eurolinkcat
establishing a linked European cohort of children with congenital anomalies

# Linking to health care databases across Europe and the UK: results and recommendations from the EUROlinkCAT project

Maria Loane

MatCHNet seminar 29-02-2024

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 733001. Jan 2017 – Dec 2021

eurolinkcat
establishing a linked European cohort of children with congenital anomalies

# Background

- Congenital anomalies (CAs) are a leading cause of mortality and morbidity in children
- European CA registries (EUROCAT) conduct epidemiological surveillance on children with anomalies, mostly up to 1 year of age
    - Includes England, Scotland and Wales
- Little information on long-term survival and morbidity outcomes
- Availability of routinely collected administrative data may bridge this gap
    - Accuracy of hospital discharge data for surveillance

# Aim of EUROlinkCAT project

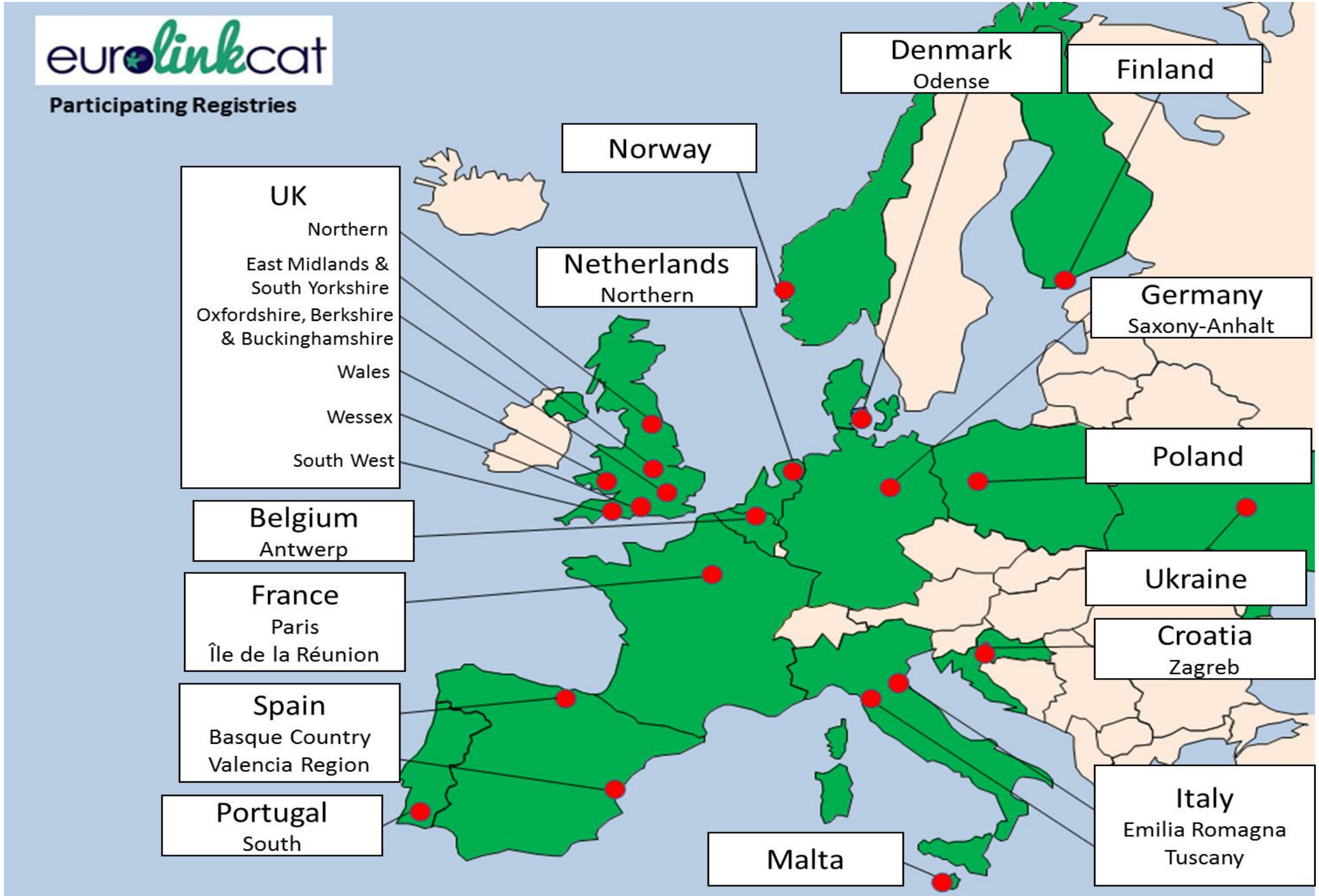
To establish a linked European cohort of children with congenital anomalies (CAs) to evaluate mortality and morbidity outcomes of these children up to the age of 10 years

