Machine learning augmented diagnostic testing

Chris Banks, Aeron Sanchez, Rowland Kao
(Roslin Institute, University of Edinburgh)
Vicki Stewart, Kate Bowen *(UK Farmcare Ltd.)*
Graham Smith *(UK Animal and Plant Health Agency)*

THE UNIVERSITY *of* EDINBURGH
The Royal (Dick) School of Veterinary Studies

29th August 2024
ModAH Conference, Nantes, France

Bovine Tuberculosis diagnostics

- Bovine Tuberculosis (bTB) is the most economically important livestock disease in GB, costing around £100m annually and having a substantial impact on animal health and welfare.
- Single Intradermal Comparative Cervical Tuberculin (SICCT / "skin") test is used for routine surveillance for bTB in the UK, Ireland, Portugal, and increasingly in France.
- Whilst highly specific, the skin test suffers from only moderate *sensitivity*, meaning many infected cattle may be missed.

©TB Hub

Skin test sensitivity

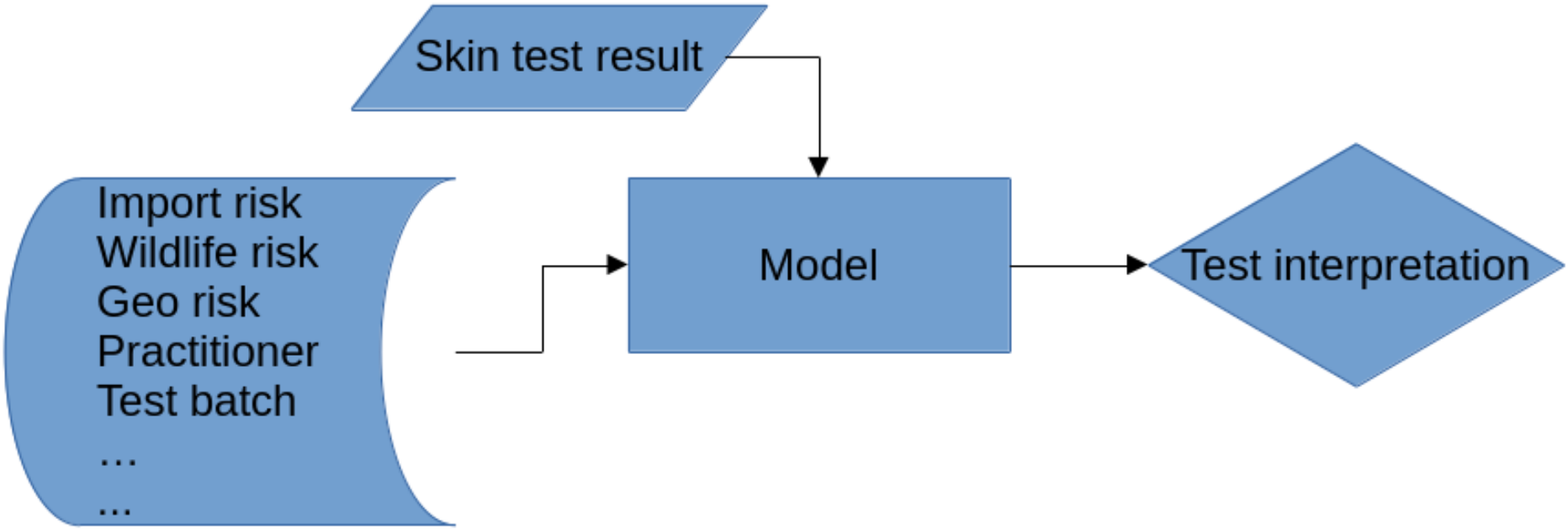
- Test sensitivity depends on many factors, many of which are hard to control.
- These can include:
  - Improperly stored or out of date tuberculin.
  - Too little bovine tuberculin injected in to the skin.
  - Tuberculin not injected in to the skin correctly.
  - Incorrect location of injection sites.
  - Injection of bovine tuberculin in the avian injection site and vice versa.
  - Reading the test too early or too late.
  - Errors in recording skin measurements.
  - Errors in identifying reactor animals.
  - ...

