Author/s:
Aitken, Z;Walmsley, S;M Bishop, G;Badji, S;Fortune, N

Title:
Methods used to construct disability indicators in linked administrative datasets: a systematic scoping review

Date:
2025-12-01

Citation:
Aitken, Z., Walmsley, S., M Bishop, G., Badji, S. & Fortune, N. (2025). Methods used to construct disability indicators in linked administrative datasets: a systematic scoping review. Population Health Metrics, 23 (1), pp.22-. https://doi.org/10.1186/s12963-025-00386-w.

Persistent Link:
https://hdl.handle.net/11343/363014

Population Health Metrics

REVIEW

Open Access

Methods used to construct disability indicators in linked administrative datasets: a systematic scoping review

Zoe Aitken[1*], Sarah Walmsley[1], Glenda M Bishop[1], Samia Badji[2] and Nicola Fortune[1,3]

Abstract

Background In this scoping review, we aimed to examine evidence on methods used to construct disability indicators in linked administrative datasets and describe the approaches used to assess the validity of the indicators.

Methods Medline (Ovid) and Embase (Ovid) were searched for studies published between January 2010 and June 2023. Original, peer-reviewed studies that aimed to construct a disability indicator using linked administrative data sources were included. Studies identifying any types of disability were included, but not those which defined the target population in terms of specific health conditions. We produced a narrative synthesis of findings related to disability indicator construction methods and validation approaches.

Results Thirty-six relevant studies were included, with 30 of those identifying a cohort of people with intellectual and/or developmental disability. Health data sources were most commonly used for indicator construction, with 33 of the studies using at least one health data source. Disability and education sector data sources were also commonly used. Diagnostic codes were used for disability identification in 34 of the 36 studies; 16 used diagnostic codes alone and 18 used diagnostic codes along with other information. A subgroup of 19 studies had a primary aim to create a disability cohort or estimate disability prevalence. Thirteen of these 19 studies compared their estimated prevalence rates with previously published estimates. Only five studies conducted testing to investigate the extent to which their derived disability indicator captured the intended target population.

Discussion We found a paucity of evidence on methods for identifying a target population of people with diverse disabilities. In the existing literature, diagnostic information is relied upon heavily for disability identification, likely due to a lack of other types of disability-relevant information in administrative data sources. Use of derived disability indicators within linked data holds potential to advance research regarding people with disability. It is crucial, however, to conduct and report validation testing to understand the strengths and limitations of the indicators and inform their use for specific purposes.

Keywords Disability, Disability indicator, Linked data, Administrative data, Scoping review

*Correspondence:
Zoe Aitken
zoe.aitken@unimelb.edu.au
[1]Present address: Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, VIC 3010, Australia
[2]Centre for Health Economics, Monash Business School, Monash University, Melbourne, VIC 3145, Australia
[3]Centre for Disability Research and Policy, The University of Sydney, Sydney, NSW 2006, Australia

Introduction

The United Nations Convention on the Rights of Persons with Disabilities (CRPD) [1] is an international human rights treaty. It affirms the need for people with disability to be guaranteed full enjoyment of human rights and fundamental freedoms across all aspects of life, recognising that "persons with disabilities continue to face barriers in their participation as equal members of society and violations of their human rights". The CRPD is nearing universal ratification, with 186 States Parties in 2023 [2]. Governments that have ratified the Convention are required to adopt "appropriate legislative, administrative and other measures" to implement the rights recognised in the Convention (Article 4) [1]. Further, they need to report on progress towards achieving these rights, and to collect statistics to inform policy implementation and to monitor the Convention (Articles 31, 33 and 35) [1].

Based on 2021 Global Burden of Disease Study data, the World Health Organization estimates that 16% of the world's population has significant disability [3]. Across the globe, people with disability continue to experience poorer health outcomes compared with the wider population [4–7]. There is growing recognition that these health disparities are largely socially produced and modifiable, driven by inequities in relation to social determinants of health [8, 9]. In addition, people with disability often experience barriers to gaining equitable access to health services, particularly preventive health care and chronic disease management [4, 10, 11]. Article 25 of the CRPD affirms the right of people with disability to enjoy the highest attainable standard of health without discrimination on the basis of disability [1]. Other articles of the CRPD address rights in relation to social determinants of health, such as housing, income, education and employment. With this perspective, it is the role of public health to tackle disability-related health inequities, with policies and interventions informed by a robust evidence base that provides insights into the underlying causes of the inequities.

