Evidence-Based Early Diagnosis

2024

The Gateway,
St Andrews

Wednesday 29 May–
Friday, 31 May

Table of Contents

EBED Team

Prof. Peter Donnelly
Director, Mackenzie Institute for Early Diagnosis
Chair in Public Health, University of St Andrews

Prof. Donnelly is the Director of the Mackenzie Institute for Early Diagnosis. Having spent many years prior working with the Scottish Government as Deputy Chief Medical Officer, Peter was, most recently, Chief Executive Officer of a large Public Health Agency in Canada. During Peters time as Deputy Chief Medical Officer between 2004 and 2008, Peter worked on pandemic influenza planning, as well as a variety of other projects including the very successful indoor smoking ban, universal sex education, and minimum pricing of alcoholic drinks based on alcohol content.

Prof. Frank Sullivan
Director of Research & Professor of Primary Care, University of St Andrews

Frank Sullivan has been an academic GP since 1984. He was appointed as the Professor of Primary Care Medicine in the University of St. Andrews in 2017 where he is also the Director of Research in the School of Medicine. He won the British Medical Association Research paper of the year in 2009 and was elected a Fellow of the Royal Society of Edinburgh in 2011 – the first family physician since 1908. His clinical practice is currently in Glenrothes.

Dr Margaret McCartney
Senior Clinical Lecturer in General Practice, University of St Andrews
Freelance writer and broadcaster

Margaret Mary McCartney is a general practitioner, freelance writer and broadcaster based in Glasgow. Dr McCartney is a vocal advocate for evidence-based medicine, and was a regular columnist at the British Medical Journal. She regularly writes articles for The Guardian and currently contributes to the BBC Radio 4 programme, Inside Health. She has written three popular science books, The Patient Paradox, The State of Medicine and Living with Dying. During the COVID-19 pandemic, She contributed content to academic journals and broadcasting platforms, personal blog, and social media to inform the public and dispel myths about COVID-19.

EBED Team

Prof. Carl Heneghan
Professor of EBM & Director, Centre for Evidence-Based Medicine,
University of Oxford
NHS Urgent Care GP

Carl Heneghan is a clinical epidemiologist with expertise in evidence-based medicine, research methods, and evidence synthesis expertise. His work includes investigating the evidence for approval of drugs and devices, assessing health claims and researching common presenting conditions in the community. He has investigated antivirals Tamiflu, acute respiratory infections and the transmission of SARs-CoV-2. He has expertise in medical device regulation, diagnosis and screening and avoidable harms.

Ms Ruth Davis
Centre Manager, Centre for Evidence-Based Medicine,
University of Oxford

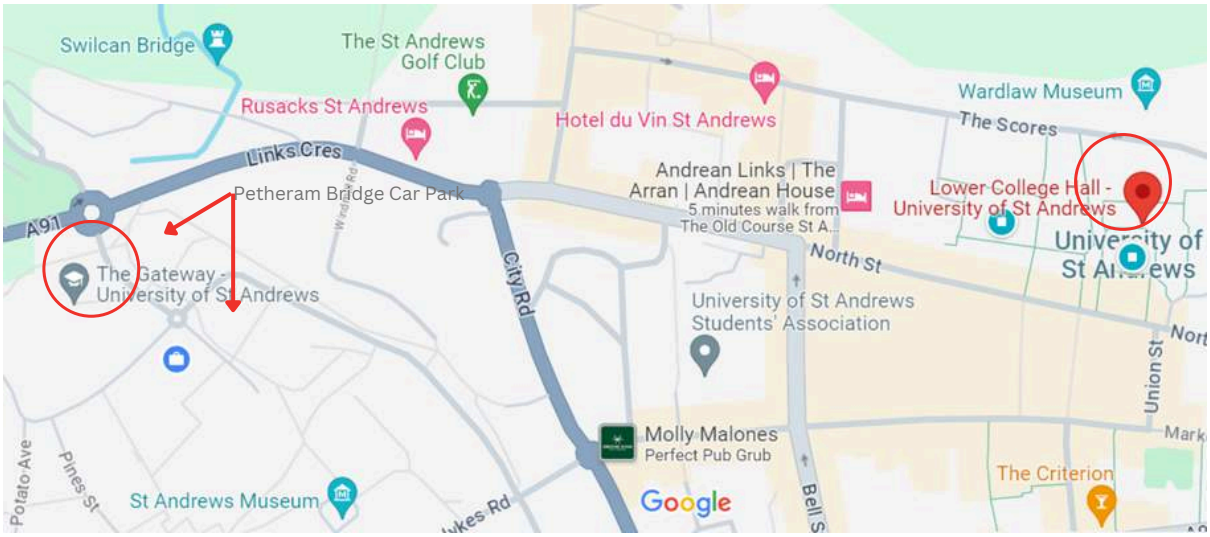
Ruth is the Centre Manager at the Centre for Evidence-Based Medicine, after spending two years as Programme Manager for the MaDOx group. She is responsible for maintaining the Centre's ability to respond to new initiatives and update its methods of interaction and dissemination. Elevating the position of all EBM and EBHC learning related activities and the relationship with the Department of Primary Care Health Sciences and Department of Continuing Education.

Prof. Jon Deeks
Professor of Biostatistics, Institute of Applied Health Research
University of Birmingham

Jon Deeks is Professor of Biostatistics and leads the Biostatistics, Evidence Synthesis and Test Evaluation Research Group in the Institute of Applied Health Research. He is also a Theme Lead within the NIHR Birmingham Biomedical Research Centre. Jon's current major focus is in test evaluation. He is the senior methodologist on numerous primary evaluations and systematic reviews of medical tests, leads the Cochrane Collaboration's test evaluation activities, and has advises the WHO on test evaluation methods. He is an NIHR Senior Investigator Emeritus and Fellow of the Academy of Medical Sciences. He is an enthusiastic teacher of statistics and research methods, and frequently runs workshops, particularly related to test evaluation, at local, national and international events.

Venue Details

Conference: The conference will be held in the Gateway, North Haugh Campus, St Andrews, KY15 9TF. Tea/coffee/lunch will be in the Well Foyer and all other sessions will be held in Lecture Rooms 3/4.



Wi-Fi access: Eduroam is available and can be accessed using the credentials of your home institute.

Conference Dinner: Thursday, 30 May at 18:45 in Lower College Hall, St Salvator's Quadrangle, North St, St Andrews KY16 9AL. The venue is walking distance from the conference.

Parking: Free parking is available in the Petheram Bridge Car Park.

Programme

Wednesday, 29 May 2024

12:00 - 13:00	**Arrival and Lunch** Gateway - Well Foyer
13:00-13:15	**Welcome and Introduction from Prof. Peter Donnelly** Gateway Lecture 3&4 combined
13:15-14:15	**Keynote Session 1** **Setting the Scene: What could success look like for the introduction of new diagnostic tests?** Gateway Lecture 3&4 combined Patient Representative - **Anne Fearfull** Industry Representative - **Andreas Halner,** Oxford Cancer Analytics (OXcan) Regulator Representative - **Joseph Burt,** Medicines and Healthcare products Regulatory Agency (MHRA) Journal Representative - **Helen Macdonald,** The *British Medical Journal*
14:15-15:30	**Keynote Session 2** **Theoretical Considerations** Gateway Lecture 3&4 combined **Dr. Margaret McCartney,** Senior Lecturer, University of St Andrews, freelance writer and broadcaster **Prof. John Brodersen**, University of Copenhagen, Department of Public Health
15:30-16:00	**BREAK** Gateway - Well Foyer

Programme

Wednesday, 29 May 2024

16:00 – 16:45 **Abstract Session - Elevator Pitches**

Gateway Lecture 3&4 combined

Theoretical Considerations

Gamuchirai Pamela Gwaza: Enhancing Early Diagnosis through Integrated Approaches: Practical Insights for Designing Interventions in LMICs - An Expert Consensus

Huw Llewelyn: Assessing the use of tests for early diagnosis: predicting at what stage of disease progression the probability of benefit from treatment exceeds harm

Clinical Perspective

Ritah Nakiboneka: Deploying host transcriptional markers for diagnosis of tuberculosis

Andrew Hall: The International Multicentre Project Auditing COVID-19 in Trauma & Orthopaedics (IMPACT): Using routinely-collected and audit-derived health data to evaluate and predict the effects of Coronavirus Disease 2019 on patients with a hip fracture

Rishma Maini: Integrating liquid biopsies into Rapid Cancer Diagnostic Services

Patient Perspective

Simon Baldwin: Birmingham Self-test Report 1: Review of rapid self-test diagnostics sold in supermarkets and chemists in the UK: their descriptions, manufacturers, distributors, and regulators.

Ridhi Agarwal: Birmingham Self-test Report 2: Are the statements of intended use and indicated medical actions for rapid self-test diagnostics sold in supermarkets and chemists appropriate?

Programme

Wednesday, 29 May 2024

16:00 – 16:45	**Bethany Hillier:** Birmingham Self-test Report 3: What is the evidence base for claims of accuracy for rapid self-test diagnostics sold in UK retail settings? **Jon Deeks:** Birmingham Self-test Report 4: Are the equipment, sampling, and instructions of rapid self-test diagnostics sold in UK retail settings correct, usable and safe, and are documents readable and in line with national and international guidelines? **Policy and Regulation** **Oscar Khawar:** Guidelines and regulations applicable to vitro diagnostic tests in the UK: a scoping review
16:45 – 17:00	**BREAK** Well Foyer - Gateway
17:00-18:30	**Keynote Session 3** **Clinical Perspective** Gateway Lecture 3&4 combined **Prof. Carl Heneghan,** University of Oxford, Centre for Evidence-Based Medicine **Prof. Susan Moug,** University of Glasgow, School of Medicine, Dentistry and Nursing **Prof. Alex Richter,** University of Birmingham, Institute of Immunology & Immunotherapy Discussion
18:30	**Drinks Reception in the Gateway, followed by walking tour of St Andrews led by University Students**

Programme

Thursday, 30 May 2024

07:15	**Beach swim/run with Dr Margaret McCartney & Prof. Frank Sullivan** Meet in Gateway Lobby
09:00 – 09:30	**Networking with Coffee, Pastries, and Bacon Rolls** Well Foyer - Gateway
09:30 – 10:30	**Keynote Session 4** **Policy and Regulations** Gateway Lecture 3&4 combined **Prof. Jon Deeks,** University of Birmingham, Institute of Applied Health Research **Dr. Stuart Hogarth,** University of Cambridge, Sociology Research Discussion
10:30 – 11:00	**BREAK** Well Foyer - Gateway
11:00 – 12:00	**Keynote Session 5** **Economic Issues** Gateway Lecture 3&4 combined **Prof. Rebecca Fitzgerald,** University of Cambridge, Department of Oncology **Prof. Bethany Shinkins,** University of Warwick Medical School Andreas Halner, Oxford Cancer Analytics (OXcan) Discussion

Programme

Thursday, 30 May 2024

12:00 – 13:15 **Lunch**

Well Foyer - Gateway

13:15 – 14:30 **Oral Abstract Session - Theoretical Considerations**

Gateway Lecture 3&4 combined

Martha Elwenspoek: Creating evidence-based optimal testing strategies for monitoring long-term conditions in primary care

Katie Charlwood: What is the impact of regular monitoring with specific blood tests in people with long term conditions on patient outcomes? Trial emulation using routinely collected primary care data.

Jacqueline Dinnes: Assessing the value of diagnostic tests: evaluation of a framework for identifying and organising test effects

Sian Taylor-Phillips: Intermediate endpoints as sufficient surrogates for cancer-specific mortality in cancer screening trials: A systematic review and meta-analysis

Stephen Bradley: Interpreting diagnostic accuracy studies based on retrospective routinely collected data

Katerina-Vanessa Savva: Real world implementation of the Biomarker Toolkit: a Tool aiming to quantifiably assess biomarker utility and guide development

---

14:30 – 15:00 **BREAK**

Well Foyer - Gateway

Programme

Thursday, 30 May 2024

15:00 – 16:00

Oral Abstract Session - Clinical & Patient Perspective

Gateway Lecture 3&4 combined

Clare Turnbull: Polygenic risk stratification for breast, colorectal and prostate cancer screening in the UK: integration of multiple national routinely collected cancer datasets for modelling of potential impact on cancer-specific mortality

Frank Sullivan: 5 year mortality in a Randomized Controlled Trial of an autoantibody biomarker for Lung cancer.

Obaid Kousha: Pragmatic and scalable diabetic retinopathy screening for lower resource settings: Binocular indirect ophthalmoscopy versus a retinal camera, including Artificial Intelligence (AI) interpretation in Indonesia

Alexandra Brandt Ryborg Jønsson: Logics of Time and Diagnosis

---

16:00 – 17:30

Meet the Expert Sessions

Associate Prof. Brian Nicholson, University of Oxford: How should we evaluate novel cancer diagnosis?

Prof. Clare Turnbull, The Institute of Cancer Research : What can polygenic testing contribute to early diagnosis

Prof. Frank Sullivan, University of St Andrews School of Medicine: A doctoral training programme in Early Diagnosis?

Mr. Chris Peters, Imperial College London: Why so few biomarkers make it into clinical practice

Prof. Clare Davenport, University of Birmingham: Guidance for the Regulation, evaluative, marketing and Monitoring of Direct to Consumer Testing

Programme

Thursday, 30 May 2024

17:30 – 18:00	**Tour of the Medical School Building** Meet in Gateway Lobby
18:45	**Conference Dinner** Lower College Hall

Programme

Friday, 31 May 2024

09:00 – 09:30	**Networking & Coffee** Well Foyer, Gateway
09:30 – 10:30	**Oral Abstract Session - Clinical Perspective & Policy and Regulation** Gateway Lecture 3&4 Combined **Sarah Mills:** Developing A Risk Prediction Tools For Near Term Mortality In Patients Who Present To Unscheduled Care In Scotland **Clare Davenport:** Developing guidance for the evaluation, regulation, marketing, and monitoring of Direct to Consumer Tests (DTCTs)- 'GUIDE DTCTs' **Allyson Pollock:** Global burden of disease estimates for Major Depressive Disorder: instruments used in studies to measure prevalence of MDD not designed for that purpose, contribute to risk of over-diagnosis and over-treatment. **James Larkin:** Payments to healthcare organisations reported by the medical device industry in Europe from 2017 to 2019: an observational study
10:30 – 12:00	**Feedback from "Meet the experts" small group discussions - Identifying gaps and how to take things forward** Gateway Lecture 3&4 Combined
12:00 – 13:00	**Summary and Future Planning**
13:00 – 14:00	**Lunch and Depart** Well Foyer, Gateway

Keynote Speakers

Prof. John Brodersen

Professor,
Centre of General Practice, University of Copenhagen
Research Unit for General Practice, Region Zealand, Denmark
Research Unit for General Practice, Department of Community Medicine, Faculty of Health Sciences, UiT The Arctic University of Norway, Tromsø

John Brodersen is general practitioner with over ten years experience in clinical practice. Dr Brodersen has a PhD in public health and psychometrics and works as an associate research professor in the area of medical screening at University of Copenhagen, Department of Public Health, Research Unit and Section of General Practice.

His research is focused on the field of development and validation of questionnaires to measure psychosocial consequences of false-positive screening results. He has employed qualitative and quantitative methods e.g. developed patient reported outcomes measures qualitatively and validated those using Rasch models to objectify subjective areas like psychosocial consequences. Dr Brodersen has published widely in peer reviewed journals.

Prof. Jon Deeks

Professor of Biostatistics
Institute of Applied Health Research
University of Birmingham

Jon Deeks is Professor of Biostatistics and leads the Biostatistics, Evidence Synthesis and Test Evaluation Research Group in the Institute of Applied Health Research. He is also a Theme Lead within the NIHR Birmingham Biomedical Research Centre. Jon's current major focus is in test evaluation. He is the senior methodologist on numerous primary evaluations and systematic reviews of medical tests, leads the Cochrane Collaboration's test evaluation activities, and has advises the WHO on test evaluation methods. He is an NIHR Senior Investigator Emeritus and Fellow of the Academy of Medical Sciences. He is an enthusiastic teacher of statistics and research methods, and frequently runs workshops, particularly related to test evaluation, at local, national and international events.

