

# NCRIS

National Research  
Infrastructure for Australia

An Australian Government Initiative

# SSB

Statistical Services Branch  
Queensland Health

## Statistical Services Branch (DLQ) – Data Linkage update

Trisha Johnston

Director, Statistical Analysis and Linkage Unit, Queensland Health



# SSB linkage team role

SSB linkage team (aka Data Linkage Queensland) is responsible for linkage of data to create an evidence base to inform health system management and service delivery

Key clients:

- Queensland Department of Health
- Hospital and Health Services
- Researchers
- government and private agencies

# What is data linkage

---

- Unique person identifiers are not often available within or across siloed data collections in Queensland and Australia
- Data linkage is used to identify persons across siloes
- Uses person-level identifying information (such as name, date of birth, address) to determine which records within a data source, or between multiple data sources, pertain to a particular individual.

# Probabilistic linkage

---

- Uses the concept of probability to produce a composite score based on weighted values for each element used for linkage
- Relies on completeness and quality of identifiers (names, date of birth, address)
- Not all identifiers have to be available to determine a match – probabilistic linkage is more resilient to data quality challenges than alternative methods

# Linkage in Queensland Health

---

- Funded by ongoing National Collaborative Research Infrastructure Strategy (NCRIS) funding, one-off Queensland Government Innovation (DEEDI) funding and Queensland Health funding since 2011
- Queensland Department of Health recurrently funds positions and various in-kind support in addition to NCRIS funds to provide an ongoing data linkage service

PHRN

DLQ  
(Qld)

CHeReL  
(NSW,  
ACT)

QGSO  
(Qld)

# NCRIS

National Research  
Infrastructure for Australia

An Australian Government Initiative

AIHW  
(National)

WADLS  
(WA)

SA NT  
Datalink  
(SA,  
NT)

CVDL  
(Vic)

TDLU  
(Tas)

ABS  
(National)

