

**GOSH BRC Applied Child Health Informatics Theme (Non-Clinical) PhD
Studentships
PhD Project Portfolio**

Deadline for applications Wednesday 15 May 2024

Health Data Science/Epidemiology/Statistics

Project 1

Project title: Health and development outcomes and their interaction for children with chronic liver disease: a population-based cohort using novel linkage between health and education records

Supervisory Team:

Prof Katie Harron, UCL Great Ormond Street Institute of Child Health (primary)

Dr Ania Zylbersztejn, UCL Great Ormond Street Institute of Child Health (subsidiary)

Dr Marianne Samyn, King's College Hospital NHS Foundation Trust (subsidiary)

Contact: Katie Harron - k.harron@ucl.ac.uk

Background:

Each year in the UK, 400 children are diagnosed with chronic liver disease (CLD), with a fifth requiring liver transplantation at some stage. Some evidence suggests that these children may have lower cognitive ability than their healthy peers, which may affect the self-management skills required for independent management of liver disease in adulthood. Earlier detection of behavioural and developmental concerns, followed by timely referral to experienced specialist services, could improve health related quality of life, educational attainment and future employment outcomes. There is a current lack of evidence on developmental outcomes for children with CLD, how these interact with health outcomes, and when is the optimal time to intervene.

Aims/Objectives:

This study will establish a national data resource for children with CLD using linked health and education records, to evaluate healthcare and educational outcomes (including need for learning support), and their interaction. It will generate evidence to inform future development of guidance on long-term developmental follow-up.

Research questions:

RQ1 How do developmental outcomes, including school attainment and special educational needs (SEN) support, of children with different liver disease diagnoses compare to each other, and to those of the general population?

RQ2 To what extent does school absence due to medical care explain the association between CLD and developmental outcomes in children with liver disease?

RQ3 How do developmental outcomes of children following liver transplantation compare to those who have not undergone liver transplantation?

RQ4 What is the effect of age at liver transplantation on school age health and educational outcomes?

Methods:

We will use the ECHILD Research Database which links hospital and school records for all children born in England since 1984. We will create phenotypes for CLD based on diagnosis and procedure codes in and use NHS Blood transfusion and Transplantation (NHSBT) data to identify a cohort of children up to age 18 with CLD (and specific diagnoses) and/or transplantation. Outcomes will include hospital admissions, school attainment, SEN and absences.

Analysis methods: We will use statistical models to evaluate outcomes, adjusting for relevant confounders (e.g. ethnicity, deprivation, birth characteristics) and clustering of outcomes within schools. We will use propensity scores to create a matched comparison group for children with and without liver transplantation, based on demographic and

clinical indicators prior to transplantation. We will extend these models to evaluate whether age at transplantation is associated with outcomes accounting for competing risks (e.g. death).

Timeline:

Months 1-6: Systematic review of cognitive outcomes for children with CLD

Months 7-12: Creation of analysis cohort; data-cleaning; identification of cases; PPIE.

Months 13-18: Analysis and write-up for RQ1-RQ2.

Months 19-24: Analysis and write-up for RQ3.

Months 25-36: Analysis and write-up for RQ4; dissemination; PPIE.

Collaborations:

Dr Jane Hartley, consultant paediatric hepatologist at Birmingham Children's Hospital.

Plans for patient and public involvement and engagement for the project/student:

The student will support ongoing PPIE activity for ECHILD and will work with a group of parents and children and young people to understand their research priorities and their views on the use of routinely collected data for research purposes. PPIE will be ongoing throughout the PhD.

Project 2

Project title: Health outcomes of children with rare or complex conditions and their families: a longitudinal cohort study using linked primary and secondary healthcare data in England

Supervisory Team:

Dr Ania Zylbersztejn, UCL Great Ormond Street Institute of Child Health

Dr Joachim Tan, UCL Great Ormond Street Institute of Child Health

Prof Mario Cortina-Borja, UCL Great Ormond Street Institute of Child Health

Contact: Joachim Tan - joachim.tan@ucl.ac.uk, Ania Zylbersztejn - ania.zylbersztejn@ucl.ac.uk, Mario Cortina-Borja - m.cortina@ucl.ac.uk

Background:

Rare or complex conditions (RCCs), such as major congenital anomalies or inherited conditions that are screened for in newborns, have serious health, functional and social consequences for affected individuals. Although individual conditions are rare, collectively the prevalence of RCCs is around 5% of the population.[1] Children with RCCs often require ongoing medical treatment or surgical intervention, and they have more frequent interactions with healthcare services than their peers. This leads to absences from schools and risk of educational underachievement, thereby compounding the effect of existing social barriers in under-served communities and puts them at greater risk of developing mental health problems.[2]

Caring for children with complex health needs makes significant demands on parents and is likely to take a toll on their physical and mental health. This may also have additional adverse impact on siblings who live in the same household. Synthesising evidence on the health needs and healthcare usage in both primary and secondary care settings enables better estimation of the total health burden of RCCs on affected children and their families. This is needed to answer calls from The England Rare Diseases Action Plan 2022 for more holistic consideration of the support needs of individuals and their families across different services.[1]

Aims/Objectives:

Our overall aim is to describe the mental and physical health of children with specific rare or complex conditions (RCCs) and their families, by following children born from 2002 up to age 21 years. We will focus on children with conditions present at birth, namely major congenital anomalies (MCAs) and inherited conditions included in the newborn blood spot screening (NBS) programme in the UK.[3]

The specific objectives are to:

- 1) Describe the prevalence of RCCs recorded in administrative data by phenotype subgroups, ethnicity, geographical region and area-based deprivation.
- 2) Compare health outcomes (including mental health, chronic conditions, death) and rates of healthcare interactions (primary and secondary) for children with RCCs and their peers, and explore variation in outcomes according to socio-demographic and geographic factors.
- 3) Compare mental and physical health outcomes and rates of interactions with primary care in mothers of children with RCCs compared to the general population of mothers.
- 4) Compare health outcomes (including mental health, chronic conditions, co-occurrence of RCCs) and interactions with healthcare (primary and secondary) for siblings of children with RCCs compared to the general population of siblings.

Methods:

This is a longitudinal cohort study. The data to be analysed is presently available and consists of primary healthcare data from the Clinical Practice Research Datalink (CPRD) linked to secondary healthcare data (Hospital Episode Statistics: admissions, outpatient and accident & emergency), as well as area-based deprivation measures, ethnicity and civil registration mortality records.[4] The study cohort includes children born from 2002 onwards registered in GP practices in CPRD, covering ~30% of the UK population. Follow-up will be from birth until age 21 years, with a minimum of 6 months. We estimate that there will be about 3.5 million children with data available for analyses. Mothers and siblings (i.e. other children linked to the same mother as the child with RCC – "index child") will be indicated using CPRD mother-baby link data.

Analysis:

Prevalence will be calculated by RCD subtypes and stratified by age, ethnic group, regions and IMD quintiles. Estimates generated from cohorts developed using established phenotype codelists will be compared against published sources.[5] For objectives 2-4, we will calculate the incidence of various study outcomes by child's age for children with RCCs, their mothers and siblings. Strength of association, adjusting for covariates, will be estimated using regression models (Poisson/negative binomial, Cox's proportional hazards, logistic or log-binomial) depending on outcome. Adjusting for year of birth, sex, region of residence, ethnicity, deprivation indices, family size and calendar year will be considered for all models within the study. For models of maternal and sibling health we will additionally consider adjusting for age of the index child, indicators of complexity of their health needs and underlying comorbidities of mothers/siblings present before the birth of index child. Propensity score matching may be used to balance the distribution of baseline covariates between mothers/siblings of children with RCCs and the general population. A combination of sensitivity analyses and/or multiple imputation will be used to address missing data as required. We will use multilevel models or other approaches (e.g. generalised estimating equations) to account for clustering on GP practice level. Besides the project's modelling components, we will explore the use of interactive visualisations which may contribute to its Patient and Public Involvement and Engagement (PPIE) aspect.

Timeline:

Months 1-6: Systematic review of health outcomes of children with RCCs and their families. Weekly meetings with principal and/or subsidiary supervisors.

Months 4-12: Development of analysis cohorts and sub-cohorts; data-cleaning; identification of cases; PPIE groundwork. Scheduled meetings every 2-4 weeks.

Months 13-21: Analysis and write-up for Objectives 1 and 2. Scheduled meetings every 1-2 months.

Months 22-30: Analysis and write-up for Objectives 3 and 4. Scheduled meetings every 2-3 months.

Months 31-36: Consolidation and final writing-up; dissemination activities and PPIE.

Collaborations:

Our collaborations comprise clinicians from GOSH who are specialists in the RCCs being studied and leads of other GOSH BRC themes. They include Professor Paul Gissen (Paediatric Metabolic Medicine) and Professor Paolo de Coppi (Paediatric Surgery) who will help review phenotype and disease code lists, interpret results and guide translation of

findings into clinical practice. The PhD candidate will be expected to acquire experience of working with PPIE groups such as the GOSH BRC Parents' and Carers' Advisory Group and the Young People's Advisory Group, who will provide feedback on research priorities, help shape analyses and dissemination strategies. There will be several opportunities to present intermediate work at monthly departmental seminars and academic conferences such as the Administrative Data Research UK and International Population Data Linkage Network. For additional learning, the student will have access to taught modules within UCL's undergraduate and MSc courses, and the Institute of Child Health also offers a range of short courses in statistical analyses, data science and writing that are often free to registered students. Lastly, the candidate will also benefit from the rich support network of researchers working in health data science across the Child Health Informatics Group (CHIG), Children and Families Policy Research Unit (CPRU) and ECHILD database team within the Population, Policy and Practice department.