Keynote Speakers

Ann Fearfull
Patient Representative

Prof. Rebecca Fitzgerald
Professor of Cancer Prevention
Founding Director of the Early Detection Institute
University of Cambridge
Hon. Consultant in Gastroenterology and Cancer Medicine
Addenbrooke's Hospital

Rebecca Fitzgerald OBE FRS FMedSci FRCP EMBO is Professor of Cancer Prevention and Founding Director of the Early Detection Institute at the University of Cambridge and practices medicine as Hon. Consultant in Gastroenterology and Cancer Medicine at Addenbrooke's Hospital. Rebecca is the Cambridge lead for the CRUK International Alliance in Early Detection (ACED). Her research aims to understand how tissues become cancerous and whether identifying pre-cancer at scale can reduce cancer morbidity and mortality, focussing on the oesophagus and stomach. Her work to develop and implement a non-endoscopic capsule sponge and related biomarker assays for detection of Barrett's oesophagus and associated dysplasia has been awarded several prizes including the Westminster Medal, an NHS Innovation prize and the Don Listwin Early Detection Prize. In 2022 Rebecca was awarded an OBE for services to cancer research. Rebecca has contributed to evidence reviews and policy work around screening including for the Department of Health in the UK and recently led a review of cancer screening for the European Commission that led to new screening policy for EU member states.

Dr Andreas Halner
President and Co-Founder
Oxford Cancer Analytics Ltd

Andreas Halner completed pre-clinical medicine training and a DPhil (PhD) in Clinical Medicine and Machine Learning at the University of Oxford. Andreas' leadership

Keynote Speakers

experience includes his role as the Chief Data Scientist of a European Clinical Research Collaboration on lung disease from 2019 onwards. From 2018-2023, Andreas has held the Head Pathology Tutor for Medicine post at St John's College, providing one third of the medical curriculum for second year medical students. He has designed multiple new mathematical and clinical paradigms for defining disease states, developing algorithms for treatment outcome prediction and treatment monitoring. Andreas is an experienced entrepreneur and has mentored numerous start-up companies in the healthcare and biotech sectors.

Prof. Carl Heneghan

Professor of EBM & Director,
Centre for Evidence-Based Medicine, University of Oxford
NHS Urgent Care GP

Carl Heneghan is a clinical epidemiologist with expertise in evidence-based medicine, research methods, and evidence synthesis expertise. My work includes investigating the evidence for approval of drugs and devices, assessing health claims and researching common presenting conditions in the community.

Dr Stuart Hogarth

Associate Professor in Sociology of Science and Technology
Fellow of Robinson College
University of Cambridge

Dr Hogarth is a Lecturer in Sociology of Science and Technology. His work focuses on biomedical innovation and his research has investigated a diverse range of emergent biotechnologies, such as stem cell therapies and synthetic biology. His primary interest is the impact of genomic science on the diagnostics sector, and he has published extensively on the political economy of diagnostic innovation, with a particular focus on regulatory governance and intellectual property rights.

Dr Hogarth uses an international comparative methodology to explore the continued salience of national institutions such as regulatory regimes and healthcare systems, in a bioeconomy which is increasingly characterised by global governance structures, international scientific collaborations and transnational flows of capital and scientific labour.

Keynote Speakers

Dr Helen Macdonald

Publication Ethics & Content Integrity Editor
British Medical Journal

Helen Macdonald graduated from Barts and The London Queen Mary's School of Medicine Dentistry, London (2006). She has worked as an editor at The BMJ since 2008 (beginning as an editorial registrar). Currently, she is the UK research editor, continues to develop The BMJ's Rapid Recommendations series, and champions aspects of our campaigns on Better Evidence and Too Much Medicine. She has previously headed the analysis and education sections of the journals, and supported the Student BMJ's editorial team.

After two years a junior doctor in London, she split her time between The BMJ and GP training until she qualified as a General Practitioner in 2014. Along the way she also did a BA in Medical Journalism (first class honours, University of Westminster) and MSc in Evidence-Based Healthcare (distinction, University of Oxford). It is blending all of these skills together, to communicate clear and helpful information for discussions about health and healthcare, which drives her work at The BMJ.

Dr Margaret McCartney

Senior Clinical Lecturer in General Practice
University of St Andrews
Freelance writer and broadcaster

Margaret Mary McCartney is a general practitioner, freelance writer and broadcaster based in Glasgow. Dr McCartney is a vocal advocate for evidence-based medicine, and was a regular columnist at the British Medical Journal. She regularly writes articles for the Guardian and currently contributes to the BBC Radio 4 programme Inside Health. She has written three popular science books: The Patient Paradox, The State of Medicine, and Living with Dying. During the COVID-19 pandemic, she contributed content to academic journals and broadcasting platforms, a personal blog, and social media to inform the public and dispel myths about COVID-19.

Keynote Speakers

Prof. Susan Moug
Honorary Professor
University of Glasgow School of Medicine, Dentistry & Nursing

Susan Moug is an academic colorectal and general surgeon in Royal Alexandra Hospital, Paisley. Her research interests are linked by the common theme of improving surgical patient outcomes: frailty, older adult, lifestyle factors in colorectal cancer and contrast ultrasound for rectal cancer staging. She collaborates nationally and internationally within her own specialty, but also beyond, including geriatricians, bioengineers, and physicists.

She currently holds a NRS Chief Scientist Office (CSO) Senior Research Fellowship and is one of two Surgical Specialty Leads for Colorectal (RCSEng). She has recently been appointed to Director of Research for ASGBI. She currently holds several grants (including CRUK) and she is local and chief investigator for several UK led trials.

Prof. Alex Richter
Clinical Immunologist
University of Birmingham

I am a Professor of Clinical Immunology, and my research is focused on the development of immunodiagnostics and establishing their use in clinical care pathways to improve patient diagnosis and outcomes. I am a practicing clinician that cares for patients with primary and secondary immunodeficiency and also Director of the Clinical Immunology Service (CIS) at the University of Birmingham. As the anchor tenant for the forthcoming Precision Health Technologies Accelerator, our ISO15189 accredited laboratory is poised to transition to the state-of-the-art Birmingham Health Innovation Campus. This strategic move will position us at the forefront of innovation, fostering dynamic collaborations among academia, the NHS, and Industry. I lead a successful development pipeline of immunodiagnostic assays and offer advisory services through the West Midlands Health Innovation Accelerator, contributing to the translation of research discoveries into tangible solutions for patient care.

Keynote Speakers

Prof. Bethany Shinkins
Associate Professor of Health Economics
Academic Unit of Health Economics
University of Leeds

Prof. Bethany Shinkins is an Associate Professor of Health Economics in the Academic Unit of Health Economics at the University of Leeds. She leads the Test Evaluation Group, a multi-disciplinary team that focuses on the economic evaluation of medical tests. She is a statistician by background and now works as both a statistician and health economist. She joined the University of Leeds in 2015 as a Lecturer in Health Economics. The vast majority of her research is focused on the evaluation of tests, spanning a wide range of diseases and clinical settings.

She sits on the Editorial Board for the *BMC Medical Research Methodology* journal and the *BMJ* Evidence-Based Medicine Journal and is a Senior Associate at the Centre for Evidence Based Medicine, University of Oxford.

Posters

1	**Enhancing Early Diagnosis through Integrated Approaches: Practical Insights for Designing Interventions in LMICs - An Expert Consensus** Gamuchirai Pamela Gwaza, Annette Pluddemann, Marcy McCall, Sabine Dittrich, Carl Heneghan	Theoretical Considerations
2	**Assessing the use of tests for early diagnosis: predicting at what stage of disease progression the probability of benefit from treatment exceeds harm** Huw Llewelyn	
3	**Defining Clinical and Biological Rationale of Biomarkers to Improve the Rate of Translation** Alice Baggaley, Katerina-Vanessa Savva, Melody Ni, George Hanna, Christopher Peters	
4	**Deploying host transcriptional markers for diagnosis of tuberculosis** Ritah Nakiboneka, Natasha Walbaum, Emmanuel Musisi, Tonney Nyirend, Marriott Nliwasa, Chisomo Msefula, Derek Sloan, Wilber Sabiiti	Clinical Perspective
5	**The International Multicentre Project Auditing COVID-19 in Trauma & Orthopaedics (IMPACT): Using routinely-collected and audit-derived health data to evaluate and predict the effects of Coronavirus Disease 2019 on patients with a hip fracture** Andrew Hall, Nick Clement, Alasdair MacLullich, Tim White, AndrewDuckworth	
6	**Integrating liquid biopsies into Rapid Cancer Diagnostic Services** Rishma Maini, Neil Cruickshank, Peter Donnelly	
7	**Rapid Antimicrobial Susceptibility Testing of Urinary Tract Infection (UTI) Bacteria Using an Innovative Technology: Scattered Light Integrated Collector (SLIC).** Hellen Onyango, Derek Sloan, Katherine Keenan, Mike Kesby, Robert Hammond	

Posters

8	**Antimicrobial Photodynamic Therapy using Organic Light Emitting Diodes: bringing light closer to the skin** Marianna de Leite Avellar, Ifor Samuel, Robert Hammond	Clinical Perspective
9	**Testing efficacy of a novel diagnostic antimicrobial susceptibility testing platform on patient bacterial isolates from a large Scottish teaching hospital** Stuart Reid, Robert Hammond	
10	**Identification of plasma markers associated with oesophageal cancer treatment outcomes utilising metabolomics** Hasnain Ahmed, David Sumpton, Alejandro Huerta Uribe, Guillaume Piessen, Michael Hisbergue, Victor H. Villar, Alan Stewart	
11	**Improving early diagnosis of terminal cancer: Identification of demographic and clinical factors associated with having a very short prognosis at their time of diagnosis with cancer** Sarah Mills, Peter Donnan, Deans Buchanan, Blair H Smith	
12	**Birmingham Self-test Report 1: Review of rapid self-test diagnostics sold in supermarkets and chemists in the UK: their descriptions, manufacturers, distributors, and regulators** Simon Baldwin, Bethany Hillier, Katie Scandrett, Ridhi Agarwal, Aditya Kale, Joseph Alderman, Trystan Macdonald, Alex Richter, Clare Davenport, Jon Deeks	Patient Perspective
13	**Birmingham Self-test Report 2: Are the statements of intended use and indicated medical actions for rapid self-test diagnostics sold in supermarkets and chemists appropriate?** Ridhi Agarwal, Katie Scandrett, Bethany Hillier, Simon Baldwin, Aditya Kale, Joseph Alderman, Trystan Macdonald, Alex Richter, Clare Davenport, Jon Deeks	

Posters

14	**Birmingham Self-test Report 3: What is the evidence base for claims of accuracy for rapid self-test diagnostics sold in UK retail settings?** Bethany Hillier, Simon Baldwin, Katie Scandrett, Ridhi Agarwal, Aditya Kale, Joseph Alderman, Trystan Macdonald, Alex Richter, Clare Davenport, Jon Deeks	Patient Perspectives
15	**Birmingham Self-test Report 4: Are the equipment, sampling, and instructions of rapid self-test diagnostics sold in UK retail settings correct, usable and safe, and are documents readable and in line with national and international guidelines?** Jon Deeks, Clare Davenport, Alex Richter, Aditya Kale, Joseph Alderman, Trystan Macdonald, Bethany Hillier, Katie Scandrett, Ridhi Agarwal, Simon Baldwin	
16	**Guidelines and regulations applicable to vitro diagnostic tests in the UK: a scoping review** Oscar Khawar, Magdalena Staworko, Frank Sullivan, Peter D Donelly, Jon Deeks, Margaret McCartney	Policy & Regulation
17	**Investigation into the incidence of co-morbidities discovered after five years of follow-up in the Early Detection of Cancer of the Lung Scotland (ECLS) study.** Nimue Lilith Romeikat	
18	**Great promise and big problems: Applied epidemiology and the new diagnostics** Peter Donnelly	
19	**Budget impact analysis of using a novel urine biomarker test to support early diagnosis of pancreatic ductal adenocarcinoma** Rosario Luxardo, Katerina-Vanessa Savva, Silvana Debernardi, Tatjana Crnogorac-Jurcevic, Melody Ni, George B Hanna	Economic Issues

Posters

20	**Blood-Based Proteomic Biomarkers for Alzheimer's Disease Classification using Gradient Boosting Machines with Selection Bias Correction** Marco Fernandes, Victor Pardo, Paul Johnston, Peter Donnelly	Other
21	**Enhancing Colorectal Cancer Mismatch Repair Biomarker Prediction in Computational Pathology: A Comparative Analysis of Domain-Specific vs General-Purpose Feature Extractors for Weakly Labelled Colorectal Cancer Whole Slide Image Classification** Craig Myles, In Hwa Um, David Harrison, David Harris-Birtill	
22	**Our Future Health: the UK's largest health research programme** Iain Turnbull, Raghib Ali	

Abstracts

Keynote Sessions

The 10 commandments of test evaluation

Jon Deeks
University of Birmingham

Rational health policy requires reliable and relevant evidence of the performance of tests, both in terms of their accuracy and the impact that they have on health. Study design, execution, reporting and scientific integrity are essential to ensure that the public can trust claims of test performance. The right studies need to be done with the right people to compare the right tests to form evidence-based policy.

I will review ten issues which are essential to provide reliable evidence to support their use. I will highlight examples from various technologies including those that were evident in the evaluation of new tests in the Covid-19 pandemic. The list may help policymakers and the public to identify claims that are trustworthy and those that are not.

Why you need to befriend a health economist – lessons from innovating a capsule sponge test

Rebecca Fitzgerald
University of Cambridge, Department of Oncology

My talk will discuss the importance of including a health economic evaluation for new diagnostic technologies and how these considerations may be different depending on where the test fits into the clinical care pathway, and depending on the health care system.

Early diagnosis and the real world of clinical practice

Carl Heneghan
Centre for Evidence-Based Medicine, University of Oxford

Accurate early diagnosis is crucial to determining the prognosis and providing effective treatments. However, the methods employed to achieve early diagnosis need to be better planned. There is a pressing need to promote informed decision-

making about medical tests for early diagnosis. Current practice is leading to significant increases in testing, which can overwhelm medical practices. We need to inform clinicians and the public about the benefits, harms, and uncertainties associated with early diagnostic tests and improve research in this area.

Evaluating and Navigating the Integration of Tests into Clinical Care Pathways

Alex Richter
University of Birmingham

Navigating the integration of tests into clinical care pathways demands a multifaceted assessment that extends beyond mere diagnostic accuracy and clinical validity. While these are crucial aspects, equally important is how patients access tests, the consideration of how test results influence clinical decision-making, patient outcomes, and the overall cost-effectiveness of care delivery. However, despite the need for a nuanced approach, it is rare to assess a diagnostic within the context of the entire care pathway. Siloed budgets and fragmented healthcare systems often hinder comprehensive evaluations, leading to suboptimal decision-making and resource allocation. Integrating tests into clinical care pathways requires breaking down these barriers and adopting a holistic perspective that considers the interconnections between diagnostics, treatments, and patient outcomes.

Abstracts

Elevator Pitch & Poster Presentations

(T) Enhancing Early Diagnosis through Integrated Approaches: Practical Insights for Designing Interventions in LMICs - An Expert Consensus

Gamuchirai Pamela Gwaza1, Annette Pluddemann1, Marcy McCall1, Sabine Dittrich2, Carl Heneghan1
1University of Oxford, Oxford, United Kingdom. 2Technische Hochschule Deggendorf, Bavaria, Germany

Objectives
This study aimed to identify and establish a consensus on essential criteria for designing integrated diagnosis interventions at the primary healthcare (PHC) level in LMICs. The focus was on utilizing technology enabling point-of-care testing and same-day delivery of results, using the same technology for multiple assays and diseases.

Method
Employing the online Delphi method, a two-part series of surveys was conducted between July and November 2023 utilizing the JISC online survey tool. A diverse group of 55 experts, representing implementers, policymakers or funders, and researchers or academic experts, participated. Predetermined consensus thresholds were set at 70% agreement on a criterion being rated as 4 or Critical to Include.

Results
The study identified 18 core criteria deemed critical, showcasing the necessity for a comprehensive health systems perspective during intervention implementation. Three overarching themes emerged: the significance of leadership and governance, the need for compatible and contextualized diagnostic tools, and a focus on improving patient health outcomes and experiences.

Conclusions
These criteria offer a valuable guide for policymakers, funders, implementers, and manufacturers in prioritizing elements when designing interventions in LMICs. Special attention should be given to ensuring critical success factors are incorporated, emphasizing a holistic approach beyond diagnosis alone. The study advocates for an integrated strategy aligned with the entire care cascade and the broader healthcare system, providing a comprehensive and patient-centred framework for enhanced healthcare in LMICs.

