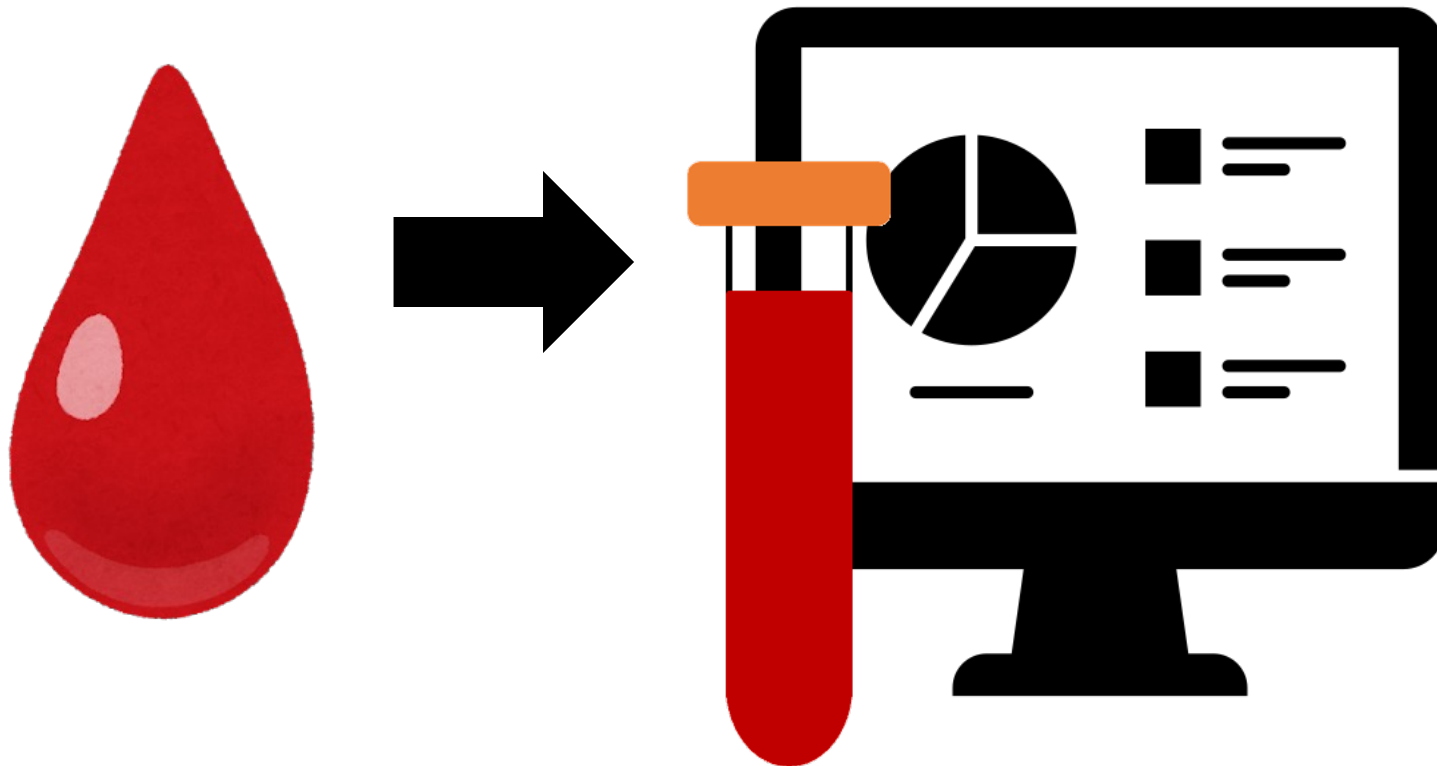
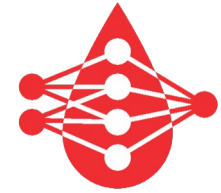


# Investigating Bias in Full Blood Count Data

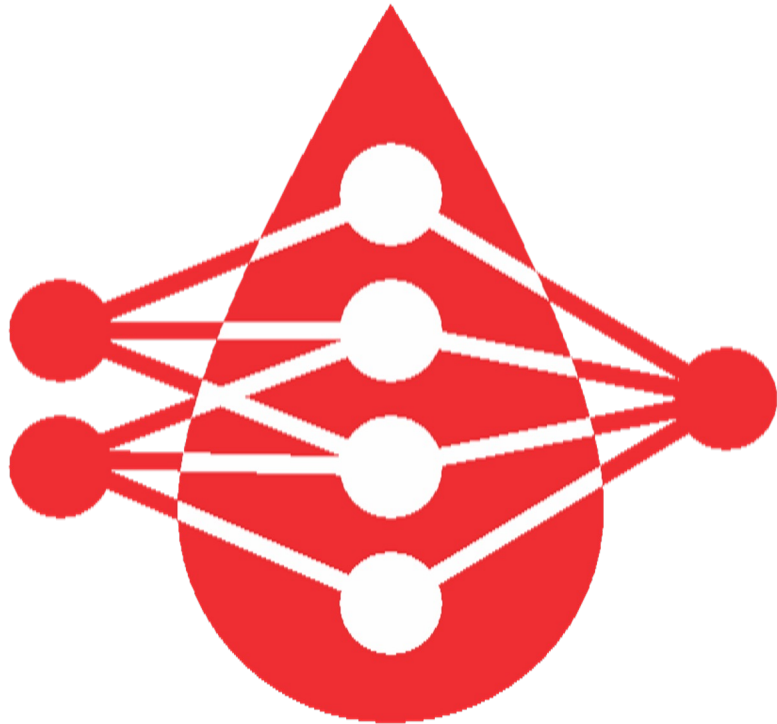
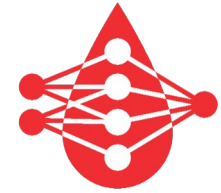
Daniel Kreuter - BloodCounts!