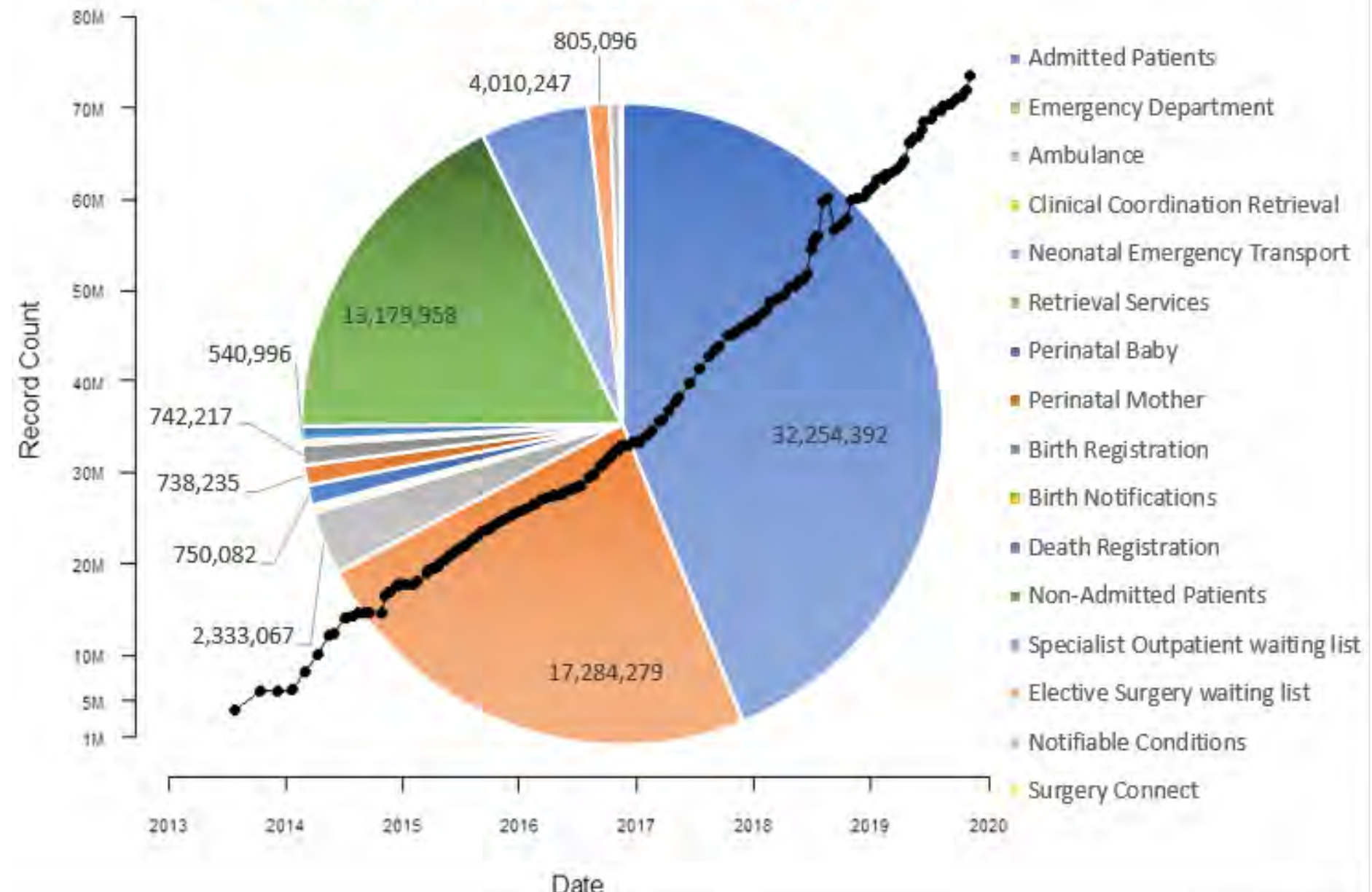
# Linkage in Queensland Health

---

- Production Linkage – creation of a Master Linkage File (MLF) with enduring linkages between core health data collections which is maintained in near-real time. The MLF currently includes approx. 73M records.
- Request linkage – provision of data from the MLF and custom linkage for ad hoc research and government requests

