On-line EQA

Steve Aldington
National Education and On-line EQA Manager

Irene Stratton
Senior Statistician

English National Screening Programme for DR

NHS
Screening Programmes
Diabetic Retinopathy

Topics for today

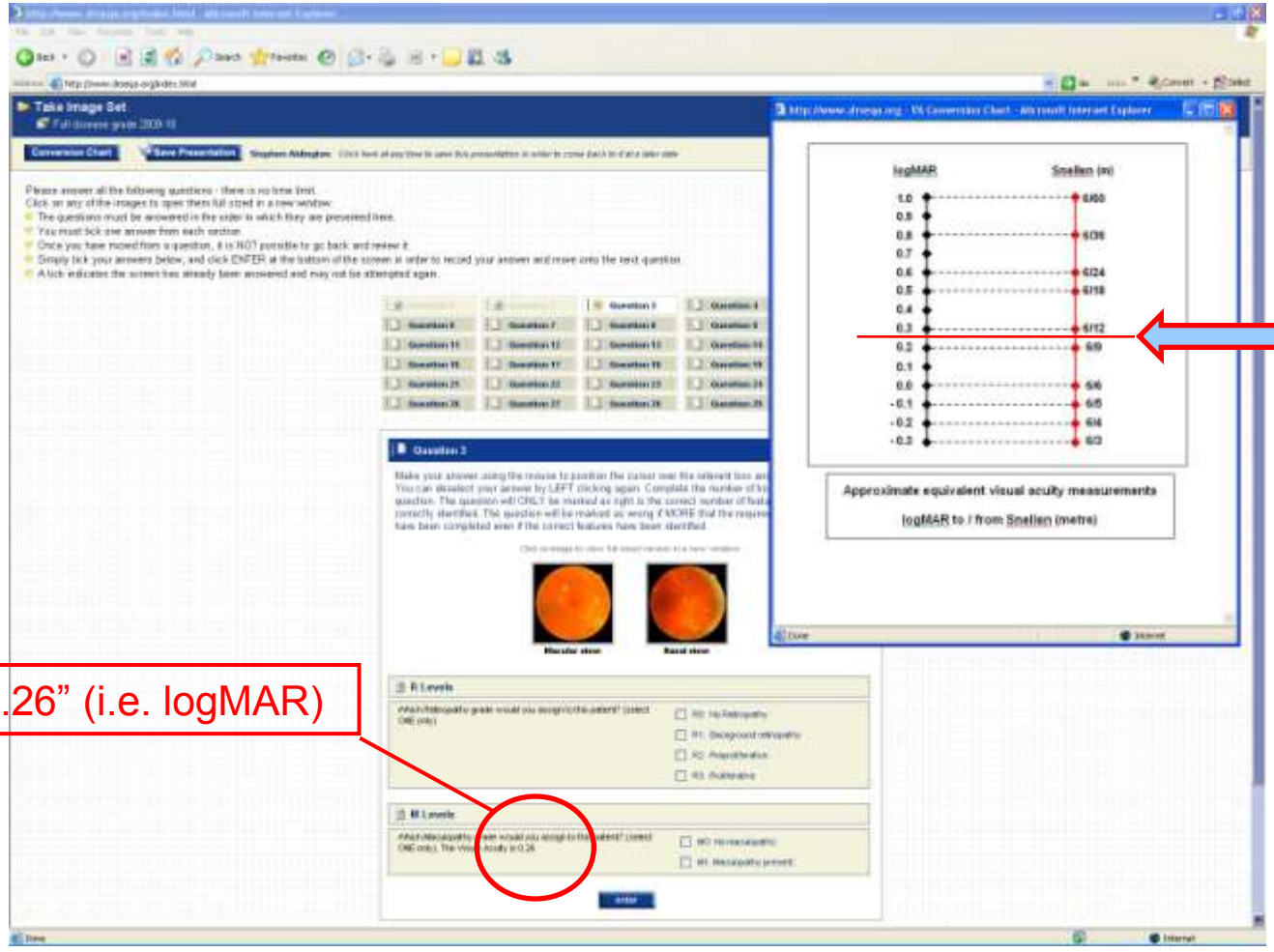
- Overview
- Visual acuity in the OL-EQA
- 2009-10 reports and main data results
- Interpretation and national standards
- Future developments

Overview

- EQA grading test set(s) were needed
- The Full Disease Grading test set was successfully piloted ending April 2009
- Nine programmes and 81 graders took part. Feedback led to improvements
- Phased roll-out to 92 English programmes between Aug-Dec ’09: 1st year ending Dec ’10
- Participation has been very good
- There have been some complaints

Visual acuity in the text

VA conversion chart



"VA = 0.26" (i.e. logMAR)

Question for the audience

- Is it appropriate to provide logMAR VA in the text and a Snellen conversion chart?
- Would Snellen VA be more appropriate?
- Should we give both in the text?
- How precisely should Snellen equivalent be given and is this valid?

Data and reporting of results:

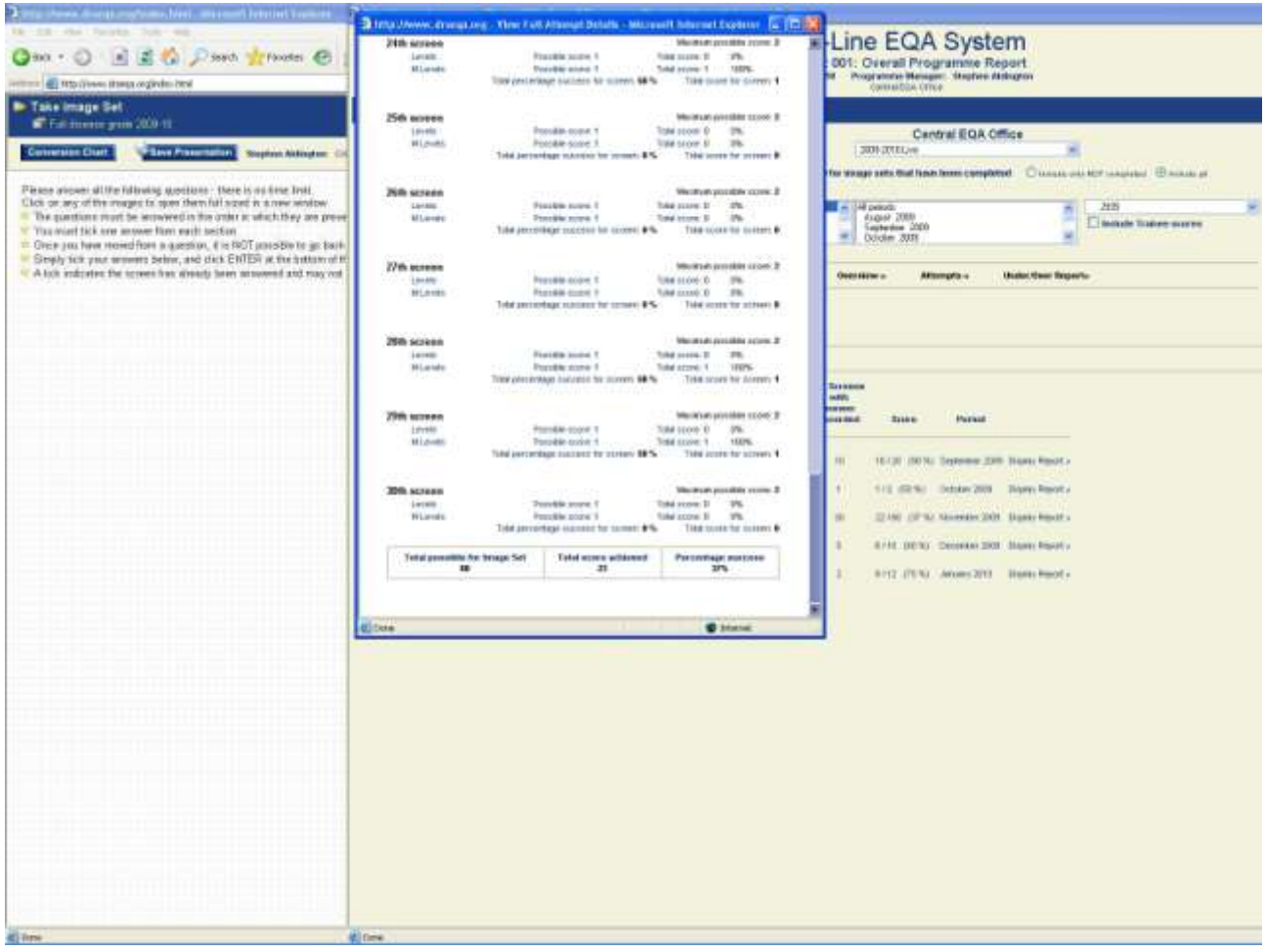
Main 2009-10 test set

Access to data and reports

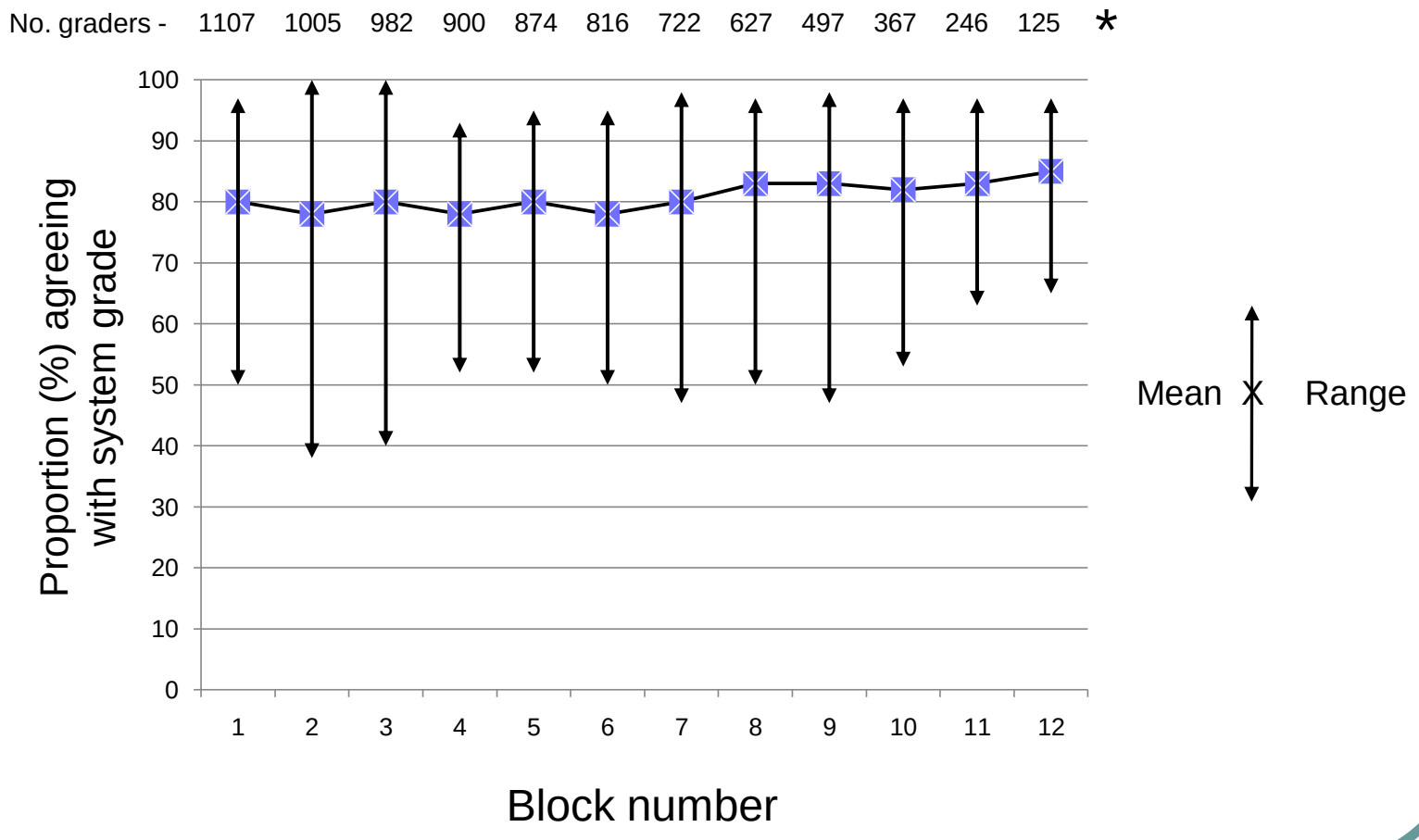
- User can currently see only their block % agreement score against system answers and an Over/Under report
- PM/ACL can access (anonymised) comparison data reports for each grader/block/month
- ? Most useful reports for PM and ACL – Over/Under (Stats) and Table of Responses

Review of images is not possible by anyone

Individual grader result



Proportional agreement score by block
(all programmes, all graders at 12/09/10)



* = total of 8268 blocks of 30 = almost 250,000 eyes!

Statistics (under/over) and Level of Agreement reports

Statistics chart – grader 1 (7 blocks)

Print Report »

This report contains only records where the image set **HAS** been completed This report does **NOT** include answers recorded by trainees

Statistics Chart

Level	No. times Presented as correct	No. times answered correctly	% times answered correctly	No. times Overgraded		% of times Overgraded	No. of times Undergraded		% of times Undergraded
R	210	144	69 %	47		22 %	19		9 %
R0	23	16	70 %	7	R1 - 5 R2 - 1 R3 - 1	30 %	0		0
R1	100	65	65 %	35	R2 - 27 R3 - 8	35 %	0	R0 - 0	0 %
R2	37	23	62 %	5	R3 - 5	14 %	9	R0 - 0 R1 - 9	24 %
R3	50	40	80 %	0		0	10	R0 - 0 R1 - 0 R2 - 10	20 %
M	210	177	84 %	32		15 %	1		0 %
M0	139	107	77 %	32	M1 - 32	23 %	0		0
M1	71	70	99 %	0		0	1	M0 - 1	1 %

Export to Excel »

Statistics chart – grader 2 (10 blocks)

Print Report »

This report contains only records where the image set **HAS** been completed This report does **NOT** include answers recorded by trainees

Statistics Chart

Level	No. times Presented as correct	No. times answered correctly	% times answered correctly	No. times Overgraded		% of times Overgraded	No. of times Undergraded		% of times Undergraded
R	300	253	84 %	16		5 %	31		10 %
R0	47	41	87 %	6	R1 - 6 R2 - 0 R3 - 0	13 %	0		0
R1	138	129	93 %	6	R2 - 4 R3 - 2	4 %	3	R0 - 3	2 %
R2	58	39	67 %	4	R3 - 4	7 %	15	R0 - 0 R1 - 15	26 %
R3	57	44	77 %	0		0	13	R0 - 0 R1 - 2 R2 - 11	23 %
M	300	276	92 %	14		5 %	10		3 %
M0	207	193	93 %	14	M1 - 14	7 %	0		0
M1	93	83	89 %	0		0	10	M0 - 10	11 %

Export to Excel »

Level of Agreement – grader 1 (7 blocks)

Level of Agreeement Export to Excel »

Block 0

EQA Grade	R0	R1M0	R1M1	R2M0	R2M1	R3M0	R3M1
R0	16	3	2	1	0	1	0
R1M0	0	35	9	9	10	6	1
R1M1	0	1	20	0	8	0	1
R2M0	0	7	1	13	3	2	0
R2M1	0	0	1	0	7	0	3
R3M0	0	0	0	4	1	10	5
R3M1	0	0	0	0	5	0	25

Not referred:
8/117 (6.9%)

Exact:
126/210 (60%)

Level of Agreement – grader 2 (10 blocks)

Level of Agreeement Export to Excel »

Block 0

EQA Grade	R0	R1M0	R1M1	R2M0	R2M1	R3M0	R3M1
R0	41	5	1	0	0	0	0
R1M0	3	78	11	2	0	1	1
R1M1	0	2	38	0	2	0	0
R2M0	0	5	0	31	0	4	0
R2M1	0	2	8	0	8	0	0
R3M0	0	2	0	4	0	17	1
R3M1	0	0	0	4	3	2	24

Not referred:
11/157 (7.0%)

Exact:
235/300 (79%)

Which metric to use?

- Total of R + M Agreement versus system
- Agreement of R or M versus system
- Referred versus not referred vs system
- Missed R3 cases (number or proportion)
- The severity of missed cases
- Over-grading as well as under-grading
- Cost to the NHS / cost to the patient
- How important is the error…

Where in the grading pathway?

- Should best graders be Primary graders
- Or Second level graders
- Or Arbitration level graders
- Where should Trainees be placed

- Implications of grader errors

Table of Responses reports

Table of Responses (PM / ACL)

Table of responses

Block 6	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
EQA Grade	R0M0	R0M0	R0M0	R0M0	R0M0	R0M0	R0M0	R1M0	R1M0	R1M0	R1M0	R1M0	R1M0	R1M0	R1M0	R1M0
	R0M0	R1M0	R0M0	R0M0	R0M0	R1M0	R1M0	R1M1	R1M0	R1M0	R1M0	R1M0	R1M0	R1M0	R1M0	R2M0
	R0M0	R0M0	R0M0	R0M0	R0M0	R0M0	R0M0	R0M0	R1M0	R1M0	R1M0	R1M0	R1M0	R1M0	R1M0	R0M0
	R0M0	R0M0	R0M0	R0M0	R0M0	R0M0	R0M0	R1M0	R1M0	R1M0	R1M0	R1M0	R1M0	R1M0	R1M0	R2M0
	R0M0	R1M0	R1M0	R1M0	R0M0	R1M0	R1M0	R1M0	R2M0	R1M0	R1M0	R1M0	R1M1	R1M0	R1M1	R2M0
	R0M0	R0M0	R1M0	R1M0	R1M0	R1M0	R1M0	R2M1	R1M0	R2M0	R1M0	R1M0	R2M0	R1M0	R2M1	R2M0