# Full Blood Count

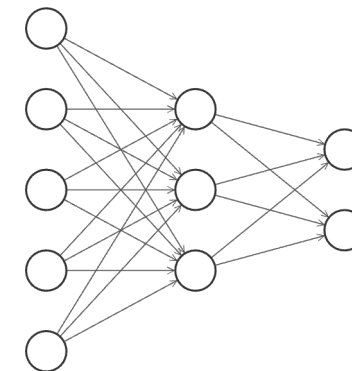
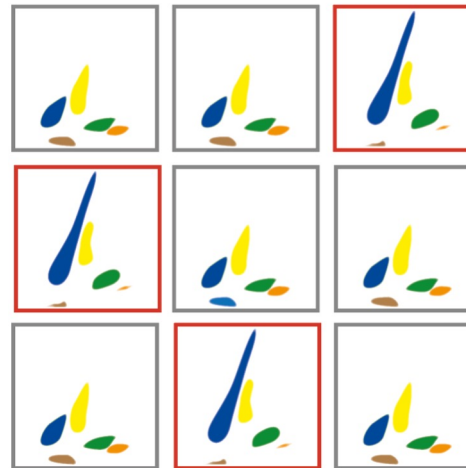


- The **most common** medical laboratory test
- **Essential** for decision making in general practice and hospital medicine
- Around 20 measurements in clinical use  
→ **100+** reported by haem. analyser!

# BloodCounts! Project

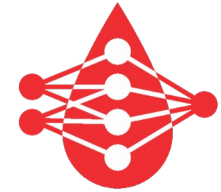


- Using AI to harness rich full blood count data
  - better medical decisions
  - early detection of new disease outbreaks
- Prize winner in the Trinity Challenge

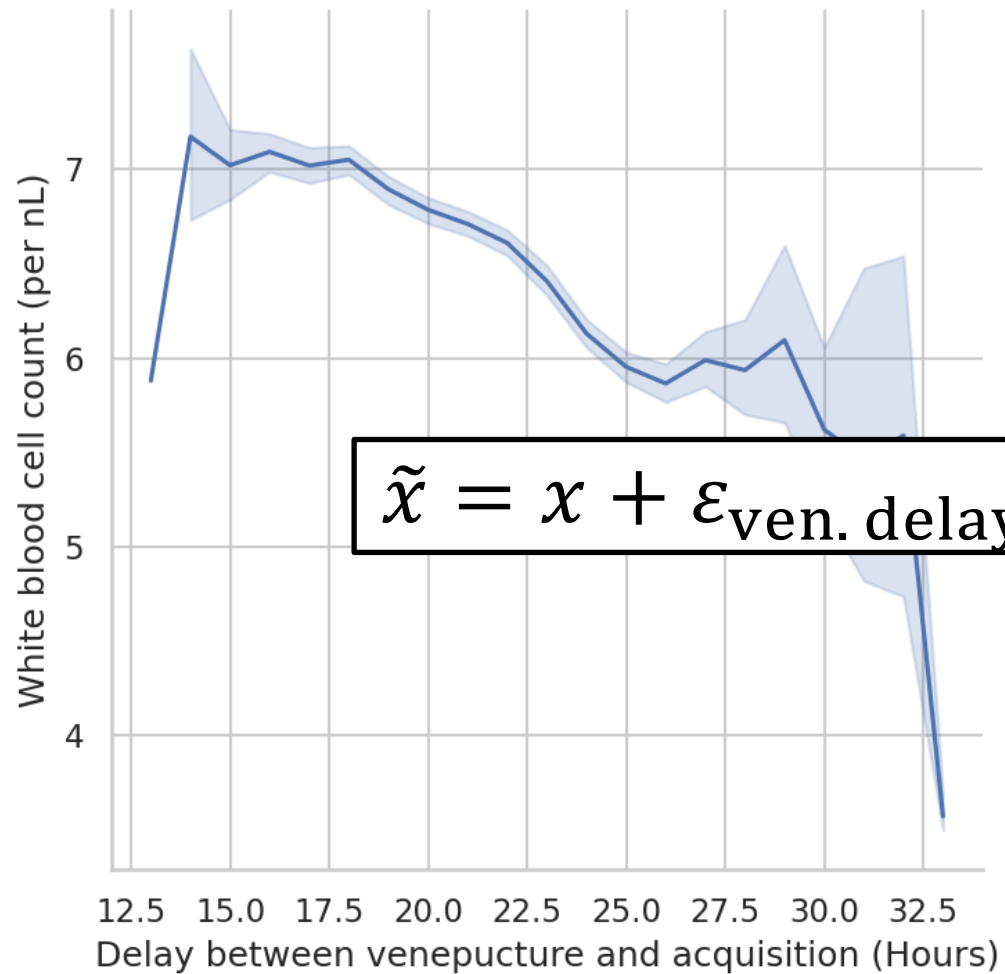


Male	Obese
40-45	Anomaly?
COVID-19 positive	

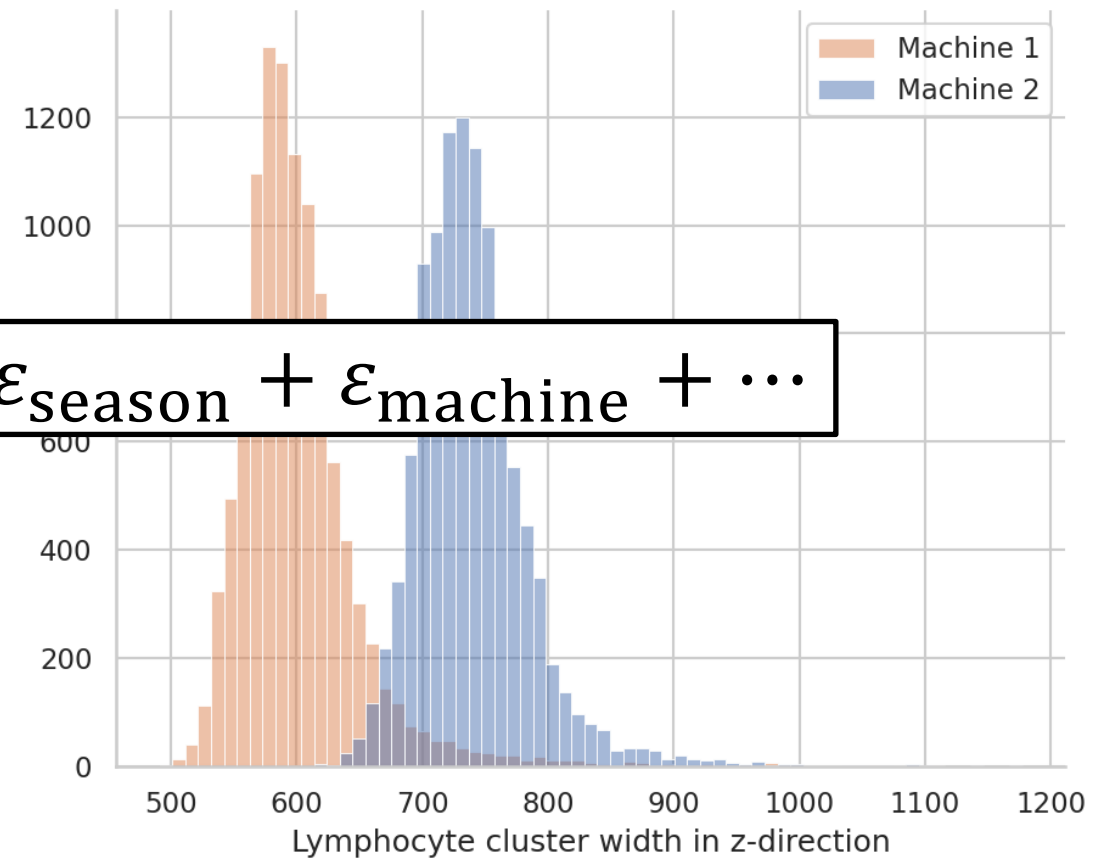
# Full Blood Count Data – Problems?