Abstracts

Elevator Pitch & Poster Presentations

(T) Assessing the use of tests for early diagnosis: predicting at what stage of disease progression the probability of benefit from treatment exceeds harm

Huw Llewelyn
Aberystwyth University, Aberystwyth, United Kingdom

Objectives

The objective is to avoid over diagnosis and over-treatment as a result of screening. Early diagnosis using new tests (e.g. genetic tests) implies that they will be used for screening patients before they become symptomatic so that they can be treated earlier. However, only very few with a positive result may ever develop symptomatic disease. Other tests are required that identify higher proportion of patients who will develop symptomatic disease later so that action can be taken at a stage when more will benefit and fewer will be treated unnecessarily and suffer adverse effects. These tests will be markers of severity of the underlying disease process and its responsiveness to treatment. They might be performed repeatedly during follow-up in those with a positive screening result. Other tests are also needed that exclude as many other conditions as possible that might mimic the target disease but not respond to its treatment.

Method

'Severity tests' of possible usefulness are assessed with an RCT. The candidate tests are performed before randomisation. The subsequent results are divided into ranges. The proportions developing the outcome of interest during the RCT are recorded within each range of the results of each test. These proportions are can then used in a logistic regression to construct curves that display the probability of the outcome for each test result on treatment and control. A separate observational study is conducted on the candidate tests by observing how their baseline results change with time. The odds ratios, risk ratios and absolute risk reductions from the RCT are then applied to the baseline probabilities of the observational study to assess the probability of benefit at each stage of disease progression. This general method is illustrated with data from an RCT of an angiotensin receptor blocker s ability to reduce the frequency of nephropathy.

Results

If a urine dipstick test is regarded as a preliminary screening test (like a genetic test) for diabetic nephropathy, there is less than 1% chance that the patient would develop nephropathy within 2 years. When the subsequent severity test used was a

yearly albumin excretion rate (AER) and treatment was an angiotensin receptor blocker, the probability of nephropathy was less than 1% for an AER less 20mcg/min even on control. Above 20mcg/min, the logic regression curves showed progressively increasing probabilities of nephropathy and larger absolute risk reductions. Other example curves illustrate the performance of severity tests that would be more and less powerful than the AER. When the risk ratio, odds ratio and risk differences from the RCT are applied to low baseline probabilities below an AER of 20mcg/min, the absolute risk reduction is very small and the NNT very high, suggesting little chance of benefit on treatment.

Conclusions

This use of severity tests avoids treating patients with positive screening tests results when there is only a slightly increased probability of disease. Because of this, the probability of benefit is low and the risk of adverse effects from treatment is significant. The severity test can also overcome the problem of lead time bias by only treating those at a known stage of severity and stage of disease progression. The performance of severity tests might be improved by basing the test on a change in the result over time. The severity test is an important concept for use with new screening tests (especially those based on genetics) if the screening test only identifies those with a slightly increased risk of disease.

(CP) Deploying host transcriptional markers for diagnosis of tuberculosis

Ritah Nakiboneka1,2, Natasha Walbaum1, Emmanuel Musisi1,3, Tonney Nyirenda2, Marriott Nliwasa2, Chisomo Msefula2, Derek Sloan1, Wilber Sabiiti1
1University of St Andrews, St Andrews, United Kingdom. 2Kamuzu University of Health Sciences College of Medicine, Blantyre, Malawi. 3Adroit Biomedical and Bioentrepreneurship Research Services, Kampala, Uganda

Objectives

Tuberculosis (TB) disease is characterised by symptoms such as cough, loss of weight, night sweats and chest pain, clinical signs similar to other respiratory diseases (ORDs). Active TB (ATB) disease progresses from the asymptomatic state of latent TB infection (LTBI). Between LTBI and ATB is the subclinical (incipient) cohort of TB cases whose clinical signs are masked and difficult to discern. There is a need to accurately diagnose and manage at the different spectrum of infection and

disease. We evaluated a panel of human gene (transcriptomic) makers for ability to diagnose LTBI and ATB and distinguish them from healthy controls (HC) and ORDs respectively. The study objectives were to- identify host genes specifically expressed during TB infection and clinically evaluate their accuracy to diagnose LTBI and ATB.

Method

Cases presenting with TB-like symptoms were enrolled at healthcare facilities in Blantyre, Malawi. ATB disease was confirmed by sputum liquid culture, and sputum bacterial load was measured using the TB-Molecular Bacterial Load Assay (TB-MBLA). Household contacts of the ATB confirmed index cases and HIV negative healthy controls (HC) were tested for LTBI using QuantiFERON-TB Gold Plus Interferon gamma release assay (IGRA). Host gene expression in whole blood was quantified using reverse transcriptase quantitative polymerase chain reaction (RT-qPCR) assay. Spearman's rho correlation and logistic regression modelling were used to assess the association between the variables.

Results

A total of 243 participants-143 presumptive cases with TB-like symptoms, 49 TB-exposed (TBExp) household-contacts and 51 HC were included in the evaluation. ATB was confirmed in 43% (61/143) presumptive cases leaving 57% (82/143) denoted ORDs. Host genes: GBP5, DUSP3, CD64, BATF2, GBP6, C1QB, GAS6, KLF2, NEMF, ASUN, DHX29 expression was higher among ATB- than ORDs-and-LTBI-participants. CD64 achieved the highest accuracy for distinguishing ATB from ORDs with a 96.5% AUC, 90.2%-sensitivity, and 95.1%-specificity. Diagnostic performance of the genes was not different by HIV-status. LTBI was confirmed in 51% (25/49) TBEx participants. Gene expression was suppressed among LTBI cases compared to HC. ZNF296 and KLF2 performed best in distinguishing people with LTBI from HC. Gene expression in 43% (22/51) of IGRA-negative HC was consistent with LTBI. Two participants exhibited gene expression consistent with incipient TB.

Conclusions

The results demonstrate the potential of host gene expression as biomarkers for accurate diagnosis of latent- and active- TB.

Abstracts

Elevator Pitch & Poster Presentations

(CP) The International Multicentre Project Auditing COVID-19 in Trauma & Orthopaedics (IMPACT): Using routinely-collected and audit-derived health data to evaluate and predict the effects of Coronavirus Disease 2019 on patients with a hip fracture

Andrew Hall1,2,3, Nick Clement3,4,5, Alasdair MacLullich6,3, Tim White4,5, Andrew Duckworth6,4,5
1School of Medicine, University of St Andrews, St Andrews, United Kingdom. 2Department of Orthopaedics & Trauma, University of Edinburgh, Edinburgh, United Kingdom. 3Scottish Hip Fracture Audit, Edinburgh, United Kingdom. 4Edinburgh Orthopaedics, Edinburgh, United Kingdom. 5Scottish Orthopaedic Research Trust into Trauma (SORT-IT), Edinburgh, United Kingdom. 6Usher Institute, University of Edinburgh, Edinburgh, United Kingdom

Objectives

Patients who suffer a hip fracture are frail, co-morbid, and have a significantly higher mortality risk than age-adjusted uninjured people. At the outset of the COVID-19 pandemic a small number of studies suggested that hip fracture patients were more vulnerable to the disease, however shortcomings in the available literature included inconsistent diagnostic criteria, reporting standards, short follow-up durations, and the use of unadjusted analyses.

The objectives were to utilise population-level routinely-collected and audit-derived health data alongside classical research methods in order to investigate the following in the context of acute hip fracture: i) prevalence of COVID-19 and patterns of disease transmission; ii) independent effects of COVID-19 on mortality risk; iii) factors associated with poor outcomes among patients with COVID-19; iv) non-lethal effects of COVID-19; v) effects of the COVID-19 pandemic on clinical services, and strategies to reduce the risk of morbidity and mortality associated with COVID-19.

Method

The IMPACT global research collaboration was established in April 2020 to coordinate a rapid and practical clinician-led research response to the COVID-19 pandemic. Seven health data-driven IMPACT Hip Fracture Projects were conducted with the support of national and international bodies.

Abstracts

Elevator Pitch & Poster Presentations

Two multicentre cohort studies assessed COVID-19 prevalence, transmission, and effects on mortality risk (IMPACT Scot & IMPACT Scot 2) in March-April 2020. An international cohort study in 112 hospitals in 12 countries provided a global perspective on mortality risk and prognostication (IMPACT Global). A subsequent nationwide study investigated the community-based prevalence and longer-term mortality risk (IMPACT Revisited). A survey-based ecological study involving 185 hospitals in 14 countries evaluated trauma service disruption (IMPACT Services Survey). A propensity score-matched study investigated non-lethal effects of COVID-19 (IMPACT Frailty). A nationwide population-level study evaluated vaccine efficacy using four government-managed healthcare databases and linked electronic health records of 13,345 hip fracture patients over two years (IMPACT Protect).

Results

Over 20,000 patients were included. COVID-19 prevalence was higher among hip fracture patients than the general population. Patients that were COVID-positive within 30 days of fracture had a higher 30-day and 365-day mortality rate (34.6% vs 9.0%, $p<0.001$, and 54.7% vs 27.2%, $p<0.001$). Older age, male sex, renal disease, and pulmonary disease were independently associated with higher mortality risk. COVID-19 diagnosed after discharge was not associated with an increased 365-day mortality risk. COVID-19 was independently associated with a greater increase in post-discharge frailty, and a four-fold increased risk of home-dwelling patients failing to return home. Unvaccinated hip fracture patients were more than twice as likely to be COVID-positive and had an almost three-fold increased 30-day mortality risk, but the additional mortality risk conferred by COVID-19 was negated if patients had been vaccinated prior to infection.

Conclusions

COVID-19 hip fracture patients were three-times more likely to die within 30 days, and twice as likely to die within a year, than COVID-negative patients. However patients diagnosed with COVID-19 after discharge from the hip fracture admission did not experience an increased mortality risk, suggesting that COVID-19 and hip fracture created a “double-hit” effect on vulnerable patients. COVID-19 exerted non-lethal affects including a greater increase in frailty and a greater likelihood that previously independent patients would require ongoing inpatient or residential care. Vaccination was effective at reducing the likelihood of contracting COVID-19, and negated the increased mortality risk conferred by COVID-19 IMPACT involved

>500 collaborators from 190 centres in 19 countries, and analysed data for >20,000 patients. It was unique in its use of population-level data-driven investigations of COVID-19 in hip fracture, guided clinical practice and national policy during the pandemic, and may guide preparation for future communicable disease outbreaks.

(CP) Integrating liquid biopsies into Rapid Cancer Diagnostic Services

Rishma Maini1,2,3, Neil Cruickshank2, Peter Donnelly1
1University of St Andrews, St Andrews, United Kingdom. 2NHS Fife, Fife, United Kingdom. 3Public Health Scotland, Glasgow, United Kingdom

Objectives
Rapid Cancer Diagnostic Services (RCDS) are being piloted in Scotland. They aim to be a one-stop shop to expedite the investigation of patients with vague symptoms that could be explained by cancer. Currently, referrals are made based on clinical judgement and most patients accepted into the service undergo a full body scan. However, more accurate triage tests are desirable to spare patients without cancer unnecessary invasive and costly investigations. Liquid biopsy tests are being developed which can detect different cancer types at an early stage, however the methods employed are known to vary widely.

We present an idea for research with the following objectives:

1. Investigating the feasibility and acceptability of integrating liquid biopsy tests to patients and healthcare staff into a RCDS; and

2. Comparing the performance of different liquid biopsy tests in correctly a) detecting patients with cancer, and identifying the cancer tissue of origin.

Method
In-depth semi-structured interviews will be undertaken with healthcare staff and patients attending the RCDS, balanced by age, sex and suspected cancer site. Questions will explore health workers' and patients' knowledge and understanding of liquid biopsy tests, as well as their thoughts around these tests being used to risk-stratify patients. Following coding of interviews, emergent themes will be identified using inductive thematic analysis.

Abstracts

Elevator Pitch & Poster Presentations

A prospective observational study which compares the performance of different liquid biopsy tests against the diagnostic outcomes of patients accepted onto the RCDS pathway will then be undertaken. The study design will enable estimates of sensitivity and specificity for different liquid biopsies to be derived and compared with one another.

Results

Thus far, we have identified several different companies whose work on liquid biopsies using different approaches is yielding excellent preliminary results. We also have established links with the RCDS in NHS Fife and NHS Lanarkshire. It is hoped that such a pump-priming study could lead to further larger scale and higher-powered trials to assess the effectiveness of a broad range of liquid biopsy tests.

Conclusions

If their effectiveness in early diagnosis of cancers can be established, liquid biopsies could be more affordable than most other diagnostic tests such as CT scans. They may also be more easily expanded to vulnerable and hard-to-reach populations, which is of import given the social gradient of cancer inequalities. In addition, by introducing these tests into an already established pathway, this presents a responsible mechanism for integrating cancer biomarkers which is unlikely to stretch demand on an already overwhelmed health service.

(PP) Birmingham Self-test Report 1: Review of rapid self-test diagnostics sold in supermarkets and chemists in the UK: their descriptions, manufacturers, distributors, and regulators.

Simon Baldwin, Bethany Hillier, Katie Scandrett, Ridhi Agarwal, Aditya Kale, Joseph Alderman, Trystan Macdonald, Alex Richter, Clare Davenport, Jon Deeks
University of Birmingham, Birmingham, United Kingdom

Objectives

To provide an indication of the scope of self-testing diagnostic kits that have recently emerged and are available for purchase from UK shops. This abstract represents part of a wider body of research into the claims, evidence, and potential harms from using self-tests for diagnosis of multiple health conditions. Here, we summarise the high-street outlets where we identified self-tests being sold, the test manufacturers, distributors and the Notified Bodies who approved their sale.

Abstracts

Elevator Pitch & Poster Presentations

Method

The sampling frame comprised of supermarkets, community pharmacies, and health and wellbeing stores. A shopping radius of 10-miles distance from the University of Birmingham Edgbaston Campus covered multiple (>10) high street outlets, spanning the key retail areas in the metropolitan boroughs of Birmingham, Dudley, Sandwell, Solihull, and Walsall – collectively 4.4% of the population of England and Wales.

We focused on self-tests where the sample is taken, tested, and results interpreted by the user. Pregnancy and ovulation tests, tests for detecting drug misuse, and test strips used in conjunction with digital monitor devices were excluded. A single example of each test kit was obtained during April 2023.

The types of tests, sample types, origins and their regulations were reported on. This information was based on the test packaging, their Instruction for Use documents (IFU) and Patient Information Sheets (PIS), as well as by inspection of the devices.

Results

We identified 35 different tests sold in the sampling frame, 30 of which were obtained (three were duplicates, two out of stock). The tests used seven different sample types: faecal, finger-prick blood, urine, semen, vaginal swab, nasal swabs and a throat swab. Fifteen of the tests were from five Chinese manufacturers; nine were from two manufacturers in Austria or France; and the remaining six were from six different companies (two in the US, the others in Australia, Israel, Denmark and France). Tests were found in nine different supermarket chains, five community pharmacies, and one health and wellbeing store. Ten of the tests were distributed by SELFCheck (Superdrug), 9 by Newfoundland (Tesco) and 3 from SureSign (ASDA). The tests were all CE IVD marked as self-tests by five different Notified Bodies; 24 of the 30 tests were CE marked by two German Notified Bodies. None of the tests were UKCA marked.

Conclusions

New self-test products have become available in multiple high-street shops, supermarkets and chemist outlets. This increases the public's access to tests, which can offer privacy and confidentiality, and potentially reduce the need for healthcare visits. Half of the tests were manufactured in China, none in the UK. The same tests may be sold by multiple distributors. The regulatory decisions on the

claims and suitability of 24 of the tests have been approved by two German Notified Bodies: TÜV SÜD 0123 and MDC MEDICAL DEVICE CERTIFICATION 0483. It is important to assess whether these new tests are suitable and safe for use by members of the public here in the UK.

(PP) Birmingham Self-test Report 2: Are the statements of intended use and indicated medical actions for rapid self-test diagnostics sold in supermarkets and chemists appropriate?

Ridhi Agarwal, Katie Scandrett, Bethany Hillier, Simon Baldwin, Aditya Kale, Joseph Alderman, Trystan Macdonald, Alex Richter, Clare Davenport, Jon Deeks
University of Birmingham, Birmingham, United Kingdom

Objectives
To assess whether Information for Use (IFU) and Patient Information Sheet (PIS) documents for self-tests identify who and when the test should be used, the conditions the test detects and diagnoses, and indicate recommended actions including obtaining medical help, treatment and further tests.