P	Q	R	S	T	U	V	W	X	Y	Z	AB	AC	AD	AE
R1M0	R1M1	R1M1	R2M0	R2M0	R2M0	R2M0	R2M0	R2M0	R2M0	R2M0	R2M0	R3M1	R3M1	R3M1
R2M0	R1M1	R1M1	R2M0	R1M0	R1M0	R1M0	R1M1	R2M0	R2M0	R2M0	R2M0	R2M1	R2M1	R2M1
R2M0	R1M1	R2M0	R2M0	R2M0	R1M0	R2M0	R1M0	R2M0	R2M0	R2M0	R2M0	R2M0	R2M1	R2M1
R2M0	R1M1	R1M1	R2M0	R2M0	R2M0	R2M0	R2M0	R2M1	R2M0	R2M0	R2M0	R2M1	R2M1	R2M1
R2M0	R1M1	R1M1	R2M0	R2M0	R2M0	R2M0	R2M1	R1M1	R2M0	R2M0	R2M0	R2M1	R2M1	R2M1
R2M0	R1M1	R1M1	R2M0	R2M0	R2M0	R2M0	R2M0	R2M1	R2M0	R2M0	R2M0	R2M1	R2M1	R2M1

Table of Responses (Steve A)

Program[illegible]e A[illegible]bing Block 10

O Id: 1744 » Title: Screen 51	P Id: 1808 » Title: Screen 62	Q Id: 1814 » Title: Screen 66	R Id: 1721 » Title: Screen 46	S Id: 1742 » Title: Screen 49	T Id: 1736 » Title: Screen 54	U Id: 1743 » Title: Screen 60	V Id: 1793 » Title: Screen 58	W Id: 1815 » Title: Screen 61	X Id: 1795 » Title: Screen 60	Y Id: 1817 » Title: Screen 65	Z Id: 1794 » Title: Screen 59	AB Id: 1807 » Title: Screen 63	AC Id: 1816 » Title: Screen 64	AD Id: 1730 » Title: Screen 53
R1M0	R1M0	R1M0	R1M1	R1M1	R1M1	R2M0	R2M0	R2M0	R2M0	R2M0	R2M0	R2M0	R2M1	R2M1
R1M0	R1M1	R1M0	R1M1	R1M1	R1M1	R2M0	R2M0	R1M0	R2M0	R2M0	R2M1	R1M0	R2M0	R1M1
R1M0	R1M1	R1M0	R1M1	R1M0	R1M1	R2M0	R2M0	R2M0	R2M0	R2M0	R2M0	R1M0	R2M1	R2M1
R1M0	R2M0	R1M1	R0M1	R1M1	R1M0	R2M0	R2M0	R1M0	R2M0	R2M0	R2M0	R1M0	R2M1	R2M1
R1M1	R1M1	R1M0	R1M1	R1M1	R1M1	R2M0	R2M0	R1M0	R2M0	R2M1	R3M1	R2M0	R2M1	R2M1
R1M0	R1M1	R1M0	R1M1	R1M1	R1M1	R2M0	R2M0	R2M0	R2M0	R2M0	R2M1	R2M0	R2M1	R2M1
R1M0	R1M1	R1M0	R1M1	R1M1	R1M1	R0M0	R2M0	R1M0	R2M0	R2M0	R2M0	R1M0	R2M1	R2M1

Program[illegible]e B[illegible]bing Block 10

O Id: 1744 » Title: Screen 51	P Id: 1808 » Title: Screen 62	Q Id: 1814 » Title: Screen 66	R Id: 1721 » Title: Screen 46	S Id: 1742 » Title: Screen 49	T Id: 1736 » Title: Screen 54	U Id: 1743 » Title: Screen 60	V Id: 1793 » Title: Screen 58	W Id: 1815 » Title: Screen 61	X Id: 1795 » Title: Screen 60	Y Id: 1817 » Title: Screen 65	Z Id: 1794 » Title: Screen 59	AB Id: 1807 » Title: Screen 63	AC Id: 1816 » Title: Screen 64	AD Id: 1730 » Title: Screen 53
R1M0	R1M0	R1M0	R1M1	R1M1	R1M1	R2M0	R2M0	R2M0	R2M0	R2M0	R2M0	R2M0	R2M1	R2M1
R2M1	R1M0	R1M1	R1M1	R0M0	R1M1	R2M0	R3M0	R2M0	R3M0	R2M0	R3M0	R1M0	R1M0	R2M1
R1M1	R1M1	R1M1	R1M1	R1M1	R1M1	R2M0	R2M0	R2M0	R2M0	R3M0	R2M1	R1M0	R2M1	R2M1
R1M1	R1M1	R1M1	R1M1	R1M1	R1M1	R2M0	R2M0	R2M0	R2M0	R3M1	R2M0	R1M0	R1M1	R2M1
R1M0	R1M1	R1M1	R1M1	R1M1	R1M1	R2M0	R2M0	R1M0	R2M0	R2M1	R3M0	R1M0	R2M1	R1M1
R2M0	R1M1	R1M1	R0M0	R1M1	R1M0	R2M0	R2M0	R1M0	R2M0	R2M0	R2M0	R2M0	R1M1	R1M1
R2M1	R1M1	R2M0	R1M1	R1M1	R1M1	R2M0	R2M0	R2M0	R1M0	R2M0	R2M0	R2M0	R1M1	R2M1

Challenges

- Reluctance / refusal to participate:
  - As individuals
  - As Programmes
  - As ‘a group’
- Cost implications (60-90 mins per test/month)
- No ‘pass mark’ or case feedback/image review
- Poor local network connection speeds
- Some definitions in R2 and M1
- Testing frequency
- Implications of poor performance

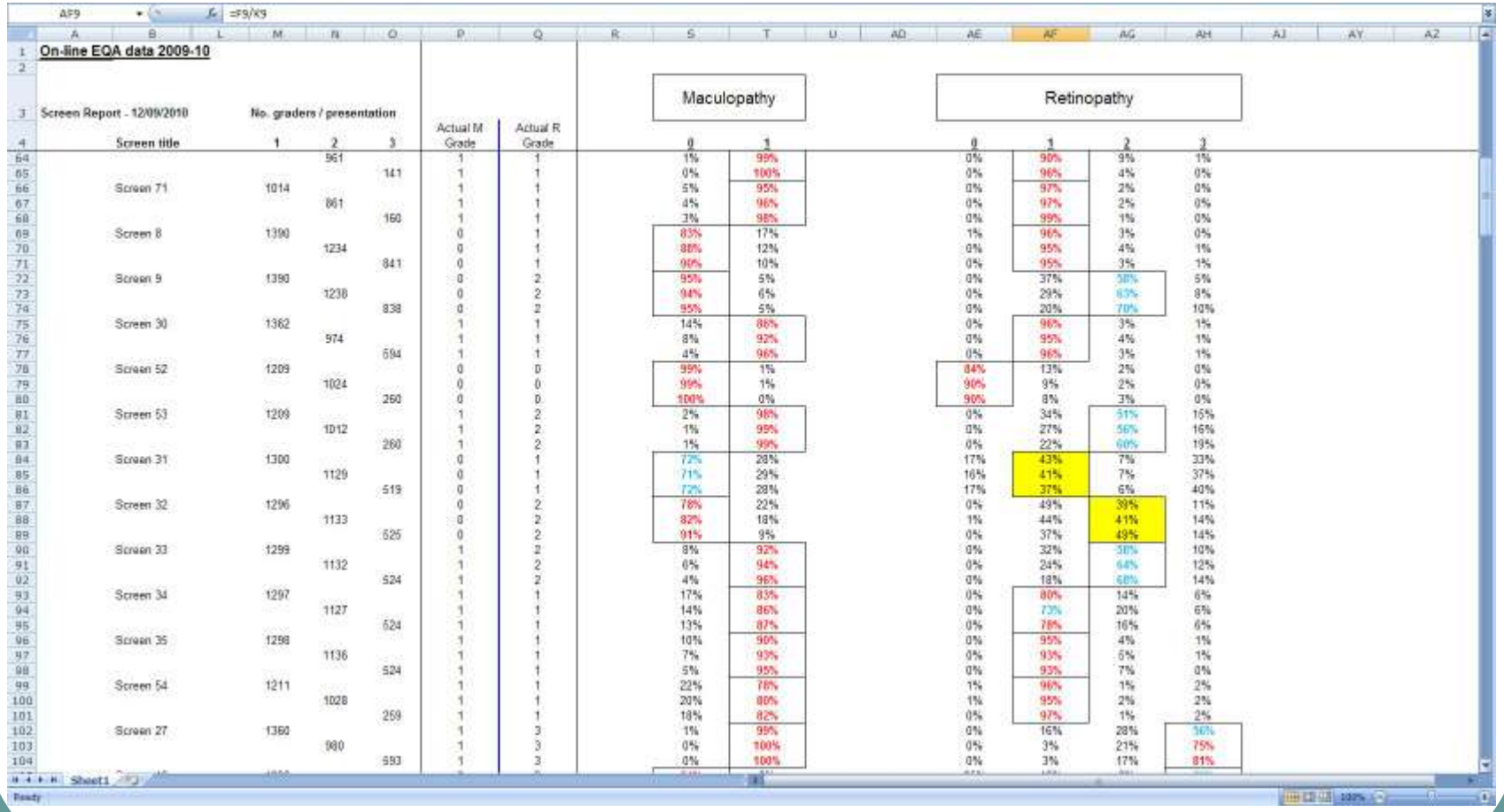
Interpretation and significance of On-line EQA test results

- Remember: a disease- and referral-positive weighted sample is being tested
- Overall agreement runs about 78-83%
- There are some cases where few people agree with the ‘system’ grade
- There are some cases where few people agree with each other

Is proportional agreement appropriate?