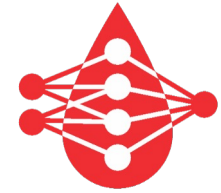
INTERVAL blood donor study



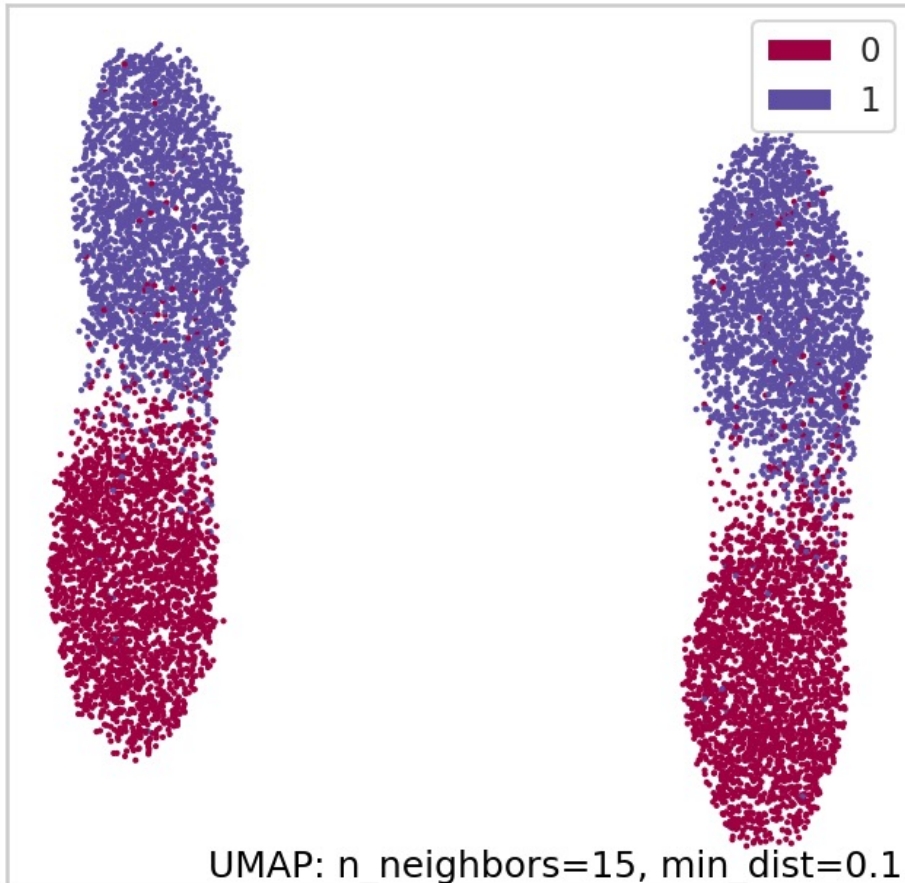
$$\tilde{x} = x + \varepsilon_{\text{ven. delay}} + \varepsilon_{\text{season}} + \varepsilon_{\text{machine}} + \dots$$



