Manchester Summer School in Digital Epidemiology

Hosted by the Arthritis Research UK Centre for Epidemiology
The Studio, Lever Street, Manchester, 18-20th July 2018

Course goal: To explore and understand the opportunities, challenges and methods for capturing and using digital data to support high-quality epidemiological research

Day 1: Electronic health records and data linkage

0900-0930 **Welcome, introductions and course overview** Will Dixon, UoM

0930-1030 **Epidemiology using EHR and linked data**

- Introduction — Will Dixon, UoM
- Show and tell: delegates' experiences of EHR and data linkage projects

Coffee

1100-1300 **What's in (and what's not in) an EHR, and how to prepare data for analysis**

- How do we get the best out of EHRs? — Liam Smeeth, LSHTM
- Drug prep algorithm (P) — Mark Lunt & Rebecca Joseph, UoM
- Linking EHR data to other data sources — David Ford, Swansea University

Lunch

1400-1520 **Analysing EHR data: traditional and novel approaches part 1**

- Handling confounding: an interactive session (P) — Mark Lunt, UoM

Coffee

1540-1700 **Analysing EHR data: traditional and novel approaches part 2**

- Machine learning and EHRs — Cian Hughes, Google Deepmind

Learning objectives Day 1:

1. Understand key issues in ensuring we get better research out of EHRs, not just more research
2. Discover some of the challenges of preparing EHR data ready for analysis
3. Understand potential valuable insights, plus the legal and cultural issues of linking health to non-health data
4. Explore practical methods for handling measured and unmeasured confounding
5. Learn about novel analysis methods including machine learning

(P) = practical session

Day 2: Patient-generated data: smartphones

0830-0930 **Epidemiology using digital patient-generated data**

- Introduction — John McBeth, UoM
- Show and tell: delegates' experience of patient-generated data projects

0930-1115 **Public involvement and designing systems for collecting patient-generated data**

- Co-design, patient and public involvement — Will Dixon & Lamiece Hassan, UoM
- Q&A with patient partners — Carolyn Gamble, Karen Staniland, Simon Stones (patient partners)

Coffee break

1145-1300 **Preparation of temporally-rich patient-generated data**

- Descriptive statistics and code for temporally rich data in populations (P)
  - Engagement states & hidden Markov models — David Selby, Uni Warwick
  - Transition matrices — Thomas House, UoM
  - Handling missing GPS data — Belay Birlie, UoM

Lunch

1400-1530 **Analytic methods – making use of longitudinal patient-generated data**

- Case only designs — Malcolm Maclure, University of British Columbia
- Time series methods for aggregated & individual data — Antonio Gasparrini, LSHTM

Coffee break

1545-1700 **Keynote lecture**

- The Population Health Monitoring Trinity: Epidemiological Methods, Informatics, and Big Digital Data — David Buckeridge, McGill University

Learning objectives Day 2:

1. Understand the importance and value of co-design in collecting patient-generated data, and why it is relevant to epidemiology
2. Recognise and learn a range of methods to describe and prepare longitudinal patient-generated data
3. Be able to use n-of-1 case-only designs to study your own health events, and understand key differences with standard study designs
4. Understand how big data and health informatics can support high-quality epidemiological research for public health benefit

Day 3: Patient-generated data: sensors and social media

0830-0930 **Epidemiology using sensor and IoT data**

- Introduction — Sabine van der Veer, UoM
- Show and tell: delegates' projects with sensor and IoT data

0930-1015 **How to collect data passively**

- What's available in wearable devices, and what might you get from them — Justin Phillips, Google

Coffee

1040-1200 **Preparing and processing sensor data**

- How to go from raw to processed sensor data (P) — Max Little, Aston University

1200-1300 **Using sensor data in epidemiological research**

- Evaluating physical activity in epidemiology studies — Soren Brage, University of Cambridge

Lunch

1400-1445 **Designing research projects that incorporate sensor and IoT data**

- Practical advice for designing research projects incorporating sensors and IoT data, including governance — John Ainsworth, UoM

Coffee

1515-1700 **Social media data from patients**

- Pharmacovigilance using social media text mining (P) — Goran Nenadic and team, UoM

Learning objectives Day 3:

1. Appreciate novel types of passively collected data including wearable sensors, possible medical insights, and the governance and data management around their appropriate use
2. Understanding of what is involved in developing and validating algorithms for data processing
3. Understand confounding in sensor data, why it matters, and how to mitigate it
4. Gain an understanding of the use of wearable sensors in physical activity epidemiology
5. Appreciate the opportunities and methods for extracting value from free text