# Use of computer aided detection to support triage for efficiency at the MBOD

Annalee Yassi, Jerry Spiegel, Steven Barker and colleagues from
School of Population and Public Health, University of British Columbia (UBC),
and Rodney Ehrlich, University of Cape Town (UCT)
in collaboration with U Wits, MBOD/CCOD and extended team

Main funding: UK Department for International Development

# CAD - general methodological considerations

- Developed as proprietary radiological software for TB by different commercial enterprises.

- Sources of bias and variation
  - Machine vs deep learning
  - How trained - How were CXRs selected, how many?
  - How validated – Source of CXRs ? Human or laboratory reference( for TB)

- *Innovation* in this project was application to a unique population:
  - High prevalence of silicosis
  - High prevalence of prior TB.
  - First studies to include silicosis as an outcome
  - First studies in a compensation adjudication setting.

# Two validation studies of CAD

- Young et al. 2020 - MBOD certification as the reference.

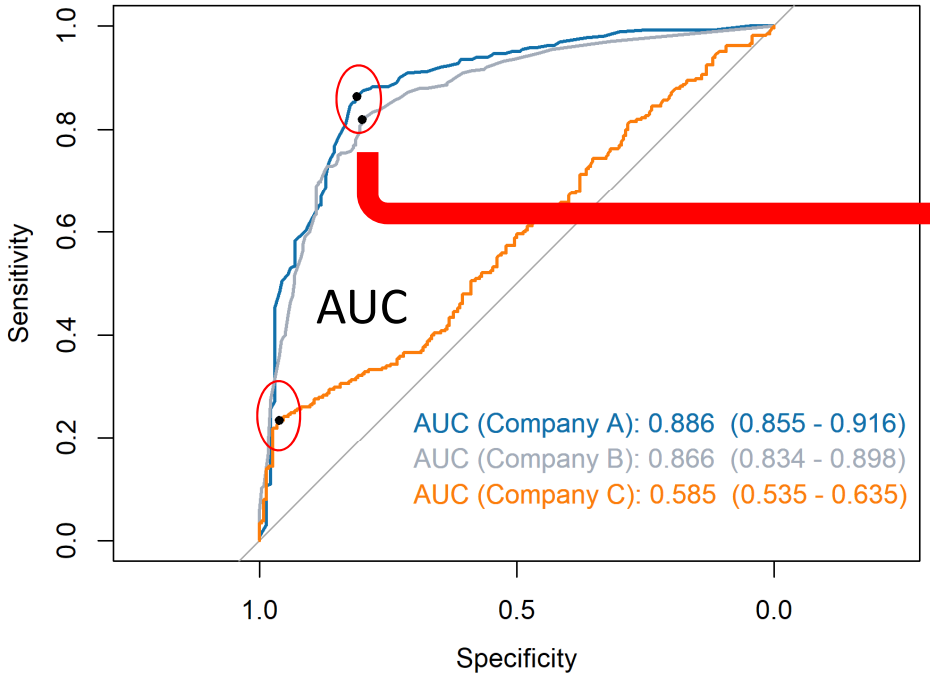
- Field study (unpublished) – Alice, Bizana, Stilfontein BMEs; 3 external readers (OM specialists) as reference.
- Three CAD systems (A, B and C), with outputs:
  - TB (0-100)
  - Silicosis (0-100)
  - "No TB or silicosis"(0-100 ) (Problematic category – not same as "normal")
- Between the 1st and 2nd studies, B and C "dropped" silicosis. Only system A directed at silicosis in 2nd study. (Also only system trained on *active* TB).
- Results of both studies used for efficiency project.

## Performance = "accuracy against a reference standard" (relative!).

- Receiver operating curve (ROC) - *plots* agreement.
- Area under the curve (AUC) *quantifies* overall agreement between tests ("readers")
- *Sensitivity* – ability of test system to identify true positives (no false negatives).

