Institution: University of Leicester		
Unit of Assessment: 10		
Title of case study: Ensuring Robust Decision Making in Medical Applications		
Period when the underpinning research was undertaken: 2014–2-21		
Details of staff conducting the underpinning research from the submitting unit:		
Name(s):	**Role(s) (e.g. job title):**	**Period(s) employed by submitting HEI:**
Evgeny Mirkes	Research Associate in Applied Mathematics	2012–Present
Alexander Gorban	Professor of Applied Mathematics	2012–Present
Period when the claimed impact occurred: 2015–2020		
Is this case study continued from a case study submitted in 2014? N		

1. Summary of the impact

NHS hospitals and services in the UK are assessed, in part, by performance-related measurements (statistics) which ultimately affect their level of resource (funding). Errors in these measurements affect their reliability and, consequently, the assessments. Research into the identification, understanding and abrogating the effects of data on predictive models at University of Leicester (UoL) underpins the impact of this case study. The impact is in an **increased awareness and understanding** of inaccuracies within data including missing data and data from disparate sources and their effects on decision making within medical and operational applications within the NHS. As a direct consequence, NHS hospitals are now being assessed based on more accurate information. The beneficiaries are NHS hospitals nation-wide, particularly via more accurately allocated Best Practice Tariff payments, and the patient population as a whole through better managed NHS resources.

2. Underpinning research

Prof Gorban and Dr Mirkes have been working in the area of data analysis and data mining for over 30 years, specifically working on the uncertainties which are inherent to any empirical data and critical assessment of standard statistical methods **[R1–R6]**. Uncertainties can arise from a variety of sources including missing values and as a result of combining data sets originating from different sources. Another challenge is the validity of standard statistical assumptions, e.g., the independence of variables, and quantifiable objectivity of data collection and recordings, such as human bias.

Missing records and values in datasets are widespread in nearly all real-life data collections. It is typical and unavoidable in NHS and medical data. Researchers at UoL have been looking at this issue since 2013 when they were presented with the challenge of analysing the Trauma Audit Research Network (TARN) database.

TARN's database is the richest injury database in Europe and holds records of the overall state of patients on their discharge date from hospital. TARN receives data on trauma events from more than 200 hospitals, including over 93% of the hospitals in England and Wales, as well as hospitals in Ireland, the Netherlands, and Switzerland. Current standard practice within the NHS assesses the quality of the patient's state after major trauma on the basis of their survival 30 days after the initial injury. However, within this period, some patients are discharged or transferred to other hospitals resulting in incomplete information. Prior to **[R5]**, records with missing survival outcomes have either been ignored or taken as survived. Both of these approaches overlook the fact that missing values may or may not be independent, potentially resulting in misleading predictions **[E1, p.6-9]**.

Due to these early transfers and discharges, outcomes for these patients for the entire duration of the 30-day window are not registered. Another source of missing data is due to the initial state of the patient. For example, patients with severe trauma may not have all the observations taken at the point of entry due to the urgency of medical intervention. Similarly, patients with moderate injuries may not require the full spectrum of observations. An example of data structure in TARN is shown in Fig 1. Cases with identified unique or most critical injuries (the top node of the tree) are split into two groups: the group of patients who were admitted to a TARN hospital within 24 hours after the injury (Main Group); and the group of patients transferred to a TARN hospital more than 24 hours after injury (Transferred). The Main Group is then split into subgroups: the group of patients for whom an outcome was known within 30 days of the injury (Available W30D), and the group of patients transferred from a TARN hospital to a non-TARN care provider during 30 days after the injury (OUT30). The Main Group is used to develop predictive modelling, and the Transferred group is used to verify modelling outcomes **[R5]**.

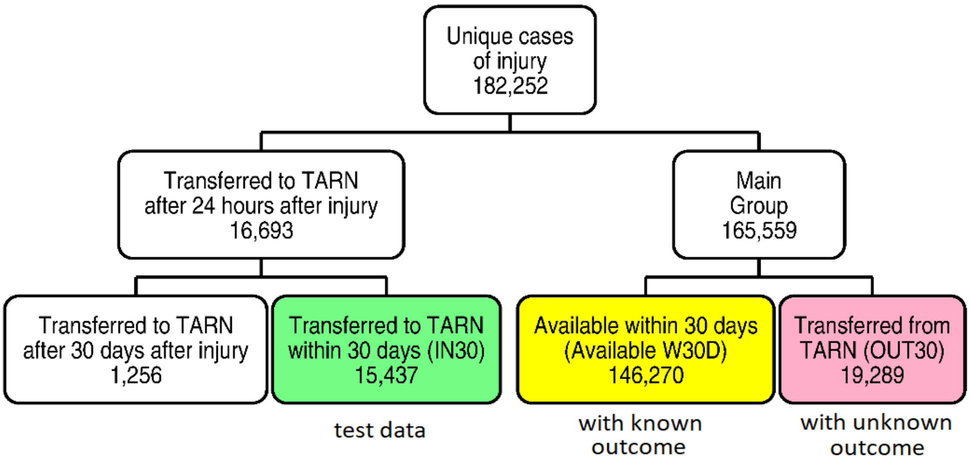

Fig. 1. Structure of TARN database (2008-2014).