Population data sources, including surveys and administrative data collections, are crucial for generating evidence on disability-related health inequities and their causes. When using population data sources for this purpose, a key consideration is how people with disability are identified in the data. The International Classification of Functioning, Disability and Health (ICF) provides a standard framework and set of classifications for capturing and organising information about functioning and disability [12]. In the ICF, disability is an umbrella term for impairments, activity limitations and participation restrictions, and denotes the negative aspects of the dynamic interaction between a person's health condition(s) and environmental and personal factors. The ICF conception of disability is consistent with the CRPD recognition that "disability results from the interaction between persons with impairments and attitudinal and environmental barriers that hinders their full and effective participation in society on an equal basis with others" (Preamble (e)) [12].

Many countries include questions in population surveys and censuses to identify people with disability, often designed to operationalise a conceptual definition of disability aligned with the ICF or CRPD [13–15]. These data sources are valuable for estimating disability prevalence, reporting on outcomes, quantifying inequalities between people with and without disability, and informing policy decisions about allocation of funding and services. However, they also have limitations. Surveys and censuses that capture cross-sectional, point-in-time data, do not provide insights into the temporal dynamics of disability and associated outcomes for individuals. Because survey data are collected from a sample of the population, there is additional uncertainty associated with estimates, some population subgroups may be missing or under-represented, and often the data are unsuitable for analyses focusing on rare outcomes or small population subgroups.

Administrative data collections are also valuable sources of data for disability statistics. They have a number of advantages, such as the inclusion of records for all individuals who have contact with the specific programs or services they are associated with, they provide information about services accessed and related outcomes, and in many cases they contain longitudinal information (either specific dates or defined time intervals such as quarterly or annually). However, in most countries, it is not possible to identify people with disability in administrative data collections for mainstream services in sectors such as health, housing, and justice, because only information critical to administering or delivering the service is collected and disability status is not usually considered necessary information for most programs or services. This limits ability to monitor equity of access to health and other services, and to compare outcomes between people with and without disability [16]. Additionally, disability information contained within administrative datasets is generally collected for a specific purpose, thus identifying a specific subgroup of people with disability, such as people accessing disability income support or using disability services, rather than identifying individuals based on a conceptual definition of disability.

Advances in data linkage in many countries offer the possibility of identifying a population of people with disability using data items present in one or more administrative data collections to derive disability indicators. The term 'data linkage' refers to bringing information from different sources together into a single file and linking records that relate to the same person or entity [17]. The

derived disability indicators can be used to conduct analyses disaggregated by disability status for data contained in diverse data sources, linked together at the individual level. This provides an opportunity to better understand the experiences of a large population of people with disability, addressing issues of statistical power and bringing together disability data with other data sources describing outcomes including health and social determinants of health (e.g., employment, health service use). For example, Australia has recently released a National Disability Data Asset, which will link together de-identified records for individuals across administrative and survey datasets, and construct a disability indicator drawing on disability-relevant information contained in multiple constituent data sources [18].

This approach of linking existing datasets to derive disability indicators holds the promise of substantially increasing the evidence base on a large number of diverse health outcomes without imposing additional data collection burden. Additionally, constructing disability indicators using disability data from multiple data sources may identify the target population of people with disability more comprehensively than would be achieved using a single administrative data source. However, it is important to understand the strengths and limitations of such derived disability indicators. For example, to what extent does the indicator align with ICF-based definitions of disability used in population surveys? Also, are there subgroups of the target population of people with disability that may be less comprehensively captured by the indicator?

Previously published reviews have addressed the use of administrative data to estimate prevalence of intellectual and developmental disabilities [19], and the development and validation of algorithms to identify reproductive-aged women with physical and sensory disabilities in administrative health data [20]. To date, no reviews have been conducted that examine approaches for using linked administrative data to construct disability indicators and describe their strengths and limitations. This is an important gap, particularly given the potential value of analysing linked administrative data to investigate health and other outcomes for people with disability, which will assist countries in meeting their obligations under the CRPD.

In this scoping review, we investigate methods for constructing disability indicators from linked administrative data sources, including both 'specific' indicators focused on a particular disability group (e.g. intellectual disability (ID)) and 'broad' indicators that include people with diverse disabilities. We focus on the approaches researchers use to assess the validity of the derived disability indicators and to understand strengths and limitations that could impact the use of such disability indicators to measure outcomes for people with disability.