References:

1. Department of Health and Social Care. GOV.UK. 2022 [cited 2023 Oct 30]. England Rare Diseases Action Plan 2022. [Available from: <https://www.gov.uk/government/publications/england-rare-diseases-action-plan-2022/england-rare-diseases-action-plan-2022>; accessed 29/01/2024]
2. Barker MM, Beresford B, Fraser LK. Incidence of anxiety and depression in children and young people with life-limiting conditions. *Pediatr Res.* 2022 Nov 11;1–10
3. Public Health England. GOV.UK. 2018 [cited 2023 Oct 26]. Newborn blood spot screening: programme overview. [Available from: <https://www.gov.uk/guidance/newborn-blood-spot-screening-programme-overview>; accessed 29/01/2024]
4. CPRD linked data. [Available from: <https://cprd.com/cprd-linked-data>; accessed 29/01/2024]
5. Denaxas S, Gonzalez-Izquierdo A, Direk K, et al. UK phenomics platform for developing and validating electronic health record phenotypes: CALIBER. *J Am Med Inform Assoc.* 2019;26(12):1545-1559. doi:10.1093/jamia/ocz105

Project 3

Project title: Harms to children and young people in the UK due to health service delays in management of eye/vision conditions.

Supervisory Team:

Prof Jugnoo Rahi, UCL Great Ormond Street Institute of Child Health (primary)

Dr Ameenat Lola Solebo, UCL Great Ormond Street Institute of Child Health (subsidiary)

Contact: Jugnoo Rahi - j.rahi@ucl.ac.uk

Background:

Childhood visual impairment is uncommon in industrialised countries, with the all-cause prevalence in the UK of 2 per 1000 and a cumulative incidence ('life time risk by age 16 years') of 5.9 (5.3-6.5) per 10000 (1). Early diagnosis is essential, for example through the UK's Newborn and Infant Physical Examination (NIPE) Screening Programme, as it enables prompt referral and thus prompt treatment. This is essential as – from Nobel Prize winning neurosciences research - we know that there is a 'critical period' for visual development in early childhood.

The Royal College Ophthalmologists has identified growing evidence of patients experiencing harm for example irreversible site loss, due to unintentional delays within NHS. But both national studies undertaken to date by the RCOphth (2) have focused exclusively on adults.

There is a need to undertake a national study to identify the current burden of harms (incidence and types of harms) experienced by children and young people with eye conditions arising through delays in the NHS attributable to pre-existing capacity issues exacerbated by the COVID-19 pandemic. Further, there is need to identify causes of delays and evaluate the utility of predictive models to identify children at increased risk of harm when a delay occurs.

This will provide the evidence base needed to inform the design and development of strategies to reduce health service delays and targeted institution of harm mitigation measures to improve outcomes for children and young people with eye conditions in the UK.

RCOphth's Paediatric Subcommittee has committed its support to this proposal and implementation of the study findings through national guidance - making this inherently translational research.

Aims

1. To assess the current national landscape and burden of harms experienced by children and young people with eye conditions arising through delays within health service provision in paediatric ophthalmology in the UK, including the additional impact of the COVID-19 pandemic.
2. To identify avoidable causes of delays, propose solutions for such delays and evaluate the utility of predictive models to identify children at increased risk of harm when a delay in healthcare occurs.

Methods

A national epidemiological study with incident 'cases' of harm due to delays identified through the well-established active surveillance scheme run by British Ophthalmological Surveillance Unit (BOSU) which over the past 25 years has provided the mechanism for a number of seminal studies of rare ophthalmic disorders or events are public health importance (3).

Newly identified incident cases are reported each month by the managing ophthalmologist and data collected using standardised questionnaires both at initial notification and at follow-up, usually one year later. Annual incidence and types of harms will be reported nationally and rates of harms investigated by key characteristics including age, socio-economic status, ethnicity and geographic location. Findings will be reported with a focus on avoidable causes of delays identifying the implications of these findings for clinical practice service provision and future research.

We will build predictive models of harm to identify the 'patient' and 'health care system' features that are most important. We will explore, if the sample size makes this feasible and appropriate, the use of Machine Learning as an approach to predictive modelling.

BOSU has already approved the study in principle as this is a priority for the RCOphth.

Enabling factors

Professor Rahi and Dr Solebo have undertaken several high impact studies through BOSU (Rahi is also BOSU's chair) and have successfully supervised prior PhD students together. As they both also have policy roles within Royal College of Ophthalmologists and the UK's National Screening Committee, they are well placed to support timely translation. This will broaden the student's academic development as there are often few opportunities for students to be involved in implementation of their research findings.

Timeline:

Formal PhD milestones (Thesis Committee meetings as per ICH schedule, Upgrade at 12 months and Submission at 36 months). Broad timelines for other activities are shown below.

Activity	0-6	7-12	13-18	19-24	25-30	31-36
Induction						
Literature Review						
Prepare and submit ethics and CAG and PBPP						
Prepare and submit BOSU application						
Data collection						

Ongoing and final Data analysis								
Final Thesis writing								
Dissemination (papers from thesis)								

Collaborations:

We will undertake the study through a collaborative study group comprising paediatric ophthalmologists in the UK that will be established for the study. This is the model we have developed and used effectively for prior national studies via BOSU.

Plans for patient and public involvement and engagement for the project/student:

We have extensive experience of PPIE in all our research and know its value and that it is an important part of the student’s research training and experience.

The student will discuss the proposed study with the Moorfields EYE-YPAG and the GOSH YPAG to seek input into ethical issues, study design and study resources/documents to inform the preparation and submission of the ethics, CAG and HRA applications. They will present interim study findings at an appropriate time point to both groups to seek input into interpretation and into dissemination plans.

References:

1. Rahi JS1, Cable N; British Childhood Visual Impairment Study Group. Severe visual impairment and blindness in children in the UK. Lancet. 2003 Oct 25;362(9393):1359-65.
2. Foot B1, MacEwen C2. Surveillance of sight loss due to delay in ophthalmic treatment or review: frequency, cause and outcome Eye (Lond). 2017 May;31(5):771-775. doi: 10.1038/eye.2017.1. Epub 2017 Jan 27.
3. Also see <https://www.rcophth.ac.uk/2017/02/bosu-report-shows-patients-coming-to-harm-due-to-delays-in-treatment-and-follow-up-appointments/>
4. <https://www.rcophth.ac.uk/standards-publications-research/audit-and-data/the-british-ophthalmological-surveillance-unit-bosu/>
5. <https://topol.hee.nhs.uk/wp-content/uploads/HEE-Topol-Review-2019.pdf>
6. <https://www.gov.uk/government/publications/using-information-technology-to-improve-the-nhs>

Project 4

Project title: Health and education outcomes of children with Sickle Cell Disease in England

Supervisory Team:

Dr Rachel Knowles, UCL Great Ormond Street Institute of Child Health (primary)
 Prof Pia Hardelid, UCL Great Ormond Street Institute of Child Health (subsidiary)

Contact: Rachel Knowles - rachel.knowles@ucl.ac.uk

Background:

SCD is a serious, life-long condition that can cause anaemia, increased infection risk, and pain.¹ Around 300 babies are born with SCD annually in the UK. SCD mainly affects people with African or Caribbean heritage. Children and Young People (CYP) living with SCD and their families may face additional challenges, including racism and barriers to healthcare access.² CYP with SCD may have multiple short periods of school absence, or longer periods of illness leading to ‘persistent absence’.³ There is a dearth of data regarding health and education outcomes for children and young people (CYP) with SCD, particularly since research into outcomes for CYP with SCD has been neglected compared to CYP living with other genetic conditions, such as cystic fibrosis.² This PhD project will seek to address this research gap, by

using a unique linked, administrative database, ECHILD,4 which connects national health, education and social care data for all CYP in England – around 15 million individuals.

Aims/Objectives:

The overall aim of the project is to describe health and education outcomes for CYP with SCD. The specific objectives are to:

- 1) Define a cohort of CYP with SCD in ECHILD, and derive suitable, matched control cohorts of CYP with other chronic conditions, and CYP with no condition recorded
- 2) Describe variations in health outcomes for CYP with SCD within ECHILD, including mortality, hospital admission rates, accident and emergency attendance rates
- 3) Describe variations in education outcomes for CYP with SCD, including absences, SEN provision and attainment (key stage results).

Methods:

The student will use the ECHILD database, which links Hospital Episode Statistics (HES), the national hospital database for England, to the National Pupil Database (education records for the 93% of children in English state schools), and social care records. For objective 1, the student will use clinical coding in HES to characterise a cohort of CYP with SCD in ECHILD, born since the mid-1990s. They will externally validate the cohort using parallel analyses of the ECHILD cohort and the National Haemoglobinopathy Register, a national register of children diagnosed with SCD via newborn screening. They will develop control cohorts of children unaffected by SCD (but potentially affected by other long-term conditions); the most useful control groups will be decided through conversations with CYP, parents and clinicians. For objectives 2&3, the student will fit appropriate generalised linear mixed models, or time-to-event analyses, utilising the longitudinal data in ECHILD to describe variations in health and education outcomes for children with SCD, compared to control children, according to English region, ethnic group, parental migration history, and socio-economic status.

Timeline: We expect the student to have defined and validated the SCD ECHILD cohort and prioritised outcomes by the end of year 1, leaving year 2 and the first half of year 3 to meet objectives 2 & 3 and present results to stakeholders (see below); and the last 6 months for writing up the thesis.

Collaborations:

Our collaborations will ensure both professional development for the student, and translational impact of the research. Dr Andrea Leigh (Red Cell consultant at UCLH), and Drs Emma Astwood and Mark Velangi (consultant haematologists at Sheffield Children's Hospital & Birmingham Children's Hospital respectively), will help define code lists, control populations and outcomes, and support translation of findings into clinical practice. Via Dr Carl Reynolds, the student will present updates and findings to the NHS Race and Health Observatory. Via Dr Knowles, Clinical Advisor to the Newborn Bloodspot Screening Programme, the student has opportunity to the present their results at programme meetings and to inform screening policy and improvements. The HDR-UK Training Team will promote this PhD opportunity via their Health Data Science Black Internship Programme Alumni Network to encourage data scientists from communities affected by SCD to apply.

Plans for patient and public involvement and engagement for the project/student:

We have discussed this project with the Sickle Cell Society (SCS) and the NHS Race and Health Observatory, and will work with them to involve parents, children and young people in this project, refine research questions, interpret results, and support dissemination.