Method
This poster is based on the sample of 30 self-tests described in our accompanying poster on rapid self-test diagnostics sold in supermarkets and chemists in the UK.

We identified Intended Use Statements included in the IFU and PIS for each test and assessed whether they state:

1) the medical purpose for which the test should be used (specifically whether it should be used for screening or diagnosis),
2) the situation in which a person should use the test (specifically the symptoms that they are experiencing or the risk factor or exposure about which they are concerned),
3) the clinical condition that the test will detect or diagnose,
4) the consequent medical or healthcare actions that are indicated, including obtaining medical help, treatments, or further tests.

Results
The Intended Use was stated in 15 tests in the "Intended Use" or "General Information sections": 7 for diagnosis, 2 for screening, 6 for both diagnosis and

screening, and 1 also claimed to have a curative effect. Intended Use statements were implied in 14 of the other tests. 29 stated the name of the biomarker, and 23 stated the positivity threshold. Only eight stated the target conditions, four did not mention them at all, and the rest were vague. Medical professional help was indicated when results were positive in 27 tests, 16 of which also indicated the use of medical professional help when results were negative. Two indicated treatments should be chosen following test results, whereas five indicated that rules and protective measures should be followed if positive. Recommendations for further testing was stated when tests were positive for 6, negative for 5, and regardless of the results for 13.

Conclusions

Statements on the medical role and the situation in which a person should use a self-test for diagnosis and screening were poorly stated in half of the tests. Whilst the name of the biomarker was routinely reported, the target condition that the self-tests aimed to detect is rarely stated. More than half of the tests indicated that professional medical help is needed regardless of the test results, which questions the value of using these tests at all. Similarly close to half of the tests indicate that further testing is needed regardless of results. Few tests recommend treatment or preventative actions based on results.

(PP) Birmingham Self-test Report 3: What is the evidence base for claims of accuracy for rapid self-test diagnostics sold in UK retail settings?

Bethany Hillier, Simon Baldwin, Katie Scandrett, Ridhi Agarwal, Aditya Kale, Joseph Alderman, Trystan Macdonald, Alex Richter, Clare Davenport, Jon Deeks

University of Birmingham, Birmingham, United Kingdom

Objectives

To assess the accuracy claims of self-tests, detailed on their packaging, Information for Use (IFU) and Patient Information Sheet (PIS) documents, and supporting evidence from manufacturers' clinical study and layperson study reports.

Abstracts

Elevator Pitch & Poster Presentations

Method

This poster is based on the sample of 30 self-tests described in our accompanying poster on rapid self-test diagnostics sold in supermarkets and chemists in the UK.

We identified performance claims in the IFU and PIS documents and packaging.

For each test, we requested the following documentation from the distributor and manufacturer: clinical (accuracy) performance studies which were used to obtain the CE IVD marking for the test by the Notified Body (clinical study reports); and studies that show the ability of these tests to be appropriately used by laypersons (layperson study reports).

Requests for reports and data were sent directly to all email addresses found in the packaging, the IFU documents, and websites of the manufacturers and distributors. The lead researcher sent requests twice, and BMJ members also contacted non-responders.

Results

Accuracy claims were made in IFUs of 25 tests: 17 for sensitivity, 16 for specificity, and 22 for accuracy. Performance ≥98% was claimed for over half of the accuracy claims (59%, 13/22), specificity claims (56%, 9/16) and 41% (7/17) of the sensitivity claims. No statements were made about prevalence and predictive values. The reference standard was stated in 17 tests: 5 used a rapid test, 4 used PCR, and 7 used a clinical laboratory method. The nature of samples or participants was stated in 12 tests, but 8/12 did not indicate whether samples were from unique participants. Clinical and layperson study reports were obtained for 12 tests, providing 9 unique reports. Requests were refused for 6 tests, and no response was received for 12. Little detail was given on the samples, participants and reference standards. Where described, the participants were often not representative of the intended population.

Conclusions

The scientific and statistical claims on self-tests are potentially misleading to the public. It is important for public health to ensure that statistical standards are implemented when evaluating tests, and that manufacturers and regulators follow guidelines for conducting and reporting clinical and lay-person studies. Issues arising from a lack of transparency and poor reporting, such as suspicion of selective patient inclusion, negatively affect the credibility and reliability of self-tests.

Abstracts

Elevator Pitch & Poster Presentations

(PP) Birmingham Self-test Report 4: Are the equipment, sampling, and instructions of rapid self-test diagnostics sold in UK retail settings correct, usable and safe, and are documents readable and in line with national and international guidelines?

Jon Deeks, Clare Davenport, Alex Richter, Aditya Kale, Joseph Alderman, Trystan Macdonald, Bethany Hillier, Katie Scandrett, Ridhi Agarwal, Simon Baldwin

University of Birmingham, Birmingham, United Kingdom

Objectives
To assess whether the equipment, the sampling method and the Instructions for Use (IFU) in rapid self-test diagnostics sold in UK retail settings are free of error, readable by members of the public, and in accordance with National and International Guidelines.

Method
This poster is based on the sample of 30 self-tests described in our accompanying poster on rapid self-test diagnostics sold in supermarkets and chemists in the UK.

We assessed the equipment, sampling process and instructions in a workshop of test experts, statisticians, clinicians and a test manufacturer. We categorised items that could cause test errors and practical challenges according to the probability of an error occurring and its potential impact.

The accessibility of the documents was measured using the font size, the Flesch Reading Ease system, and the Flesch-Kincaid Grade. Comparative control documents were the first 1000-1500 words of the Highway Green Cross Code and the first and final books from the Harry Potter series. We identified guidelines for each test in NICE, the NSC and the WHO that match the conditions most related to the intended use claims for the tests (the 30 tests were matched to 19 conditions).

Results
17 tests had high-risk concerns: 11 in equipment, 10 in sampling, and 15 in instructions and interpretation. Equipment concerns included dipsticks with no orientation mark, no labelling of T and C marks, no sterile pot provided and pipette labelling errors. 8 tests had serious concerns in the sampling capillary finger prick process.

Inappropriate choice, description or poor use of thresholds and reference ranges was a high concern in 10 tests. 20 of the 30 tests had Flesch Reading Ease scores of 40-59, above the required reading level of the average 13 to 14-year-olds. Documents were harder to read than all control documents. The minimum font size of 9 points (as required for leaflets on drugs) was met for three tests; IFUs for 15 tests were printed in font sizes less than 7. The intended use was contrary to their use in guidelines for twelve of the 19 conditions (18 tests).

Conclusions

Over half of the self-tests we evaluated were identified as having high-risk concerns, which could cause test errors and harm to test users. Particular problems exist in tests that measure concentration (rather than identify infectious diseases) where there is no single threshold which is suitable for use. Laboratory tests account for differences such as age, condition (e.g. pregnancy) and sex through use of reference ranges. Similarly, screening tests may alter the positivity threshold to balance benefit and harm due to differences in prevalence. The use of a fixed lateral flow test in these situations can lead to wrongly misclassifying individuals as positive or negative. The documentation of many tests is not fit for purpose, and tests are contrary to equivalent near-patient and laboratory national and international guidelines. It is unclear how many of these tests have been approved by regulators, as many have issues which could cause harm.

(PR) Guidelines and regulations applicable to vitro diagnostic tests in the UK: a scoping review

Oscar Khawar1, Magdalena Staworko1, Frank Sullivan1, Peter D Donelly1, Jon Deeks2, Margaret McCartney1
1University of St Andrews, St Andrews, United Kingdom. 2Univeristy of Birmingham, Birmingham, United Kingdom

Objectives

To describe recommendations applicable to new diagnostic and screening tests brought to market in the United Kingdom as of 01/06/23; and extract agreements, disagreements and gaps.

Abstracts

Elevator Pitch & Poster Presentations

Method

PubMed, Web of Science Core Collection and Scopus; grey literature via EuropePMC and Google, government regulations and guidelines, and relevant professional societies were searched for relevant sources using the following criteria: Extant regulations, recommendations and guidelines for new diagnostic and screening tests applicable to new products placed in the UK market as of 01/06/2023. Non-English and references not applicable to new tests seeking market access in the UK on 01/06/2023 were excluded. References of relevant included data were scanned for includable articles. Resultant data was thematically analysed and presented as a narrative scoping review.

Results

943 items were initially identified with 892 excluded. Reference searching located a further 31 papers and 82 items were analysed. Seven themes were identified: regulation, companion diagnostics and lab developed tests, safety and evidence, test specific recommendations, data, innovation, and recommendations for patients/the public. Wide agreement included the need to reduce bureaucracy and duplication; to mitigate to avoid unintended consequences of IVDR. Disagreement over whether high quality evidence should precede regulatory approval, or could be gathered as part of post marketting surveillance emerged.

Conclusions

Industry, regulators, academics, patients representing a variety of views, should collaborate to work through areas of disagreement.

Abstracts

Oral Presentations

(T) Creating evidence-based optimal testing strategies for monitoring long-term conditions in primary care

Martha Elwenspoek, Katie Charlwood, Lewis Buss, Rachel O'Donnell, Benedita Deslandes, Tom Harding, Mary Ward, Howard Thom, Alice Malpass, Jonathan Banks, Clare Thomas, Hayley Jones, Jonathan Sterne, Alastair Hay, Jessica Watson, Penny Whiting
Population Health Sciences, University of Bristol, Bristol, United Kingdom

Objectives
Patients with long term conditions (LTC) have regular monitoring appointments including blood tests. These tests aim to monitor disease progression, response to treatment, and detection of complications. The evidence base for current testing recommendations is weak because measuring patient benefits or harms of regular monitoring is challenging and are dependent on what is done in response to the test result. The aim of the Optimal Testing project is to develop a methodology for creating evidence-based testing strategies to monitor people with LTCs and accompanying patient and clinician materials.

Method
We identified a list of commonly used tests. We defined a series of filtering questions to determine whether there was evidence to support the rational of monitoring, such as 'Is this patient population at greater risk than the general population to get a complication which can be picked up by a candidate test', and 'can the GP do anything in response to an abnormal test result?'. Through a series of rapid reviews we identified evidence to answer each question. The evidence was presented at a consensus meeting where clinicians and patients voted for inclusion, exclusion, or further analysis. A process evaluation was performed alongside this. Further analyses are performed using routinely collected healthcare data, by performing incidence analyses and emulating RCTs. The optimal frequency of monitoring will be determined by disease progression modelling and health economic modelling.

Results

We tested this methodology on three common LTCs chronic kidney disease (CKD), type 2 diabetes mellitus (T2DM), and hypertension. We found sufficient evidence to include HbA1c and eGFR for monitoring T2DM patients; haemoglobin and eGFR for CKD; and eGFR for hypertension. The consensus panel voted to exclude haematinics, clotting tests, brain natriuretic peptide, c-reactive protein, and erythrocyte sedimentation rate. However, for the majority of tests, there was not enough evidence to include or exclude from the testing panel, which were selected for further analyses. Incidence analyses suggested patients with CKD and T2DM do not have a higher risk of abnormal thyroid function or haemoglobin, respectively, than matched controls and monitoring is therefore not necessary. The emulated RCTs aim to investigate the effect of regular testing with certain tests on patient outcomes among routinely monitored patients (ongoing). The final testing panel will be decided in an upcoming consensus meeting.

Conclusions

Many tests that are currently done in primary care may not have a strong rationale and may not lead to any patient benefit. The methodology we have developed may be used to optimise disease monitoring of other chronic conditions. We currently developing an intervention package and are planning to run a feasibility trial to test the cost-effectiveness of the evidence-based testing panels. This programme of work has the potential to change how LTCs are monitored in primary care, ultimately improving patient outcomes, and leading to more efficient use of NHS resources.

Abstracts

Oral Presentations

(T) How should we include ethnicity in prediction models that guide clinical decision making? A multidisciplinary reflection on potential and risk

Tanvi Rai1, Sharon Dixon1, Ashley Clift2, Jacqui Gitau3, Ann Marie Raphael3, Diana Withrow1
1University of Oxford, Oxford, United Kingdom. 2Early Health, London, United Kingdom. 3African Families in the UK (AFiUK), Oxford, United Kingdom

Objectives
Risk prediction tools are predicated on the belief that they can improve health outcomes for all through directing limited resources effectively. There is wide awareness that race/ethnicity is a significant determinant of health and access to health care, and as a result the variable is present in many of the data sources that feed risk prediction models. However, the recording of ethnicity is problematic at both epistemic and operational levels. Risk prediction tools are being produced and implemented at a rapid pace, but we are concerned that if ethnicity data are included in presently available forms, existing ethnic health disparities may be exacerbated.

Method
In this presentation, we will use examples from our own research and from the literature to reflect on the risks from the use of incomplete and inconsistent ethnicity variables in risk prediction tools and their potential consequences. Specifically, we will outline the problem, the potential implications, and identify key questions we must address before we can move forward to build safe, equitable and effective prediction tools.

Results
Risk prediction tools have an inherent input problem. Ethnicity data, in particular, have a high degree of missingness, are collected and coded heterogeneously across health settings, and change over time, and based on who records it. Furthermore, ethnicity is an ill-defined concept acting as a proxy for a complex mix of social, structural and systemic factors that apply differently across individuals and contexts. Given these operational and epistemic problems with the data, any tools that include ethnicity may also be suboptimal. In this imperfect data landscape we must pause to reflect on key questions such as: when is it appropriate to impute ethnicity data? Should we publish tools even if they suggest reducing access/action for ethically minoritized groups? Is voluntary reporting of the limitations of the data and tools enough? Should reporting be mandated? Could a standardized process be in place for evaluating potential for harm?

Conclusions
As scientists, clinicians, public health practitioners, patients, service users, advocates, and policy makers, we would like to spark discussion about what can we do to develop, evaluate and implement these tools responsibly. There are serious ethical and public health implications for using tools that risk exacerbating existing ethnic health inequalities. Until we address these concerns, and come to a workable consensus, we are at risk of thwarting our best efforts to improve health equitably.

(T) What is the impact of regular monitoring with specific blood tests in people with long term conditions on patient outcomes? Trial emulation using routinely collected primary care data.

Katie Charlwood, Martha Elwenspoek, Jessica Watson, Jonathan Sterne, Penny Whiting
University of Bristol, Bristol, United Kingdom

Objectives
It is generally accepted that people with long term conditions benefit from regular monitoring after diagnosis. However, the evidence base for the optimal monitoring strategies, including which test should be used at what frequency, is weak. Current practice is largely based on expert opinion and local protocols vary, which has led to substantial variation in blood test use within the UK. We aim to investigate whether regular monitoring in people that have recently been recently diagnosed with type 2 diabetes mellitus (T2DM), hypertension, or chronic kidney disease with certain blood tests impacts health outcomes using routinely collected primary care data.

Method
We are developing analyses to emulate a target trial using primary care electronic health records from Clinical Practice Research Datalink (CPRD) and Hospital Episode Statistics (HES). We are using a sequential trial approach to estimate the effects of regular testing with commonly used blood tests (including liver function tests, renal function tests, and lipid profile) on patient outcomes. We will compare patients who have received regular testing with the candidate test to patients who have not received these tests. The primary outcomes are events that could be

prevented with regular monitoring such as unplanned hospital admissions. We will censor patients when they deviate from their assigned strategy, pooling data from the trials to use pooled logistic regression to calculate outcome cumulative incidence and risk difference. Time-varying confounding will be accounted for by applying time updating inverse probability weights.

Results

We are developing the analysis to evaluate the effects of monitoring liver function in patients with newly diagnosed T2DM. Patients were eligible if they had a T2DM diagnosis and HbA1c record within 30 days of diagnosis between 2004 and 2019, were not pregnant during the study period, and had no history of liver disease. 47,344 patients were eligible for recruitment, and were recruited on the date of their first HbA1c test 12 weeks after diagnosis . Patients were assigned to the testing strategy compatible with their data on this date. 28,993 patients had liver function testing on this date and were assigned to the intervention group, and 18,351 people did not have liver function testing and were assigned to the control group. Eighty percent of people in the control group and 56% of the intervention group switched monitoring strategy during follow-up and were censored.

Conclusions

We aim to apply these methods to other test and condition combinations once finalised, and use these findings to decide whether to recommend regular monitoring with certain blood tests in patients with T2DM, hypertension, or chronic kidney disease. Challenges developing these methods include accounting for residual confounding, high censoring rates, and limitations associated with routine data.