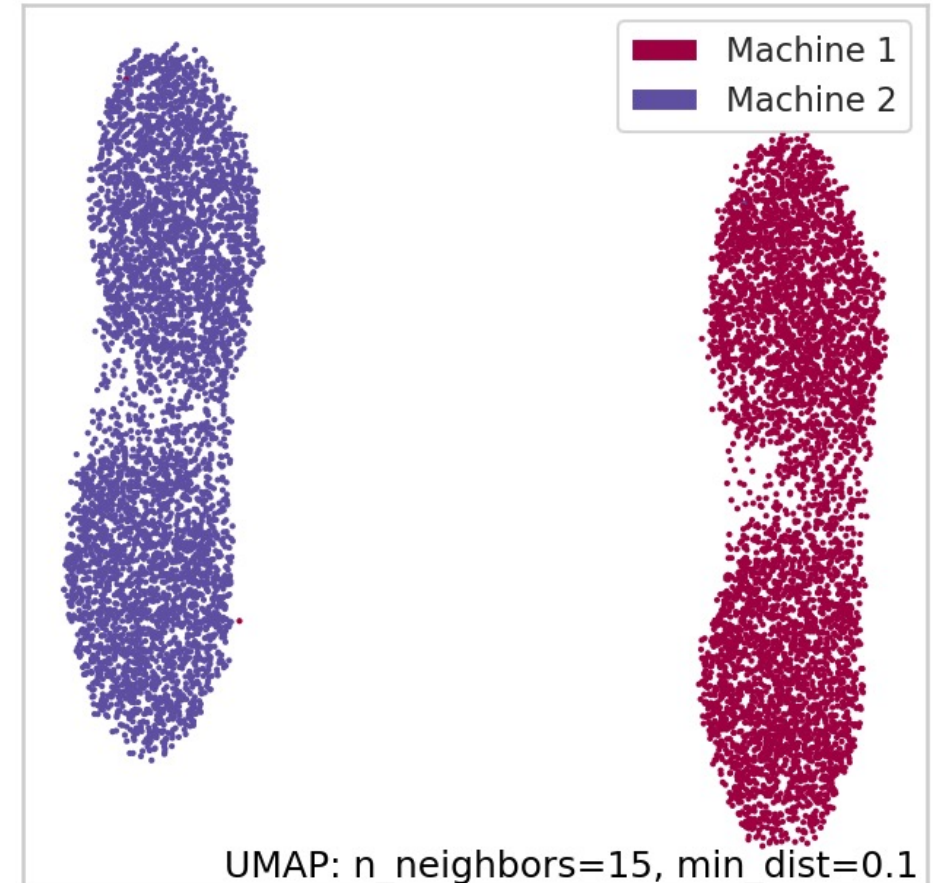
# Entangled Domains



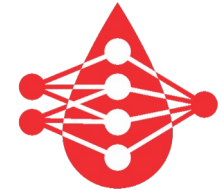
Classes



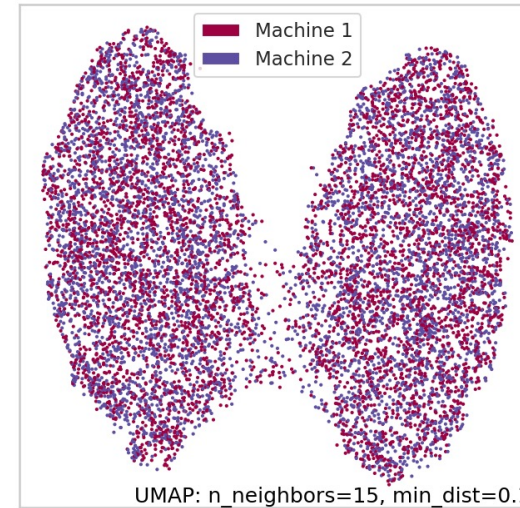
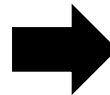
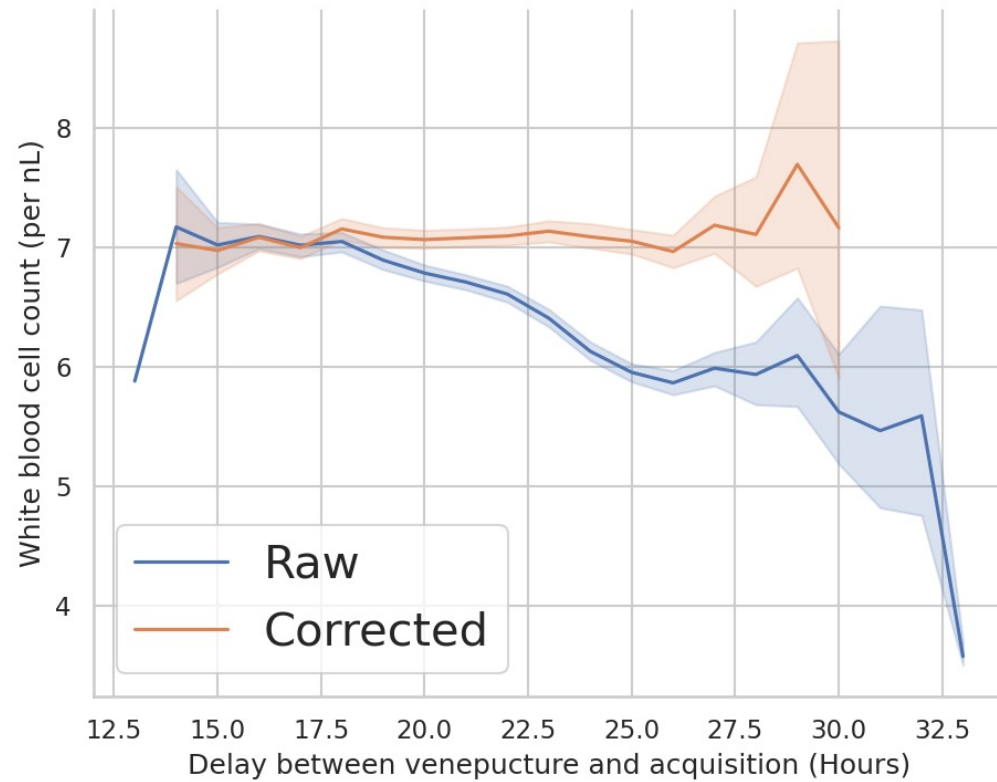
Machines



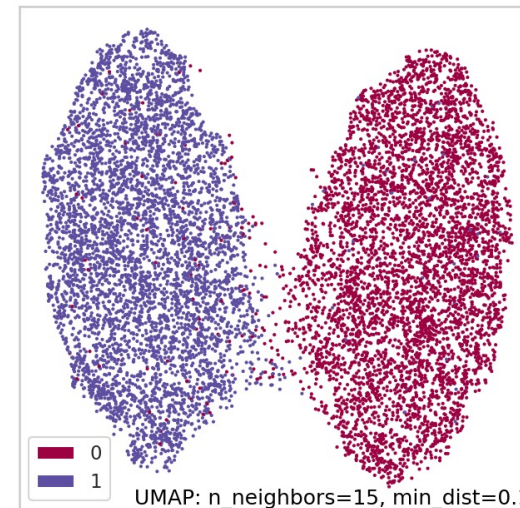
# One way...



