

Factors Associated with Time Taken to Complete a Linkage Request

Risk factor	Shorter	Medium	Longer
Data quality:			
Quality of person identifiers in data collections being linked	Good quality data with no missing or misspelt details - little or no manual clerical review required	Some errors or missing data - a moderate amount of links need manual review	Missing identifiers and/or many errors in identifiers - large amount of manual clerical review required
Format of data provided for linkage	All data items able to be easily read in to the linkage unit data systems with no issues with format or misplaced delimiters	Some intervention required to import data into data linkage unit systems	Data not readily able to be imported into data linkage unit systems with manual input required to correct formats and ensure all data items are read in correctly
Size of request and amount of linkage required:			
Data collection inclusion in Master Linkage File (MLF)(prelinked data)	All data collections requested are included in MLF	Some data collections requested are included in the MLF	All data collections need to be specially linked for the project
Number of records that require linking for a project	No linkage required (all in MLF)	Some custom linkage required and datasets are moderately sized (hundreds of records)	Thousands or millions of records required to be linked
Time period requested	Short, recent time period (eg. 1-2 years)	Longer, but recent time period (eg last 5 years)	Long time period, especially if multiple versions of ICD codes are included and including years of data going back in time that may be stored in different systems Very recent data that has not yet been submitted by those collecting the information or validated
Complexity of request:			
Number of jurisdictions involved	Single state or territory		Multiple jurisdictions or inclusion of Commonwealth data (eg. MBS, PBS)
Number of data collections involved	One dataset or single custodian responsible for approving and extracting all data collections requested		Multiple data collections and custodians need to approve data access and extract data for linkage
Study design	Simple cohort specifications		Control group or complex cohort specifications, for example involving characteristics from across multiple data collections or requiring large amounts of coding to describe inclusion and exclusion criteria or to generate data items in the required format

	Shorter	Medium	Longer
Capacity issues:			
Data linkage team capacity and competing priorities	Few other requests and full staff available		Staff on leave, many other requests to process, urgent departmental requests
Data custodian experience with providing data extracts	Custodian has staff experienced in extracting and formatting data	Some experience but limited capacity	Data collected for operational purposes and no resource allocated for extracting and formatting data
Clarity of the request:			
Description of requirements provided	Variable names clearly specified using the terminology used by the data custodian, research questions and proposed analysis and use of the data clearly described and/or required format of data items clearly described		List of data items described in generic terms with no link between what is requested and how it will be used provided
ICD and ACHI (diagnosis and procedure codes) provided	All provided, including accounting for changes in edition and version for all time periods requested	Some ICD/ACHI codes provided but no consideration of potential changes over time	No ICD/ACHI codes provided - just broad descriptions of conditions/procedures of interest

[For more information see Data Linkage In Queensland](#)