Methods

We used scoping review methodology because our objective was to scope the body of literature describing methods used to construct specific disability or broad disability indicators in linked administrative datasets. Scoping reviews are an appropriate approach when the purpose is to examine how research is conducted on a certain topic, rather than to appraise and synthesise research findings on a particular question [21]. As scoping reviews provide an overview of evidence on a given topic, included studies are generally not assessed for methodological quality or risk of bias [22].

We systematically identified studies and extracted relevant information, following the five-step process set out by Arksey and O'Malley [23] and refined by Levac et al. [24]: identifying the research question/s, identifying relevant studies, study selection, data extraction, and collating, summarising and reporting results.

Identifying the research questions

We set out to address two research questions:

1. What are the characteristics of published methods for constructing specific and broad disability indicators in linked administrative datasets, in terms of disability cohort identified, data sources and data items used, and algorithms applied?
2. What approaches have authors used to validate methods used for constructing disability indicators in linked administrative datasets?

Identifying relevant studies

Articles were identified through searches of Medline (Ovid) and Embase (Ovid). The search strategy was developed by the authors with advice from an academic librarian, combining search terms for disability and disability groups together with search terms for administrative or linked data. The search terms are detailed in the Supplementary Table 1.

A first search identified articles published from January 2010 to August 2021. A second search identified articles published from August 2021 to June 2023. The searches were limited to articles written in English. The Embase search was further limited to full text or in press articles to exclude conference abstracts.

Study selection

We adopted the schema set out in the JBI Manual for Evidence Synthesis to define inclusion criteria [22], as follows:

Population People with disability. We are conceptualising disability as a population of interest rather than an outcome, recognising that health disparities for this population are often not caused by underlying disability, and instead relate to avoidable disadvantage and inequity associated with social determinants of health [5, 9]. We included studies focused on a disability cohort encompassing diverse disabilities or specific disability groups (e.g., ID). We excluded studies in which the cohort was not defined by disability but instead used other criteria based on research questions where disability was an outcome or a covariate because our scoping review was focussed on studies that deliberately identified people with disability in the initial cohort. Studies in which the cohort was defined in terms of specific health conditions (e.g., spina bifida, epilepsy) were excluded. We included studies focusing on 'autism spectrum disorder'; while autism is a diagnosis, it is also regarded as a disability group in many countries.

Concept Construction of a specific or broad disability indicator using disability information from two or more linked administrative data sources. We defined administrative data as jurisdiction-wide data collections produced routinely in the course of delivering services or programs.

Context Studies using data at national or state/province/region level, where the disability cohort was identified for use in statistical analysis (e.g., to estimate prevalence of disability or to examine outcomes for people with disability).

Types of evidence sources Original peer-reviewed studies. We excluded conference abstracts, study protocols, letters, reviews, editorials, and book chapters reporting on previously published studies.

Selection of articles

The first and second searches combined identified 2,175 articles (Fig. 1). After the removal of 599 duplicates, 1,576 articles were retained for screening. The data management software Covidence was used to assist the article selection process. Title and abstract screening were conducted independently by ZA and NF for all articles, applying the inclusion criteria above. All disagreements were discussed until consensus was reached.

Full-text review was conducted independently by both ZA and NF for all of the 156 articles retained following title and abstract screening, with disagreements discussed between ZA and NF to reach consensus and ensure a consistent approach. The most common reason for exclusion at the full-text review stage was use of a single data source to identify the disability cohort (67% of excluded studies).

Reference lists of articles retained following full-text review were searched to identify additional eligible studies. Title and abstract screening and full-text review for these articles were conducted in the same way as for articles identified in the initial searches.

Upon review of the full set of included articles in preparation for data extraction, the authors decided to conduct a second round of article exclusion to focus only on articles describing novel methods for constructing disability indicators in linked administrative datasets. Several of the included studies used established methods described in previously published articles. In these cases, only the first published article was retained, to avoid duplication of information. Subsequently-published articles were included if they described a substantial modification to a method previously described (e.g., adding a new data source) or reported new findings concerning the method or validation testing. This process resulted in 36 articles being deemed eligible for inclusion.

Data extraction

Following discussion among ZA and NF, the key information for extraction was determined, which included: data on country and year of publication; purpose of the study; disability cohort identified; number and type of administrative data sources used for constructing the disability identifier; methods used to construct disability identifier (including data items utilised and algorithms applied); nature of validation (if conducted); key findings concerning method for identifying disability cohort (including results of validation, if conducted). For the first 5 articles, extraction was conducted independently by ZA and NF using an Excel spreadsheet template. The results were compared and discussed to ensure consistency and relevance to the research questions. Data from the remaining 31 articles were then extracted by NF.