Correction using p-splines



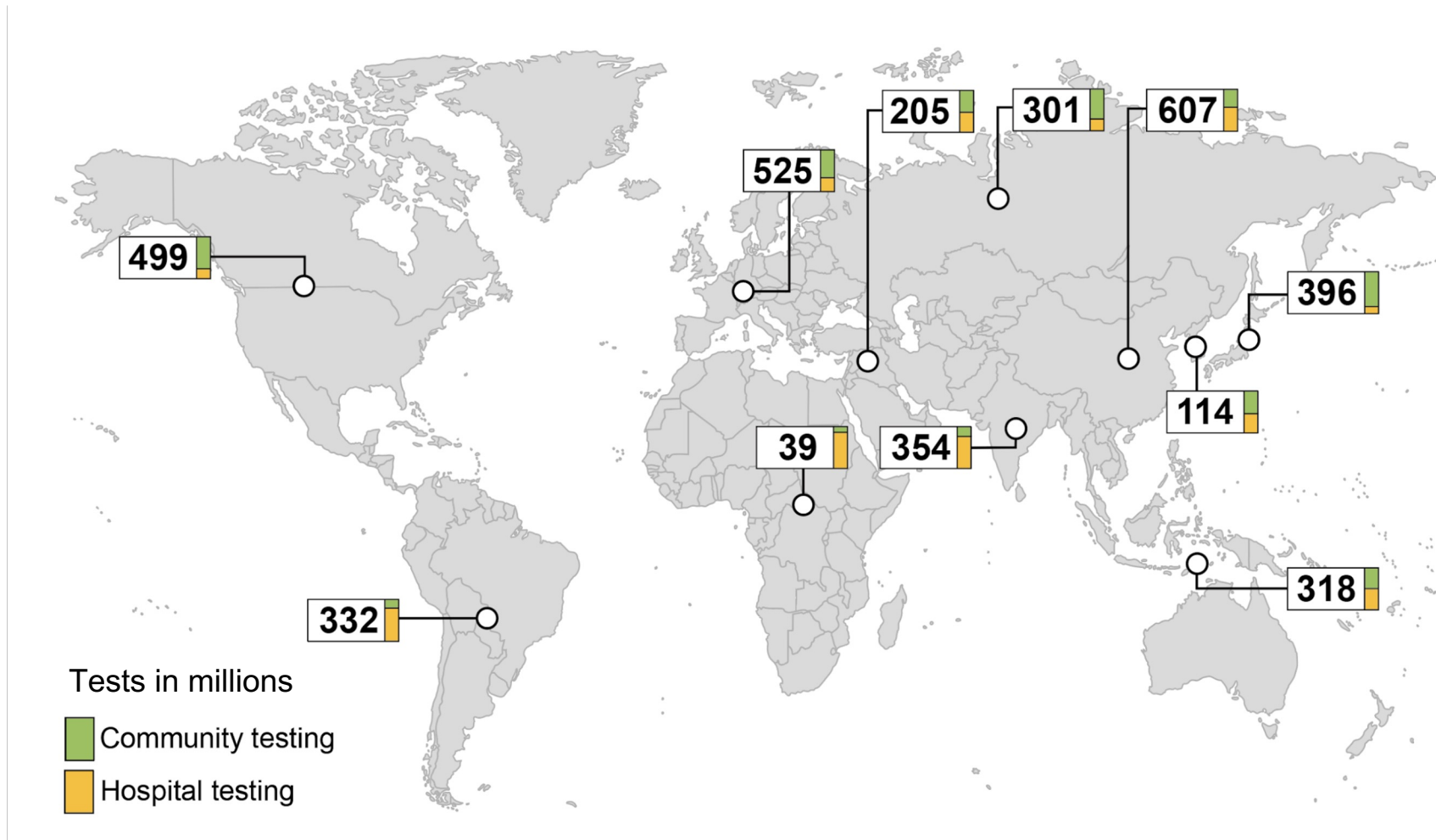
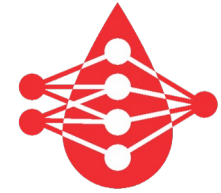
Machines



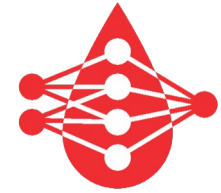
Classes



# Scalability & Comparability



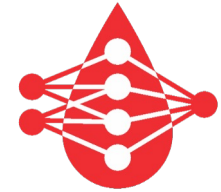
# Proposition



	Conventional methods	Desired method
<b>Model bias effects</b>	✓	✓
<b>Scalable</b>	✗	✓
<b>Comparable</b>	✗	✓
<b>Reproducible</b>	✗	✓
<b>Adaptable</b>	✗	✓



# BloodCounts! Team



Dr. Michael Roberts



Dr. Nicholas Gleadall



Dr. Julian Gilbey



Samuel Tull



UNIVERSITY OF  
CAMBRIDGE



Cambridge University Hospitals  
NHS Foundation Trust



University College London Hospitals  
Biomedical Research Centre



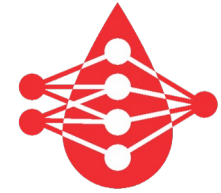
Cambridge Biomedical  
Research Centre



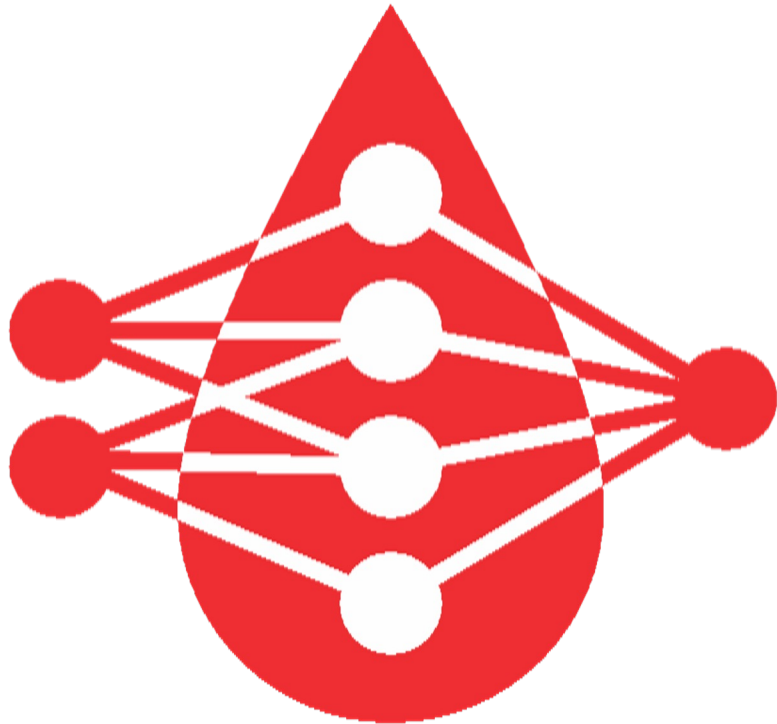
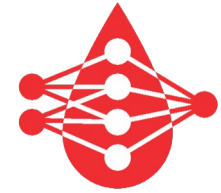
*Blood and Transplant*



