Inter-observer Agreement of FDG PET/CT visual Herder Score for the assessment of solitary pulmonary nodules (SPN)

K Ordidge, N Gandy, M Arshad, N Soneji, K Wallitt, S Khan, T Barwick

Department of Nuclear Medicine, Charing Cross Hospital, Imperial College Healthcare NHS Trust, London, UK
Background: Herder Visual Score for Solitary Pulmonary Nodules

- British Thoracic Society (BTS) guidelines recommend FDG PET CT for pulmonary nodule follow up in high risk individuals or if the Brock score is >10% [1]
- Herder score was originally proposed by Herder et al. in 2005 in a retrospective study of 106 patients with indeterminate solitary pulmonary nodules evaluated by FDG PET CT [2]
- Uptake in the nodule is assessed visually, in comparison to the surrounding lung tissue and mediastinal blood pool

References:
BTS Pulmonary Nodule Risk Prediction Calculator

- Herder score from the FDG PET CT is used to calculate the probability of malignancy, also taking into account other patient demographics - age, smoking history etc.
- Herder model probability is used to guide patient management
• Differences in image interpretation using the visual scale will affect the Herder model risk prediction calculation
• This could potentially lead to changes in patient management

Reference:
Aim- Assessing Inter-observer Agreement

- Herder score has been reported as having high diagnostic validity
- Visual scale and needs to be reproducible between observers
- No studies to date looking at inter-observer agreement in using the Herder score
- High inter-observer agreement has been reported using the Deauville criteria in lymphoma [1,2,3]

References:
Methods

- Single centre retrospective study with local ethical approval
- Imperial FDG PET CT database was searched under IC10 code IC34 JF (lung, characterisation)
- Identified 100 consecutive patients having FDG PET CT for the evaluation of a solitary pulmonary nodule
- Exclusion criteria: documented malignancy in the past 5 years
- Anonymised images were reviewed separately by 3 consultant Nuclear Medicine Radiologists using a dedicated Hermes workstation
- Herder score was recorded, along with a confidence score graded 1-3
- Inter-observer agreement was assessed using Interclass Correlation Coefficient modelling
Results - inter-observer agreement

- Complete reviewer agreement in **81% of cases**
- Interclass correlation with Cronbach’s Alpha was excellent at **0.963** (95% CI 0.932-0.982) for the first 29 cases reviewed
- The agreement between pairs of reviewers was good (Kappa scores for reviewer 1v2= 0.722, 1v3= 0.807, 2v3 =0.625)

- For the remainder of the cases Cronbach’s Alpha was **0.977** (95% CI 0.966-0.985)
- Kappa scores for reviewers 1v2= 0.782, 1v3= 0.901, 2v3= 0.801
- This suggests that with increasing case numbers, the inter-observer agreement improved

- Overall, for the total cases assessed, Cronbach’s Alpha was **0.973**
Results- 19 discrepant cases

- 8 patients Herder risk calculations meant unchanged management plan
- Clinically significant difference in risk resulting in different management plans in 11 patients- these patients were followed up

- 6 cases absent/faint discrepancy
- 7 cases faint/moderate
- 6 cases moderate/intense
Discrepant case example- faint vs moderate
Follow up

- Patient too frail for lung biopsy
- In practice this case was followed up with interval CT studies at 3/12 and 12/12
- Stable volumetry

- In most of the discrepancy cases, there was additional patient history that decided the management plan in the MDT setting
Conclusion

- Inter-observer agreement is excellent, with 81% complete agreement and Cronbach’s Alpha of 0.973
- Inter-observer agreement improved with increasing number of cases reviewed
- Differences in visual interpretation can lead to clinically significant difference in the calculated risk, changing the suggested management
- In practice, the clinical history is used to guide management, particularly in cases where there is a previous history of cancer
- Equivocal cases were most likely to be followed up with interval CT at 3/12 and 12/12
Further work

- Could inter-observer agreement improve further using a semi-quantitative method to assess lesion SUVmax and mediastinal blood pool SUVmax?
- Multiple studies using FDG PET & FDG PET CT suggest that semi-quantitative methods do not increase accuracy of diagnosis [1,2], even for lesions with low SUVmax [3]
- ultraHD, Q.Clear- potential that things which were absent become faint on newer reconstruction methods- affect on inter-observer agreement?

References:
Thank you

Supervisor: Dr Tara Barwick

Consultant reviewers: Dr Neil Soneji, Dr Sam Khan, Dr Kathryn Wallitt

Colleagues from Nuclear Medicine Department at Charing Cross Hospital