Collating, summarising and reporting the results

The extracted data were used to produce a descriptive summary of each article. For each of the information fields listed above, we reviewed the data recorded and devised a simple set of categories to facilitate a summary overview of results, as reported in the following section. For the last three information fields (methods used to construct disability identifier; nature of validation; key findings concerning method for identifying disability cohort), we undertook a narrative synthesis, identifying key themes and similarities and contrasts between the included studies, which we refined through discussion.

Results

In this section, we first describe the characteristics of all included articles. Then, for studies whose primary aim was to construct a cohort of people with disability, we

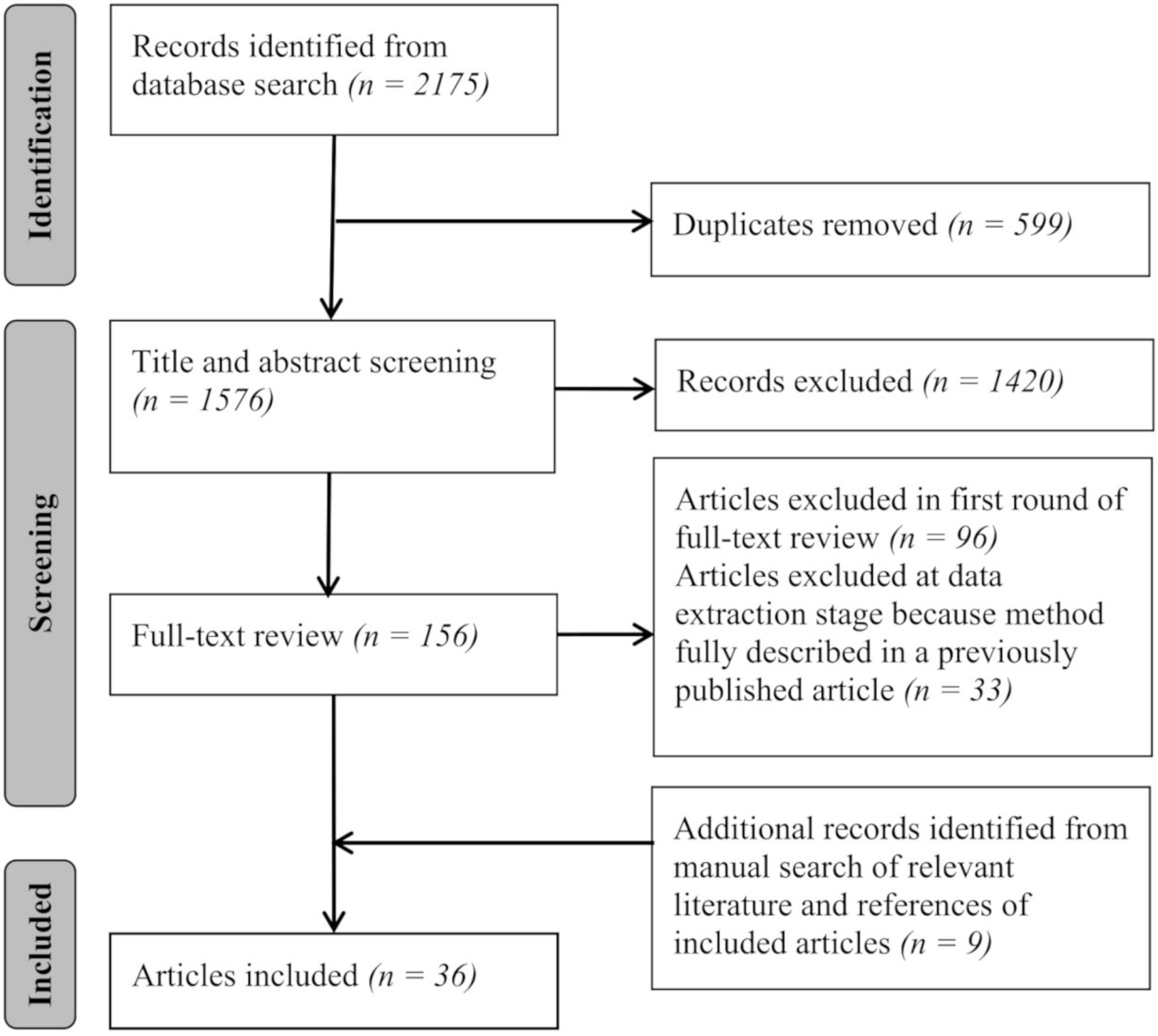

Fig. 1 Literature search and screening flow chart

describe findings relevant to the methods used to construct the disability indicator (research question 1) and outline the approaches used by authors to validate the disability indicators (research question 2).

