University Hospitals Bristol and Weston NHS Foundation Trust

DEVELOPMENT OF A VISUAL GRADING CHARACTERISTICS (VGC) ANALYSIS TOOL

Niamh Banks
Trainee Clinical Scientist

Anne Hill
Principal Clinical Scientist

University Hospitals Bristol and Weston NHS Foundation Trust

Contents

- What we have done for previous image quality investigations
- Why we wish to move towards a new method
- What is VGC and why have we chosen this method
- How we have developed the VGC methodology into a usable tool
- Results of the VGC tool with previously acquired CT image quality data

Previous method of image quality analysis

- The Physics team at UHBW have worked with reporting staff to score clinical images
- Images were given an image quality rating of Poor – Excellent
- The mean score for each set-up was used as a comparative image quality measure

1	Poor / non-acceptable
2	Acceptable but not fully diagnostic
3	Suboptimal
4	Adequate
5	Excellent

What are the flaws in this approach?

- The data may not be normally distributed: the mean value may not be the most representative value
- Taking a mean value of an ordinal dataset is not statistically valid

1	Poor / non-acceptable
2	Acceptable but not fully diagnostic
3	Suboptimal
4	Adequate
5	Excellent

Our criteria for a new image quality analysis method

University Hospitals Bristol and Weston NHS Foundation Trust

- Easy to implement and use
  - Widely used
  - Doesn't require additional statistics programs
- Minimises inter-observer variation
  - A clearly defined set of criteria
- Capable of performing statistical analysis
  - Determine statistical significance of results
- Requires a relatively small sample size to minimise reporter workload

What is Visual Grading Characteristics (VGC)?

What is VGC analysis?

- VGC is a comparative image quality analysis method
- Requires clinical images to be scored against a set of clinically relevant criteria
  - Minimises inter-observer variation
  - European Quality Criteria are deemed suitable for use in Visual Grading http://www.drs.dk/guidelines/ct/quality/index.htm
- The statistical analysis is valid for an ordinal scoring scale

Visual grading characteristics (VGC) analysis: a non-parametric rank-invariant statistical method for image quality evaluation, Bath and Mansson, 2006

European Quality Criteria for CT

University Hospitals Bristol and Weston NHS Foundation Trust

LIST OF QUALITY CRITERIA FOR COMPUTED TOMOGRAPHY

Pa

CRANIUM

- Brain, General
- Skull Base

FACE AND NECK

- Face and Sinuses
- Petrous Bone
- Orbits
- Sella and Hypophysis
- Salivary Glands (Parotid and Submandibular)
- Pharynx
- Larynx

SPINE

- Vertebral and Paravertebral Structures
- Lumbar Spine, Discal Herniation
- Spinal Cord

1. DIAGNOSTIC REQUIREMENTS

Image criteria:

1.1 Visualization of

1.1.1 Whole cerebrum

1.1.2 Whole cerebellum

1.1.3 Whole skull base

1.1.4 Vessels after intravenous contrast media

1.2 Critical reproduction

1.2.1 Visually sharp reproduction of the border between white and grey matter

1.2.2 Visually sharp reproduction of the basal ganglia

1.2.3 Visually sharp reproduction of the ventricular system

1.2.4 Visually sharp reproduction of the cerebrospinal fluid space around the mesencephalon

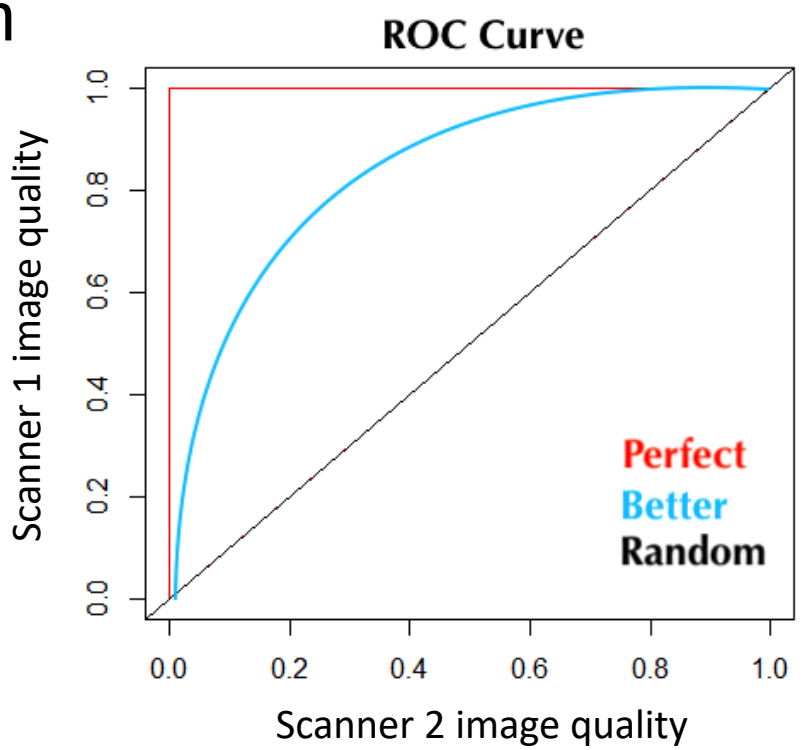
1.2.5 Visually sharp reproduction of the cerebrospinal fluid space over the brain

1.2.6 Visually sharp reproduction of the great vessels and the choroid plexuses after intravenous contrast media

http://www.drs.dk/guidelines/ct/quality/index.htm

VGC curve

- VGC analysis plots a curve similar to a ROC curve to directly compare two system
- Area under curve (AUC) determines which set-up is superior
- AUC 95% confidence intervals determine if the result is statistically significant
  - Do the CI's of the AUC cross the 0.5 indicator of equal performance?



How have we developed the VGC method into a usable tool?

VGC scoring sheet

University Hospitals Bristol and Weston NHS Foundation Trust

- The VGC method has been set-up in an Excel spreadsheet
- The scoring sheet is for reporters to enter image quality scores
- Clinical criteria will be set in conjunction with clinical staff

	Setup	Scoring Criteria		
Reporter Name	Scanner	Visualisation of whole of cerebrum	Visualisation of vessels after IV contrast	Visually sharp reproduction of border between white and grey matter
Reporter 1	Scanner 1	Adequate	Suboptimal	Adequate
Reporter 1	Scanner 1	Excellent	Adequate	Excellent
Reporter 1	Scanner 1	Adequate	Adequate	Suboptimal
Reporter 1	Scanner 2	Excellent	Excellent	Excellent
Reporter 1	Scanner 2	Excellent	Adequate	Adequate
Reporter 1	Scanner 2	Excellent	Adequate	Suboptimal

VGC statistics sheet

University Hospitals Bristol and Weston NHS Foundation Trust

- Minimum sample size of 50 scores per set-up
  - This can be divided between several reporters

Sample size calculator		
# images per setup:	10	*e.g. 10 per scanner*
# scorers per image:	3	
# answers per setup:	30	*small sample size (<50)

- Can select either a single criterion or can perform analysis across all criteria

Results of VGC analysis on previously acquired CT image quality scores

2017 CT image quality investigation

University Hospitals Bristol and Weston NHS Foundation Trust

- In 2017, an image quality investigation was conducted to compare CT performance at UH Bristol
- 4 CT scanners
- 3 radiologists scored 10 abdo-pelvis scans from each scanner
  - 30 scores per scanner: less than the required sample size of 50
- The radiologists scored from 1 – 5 (Poor – Excellent)
  - Criteria were not documented

Image quality scores

University Hospitals Bristol and Weston NHS Foundation Trust

CT Image Quality Scores

NB / AH 16.04.2022

		Setup	
Image number	Reporter Name	Scanner	Image quality score
S1.Im1	Reporter 1	Scanner 1	Adequate
S1.Im1	Reporter 2	Scanner 1	Adequate
S1.Im1	Reporter 3	Scanner 1	Adequate
S1.Im2	Reporter 1	Scanner 1	Excellent
S1.Im2	Reporter 2	Scanner 1	Adequate
S1.Im2	Reporter 3	Scanner 1	Adequate
S1.Im3	Reporter 1	Scanner 1	Adequate
S1.Im3	Reporter 2	Scanner 1	Suboptimal
S1.Im3	Reporter 3	Scanner 1	Suboptimal