Abstracts

Oral Presentations

(T) Assessing the value of diagnostic tests: evaluation of a framework for identifying and organising test effects

Jacqueline Dinnes1, Clare Davenport1, Isobel Harris1, Lavinia Ferrante di Ruffano2, Chris Hyde3, Sue Mallett4, Jonathan Deeks1
1University of Birmingham, Birmingham, United Kingdom. 2University of York, York, United Kingdom. 3University of Exeter, Exeter, United Kingdom. 4University College London, London, United Kingdom

Objectives
The value of a diagnostic test goes beyond simple measures of its accuracy. Evaluating the impact of a test requires consideration of the clinical pathway in which the test will be used, identification of important ways in which the test might affect that pathway, and appropriate selection of outcomes that adequately assess whether the test's introduction will realise clinical benefit.

We aimed to evaluate how well our previously published test evaluation framework (TEF) captures the intended and unintended effects of diagnostic technologies typically evaluated in HTAs. Evidence reviews for HTAs of tests were analysed to examine how changing a test, or introducing a new test strategy, were considered to impact on patient health or healthcare delivery. Our objectives were to obtain empirical evidence of the frequency and importance of different test effects, including both intended and unintended effects, and to identify patterns in non-accuracy-based effects according to type of test.

Method
We collated a catalogue of HTAs of tests used for either diagnosis or staging of disease from seven HTA organisations, published in English from 2010-2020. Key characteristics about technology types, test comparison type, target condition, clinical setting, and availability of evidence to answer the review question was abstracted from the evidence reviews for each HTA.

Two HTA organisations (NICE and MSAC) were identified as providing full evidence reviews specifically focused on test evaluation questions. Reviews from these organisations were purposively sampled to ensure a wide distribution of diagnostic

topics. For each review, we mapped out the patient management strategies being compared, used the TEF to identify all likely test effects, and any additional potential mechanisms that might impact on the test's (potential or realised) ability to effect downstream outcomes. All extractions were conducted by one reviewer, checked by a second and discussed at roundtable project meetings.

Results

We included 45 reviews reporting 50 review questions of which 38 (75%) compared an index and comparator strategy using the same type of diagnostic technology (e.g. IVD versus IVD). The clinical claim for the test was reported in a dedicated section in 28 reviews (56%). All test effect mechanisms (from the TEF) in the claim for the test were translated to identifiable outcomes in 26 reviews (52%); 82% specified additional outcomes unrelated to the claim for the test. Reviews most often (49, 98%) considered at least one mechanism related to impact (accuracy in 98%; therapeutic yield, 84%; and treatment effectiveness, 94%), or to feasibility and interpretation (40, 80%) (acceptability, 42% and/or test failure rates, 48%). Timing mechanisms were considered by 35 (70%) reviews (time to produce a result, 36%; speed of diagnosis, 46%; time to treatment, 46%), and were most often considered in evaluations of IVDs or endoscopic tests.

Conclusions

The TEF provides a useful tool for elicitation of intended and unintended effects of introducing a new test or changing a testing strategy. Although HTAs were relatively comprehensive in identifying important test effects in the claim for the test, these were frequently not translated to measurable outcomes. Mechanisms such as accuracy or treatment effectiveness, appear to be included as 'standard' outcomes to be measured even when not identified as an important mechanism of effect. Conversely, mechanisms arguably relevant to the majority of test evaluations (e.g. test failure rates) were identified by less than half of reviews (48%). Mechanisms related to decisional impacts (clinician or patient confidence in the diagnosis or treatment) were identified in a minority of reviews (<20%), but were considered more often in reviews of IVD-based strategies compared to other types of tests. These findings suggest HTA's may miss evaluating key aspects of a diagnostic's potential value.

Abstracts

Oral Presentations

(T) Intermediate endpoints as sufficient surrogates for cancer-specific mortality in cancer screening trials: A systematic review and meta-analysis

Julia Geppert1, Nefeli Kouppa2, Matejka Rebolj2, Adam Brentnall3, Bethany Shinkins4, Chris Stinton4, Karoline Freeman1, Sian Taylor-Phillips1
1Universty of Warwick, Coventry, United Kingdom. 2Kings College London, London, United Kingdom. 3Queen Mary University of London, London, United Kingdom. 4University of Warwick, Coventry, United Kingdom

Objectives
To investigate whether novel cancer screening can be implemented earlier based on surrogate outcomes rather than waiting for follow up to mortality outcomes. This requires evidence of 'sufficiency' of the surrogate. Using all eligible historical randomised controlled trials (RCTs) of screening, we investigated the relationship between the effect of the screening interventions (compared to no screening or different screening) on potential surrogate outcomes and on cancer-specific mortality (overall, by cancer site, by modality, by site-modality combination). The intermediate endpoints evaluated include absolute incidence of late-stage cancer, proportion of cancers detected at a late stage, absolute incidence of early-stage cancer, proportion of cancers that are screen-detected, proportion of aggressive cancers that are screen-detected and diagnostic yield at screening.

Method
Using a systematic review approach, we searched the literature in two stages to identify all cancer screening RCTs reporting mortality outcomes, and subsequently, for each trial identified, to find all publications that reported relevant intermediate outcomes or further mortality outcomes. We used the "main" mortality timepoint of each trial (as defined in study protocol, statistical analysis plan or similar) as well as the "best" available intermediate endpoint (i.e. early-stage and late-stage as defined by the trialists in the original publications; preferably measured closest to the midway point between the end of the intervention period and the "main" mortality timepoint). If applicable, we obtained the estimated rate ratio or relative risk (RR) and 95% confidence intervals (CIs) between the intervention and control groups. Using a weighted fixed-effects linear model, we evaluated the association between the intermediate endpoint (logRR or proportion) and log(RR mortality). We performed sensitivity and subgroup analyses.

Results

In 57 RCTs (14 lung/13 breast/11 bowel/5 prostate/4 cervical/3 ovarian/2 liver/5 other), the correlation between the RR for late-stage cancer incidence and the RR for cancer-specific mortality was 0.69 (95%CI 0.47-0.84; R2=0.47). This varied between cancer types, being 0.58 (95%CI 0.27-0.93; R2=0.34) for bowel, 0.79 (95%CI 0.49-0.94; R2=0.62) for breast, and 0.91 (95%CI 0.84-0.96; R2=0.83) for lung. In 95% (n=18/19) trials with late-stage cancer incidence reported at a time point before mortality, the 95% CIs of the treatment effect on late-stage cancer incidence accurately encompassed the final treatment effect on cancer-specific mortality. However, only 11% (n=2/19) of trials could have been stopped early with a correct prediction of benefit for mortality from the treatment effect on late-stage cancer incidence. In the remaining trials there was either no benefit or uninformatively wide CIs. For the other potential surrogates there was moderate to strong evidence that they would be a poor surrogate.

Conclusions

This is the largest and most comprehensive meta-analysis of cancer screening surrogate endpoints to date. The observed correlation between advanced cancer incidence and cancer-specific mortality within and between cancer types and cancer screening tests suggests it may be considered for use as a surrogate endpoint, but with a range of caveats and further research requirements. Direct evidence was strongest for lung, breast and colorectal cancer screening. The evidence for generalising from one cancer to another, for example in MCEDs, is less well developed. The proportion of cancers diagnosed at late stage is an inferior surrogate endpoint, and we do not recommend it. The other evaluated intermediate outcomes have no potential as surrogates for sufficiency due to being impacted by overdiagnosis. Modelling might be used to predict cancer-specific mortality from a trial by using the observed difference in late-stage incidence with other evidence, such as on treatment adherence and stage-specific survival.

Abstracts

Oral Presentations

(T) Interpreting diagnostic accuracy studies based on retrospective routinely collected data

Stephen Bradley1, Gary Abel2, Bethany Shinkins3, Matthew Callister4
1University of Leeds, Leeds, United Kingdom. 2University of Exeter, Exeter, United Kingdom. 3University of Warwick, Coventry, United Kingdom. 4Leeds Teaching Hospitals Trust, Leeds, United Kingdom

Objectives

Observational evidence from routinely collected data is frequently used to generate estimates of diagnostic test accuracy. Such studies offer advantages of convenience and retrospective, rather than prospective, identification of positive cases helps ensure estimates of test performance can be undertaken with satisfactory statistical power. Prospective interventional studies with recruitment of patients to undergo specific tests are costly, time-consuming and may not be ethically tenable. Therefore prospective studies, particularly in the symptomatic context are rarely performed. However, diagnostic accuracy studies using observational data often make little reference to important assumptions involved in estimating test performance using such methods.

This presentation will:

1) Explain with examples the underlying assumptions and potential for bias in using observational evidence to estimate diagnostic test performance.
2) Demonstrate that using data from screening studies which evaluate two or more diagnostic tests can be used to generate additional evidence on comparative test performance

Method

In such studies clinical diagnosis within a specified period of test is often used as a reference standard. This requires two assumptions:

1) If undetected by the test, the target condition would subsequently be diagnosed within the specified time period i.e. the disease would progress and manifest with persisting or worsening symptoms.

2) That the target condition was present when the test was performed and did not arise de novo within the specified time period..

Violation of the first assumption will lead to over-estimation of diagnostic accuracy. This is because a negative test result will be incorrectly labelled as true negative if the disease is present but not diagnosed within the specified period.

Violation of the second assumption will lead to under-estimation of diagnostic accuracy. This is because a negative test result will be incorrectly labelled as false negative because the disease has been diagnosed during the specified time period.

Results

Examination of data from the National Lung Cancer Screening Trial (NLST) shows just how far estimates of test accuracy may be distorted by such biases. NLST reported sensitivity and specificity of both chest x-ray (CXR) and computed tomography (CT) in asymptomatic populations. The prevalence of lung cancer in targeted asymptomatic populations is comparable to symptomatic populations who are investigated with chest x-ray (0.7% versus 1.1%). NLST used clinical diagnosis within one year as reference standard. CXR sensitivity in NLST was reported as 75.4% (95%CI 67.5-83.3), similar to that demonstrated elsewhere for symptomatic patients (73.5%, 95%CI 67.2-79.8). Since NLST randomised patients to CXR or CT we can compare numbers of cancers detected in both arms to estimate CXR sensitivity (CT arm as a reference standard) with bootstrapping to determine confidence intervals. This leads to a sensitivity estimate of 50.4% (95%CI 40.0-60.8)-considerably lower than the 73.5% reported in retrospective observational research.

Conclusions

Where tests are already used in clinical practice, analysis of retrospective routinely collected data can help evaluate diagnostic accuracy in settings where disease prevalence is low and prospective studies would be expensive and require large numbers of participants. This type of data is also likely to be increasingly used for post-market evaluations of test performance to meet new regulatory requirements. But the limitations of this type of data, particularly the assumptions required to use clinical diagnosis within a specified time period as the reference standard, needs to be appreciated. The case for prospective diagnostic accuracy studies remains

strong, despite logistical and governance barriers. We recommend that reporting of diagnostic accuracy studies should make explicit reference to the limitations of observational data and sensitivity analyses (i.e. varying the length time period for a clinical diagnosis) are conducted to explore the potential impact that this decision is having on diagnostic accuracy estimates.

(T) Real world implementation of the Biomarker Toolkit: a Tool aiming to quantifiably assess biomarker utility and guide development

Katerina-Vanessa Savva1, Alice Baggaley1, Silvana Debernardi2, Tatjana Crnogorac-Jurcevic2, Melody Ni Zhifang1, George B. Hannna1, Christopher J. Peters1
1Imperial College London, london, United Kingdom. 2Barts Cancer Institute, london, United Kingdom

Objectives
Increased resources have been spent on cancer biomarker discovery, for both prognostic and diagnostic purposes, but very few of these biomarkers have been clinically adopted. To bridge the gap between biomarker discovery and clinical use, we have previously developed and validated the Biomarker toolkit (Savva et al., 2023). This tool aims to assess biomarker potential and then, more importantly, guide its further development. In this study we aim to apply the Biomarker toolkit to early-phase cancer biomarkers through collaboration with CRUK Horizons and the Pancreatic Cancer Group (PCG), at Barts Cancer Institute. The toolkit output will assist in identifying research gaps and guide subsequent research trajectory of the biomarker candidates. Simultaneously, we aim to evaluate the real-world impact and usability of our tool.

Method
Our tool was developed using mixed methodology, including systematic literature search, semi-structured interviews and a two-stage Delphi-Survey with biomarker experts (i.e.,scientists/clinicians/industry). Validation of the checklist was achieved by independent systematic literature searches using keywords/subheadings related to successfully clinically implemented and stalled breast and colorectal cancer biomarkers. Aggregated scores were generated for each selected publication based

on the presence/absence of a characteristic listed in the toolkit. Aggregated scores reflect i)biomarker's clinical implement ability and ii)provide useful guidance for end-users. During real-world implementation we aim to apply the toolkit in collaborator's developing biomarkers, at different development stages. Specifically, we are working with three biomarker groups within the CRUK Horizon portfolio, and PCG. We aim to gather all published/unpublished data regarding a specific biomarker, apply the toolkit and score biomarker of interest. The utility of the toolkit will be mainly evaluated using stakeholders semi-structured interviews.

Results

The PCG developed a urine-based biomarker panel, for pancreatic cancer early detection. Initially, we identified relevant literature through systematic literature searches (databases:Medline&Embase) and internal reports. We then applied the Biomarker toolkit as described in Savva et al.(2023). Our analysis revealed that the biomarker scored 40.25% for clinical-validity ((Reference-Scores:Succesful-biomarkers:41.51%(STDEV:±1.08)/Stalled-biomarkers:36.47%(STDEV:±4.2)), 49.35% for analytical-validity (Reference-Scores:Succesful-biomarkers:49.35%(STDEV:±4.3)/Stalled-biomarkers:46.42%(STDEV:±5.25)), and 9.62% for clinical-utility (Reference-Scores:Succesful-biomarkers:54.16%(STDEV:±10.7)/Stalled-biomarkers:15.82%(STDEV:±1.4)). Upon comparing these scores with reference standards for successful and stalled biomarkers, we identified several gaps that needed to be addressed Key recommendations included conducting Human Factor and budgetary impact analysis. The PCG said that "the report provided was highly valuable in identifying areas of strength and areas needing improvement to enhance clinical utility". This feedback guided their collaboration with the London In-Vitro Diagnostic Co-operative to address identified research gaps. CRUK Horizon results are underway.

Conclusions

This novel study applied the Biomarker toolkit, a tool used to mediate the successful translation of Biomarkers from lab to clinic in real world early detection biomarkers. This theoretical framework provided by our tool identifies gaps in research in early detection biomarkers, at any stage in the biomarker pipeline. This would undoubtedly direct research trajectory toward clinical utility. The toolkit could be used i) to detect biomarkers with the highest potential of being clinically

implemented, ii) to shape how biomarkers studies are designed/performed and iii) provide a framework at an early stage to inform researcher towards more impactful research. This will naturally support the translation of diagnostic biomarkers; thus, reducing costs associated with excessive discovery while promoting patient early diagnosis.

CP) Polygenic risk stratification for breast, colorectal and prostate cancer screening in the UK: integration of multiple national routinely collected cancer datasets for modelling of potential impact on cancer-specific mortality

Catherine Huntley1, Bethany Torr1, Richard Houlston1, Michael Jones1, Amit Sud1, Aroon Hingorani2, Clare Turnbull1
1Institute of Cancer Research, Sutton, United Kingdom. 2University College London, London, United Kingdom

Objectives
It has been proposed that stratification using polygenic risk scores (PRS) could increase efficiency of cancer screening, enabling extension into new age-groups. We integrated multiple national routinely-collected National Disease Registration Service (NDRS) cancer datasets to model the impact of introducing hypothetical new cancer screening programmes for age ranges that currently fall outside of UK national screening programmes: women aged 40-49 (breast), men aged 60-69 (prostate), and the population aged 50-59 (colorectal). We considered screening the PRS-defined high-risk quintile (20%), versus screening the oldest quintile, versus screening the full population in that age-group.

Method
We built our model using published PRS metrics for breast, prostate, and colorectal cancers (AUC=0.64, 0.70, 0.62), population size estimates from the Office for National Statistics (ONS), and cancer incidence, ten-year cancer-specific survival and routes-to-diagnosis data from the National Cancer Registration Dataset (NCRD). We estimated numbers of screen-detected cancers in each hypothetical scenario and reassigned route-, age- and stage-specific survival on this basis. The model utilised multiple favourable assumptions which would maximise survival estimates (e.g., full uptake of screening, full Western European ancestry and no interval cancers).