©Tim Scrivener, Farmers Weekly

Skin test sensitivity

- Sensitivity can be improved in “high risk” situations using a “severe” interpretation.
- The measurement cut-off usually favours high specificity to avoid false positives.
- The cut-off is shifted in high risk situations to improve sensitivity.
- High risk situations include where there is an ongoing TB outbreak or on farms where animals have been moved before detecting an outbreak at the source.

Machine learning augmented diagnostics



- Here we take a more holistic view of the "severe" interpretation.
- Multiple potential risk factors integrated into a model for interpreting a skin test result.

bTB diagnostics project

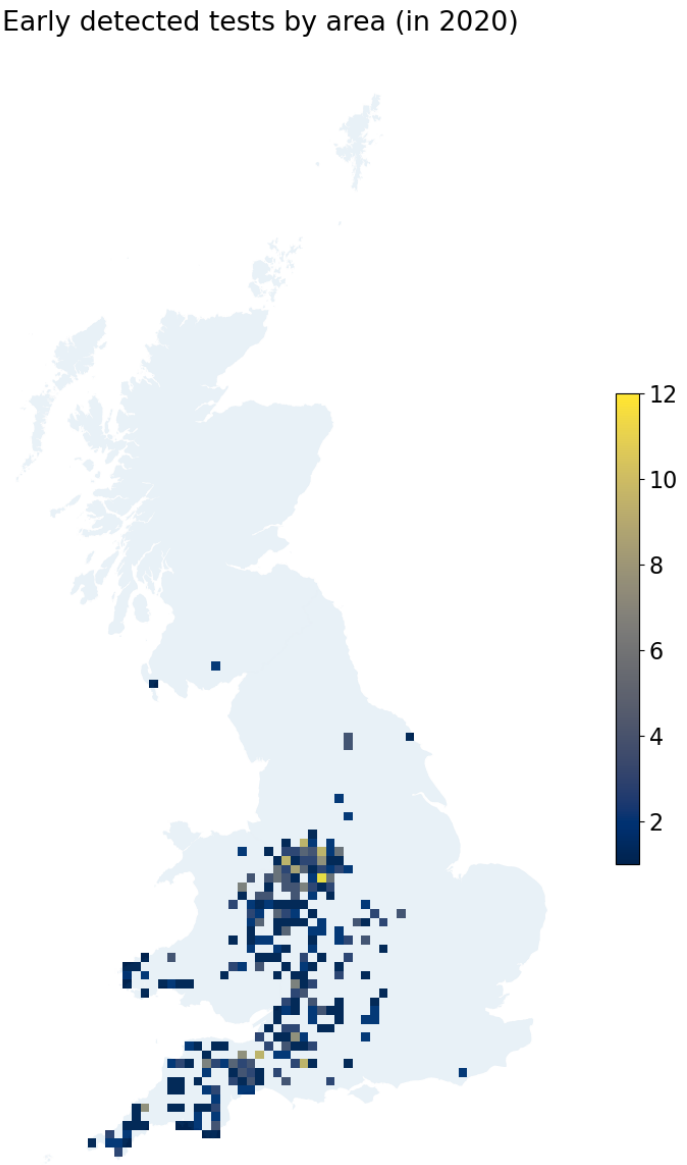
- Goal:
  - Augment the results of the Single Intradermal Comparative Cervical Tuberculin (SICCT) test with surrounding epidemiological risk factors:
    - to improve herd-level test sensitivity,
    - thus improve early detection and reduce onward transmission.
  - Integration of a large number of risk factors, including:
    - Farm characteristics, movements, testing, wildlife abundance, *vet practice*, *tuberculin batch*, ...
- Outcomes:
  - Herd-level prediction of bTB outbreaks using a machine learning model.
  - Risk factor breakdown and analysis using feature importance.
  - Individual-based transmission model to asses onward transmission.

Model and Data

- Data extracted from APHA bTB surveillance database (SAM) and the Cattle Tracing System (CTS) database.
  - Every recorded SICCT test event between January 2012 and September 2021 (1.3m records)
  - Metadata on herd, location, movements, ...
- Data from UKFarmcare on vet and tuberculin batch for some tests:
  - 400 vet practices, covering 120k tests
  - 650 tuberculin batches, covering 57k tests
- Data from APHA on badger abundance.