References:

1. Dormandy E, James J, Inusa B, Rees D. How many people have sickle cell disease in the UK? *J Public Health* 2017;40(3):e291-e95. doi: 10.1093/pubmed/fdx172
2. Lee L, Smith-Whitley K, Banks S, Puckrein G. Reducing Health Care Disparities in Sickle Cell Disease: A Review. *Public Health Rep* 2019;134(6):599-607. doi: 10.1177/0033354919881438

3. Sick Cell Society. Sick Cell Disease in Childhood: Standards and Recommendations for Clinical Care. 2019. https://www.sickcellssociety.org/wp-content/uploads/2019/11/SCD-in-Childhood_Final-version-1.pdf
4. Mc Grath-Lone L, Libuy N, Harron K, Jay MA, Wijlaars L, Etoori D, Blackburn R. Data Resource Profile: The Education and Child Health Insights from Linked Data (ECHILD) Database. *Int J Epidemiol* 2022;51(1):17-17f. doi: 10.1093/ije/dyab149

Project 5

Project title: Developing statistical approaches to analyse and report reinterventions in children who have heart surgery for national benchmarking and quality improvement.

Supervisory Team:

Dr Deborah Ridout, UCL Great Ormond Street Institute of Child Health (primary)
 Prof Katherine Brown, Great Ormond Street Hospital (subsidiary)

Contact - Deborah Ridout - d.ridout@ucl.ac.uk, Katherine Brown - Katherine.brown@gosh.nhs.uk

Background:

Congenital heart disease (CHD): CHD affects approximately 5600 live-born children annually in England and Wales and is the commonest cause of infant death due to congenital anomalies in the United Kingdom (UK). Annually, 7000 paediatric cardiac procedures are undertaken in the UK, 58-60% of them in children under a year old. In very complex CHD, with only one functional ventricle, children need serial surgeries to achieve stable palliation consisting of at least two in infancy. For many CHD types with two functional ventricles, one-stage repair in infancy is standard practice and more than one operation may be required for smaller or more complex patients since they need a staging surgery; or as children grow, they may need reintervention to accommodate their increasing size.

The national congenital heart diseases audit (NCHDA): The research analysis will focus on the NCHDA data with linkage to other important dataset providing long term survival status and hospital care episodes. NCHDA data is of high quality, as evidenced by annual independent validation of children's heart surgeries at the 11 specialistic centres in the UK. CHD and paediatric cardiac operations are very diverse and complex and are described in NCHDA using a special coding scheme called the International Pediatric Congenital Cardiac Code (IPCCC). The specific primary CHD diagnoses and paediatric cardiac operations are described in terms of their complexity or risk of death based on code combinations, using previously developed and tested methodology from our research group.

Paediatric cardiac surgery outcomes: In children undergoing paediatric cardiac surgery the early post-operative mortality rate is currently <2%, and this means that to understand disease and treatment impacts better, with a view to further care improvements, we need to look beyond this early post-operative time window. In our recent CHAMPION project (to which this proposal is aligned), key datasets were linked for ~ 60,000 children with CHD and longer-term survival was ascertained. For the most complex CHD, hypoplastic left heart syndrome survival is 54% at 10-years of age and for a less complex condition, tetralogy of Fallot survival is 95% at 10-years of age. We found that the range of procedures that children undergo varies. For hypoplastic left heart syndrome children have a median (IQR) of 3 (2,4) procedures by the age of 5-years and 44% have at least one 'off pathway re intervention'. With tetralogy of Fallot, children have a median (IQR) of 1 (1,2) procedure by the age of 5-years, and 20% have at least one re intervention. As we detailed in the PPI section, patients and users have identified the occurrence of reinterventions as a key metric for people living with CHD. Reinterventions lead to more time in hospital and important health burdens for patients. Quality improvement initiatives directed at reducing rates and impacts of reinterventions require that these are measured and recorded in a coherent and robust way. There are important methodological and statistical challenges to meet before this can be undertaken.

Aims/Objectives:

The aims are to describe the burden of cardiac reinterventions for the national cohort of children with complex CHD; to explore applicable statistical approaches for future analyses of cardiac reinterventions as competing events in

national audit and benchmarking, and to assess the impact of cardiac reinterventions early in life on subsequent survival.

Objectives

1. To identify cardiac unplanned reinterventions meaning off pathway surgeries and interventional catheterisations undertaken over and above the interventional treatment pathway amongst children with complex sentinel CHDs
2. To explore a range of statistical techniques applicable to the analysis of unplanned reintervention, accounting for competing events, for children surviving with complex sentinel CHDs
3. To identify risk factors for cardiac reinterventions undertaken over and above the established treatment pathway for children with complex sentinel CHDs
4. To explore links between early cardiac re interventions undertaken over and above the established treatment pathway in complex sentinel CHDs and the outcome of long-term survival

Methods:

Study design:

Retrospective cohort study based on linked electronic health care record data for ~ 60,000 children and adolescents with congenital heart disease.

- a) National Congenital Heart Diseases Audit, (NCHDA)
- c) Hospital Episode Statistics (HES),
- d) Deaths from Office of National Statistics (ONS).

Inclusion criteria: The CHD types were selected (by clinicians and patients) as part of the CHAMPION project, as major conditions that affect young children, necessitate intervention at a very early age to survive, and commonly require more than one procedure: single ventricle - hypoplastic left heart syndrome, functionally univentricular heart conditions (double inlet left ventricle and tricuspid atresia), single and two ventricle - pulmonary atresia all types, two ventricle defects - transposition of the great arteries, tetralogy of Fallot, atrioventricular septal defect, congenital aortic stenosis, significant ventricular septal defect, and coarctation of the aorta.

Study period: We will include children who were born and were recorded as having any surgical procedure for one of the selected CHD types between 2000 and 2022, with complete follow up data available to 2023.

Exposures and comorbidities: gender, ethnicity, deprivation, region, age, preterm birth, non-cardiac underlying health conditions as previously defined.

Study outcomes: for objectives 1-3) the outcome is cardiac reintervention and for objective 4) the outcome is survival.

Statistical methods

Objectives 2 and 3 – Reinterventions will be explored using both time to event analysis and competing risks analysis (competing events include death, heart transplant etc.). It is likely that conditional models will be explored, based on whether patients have achieved or completed different stages of their planned clinical pathway and the use of multi-phase models can be investigated. Models will be adjusted for known, pre-specified covariates and time varying covariates will be considered as appropriate.

Objective 4 - the use of novel causal mediation methods will be explored, to understand the relationship between case complexity, reintervention as a mediator and the outcome of long term survival.

Timeline:

Year one – descriptive analysis objective 1)

Year two – develop time to event and competing risk analysis objectives 2) and 3)

Year three – exploring links between reinterventions and outcome objective 4)

Collaborations:

Anusha Jegatheeswaran

Locum Consultant and Honorary Associate Professor Cardiothoracic Surgery recently joined GOSH from Toronto, where she was undertaking research into cohorts with the paediatric cardiac surgery multi-centre registry based there at the Congenital Heart Surgeons' Society Data Centre.

Nigel Drury, DRURY

Associate Professor of Paediatric Cardiothoracic Surgery, Birmingham Children's Hospital and University of Birmingham, will provide clinical interpretation of the data from the perspective of another large children's cardiac programme.

Sonya Crowe

Professor of Operational Research, Department of Mathematics, University College London, will provide additional analytical support with the data as a second analytical supervisor to Ms Ridout.

Plans for patient and public involvement and engagement for the project/student:

Working with CHD user groups within the conduct of the CHAMPION project, we identified CHD types that will be used in the proposal. Clinical and lay advisors reached a consensus that the following issues were most important: these CHD conditions are the most prevalent, most likely to require one or more operations in infancy, and most likely to be linked to mortalities. Reinterventions were identified as a key outcome metric within a patient engagement process. Patient user groups individually set up and moderated closed, asynchronous, online discussion groups via their Facebook pages. There were separate forums for adult patients/carers, adolescent patients and parents/carers of children and young people with CHD. Questions focused on participants' perceptions of important outcomes related to CHD, including how and what outcomes should be reported. Forums ran for 12-24 weeks in 2021.

References:

1. Risk Factors for Reintervention With Functionally Single-Ventricle Disease Undergoing Staged Palliation in England and Wales: A Retrospective Cohort Study. Huang Q, Ridout D, Tsang V, Drury NE, Jones TJ, Bellsham-Revell H, Hadjicosta E, Seale AN, Mehta C, Pagel C, Crowe S, Espuny-Pujol F, Franklin RCG, Brown KL. *Circulation*. 2023 Oct 24;148(17):1343-1345. doi: 10.1161/CIRCULATIONAHA.123.065647. Epub 2023 Oct 23.
2. Hospital resource utilization in a national cohort of functionally single ventricle patients undergoing surgical treatment. Dorobantu DM, Huang Q, Espuny Pujol F, Brown KL, Franklin RC, Pufulete M, Lawlor DA, Crowe S, Pagel C, Stoica SC.
3. *JTCVS Open*. 2023 Apr 8;14:441-461. doi: 10.1016/j.xjon.2023.03.014. eCollection 2023 Jun. Long-term survival and center volume for functionally single-ventricle congenital heart disease in England and Wales. Brown KL, Huang Q, Hadjicosta E, Seale AN, Tsang V, Anderson D, Barron D, Bellsham-Revell H, Pagel C, Crowe S, Espuny-Pujol F, Franklin R, Ridout D. *J Thorac Cardiovasc Surg*. 2023 Aug;166(2):306-316.e3. doi: 10.1016/j.jtcvs.2022.11.018. Epub 2022 Nov 25.
4. The Relationship Between Postoperative Complications and Children's Quality of Life: A Mediation Analysis to Explore the Role of Family Factors. Brown KL, Moinuddin M, Jones A, Sheehan K, Wellman P, Rodrigues W, Ridout D, Wray J. *Pediatr Crit Care Med*. 2023 Mar 1;24(3):194-203. doi: 10.1097/PCC.0000000000003137. Epub 2022 Dec 30
5. Factors associated with unplanned reinterventions and their relation to early mortality after pediatric cardiac surgery. Dorobantu DM, Ridout D, Brown KL, Rodrigues W, Sharabiani MTA, Pagel C, Anderson D, Wellman P, McLean A, Cassidy J, Barron DJ, Tsang VT, Stoica SC. *J Thorac Cardiovasc Surg*. 2021 Mar;161(3):1155-1166.e9. doi: 10.1016/j.jtcvs.2020.10.145. Epub 2020 Nov 27

Applied Machine Learning and Informatics

Project 6

Project title: Disease modelling to understand long-term progression and treatment response in Spinal Muscular Atrophy and Duchenne Muscular Dystrophy.