What is an appropriate ‘pass mark’?

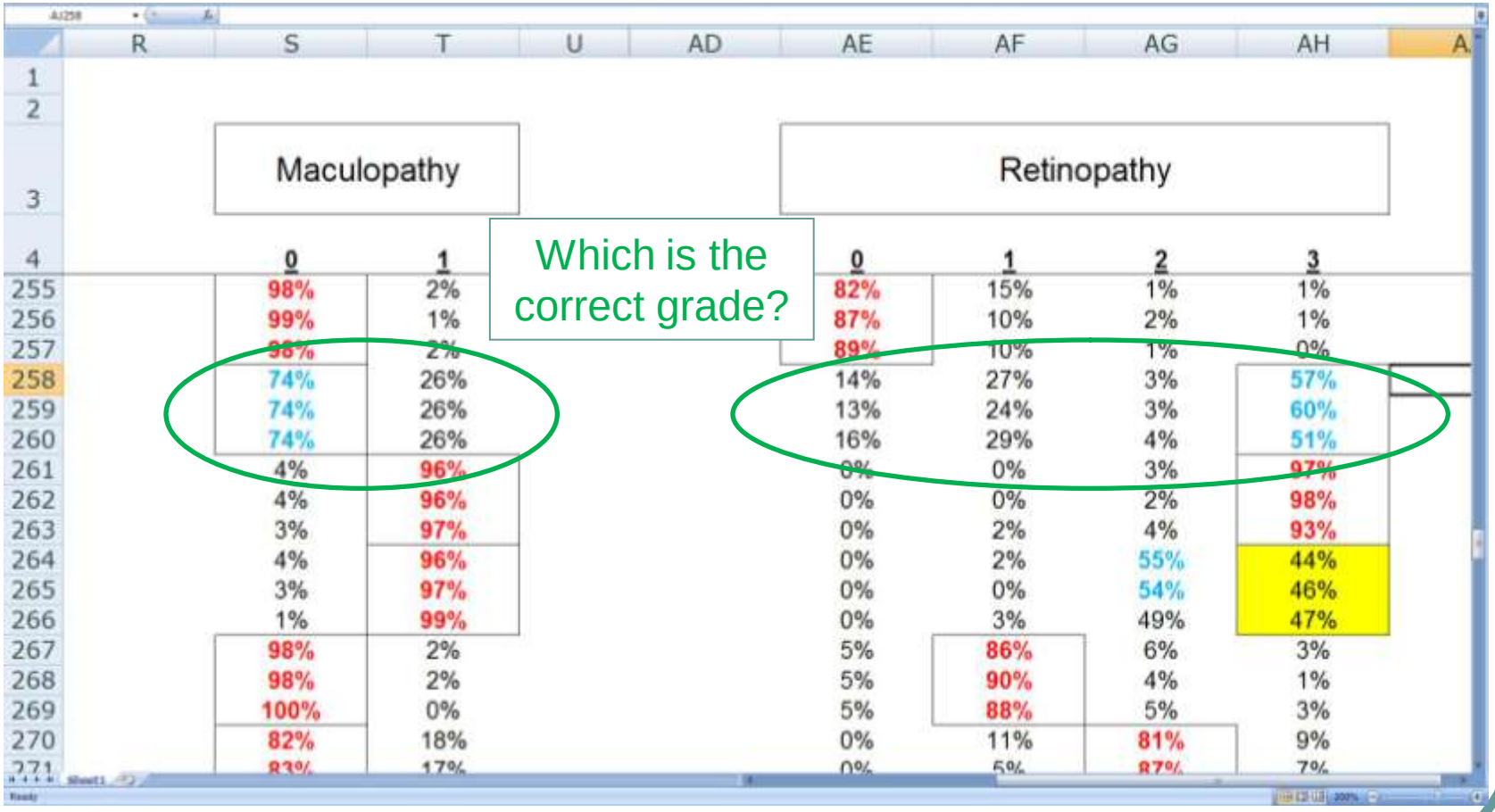
National agreement data



On-line EQA data 2009-10

Screen Report - 12/09/2010		No. graders / presentation			Actual M Grade	Actual R Grade	Maculopathy		Retinopathy			
	Screen title	1	2	3			0	1	0	1	2	3
64			961		1	1	1%	99%	0%	90%	9%	1%
65				141	1	1	0%	100%	0%	96%	4%	0%
66	Screen 71	1014			1	1	5%	95%	0%	97%	2%	0%
67			861		1	1	4%	96%	0%	97%	2%	0%
68				160	1	1	3%	98%	0%	99%	1%	0%
69	Screen 8	1390			0	1	83%	17%	1%	96%	3%	0%
70			1234		0	1	88%	12%	0%	95%	4%	1%
71				841	0	1	90%	10%	0%	95%	3%	1%
72	Screen 9	1390			0	2	95%	5%	0%	37%	58%	5%
73			1238		0	2	94%	6%	0%	29%	63%	8%
74				838	0	2	95%	5%	0%	20%	70%	10%
75	Screen 30	1362			1	1	14%	86%	0%	96%	3%	1%
76			974		1	1	8%	92%	0%	95%	4%	1%
77				594	1	1	4%	96%	0%	96%	3%	1%
78	Screen 52	1209			0	0	99%	1%	84%	13%	2%	0%
79			1024		0	0	99%	1%	90%	9%	2%	0%
80				260	0	0	100%	0%	96%	8%	3%	0%
81	Screen 53	1209			1	2	2%	98%	0%	34%	51%	15%
82			1012		1	2	1%	99%	0%	27%	56%	16%
83				260	1	2	1%	99%	0%	22%	60%	19%
84	Screen 31	1300			0	1	72%	28%	17%	43%	7%	33%
85			1129		0	1	71%	29%	16%	41%	7%	37%
86				519	0	1	72%	28%	17%	37%	6%	40%
87	Screen 32	1296			0	2	78%	22%	0%	49%	39%	11%
88			1133		0	2	82%	18%	1%	44%	41%	14%
89				525	0	2	91%	9%	0%	37%	49%	14%
90	Screen 33	1299			1	2	8%	92%	0%	32%	58%	10%
91			1132		1	2	6%	94%	0%	24%	64%	12%
92				524	1	2	4%	96%	0%	18%	68%	14%
93	Screen 34	1297			1	1	17%	83%	0%	80%	14%	6%
94			1127		1	1	14%	86%	0%	73%	20%	6%
95				524	1	1	13%	87%	0%	78%	16%	6%
96	Screen 35	1298			1	1	10%	90%	0%	95%	4%	1%
97			1136		1	1	7%	93%	0%	93%	6%	1%
98				524	1	1	5%	95%	0%	93%	7%	0%
99	Screen 54	1211			1	1	22%	78%	1%	96%	1%	2%
100			1028		1	1	20%	80%	1%	95%	2%	2%
101				259	1	1	18%	82%	0%	97%	1%	2%
102	Screen 27	1360			1	3	1%	99%	0%	16%	28%	56%
103			980		1	3	0%	100%	0%	3%	21%	75%
104				593	1	3	0%	100%	0%	3%	17%	81%

Non-uniformity of some results



	R	S	T	U	AD	AE	AF	AG	AH	A
1										
2										
3		Maculopathy				Retinopathy				
4		0	1			0	1	2	3	
255		98%	2%			82%	15%	1%	1%	
256		99%	1%			87%	10%	2%	1%	
257		98%	2%			89%	10%	1%	0%	
258		74%	26%			14%	27%	3%	57%	
259		74%	26%			13%	24%	3%	60%	
260		74%	26%			16%	29%	4%	51%	
261		4%	96%			0%	0%	3%	97%	
262		4%	96%			0%	0%	2%	98%	
263		3%	97%			0%	2%	4%	93%	
264		4%	96%			0%	2%	55%	44%	
265		3%	97%			0%	0%	54%	46%	
266		1%	99%			0%	3%	49%	47%	
267		98%	2%			5%	86%	6%	3%	
268		98%	2%			5%	90%	4%	1%	
269		100%	0%			5%	88%	5%	3%	
270		82%	18%			0%	11%	81%	9%	
271		83%	17%			0%	5%	87%	7%	

Which is the correct grade?