  *Specificity* – ability of a test system to identify true negatives (no false positives)
- Each point on the ROC is a unique combination of sensitivity and specificity
  - Sensitivity and specificity are *complementary* – one comes up, other goes down.

# Detecting TB using CAD (field study)[1,2]



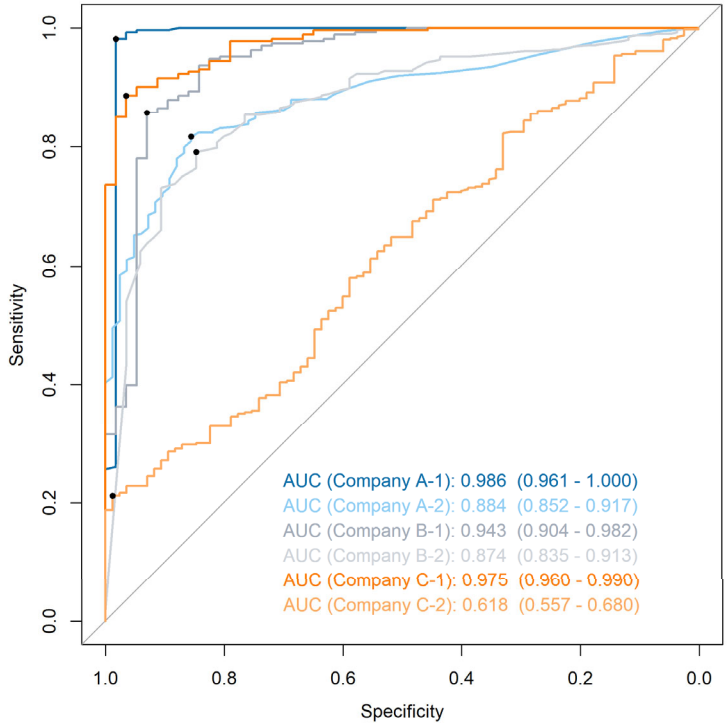

| System | Sensitivity | Specificity |
|---|---|---|
| At Youden-index[3] maximum: | | |
| A | 81.6% | 85.5% |
| B | 79.1% | 84.7% |
| C | 21.2% | 98.8% |
| At fixed minimum 90% sensitivity: | | |
| A | 91.1% | 55.4% |
| B | 90.1% | 58.8% |
| C | 90.1% | 17.6% |

[1] External reader 2 as the reference standard.

[2] A and B had reasonably good performance when detecting TB image, and not significantly different from one another

[3] Youden index is the point at which the combination of [sensitivity + specificity – 100] (%) is maximised.

# System performance declined between trials – example: detection/ exclusion of "abnormality")[1,2]




| System | Study | Sensitivity | Specificity |
|---|---|---|---|
| **Fixed at minimum 90% sensitivity:** | | | |
| A | 1: (Young 2020) | 90.1 % | 98.2% |
| A | 2: (Field study) | 91.1% | 55.4% |
| B | 1: (Young 2020) | 90.1% | 84.2% |
| B | 2: (Field study) | 90.1% | 58.8% |
| C | 1: (Young 2020) | 90.10% | 94.70% |
| C | 2: (Field study) | 90.10% | 17.60% |

[1] External reader 2 as reference standard in field study. MBOD classification reference in Young et al.

[2] Statistically significant performance decline across all systems.