Supervisory Team:

Prof Giovanni Baranello, UCL Great Ormond Street Institute of Child Health (primary)
Dr Deborah Ridout, UCL Great Ormond Street Institute of Child Health (subsidiary)

Contact - Giovanni Baranello - g.baranello@ucl.ac.uk, Deborah Ridout - d.ridout@ucl.ac.uk

Background:

This project addresses the urgent medical need to collect and analyse big data sources on the natural history disease progression of children with Spinal Muscular Atrophy (SMA) and Duchenne Muscular Dystrophy (DMD). Currently some treatments are available for both these rare neuromuscular conditions in the UK, and many clinical trials are ongoing, with further therapeutic options, including genetic therapies, that could become available to patients in the near future. There is a strong need to understand how different aspects, including motor function, growth patterns, orthopaedic comorbidities, respiratory and bulbar function evolve, both in untreated and in treated patients, to improve outcomes and personalize treatment. Several studies have shown that patients with SMA and DMD tend to follow different trajectories based on contributing factors that are only partially understood. In the case of DMD, our group has published several studies showing that steroids regimen (daily versus intermittent), baseline motor function score, height and weight, as well as specific clusters of gene deletions can partly explain the variation of phenotype and disease progression in individual patients. Similarly, in the instance of SMA, three genetic disease modifying treatments are now available through NHS; however, patients respond differently and show differential responses in different domains like motor, respiratory and bulbar function that is not easy to predict. Machine learning and data modelling techniques can be applied to better characterize the pattern of disease progression and predict response to treatment in these rare and complex neuromuscular conditions and will inform precision medicine approaches to identify the optimal therapeutic protocols and management for patients treated with approved disease-modifying medications. This can ultimately improve patients' outcome and allow a more efficient use of economic and public health resources to deliver novel therapies to patients with rare neuromuscular conditions.

Aims/Objectives:

Over the past three years our team has been working to expand the data collection and improve data quality and cleaning of the two large national disease specific databases for DMD and SMA (the North Star and the SMA REACH databases, respectively). The systematic collection of biosamples and muscle imaging has also been implemented. A PhD student whose completion is expected in March 2024 has been working on piloting a joint modelling and machine learning approach to investigate multiple disease outcomes simultaneously. The aim of this PhD project application is to continue this pilot work and further collate and use advanced statistical modelling to analyse both retrospective and prospective clinical and biomarker data in SMA and DMD patients, to define a model to predict disease progression and clinical response to treatment.

Methods:

Currently more than 1200 DMD patients are registered in the North Star database and more than 600 SMA patients treated with disease modifying treatments in the SMA REACH database. There is a wealth of opportunities to build on previous research in this field. Approaches such as longitudinal trajectory analysis and joint modelling could form the basis of this project. Other objectives could be explored considering more novel applications of latent class models, or machine learning techniques, in order to provide essential information for evaluating the impact of the underlying conditions and comorbidities on clinical outcomes.

Timeline:

Months 0-6: research passport, review of the existing literature, identification of key questions to address during the project and selection of statistical methods to apply.

Months 7-12: access to the DMD and SMA national databases; data cleaning and quality control; extension of the pilot on joint modelling and machine learning.

Months 12-24: Data processing: training with different algorithms; validation.

Months 24-36: Evaluation of model performance and project completion.

Collaborations:

Prof. Baranello is the Co-Lead of the North Star and the SMA REACH national networks and databases. The PhD student will have access to this wealth of data, most of them already cleaned and verified for a previous PhD project. The strong connection with other sites across the national network will ensure access to additional data through an audit submission to the steering committee of the North Star and the SMA REACH networks, should this additional required data for the project not be recorded on the database. Dr. Ridout will provide expert advice on statistical approaches to use for the big data analysis. Advice from the ACHI BRC Theme and other collaborations, including the Turing Institute, for machine learning approaches will be available to the PhD student.

Plans for patient and public involvement and engagement for the project/student:

The student will work in close collaboration with patient representatives and advocacy groups from MDUK and SMA UK to discuss priorities and clinical needs from PPI perspective to guide their research project. The student can also be invited to be part of the steering committees of the North Star and SMA REACH networks that comprise different stakeholders including public and patients representatives.

References:

1. Muntoni F et al. DMD Genotypes and Motor Function in Duchenne Muscular Dystrophy: A Multi-institution Meta-analysis With Implications for Clinical Trials. *Neurology*. 2023 Feb 1;10.1212.
2. Coratti G et al. Predictive models in SMA II natural history trajectories using machine learning: A proof of concept study. *PLoS One*. 2022 May 5;17(5):e0267930.
3. Weststrate H et al. Evolution of bulbar function in spinal muscular atrophy type 1 treated with nusinersen. SMA p-FOIS Working Group. *Dev Med Child Neurol*. 2022 Jul;64(7):907-914.
4. Chesshyre M et al. Investigating the role of dystrophin isoform deficiency in motor function in Duchenne muscular dystrophy. *J Cachexia Sarcopenia Muscle*. 2022 Apr;13(2):1360-1372.
5. Baranello G et al. Risdiplam in Type 1 Spinal Muscular Atrophy. *N Engl J Med*. 2021 11;384(10):915-923.

Project 7

Project title: Using artificial intelligence and machine learning techniques to improve diagnosis and predict outcomes in children with heart muscle disease.

Supervisory Team:

Prof Juan Pablo Kaski, Great Ormond Street Hospital
Dr Gabrielle Norrish, Great Ormond Street Hospital

Contact: Juan Pablo Kaski - j.kaski@ucl.ac.uk

Background:

Paediatric cardiomyopathies, or heart muscle diseases, are a group of rare and heterogeneous conditions with highly variable outcomes. However, they share certain characteristics including a need for life-long serial follow up, and increased risk of arrhythmias, heart failure and, or sudden death(1). The 2023 ESC Guidelines for the management of cardiomyopathies have emphasised the importance of determining the underlying cause for the diagnosis, management and follow up of patients with heart muscle disease. However, for some types of childhood cardiomyopathies (eg dilated cardiomyopathy) aetiology remains undetermined in up to 60% meaning these patients are not able to benefit from aetiology-specific therapies. Population based studies have provided useful information on the natural history and outcomes of cardiomyopathies presenting in childhood. However, our understanding of disease progression remains limited and we are unable to predict likely disease trajectory despite collecting a wealth of serial data from regular clinical reviews. We have recently developed and validated the first paediatric specific risk prediction model for SCD in childhood HCM using routinely collected, non-invasive baseline clinical investigations (HCM Risk-Kids) that allows clinicians to calculate individualised estimates of 5-year risk for the first time. However, the model remains imperfect and fails to capture dynamic changes in risk over time. Additionally, no risk prediction model exists for other types of paediatric cardiomyopathies. Combined, these evidence gaps prevent aetiology specific management and prognostication for young patients with heart muscles disease.

One of the challenges of rare disease research is having access to sufficient patient numbers. This project will use two different childhood onset cardiomyopathies, hypertrophic cardiomyopathy (HCM) and Dilated cardiomyopathy (DCM), as exemplar conditions to explore how machine learning methods could improve the diagnosis, management and outcome prediction of children with rare diseases. Taking advantage of the largest clinical population of childhood HCM in the world, this project represents an opportunity to explore machine learning and deep learning algorithms for rare disease outcome prediction as well as the use of deep learning models such as Generative Adversarial Networks (GANs) for data augmentation. If successful, the techniques developed could be translated to other rare diseases.

Study population

Cohort 1: Childhood Dilated Cardiomyopathy

A deeply-phenotyped cohort of young people (0-<25 years) with a diagnosis of DCM will be recruited from Inherited Cardiovascular Diseases and Heart failure clinics at Great Ormond Street Hospital and St Bartholomew's Hospital through collaborations with Professor Juan Pablo Kaski (Great Ormond Street Hospital), Professor Michael Burch (Great Ormond Street Hospital) and Professor Perry Elliott (St Bartholomew's Hospital). Patients will be eligible for inclusion if they meet diagnostic criteria for DCM (defined as the presence of LV dilatation and global or regional systolic dysfunction unexplained by abnormal loading conditions).

Cohort 2: Childhood Hypertrophic Cardiomyopathy

In 2016, the principal supervisor and his team established the International Paediatric Hypertrophic Cardiomyopathy Consortium (IPHCC) to develop the first paediatric specific risk prediction model for SCD in childhood HCM using routinely collected, non-invasive baseline clinical investigations (HCM Risk-Kids). The consortium currently includes data on 1765 patients with HCM ≤ 16 years from 45 expert centres worldwide, including serial, repeated measure data on clinical and outcome parameters. The original IPHCC cohort will be expanded to include additional baseline and longitudinal information as well as recruiting newly diagnosed children.

Aims/Objectives:

- 1) Explore the use of machine learning to identify phenotypically similar groups of patients with paediatric heart muscle disease meaning care can be personalised even if the aetiology is unknown
- 2) Explore the use of machine learning to describe phenotype progression in paediatric heart muscle disease to allow clinicians to accurately prognosticate and counsel families
- 3) Develop a prognostic model for paediatric heart muscle disease using machine learning and serial non-invasive clinical data to allow clinicians to accurately prognosticate and counsel families

Project workstreams:

WS1: Identify phenotypic clusters in patients with childhood onset cardiomyopathy using supervised and unsupervised machine learning techniques

Hypothesis Machine learning approaches can identify phenotypically homogenous groups of patients that may share a similar aetiology and benefit from similar treatments or management strategies

Study design and analysis

The baseline morphological and functional characteristics of patients with paediatric onset DCM (*cohort 1*) will be explored to identify phenotypic clusters. Detailed phenotyping data will be used including routine clinical investigations (eg 12 lead ECG, echocardiography, magnetic resonance imaging); blood biomarkers (eg Pro-BNP, Troponin I) and genomic data. Different supervised (eg Random Forest) and unsupervised (eg hierarchical clustering) methods will be used to identify phenotypic clusters and results compared. To identify missing heritability and detect novel-disease causing genes, phenotypic subgroups will be integrated with whole exome data.