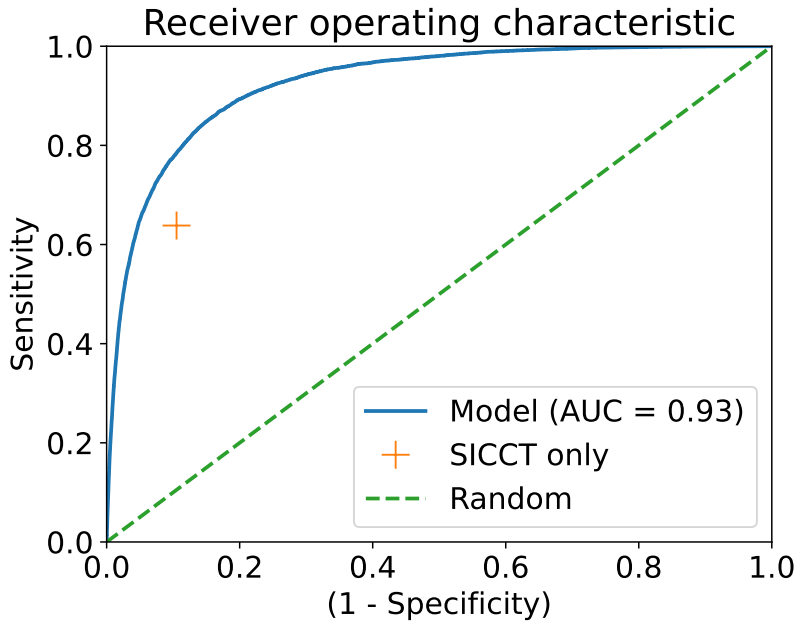
Results – Overview

- Model goal is to predict a new confirmed breakdown within 90 days of testing.
- Balance of sensitivity vs. specificity can be tuned in the model.
- Here we take the same herd-level specificity as the skin test alone, maximising sensitivity.
- **Increase in herd-level sensitivity 14.6%.**
- Over one year (2020) we find **618 negative herd-level tests**, that went on to have breakdowns, but were **identified as positive by the model**.