# Efficiency gains - different years of service thresholds vs CAD[1]

| | Referral rule or threshold | No. of expert reviews | Efficiency gain [2] |
|---|---|---|---|
| 1 | No triage (reference category): All cases go the full panel | 47,275 | - |
| 2 | Hypothetical maximum: all TB-flagged claims are sent to the TB panel, all normal claims (non-compensable without disease) are sent to the reduced panel, and all other claims are sent to the full panel. | 29,144 | 38.4% |
| 3 | 10-year service threshold, no CAD (actual triage, not simulated) | 37,657 | 20.3% |
| 4 | 15-year threshold, no CAD | 36,747 | 22.3% |
| 5 | 20-year threshold, no CAD | 35,865 | 24.1% |
| 6 | No CAD, no triage based on years of service: **All TB cases routed to TB panel, all others to reduced panel.** | 34,015 | 28.0% |
| 7 | CAD triage: sensitivity 98.2%, specificity 98.2% (Youden)[3] | 30,231 | 36.1% |
| 8 | CAD triage: sensitivity 100%, specificity 86.0% (Sensitivity fixed) | 31,504 | 33.4% |
| 9 | CAD triage: sensitivity 90.4% and specificity 60.6% (Youden)[4] | 34,702 | 26.6% |
| 10 | CAD triage: sensitivity 70.5% and specificity 90.4% (Specificity fixed) | 32,214 | 31.9% |

[1] All TB cases to TB panel, all deceased cases to full panel. [2] CAD efficiency gains are an *alternative* to the routing options presented separately. The two are not additive. [3] Young et al. [4] Field trial, system B vs reader 2

# Decisions facing users re CAD

- Choice of CAD system. Need to understand criteria include performance metrics and potential bias – on whom trained, against whom/what validated.
  - In the field trial, the agreement between CAD and external readers for "any abnormality" for "TB" met or were close to WHO screening tests guidelines (90/60) for systems A and B.
  - C performed very poorly.
  - Only A reads for silicosis

- Choice of optimisation point - trade-off of false negatives against false positives [cost and time efficiency as well as "risk" (ethical, legal, reputational)]. E.g. 100% efficiency, with substantial false positives. Or lower sensitivity, leaving it to reduced panel to pick up false negatives?

# Other considerations re CAD

- Manage CAD- medical adjudicator interface.  How will it be used in day to day practice? Triage or backup?
- Need continuous monitoring of quality (by radiologist and MOs) and consider incremental improvements over time,  i.e. not a once off decision.
- Manage relationship with vendors
  - Facilitate further machine learning by providing images
  - Manage IP issues
- Are there legal issues in use of CAD?

# Recommendations re pre-panel improvements in efficiency

- Educate referring physicians
  - "TB wage loss only" to be clearly marked on the claim documentation.
  - Clinical skills in diagnosing silicosis and TB (prior or active) in miners and ex-miners
  - Understanding of ODMWA, including possible discretion in not submitting the results of medicals where there is evidence of compensable disease. (Needs legal clarification of sections 32 and 33)
- Train MBOD clerical staff

  Ensure that TB claims are accurately routed to TB panel.