Human bias and subjectivity often occurs at the patient's point of entrance into the NHS system. The reliability of common decision-making practices employed by NHS hospitals need to be scientifically justified. Nurses' initial patient assessments can be subject to bias, especially in A&E departments. This bias can affect prioritisation of a patient and survival outcomes. For example, in 12 hospitals and 4 ambulance services in the UK, clinical decisions to refer a child to primary care or self-care are based on the Paediatric Observation Priority Score (POPS). POPS supports clinical decisions, especially for staff not familiar with dealing with a child. POPS uses a combination of 8 physiological, behavioural, and known-risk parameters: oxygen saturations, level of alertness, extent of breathing difficulty, background history, nurse gut feeling, heart rate, respiratory rate and temperature. Some of these variables (e.g., nurse gut feeling) are clearly subjective, have human bias, and have never been assessed in a robust and rigorous way. The UoL study evidenced that despite the nature of the variables, POPS score is a sufficient indicator of the patient's length of stay, thus demonstrating the reliability of the POPS score as a decision-making tool **[R6]**.

Gathering profiles of physical activity is key for monitoring health and well-being of the general public, and patients at risk of cardiometabolic disease, obesity, and musculoskeletal health. Accelerometer data is increasingly used to collect this information. There are three widely used types of research-grade accelerometers: GENEActiv, Axivity, and ActiGraph; with varying, possibly non-fixed sampling frequencies. To use data from all devices, to increase the capability of data collection, and increase the power of health care research through increased sample sizes of studies, harmonisation is required. The research conducted by Dr Mirkes enables harmonisation of accelerometer data across different platforms, studies, and pooling of data **[R3, R4].**

3. References to the research

[R1] Gorban, A.N., Golubkov, A., Grechuk, B., **Mirkes, E.M.** and Tyukin, I.Y., 2018. “Correction of AI systems by linear discriminants: Probabilistic foundations”. *Information Sciences,* 466, pp.303-322

[R2] Manso A.S., Chai M.H., Atack J.M., Furi L., De Ste Croix M., Haigh R., Trappetti C., Ogunniyi AD., Shewell L.K., Boitano M., Clark T.A., Korlach J., Blades M., **Mirkes E., Gorban A.N**., Paton J.C., Jennings M.P., Oggioni M.R..2014. “A random six-phase switch regulates pneumococcal virulence via global epigenetic changes”. *Nature Communications*, 5, p.5055.

[R3] Rowlands, A.V., **Mirkes, E.M.**, Yates, T., Clemes, S., Davies, M., Khunti, K. and Edwardson, C.L., 2018. “Accelerometer-assessed physical activity in epidemiology: are monitors equivalent?” *Medicine and Science in Sports and Exercise*, 50(2), pp.257-265.

[R4] Rowlands, A., Edwardson, C., Plekhanova, T., **Mirkes, E.,** Davies, M., Khunti, K. and Yates, T., 2018, October. “Innovations in the use of raw accelerometry in epidemiology: A basis for harmonisation of physical activity outcomes across international datasets”. *Journal of Physical Activity and Health*, 15 (10), pp.S18-18.

[R5] Mirkes, E.M., Coats, T.J., Levesley, J. and **Gorban, A.N.,** 2016. “Handling missing data in large healthcare dataset: A case study of unknown trauma outcomes”. *Computers in Biology and Medicine*, 75, pp.203-216. DOI: 10.1016/j.compbiomed.2016.06.004

[R6] Roland, D., Suzen, N., Coats, T., Levesley, J., **Gorban, A.N**., & **Mirkes, E**., 2020. “What can the randomness of missing values tell you about clinical practice in large data sets of children’s vital signs?” *Pediatric Research*, pp.1-6.

4. Details of the impact

The impact from the research is in terms of **increased awareness and understanding**.

Leicester Diabetes Centre
The Leicester Diabetes Centre is a world-renowned, applied health research unit committed to improving the lives and care of people with diabetes and other long-term conditions. Research on combining data from different disparate sources, subsequently reported in **[R3, R4]**, enabled clinicians working at the Diabetes centre *“to process accelerometer data from different devices using the same software [which] crucially reduced uncertainty in comparison of results found in different studies”* **[E5]**. The research *“shows which parts of the data collected by different type of accelerometers*

can be considered equivalent and thus considered inside of one study" **[E5]**.