Characteristics of included studies

The characteristics of the 36 included studies [25–60] are summarised in Table 1. Of the 36 included studies, 10 used data from Europe (Denmark [37, 40], England [54], Finland [56, 60], Italy [42], Netherlands [46], Sweden [30, 48], Wales [26]), 17 from Northern America (Canada [27, 32, 35, 36, 38, 41, 45, 47, 49, 51–53, 57], United States of America [28, 29, 34, 39]) and 9 from the Pacific region (Australia [25, 31, 33, 50, 55, 58, 59], Fiji [44], New Zealand [43]).

Thirty of the 36 studies identified disability cohorts that could be defined as 'intellectual and developmental disabilities', comprising fifteen studies focusing on people with ID [25, 29, 30, 34, 37, 39, 40, 46, 50, 54, 56–60], six on autism [31, 43, 45, 47, 48, 55], three on developmental disability (DD) [32, 35, 36], five including people with ID and either DD or autism [26, 27, 51–53], and one focused on DD and autism [41]. Of the remaining studies, two focused on dementia [42, 49], one focused on visual impairment [44], and three included diverse disability types [28, 33, 38].

Of the 36 studies, 10 included children only (below 18 years) [26, 29, 34, 39, 44, 45, 47, 50, 54, 55], 9 included adults only (18 + years) [28, 32, 36, 42, 46, 49, 51, 52], 10 included children and adults [30, 31, 37, 38, 40, 56–60],

Table 1 Characteristics of included studies

Study	Country	Age	Disability cohort[1]
Bacigalupo et al. 2022	Italy	50 + years	Dementia
Bowden et al. 2020	New Zealand	0–24 years	Autism
Brameld et al. 2018	Australia	20 + years	ID
Brophy et al. 2018	Wales	< 18 years	ID, Autism
Calver et al. 2021	Canada	0–19 years	ID, DD
Cama et al. 2010	Fiji	0–15 years	Visual impairment, low vision, blindness
Coo et al. 2017	Canada	2–14 years	Autism
Cuypers et al. 2021	Netherlands	18 + years	ID
Deroche et al. 2017	USA	50–75 years	Blind/low vision, ID, spinal cord injury
Dodds et al. 2009	Canada	3–16 years	Autism
Griffith et al. 2011	USA	5–11 years	ID
Hirvikoski et al. 2021	Sweden	22–33 and 0–81 years	ID
Hwang et al. 2019	Australia	5–64 years	Autism (with/without ID)
Idring et al. 2012	Sweden	4–23 years	Autism
Kosteniuk et al. 2015	Canada	45 + years	Dementia
Leonard et al. 2003	Australia	6–15 years	ID
Lin et al. 2013	Canada	18–64 years	ID, DD
Lin et al. 2014	Canada	18–64 years	ID, DD
Lloyd et al. 2022	Canada	0–19 years	ID, DD
Lunsky et al. 2012	Canada	18–64 years	DD
Maclean et al. 2017	Australia	0–20 years	Disability
Madley-Dowd et al. 2022	UK	Children	ID
Mann et al. 2013	USA	3–6 years	ID
Marquis et al. 2020	Canada	0–19 years	DD
Nielsen et al. 2023	Australia	2–15 years	Autism
Nurminen et al. 2022	Finland	All ages	ID
Ouellette-Kuntz et al. 2009	Canada	All ages	ID
Reppermund et al. 2017	Australia	All ages	ID
Reppermund et al. 2019	Australia	All ages	ID
Shooshtari et al. 2011	Canada	20 + years	DD
Su et al. 2021	Denmark	0–36 years	ID
Tarasoff et al. 2020	Canada	Women; 15–44 years	Physical, sensory, ID, DD
Wang et al. 2015	USA	3–9 years	ID
Wang et al. 2023	Denmark	2–40 years	ID
Weise et al. 2018	Canada	18–24 years	DD, Autism
Westerinen et al. 2007	Finland	All ages	ID

1. Disability groups in the cohort (DD = developmental disability; ID = intellectual disability)

one included young people only (18 to 24 years) [41] and 6 included children and young adults [27, 33, 35, 43, 48, 53].