# Master Linkage File (MLF)

## 73+ Million Records





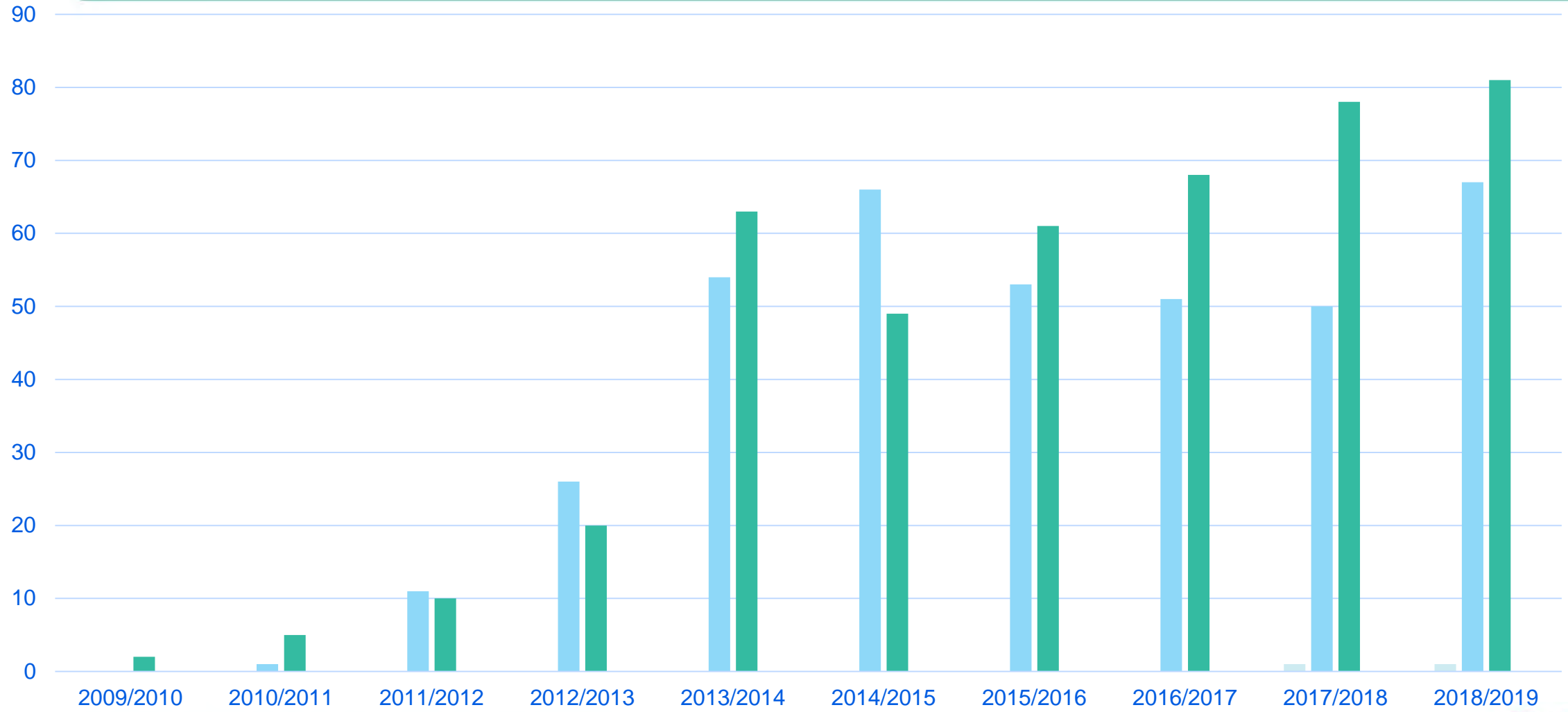
## DLQ MLF

- Routinely brings together information from multiple data collections and registries in near real time
- Groups together event-level records belonging to the same person (identified by an enduring linkage key)
- Client requirements have determined MLF inclusions
- Encrypted MLKs provided directly to approved QH users for evidence-based policy, service planning, management, monitoring and evaluation
- MLF is used as the basis for ~80% of linkage requests and enables much faster response times

