

ACCURACY AND RELIABILITY OF GLEASON SCORES IN SEER DATA

Michael Goodman, MD, MPH

Kevin Ward, MPH, PhD

Emory University

Rollins School of Public Health

Georgia Center for Cancer Statistics

November 6, 2009

Background

- In 2004 SEER registries began collecting **Gleason's score** (GS) information
- **GS** is a measure of prostate cancer grade, and a determinant of patient treatment & prognosis
- GS is a subjective measure; its accuracy depends on type of specimen and physician's experience
- GS derived from prostatectomy specimens are more accurate, but from patient care point of view GS of a biopsy specimen is more important

Gleason Scale

Well differentiated



- 1 Small, uniform glands
- 2 More space between glands
- 3 Infiltration of cells from glands at margins
- 4 Irregular masses of cells with few glands
- 5 Lack of glands, sheets of cells

Poorly differentiated



- Gleason's system assigns histologic grade to predominant (primary) and lesser (secondary) pattern of tumor
- The grade numbers of the two patterns are added to obtain the Gleason score, which may range from 2 to 10

Study Aims

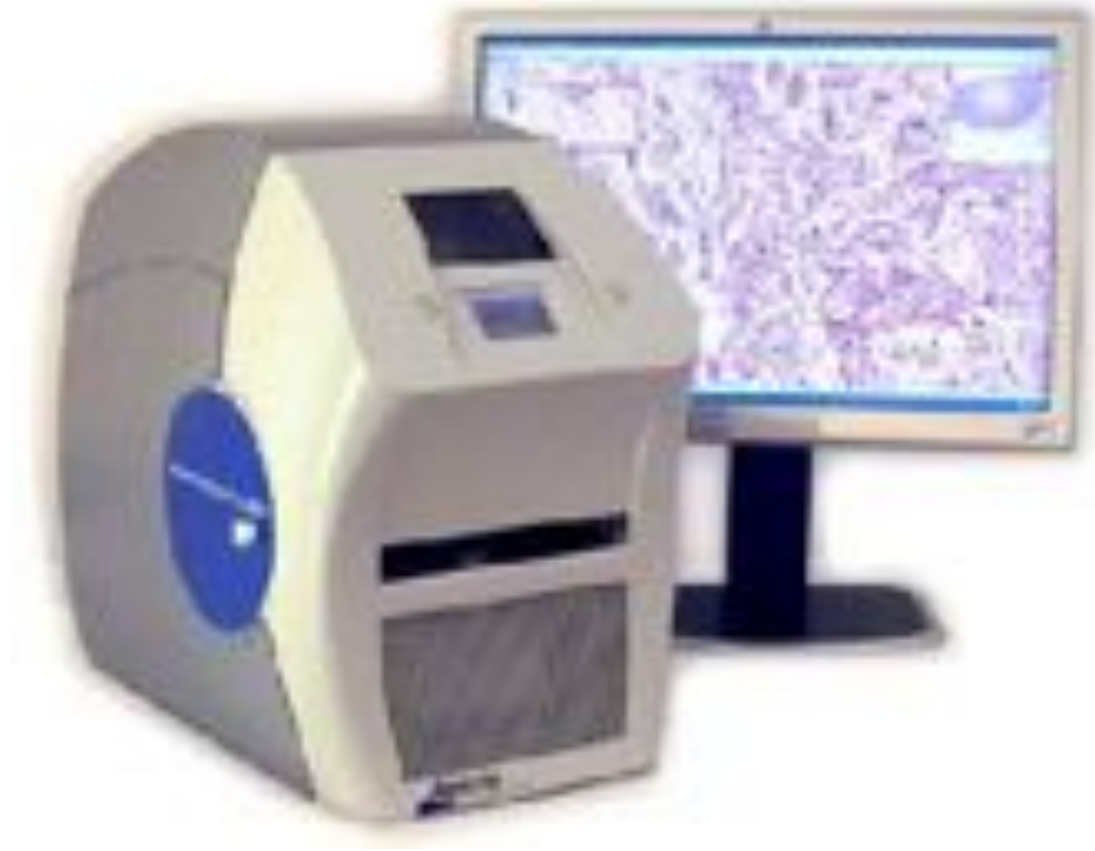
- Aim I: Examine the level of agreement between Gleason scores reported at diagnosis and those assigned by the expert review.
- Aim II: Determine the feasibility of linking registry data to pathology images and explore the opportunities for long-distance pathology consultation and conferencing.

