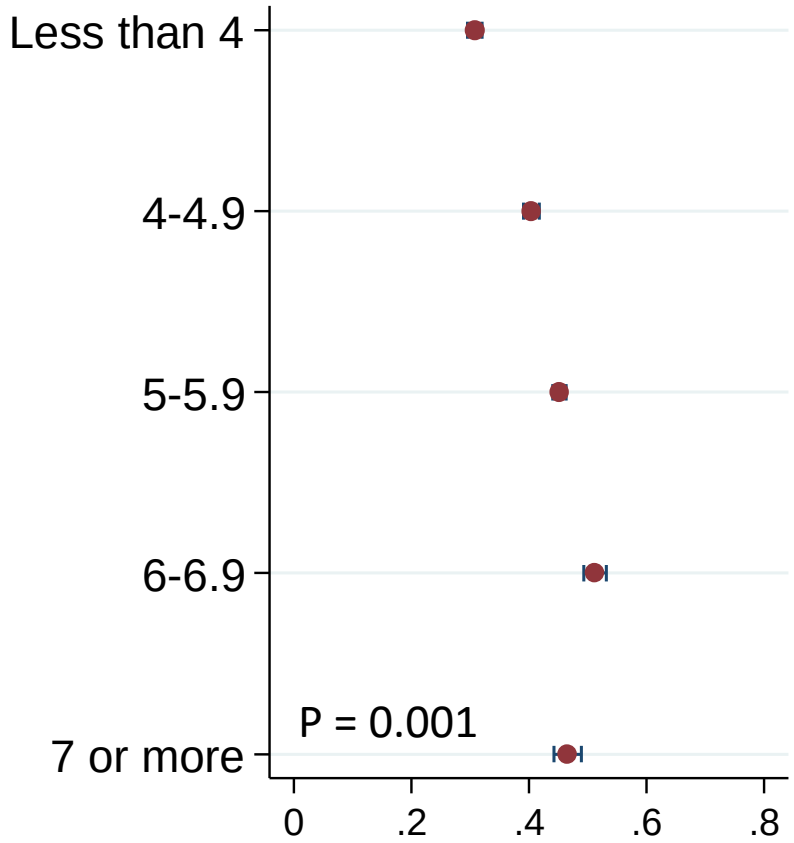
Average Number of Target Lesions



# Trials	# Patients	Proportion Discordant
1	116	45%
7	1,459	35%
20	6,499	46%
31	11,367	40%
13	2,500	40%
6	882	38%

% of patients for whom adjudication is required with 95% confidence intervals

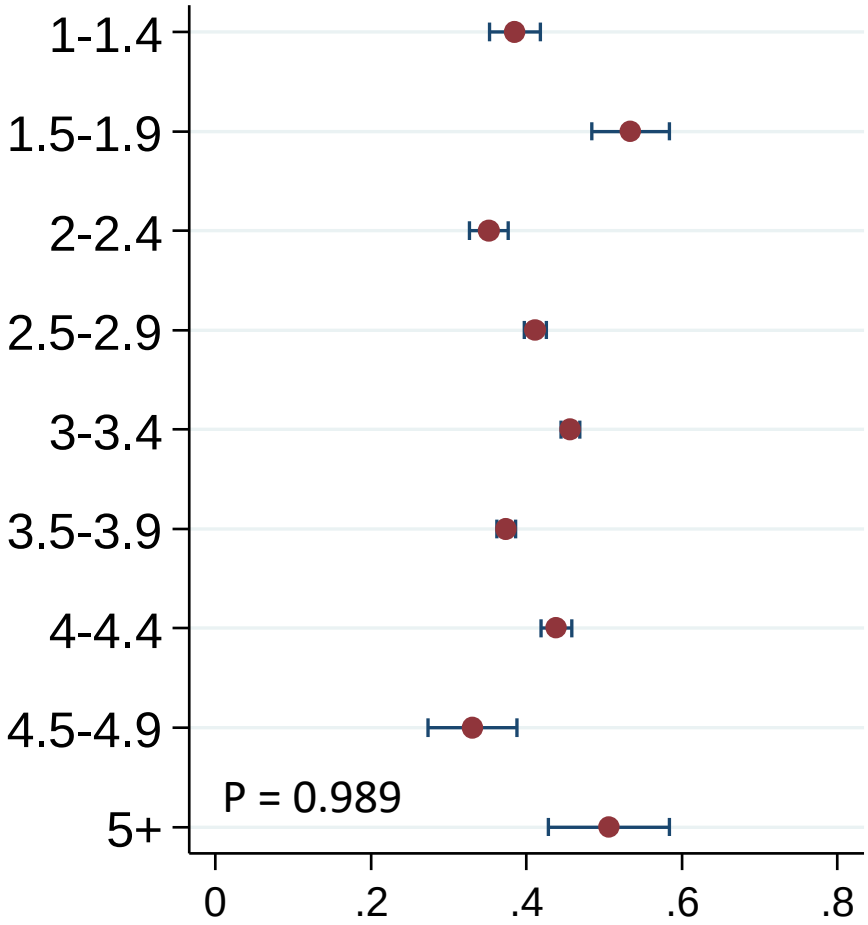
Average Number of Time Points



	# Trials	# Patients	Proportion Discordant
Less than 4	26	5,444	31%
4-4.9	16	5,362	40%