# Linkage requests

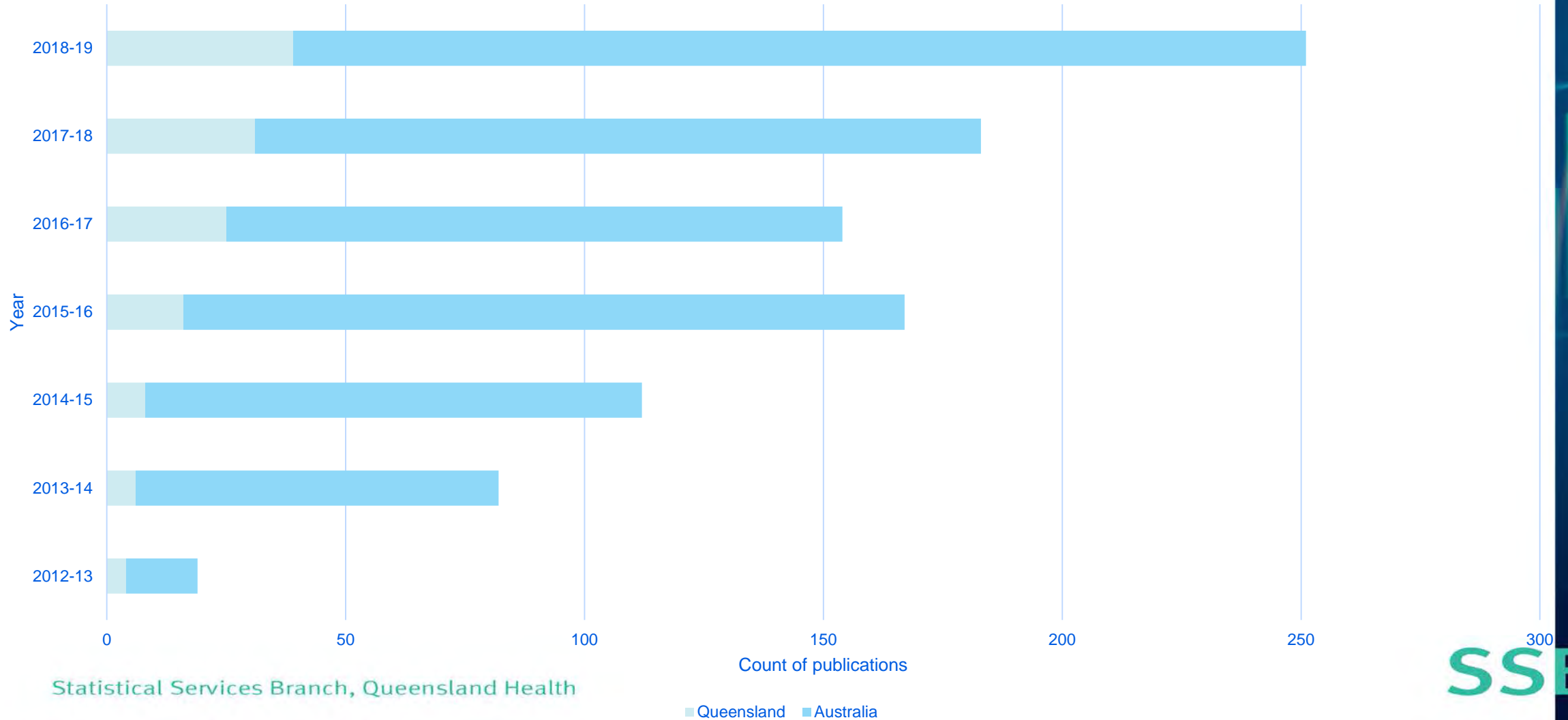


# Linkage requests by year and type

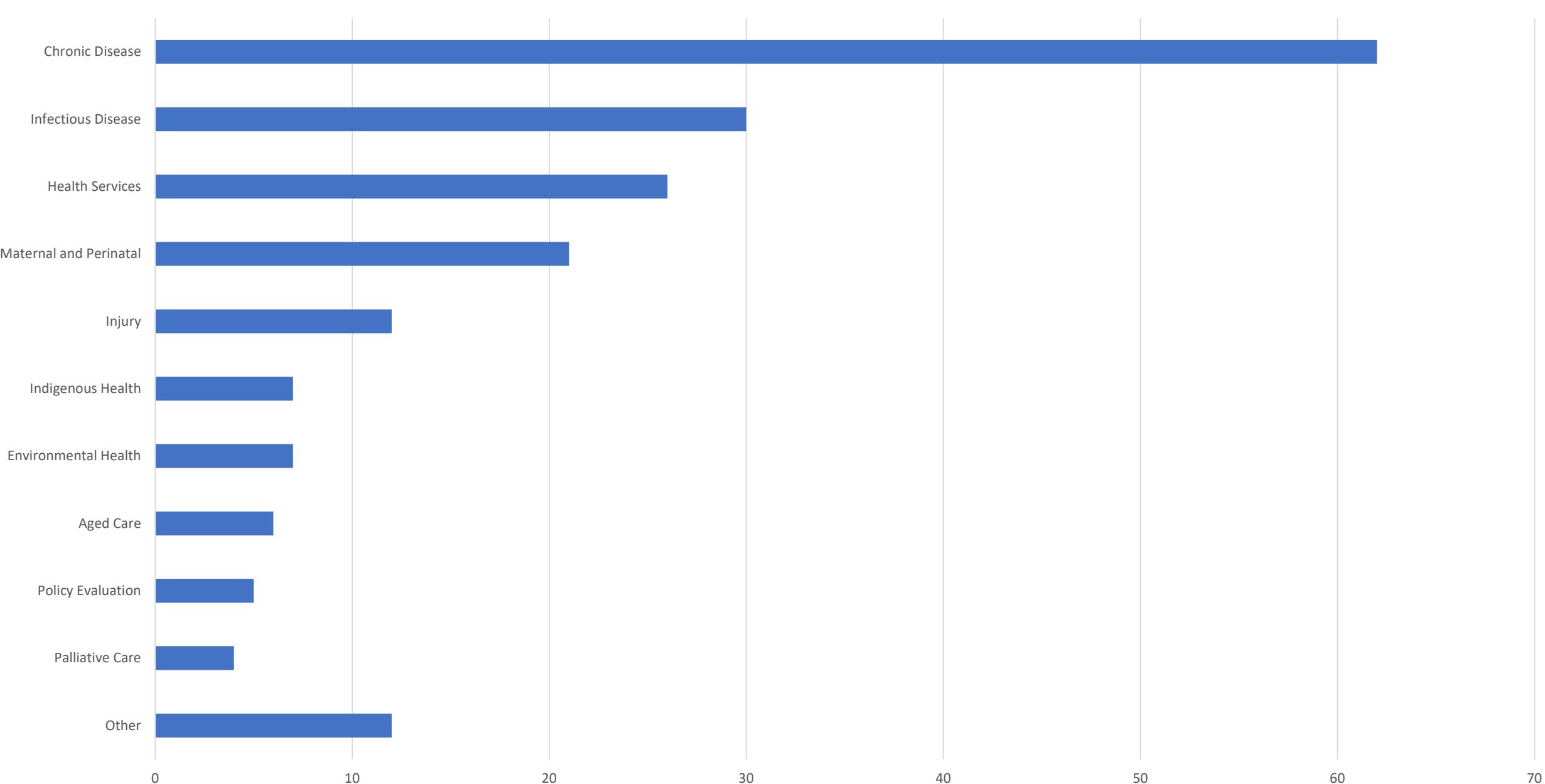


# Research use of linked data

Annual count of peer reviewed publications based on linked data



# Research topics 2010-2019



# Use of linked data in Queensland Health

---

- Person-level data rather than event-level data enables:
  - Insight into post-discharge outcomes (e.g., readmissions, deaths), patient journeys, scope of activity for an individual
  - Accurate cost estimates for models of care and care optimisation. Enables evidence-based service funding.