# BloodCounts! Team



# Thank You!



# UNIVERSITY OF CAMBRIDGE

**NIHR** | University College London Hospitals  
Biomedical Research Centre



**Cambridge University Hospitals**  
NHS Foundation Trust

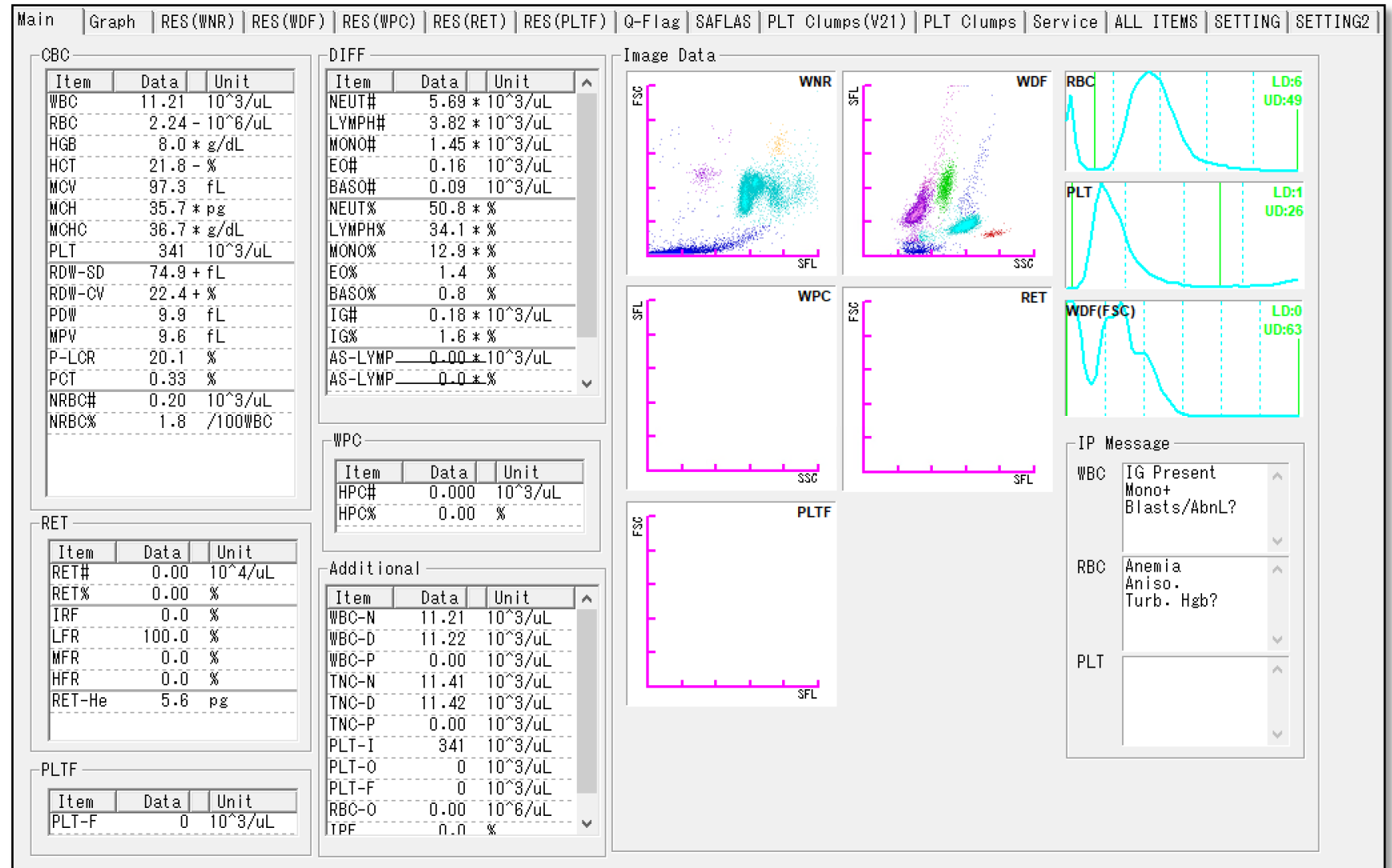
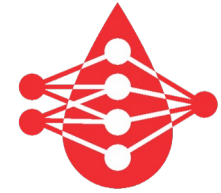
**NIHR** | Cambridge Biomedical  
Research Centre



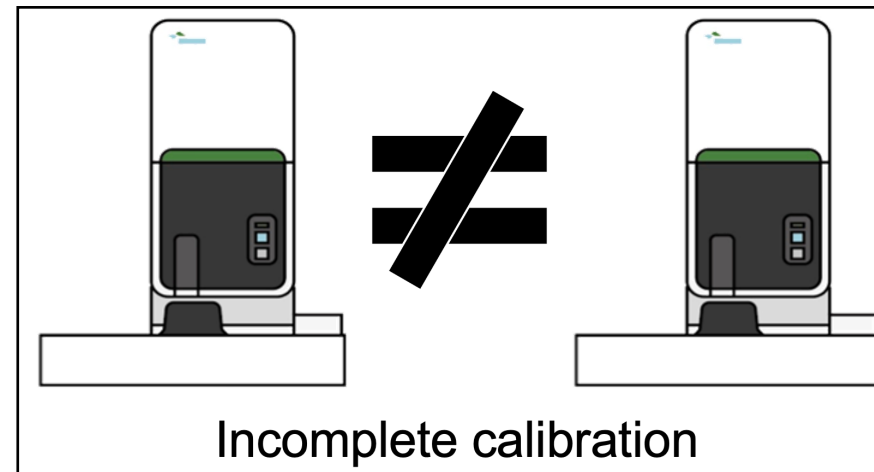
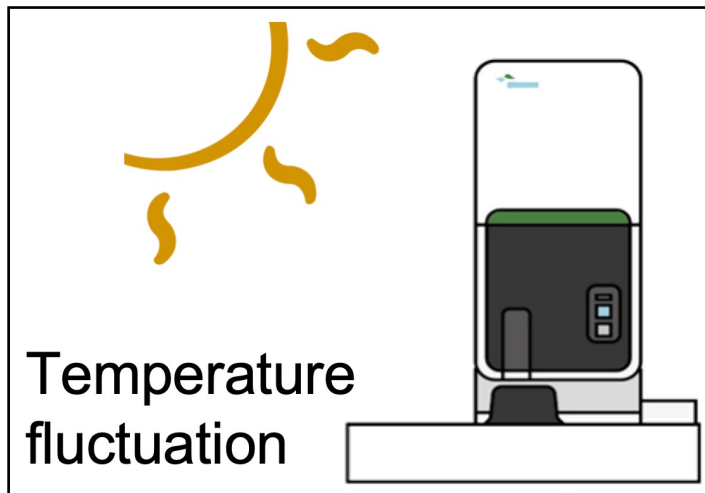
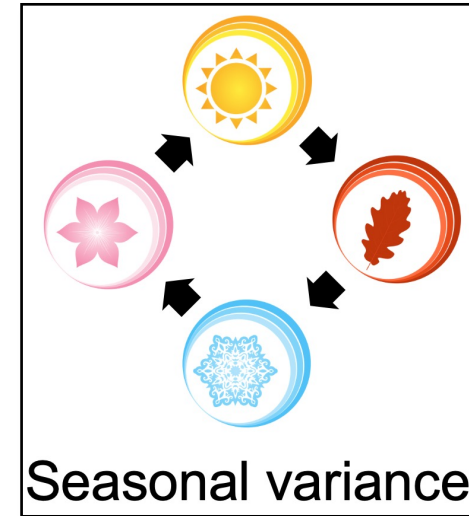
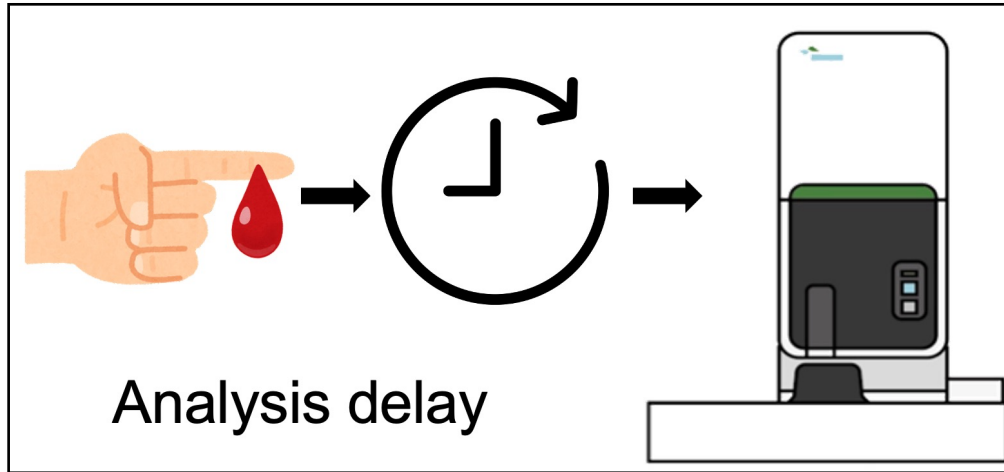
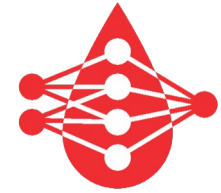
**Blood and Transplant**