5-5.9	20	7,693	45%
6-6.9	8	2,611	51%
7 or more	8	1,713	47%

% of patients for whom adjudication is required with 95% confidence intervals

Average Number of Exams



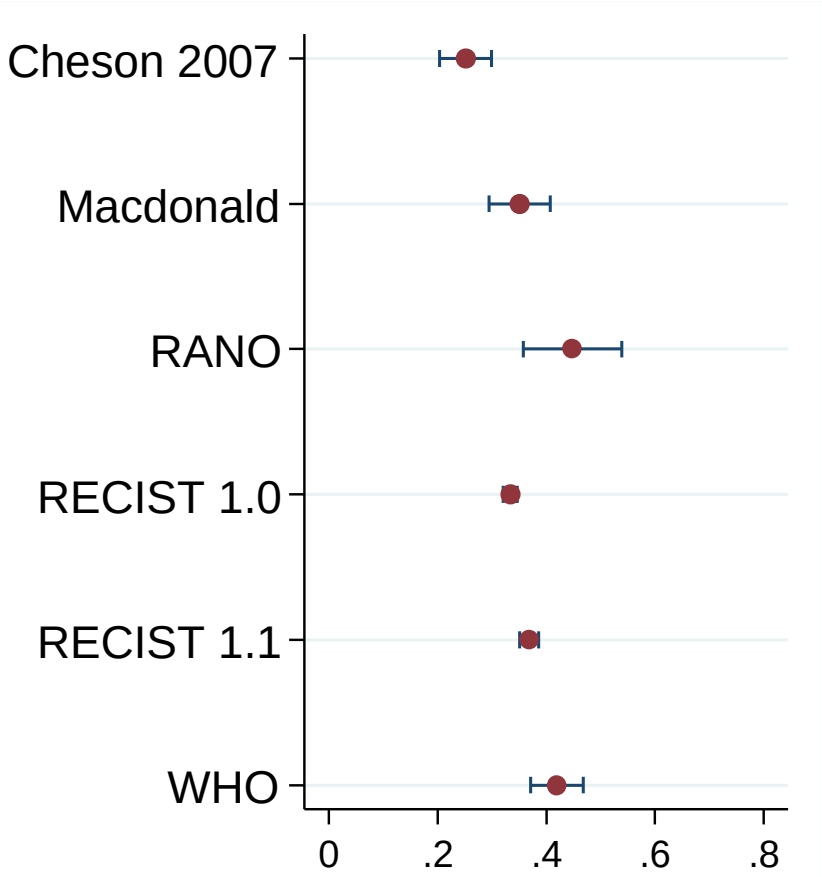
	# Trials	# Patients	Proportion Discordant
1-1.4	4	852	38%
1.5-1.9	3	380	53%
2-2.4	4	1,403	35%
2.5-2.9	14	4,669	41%
3-3.4	19	6,622	46%
3.5-3.9	20	6,139	37%
4-4.4	11	2,340	44%
4.5-4.9	1	260	33%
5+	2	158	51%