Trauma Audit Research Network (TARN)

Researchers at UoL have been working with TARN to improve the reliability of the modelling outcomes using their database. Prior to the underpinning research, records with missing survival outcomes have either been ignored or taken as survived. The research has revealed that ignoring missing data, or assuming that the data in the TARN dataset are missed at random, adversely affects predictive modelling of mortality rates. The UoL team developed a family of new models where mortality is based on non-stationary Markov chains. These models take explicit account for missing data in clinical records. A distinctive feature is the presence of two 'lotteries': the *"lottery of death/discharge"* and the *"lottery of transfer"* **[R5]**. The implemented models allow explicit simulation of recovery, death, and transfer between TARN and out-of-TARN hospitals.

The team's research on the TARN database revealed that the data on more than 11% (around 19,289 patients) of the patients are not accounted for in mortality estimates, and hence lead to a bias in NHS resource management. The fundamental research carried out by the team at UoL, through **increased awareness and understanding** of the importance of the effects of missing data, has enabled TARN and NHS resource managers to better manage and predict resource requirements. This is evidenced by practitioners within TARN who have recognised the benefits of the improved predictions **[E2]** and more general NHS practitioners.

The team's research on the missing data and more accurate modelling, underpinned by the research subsequently reported in **[R5, R6]**, has directly improved the accuracy of hospital's performance assessment nation-wide. In particular, the UoL team showed, by using more accurate modelling of mortality rates, that ignoring missing data overestimates the rate of mortality by 12% (200 deaths per year). This research has *"increased [TARN] awareness and understanding of the fact that a missed attribute within the data may not be missed 'at random' or not independent on other attributes. . . . With this new level of understanding … TARN gained access to true mortality … TARN data is embedded in a novel system of quality-based commissioning, being used to determine 'Best Practice Tariff' payments to hospitals nationally"* **[E2]**. As a direct consequence, NHS hospitals across the UK are now being assessed based on more accurate information and 'Best Practice Tariff' payments are more accurately allocated across the NHS.

Paediatric Observation Priority Score (POPS)

The research on POPS confirmed, for the first time, that despite the variables and measurements contributing to the POPS score being highly uncertain and having a fundamental variability in that they depend on the individual's feeling about a patient's state, the final POPS score is a reliable indicator of the patient's length of stay and subsequent resources requirements. *"POPS is a unifying clinical decision tool, especially for staff not familiar with dealing with children"* **[E3]**. The impact of research is nation-wide affecting 12 NHS hospitals and 4 ambulance services across the UK that use POPS, who collectively see around 300,000 children each year **[E3]**. *"Scoring systems in Emergency Departments (EDs) in the UK are rarely validated. POPS is a method of quantifying patient acuity, which is being used by 12 NHS Hospitals and 4 Ambulance Services around the UK. POPS identifies a range of severity of childhood illness, to support staff in taking the decision to redirect the child to primary care or discharge to self-care and to help them in expediting senior/specialist assistance for deteriorating children. Importantly, POPS is not intended to determine a course of action but to support clinical decisions, especially for* *staff not familiar with dealing with children. . . . As a consequence of this increased awareness POPS is now more robust to the effects of missing data and uncertainty."* **[E3]**.

Reliable data is always important in healthcare, but the confidence in POPS as an assessment tool, provided by research such as **[R6]** is increasingly important in the Covid-19 pandemic. A recent study has reported that the use of POPS has led to an improved detection rate of Covid-19 virus in children. *"Existing triage tools have good but not excellent prediction for adverse outcome in children with suspected COVID-19. POPS … could achieve an appropriate balance of sensitivity and specificity for supporting decisions to discharge home by considering any score above zero to be positive."* **[E4]**.

5. Sources to corroborate the impact

[E1] Hand, DJ. *Dark Data: Why What You Don't Know Matters*. 2020, Princeton University Press. Princeton and Oxford. pp6-9.

[E2] Letter of Acknowledgement from TARN.

[E3] Letter of Acknowledgement from POPS.

[E4] Biggs, K., Thomas, B., Goodacre, S., et. al., 2020. *Prognostic accuracy of emergency department triage tools for children with suspected COVID-19: The PRIEST observational cohort study*. medRxiv.

[E5] Letter of Acknowledgement from Melanie J Davies, Professor of Diabetes Medicine, NIHR and Senior Investigator Emeritus Director of NIHR, Leicester BRC.