Results
The PRS-defined high-risk quintile and oldest quintile respectively capture 37% and 29% of breast cancers in women aged 40-49; 46% and 28% of prostate cancers in men aged 60-69; and 34% and 28% of colorectal cancers in the population aged 50-59. Annual screening of the PRS-defined high-risk quintile / oldest quintile / full population of women aged 40-49 would identify 1,968 (26%) / 1,538 (20%) / 5,273 (70%) of the incident 7,533 annual breast cancers (using digital mammography: sensitivity 70%, specificity 92%), corresponding to improvement in survival for this age group of 1.4%, 1.1%, or 3.6%. For men aged 60-69, respectively 2,473 (15%) / 1,494 (9%) / 5,393 (32%) of the incident 16,853 annual prostate cancers would be detected (using PSA 3ng/mL: sensitivity 32%, specificity 85%), with improvement in survival of 0.9%, 0.6%, or 2.0%.

Conclusions
For men and women aged 50-59, respectively 1,192 (24%) / 982 (19%) / 3,536 (70%) of the incident 5,052 annual colorectal cancers would be detected (using FIT 25-50 ug/g: sensitivity 70%, specificity 95%), with improvement in survival of 3.7%, 3.1%, or 11.0%. Conclusion Even under multiple favourable assumptions, PRS-stratified screening offers marginal improvement in cancer-specific mortality (averting of deaths from cancers). However, these modelled gains will likely be substantially attenuated when we take into account incomplete population uptake of PRS profiling, incomplete population uptake of cancer screening, interval cancers, non-European ancestry, all-cause mortality, and other factors. Only with cluster-randomised trials with long term follow-up can we quantify the impact of these factors on cancer outcomes.

(CP) 5 year mortality in an Randomized Controlled Trial of an autoantibody biomarker for Lung cancer

Frank Sullivan
University of St Andrews, St Andrews, United Kingdom

Objectives
To report five year follow up data on lung cancer and all cause mortality in the Early Diagnosis of Lung Cancer Scotland (ECLS) trial.

Method

ECLS was a pragmatic randomized trial involving 12 208 high-risk participants recruited through general practice and community-based recruitment strategies in Scotland. Recruitment occurred between April 2013 and July 2016 with follow up undertaken 60 months after randomization for each participant: adults aged 50–75 considered at increased risk of developing lung cancer compared to the general population. Our earlier publication from this trial reported outcomes after two years showing a significant reduction in late stage presentation, with a hazard ratio for stage III/IV presentation of 0.64 (95% CI 0.41–0.99), but no significant difference in lung cancer or all-cause mortality at 2 years follow-up. This presentation will present five year follow up per protocol analysis on lung cancer and all-cause mortality.

Results

77 077 invitation letters were sent to people fulfilling the record search criteria from 166 general practices and 16 268 responded (21.1%). 12 241 were invited to an in-person screening appointment, and 12 208 were randomised and followed up. The recruitment rate of people identified as potential study participants from family practice records was 13.4%; and the recruitment rate from self-referral was 79.1%. Participant characteristics were balanced between the intervention and control groups . 28.5% of participants lived in the most deprived quintile in Scotland, the mean age at recruitment was 60.5 years (S.D. 6.58), and the mean pack years smoked was 38.2 (S.D. 18.58). The main findings are undergoing peer review by a major journal at present but will be available outside their embargo in time for the conference

Conclusions

Blood tests or other biomarkers could substantially reduce the number of people requiring imaging investigations depending upon where the cut-off for sensitivity and specificity is set. This may have globally significant implications for case finding and screening for lung cancer in people at high risk of the disease. Whether blood based biomarkers should be used as one method to reduce lung cancer mortality requires further elucidation.

Abstracts

Oral Presentations

(CP) Pragmatic and scalable diabetic retinopathy screening for lower resource settings: Binocular indirect ophthalmoscopy versus a retinal camera, including Artificial Intelligence (AI) interpretation in Indonesia

Obaid Kousha, Andrew Blaikie, John Ellis
University of St Andrews, St Andrews, United Kingdom

Objectives

Diabetic retinopathy (DR) is a major cause of preventable blindness affecting more than 100 million people globally with the majority living in low-to-middle-income countries (LMICs). Applying high-income countries' (HICs) camera-based DR screening to LMICs has failed to gain widespread traction due to several pragmatic strategy that is potentially more scalable: the use of a low cost binocular indirect ophthalmoscope (BIO) with a 14 dioptre (14D) condensing lens in screening for vision-threatening DR (VTDR). We compared the results of the BIO with an orthodox tabletop fundus camera.

Method

After piloting and sample size calculation, we recruited 152 participants (304 eyes) suffering from diabetes attending an endocrine clinic at Hasanuddin University Hospital, Makassar, Indonesia. Two graders independently graded the eyes with dilated pupils using the BIO and a 14D lens. Subsequently, 45-degree fovea centred fundus images were taken using a tabletop camera. Both graders subsequently discussed their BIO examination findings, reviewed the camera fundus images, and if necessary, re-examined the participant with the BIO or captured eccentric fundal images to agree on a final reference grade. All the images were additionally graded by an artificial intelligence (AI) algorithm looking for VTDR.

Results

The prevalence of VTDR was 35.8% based on clinical examination and fundal images assessed by two graders. Ungradable examination rates with the camera were 24.0%, with 14 VTDR eyes missed. Only 2% of eyes were ungradable with the BIO. BIO sensitivity and specificity for DR were 98.1% (95% confidence interval 94.6%-100%) and 97.9% (95.1%-100%) respectively for grader 1, and 88.9% (80.5%-97.3%) and 100% respectively for grader 2. Intergrader agreement Cohen's kappa was 0.87 (0.78–

0.95). BIO sensitivity and specificity for maculopathy were 88.8% (80.5%-97.2%) and 98.0% (96.0%-99.9%) respectively for grader 1, and 87.0% (78.1%-96.0%) and 97.5% (95.3%-99.7%) respectively for grader 2. Intergrader agreement Cohen's kappa was 0.96 (0.91–0.99). The AI algorithm had sensitivity of 97.6% (92.8%-89.4%) and specificity of 83.1% (74.4%-91.8%) for VTDR in the worse eye, when compared to human fundal image grading. However, this excluded the ungradable images and images with pathology outside of 45-degree fundal image.

Conclusions
Our results demonstrate superior performance in screening for and diagnosing VTDR using a BIO compared to a camera. Use of a BIO some important additional advantages: immediate synchronous grading, diagnosis of coincident eye disease, direct patient communication and potential same-day treatment. These benefits can reduce 'leakage' experienced in the orthodox camera-based HIC referral chain as well as more broadly strengthen eye care delivery. A health care professional delivered screening using a low cost BIO can consequently be seen to be a contextually relevant and scalable DR screening strategy in LMICs, overcoming many of the limitations of orthodox camera-based approaches. While AI image grading shows potential in HICs, high rates of ungradable images in LMIC settings, due to higher prevalence of cataract and corneal opacity, highlights the relevance of a BIO approach. In conclusion, BIO-based DR screening is a promising, economically viable and scalable solution in LMICs.

(CP) Developing A Risk Prediction Tools For Near Term Mortality In Patients Who Present To Unscheduled Care In Scotland

Sarah Mills, Colin McCowan
University of St Andrews, St Andrews, United Kingdom

Objectives
Unscheduled care services are at a critical juncture, grappling with surging patient demand and an overflow from in-hours NHS services that threatens sustainability. In Scotland, the system encompasses General Practice Out-of-Hours (GPOOH), Emergency Departments (ED), NHS24, and the Scottish Ambulance Service (SAS). The

last year of life in Scotland sees 75-90% of individuals utilizing unscheduled care services. Many of these people are not identified as being near death. We aim to develop and validate an NHS Unscheduled Care risk prediction tool that will identify which patients using unscheduled care are at high risk of death within the next 6 months. Using artificial intelligence, including machine learning models, can we identify who may be in their last year of life and predict their future use of unscheduled care?

Method

The study dataset will include all Unscheduled Care attendances in Scotland, by adults aged ≥65 years old, between 01/01/2017 and 31/12/2021. Patients will be identified by the unique-patient-identifying Community Health Index(CHI) number. CHI-linked demographic and clinical data from national datasets, including General Practice Out-of-Hours(GPOOH), Emergency Department(ED), Scottish Ambulance Service (SAS) and NHS24. Analysis will compare risk predictors in those who died within 6 months of Unscheduled Care attendance with those who did not. We will use machine learning to analyse their health data to develop a predictive model. Area Under the Receiver-Operating Characteristic (AUROC) calibration will be used to compare the performance of the risk prediction tool to other relevant indices. Reproducibility will be tested against an external validation dataset.

Data will be obtained, cleaned, anonymised, stored and analysed in the Electronic Data Research and Innovation Service (eDRIS) Trusted Research Environment. Analyses will be conducted in R.

Results

In the initial phase of our project, the feature engineering component will entail a systematic process encompassing multiple technical elements including: 1. Data Cleaning: • Handling Missing Values: imputation or deletion to deal with missing data. • Outlier Detection and Treatment 2. Data Transformation: • Scaling: Normalizing or standardizing feature values to bring them within a comparable range. • Encoding: Converting categorical data into a format that can be provided to machine learning algorithms 3. Feature Construction: • Feature Interaction: Creating new features based on interactions between existing features. • Polynomial Features: Generating new features by considering polynomial combinations 4. Time Series Data Handling: • Seasonal Decomposition: Separating time series data into components like trend,

seasonality, and noise. • Lag Features: Creating lag features to incorporate past observations 5. Geospatial Data Handling: Creating features based on geographical data. 6. Domain-Specific Feature Engineering: Creating features based on domain knowledge 7. Data Aggregation: Aggregating data to create summary features.

Conclusions

This work has recently been funded as part of a large programme grant from the Chief Scientist Office (CSO). Work will begin at the University of St Andrews in September 2024. This initial work seeks to develop the project strategy for feature engineering in order to create a model for risk prediction tools using NHS data. We anticipate that the eventual clinical risk prediction tools will facilitate objective, unbiased identification of people at risk of dying in the six months following unscheduled care use, which will enable optimisation of their medical management and access to anticipatory and palliative care planning.

(PP) Logics of Time and Diagnosis

Alexandra Brandt Ryborg Jønsson

Time is integral to how cancer research, policies, and prevention is practiced today. Despite conflicting evidence, the prevailing 'the sooner the better' approach remains unchallenged by research, policy or society. One explanation for this comes from the linear perception of time and societal traces of neoliberalism and acceleration in the Global North that affects societal and epidemic discourse on cancer as something that requires acute action. In this presentation, I discuss how different notions of time and linearity are essential in today's research ontology of cancer, describe the individual and societal consequences of such ontology, and invite a rethinking of time in cancer. Drawing on theoretical concepts of time together with cancer epidemiological, historical and ethnographical data, I analyse how the logic of early diagnosis has been established as a stable concept and make more people patients unnecessarily.

This and Professor John Brandt Brodersen presentation build on Damhus, C. S., Risør, M. B., Brodersen, J. B., & Jønsson, A. B. R. (2024). Rethinking the logic of early diagnosis in cancer. Health (London, England : 1997), 13634593241234481. Advance online publication. https://doi.org/10.1177/13634593241234481

Abstracts

Oral Presentations

(PR) Developing guidance for the evaluation, regulation, marketing, and monitoring of Direct to Consumer Tests (DTCTs)- 'GUIDE DTCTs'

Clare Davenport1, Steven Blackburn1, Aditya Kale1, Finlay MacKenzie2, Rachel Marrington2, Alex Richter1, Jessica Watson3, Jon Deeks1
1University of Birmingham, Birmingham, United Kingdom. 2University Hospitals Birmingham, Birmingham, United Kingdom. 3University of Bristol, Bristol, United Kingdom

Objectives
Direct To Consumer Tests (DTCTs) have the potential for significant positive impacts on individuals, populations and the health economy through facilitating early diagnosis whilst increasing autonomy and the acceptability of testing. There is evidence that current IVDD regulations for DTCTs are not robust, transparent, or fit for purpose. Further they do not cover marketing to consumers. This has the potential to result in misinformed testing choices by the public leading to unnecessary anxiety, delays in diagnosis and further strain on healthcare services.

GB are currently in a period of transition from EU to GB specific medical devices regulation. This creates an opportunity for cross sector collaboration to minimise harm and maximise the potential health benefits of DTCTs.

The objectives of this programme of research are to:

-Develop good practice guidance covering the evaluation, regulation, marketing and post market surveillance of DTCTs

-Identify priorities for further research

Method
A stakeholder group representing the medical and pharmaceutical professions, regulation, retailers, policy makers and the public will be convened to oversee:

-Scoping reviews and semi structured interviews with relevant stakeholders to map the regulation process, evaluation and marketing of DTCTs

-Scoping reviews to capture empirical literature, theoretical perspectives and commentaries concerned with the use and impact of DTCTs.

-Cross sectional studies to scrutinise the retail and online DTCT markets regarding fitness for (claimed) purpose and potential impact to consumers and the UK health care system.

-Focus groups with healthcare professionals, regulators, manufacturers, retailers and the public to capture perceived opportunities and challenges, and expectations and personal experiences of DTCTs.

-Consensus development of Good Practice Guidance for DTCT development, evaluation, regulation, marketing and post market surveillance and priorities for future research.

Results

A key output from this research will be the convening of a multi-disciplinary stakeholder group (DTCT special interest group) to progress the research and policy landscape of DTCTs. This research will provide critical evidence and guidance required to inform the development of new regulatory frameworks to avoid the potential harm and financial burden to the public and NHS of the current DTCT landscape. Our outputs will inform the development of standards for evidence requirements for risk classification of devices by regulatory bodies, standards for labelling, including reporting of test claims, instructions for use at the point of sale, interpretation of test results and any subsequent action to be taken including interaction with health services.

Abstracts

Oral Presentations

(PR) Global burden of disease estimates for Major Depressive Disorder: instruments used in studies to measure prevalence of MDD not designed for that purpose, contribute to risk of over-diagnosis and over-treatment.

Petra Sevcikova1, Lisa Cosgrove2, Gianna Gianna D'Ambrozio3, Farahdeba Herraawi2, Allyson Pollock4, Elia Abi Jaoude5
1Newcastle University, Newcastle, United Kingdom. 2Univ of Massachussets, Boston, USA. 3University of Massachusetts Boston, Boston, USA. 4newcastle Univ, Newcastle, United Kingdom. 5Univ of Toronto, Toronto, Canada

Objectives
Global Burden of Disease (GBD) estimates, produced by the Institute for Health Metrics and Evaluation (IHME) in collaboration with the World Health Organization, have significant policy implications nationally and internationally. In 2007, calls by the Lancet Global Mental Health Group to "scale up" services for people with mental disorders were followed by the launch of the Movement for Global Mental Health with the aim of "clos(ing) the treatment gap for people living with mental disorders worldwide". Recently, the WHO ranked depression as the single largest contributor to global disability. However, there is no standardized method for collecting prevalence data on mental disorders and global burden of disease estimates for depression have been shown to be highly unreliable especially at country level. The aim of this study is to examine the appropriateness of the instruments most commonly used to measure prevalence of depressive orders in studies underpinning GBD estimates.

Methods
The instruments used in the 566 country studies which underpin the (2019) GBD estimates for Major Depressive Disorders (MDD) worldwide were extracted and categorized. We then analysed the five most commonly used instruments with respect to their sensitivity, reliability and diagnostic application for measuring prevalence of MDD as reported in the literature and by the developers of the instruments.

Results
Of the 566 country studies, 98 (17.3%) used dedicated depression screeners, 356 (62.9%) used general mental health screeners or structured/semi-structured

interview guides, and 112 (19.8%) used other tools for assessing depression. Of the five most commonly used instruments, only the patient health questionnaire, full form (PHQ-9) and the Mini International Neuropsychiatric Interview (MINI) have been shown to meet minimum criteria for sensitivity and specificity for screening; these were used in 11% of the studies. The PHQ-9 was not designed to make a diagnosis of depression, and the MINI was designed to be overinclusive.

Conclusions
GBD estimates of prevalence of MDD are underpinned by a majority of studies which use screening instruments that are not designed to make a diagnosis of or assess MDD, and/or are known to overestimate prevalence of MDD. Our results are congruent with and extend previous research that has identified critical flaws in the data underpinning the GBD prevalence estimates for MDD. It has previously been reported that using screening tools 'in this way distorts prevalence estimates, often substantially, and does so disproportionately in low-prevalence populations". The use of screening questionnaires in epidemiological studies of prevalence, especially when devoid of socioeconomic and cultural context, will likely lead to the overdiagnosis and misguided 'treatment' of depression and misallocation of resources. Policy makers should reconsider their usefulness and application when setting priorities.

