

Development of, and early findings from, the Wales Electronic Cohort for Children (WECC)

Professor Ronan Lyons

Centre for Health Information, Research and Evaluation (CHIRAL)

Institute of Life Science

School of Medicine

Swansea University

Health Information Research Unit (HIRU)

- Funded by NISCHR
- Part of the strategic research infrastructure in Wales and UK
- Institute of Life Sciences: £50M development; IBM HPC
- Created the Secure Anonymised Information Linkage (SAIL) system

The essence of HIRU's work . .

Linking anonymised data, at individual and ecological levels, across multiple datasets, drawn from operational systems in health services, national databases, clinical datasets and databases in social care, education, housing, etc.

Anonymised Linkage Fields (ALFs)

- Split files, trusted third parties, and double encryption to protect confidentiality
- ALF – encrypted NHS number
- RALF – encrypted address

Wales Electronic Cohort for Children (WECC)

An all-Wales public health and paediatric collaboration

Ronan Lyons, David Fone, John Gallacher, David Ford, Caroline Brooks, Sinead Brophy, Frank Dunstan, Mike Gravenor, Kerina Jones, Sailesh Kotecha, Gareth Morgan, Shantini Paranjothy, Sarah Rodgers, Rhys Williams, Gareth John.

Wales Electronic Cohort for Children (WECC)

Anonymised data from 730,000 children

- Platform for translating information into child health population policy:
- What factors determine the future health service need for individuals that are vulnerable at birth, and inform the development of interventions to reduce health inequalities in these groups?
- What is the influence of the social and physical environment on childhood obesity?

Anonymised datasets to be included in WECC

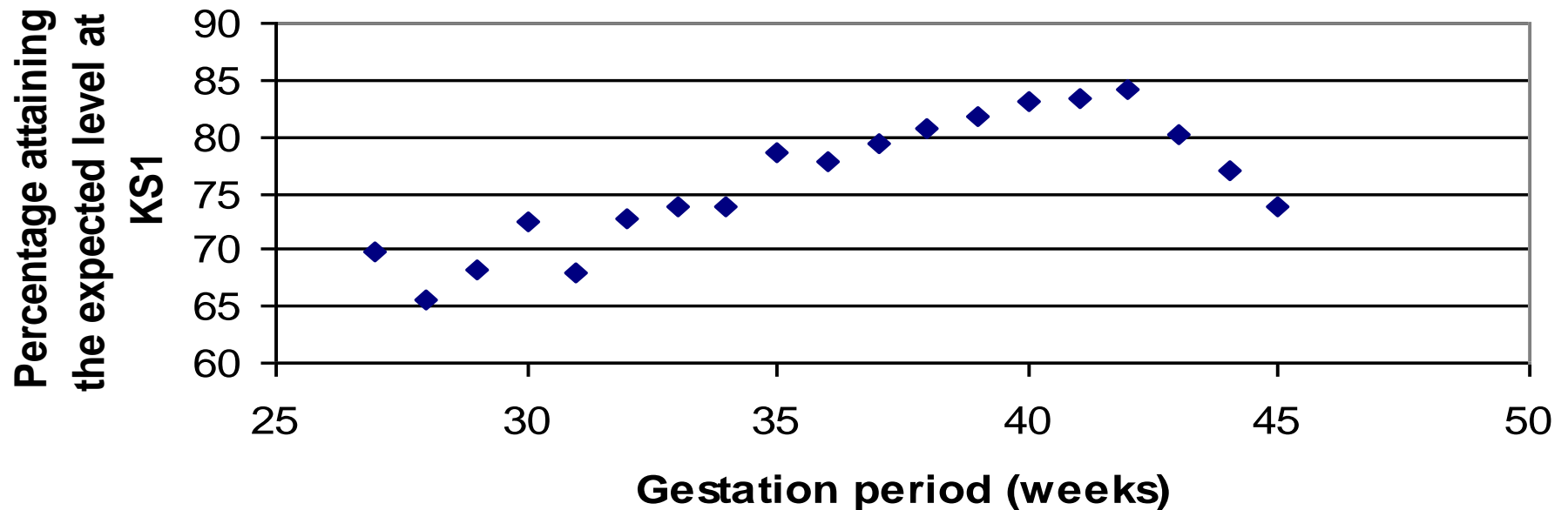
- NHSAR – population register
- NCCHD - Community Child Health
- ONS - births and deaths
- AWPS – Perinatal Survey
- PEDW – inpatients
- OPMD – outpatients
- CARIS – congenital anomalies
- GP data – depending on availability
- DCELLS – educational attainment
- Environmental data (RALF and small area derived)

Data cleaning

- Some data are incomplete
- Coding variability: Asthma 16 and 599 READ codes
- Multiple datasets, even with v high accuracy, results in data conflicts
- Data are refreshed - don't make permanent changes
- Huge task!