Expected value This study will identify phenotypically-similar groups of patients that may benefit from particular clinical investigations, management strategies and therapies meaning care can be personalised even if the aetiology is unknown. Such strategies could be extended to other rare diseases.

WS2: Determine the relationship between time-varying phenotypic characteristics and adverse outcomes using Long Short-Term Memory Networks (LSTM)

Hypothesis Machine learning approaches can identify phenotypically homogenous patients that share similar disease trajectories (progression or resolution).

Study design and analysis

The prevalence, evolution and natural history of pre-selected phenotypic characteristics (eg left ventricular wall thickness, left ventricular outflow tract obstruction, systolic and diastolic dysfunction) will be explored in a cohort of children with hypertrophic cardiomyopathy (cohort 2). These results will enable us to understand the patient population and describe the development and progression (or resolution) of the clinical phenotype of childhood HCM. The primary outcome of interest will be Major Arrhythmic Cardiac Event (MACE; defined as SCD or an equivalent event, including resuscitated cardiac arrest, sustained ventricular tachycardia with haemodynamic compromise or appropriate ICD therapy for ventricular tachyarrhythmia). The relationship between baseline and time-varying risk predictors and the primary time-to-event outcome (MACE) will then be explored using LSTM networks, which are a type of recurrent neural network for sequence prediction problems to learn long-term dependencies. Additional time-series regression approaches (Auto-regressive integrated moving average [ARIMA]) can be explored to compare results arising from the LSTM network approach.

Expected value This study will provide an understanding of disease specific progression allowing clinicians to accurately prognosticate and counsel families. Such strategies could be extended to other rare diseases.

WS3: Develop a prognostic model for SCD in childhood HCM using serial non-invasive clinical investigations

Hypothesis Different machine learning approaches can be used to improve risk prediction in rare diseases using routinely available clinical investigations.

Study design and analysis

Patient assessment and clinical outcomes: Serial anonymised, non-invasive, clinical data will be collected from baseline and yearly over follow up from cohort 2. Primary outcome will be Major Arrhythmic Cardiac Event (MACE) as in WS2.

(a) Data augmentation using Generative Adversarial Networks (GANs).

Given that the student will be working on a rare disease cohort, they will investigate the potential of a deep learning approach, GANs, to generate additional synthetic data for model development.

(b) Evaluation of machine learning approaches for rare event prediction.

The primary outcome (MACE) is a rare event among children with HCM, meaning that there will be class imbalance of the target feature. The student will explore different approaches for rare event prediction:

- 1) Synthetic Minority Oversampling Technique (SMOTE) – to generate new instances of the rare events. This new data will be used as input for supervised tree-based algorithms (e.g., random forest and XGBoost). The student will learn to impute missing values, create one-hot encoding variables, perform feature selection using recursive feature elimination, train and tune models and perform hyperparameter tuning to optimise the model.
- 2) Downsampling and upweighting of the majority class, as an alternative to SMOTE.
- 3) Exploring penalised algorithms such as BRC deleted following this as it exceeded the word limit

Timeline

	Y1				Y2				Y3			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Completion of data collection												
WS1												
WS2												

WS3(a)													
WS3(b)													
WS3(c)													
Write-up and PhD submission													

References:

1. Arbelo E, Protonotarios A, Gimeno JR, Arbustini E, Barriales-Villa R, Basso C, et al. 2023 ESC Guidelines for the management of cardiomyopathies. Eur Heart J. 2023.
2. Norrish et al. Development of a Novel Risk Prediction Model for Sudden Cardiac Death in Childhood Hypertrophic Cardiomyopathy (HCM Risk-Kids). JAMA Cardiol 2019;4:918-92
3. Norrish et al. External validation of the HCM Risk-Kids model for predicting sudden cardiac death in childhood hypertrophic cardiomyopathy. Eur J Prev Cardiol 2022;29 :678-686
4. Krittanawong, C., Johnson, K.W., Rosenson, R.S., Wang, Z., Aydar, M., Baber, U., Min, J.K., Tang, W.W., Halperin, J.L. and Narayan, S.M., 2019. Deep learning for cardiovascular medicine: a practical primer. European heart journal, 40(25), pp.2058-2073.
5. Lan, L., You, L., Zhang, Z., Fan, Z., Zhao, W., Zeng, N., Chen, Y. and Zhou, X., 2020. Generative adversarial networks and its applications in biomedical informatics. Frontiers in public health, 8, p.164.

Project 8

Project title: Precision Diagnosis for Congenital Anomalies – using AI & leveraging multimodality data to help counsel parents appropriately in rare diseases.

Supervisory Team:

Dr Susan Shelmerdine, Great Ormond Street Hospital
 Prof Ivana Drobnjak, UCL Department of Computer Science
 Prof Owen Arthurs, UCL Great Ormond Street Institute of Child Health

Contact: Susan Shelmerdine - susan.shelmerdine@gosh.nhs.uk

Background:

Skeletal dysplasias are a heterogenous group of congenital bone and cartilage disorders with genetic etiology. There are currently 461 recognised diseases that fall into this category. Individually each one is rare, but collectively these disorders are common, and unfortunately a subgroup of these diseases are lethal - either in utero or very soon after delivery. A correct diagnosis can help establish the likelihood for recurrence in future pregnancies and potentially enable improved prenatal identification, and better family planning guidance. At present, post-mortem investigations for skeletal dysplasia across the country is haphazard and there are a lack of qualified pathologists and radiologists trained for this.

Currently, when a pregnancy ends in termination or miscarriage, a diagnosis of skeletal dysplasia is made at post-mortem using a variety of different methods as follows: review of the maternal and obstetric history including any prenatal imaging is reviewed, followed by an X-ray of the whole fetus (called a babygram). Specialist paediatric radiologists (trained to recognise features of skeletal dysplasias) review the imaging and where possible make a diagnosis or in some cases, further imaging with micro-CT (only at GOSH) and genetic testing will be performed. There are several issues with this process – nationally there are a lack of paediatric radiologists with skeletal dysplasia training to review the Xrays, GOSH is the only centre in the UK that can provide high resolution micro-CT of fetuses to give detail on the internal anatomy of the fetuses (as many have associated renal or cardiac anomalies), and genetic analysis is expensive and time consuming to do.

If an AI tool, trained on thousands of prior datasets with all this multimodal information, were available it could generate a list of plausible differential diagnosis and likely recurrence rates for future pregnancies from review of the

initial fetal X-ray (babygram) thereby creating a streamlined, cost-effective and efficient way to generate personalised results for families who have experienced a pregnancy loss.

Aims/Objectives:

To evaluate whether an AI, trained on thousands of multimodal datapoints, can be used to generate personalised predications on type of lethal skeletal dysplasia from babygrams, and provide a likely recurrence rate in future pregnancies to aid improved genetic counselling and future family planning.

Methods:

All GOSH cases referred for fetal post mortem investigation over the last 10 years will be eligible for inclusion, and already provide informed consent for their results to be used for education and research purposes.

Imaging data (radiographs and where applicable, micro-CT images) will be collected, together with associated information (e.g. maternal, obstetric, pathological clinical notes) and autopsy investigations including placental histology and genetic information. All data will be de-identified before uploading to an XNAT server hosted by GOSH DRIVE. A secure data sharing link will facilitate imaging and clinical data from other centres where maternal or obstetric history/imaging is not available in the GOSH health records.

Computer and health data scientists at UCL will work on the de-identified data to create a multimodal AI algorithm that collates all the information provided from different sources and produce a list of possible differential diagnoses based on the fetal babygram data (as the input). This algorithm will then be tested and evaluated on a subset of unseen data from the main dataset for accuracy and also against international world experts in skeletal dysplasia for the plausibility of outputs prior to trailing its implementation in clinical practice locally as an adjunct to usual practice.

Our dataset will include both normal and those of skeletal dysplasias to ensure the AI algorithm is not biased and can recognise differences between them. Incomplete data will be excluded. Key statistical methods employed will include those for diagnostic accuracy for correct skeletal dysplasia diagnosis.

Timeline:

Year 1: Set up necessary data sharing agreements between GOSH and UCL. Complete data collection, standardisation, de-identification of electronic and imaging data records. Retrieval of genetic and autopsy (if available and applicable) information. Set up a secure data research environment for storage of data.

Year 2: Develop a predictive AI model to provide differential diagnoses for skeletal dysplasias from babygram, and suggest likelihood factor for recurrence. Conduct a multireader multicase study amongst internationally recognised experts in skeletal dysplasias to understand likely pitfalls of the algorithm based on plausibility of outputs generated.

Year 3: Implement the AI solution into clinical practice and evaluate real-world prospective outcomes and gain feedback from fetal medicine experts, paediatric pathologists, geneticists and PPIE regarding applicability and acceptability. Discussions to involve ethics of use and explain-ability of results, including any necessary changes to the user interface.

Collaborations:

We have strong established collaborations already from our prior work on children's imaging and AI with the following informatics/ innovations and computer science groups at UCL and GOSH:

UCL GOS ICH Radiology – Dr Ali Calder (dysplasia expertise), Dr Susan Shelmerdine (AI expertise)

UCL GOS ICH Genetics – Dr Alison Male (genetics expertise)

UCL GOS ICH Pathology – Dr Ciaran Hutchinson, Thivya Sekar (fetal autopsy)

GOSH DRIVE – Prof Neil Sebire (AI expertise, fetal autopsy expertise)

UCL Advanced Research Computing Centre – Prof Graca Carvalho (computer science expertise)

UCL CMIC – Prof Ivana Drobnjak (AI, mathematical modelling expertise)

In terms of clinical collaborations for international expertise and knowledge for how this innovative would could be integrated into clinical practice, help with focus group work or multi-centric data collections in the the future, we have established links as a group to the following societies:

Skeletal Dysplasia Group (British society for skeletal dysplasias)

ISDS – International skeletal dysplasia society

One of the subsidiary supervisors (OA) is the chair for the European Society of Paediatric Radiology (ESPR) postmortem imaging task force

The main supervisor (SS) is the chair of the ESPR AI taskforce.