% of patients for whom adjudication is required with 95% confidence intervals

Multivariate Model

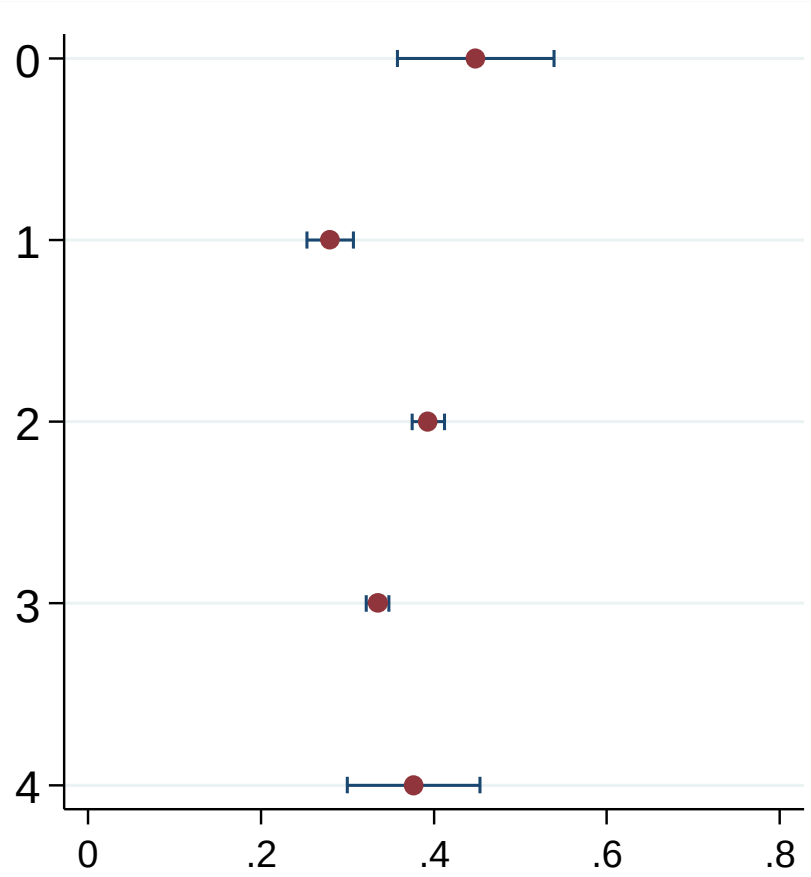
	Odds Ratio	95% CI	p-value
Avg. number of time points			<0.001
Linear term	1.66	(1.31, 2.10)	
Quadratic term	0.97	(0.95, 0.99)	
Avg. number of target lesions, ≥ 2	0.82	(0.71, 0.94)	0.007
Number of charter adjudication variables			<0.001
1	Ref	---	
2	1.46	(1.23, 1.72)	
3+	1.69	(1.43, 1.99)	

Date of Progression Only: Read Criteria



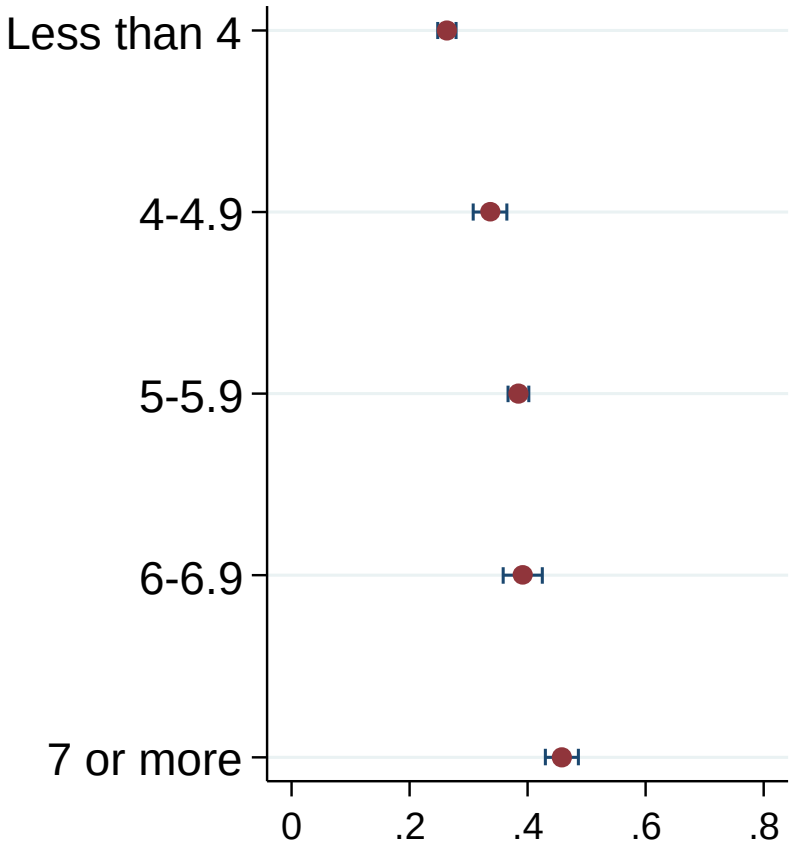
	# Trials	# Patients	Proportion Discordant
Cheson 2007	2	322	25%
Macdonald	1	276	35%
RANO	1	116	45%
RECIST 1.0	13	5151	33%
RECIST 1.1	8	2867	37%
WHO	1	398	42%