National proportional distribution

Of the 120 screens, each now viewed 3x:

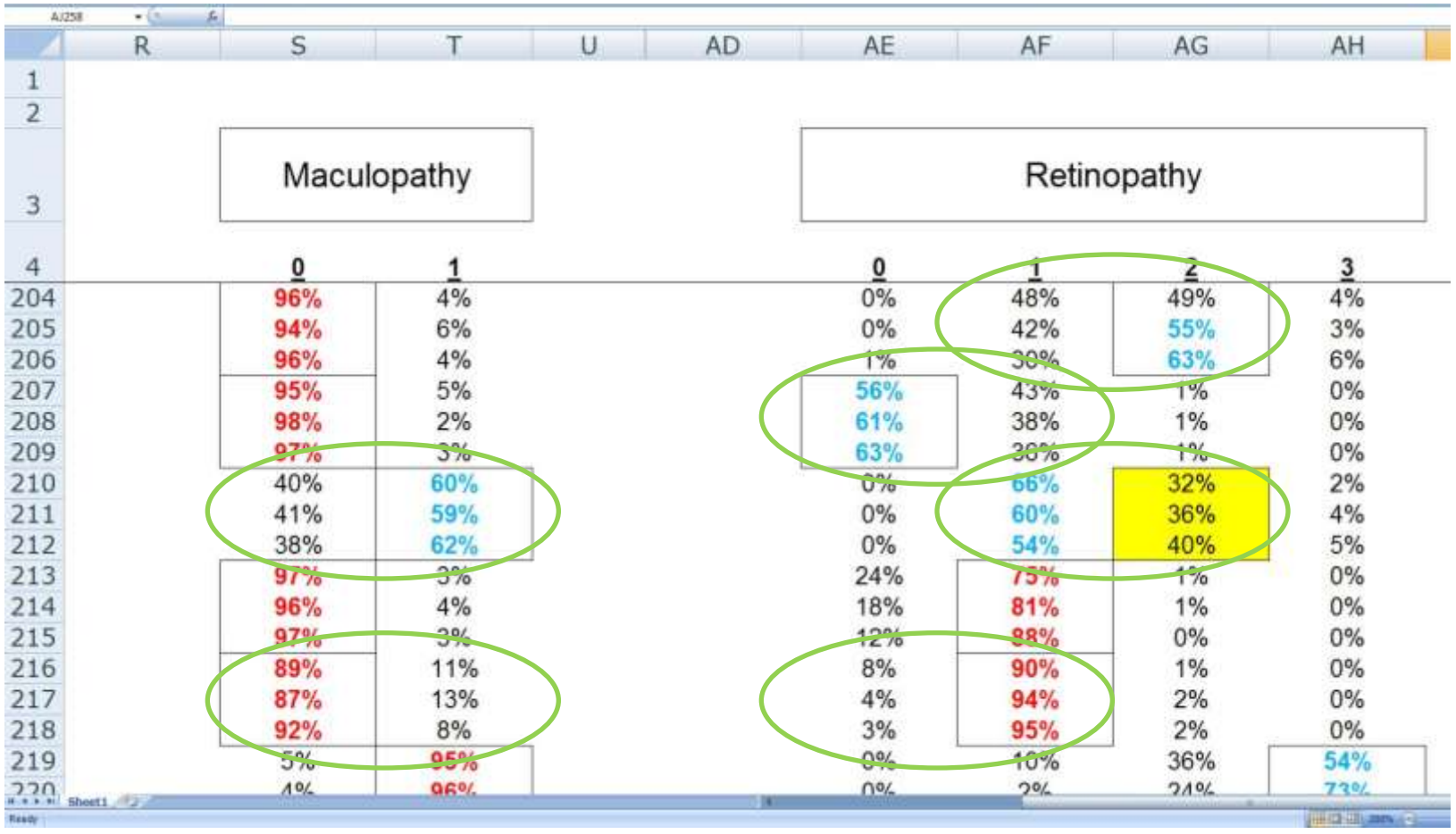
- R level ‘agreement’ by >50% graders:
  - 1st presentation: 109 / 120 (91%)
  - 2nd presentation: 114 / 120 (95%)
  - 3rd presentation: 109 / 120 (91%)
- R level ‘agreement’ by >75% graders:
  - 1st presentation: 67 / 120 (56%)
  - 2nd presentation: 67 / 120 (56%)
  - 3rd presentation: 69 / 120 (58%)

National proportional distribution

Of the 120 screens, each now viewed 3x:

- M level ‘agreement’ by >50% graders:
  - 1st presentation: 120 / 120 (100%)
  - 2nd presentation: 120 / 120 (100%)
  - 3rd presentation: 120 / 120 (100%)
- M level ‘agreement’ by >75% graders:
  - 1st presentation: 100 / 120 (83%)
  - 2nd presentation: 102 / 120 (85%)
  - 3rd presentation: 105 / 120 (85%)

Variable agreement with ‘system’



	R	S	T	U	AD	AE	AF	AG	AH
1									
2									
3		Maculopathy				Retinopathy			
4		0	1			0	1	2	3
204		96%	4%			0%	48%	49%	4%
205		94%	6%			0%	42%	55%	3%
206		96%	4%			1%	30%	63%	6%
207		95%	5%			56%	43%	1%	0%
208		98%	2%			61%	38%	1%	0%
209		97%	3%			63%	36%	1%	0%
210		40%	60%			0%	66%	32%	2%
211		41%	59%			0%	60%	36%	4%
212		38%	62%			0%	54%	40%	5%
213		97%	3%			24%	75%	1%	0%
214		96%	4%			18%	81%	1%	0%
215		97%	3%			12%	88%	0%	0%
216		89%	11%			8%	90%	1%	0%
217		87%	13%			4%	94%	2%	0%
218		92%	8%			3%	95%	2%	0%
219		5%	95%			0%	10%	36%	54%
220		4%	96%			0%	2%	24%	73%

Is proportional agreement with ‘system’ grade (ever) the best measure?

Is there a better metric for measuring and reporting ‘performance’?

Future Direction of the On-line EQA tests

Key factors for 2011 and beyond

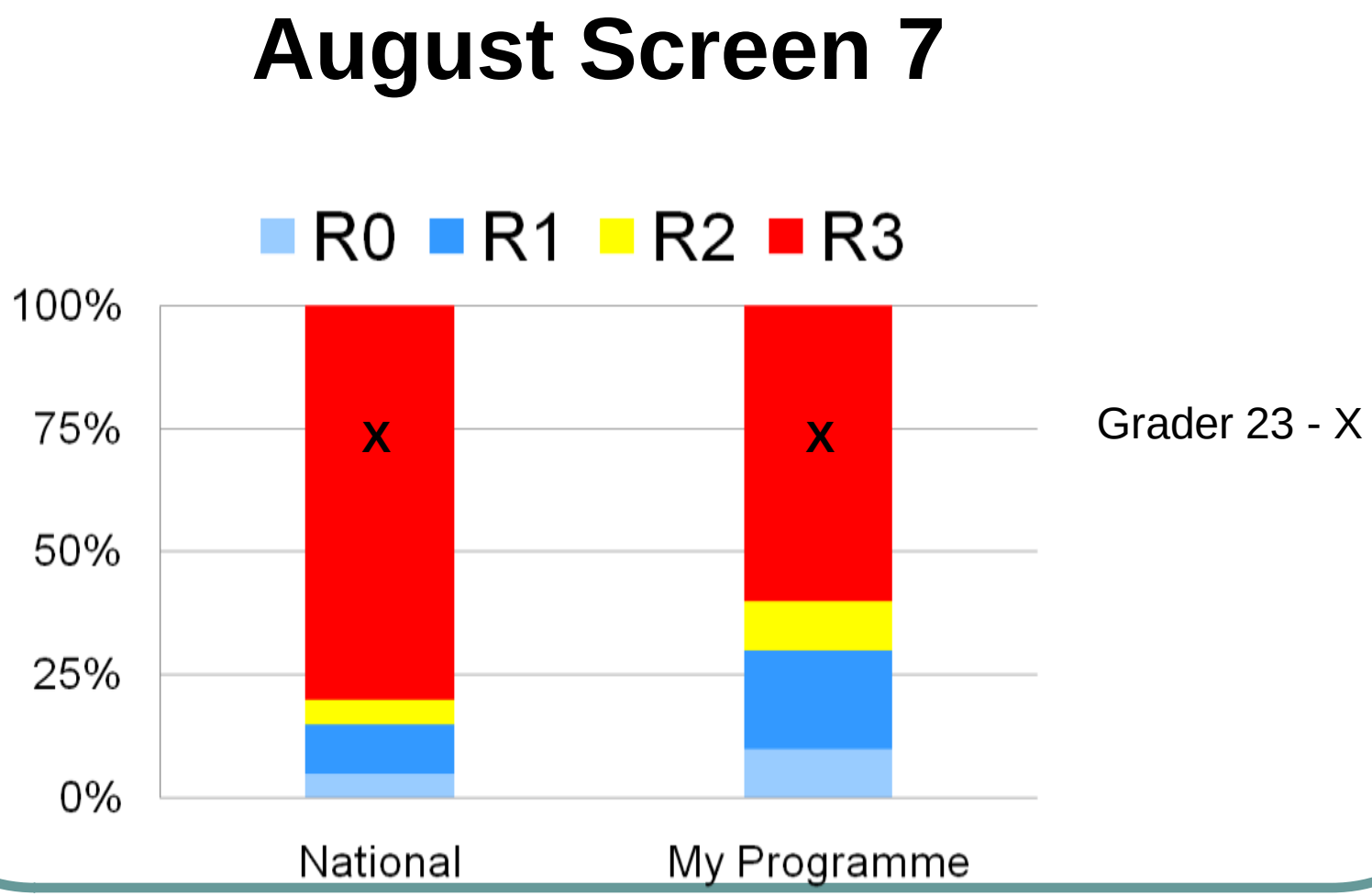
- Screeners and graders want to get it right
- No access to images is a major constraint
- Simple score for agreement against system grade may be problematic
- Difficult for ophthalmology / HES to help provide remedial staff training
- How do we do more to support training?