Plans for patient and public involvement and engagement for the project/student:

We have extensive experience with PPIE in challenging topics including miscarriage, post mortem imaging and stillbirth. Over 10 years of close working with key third sector organisations and main charities (SANDS, The Miscarriage association, Lullaby Trust, Daddies for Angels) we have integrated PPIE into all of our research including co-authoring BMJ articles (e.g. Arthurs OJ, Bevan C, Sebire NJ. Less invasive investigation of perinatal death. *BMJ*. 2015 Jul 8;351:h3598).

The PhD student will be embedded into PPIE from the beginning of the study, with representative members from these groups invited to form part of our Study PPIE Steering Committee. They will inform how the study should be reported, review lay summaries and what patients and families may want from an AI tool when implemented in practice, including whether informed consent should be provided prior to use of an AI in the future.

The student will be assisted by our research imaging team in setting up the PPIE steering group and in all conversations and interviews with the group. The GOSH PPIE lead is also experienced in dealing with non-medical lay members of public and available to provide support where needed. We have co-designed recruitment material, dissemination animations on You Tube to help understand our research (e.g. <https://www.youtube.com/watch?v=nV16GazlcGA>) and lay summaries, and our PPIE activities are well funded from existing grants. We encourage our students and PPIE members to access the NIHR UCLH Biomedical Research Centre for support and training programmes, and support PPIE members to attend training and conferences.

References:

1. Nishimura G, Handa A, Miyazaki O, Tsujioka Y, Murotsuki J, Sawai H, Yamada T, Kozuma Y, Takahashi Y, Ozawa K, Pooh R, Sase M. Prenatal diagnosis of bone dysplasias. *Br J Radiol*. 2023 Jul;96(1147):20221025. doi: 10.1259/bjr.20221025
2. Stembalska A, Dudarewicz L, Śmigiel R. Lethal and life-limiting skeletal dysplasias: Selected prenatal issues. *Adv Clin Exp Med*. 2021 Jun;30(6):641-647. doi: 10.17219/acem/134166.
3. Groza T, Hunter J, Zankl A. The Bone Dysplasia Ontology: integrating genotype and phenotype information in the skeletal dysplasia domain. *BMC Bioinformatics*. 2012 Mar 26;13:50. doi: 10.1186/1471-2105-13-50.
4. Rassmann S, Keller A, Skaf K, Hustinx A, Gausche R, Ibarra-Arrelano MA, Hsieh TC, Madajieu YED, Nöthen MM, Pfäffle R, Attenberger UI, Born M, Mohnike K, Krawitz PM, Javanmardi B. Deeplasia: deep learning for bone age assessment validated on skeletal dysplasias. *Pediatr Radiol*. 2024 Jan;54(1):82-95. doi: 10.1007/s00247-023-05789-1.
5. Tang J, Han J, Xie B, Xue J, Zhou H, Jiang Y, Hu L, Chen C, Zhang K, Zhu F, Lu L. The Two-Stage Ensemble Learning Model Based on Aggregated Facial Features in Screening for Fetal Genetic Diseases. *Int J Environ Res Public Health*. 2023 Jan 29;20(3):2377. doi: 10.3390/ijerph20032377

Project 9

Project title: Using big data to better define disease types and predict outcome in childhood myositis.

Supervisory Team:

Prof Lucy R Wedderburn, UCL Great Ormond Street Institute of Child Health

Prof Mario Cortina-Borja, UCL Great Ormond Street Institute of Child Health

Dr Merry Wilkinson, UCL Great Ormond Street Institute of Child Health

Contact: Merry Wilkinson Meredyth.wilkinson.14@ucl.ac.uk, Lucy Wedderburn l.wedderburn@ucl.ac.uk, Mario Cortina-Borja - m.cortina@ucl.ac.uk

Background:

Juvenile Dermatomyositis (JDM) is a rare serious childhood autoimmune disease presenting with muscle and skin inflammation, which can affect any organ, and carries burden of serious morbidity and even mortality. Current treatment involves long-term immunosuppression (steroids, methotrexate), but is not targeted, since specific pathological mechanisms of JDM are unknown(1). AT GOSH/ICH we host the largest prospective cohort study of JDM (so-called JDCBS;17 UK centres), with linked longitudinal biological and clinical data (mean follow-up 7.6yrs) on 700 cases of childhood myositis.

Work of Dr Merry Wilkinson (**MW**, previous BRC Catalyst fellow, proposed co-supervisor) recently identified that in addition to the well-documented interferon signature in JDM, there is a strong signal of **dysfunctional mitochondria**, detected by transcriptional analysis of immune cells, which is most marked in blood monocytes(2). Our data suggest that oxidised mitochondrial DNA is released and detected intracellularly (via nucleic acid sensing or TLR ligation). Interestingly this signal **does not resolve on immunosuppression**, even when IFN signature normalises (Wilkinson et al 2023(2)). MW has replicated the MGS signature in a large validation JDM cohort ($n=60$) and started spatial transcriptome analysis on muscle tissue, from matched patients. We have defined the key leading genes which best represent this Mitochondrial-gene-signature (MGS) and optimised an NCounter (Nanostring) high-throughput assay for measuring this. Pilot results show a high degree of reliability.

There is an urgent need in GOSH/ICH for capacity building in data science: this is an added-value aspect of the project. We address this need by training a biomedical/data scientist in both the omics and clinical data in this project and in current modern statistical and Machine Learning (ML) methods. The student will become expert in analysis of a range of high-dimensional, longitudinal data and will apply methods that are well-suited for this task. The project will cover the two central aspects of modern statistical and machine learning methods, namely estimation and prediction/validation. The student will be trained to an advanced level of programming in the R statistical language, and will benefit from the supervisory team's links with the Royal Statistical Society, colleagues at PPP and UCL Statistical Science Department, as well as the growing community of spatial tissue biologists at UCL and beyond. **All recruited patients, clinical data samples and approvals required for this project are already in place, through JDCBS.** Student will be well supported by the JDCBS Study team and wider collaborative myositis scientific network.

Aims/Objectives/Methods:

AIM 1: Investigate if the mitochondrial gene signature can characterise groups of patients, and predict disease course or outcome.

Advisor to this analysis: Professor Cortina-Borja, input from Professor Sebire. The student will clean, quality-assure and curate a detailed longitudinal clinical dataset for analysis; this will include demographics, MSA autoantibody status, disease activity scores, specific severe disease features, and medication. Missing values will be dealt with using multiple imputation methods, and inference will be performed in the Bayesian framework(3,4).

1.1 Trajectory analysis In JDM patients in whom expression of mitochondrial-gene-signature (MGS) in PBMC and monocytes at onset is already quantified. Student will build and deliver an analysis plan to ask if MGS predicts course, disease severity or features in JDM. Visualisations based on methods focused on reduction of dimensionality will play an important role. Linear and nonlinear mixed-effects regression models will be fitted to analyse trajectory scores for each measure per patient, providing a trajectory of disease course over time. Given relatively low number of patients with data on a large number of variables, avoiding collinearity and overfitting by focusing on building parsimonious models will be a crucial aspect of the analyses.

1.2 In a large validation cohort of JDM (100) and controls (20) with MGS score (nCounter) the student will build statistical models to correlate disease activity, course and features with MGS score to stratify JDM subtypes and outcome. This new cohort of patients have longitudinal samples (typically 3-5 time points over 4 years) with

matched complete clinical data. Analysis will test whether MGS score can predict severe features, outcome, treatment response or disease flares.

- 1.3 **Machine learning:** in parallel with the curated JDCBS clinical data, the student will build and validate ML models(5) based on the full GOSH DRE data, directly pulled from the EPIC record which we have previously shown is more extensive (Hamilton *et al* in prep). These models will be useful to predict outcomes for future individual patients given their trajectories over many explanatory variables, and to refine these models and their predictions as new observations are obtained.

AIM 2: investigate whether mitochondrial dysfunction is mirrored in disease tissue.

We have generated world-first spatial transcriptomic (NanoString-GeoMx) and single cell tissue sequencing (NanoString-CosMx) datasets, in 3 JDM and 3 child healthy control muscle tissues. Pilot analysis suggests that again there is mitochondrial dysfunction, clearly detectable in muscle tissue in JDM compared to healthy muscle. Advisor to this analysis: Professor Croft collaborator, and team with input from Professor Castellano.

- 2.1 Define depth of information that can be obtained from the NanoString **spatial-omic data**, developing a pipeline in R to clean, quality assure and curate both datasets; this will include adapting methods for outlier detection in high-dimensional datasets and implementing appropriate QC procedures;
- 2.2 Develop PCA and U-maps to compare JDM to control muscle, perform **pathway/gene set enrichment analysis**; develop and run pipelines to analyse cell types and cell-cell interactions/nearest neighbour analysis;
- 2.3 Analyse **both spatial datasets** (GeoMx, CosMx) comparing JDM to healthy muscle to establish if mitochondrial dysfunction is evident in muscle fibres, and in inflammatory infiltrate;
- 2.4 Perform **focused analysis** on tissue macrophages (CD68+) (CosMx data) to test hypothesis that the abnormalities we have seen in blood are present in tissue.

AIM 3: data integration and translation.

Student will design strategies to integrate spatial-omic data with RNA-seq data

- 3.1 **Develop methods** in R to visualise, analyse, combine and compare the different transcriptomic data-sets
- 3.2 **Corresponding cell types** in CosMx tissue data will be annotated using scRNAseq data-from JDM blood monocytes and GeoMx data from JDM muscle, defining gene modules that identify cell populations in the scRNAseq/GeoMx data that also appear in the CosMx tissue data. Spatial analysis will be performed to localise disease-associated mitochondrial gene signature to specific cell populations in muscle tissue, and neighbour-proximity analysis to define cellular niches in muscle and identify disease critical cellular interactions.
- 3.3 **Translational potential:** Test if MSG blood signature is mirrored in CD68+macrophages in muscle, and whether MGS biomarker can be used to reflect muscle pathology and disease activity; if robust such biomarker could be used in clinical trials we will propose.