(PR) Payments to healthcare organisations reported by the medical device industry in Europe from 2017 to 2019: an observational study

James Larkin1, Shai Mulinari2, Piotr Ozieranski3, Kevin Lynch4, Tom Fahey1, Akihiko Ozaki5, Frank Moriarty1
1RCSI University of Medicine and Health Sciences, Dublin, Ireland. 2Lunds Universitet, Lund, Sweden. 3University of Bath, Bath, United Kingdom. 4Unaffiliated, Mullingar, Ireland. 5Medical Governance Research Institute, Tokyo, Japan

Objectives
The medical device industry makes payments to healthcare organisations. Industry payments can create conflicts of interest. Conflicts of interest occur in situations where a healthcare professional's or healthcare organisation's primary interest, the

care of patients, is in conflict with a secondary interest; potential financial gain. Conflicts of interest have the potential to negatively affect patient care increasing unnecessary testing and unnecessary treatment. An initial step in addressing this issue is by enhancing transparency of industry financial support to healthcare organisations. MedTech Europe, a medical device trade body, operate a system of disclosure of education payments to European healthcare organisations. This study aimed to characterise payments reported in this database and to evaluate the accessibility and quality disclosure system.

Method

An observational study of the education-related payments disclosure website transparentmedtech.eu was conducted. Transparentmedtech.eu is the disclosure website for MedTech Europe, a European medical technology trade association. On the website transparentmedtech.eu, medical device companies publish details of 'educational grants' provided to healthcare organisations. These educational grants are described by MedTech Europe as supporting "Healthcare Professionals' independent Medical Education." Payments to healthcare organisations were summarised by year, country, and medical device company. Database accessibility and quality attributes were examined. Attributes included database structure, format, searchability, downloadability and availability of summary statistics, among others.

Results

Overall, 116 medical device companies reported education-related payments in 53 European and non-European countries, valuing over €425 million between 2017-2019, increasing in value between 2017-2019, from €93,798,419 to €175,414,302. Ten countries accounted for 94% of all payments. Switzerland made up 41.8% of the total value of payments, followed by Spain (20.2%). Ten companies accounted for 80% of all payments. This high degree of concentration was largely driven by one medical device company, Johnson & Johnson Medical, who accounted for 43.3% of all payments. The accessibility, availability and quality of the database rated low for six measures, medium for six measures, and high for three measures.

Conclusions

There is a large amount of education-related payments from medical device companies to European HCOs, creating substantial potential for conflicts of interest.

Abstracts

Oral Presentations

MedTech Europe's disclosure system has many shortcomings and the figure of €425 million likely underestimates the true extent of medical device industry payments. 2,3,5,16,56 Overall the shortcomings of this database are reflective of issues seen with self-regulation across several industries, such as pharmaceutical, nutrition and alcohol. This hihglights the need for alternative governance approaches to self-regulation. A European-wide publicly mandated disclosure system for both the medical device and pharmaceutical industries should be introduced.

(T) Defining Clinical and Biological Rationale of Biomarkers to Improve the Rate of Translation

Alice Baggaley, Katerina-Vanessa Savva, Melody Ni, George Hanna, Christopher Peters
Imperial College London, London, United Kingdom

Objectives

Most biomarkers fail to move from discovery to translation and adoption. Our group has found a translation rate of only 0.16% for diagnostic colorectal cancer biomarkers in the literature. Early diagnosis of disease (in particular cancer) is a stated aim of NHS/CRUK/funders, and has been shown to be cost-effective. The Biomarker Toolkit has been created by our Biomarker Translation group to assess biomarker potential and then, more importantly, guide their further development (Savva et al, 2023). The Toolkit can be applied to biomarker research (published and unpublished) and outputs an overall score as well as subdomain scores encompassing analytical validity, clinical validity and clinical utility. The objective of this project is to expand the Toolkit to include clinical and biological Rationale, resulting in a checklist of attributes to score biomarker research for these domains.

Method

An extensive literature search was undertaken via Embase and MEDLINE databases to extract reviews/commentaries/guidelines/editorials regarding clinical and biological rationale for translated biomarkers. Abstract screening, full text screening, and data extraction was performed using a novel semi-automated software platform (AutoLit, Nested Knowledge). Grey literature, including Government recommendations and guidelines, approval regulations and industry literature was also sought out.

The resulting Rationale checklist covers rationale attributes associated with successfully translated biomarkers. The next stage of the project (in-progress) is to gain expert stakeholder consensus on this checklist, via semi-structured interviews until thematic saturation is reached, before taking forward these insights for a Delphi survey. Stakeholders include biomarker academics, clinicians, industry partners, funders, commissioners, and patients.

Results

The results of the literature search are demonstrated via the PRISMA diagram. From 2620 screened papers, 24 were included in the final analysis. From secondary

sources, 5 reporting guidelines were included, as well as internal industry biomarker classification systems, regulatory frameworks, and recommendations from national/international biomarker consortiums. Biological Rationale covers whether a biomarker makes sense with regard to disease pathogenesis, whereas clinical Rationale encompasses whether a biomarker is needed within a clinical pathway, and how it will impact management/decision-making. Within the domain of Biological Rationale, themes covered by the checklist include mechanism of action/biological plausibility, specimen type and location, intended disease stage, and biospecimen invasiveness. Themes within the Clinical Rationale domain include the Context of Use, biomarker type, intended use population, clearly stated unmet clinical need, industry or consortium collaboration, and clinical pathway mapping.

Conclusions

We have created a novel Rationale checklist, covering clinical and biological rationale, to be integrated into our Biomarker Toolkit. Use of this Toolkit aims to assess for biomarker potential, allowing better use of limited resources and to improve the translation of biomarkers, ultimately allowing increased access to biomarkers for patients. The Rationale checklist has been created with literature approved attributes and will be validated with expert consensus in ongoing work. Future work will involve validating this new version of the Biomarker Toolkit with independent datasets of biomarkers and demonstrating its potential health economic impact to the aforementioned stakeholders.

(CP) Rapid Antimicrobial Susceptibility Testing of Urinary Tract Infection (UTI) Bacteria Using an Innovative Technology: Scattered Light Integrated Collector (SLIC).

Hellen Onyango, Derek Sloan, Katherine Keenan, Mike Kesby, Robert Hammond
St Andrews, St Andrews, United Kingdom

Objectives

The lack of rapid diagnostics and increased incidence of antimicrobial resistance (AMR) among uropathogens has contributed to inadequate antimicrobial therapy among UTI patients. Accurate, timely and cost-effective determination of antimicrobial susceptibility among clinically relevant UTI bacteria is critical for patient care and antimicrobial stewardship. This study evaluates the performance of SLIC in antimicrobial susceptibility testing of UTI bacteria.

Method
A hospital-based, cross-sectional study was conducted among symptomatic UTI patients. Midstream urine was cultured on Cystine-Lactose-Electrolyte-Deficient agar and incubated at 370C overnight. AST was performed both on Kirby-Bauer disk diffusion method and SLIC, and interpretations were based on CLSI 2022 guidelines and 50% growth inhibition respectively. AST profiles generated from the two methods were compared and overall categorical agreement determined. The turnaround times (TAT) of the two AST methods was also compared.

Results
Overall, 552 participants were recruited with a median age of 29 years (IQR:24-36). The majority were female; 398 (72%). A total of 620 antimicrobial bacterial combinations were analysed. Using culture as a gold standard, the overall categorical agreement was 580/620 (94.4%), very major error rates of (4.1%), and major error rates of (6.8%). The sensitivity of SLIC in detecting AMR was 96% (CI 92.6-98) and ranged between 92-100% for individual antibiotics. There was no pattern in the errors observed, as they cut across all the bacteria and antibiotics tested. The average TAT for disk diffusion was 72.47 ± 1.05hrs. Introduction of SLIC accelerated time to AST results by 46.45 hours (26.02 ± 0.25), a 64.1%-time reduction compared to the culture dependent AST.

Conclusions
SLIC offered fast and accurate AST for UTI bacteria in a clinical setting. The results were highly comparable to the gold standard culture method. The technology has potential to provide AST results within a clinically relevant time frame, particularly in resource limited, high infectious disease settings such as Sub-Saharan Africa.

(CP) Antimicrobial Photodynamic Therapy using Organic Light Emitting Diodes: bringing light closer to the skin

Marianna de Leite Avellar, Ifor Samuel, Robert Hammond
St Andrews, St Andrews, United Kingdom

Objectives
Photodynamic therapy (PDT) is a form of treatment that uses a light source and a chemical compound (photosensitiser) which is activated by light and, in the

presence of Oxygen, generate reactive oxygen species. The antimicrobial application of PDT (APDT) is advantageous due to the non-selective oxidative damage on cellular structures, acting on multiple targets determined by the diffusion of the photosensitiser within the cell.

For PDT, light sources currently used are large and expensive machines exclusively in hospitals, limiting access to the treatment. Organic Light Emitting Diodes (OLEDs) are electroluminescent devices that can be produced as small, thin, and lightweight wearable devices, and still provide the homogeneous and efficient light output to be used as the light source for PDT. In this experiment, we show the efficiency of OLED-based APDT in vitro using methylene blue as photosensitiser on Staphylococcus aureus.

Method

A novel technology that optically measures bacterial growth on real time, the Scattered Light Integrating Collector (SLIC) (Hammond et al., 2022), was used for rapid screening of the treatment. Red-emitting 1.4 cm by 1.4 cm OLEDs were used at irradiance of 4 mW/cm2, and methylene blue, in a series of different concentrations, was used as the photosensitiser for PDT on S. aureus.

Results

OLED-based APDT using methylene blue showed significant growth inhibition of S. aureus (>3 log10) even at low concentrations of the photosensitiser.

Conclusions

Use of OLEDs for APDT shows efficacy on the treatment and a promising advancement of the field, especially with the advancement of antimicrobial resistance. OLED-PDT could represent a major advantage for treatment of skin and soft tissue infections as portable devices to allow ambulatorial treatment. Keywords: New and non-traditional drugs; Antimicrobial resistance; Antimicrobial Photodynamic Therapy.

Abstracts

Poster Presentations

CP) Testing efficacy of a novel diagnostic antimicrobial susceptibility testing platform on patient bacterial isolates from a large Scottish teaching hospital

Stuart Reid, Robert Hammond
St Andrews, St Andrews, United Kingdom

Objectives

The development of novel, rapid, AST methods are essential to combat the rise of multi drug resistant organisms. This project evaluates a novel technology with the potential to replace current laboratory AST methods. A range of bacterial isolates from patients including CPE, MRSA, VRE and Cystic Fibrosis respiratory isolates were analysed by a novel technology developed by the University of St Andrews. The Scattered Light Integrating Collector (SLIC) has previously been shown to allow the rapid calculation of Minimum Inhibitory Concentrations (MIC) of bacteria.

Method

Isolates were incubated for 16 hours on blood agar plates in aerobic conditions, then inoculated into Mueller-Hinton broth and incubated for 2 hours. The density of suspension was calculated using a spectrophotometer. The bacterial suspensions were inoculated into cuvettes and a concentration of antimicrobial was added. A growth control with no antimicrobials was included in each set of tests. The bacterial growth in the cuvettes containing the antimicrobial was measured using laser scatter methods and compared to the growth control. The effect of the specific concentration of antimicrobial was measured and the MICs compared. Discrepant results were retested using standard micro-broth dilution (BMD) methods.

Results

144 combinations of bacterial isolate/antimicrobial were tested. The SLIC results were compared with those generated by VITEK2 analyser (bioMérieux). SLIC gave concordant results in 139/144 tests (96.5%). 5 discrepant results were detected (4%).One of these isolates failed to grow on BMD retest – this was Ciprofloxacin with an isolate of Klebsiella pneumoniae. A further 3 isolates appeared to be discrepant, however, retesting via BMD indicated an MIC in concordance with SLIC and subsequently these were not classed as discrepancies.

Conclusions

The SLIC analyser was shown to replicate results from the current methods with 96.5%

accuracy and is rapid and cost saving. Although unable to be retested, a Ciprofloxacin MIC of 0.5 mg/L with Klebsiella pneumoniae is classed as an area of technical uncertainty (ATU) by EUCAST and confirmation of the result with another test method is recommended.

(CP) Identification of plasma markers associated with oesophageal cancer treatment outcomes utilising metabolomics.

Hasnain Ahmed1, David Sumpton2, Alejandro Huerta Uribe2, Guillaume Piessen3, Michael Hisbergue3, Victor H. Villar1, Alan Stewart1
1University of St Andrews, St Andrews, United Kingdom. 2Cancer Research UK Scotland Institute, Glasglow, United Kingdom. 3University of Lille, Lille, France

Objectives
Oesophageal cancers, including oesophageal adenocarcinoma (OAC) and oesophageal gastric junction cancer (OGJ), are one of the leading causes of cancer mortality worldwide due to having a minimal five-year survival of <15%. Curative therapy consists of surgery, either alone or in combination with adjuvant or neoadjuvant chemotherapy or radiation, or combinational chemoradiotherapy regimens. There is an urgent need to improve OAC disease management and treatment strategies. Current chemotherapeutic strategies only benefit a minority (20-30%) of patients. There are currently no clinico-pathological means of predicting which patients will benefit from chemotherapeutic treatments. We will analyse and compare the plasma metabolome from OAC patients prior to chemotherapy or chemo-radiotherapy and correlate the differential metabolites with the therapy outcome referred to as the tumour regression grade (TRG). The main objective of this study is to identify metabolites that may be used as predictive tools and potentially direct future treatment choices.

Method
Liquid Chromatography-Mass Spectrometry (LC-MS)-based metabolomics was performed on plasma samples from 112 patients with locally advanced OAC prior to treatment. Individuals were grouped based on their TRG grade acquired post treatment in which TRG1/2/3 groups had none or little residual tumour left, and TRG4/5 groups had an absence of regressive changes. We performed targeted

metabolomics analysis. The peak areas of the metabolites were determined by using the m/z from singly changed ions (extracted ion chromatogram ± 5 ppm) and the retention time from our in-house metabolite library (Skyline Software 23.1.0.238). The untargeted metabolomic analysis was performed using Compound Discoverer (Thermo Scientific v3.2). This allowed the identification of unknown metabolites with differential levels between different groups. The structure elucidation and identification of unknown hits will be performed using their fragmentation pattern (LC-MS/MS) and online data bases.

Results

Plasma samples within the chemotherapy treatment arm showed significantly decreased levels of pyruvate and lactate in TRG4/5 patients compared to TRG1/2/3 patients. Expansion into the untargeted chemotherapy group analysis also revealed an unknown metabolite with the predicted formula C5H11NO3 (134.0811 ± 5ppm) as the most significantly different metabolite between TRG1/2/3 vs TRG4/5, with the metabolite being lower in TRG4/5 group. Current efforts are being made into elucidating the structure of this metabolite. The targeted metabolomics analysis in patients that underwent chemo-radiotherapy revealed higher levels of α-ketoglutarate and arginine in TRG 4/5 patients.

Conclusions

The metabolome comparison between therapy responders and non-responders showed that pyruvate, lactate and one unknown metabolite (C5H11NO3) have the potential to predict neoadjuvant chemotherapy outcome. This type of analysis also showed that lower levels of α-ketoglutarate and arginine have the potential to predict beneficial outcomes for patients undergoing chemo-radiotherapy treatments. This study showcases that plasma metabolites have the potential to direct effective therapeutic strategies as well as the potential to preventing unnecessary cytotoxic effects from ineffective treatment strategies to preserve the quality of life of oesophageal cancer patients.

Abstracts

Poster Presentations

(CP) Improving early diagnosis of terminal cancer: Identification of demographic and clinical factors associated with having a very short prognosis at their time of diagnosis with cancer.

Sarah Mills1, Peter Donnan2, Deans Buchanan2, Blair H Smith2
1University of St Andrews, St Andrews, United Kingdom. 2University of Dundee, Dundee, United Kingdom

Objectives
In order to deliver appropriate and timely care planning and minimise avoidable late diagnoses, clinicians need to be aware of which patients are at higher risk of receiving a late cancer diagnosis. We aimed to determine which demographic and clinical factors are associated with receiving a 'late' cancer diagnosis.