Image quality scores

University Hospitals Bristol and Weston NHS Foundation Trust

In the stats sheet, the scores from all images and reporters are collated for each scanner:

		Occurrence of scores				
Setup	Scanner	Excellent	Adequate	Suboptimal	Acceptable (but not fully diagnostic)	Poor
1	Scanner 1	3	12	13	2	0
2	Scanner 2	10	19	1	0	0
3	Scanner 3	8	20	2	0	0
4	Scanner 4	13	15	2	0	0

Image quality scores

University Hospitals Bristol and Weston NHS Foundation Trust

The results are converted into “cumulative results” by dividing by the total number of scores and summing to 1

e.g. for row 1: 50% of the scores are adequate or better

		Occurrence of scores				
Setup	Scanner	Excellent	Adequate	Suboptimal	Acceptable (but not fully diagnostic)	Poor
1	Scanner 1	3	12	13	2	0
2	Scanner 2	10				
3	Scanner 3	8				
4	Scanner 4	13				

Cumulative results				
Excellent	Adequate	Suboptimal	Acceptable (but not fully diagnostic)	Poor
0.1	0.5	0.933333333	1	1
0.333333333	0.966666667	1	1	1
0.266666667	0.933333333	1	1	1
0.433333333	0.933333333	1	1	1

VGC plot: Scanner 1 vs Scanner 2

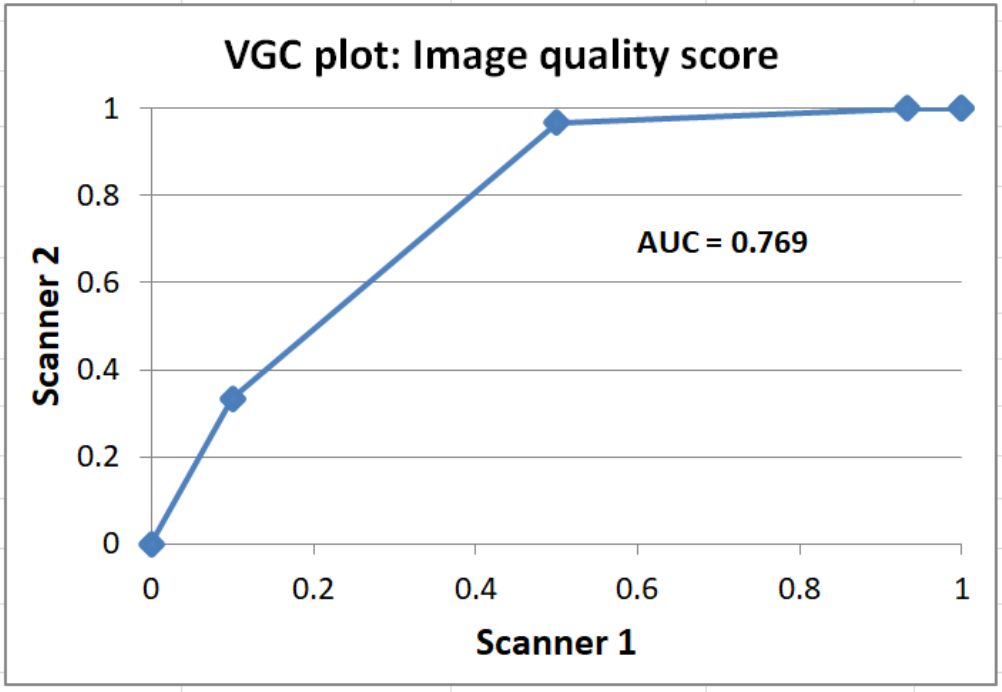
University Hospitals Bristol and Weston NHS Foundation Trust

- Two scanners are compared at a time on the VGC curve
- In this example, Scanner 2 has the better image quality over Scanner 1