# Next steps

1. Decision making by CCOD/MBOD

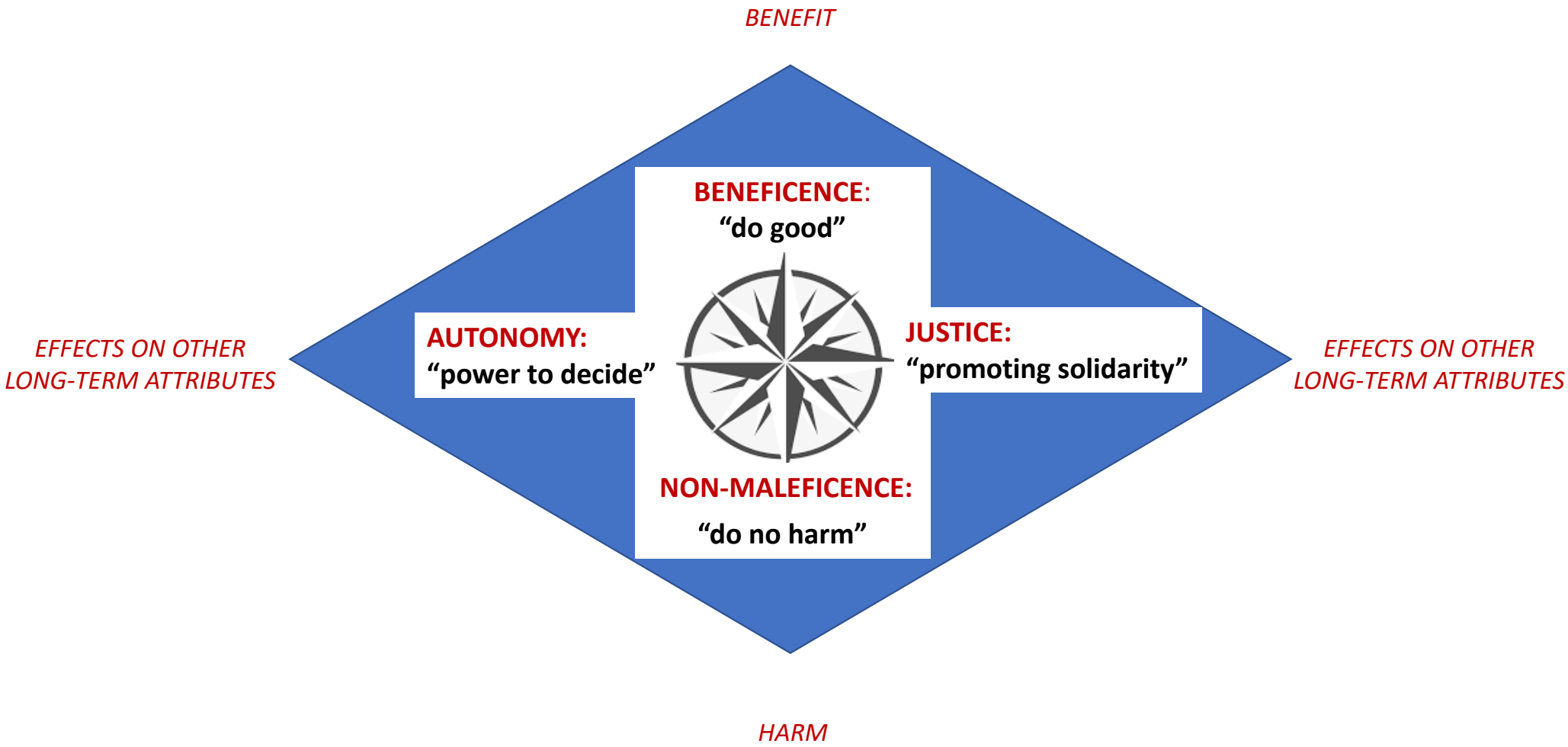
- Primary routing - Whether a change in living claims processing should be implemented, whereby, for example, *all* claims are routed to either the TB panel or the reduced panel initially, *with only those determined by those two* 2-person panels to needing the full-panel assessment referred upwards.

- CAD – as above

2. Submission of draft publication for peer review. (WHO funded field study article will follow).

3. A cost or cost-benefit analysis of the financial savings in adopting various options.

# Principles of Bio-ethics



# Reflecting on principles of Bio-ethics in relation to our CAD application

Will anyone eligible for benefits be excluded?
Minimisation of "false negatives", i.e. incorrect rejection of eligible claims

*Choice of sensitivity/specificity combination, monitoring by users of CAD outcomes*

**BENEFICENCE**:
**"do good"**

Is intellectual property creation managed?

Will public sector retain access?

Will experts lose roles, be deskilled?

*Clarity re triage and role of adjudicators*

**AUTONOMY:**
**"power to decide"**


**JUSTICE:**
**"promoting solidarity"**

*A priori* agreement

Is there transparency & accountability of procurement, application and performance of system?

*Regular reporting by management*

**NON-MALEFICENCE:**

**"do no harm"**

Could information be misused?

*Provisions for privacy & security*

Contribute to ongoing training of CAD?

*Share databases of images and readings*