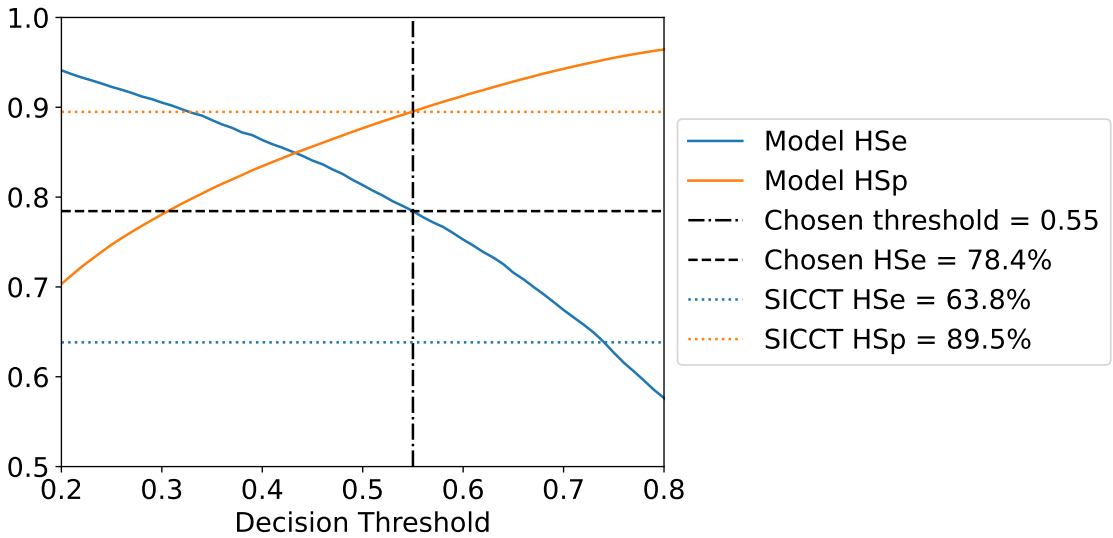


Results – Tuning

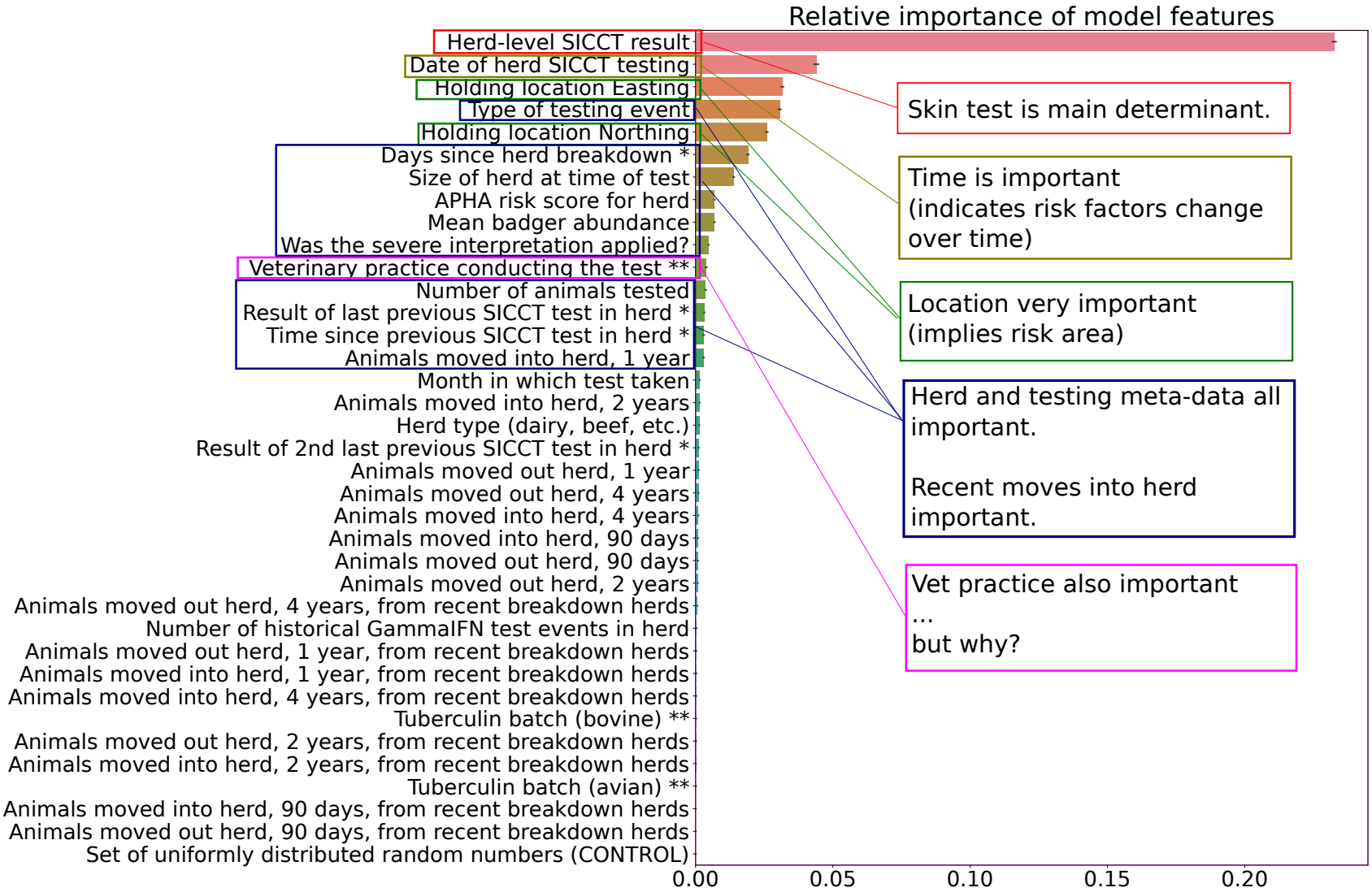
- Model is hyperparameter tuned and 10-fold cross-validated.
- Model herd-level accuracy greater than SICCT at all thresholds.



- Threshold chosen to maximise sensitivity, maintaining SICCT-level specificity.