Data sources and data items used to produce the disability indicators for each study are described in Supplementary Table 2. A range of types of administrative data sources were linked for cohort identification. Health data were the most common sources (used in 33 studies, including general practitioner records, hospital records, and clinical billing records, e.g., physician health insurance claims), with 10 studies using data only from health sources [27, 28, 32, 35, 37, 38, 40, 47, 49, 51]. This was followed by disability data sources (22 studies, including disability support, disability income support, long-term care, and community-based rehabilitation sources), and education data sources (15 studies).

Most studies (34 out of 36) used diagnostic codes (International Classification of Diseases (ICD), Diagnostic and Statistical Manual of Mental Disorders (DSM), and general practitioner diagnosis codes) for disability identification [25–28, 30–45, 47–60]. Of these, 16 studies used diagnostic codes alone [26–28, 32, 33, 35, 37, 38, 40, 41, 43, 44, 47, 51, 52, 56], and 18 used diagnostic codes in conjunction with additional information such as disability category, or information regarding income and education support [25, 30, 31, 34, 36, 39, 42, 45, 48–50, 53–55, 57–60]. Only two studies did not use diagnostic codes for disability identification: one identified disability based on whether a child received special education

within the public school system or ID-related services [29], the other used reasons for entitlement to chronic care services and welfare benefits [46].

Seventeen of the included studies had a primary aim of analysing specific outcomes for a disability cohort [25–41]. Eleven of those studies explored health-related outcomes, including health service usage [27, 28, 32, 41], mortality [25, 30, 31], prescription usage [26], health characteristics or disease prevalence rates [36, 38, 40]. Four studies examined associations between maternal pregnancy characteristics and risk of child ID [29, 34, 37, 39], one explored risk of maltreatment within the child protective system [33], and one explored mental health outcomes of the parents of children with DD [35]. These 17 studies focused on reporting findings on the outcomes explored but did not report findings related to methods used to identify the disability cohort or to validate their disability indicator, which was a key research question of the present scoping review.

The primary aim of the other 19 studies was to create a disability cohort or estimate disability prevalence [42–60]. Details of the method used to construct disability indicators, findings relevant to the method for constructing the disability cohort, and the validation approach for these 19 studies are outlined in Table 2. We focus on these studies in the following section, where we report findings concerning methods for identifying the disability cohort and validation approaches.

Methods used to construct disability indicators

The disability cohorts included in these 19 studies were autism (5 of 19) [43, 45, 47, 48, 55], visual impairment (1 of 19) [44], dementia (2 of 19) [42, 49], ID (8 of 19) [46, 50, 54, 56–60], and ID with DD (3 of 19) [51–53]. No studies aimed to identify a target population of people with diverse disability. The number of administrative data sources used to derive the disability indicator ranged from two to eight, with 13 of the 19 studies linking four or more sources.

Thirteen of the 19 studies used methods that only required a single instance of identification in any of the constituent data sources for disability cohort inclusion [43, 44, 46, 48–50, 53, 55–60]. One study required two or more of eight specified criteria be met across constituent data sources [54]. Another study required two or more prescriptions of drugs for dementia within 12 months within the prescription drug data source, or at least one instance of identification in the three other data sources [42]. Another study required at least two ID-related physician claims, or at least one instance of identification in any of the four other data sources [52].

The remaining three studies created and tested the efficacy of multiple algorithms for disability cohort identification [45, 47, 51]. The basis for selecting preferred algorithms varied in these three studies. One tested five algorithms to identify autistic children and youth, with two of those algorithms requiring two or more instances of identification in medical services data [45]. They recommended algorithms that utilised physician claims and education data only, as few additional cohort members were identified in hospital and mental health services data [45]. Another compared seven algorithms for identifying autism diagnoses within hospital, physician billing and mental health data sources, with different combinations of requirements, including single or multiple codes from one or more of three administrative data sources [47]. They found that the algorithm with the best test characteristics identified the cohort based on at least one autism related code in any one of the three constituent sources [47]. This algorithm yielded an overall sensitivity of 69% and a specificity of 77%, with specificity being highest within the hospital data source and sensitivity highest within the physician billing data source [47]. The final study compared broad, intermediate and narrow algorithms to identify a cohort of people with intellectual and developmental disabilities [51]. The intermediate algorithm, which required a single contact with either an inpatient or emergency department or a minimum of two physician visits, was preferred as the resulting prevalence was closest to that reported in two comparison target populations (0.50%), suggesting that the broad algorithm captured higher rates of false positives, and the narrow algorithm higher rates of false negatives [51].