# Methods

- 22 EUROCAT registries in 14 European countries planned to participate in the study
  - 3 did not get ethical approval in time so they were unable to participate
- All **liveborn** children with CAs born 1995-2014
- Children without CAs (reference children) were selected from the same population covered by the EUROCAT registry, and from the same time period
- Children were followed up to their 10$^{th}$ birthday, or to 31st Dec 2015, to have at least 1-year follow-up information

eurolinkcat

**Participating Registries**



# Steps to create European cohort

- Registries applied for ethical approval to link to administrative data
- Create a common data model (CDM) to **standardise** administrative data
  - Mortality data
  - Hospital discharge data
  - Prescription data
- Create linked standardised dataset on a cohort of children with CAs and reference children

# STEPS INVOLVED

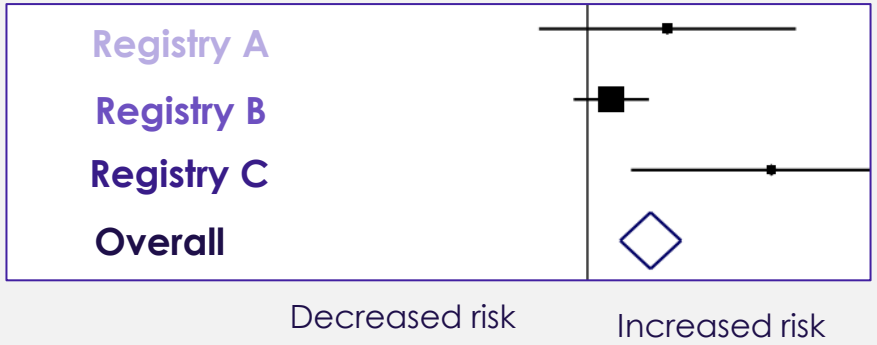
Linked standard dataset

Syntax scripts

## Individual Registry Analysis

| Registry A | Registry B | Registry C |
|---|---|---|
| Aggregate data and results | Aggregate data and results | Aggregate data and results |

## Meta-analysis




eurolinkcat

# Linkage methods

- Deterministic linkage – unique personal identification (ID) number
- Probabilistic linkage – combination of variables e.g. DOB, gender, residence code, mother's DOB
- Combination of deterministic and probabilistic methods
- Manual linkage

# Linkage methods (survival study)

- Deterministic linkage – unique personal identification (ID) number (n=7 registries)
- Probabilistic linkage – combination of variables e.g. DOB, gender, residence code, mother's DOB (n=2 registries)
- Combination of deterministic and probabilistic methods (n=6 registries)
- Manual linkage (n=3 registries)

eurolinkcat
establishing a linked European cohort of children with congenital anomalies

# Survival

# Survival – (CAs only)

- Linked CA data to national vital statistics (n=11 registries)
  - Successfully linked 96.5% of children with CAs
- Linked CA data to death certificates only (n=7 registries)
  - Not possible to assess linkage success
- Of the 18 registries:
  - One was excluded as <85% of children had identifiers to enable linkage
  - One was unable to complete the study as only COVID-19 research was allowed

# Survival – Description of cohort (CAs only)

| | Linkage to Mortality Records Only | Linkage to Vital statistics |
|---|---|---|
| **Registries (n)** | 5 | 11 |
| **Registry name** | Belgium (Antwerp), Germany (Saxony Anhalt), Malta, Spain (Valencian Region), Ukraine | Denmark (Funen), Finland, France (Paris), Italy (Emilia Romagna, Tuscany), Netherlands, Norway, UK (Wales, Thames Valley, East Midlands, Wessex) |
| **Live births (gestational age ≥ 24 weeks)** | 26,670 | 149,000 |
| **Completeness of data** | Estimated 91% | 97% |
| **Deaths** | 1,225 deaths | 7,767 deaths |

eurolinkcat
establishing a linked European cohort of children with congenital anomalies

# Morbidity

# Morbidity

- Hospital admissions due to obstetric stay were excluded i.e. births
- Reference children available
  - 5 registries: whole population
  - 2 registries:  10%/20% random sample of population (Tuscany, Netherlands)
  - 4 registries: no reference children (Croatia, 3 English registries)
- Linkage success = 96.8% for children with CAs
- Linkage success =95.2% for reference children

# Morbidity – Description of cohort

| | Linkage to Hospital Records (Children with CAs) | Linkage to Hospital Records (Reference children) |
|---|---|---|
| **Registries (n)** | 11 | 7 |
| **Registry name** | Denmark (Funen), Finland, Italy (Emilia Romagna, Tuscany), Netherlands, Spain (Valencian Region), UK (Wales, Thames Valley, East Midlands, Wessex) Croatia (Zagreb). | Denmark (Funen), Finland, Italy (Emilia Romagna, Tuscany), Netherlands, Spain (Valencian Region), UK (Wales) |
| **Live births (gestational age ≥ 23 weeks)** | 99,416 | 2,021,772 |
| **Completeness of data** | >95% | >97% |