Compare		
Setup:	1	*Scanner 1*
with Setup:	2	*Scanner 2*

Data for VGC plot

Setup 1	Setup 2
0	0
0.1	0.33333333
0.5	0.96666667
0.93333	1
1	1
1	1

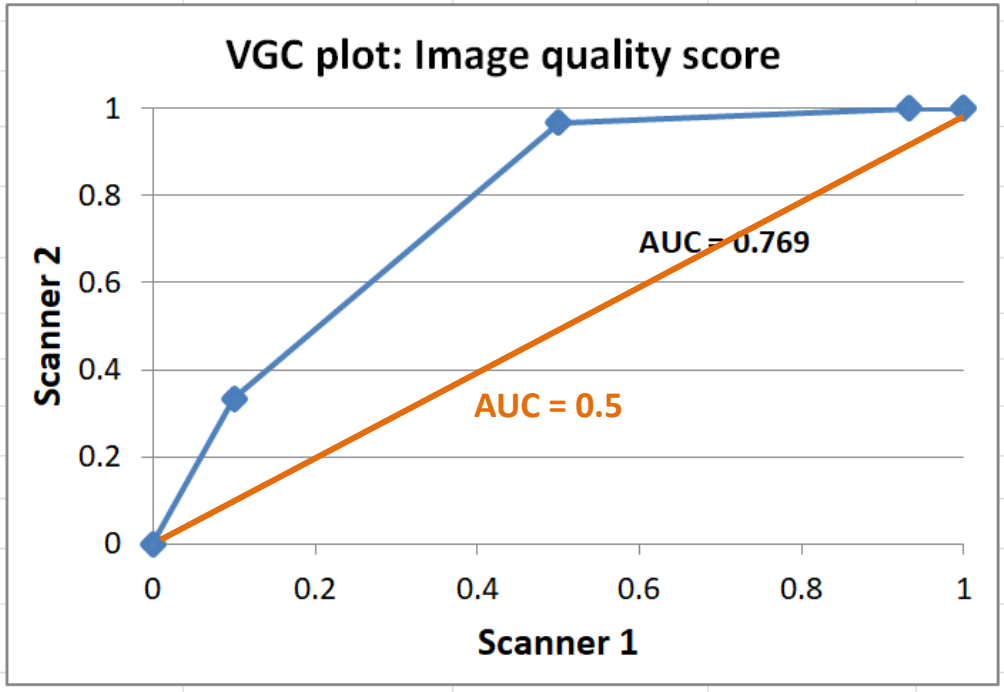


Statistical significance

University Hospitals Bristol and Weston NHS Foundation Trust

- Image quality comparison is considered significant if 95% confidence bounds of AUC does not include 0.5 (representing equal performance)

AUC	**0.769**
95% confidence bounds:	0.65-0.89
Significance:	significant
'Scanner 2' has better image quality	
*small sample size (<50)	



We are supportive respectful innovative collaborative. We are UHBW.

AUC results

University Hospitals Bristol and Weston NHS Foundation Trust

VGC plot (x vs y)	AUC	95% confidence intervals
Scanner 1 vs Scanner 2	0.77	0.65-0.89
Scanner 1 vs Scanner 3	0.74	0.61-0.87
Scanner 1 vs Scanner 4	0.78	0.66-0.90
Scanner 2 vs Scanner 3	0.46	0.31-0.60
Scanner 2 vs Scanner 4	0.54	0.39-0.68
Scanner 3 vs Scanner 4	0.58	0.43-0.72

= significant result (if n > 50)

Mean scores

- Scanner 1 has a much lower mean score
- The standard deviations of the mean scores overlap - the results do not appear significantly different

Scanner	Mean score	Standard deviation
Scanner 1	3.53	0.76
Scanner 2	4.30	0.53
Scanner 3	4.20	0.54
Scanner 4	4.37	0.60

Conclusions

- Setting up the VGC method in Excel was found to be an easy and practical implementation
- VGC analysis is capable of identifying significant differences in image quality performance
  - Only if sample size is > 50
- Clinical staff engaged with the tool when we presented it at a multi-departmental meeting
- There are plans to use the tool for an investigation into Safire iterative settings for CAP scans

Thank you for listening!
Any questions?

niamh.banks@uhbw.nhs.uk anne.hill@uhbw.nhs.uk

References

European Quality Criteria for CT: http://www.drs.dk/guidelines/ct/quality/index.htm

Visual grading characteristics (VGC) analysis: a non-parametric rank-invariant statistical method for image quality evaluation, Bath and Mansson, 2006