The period of time over which cohort members were identified in the constituent data sources varied from one year to 23 years. In three of the 19 studies, the time period varied between data sources [48, 59, 60]. In three studies it was not possible to determine the data period from the information provided in the paper.

Fourteen of the 19 studies reported the percentage of the cohort identified in each data source [42–44, 46, 49–53, 55, 57–60]. Ten studies reported the percentage of the identified cohort uniquely identified in each data source [42, 49–53, 55, 57, 59, 60], that is, the percentage that would be missed if the data source were excluded. A number of those studies identified a high proportion of individuals uniquely identified in each of their data sources, indicating that the inclusion of additional sources contributed significantly to cohort identification. For example, a study that linked data from a government disability agency with education data found that approximately half of the people identified with ID in each data source were unique to that source [50]. Another study, which linked health, education and income assistance data sources to identify a cohort of people with ID, reported high proportions (ranging from 19 to 36%) of uniquely identified cohort members across the sources [57], and found that the hospital data source was the

Table 2 Studies with a focus on constructing cohort or estimating disability prevalence

Study	Disability cohort[1]; age	Method used to construct disability indicator	Findings relevant to method for constructing disability indicator	Validation approach
Baciga-lupo et al. 2022	Dementia; 50 + years	Identification in any of 4 data sources: Hospital discharge; Drug prescription; Exemption registry; Long-term care (≥ 2 prescriptions within 12 months; ≥ 1 record in other data sources). 5 year data period.	Case identification by source: drug prescriptions (65%; 42% unique), hospital discharge (19%; 2.4% unique), exemption from health care co-payment (7.4%; 0.3% unique), long-term care (4.1%; 0.2% unique). 68% of cases were correctly identified by algorithm; 3.8% of controls identified as cases. Algorithm performance varied between age groups (higher for 75 + years age group) and regions. Variables associated with false negatives included younger age, lower education, milder impairment, shorter disease duration and lower use of medication.	Algorithm tested against reference population of cases and controls determined by clinical evaluation. Comparison with previously published prevalence estimates.
Bowden et al. 2020	Autism; 0–24 years	Identification in any of 3 data sources: Mental health services; Hospital discharge; Disability support services. 6 year data period.	Case identification by source: disability support services (77%), mental health services (25%), hospital discharges (22%). Variation in distribution of data source by age, ethnicity and deprivation. Changes to disability support eligibility criteria associated with differing rates of co-occurring intellectual disability. Estimated approximately 40% undercount.	Comparison with previously published prevalence estimates and demographic profiles.
Cama et al. 2010	Visual impairment, low vision, blindness; 0–15 years	Identification in any of 4 data sources: Community Based Rehabilitation; School for the Blind; Eye clinic; Screening programs. Health care professional 'key informants' used to identify new case. 16-month data period.	Case identification by source: eye clinic records (36%), School for the Blind (25%), Community Based Rehabilitation (21%), screening programs (15%), key informants (4%). 12% identified through 2 data sources.	Comparison with previously published prevalence estimates.
Coo et al. 2017	Autism; 2–14 years	Identification in following data sources: Physician claims; Hospital discharge; Education; Mental health services. Multiple algorithms examined, varying number of sources and number of records required in physician data. 15-year data period.	The 10–12% of service provider-reported cases not identified in administrative data were more likely to be born outside province, diagnosed at older age, and not have ID or DD. Recommended algorithms using physician claims and education data only; hospital and mental health services data identified few additional cases.	Sensitivity of different algorithms determined by comparing to cases reported by service providers. Information sought from parents to verify autism diagnoses for individuals only identified in administrative data (low response rate).
Cuypers et al. 2021	ID; 18 + years	Identification in either of 2 data sources: Chronic care database; Welfare database. Single reference year.	Case identification by source: chronic care database (66%), welfare database (35%).	Comparison with previously published prevalence estimates.
Dodds et al. 2009	Autism; 3–16 years	Seven algorithms were derived from combinations of requirements for single or multiple autism codes from one or more of three administrative databases (hospital discharge, physician billing, mental health outpatient). 16-year data period.	The algorithm with the best test characteristics was based on one autism code in any of the three databases (sensitivity 69%, specificity 77%). Of the three databases, specificity was highest for hospital data and sensitivity was highest for physician billing data. The sensitivity of the administrative data in identifying an autism diagnosis was similar across most maternal and neonatal characteristics investigated.	Accuracy of each algorithm assessed by comparing against 'gold standard' clinical autism database.