Working with world first datasets in this rare complex disease, the student will gain a thorough training in handling and analysis of many types of high-dimensional data, and how results would translate to patient benefit.

Timeline: AIM1:months 0-18; AIM2:months 6-24; AIM3:months 18-30. Data integration, papers, months 24-32. Upgrade planned at ~12m; last 6 months writing up.

Gantt Chart		year	1	2	3
	Initial training: R, statistical and analysis methods				
AIM 1	Investigate if mitochondrial gene signature characterises groups of patients, predicts disease course or outcome				
	Trajectory analysis using baseline RNAseq data				
	Disease course and course analysis vs nCounter MGS data				
	Machine learning analysis				
AIM2	Investigate whether mitochondrial dysfunction is seen in muscle tissue				
	Curate and clean /QC Spatial transcriptomic data, initial UMAPs				
	Test mitochondrial signature (GeoMx, CoxMx) JDM vs control muscle				
	Targeted analysis of tissue macrophages, single cell spatial data				
AIM 3	Data integration, translational potential				
	Build models for intergation of clinical, blood and muscle data				
	Tissue atlas spatial data analysis, neighbour analysis, cellular niches				
	Analyse blood vs muscle data and model against clinical outcomes				
Integration	Data analysis, paper writing/revisions				
PPIE	Regular focus group meetings, presentations and interaction				
PhD steps	PhD milestones: Thesis committee, upgrade, write up and viva			upgrade	write up

Collaborations:

GOSH/ICH/UCL: Professor Sebire (BRC-ACHI Theme lead), Dr Papadopoulou(GOSH JDM Clinic lead), Professor Castellano (UCL-Genomics)

University of Birmingham: Professor Croft and bioinformatics team, spatial transcriptomics analysis

Liverpool AlderHey Children's Hospital : Dr McCann, expert on JDM clinical outcome analysis.

Plans for patient and public involvement and engagement for the project/student: student will discuss the project regularly with our active PPIE-JDCBS group, sharing results and discussing implications for patients/families. They will provide updates at National JDCBS research day, in our regular patient newsletter, and websites. They will have many opportunities to discuss progress with young people/families through The Centre for Adolescent Rheumatology at UCL, and Young scientists days. All members of the JDM team receive PPIE training.

References:

1. Papadopoulou, C., Chew, C., Wilkinson, M.G.L.,McCann, L. and Wedderburn, L.R. (2023). Juvenile idiopathic inflammatory myositis: an update on pathophysiology and clinical care *Nature Rev Rheumatol* 19(6):343-362
2. Wilkinson M. G. L., Moulding, D., McDonnell, T. C. R., Orford, M., Wincup, C., Ting, J., Otto, G. W., Restaudi, R., Kelberman, D., Papadopoulou, C., Castellano, S., Eaton, S., Deakin, C. T., Rosser, E. C., Wedderburn, L. R. (2023) The role of CD14+ monocyte-derived oxidised mitochondrial DNA in the inflammatory interferon type 1 signature in Juvenile Dermatomyositis. *Annals Rheum Diseases* 82 (5) 658-669
3. James G., Witten D., Hastie T., Tibshirani R (2023). An Introduction to Statistical Learning : with Applications in R. New York: Springer, 2nd edition.
4. Shoop-Worrall, S.J.W., Hyrich K.L, Wedderburn L.R., Thomson, W., Geifman N. on behalf of CAPS and the CLUSTER Consortium (2021). [Patient-reported wellbeing and clinical disease measures over time captured by multivariate trajectories of disease activity in individuals with juvenile idiopathic arthritis in the UK: a multicentre prospective longitudinal study](#) *Lancet Rheumatol*, 3(2):e111-e121
5. Gelman A., Hill J., and Vehtari A (2020). Regression and Other Stories. Cambridge: Cambridge University Press

Project title: Multimodal Artificial Intelligence for the Detection and stratification of Necrotising Enterocolitis in premature born infants (MAIDNEC)

Supervisory Team:

Prof Simon Eaton, UCL Great Ormond Street Institute of Child Health (primary)

Dr Evangelos Mazomenos, UCL Department of Med Phys & Biomedical Eng (subsidiary)

Contact: Evangelos Mazomenos - e.mazomenos@ucl.ac.uk

Background:

Necrotising Enterocolitis (NEC) is a severe gastrointestinal inflammatory condition affecting neonates, most commonly premature infants although up to 25% is observed among full-term babies¹. Nearly 12% of infants born weighing less than 1500 g will develop NEC, with overall mortality between 18-63% and major long-term developmental complications (inflammatory strictures, bowel obstruction, poor neurodevelopment)². In 2011, the National Confidential Enquiry into Perioperative Death (NCEPOD) in Children's Surgery found that NEC accounted for a quarter of deaths in children following surgery. A recent prospective study by the British Association of Paediatric Surgeons Congenital Anomalies Surveillance System (BAPS-CASS) on 40 babies receiving surgery for NEC found that overall mortality was 18% at 28 days and 29% at one year. Almost half of babies born with NEC in England die in the first year of life³. NEC requires specialist management care at a tertiary NICU with a multidisciplinary team. The economic cost of NEC is high, accounting for approximately 19% of neonatal expenditures and an estimated \$5-6 billion per year for hospitalizations in the United States (there is no readily available UK based data for this).

Medical management of NEC includes gut rest, intravenous nutrition (Total Parenteral Nutrition) and antibiotics. Many infants though, will require surgical intervention (intestinal resections and stoma formation), with severe cases being an emergency. Early diagnosis from abdominal X-rays (AXR) and subsequent surgical referral are vital as delays can negatively impact outcomes. Confounding factors such as variability in presentation and similarities to other conditions (neonatal sepsis) pose great challenges to radiologists, paediatric surgeons, and neonatologists for correct diagnosing and treatment decisioning. This is compounded by the challenges of interpreting neonatal imaging, particularly for non-specialists, while support from neonatal radiologists can be limited especially out of hours. Furthermore, not all Neonatal units have access to Paediatric Surgeons on site (20 of the 47 NICUs do not have a co-located paediatric surgical centre). This results in delayed diagnosis and patient transfer to a paediatric surgical centre, ultimately delaying initiation of medical treatment or surgical intervention. This has two consequences: 1) patients with NEC are diagnosed and treated late, 2) patients without NEC (i.e. patients with sepsis or other pathology) unnecessarily undergo medical management for NEC due to the inability to differentiate between the two conditions at early stages. In many situations, extended management with Parenteral Nutrition is followed which is both costly and may affect liver function. There is consensus that earlier identification of NEC will improve treatment and outcomes, however even if NEC is identified early, it is often challenging to identify which patients would benefit from longer medical treatment and which patients require earlier surgical intervention. At present the only absolute indication for surgery in NEC is a perforation and all other indications rely on the judgement of the clinician based on serial assessments of the patient's condition and the progress of the disease.

It has been identified that scoring and classification systems would be useful to diagnose the disease, predict its course and help identify ideal moments for surgical intervention. The most widely known system is Bell's classification of NEC which is based on general clinical parameters and radiologic findings. It ranges from stage 1 (suspected disease) to stage 3b (Severe disease with evidence of perforation). This system was devised in 1978 and has been a useful tool for clinicians, however it is somewhat outdated now. Further attempts at creating a more modern classification/scoring system have not successfully gained traction. With the technological advances of this decade, one potential solution to this problem would be the exploration of Artificial Intelligence (AI). Specifically in radiography, it has been demonstrated that AI solutions can achieve expert-level signs detection leading to full clinical diagnosis in mammography images⁴ and chest radiographs⁵.

Aims/Objectives:

The initial objective (year 1) is to develop AI methods to process abdominal X-rays (AXR) and detect NEC against confounding conditions (e.g. sepsis) of non-NEC cases. The main research challenge is to structure and train AI models that are capable to extract and detect the subtle characteristic patterns of NEC in radiographs, effectively distinguish it from confounders. The long-term goal (years 2-3) is to devise AI methods for automatically stratifying the severity of NEC and identifying patients that require urgent surgical intervention as opposed to cases that can be managed with medical treatment alone (conservative management). To achieve this, the project will pioneer multimodal learning, whereby additional to AXR data sources (physiological and clinical data) will be used for developing AI technology for NEC severity stratification. To increase trustworthiness of the decision support system and promote adoption of the developed technology by end-users (medical personnel), the developed AI models will embed Explainable AI (XAI) methods to rationalise machine understanding with respect to human perception and provide causal interpretation of system outputs. At its completion, the project will deliver an AI-based decision support system for early NEC diagnosis and severity stratification validated on prospective data with the potential to contribute towards reducing NEC-associated long-term morbidity and mortality. We expect mortality of surgical NEC cases (currently at ~50%) to be reduced by 10%. Overall, the proposed research creates opportunities for broad scientific impact both in the technical (specialised AI models combining heterogenous data sources) and clinical domain (NEC diagnosis and management, deeper understanding of disease progression/severity via novel multimodal features).

Methods:

Ongoing investigations are supported by a collaboration between GOSH/WEISS (registered study GOSH CRAC, Protocol Number: 21DS17) and a database of 800 cases (450 NEC, 350 no-NEC/control). Ethical approval for multimodal prospective data collection and research agreement are in place. Manual annotation and data curation is taking place to facilitate further development of AI methods. Data sharing and GPU computing equipment are available through WEISS. Data are stored in secured WEISS-owned space using the XNAT protocol for anonymisation, sharing, storage and management between GOSH and WEISS. Technical support is provided by UCL ARC. The project's database is jointly managed by the investigators team. The project's methodology is summarized in the work packages (WP) below:

WP1 - Automated detection and discrimination of NEC in AXR imaging (2024-2025)

Novel AI approaches for AXR analysis based on Fine-Grained Visual Classification (FGVC), a recently proposed object detection method that detects and classifies highly specific and detailed categories (different types of flowers or cars) of natural images will be developed. FGVC presents many similarities to NEC diagnosis, which requires radiologists to identify subtle, occasionally difficult to discriminate, findings in AXRs. We will thus formulate NEC detection/classification problem as a weakly-supervised FGVC task and propose customised AI architectures to address it.