Analysis – Risk factor importance

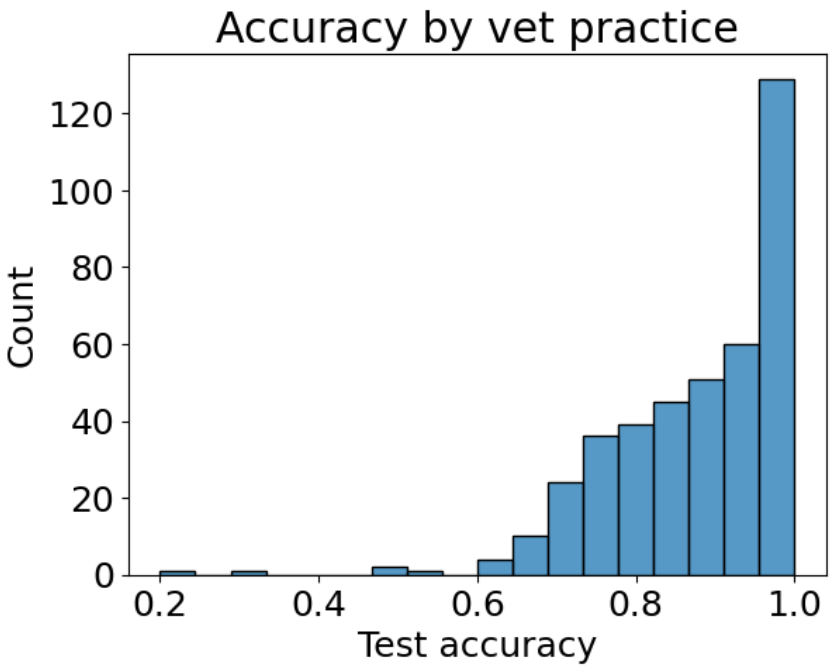


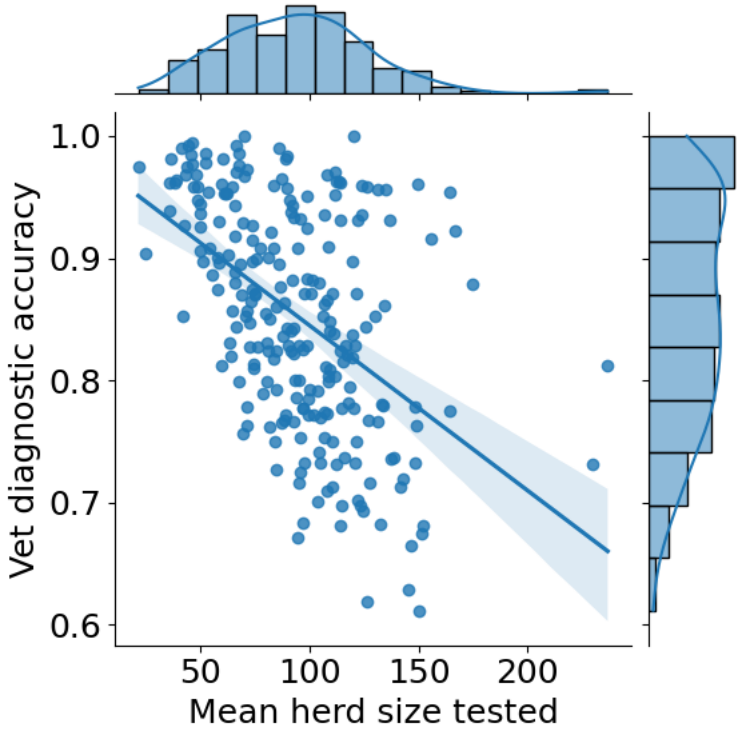
- Risk factors ranked by permutation-based importance testing.
- Bar length shows accuracy reduction when factor randomised.

Vet practice importance

- Can we further explain risk factor importance of vet practice?
- Data coverage may be reducing effectiveness:
  - Practice data covers only 10% of test records.
  - In a model using only records with vet data, practice becomes 3rd most important factor.
  - Random control models of 10% of records show little change in importance, and only a minor drop in overall accuracy.