# Comparison of linked versus unlinked children

- Children born preterm less likely to be linked
- Children with CAs born to teenage mothers less likely to be linked
- Reference children born to older mothers less likely to be linked

eurolinkcat
establishing a linked European cohort of children with congenital anomalies

# Prescription data

# Prescriptions – Methods

- Prescriptions for antibiotics and chronic diseases in children
  - Chronic diseases=asthma, diabetes, epilepsy, and cardiac conditions
- Children with CAs and reference children born **2000-2014**
  - Data collected up to 31st December 2015, so 1-year follow-up
- 7 registries included
- Linkage success = 94.9% for children with CAs (based on 6 registries)
- Linkage success =95.1% for reference children (based on 6 registries)
- Cohort - linked prescription data available for:
  - 60,662 children with CAs, 1,722,912 reference children

eurolinkcat
establishing a linked European cohort of children with congenital anomalies

# Accuracy of hospital data for ascertaining CAs

# Accuracy study

- 11 EUROCAT CA registries in 8 European countries
- All liveborn children (2010 -2014) recorded in the EUROCAT registries (gold standard) who were linked to hospital databases
- Children in the hospital databases with any ICD9-CM or ICD10 code for CA in the first year of life
- Compared codes in the hospital database to codes recorded in the CA registries data for 17 specific CAs
- Calculated sensitivity (isolated CA cases only) and positive predictive value (PPV) for each specific anomaly

# Main results

- European hospital databases accurately record  a limited number of anomalies in live born children e.g. cleft lip with or without cleft palate, gastroschisis and Down syndrome
- CAs that do not require hospitalization or surgery are underreported in hospital discharge databases
- Pregnancies resulting in termination following prenatal diagnosis of fetal anomaly are often missing in hospital databases - relevant for anomalies with a high termination rate

# Conclusion

- Hospital databases can be an additional ascertainment source for CA registries, but they cannot replace CA registries

eurolinkcat
establishing a linked European cohort of children with congenital anomalies

# Strengths & limitations

# Strengths of the cohort

- Population-based European cohort
  - Data available on >100 specific types of congenital anomaly
- The common data models for standardising the data are publicly available
  - Can be used in other populations as the variables in the health care databases are the same regardless of the population studied

# Limitations of the cohort

- Only aggregate tables/results are available
  - Linked individual case data remains at local registry level
- Some bias in the “not-linked” children
  - children born preterm and those born to mothers aged <20 and ≥35 years were less likely to be linked
- Only 3 registries had data from 1995
  - Unable to follow-up children to 10 years of age for the other registries

# Recommendations

- Assign a permanent unique identification (ID) number to each baby at birth to enable accurate linkage to administrative healthcare databases.
- Systems for classifying and reporting anomalies diagnosed in fetuses who undergo a termination.
- Implement algorithms to discriminate between major CAs and suspected or minor anomalies in healthcare databases.

# Conclusion

- Linking to administrative data allowed us to create a European cohort of children to estimate long-term survival and morbidity outcomes
- Survival cohort: ~175,000 children with CAs
- Morbidity cohort: ~100,000 children with CAs and ~2 million reference children
- Prescriptions cohort: ~61,000 children with CAs, 1.7 million reference children
- Hospital databases can be an additional ascertainment source for CA registries
- CA registries are still the most appropriate data source for surveillance of CAs.

eurolinkcat
establishing a linked European cohort of children with congenital anomalies

# Thank you!

**http://www.EUROlinkCAT.eu/publications**

- Morris JK, Garne E, Loane M, et al. EUROlinkCAT protocol for a European population-based data linkage study investigating the survival, morbidity and education of children with congenital anomalies. BMJ Open. 2021 Jun 28;11(6):e047859. https://doi.org/10.1136/bmjopen-2020-047859

- Loane M, Given JE, Tan J, et al. (2021) Linking a European cohort of children born with congenital anomalies to vital statistics and mortality records: A EUROlinkCAT study. PLoS ONE 16(8): e0256535. https://doi.org/10.1371/journal.pone.0256535

- Loane M, Given JE, Tan J, et al. (2023) Creating a population-based cohort of children born with and without congenital anomalies using birth data matche20d to hospital discharge databases in 11 European regions: Assessment of linkage success and data quality. PLoS ONE 18(8): e0290711. https://doi.org/10.1371/journal.pone.0290711

- Bakker MK, Loane M, Garne E et al. Accuracy of congenital anomaly coding in live birth children recorded in European health care databases, a EUROlinkCAT study. *Eur J Epidemiol* (2023). https://doi.org/10.1007/s10654-023-00971-z