WP2 – Multimodal learning for NEC severity stratification (2025-2026)

Automated NEC stratification and treatment decisioning will be carried out with novel AI models structured to process and learn from heterogenous medical data (imaging-AXR, time series-physiological data/lab tests, text-medical reports). Disease stratification labelling for driving the development of AI technology will be carried out using established clinical classification systems (e.g. Bell's classification of NEC established in 1978) based on general clinical parameters and radiologic findings. Bell's system ranges from stage 1 (suspected disease) to stage 3b (severe disease with evidence of perforation).

WP3 –Integration and evaluation into the clinical NEC workflow (2026-2027)

The final WP will integrate outputs from WP1 and WP2 into a single AI system. This will be deployed alongside the standard NEC clinical workflow to evaluate the benefits that automated NEC detection and stratification yields with prospective data. A user-study investigation will take place on the best strategies to present system outputs, integrating explainable AI components, to medical staff thereby optimising further clinical adoption. This will initially BRC deleted following text over the word limit

References:

6. Neu J, Walker WA. Necrotizing enterocolitis. *New Eng J Med*. 2011;364:255–264.

7. Eicher C, Seitz G, Bevot A, et al.. Surgical management of extremely low birth weight infants with neonatal bowel perforation: a single-center experience and a review of the literature. *Neonatology*. 2012;101(4):285–292.
8. Allin B, Long AM, Gupta A, Knight M, Lakhoo K; British Association of Paediatric Surgeons Congenital Anomalies Surveillance System Necrotising Enterocolitis Collaboration. A UK wide cohort study describing management and outcomes for infants with surgical Necrotising Enterocolitis. *Sci Rep*. 2017 Jan 27;7:41149.
9. McKinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafian H, Back T, Chesus M, Corrado GS, Darzi A, Etemadi M, Garcia-Vicente F, Gilbert FJ, Halling-Brown M, Hassabis D, Jansen S, Karthikesalingam A, Kelly CJ, King D, Ledsam JR, Melnick D, Mostofi H, Peng L, Reicher JJ, Romera-Paredes B, Sidebottom R, Suleyman M, Tse D, Young KC, De Fauw J, Shetty S. International evaluation of an AI system for breast cancer screening. *Nature*. 2020 Jan;577(7788):89-94.
10. Tiu E, Talius E, Patel P, Langlotz CP, Ng AY, Rajpurkar P. Expert-level detection of pathologies from unannotated chest X-ray images via self-supervised learning. *Nat Biomed Eng*. 2022 Dec;6(12):1399-1406.

Project 11

Project title: AIDE: an Artificial Intelligence aiDed clinical support system on Edge for emergency transport of critically ill children

Supervisory Team:

Prof Mark Peters, UCL Great Ormond Street Institute of Child Health (primary)
Dr Kezhi (Ken) Li, UCL Institute of Health Informatics (subsidiary)
Dr Philip Knight, Great Ormond Street Hospital (subsidiary)

Contact: Kezhi (Ken) Li - ken.li@ucl.ac.uk and Mark Peters mark.peters@ucl.ac.uk

Background:

Paediatric intensive care is a highly specialised, resource-intensive NHS England commissioned service available at just 25 UK hospitals. Paediatric retrieval teams currently transport over 5000 critically ill children to PICUs each year, making up half of all emergency admissions to UK PICUs (1). Retrieved children are much sicker than similar emergency PICU admissions from wards; they face significant morbidity, frequently requiring prolonged invasive support for vital organs such as the lungs, heart and kidneys. While overall PICU mortality is low (3-4%), transported admissions have worse mortality compared to other admissions (8%) (2). In those who survive PICU, the after-effects of critical illness can often be severe and pervasive (affecting physical, functional, developmental, emotional and psychosocial domains), resulting in significant long-term health and social care costs (3). Improving outcomes for children retrieved for intensive care is an urgent national priority.

Unlike hospital ICUs, however, computers are rarely used during retrieval or to help retrieval clinicians make important decisions such as whether a child is getting sicker and whether they will respond to particular treatments. These decisions can be hard to make in sick children with a wide range of ages and diagnoses, especially in the fast-paced and high-stakes retrieval environment. Research from other intensive care areas suggests that AI software offers great potential to help clinicians make better decisions and improve patient outcomes.

As a group of researchers and doctors, our aim is to develop AIDE: an AI-aided clinical support system to help clinicians manage multiple risks under extreme environments during emergency transport of critically ill children. It allows clinicians to spot early deterioration promptly and start the right treatment at the right time.

In previous grant funded pilot work (UCL - Rosetrees Trust Healthcare Engineering Award), we have cleaned the multi-modal time series, extracted features from physiological patterns, and explored correlations between 'physiotypes' and clinical deterioration from the unique dataset (waveform and numerical) captured by first-of-kind telemetry software (Medivue) since 2017 implemented in Children's Acute Transport Service (CATS). We will make use of these results, and develop a resource-efficient system that can estimate real-time personalised risks with clinical explanations. This

system can provide more effective early warnings for clinical deterioration compared to traditional threshold-based warning systems with a better precision-recall curve.

Aims/Objectives:

In our pilot study we have prepared raw data into a well-formatted dataset, and extracted underlying patterns of physiological measurements ('physiotypes') within- and between-groups of patients through statistical and time-series analysis. In this follow-on project, our objectives include:

- 1) Develop prototype prediction models using supervised learning to evaluate the risk of patient deterioration by exploring correlations between 'physiotypes' and deterioration (e.g., acute hypoxia), interventions (e.g., fluid resuscitation) and outcomes (e.g., in-hospital mortality).
- 2) Develop a real-time risk estimation system with clinical explanations based on their personal characteristics (age, gender, medical history, etc) and the trend of vital signs (such as heart rate, oxygen saturations and blood pressure).
- 3) Tackle the bias in AI to mitigate the discrimination for improved intensive care (e.g. ethnicity subgroup and imbalance data)
- 4) Design and build a resource-efficient (in terms of computation and power) edge computing API and UI, making the system user-friendly and ready to use for the current software.

Methods:

The combination of our previous experience and recent advances in Edge computing puts us in a strong position to bring the power of AI into the ambulance: Our team has collaborated since 2018 to develop prototype AI-based risk models to predict the occurrence of a critical incident during transport and death in PICU by tapping into high-resolution vital signs monitoring data (numeric data every 1 second on heart rate, blood pressure, oxygen saturation, end-tidal carbon dioxide and temperature; waveform data at 120-300 Hz from ECG, capnograph, arterial pressure and oxygen saturation) collected from nearly 2000 children transported by the CATS retrieval service at GOSH.

Various deep learning techniques will be applied to the clinical multimodal data to extract useful information from the real-time time series data, and recognise the useful 'physio-type' patterns from them to instruct the early diagnosis and intervention. Different interventions during the transports are to be examined carefully, and investigate to optimal approaches to cope with the health conditions of children patients under different scenarios and incidents. To deliver the application, edge computing will be utilised and allow us to seamlessly implement these AI models in the ambulance just-in-time and without relying on network connectivity[4]. An API and user friendly interface will be designed and integrated to the current software system.

Timeline:

We will build on the strengths of our team's previous work to develop, refine and test the clinical usability and acceptability of MediSense. Data from nearly 3000 patient transports (~400GB) has been collected since 2017. UCL data scientists have built the data pipeline and developed prototype AI models using patient monitoring data from ~2000 CATS transports to predict the occurrence of a critical incident and death in PICU.

Over the PhD study (36 months), we will achieve the project objectives in three overlapping work packages (WP):

- WP1 (months 1-3):

WP1a: Mobilise stakeholders and agree on governance.

WP1b: Focus on exploring correlations between 'physiotypes' and deterioration (e.g., acute hypoxia), interventions (e.g., fluid resuscitation) and outcomes (e.g., in-hospital mortality) of patients in demographic sub-groups.

-WP2 (months 4-15): A real-time risk management system with clinical explanation will be developed based on personal characteristics (age, gender, medical history, etc) and trends of vital signs.

-WP3 (month 15-24): Optimise the system in terms of computation to build a resource-efficient edge computing API with user-friendly UI

-WP4 (month 15-36): Build the prototype product, implement the system to the real-time use as a stand-by component, test it and received feedback for the next phase of wider applications.

-WP5 (months 30-40): Wrap up the project, write papers/reports and future funding applications.

-WP6 (months 2, 15, 24) Consult clinicians and organise PPI activities to engage with patients and the public.

Collaborations:

Prof Christina Pagel: Professor of Operational Research, Director of UCL Clinical Operational Research Unit. Christina has worked with GOSH ICU teams since 2006, including a 5-year researcher-in-residence with CATS. She is Co-CI at the CHiMERA (Collaborative Healthcare Innovation through Mathematics, Engineering and AI) hub at UCL

Prof Mark Peters: the UK's first Professor of Paediatric Intensive Care (UCL), and is also an Honorary Consultant Paediatric Intensivist (GOSH). He is Deputy Chair of the NIHR HTA General Funding committee, part of the CHiMERA hub and Chief Investigator on the Oxy-PICU, GASTRIC-PICU and PIVOTAL NIHR HTA multiple centre trials.

Kinseed: Industry partner (Kinseed) is currently working with 80% of the paediatric/neonatal (and more recently adult) 'mobile ICU' transport services across England and Northern Ireland, providing a clear route to market for MediSense

Flock.id: Flock.io will provide support in patient privacy preserved techniques to protect patient sensitive data in the process of transports and data analysis.

Plans for patient and public involvement and engagement for the project/student:

Mobile ICU settings are, by their nature, difficult areas to engage patients directly, especially in the paediatric retrieval environment where most children are too sick to remember any of the transport experience. Members of the research team however have extensive experience in successfully engaging parents/family members in previous NIHR-funded research in children's critical care. Christina's work in particular (developing a website for the public to understand heart surgery in children) has gained international recognition (<https://bit.ly/3aJMzjw>).

References:

1. PICANet Annual Report, 2020...
2. Ramnarayan P, et al. *Pediatr Crit Care Med*. 2018
3. Manning JC, et al. *Pediatr Crit Care Med*. 2018
4. Greco, L., et al, *Pattern Recognit Lett*, 2020