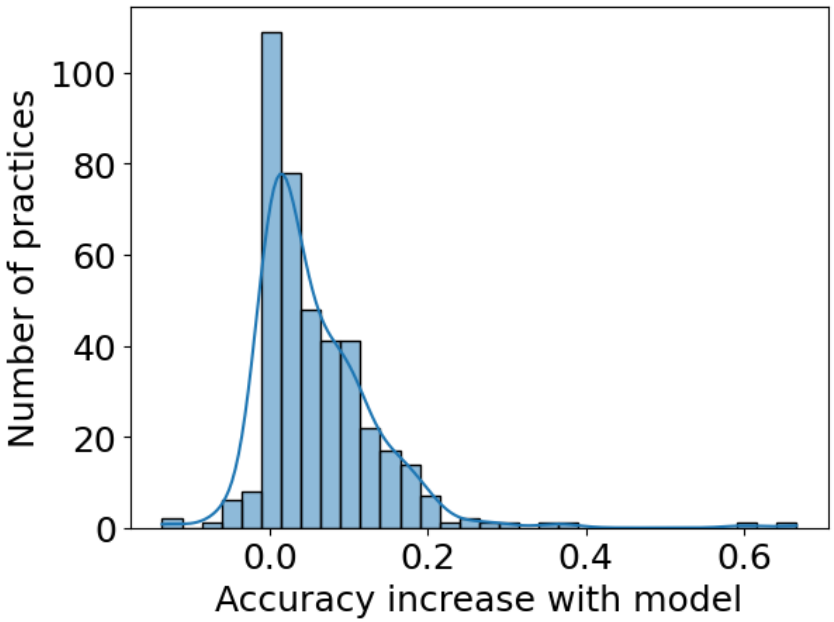
Vet practice importance – what can we say?

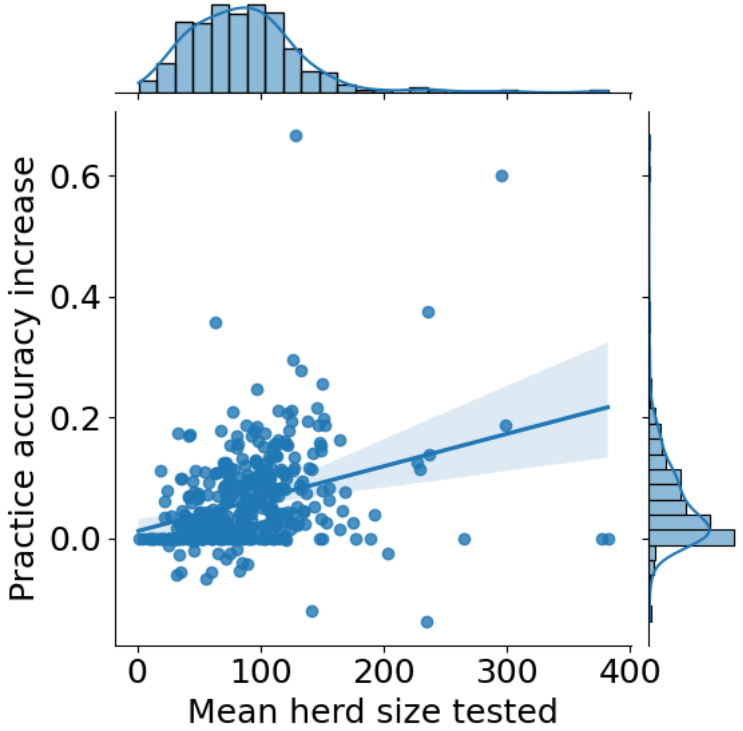




- "Accuracy" is whether a herd-level test result predicts a confirmed breakdown within 90 days.
- Majority of practices have high accuracy, a few are lower.
- **This could be a result of many confounding factors.**
- For example, those that test larger, riskier herds.

Vet practice importance – how does the model help?





- Model increases vet accuracy on the whole.
- Accuracy increases with greater proportion for larger herds, for example.

Conclusions

- The model, trained on extensive historical data, accurately identifies situations where a negative test has sufficient prior probability of an outbreak that the test can be interpreted more severely.
- We show that a significant number of herds could have had earlier identification of an outbreak using this method.
- Using a simulation model testing two areas of the UK with different risk profiles, we also showed the impact of this increase in herd-level sensitivity:
  - an overall decrease in confirmed breakdowns and individual reactors.
- Feature importance identifies risk factors, but some caveats are worth noting:
  - a feature with missing data, like vet practice, will be underpowered in the model (but still useful!)
  - correlated features may also be underpowered (but also still useful!).
- Unpicking features with many potential confounders can be tricky, e.g. vet practice.

Where next?

- Possible extension to use skin test measurements instead of binary results
- How could this be used in the field?
  - App interface to model for vets to use at testing?
  - A guide to re-testing high-risk herds?
- Other situations: BVD, Johne's.

Thank you!