Date of Progression Only: Average Number of Target Lesions



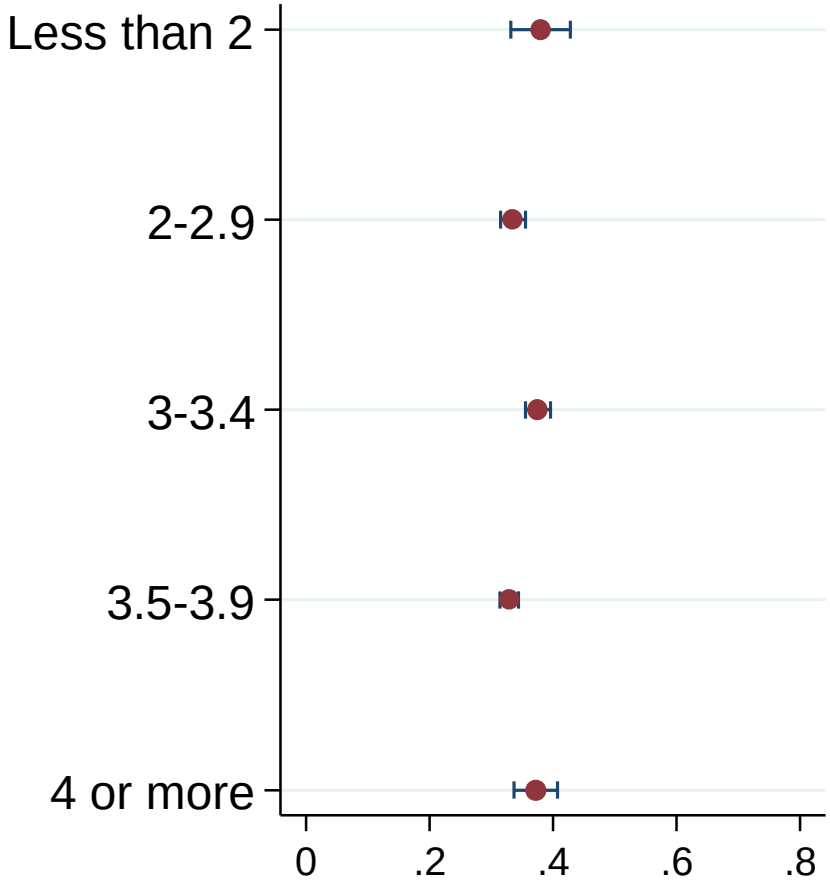
	# Trials	# Patients	Proportion Discordant
0	1	116	45%
1	4	1,072	28%
2	7	2,654	39%
3	13	5,134	33%
4	4	154	38%

Date of Progression Only: Average Number of Time Points



# Trials	# Patients	Proportion Discordant
8	3,158	26%
5	1,052	34%
7	2,859	38%
3	834	39%
3	1,227	46%

Date of Progression Only: Average Number of Exams



	# Trials	# Patients	Proportion Discordant
Less than 2	2	392	38%
2-2.9	7	2,135	34%
3-3.4	5	2,213	38%
3.5-3.9	10	3,651	33%
4 or more	2	739	37%

Conclusions

- There are several factors that may modify agreement between radiologists' assessment of clinical trial endpoints
  - Adjudication variable (endpoint), indication, number of lesions, number of time points
- These sources of variability may exist even in the absence of true errors in measurement
- Should aim to optimize study design and primary endpoint definitions so that variability in endpoint determination is reduced

Collaborators

BioClinica, Inc.

Robert Ford

Michael O'Neal

John Fraunberger

Disclosure: Funding provided by BioClinica, Inc.

Spring in New York?