Jean Golding Institute – Seed corn funding
Application Form

Title (max 20 Words)
Elements of free text used in decision making: an exemplar from death reviews in prostate cancer and learning disabilities.

Start Date:	1st February 2020
Duration:	Feb-June 2020

Does your project involve human studies? *If so you will need to provide evidence of your plans to obtain the necessary ethical permissions for your studies.*

YES	*X*	*NO*	

- **LeDeR and CAP already have CAG section 251 and MREC approvals in place that cover this application. References and hard copies available on request.**

Lead applicant:	
Name	Dr. Avon Huxor
Post	Senior Research Associate
Department	School of Policy Studies
Institution	University of Bristol
Telephone	0117 33 10556
E-mail	phzaph@bristol.ac.uk
Co- applicant (1)	
Name	Dr. Emma [redacted]
Post	Research Fellow
Department	Bristol Medical School – Population Health Sciences
Institution	University of Bristol
Telephone	[redacted]
E-mail	
Co- applicant (2)	
Name	Ms. Eleanor [redacted]
Post	Senior Research Associate
Department	Bristol Medical School – Population Health Sciences
Institution	University of Bristol
Telephone	[redacted]
E-mail	
Co- applicant (3)	
Name	Dr Raul [redacted]
Post	Senior Lecturer in Data Science and Intelligent Systems
Department	School of Computer Science, Electrical and Electronic Engineering, and Engineering Maths
Institution	University of Bristol
Telephone	[redacted]
E-mail	

Please describe how the funding will be used and what specific goal/milestone your project will achieve.
- Describe key challenge.
- Milestones and/or Deliverables

(Up to 1 page)

Key Challenge:
Accurately assigning contributory factors and cause of death is vital to understanding health outcomes in the population and improving health care provision. However, these are not always recorded accurately on the death certificate (Turner et al, 2016; 10.1038/bjc.2016.162). The a) Learning Disability Mortality Review programme (LeDer - Huxor) and b) the Cancer Research UK Cluster randomised trial of PSA testing for prostate cancer (CAP - Turner/Walsh) are two separate projects with similar challenges in assigning cause of death accurately: (a) LeDeR is a NHS England initiative that collects cause of death information and summaries of the health care practice experienced by individuals with learning difficulties to directly impact on service improvement; (b) CAP is a long-running pragmatic randomised control trial that collects cause of death information to assign disease-specific mortality as part of trial outcomes to investigate the effectiveness of prostate cancer screening.
Both projects have generated over 4,000 death reviews (free text vignettes) which are used by volunteer experts to determine a) good or poor health care practice, and b) cause of death of individuals in their respective cohorts in England. The analyses from these records have already had impact on the health care system and the quality of patient outcomes (LeDeR, 2019; Martin et al, 2018; 10.1001/jama.2018.0154).
However, the data collection and decision process are expensive and time consuming, collectively involving over 200 trained reviewers, taking an average of a week for each review visiting health care providers, obtaining health records and summarising patients' medical and social history for volunteer clinical experts to assign a) good or poor health care practice or b) cause of death.
We will use machine learning techniques to identify the key elements (words and phrases) that are good predictors of a) good/poor health care practice or b) cause of death in the **thousands** of death reviews already collected from both LeDeR and CAP. Identifying these key elements will be used to guide the reviewers to target information in health records and extract the most important themes, making better use of this expensive resource. We envisage that the machine learning technique can also be used to semi-automate the cause of death decision process to assign cause of death directly for CAP, reducing the input from volunteer clinical experts.

We have already demonstrated that standard machine learning techniques can provide good (>90%) accuracy in predicting cause of death from the CAP reviews. The challenge now is to refine these techniques to improve accuracy, identify the key elements of these predictions and apply this to the LeDeR reviews. Often machine learning techniques focus on the value of the predictions made, rather than the elements that have generated the prediction itself. By identifying the key elements that account for the machine learning predictions we will also provide confidence in the techniques, making it transparent and explainable. Ultimately, this will better indicate which patients have a) experienced good/poor health care practice (LeDeR) or have b) died with/from prostate cancer (CAP). This will also be built into a generic, open-access tool that can explain the role of the key elements in the machine learning techniques that will be made available to the wider research community.

We envisage applying for future funding to utilise this machine learning technique to identify key elements in electronic health records, such as those collected by Public Health England and NHS Digital. This would be to develop a method to automatically generate reviews from the electronic health records using machine learning, further reducing the need for direct data collection from the hospital records. This would be directly exploited by a) LeDeR and b) CAP, but also applied to the collection of health outcomes in future randomised trials and large cohort studies.

Milestones/Deliverables:
1. Feb 2020: Recruit researcher/programmer
2. April-May 2020: Six-week period of intense research and programming work by researcher. Develop targeted review method and generic, open-access machine learning tool.
3. June 2020: Write report for JGI, publication of open-access machine learning tool and applications for external funding

We are also happy to present our techniques and results as part of the JGI Data Week 2020, for a session on machine learning in text analysis.

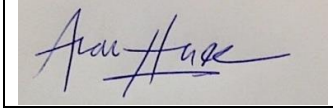
Please describe how this project will enable you to develop your research activity and what your plans are to seek follow on funding. (Up to ½ page)
The project will provide resources for a pilot study into 'explainable machine learning' as applied to free text data. It can be immediately exploited by our reviewers in the field, improving the speed at which important records are sought, and hence the quality of the ongoing work packages being undertaken by a) LeDeR and b) CAP. This project will provide a complete proof-of-concept to enhance future funding applications, as previous experience has suggested that the funders are often averse to funding exploratory work. Future funding would be sought to use the machine learning technique to identify key elements in electronic health data, collected by Public Health England and NHS Digital, for example, that can be used to generate reviews of health care practice and death outcomes, further reducing the need for direct data collection from medical record. We will seek future funding from (at least): 1. The National Institute of Health Research (https://www.nihr.ac.uk/) Cross-programme grant, Nov. 2020 2. The Health Foundation (https://www.health.org.uk/) Programme grant, Dec. 2020 It will also build cross-faculty relationships between: Social Sciences and Law; Health Sciences; and Engineering and will facilitate future collaborations involving external funding.

Please describe how the funds will be used. Please list any staff supported under this proposal and justify any consumables associated with this proposal. (Up to ½ page)
The current funding for both LeDeR and CAP cover work to deliver the main goals of these projects. Therefore, there is no scope to develop exploratory work. The funds sought from JGI will be used to employ a research associate (RA) to undertake the research and programming into explainable machine learning techniques using our free-text death review data and develop an open-access tool that can be used to apply this technique to other health research areas. The oversight of the technical work undertaken by the RA will be managed by Dr Raul Santos-Rodriguez, from Engineering Mathematics, with support from the rest of the project team.

Please provide a breakdown of the costs associated with this project. ***This funding cannot pay indirects and overheads.*** The funding will cover 100% direct costs.

FEC reference	282672: Elements of free text used in decision making: an exemplar from death reviews in prostate cancer and learning disabilities.
Costs breakdown	
Staff Costs: £	3,504
Consumables: £	-
Travel: £	-
Other: £	-
Total requested:	3,504

I confirm that I wish to apply for the Jean Golding Institute seed corn funding. The information given in this application and any accompanying material is accurate to the best of my knowledge. All institutions that are part of this proposal have approved this submission. I confirm that by submitting the application, that all costs have been checked by the Finance team in my school/department.

Signature of Lead Applicant:	Date:
	25th November 2019

Email jgi-admin@bristol.ac.uk if you have any queries and to submit your application. Deadline 17:00, 25 November 2019.