# Full Blood Count

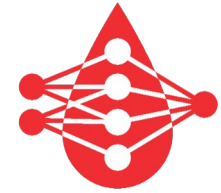


# Underlying Biases



# Problem

---

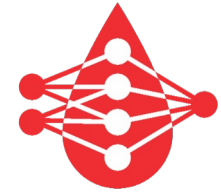


$$\tilde{x} = x + \varepsilon_{\text{ven. delay}} + \varepsilon_{\text{season}} + \varepsilon_{\text{machine}} + \dots$$

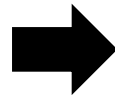
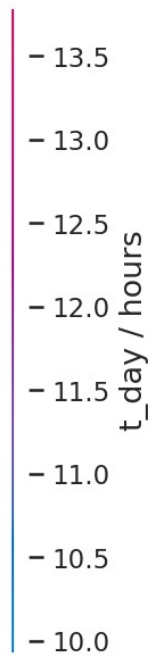
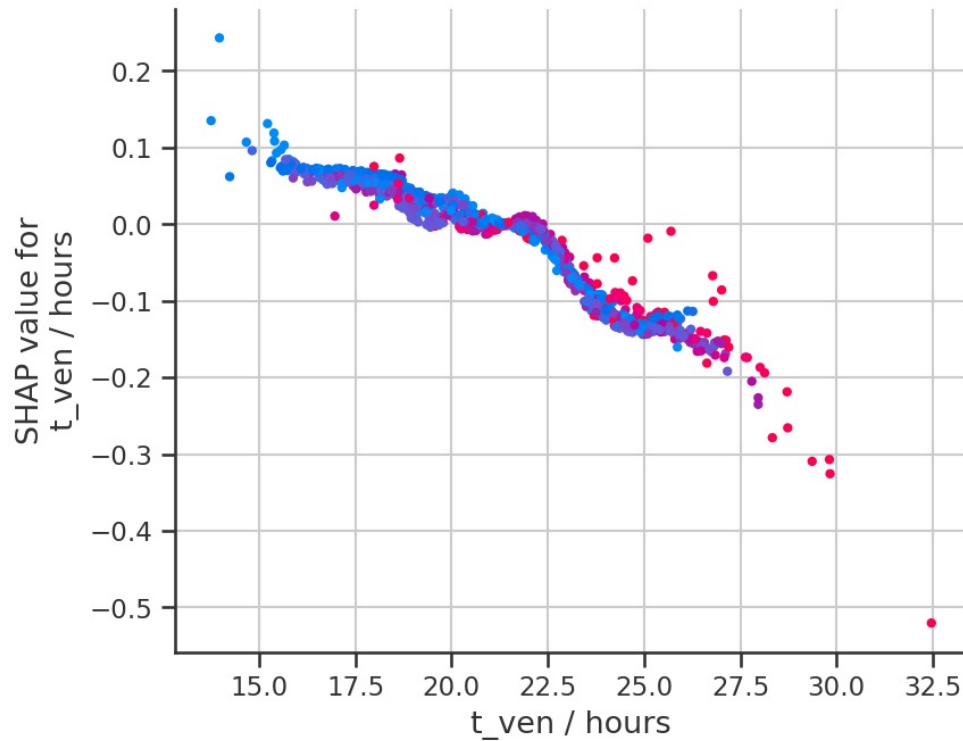
- Distorted data
- More biases could (and probably do) exist