There is no agreed definition of 'late diagnosis' in cancer care, with previous publications suggesting definitions ranging from a few weeks to one year before death. For the purposes of this study 'late diagnosis' was defined as a diagnosis of cancer within the last 12 weeks of life.

AIM: To identify any associations between demographic factor and cancer type, with receiving a late diagnosis of cancer, in a population of people who go on to die from cancer.

Method
Retrospective cohort study of 2,443 people who died from cancer ('cancer decedents') in 2013–2015. Demographic and cancer registry datasets linked using unique patient-identifying Community Health Index (CHI) numbers. Demographic data at time of diagnosis were obtained from the Cancer Registry (Scottish Morbidity Records), Scottish Executive Urban Rural Classification (SEURC, which classifies postcodes in terms of remoteness and rurality), and Scottish Index of Multiple Deprivation (SIMD, which categorises deprivation into quintiles from SIMD 1 [most deprived] to SIMD 5 [least deprived]), and linked using CHI numbers. Analysis used binary logistic regression, with univariate and adjusted odds ratios (SPSS v25).

Results
One third (n = 831,34.0%) received a late diagnosis. Age and cancer type were significantly associated with late cancer diagnosis ($p < 0.001$). Other demographic factors, including gender, rurality and deprivation, did not meet the threshold for

significance on univariate or multivariate logistic regression. Cancer decedents with lung cancer were more likely to have late diagnosis than those with bowel (95% Confidence Interval [95%CI] Odds Ratio (OR)1.52 (OR1.12 to 2.04)), breast or ovarian (95%CI OR3.33 (OR2.27 to 5.0) or prostate (95%CI OR9.09 (OR4.0 to 20.0)) cancers. Cancer decedents aged >85 years had higher odds of late diagnosis (95%CI OR3.45 (OR2.63 to 4.55)), compared to those aged <65 years. While gender was not statistically significant there was an observed tendency for women to be more likely to have a late diagnosis than men. Though it narrowly missed statistical significance, on multivariate analysis women were 16% more likely to receive a late diagnosis than men (95%CI OR0.84 (OR0.70 to 1.00).

Conclusions
Increased age and having lung cancer were strongly associated with patients having increased odds of having a late cancer diagnosis, in a population of patients who went on to die from cancer. Practice and policies aimed at addressing those at higher risk of receiving a late cancer diagnosis could have greater impact if they focused on older people and those with lung cancer symptoms.

(PR) Investigation into the incidence of co-morbidities discovered after five years of follow-up in the Early Detection of Cancer of the Lung Scotland (ECLS) study.

Nimue Lilith Romeikat
University of St. Andrews, St. Andrews, United Kingdom

Objectives
One of the main reasons behind delayed implementation of national lung cancer screening (LCS) are uncertainties surrounding cost and the potential burden that incidental findings of co-morbidities could have on the NHS.

This research aims to establish the incidence of co-morbidities, discovered after 5 years of follow-up, in the intervention arm (group receiving an EarlyCDT test and if positive, low dose CT) compared to the control arm (group receiving current standard NHS clinical care) of the ECLS study. Adding biomarkers to LDCT may correlate to greater specificity and fewer co-morbidities than LDCT alone.

The study will also describe the types of co-morbidities (cardiovascular, pulmonary etc.) and corresponding patient demographics (age, SIMD, smoking-history) in which these co-morbidities are discovered. This will contribute to determining the proportion of people screened who may be ineligible for any intervention, as well as how co-morbidities discovered during screening may affect treatment.

Method

The data for this project will be obtained from the ECLS study. For data protection purposes it will be kept in Dundee's HIC Safe Haven and 'R' will be used for statistical analysis.

Following an inspection of the database and collation of the relevant subsets, there will be two steps to the dataset analysis. Firstly, to establish whether there is a higher incidence of co-morbidities discovered at follow-up in the intervention group compared to the control group. Comparisons will be explored using the appropriate tests, such as the Chi-squared Test and, where numbers are too small, Fisher's Test.

Secondly, a correlation will be investigated between individual patient characteristics (age, SIMD, smoking-history) and the incidence of co-morbidities discovered, which will be ordered into specific subgroups of co-morbidity (cardiovascular, pulmonary, malignancy, haematological, autoimmune, renal, orthopaedic). If a correlation is discovered, Spearman's and/or Pearson's correlation will be used to quantify its significance.

Results

The project is due to commence at the start of May. Preliminary descriptive data will be available by the time of the conference.

Conclusions

Tentative suggestions based on the scientific literature and the early results will be presented at the conference.

Abstracts

Poster Presentations

PR) Great promise and big problems: Applied epidemiology and the new diagnostics

Peter Donnelly
University of St Andrews, St Andrews, United Kingdom

Objectives

Developments in multi-omic technology and AI driven signal recognition are enabling the development of new diagnostic tests for cancer which promise earlier diagnosis. Regular cancer testing for all is a politically attractive proposition which if implemented without care will pressure the NHS.

One piece of theoretical epidemiology can serve as an example. The purpose of this poster is explain the danger of implementing tests developed in artificial high prevalence environments in real world situations without adequate forethought.

Method

Consider test development in a group of people, half of whom have the disease and half of whom don't. Assuming a sensitivity and specificity each of 90%, then the positive predictive value would also be 90%. You still have 100 false positives per 2000 tests to investigate and reassure but for those who test positive there is a 90% chance that they have the disease. But what happens when prevalence is at a more realistic level. (All these results are shown in the poster in a table.)

Results

When prevalence is a rather more realistic 5% (think out patient referrals) then we have 190 false positives per 2000 people tested and a PPV of only 32%. This feels less useful but may have some benefit. However, at a prevalence of 0.5% (think screening) then the number of false positives per 2000 tested is 199 and now the PPV is only 4%. In other words, those who test positive almost certainly don't have the disease. The test has left you with 208 who think they have cancer when only 9 of them do. This feels unsustainable in terms of healthcare resource usage and irresponsible in terms of patient wellbeing.

Conclusions

Tests are developed in situations of artificially high prevalence. As we move towards disease prevalence rates more typical of primary care or screening settings then the

positive predictive value plummets. This has ethical, operational and financial implications for the NHS. New diagnostic tests must be introduced with care. Assessing test impact prior to implementation, maximising pre test probability and undertaking patient informed public communications are all strategies which could mitigate against adverse effects.

(EC) Budget impact analysis of using a novel urine biomarker test to support early diagnosis of pancreatic ductal adenocarcinoma.

Rosario Luxardo1, Katerina-Vanessa Savva1, Silvana Debernardi2, Tatjana Crnogorac-Jurcevic2, Melody Ni1, George B Hanna1
1Imperial College London, London, United Kingdom. 2Pancreatic Cancer Group at Barts Cancer Institute, London, United Kingdom

Objectives
The poor prognosis of Pancreatic ductal adenocarcinoma (PDAC) is mostly related to late diagnosis. The Pancreatic Cancer Group developed a urine-based biomarker panel and affiliated PancRISK for early detection of pancreatic cancer. Increased resources have been spent on cancer biomarker discovery, and different methods for health economics evaluation allow to identify, measure, value, and compare the costs and consequences of different public health interventions.

Budget impact analyses (BIAs) are an essential part of a comprehensive economic assessment of health care interventions and are increasingly required by reimbursement authorities as part of a listing or reimbursement submission. It assesses the likely financial impact of the technology before implementation and determines whether the technology will be affordable within the decision maker's budget holder. The objective of our work is to analyse the difference in costs and savings of introducing the novel urine biomarker test to support early diagnosis of PDAC compared with the current care pathway.

Method
The development of a BIA involves different steps.1 First step is to specify the target population likely to be impacted by the urine-based biomarker panel. This analysis will encompass test use in both symptomatic patients and in surveillance of

asymptomatic populations. The second step is to define the direct costs occurred during the current and new clinical pathway once the new technology is implemented. To define the costs and benefits a stakeholder's analysis and care pathway mapping is carried out. The perspective of the budget will be the NHS and the analysis time horizon is projected to five years. Sensitivity analyses will be conducted to assess the robustness of the model.

Results
The urine-based biomarker panel is an add-on diagnostic test. The preliminary findings of the care pathway mapping shows that the new test could be implemented in two different settings: the general practice premises and the Gastrointestinal Clinics for patients with unspecified symptoms. For patients with genetic predisposition the test would be performed as part of the surveillance programme. The main stakeholders were identified and will be invited to participate in structured interviews to discuss the proposed new care pathway and identify possible barriers related to it. The target population for the BIA will be estimated from England pancreas cancer databases and the source for the costs will be NHS.2The model is under process, we expect to have results by the time of the conference.

Conclusions
The main impact of the upcoming results from the budget impact model will be the estimation of the likely change in expenditure for the NHS as well as it will help understand the costs incurred and saved by implementing the urine-based biomarker panel.

(O) Blood-Based Proteomic Biomarkers for Alzheimer's Disease Classification using Gradient Boosting Machines with Selection Bias Correction.

Marco Fernandes, Victor Pardo, Paul Johnston, Peter Donnelly
School of Medicine, University of St Andrews, St Andrews, United Kingdom

Objectives
This study aims to identify proteomic signatures in blood plasma, measured using the SomaScan platform, to classify Alzheimer's disease (AD), mild cognitive impairment (MCI), and healthy controls from the Bio-Hermes trial cohort. Gradient boosting machines (GBM) models will be developed to predict disease status using

blood-based proteomics. Additionally, the study will address potential selection bias arising from comorbidities and medication use in AD and MCI cases. The same procedure will be applied to the UK Biobank (UKBB) OLINK proteomic dataset as a replication study cohort.

Method

We will develop GBM models using plasma-based proteomics adquired with SomaScan from the Bio-Hermes trial. The validation step encompasses a leave-one-out cross-validation using the area-under-the-curve (AUC) for model evaluation and applying a one-vs-rest (OvR) strategy for the multi-label problem. Then, SHapley Additive exPlanations (SHAP) analysis will assess the contribution of individual protein markers and their non-linear interactions. Age and ApoE4 status will be incorporated as covariates in the GBM models and used to benchmark against the main model. To address selection bias, logit models will be built for the top-10 most correlated comorbidities and medications with AD and MCI, while adjusting for age, ApoE4 status, education years, race/ethnicity, assessment center, and five principal components from the proteomic data. Afterwards, mahalanobis distances, calculated on the residuals of these logit models, will then be included in the GBM models for bias correction.

Results

This study proposes a novel approach for classifying AD, MCI, and healthy controls using blood-based proteomics and GBM models with selection bias correction. Due to the ongoing nature of the investigation, results regarding model performance (AUC), identification of key protein markers via SHAP analysis, and the effectiveness of the bias correction method are not yet available.

Conclusions

Upon its conclusion, this study holds promise in uncovering blood-based proteomic markers for Alzheimer's disease (AD) classification. The implemented selection bias correction has the potential to enhance the generalizability of the model compared to conventional methods. However, conclusive assessments regarding the effectiveness of this approach will be contingent upon the analysis of the gathered data. Disclaimer: It's crucial to acknowledge that this study is presently ongoing. The Results and Conclusions sections provided above are formulated based on the proposed methodologies and do not represent actual findings.

Abstracts

Poster Presentations

(O) Enhancing Colorectal Cancer Mismatch Repair Biomarker Prediction in Computational Pathology: A Comparative Analysis of Domain-Specific vs General-Purpose Feature Extractors for Weakly Labelled Colorectal Cancer Whole Slide Image Classification

Craig Myles, In Hwa Um, David Harrison, David Harris-Birtill
St Andrews, St Andrews, United Kingdom

Objectives

Colorectal cancer (CRC) is the second leading cause of cancer-related death worldwide. Our study investigates the efficacy of state-of-the-art in-domain feature extractors against the widely used ImageNet-trained model ResNet50 for predicting mismatch repair (MMR) status in CRC, from fewer than 500 hematoxylin and eosin stained Whole Slide Images (WSIs). MMR status plays an important role in informing patient cancer prognosis and is an indicator of positive response to immunotherapy. We aim to assess how domain-specific models, including Phikon and CTransPath, can surpass general-purpose models to maximise smaller dataset utility. The primary objective is to ascertain which feature extraction method yields the most effective feature vectors for training a Transformer model—an architecture known for natural language modelling, now adapted for complex pattern recognition—for precise slide-level MMR prediction. This research seeks to identify optimal feature extraction and model training strategies, ultimately enhancing diagnostic accuracy and treatment decision-making in colorectal cancer.

Method

Our methodology involved an extensive hyperparameter search, optimising across 144 machine learning runs, focusing on three distinct patch-level feature extractors: ResNet50, Phikon, and CTransPath, specifically for the prediction of mismatch repair (MMR) status on whole slide images (WSIs). Training and validation were executed on separate sets of WSIs, with performance evaluated based on validation Area Under the Receiver Operating Characteristic (AUROC) scores. This approach ensured a comprehensive assessment of each model's predictive capability. Employing a grid search approach, we explored various hyperparameters including activation functions, dropout rates, learning rates, and feature extractors to identify configurations which yield the most accurate MMR biomarker predictions. Particular emphasis was placed on the analysis of the top quartile of performers, identified through their AUROC scores during the validation phase, to pinpoint the most effective models. The statistical significance of performance differences among these top models was assessed using the Mann-Whitney U test.

Results
Evaluation of the top quartile performers revealed that domain-specific models, CTransPath and Phikon, significantly outperformed the ImageNet pretrained ResNet50, with CTransPath showing a notable 5.93% increase in average validation AUROC. CTransPath led with an average AUROC of 0.9187 (SD=0.0122), with the best performing model achieving 0.9466, demonstrating its superior capability for MMR status prediction in WSIs. Phikon followed with an average AUROC of 0.8924 (SD=0.0173), and a best performance of 0.9177. ResNet50 had an average AUROC of 0.8673 (SD=0.0307), with its highest at 0.9124. The statistical significance of these performance differences was validated by the Mann-Whitney U test ($U=8.0$, $p<0.001$), between CTransPath and ResNet50. This analysis underscores the critical advantage of domain-specific feature extractors in enhancing the accuracy of biomarker predictions in computational pathology.

Conclusions
These results highlight the efficacy of domain-specific feature extractors in targeted downstream tasks, using a small dataset of fewer than 500 cases for mismatch repair (MMR) status prediction. Such models, exemplified by CTransPath, efficiently process datasets to significantly improve diagnostic accuracy for specific diagnostic and prognostic biomarkers in computational pathology. This methodology not only underscores the value of domain-specific tools but also highlights their potential to enhance patient care. By informing testing and treatment decisions, machine learning models based on domain-specific feature extractors can accelerate the delivery of critical results, ultimately advancing patient outcomes.

(O) Our Future Health: the UK's largest health research programme

Iain Turnbull1,2, Raghib Ali1,3,4
1Our Future Health, London, United Kingdom. 2Nuffield Department of Population Health, University of Oxford, Oxford, United Kingdom. 3Oxford University Hospitals NHS Trust, Oxford, United Kingdom. 4Office for Health Disparities and Improvement, London, United Kingdom

Objectives
Our Future Health is a new UK-wide prospective cohort study supported by the UK Government, industry and charity sectors with the aim of recruiting 5 million UK adult

residents to facilitate aetiologic and translational research. Our aim is to help people live longer, healthier lives through the discovery and testing of more effective approaches to prevention, earlier detection and treatment of diseases.

Method

The sample frame is the total UK adult population using open enrolment. Participants are being genotyped using a custom array with a genome-wide backbone and coverage of disease- and phenotype-associated variants. High priority data linkages include primary and secondary, and cancer and death registration in each of the four devolved nations of the UK.

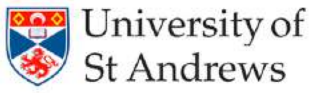
To date, 1.3 million participants have consented to be part of Our Future Health of which over 900,000 have completed the broad baseline questionnaire and over 600,000 have donated blood for genotyping and biobanking of plasma, buffy coat and DNA.

Conclusions

We aim to recruit a cohort that is reflective of the diversity of the UK population in terms of age, ethnicity, socioeconomic status and geography. Key translational research aspects of the study include the forthcoming ability for study proposals to include access to baseline blood samples, and for investigator-led recontact studies to recruit selected participants based on demographics, phenotypes, genotypes or disease-risk for risk-stratified research studies.

Thank You

The EBED Team would like to thank the following organisations for sponsorship of the conference:



We would like to extend a huge thank you to all of you for attending #EBED2024, especially the attendees who have had to travel long distances.