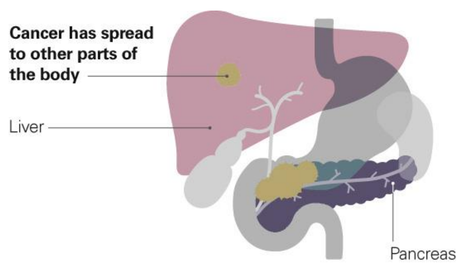
Can we screen for pancreatic cancer?

Identifying a sub-population of patients at high risk of subsequent diagnosis using machine learning techniques applied to primary care data

Ananya Malhotra; Morghan Hartmann; Bernard Rachet; Audrey Bonaventure; Stephen P Pereira; Andrew Green; Han-I Wang; Laura M Woods

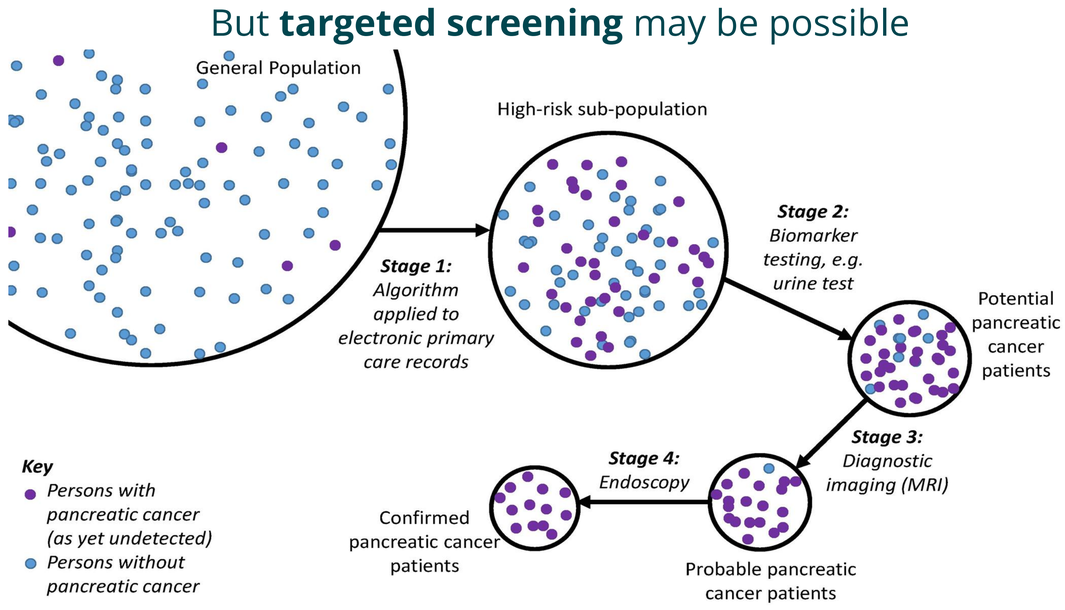
1. Background

Pancreatic cancer **diagnosed late** as no early-stage symptoms

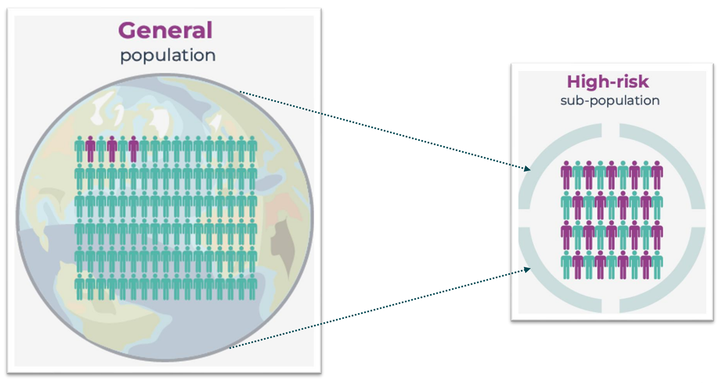


Population level screening impractical:
low incidence | high cost

But **targeted screening** may be possible



2. Pilot study[1] conducted using CPRD GOLD



1139 persons diagnosed with pancreatic cancer 2005-10

4556 cancer patient controls

Logistic regression (machine learning setting)

Population identified for screening reduced by 58%:

15-60 year olds
20 months before diagnosis
AUC = 66%
63% cancers detected early[2]

61-99 year olds
17 months before diagnosis
AUC = 61%
57% cancers detected early[2]

Take home message:

Targeted screening + **Biomarker testing** = Around **60% of pancreatic cancers** could be **diagnosed earlier** than currently detected

But… some limitations:

- Relatively small sample size (n≈1000)
- Non-pancreatic cancer population controls
- Older data (<2010)

3. Current project

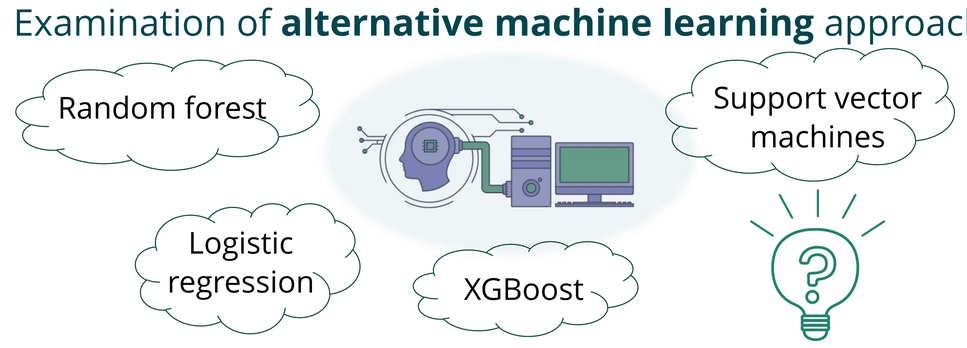
Full-scale case-control study and economic evaluation

Data

- CPRD **GOLD** & **AURUM** linked to cancer registration data
- **Cases -** Aged 15-99 years | Diagnosed 2010 – 2022
- **Controls -** Population-based | 4 per case

Methods

- Age, sex, GP practice **matched controls**
- **57+ risk factors** from **0-5 years** prior to cancer diagnosis
- Examination of **alternative machine learning** approaches

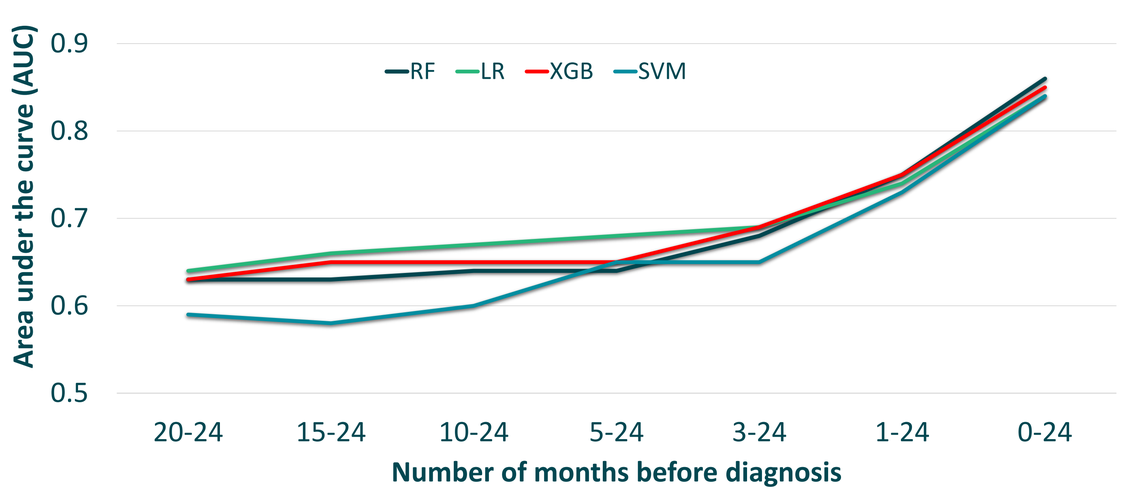


- Stratified analysis among **smokers** & **diabetics**
- Assessment of effectiveness in **'real time'**
- Economic evaluation of **heath service cost-effectiveness** NHS

4. Progress so far

- **New data** application approved | December 2022
- Literature search for **newly identified risk factors**
- **Updating diagnostic code lists** from CPRD
- **Patient and public involvement** sessions
  9 conversations about the diagnostic journey
- **Screencast** being developed based on people's lived experience

- Testing various **machine learning models**



References

1. Malhotra A, Rachet B, Bonaventure A, Pereira SP, Woods LM. Can we screen for pancreatic cancer? Identifying a sub-population of patients at high risk of subsequent diagnosis using machine learning techniques applied to primary care data. PLoS One. 2021;16(6):e0251876.
2. Radon TP, Massat NJ, Jones R, Alrawashdeh W, Dumartin L, Ennis D, et al. Identification of a Three-Biomarker Panel in Urine for Early Detection of Pancreatic Adenocarcinoma. Clinical Cancer Research. 2015;21(15):3512–3521.