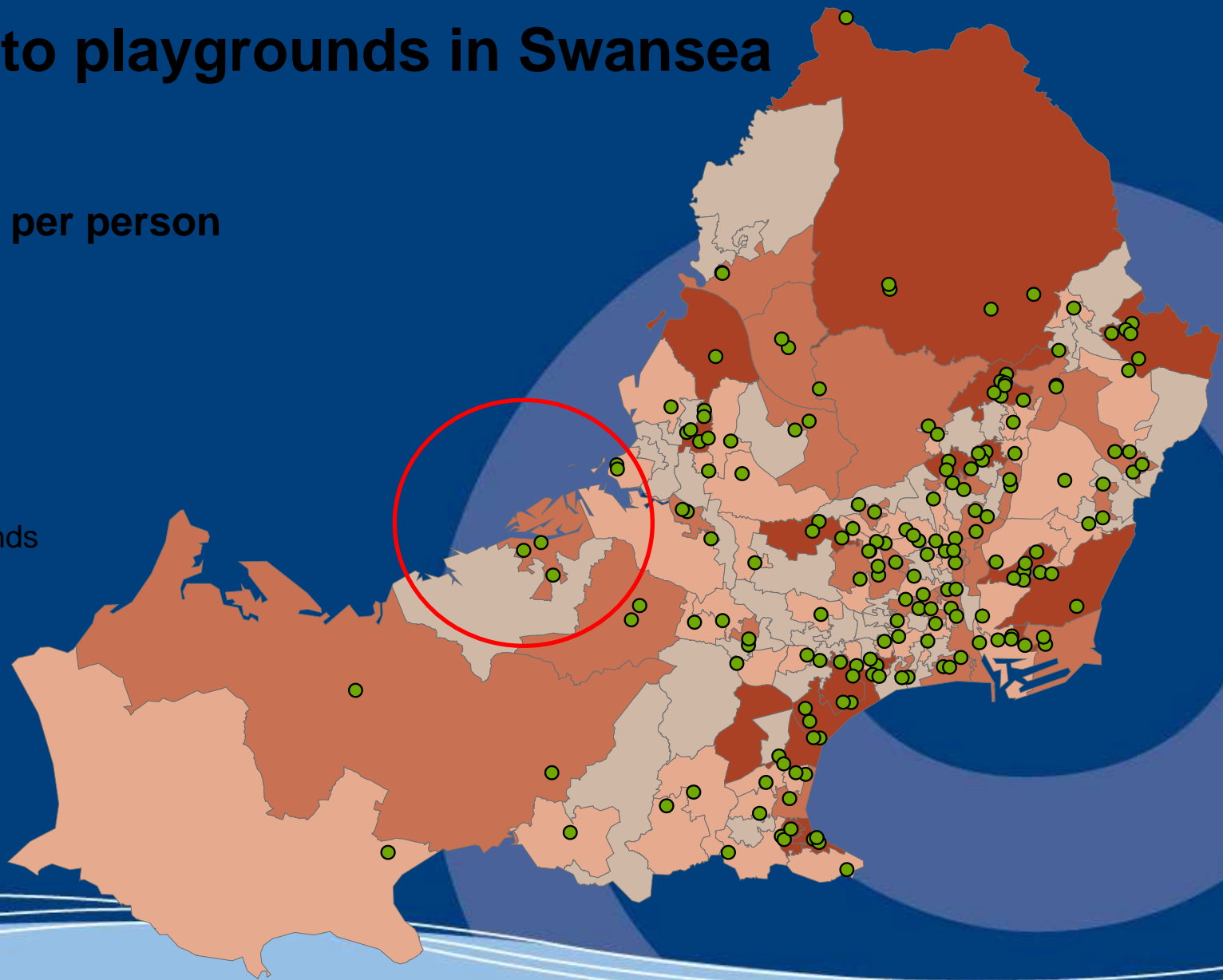
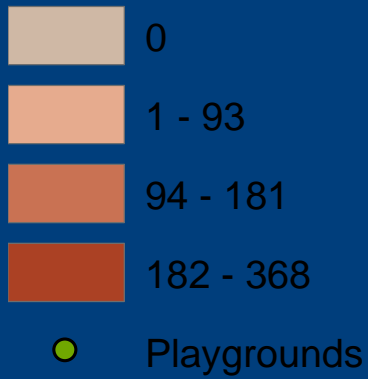
Replication of known associations: WECC, n=123,633