The future for OL-EQA

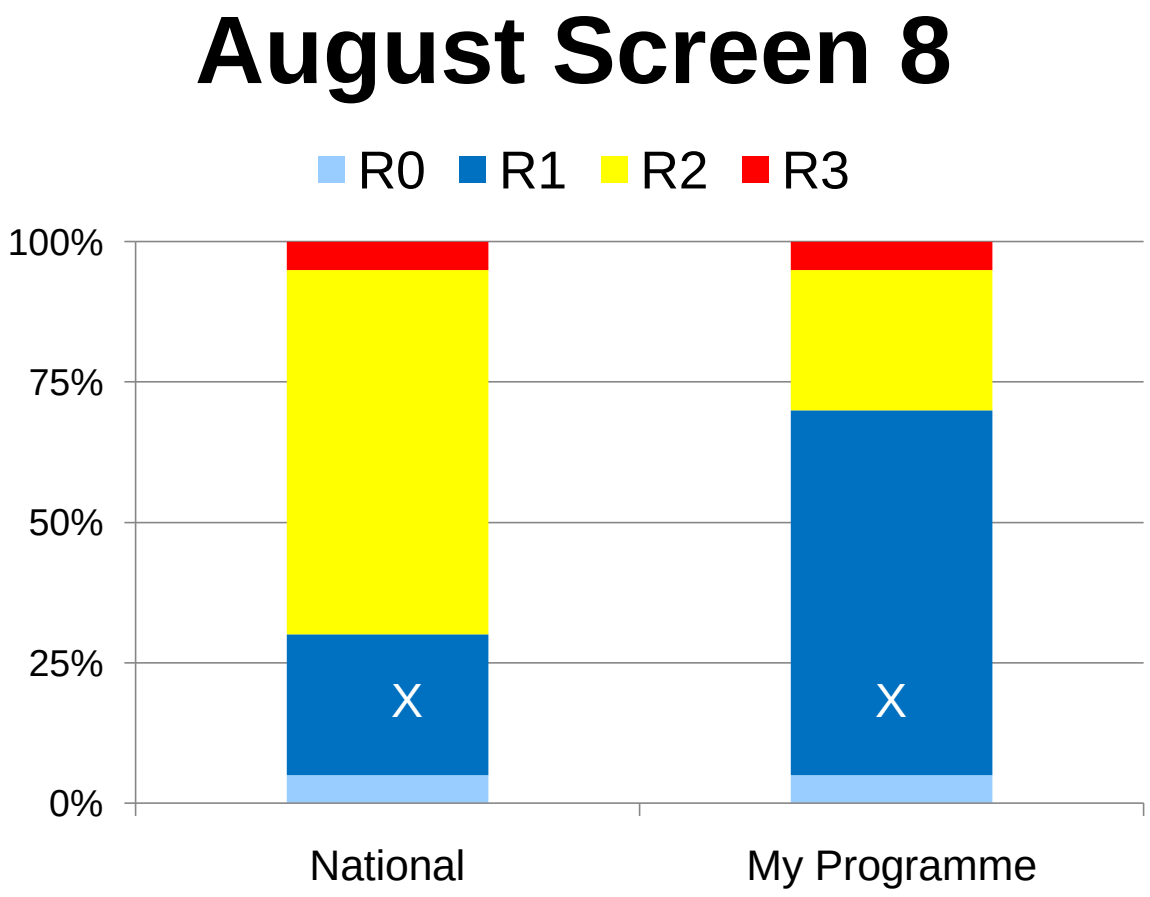
Enhancing the usefulness of OL-EQA:

- Concept shift from EQA to ‘Test and Training’
- Fixed monthly sets April ’11- March ’12 for all
- Fewer cases / month (20 in each)
- Results against PEER opinion and ‘system’
- Results and (most) images will be given immediately following month end
- Launch date for this: 1st April 2011