Study overview

- ~300 prostate cancer cases selected for scanning and review of pathology slides
- Slides requested after review of pathology reports
- Sampling goal to increase representation of small and large facilities and white and African-American patients
- Original GS extracted from pathology reports
- Two urologic pathology experts reviewed the slides and assigned their own Gleason scores

Data collection steps

1. Slides brought to Emory's Winship Cancer Institute
2. Slides scanned using ScanScope® technology,
3. Slides returned to their facility of origin
4. Expert pathologist view electronic files and reassign the Gleason scores





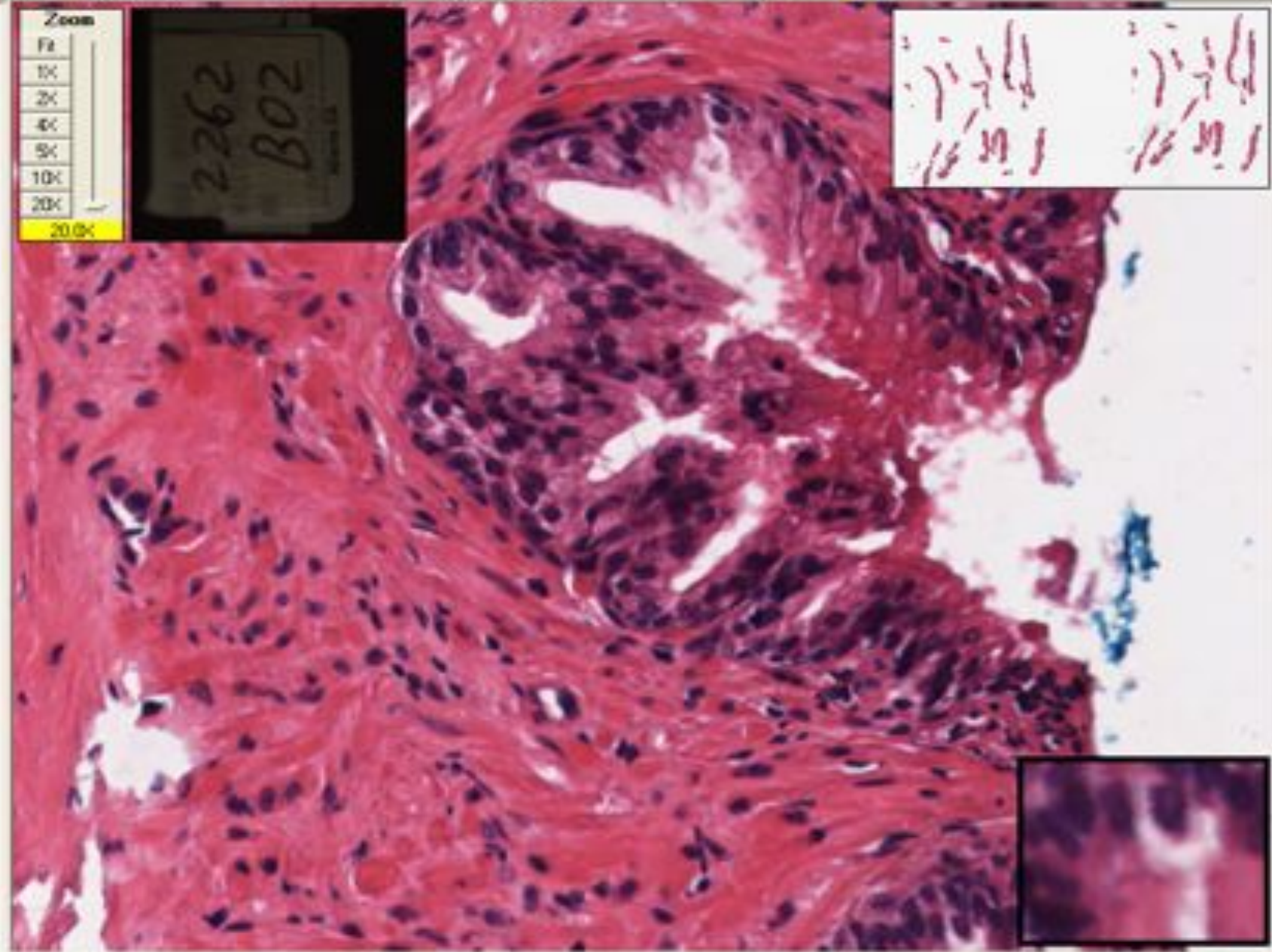
2262801.svs



2262802.svs

Zoom

Fit
1K
2K
4K
5K
10K
20K
200K



Characteristics of the study sample

Patient Characteristics	All samples (n=294)	Biopsies (n=195)		Prostatectomies (n=99)		
	Number	Number	Percent	Number	Percent	
Diagnosis age						
<50	30	20	10.3%	10	10.1%	
50-59	104	57	29.2%	47	47.5%	
60-69	116	75	38.5%	41	41.4%	
70-79	34	33	16.9%	1	1.0%	
80+	10	10	5.1%	0	0.0%	
Race						
White	144	92	47.2%	52	52.5%	
Black	150	103	52.8%	47	47.5%	

Characteristics of the study sample

Patient Characteristics	All samples (n=294)	Biopsies (n=195)		Prostatectomies (n=99)		
	Number	Number	Percent	Number	Percent	
Disease stage						
Localized	248	165	84.6%	83	83.8%	
Regional	33	17	8.7%	16	16.2%	
Distant	13	13	6.7%	0	0.0%	
Facility type						
University-affiliated hospitals	93	58	29.7%	35	35.4%	
Large community hospitals	86	36	18.5%	50	50.5%	
Commercial laboratories	69	69	35.4%	0	0.0%	
Small community hospitals	46	32	16.4%	14	14.1%	

Biopsy Gleason scores assigned by different reviewers

Biopsy specimens (n=195)	Pathology reports		Reviewer #1		Reviewer #2	
	Number	Percent	Number	Percent	Number	Percent
2	0	0.0%	0	0.0%	0	0.0%
3	2	1.0%	0	0.0%	0	0.0%
4	0	0.0%	0	0.0%	0	0.0%
5	0	0.0%	0	0.0%	0	0.0%
6	107	54.9%	99	50.8%	71	36.4%
7	60	30.8%	72	36.9%	83	42.6%
8	11	5.6%	13	6.7%	16	8.2%