Relationship between Gestation period classes (27 to 45 weeks) and KS1 performance



Access to playgrounds in Swansea

playgrounds per person



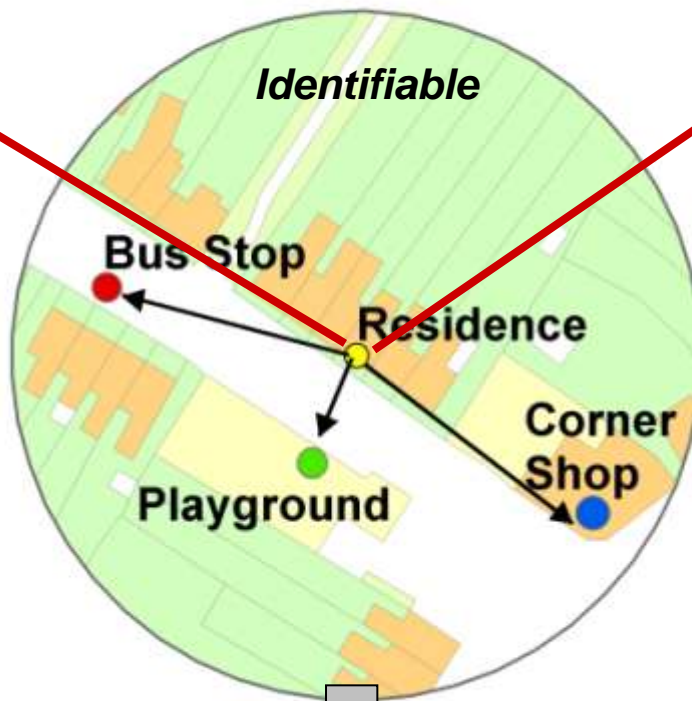
Network analysis: 264m vs 45m



Environmental metrics and links to health and social data – fictional example

David Williams
8 Main St, Swansea

Male
age 8
Asthma
--
Free School Meals



Margaret Williams
8 Main St, Swansea

Female
age 35
Diabetes
Smoker

Anonymous

ALF	RALF	Medical 1	Medical 2	Social 1	Environment1
11223387	5448893	Asthma	--	FSM	5.85
11238889	5448893	Diabetes	Smoker	--	5.85

Same RALF: cohabiting

GP Database

Same environmental metric
e.g. dwelling density,
land use mix,
metres to bus stop or GP surgery

Social Database

Potential other research questions - WECC

- Wide variety possible, e.g. impact of :
 - health conditions on educational outcomes
 - maternal depression and treatment on injury risk
 - treating ADHD on educational achievement
 - housing quality and health
 - household infections in early pregnancy on birth anomalies (antibiotic use as proxy within RALFs)

Relevant methodology papers

- Ford DV, Jones KH, Verplancke J-P, Lyons RA, John G, Brown G, Brooks C, Bodger O, Couch T and Leake K. **The SAIL Databank: building a national architecture for e-health research and evaluation.** *BMC Health Services Research* 2009, 9:157 (4 September 2009)
- Lyons RA, Jones KH, John G, Brooks CJ, Verplancke J-P, Ford DV, Brown G and Leake K. **The SAIL databank: linking multiple health and social care datasets.** *BMC Medical Informatics and Decision Making* 2009, 9:3 (16 January 2009)
- Rodgers SE, Lyons RA, Dsilva R, Jones KH, Brooks CJ, Ford DV, John G and Verplancke J-P. **Residential Anonymous Linking Fields (RALFs): A Novel Information Infrastructure to Study the Interaction between the Environment and Individuals' Health.** *Journal of Public Health*, 2009, pp. 1-7.