Agreement against peers

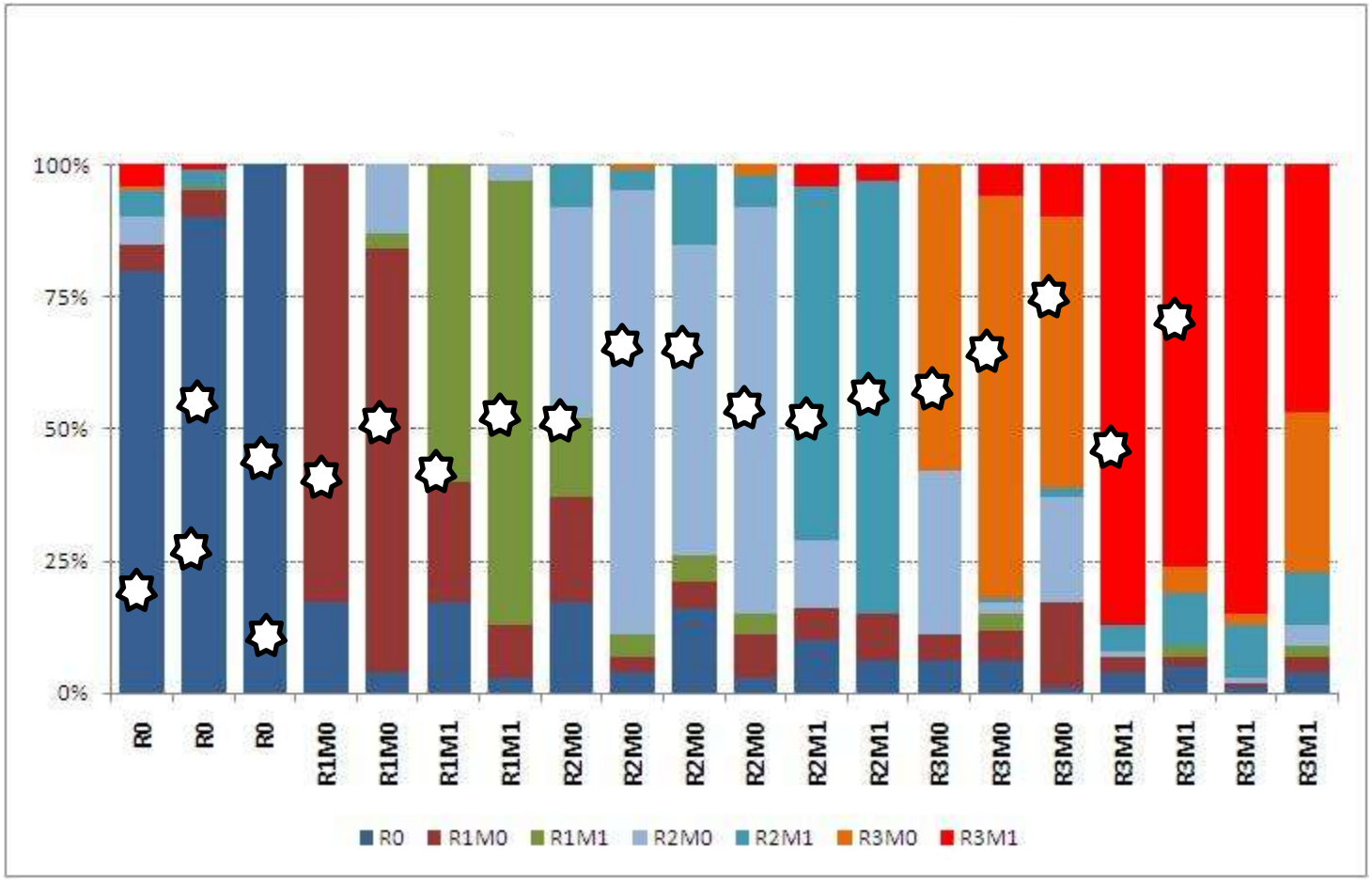


Agreement against peers

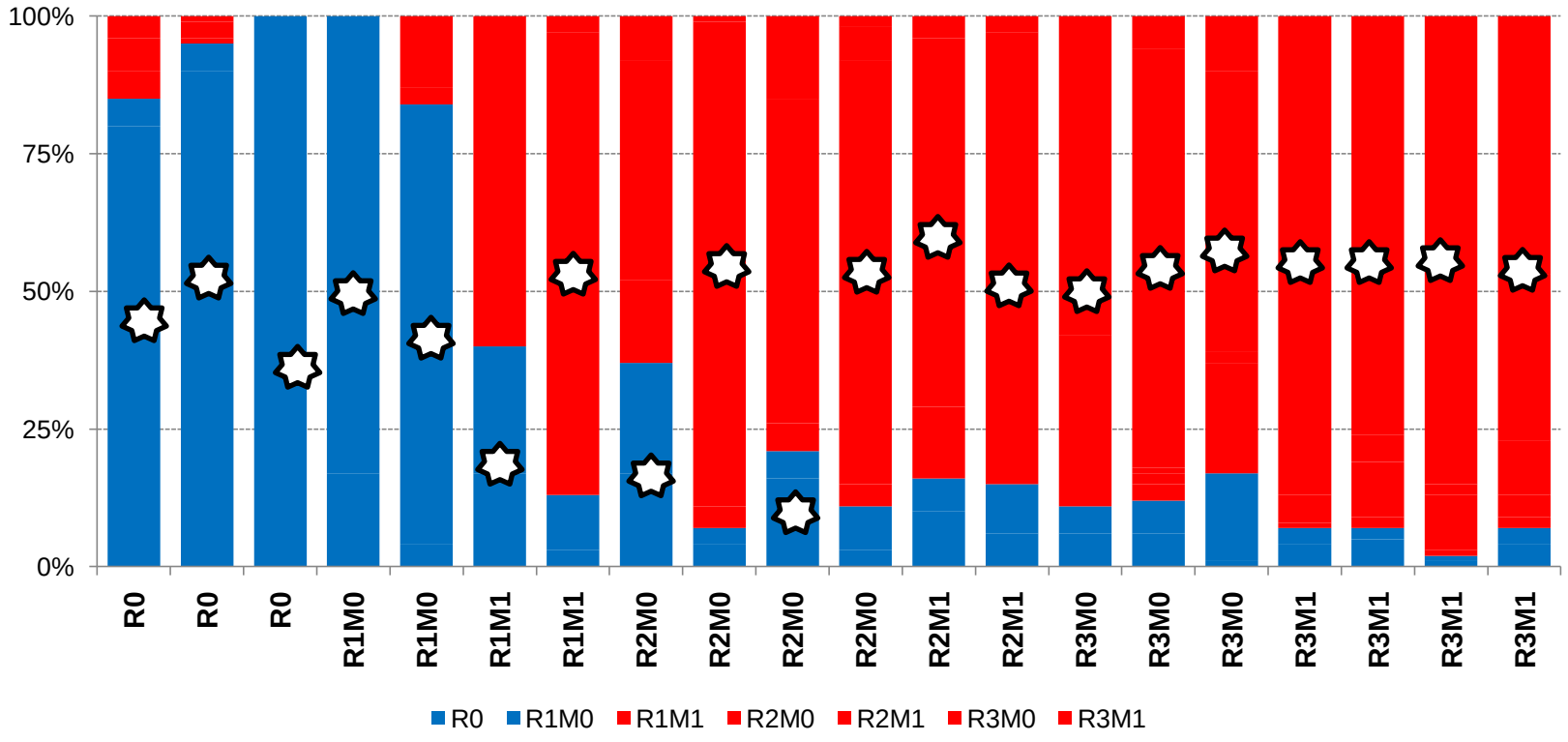


Grader 23 - X

Agreement against peers- April 2011



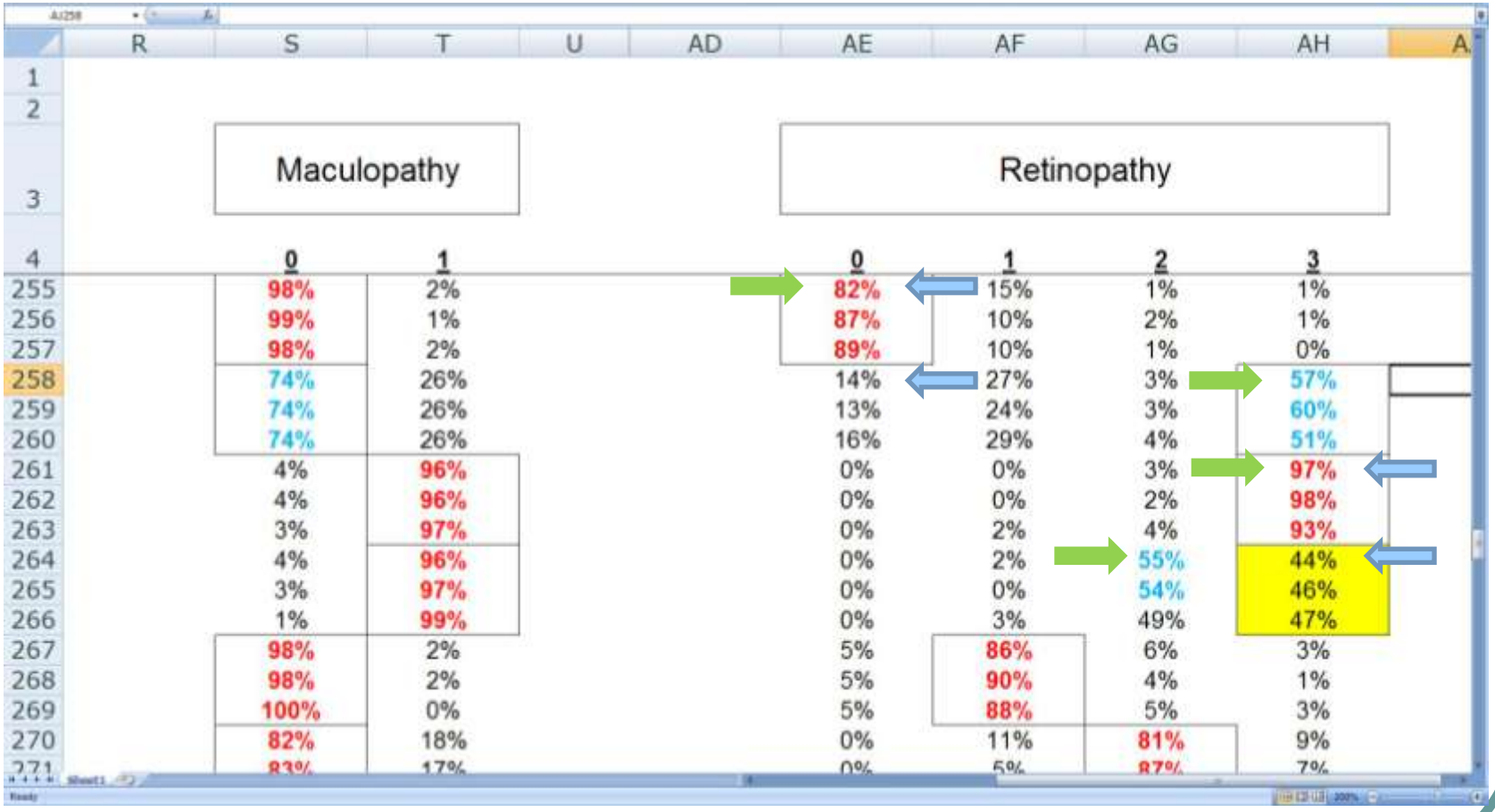
Agreement against peers- April 2011



Question for the audience

- Would you still want to know what was the identified ‘system’ grade for each case?
- How do we combine grades from 20 cases into a single score, if not using agreement against system grade?

Performance of graders 23 and 24



	Maculopathy		Retinopathy			
	0	1	0	1	2	3
255	98%	2%	82%	15%	1%	1%
256	99%	1%	87%	10%	2%	1%
257	98%	2%	89%	10%	1%	0%
258	74%	26%	14%	27%	3%	57%
259	74%	26%	13%	24%	3%	60%
260	74%	26%	16%	29%	4%	51%
261	4%	96%	0%	0%	3%	97%
262	4%	96%	0%	0%	2%	98%
263	3%	97%	0%	2%	4%	93%
264	4%	96%	0%	2%	55%	44%
265	3%	97%	0%	0%	54%	46%
266	1%	99%	0%	3%	49%	47%
267	98%	2%	5%	86%	6%	3%
268	98%	2%	5%	90%	4%	1%
269	100%	0%	5%	88%	5%	3%
270	82%	18%	0%	11%	81%	9%