9	13	6.7%	11	5.6%	24	12.3%
10	2	1.0%	0	0.0%	1	0.5%

Friedman test for three related samples: $P < 0.0001$

Prostatectomy Gleason scores assigned by different reviewers

Prostatectomy specimens (n=99)	Pathology reports		Reviewer #1		Reviewer#2	
	Number	Percent	Number	Percent	Number	Percent
2	0	0.0%	0	0.0%	0	0.0%
3	0	0.0%	0	0.0%	0	0.0%
4	0	0.0%	0	0.0%	0	0.0%
5	2	2.0%	0	0.0%	1	1.0%
6	51	51.5%	56	56.6%	44	44.4%
7	39	39.4%	37	37.4%	44	44.4%
8	3	3.0%	2	2.0%	7	7.1%
9	3	3.0%	4	4.0%	3	3.0%
10	1	1.0%	0	0.0%	0	0.0%

Friedman test for three related samples: P=0.09

Accuracy and reliability of biopsy-derived grades using Reviewer 1 as gold standard

Reviewer 2	Reviewer 1 (gold standard)		Total		
	Poorly differentiated	Moderately differentiated			
Poorly differentiated	88	36	124	Sensitivity	0.92
Moderately differentiated	8	63	71	Specificity	0.64
	96	99	195	Kappa	0.55

Pathology reports	Reviewer 1 (gold standard)		Total		
	Poorly differentiated	Moderately differentiated			
Poorly differentiated	72	14	86	Sensitivity	0.75
Well-Moderately differentiated*	24	85	109	Specificity	0.86
	96	99	195	Kappa	0.61

* Includes two specimens with Gleason score of 3

Accuracy and reliability of prostatectomy-derived grades using Reviewer 1 as gold standard

Reviewer 2	Reviewer 1 (gold standard)		Total		
	Poorly differentiated	Moderately differentiated			
Poorly differentiated	38	16	54	Sensitivity	0.88
Moderately differentiated	5	40	45	Specificity	0.71
	43	56	99	Kappa	0.58

Pathology reports	Reviewer 1 (gold standard)		Total		
	Poorly differentiated	Moderately differentiated			
Poorly differentiated	33	13	46	Sensitivity	0.77
Moderately differentiated	10	43	53	Specificity	0.77
	43	56	99	Kappa	0.53

Conclusions

- A substantial proportion of biopsy and prostatectomy specimens have different Gleason scores assigned by different pathologists
- The magnitude of disagreement with respect to grade is modest
- Our study also demonstrates the feasibility of linking registry data with digitized pathology slides