→ Need method to disentangle bias domain from usable task domain

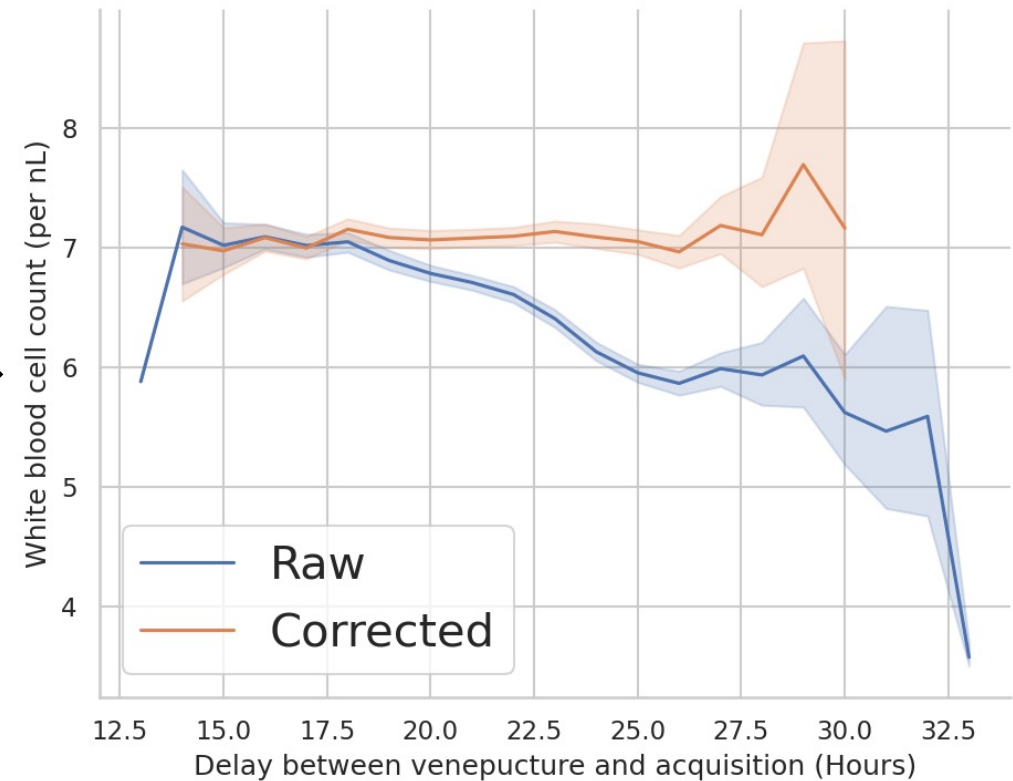
# One way...



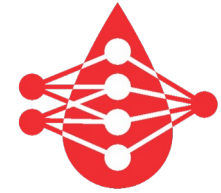
Spline fit



Correction

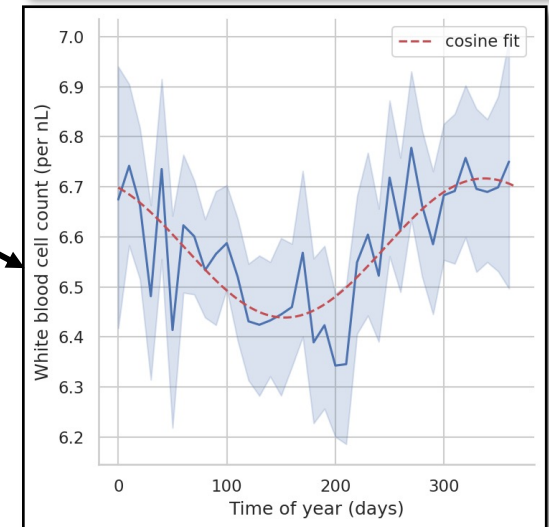
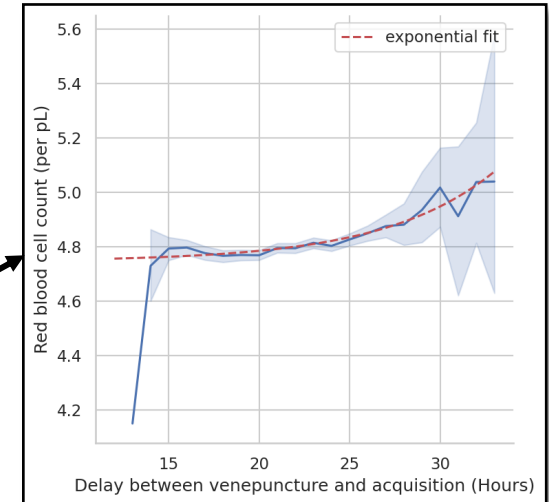
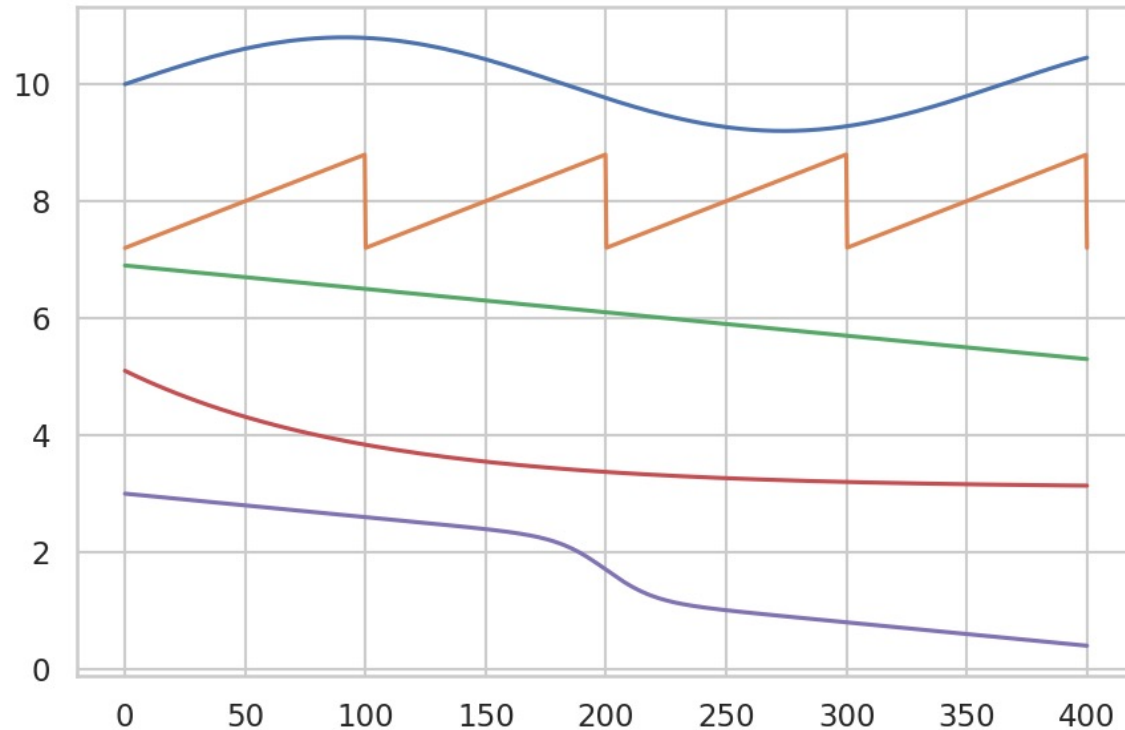






# Modelling Biases

Bias functions



# Icons