# Use of linked data in Queensland Health

---

- Supply to units across all QH Divisions
  - patient safety monitoring (CED)
  - purchasing decision-making (HPSP)
  - service and system planning (SPPD, CED)
  - monitoring and reporting patient outcomes (CED)
  - service evaluation (CHO, CED, HPSP, SPPD)
  - Auditing service use (various)

# Use of linked data in Queensland Health

---

- Use for predictive analytics to target patient care for initiatives such as the Nurse Navigator program and other integrated care programs (supply of high service use patient lists (SCaNNR) to HHSs)
- Identification of deceased patients prior to patient follow up contact and for supply to HHSs for waiting list and patient master list management



# Use of linked data in Queensland Health

---

- Monitoring the completeness and accuracy of key QH data collections and reconciliation of inconsistent information
- Maintaining data repositories across data collections for analytics and reporting on congenital anomalies, terminations of pregnancy, injury and maternal and perinatal reporting
- Feeding into BI initiatives – e.g., chronic disease and congenital anomaly dashboards

# Working with other government agencies

---

- Cost effectiveness of **homelessness** interventions
- **Black lung** incidence
- Impact of mental illness on **criminal offending** and **child welfare**
- **Transport accident** incidence and outcomes
- Impact of **paediatric infections** on long term **education outcomes**

# Working with other government agencies

---

- Improving **Indigenous identification** on National Death Registration data
- Informing a national registry of **congenital anomalies**
- Various national projects through involvement in National Integrated Health Services Information (**NIHSI**) project

# Client services team

---

- Client Services team added in 2018
- Ongoing 100% satisfaction with service in annual client survey 😊
- Website is being continually updated – programming code and tips for working with linked data, technical reports on data quality issues, dataset information

# Data Linkage Training Resources



These resources are provided by the Statistical Analysis and Linkage Unit, Queensland Health for researchers and users of linked data.

## Counting number of unique persons

Standard SQL / SAS Proc SQL	SAS Data step	R (Base R)	Stata
<pre>select  count(distinct id) from    dset;</pre>	Use Proc SQL	<pre>length(unique(dset\$id)) #or length(unique(dset[["id"]]))</pre>	<pre>by id, sort: gen nvals = _n == 1 count if nvals</pre>

## Introduction

- The ultimate aim of working with linked data...

## Relational data

- Example
- 'Tidy data'

## Format, coding and efficiency

- 'Wide' versus 'Long' data
- Analysis efficiency
- Coding and logical efficiency
- Storage efficiency for larger datasets
- Visual efficiency

## Basic Joins

Type of Join	Standard SQL / SAS Proc SQL	SAS Data step	R (Base R)	Stata
Inner Join	<pre>create table mergeddata as select  d1.id,         d2.var from    dset1 d1 inner join dset2 d2 on      d1.id = d2.id;  /* or */  create table mergeddata as select  d1.id,         d2.var from    dset1 d1,         dset2 d2</pre>	<pre>data mergeddata; merge  dset1 (in = d1)        dset2 (in = d2); by    id; if    d1 and d2; run;</pre>	<pre>mergeddata &lt;- merge(   dset1,   dset2,   by="id" )</pre>	<pre>merge 1:1 id using "c:\data\dset2.dta"</pre>

## Counting and units

- Establish counts of records and unique records
- Unit of analysis

## Merging/joining data

- Inner join
- Left join (also known as a left outer join)
- Full join (also known as a full outer join)
- Many-to-one and one-to-many joins

## Sequencing events

- Example

## Code snippets

- Counting rows and unique persons
- Basic joins
- Aggregating

## References

- Further reading

## Unit of analysis

Identifying the appropriate 'unit of analysis' is an important step in analysing linked data.

To give a simple example, consider the datasets introduced earlier.

You can see that there are 6 people in the study cohort, of which only 4 are represented in the hospital admissions data. These 4 people represent a total of 6 admissions.

If you were to calculate statistics related to deaths based on this data, you could say:

- 1/6 = 16.7% of the *cohort* died, or
- 1/4 = 25.0% of the *admitted people* died, or
- 1/6 = 16.7% of all *admission episodes* were recorded as ending in a death.

The meaningfulness of these calculations depends on your research question. You must however be clear about both the numerator and denominator of your analysis when dealing with linked data that is measured at different levels.

# More information and to request linked data

---

## Qld Health data linkage unit

- <https://www.health.qld.gov.au/hsu/link/datalink>
- email: [HSBresearch@health.qld.gov.au](mailto:HSBresearch@health.qld.gov.au)

## Qld Health research ethics and governance unit

- [https://www.health.qld.gov.au/hiiro/html/regu/regu\\_home](https://www.health.qld.gov.au/hiiro/html/regu/regu_home)
- email: [HIIRO\\_REG@health.qld.gov.au](mailto:HIIRO_REG@health.qld.gov.au)