0.82 + 0.57 + 0.97 + 0.55 = 2.91

0.82 + 0.14 + 0.97 + 0.44 = 2.37

Results against peers

Data against peer opinion will be shown and reported

A monthly combined performance figure will be calculated and reported

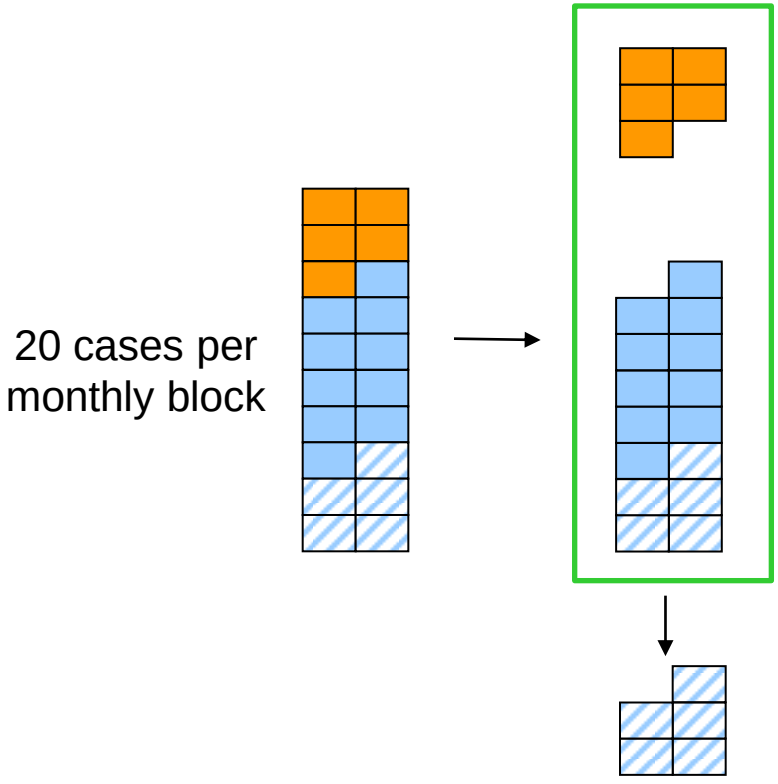
This accounts for agreement cases and difference from majority agreement cases

Results against system

Data against ‘system’ grade will also still be reported. But why?

Because it is of course possible that the ‘majority’ get it wrong!

Monthly block composition



5 cases without image release (for CUSUM comparison)

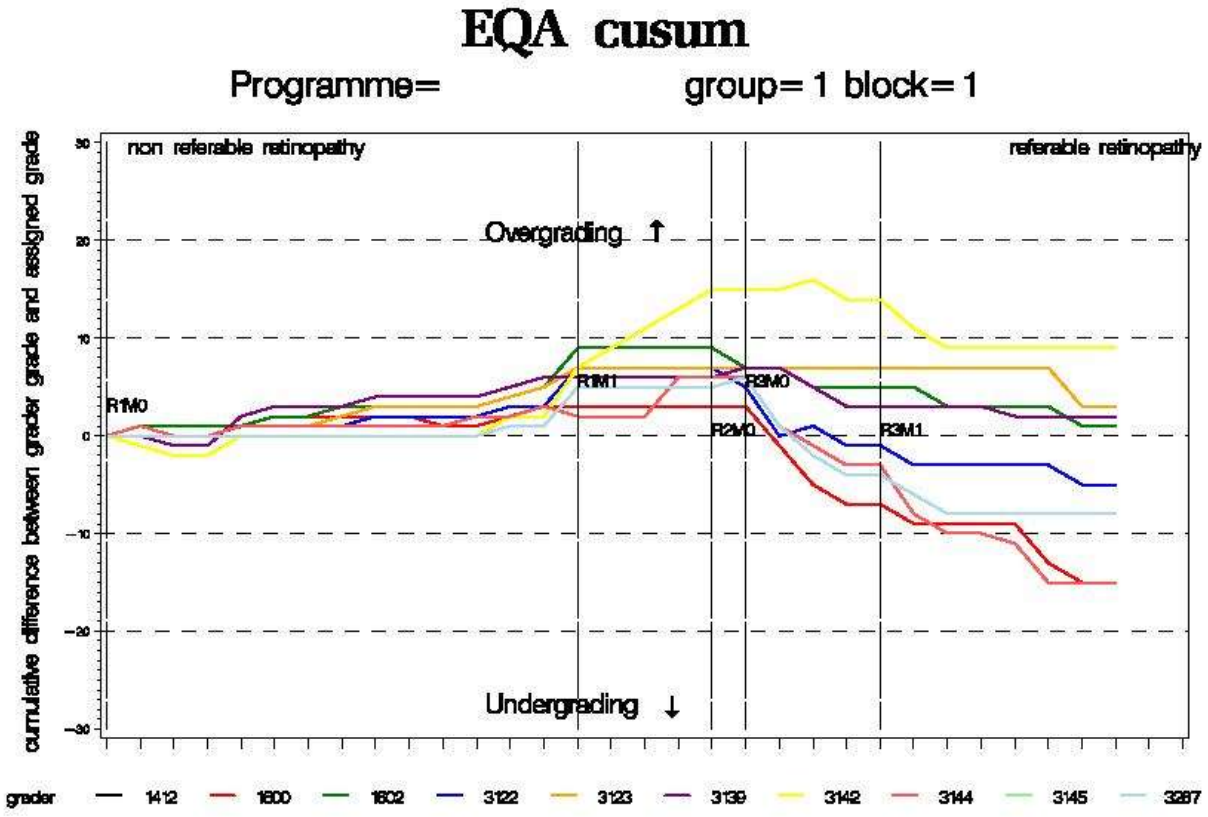
15 cases with full data and image release (training set)

Data against peers from all 20 cases

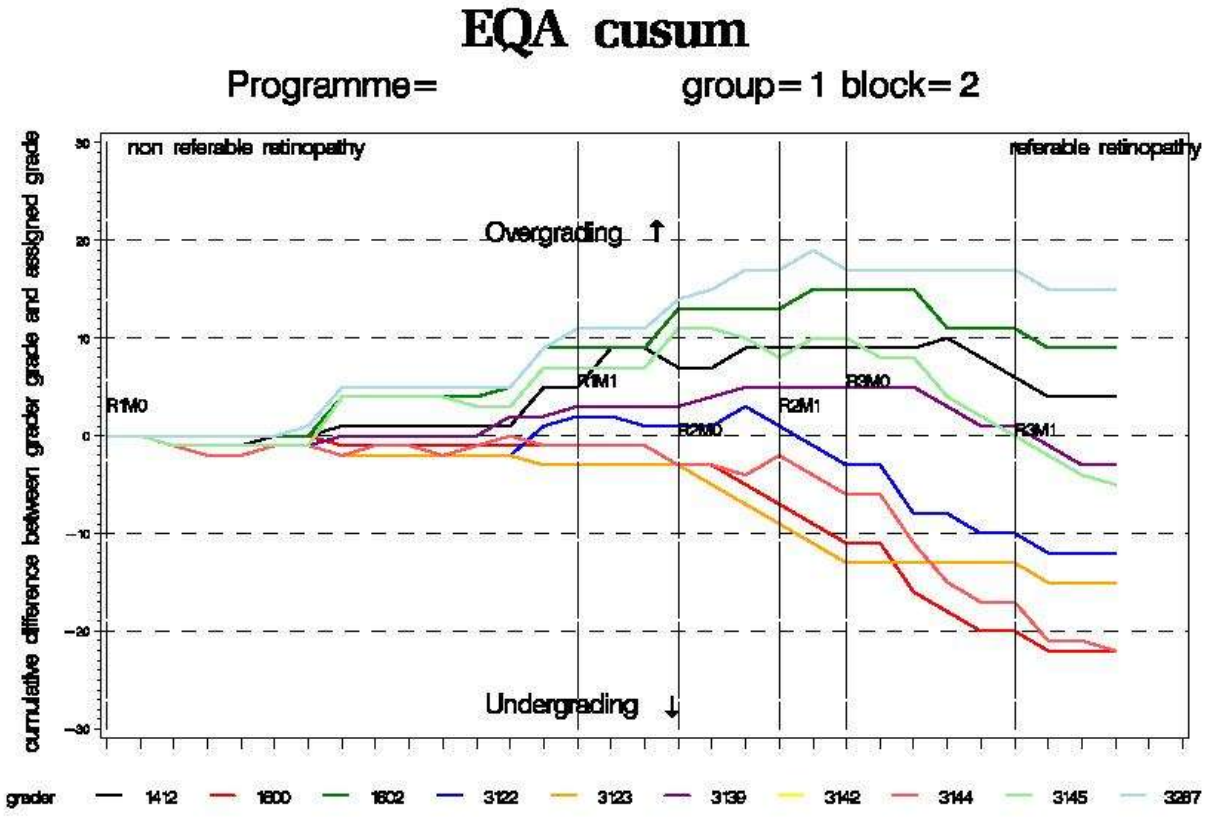
5 cases with annotated data (enhanced training set)

(at end of each month)

Cusum – block 1



Cusum – block 2



Further development - annotation

New - Lesion Annotation Tool

Lesion Annotation Tool

- Provided to support additional T.A.T.
- Is not mandated in any way
- Helps identify exactly what graders see
- Produces permanent DICOM images and auto-comparable ASCII text data
- Grader annotations can be compared against marked-up images from the system, local peers, HES, ?nationally

The future for OL-EQA

- Concept shift from EQA to ‘Test and Training’
- Fewer cases / block (20 in each)
- Fixed monthly blocks April-March for all
- Primary results against PEERS not system
- Images and results will be (mostly) visible – and at each month-end
- Lesion Annotation Tool available
- More improvements to reports incl. ANOVA
- Launch date 1st April 2011
- Pilot sites needed for Jan – March 2011!!

Thanks for your time.

Any comments or questions?