Table 2 (continued)

Study	Disability cohort[1]; age	Method used to construct disability indicator	Findings relevant to method for constructing disability indicator	Validation approach
Idring et al. 2012	Autism; 4–23 years	Identification in any of 4 data sources: Habilitation Register, Clinical Database for Child and Adolescent Psychiatry, Public healthcare services, National Patient Register. Data period from 6 to 23 years, depending on data source.	Case identification by source: Habilitation Register (68%), Clinical Database for Child and Adolescent Psychiatry (58%), public healthcare services (44%), National Patient Register (14%). 96% of reviewed case-notes consistent with autism diagnosis; Autism diagnoses confirmed in 85% of affected twins.	Case-note review (3.4% sample). Cross validation with co-existing cases in a national twin study (0.5% of cases). Comparison with previously published prevalence estimates.
Kosteniuk et al. 2015	Dementia; 45 + years	Identification in any of 4 data sources (Hospital; Physician; Prescription drugs; Long-term care). 12-year data period (12-mth incidence and prevalence calculated).	Greatest proportion of incident cases first identified in long-term care data (35%), followed by physician (30%), hospital (29%), and prescription data (6.5%). Greatest proportion of prevalent cases first identified in physician data (40%), followed by long-term care (25%), hospital (24%) and prescription data (11%).	Comparison to incidence and prevalence estimated by other studies.
Leonard et al. 2003	ID; 6–15 years	Identification in Disability Services Commission (DSC) records, or cases registered in single year with any of 3 education agencies.	Case identification by source: DSC (67%; 45% unique), education agencies (76%; 54% unique).	Comparison with previously published prevalence estimates.
Lin et al. 2013	ID, DD; 18–64 years	3 nested algorithms using 5 data sources: Psychiatric inpatient; Acute care inpatient; Same-day surgery (SDS); ED visits; Physician visits. Broad algorithm: identification in any inpatient or SDS discharge or any ED visit or any physician visit, from database inception. Intermediate algorithm: identification in any inpatient or SDS discharge or any ED visit or ≥ 2 physician visits, from database inception. Narrow algorithm: identification in any inpatient or SDS discharge or any ED visit, 3-year data period.	Case identification by source (broad algorithm): Psychiatric inpatient (6.6%; 3.4% unique); Acute impatient (26%; 11% unique); SDS (11%; 2.4% unique); ED visits (5.7%; 0.8% unique); Physician visits (79%; 62% unique). Compared to broad algorithm, intermediate algorithm identified 66% of individuals and narrow algorithm identified 23% of individuals. Compared to intermediate cohort: narrow cohort had older age profile, more rural residents, higher rates of psychiatric co-morbidity and diabetes, and lower rate of asthma; broad cohort had more women and urban residents, and lower rate of psychiatric co-morbidity.	Comparison with previously published prevalence estimate.
Lin et al. 2014	ID, DD; 18–64 years	Identification in any of 5 data sources (from inception): Psychiatric inpatient; Acute care inpatient; ED visits; Physician claims (≥ 2); Disability income support (12-month data period).	Case identification by source: Disability income support (63%; 34% unique); Health data (66%, 37% unique; Hospital discharges 13% unique, ED visits 0.5% unique, Physician claims 17% unique). Profiles of cohorts identified separately by health and disability income support data differed in terms of age, neighbourhood affluence, psychiatric comorbidity, receipt of income support, and use of some health services.	Comparison with previously published prevalence estimates.
Lloyd et al. 2022	ID, DD; 0–19 years	Identification in any of 5 health data sources (Psychiatric inpatient; Acute care inpatient; SDS; ED visits; Physician visits) or in Special Olympics data.	Case identification by source: Special Olympics (15%; 6.4% unique); Health data (94%; 85% unique). Cohort identified uniquely in Special Olympics data compared with cohort identified in health data: older mean age; greater proportion female; lower health services utilisation.	–
Madley-Dowd et al. 2022	ID; Children	Identification if ≥ 2 criteria met across following data sources: Longitudinal cohort study (IQ scores and free text); Education; Hospital episodes; General Practitioner (GP) records; Mental health services (≥ 2); Previous study to identify cohort members with developmental delay.	Of the administrative data sources, cases were most frequently identified in education and GP data, with fewer than 5 cases identified in mental health services data. Identified ID cases had IQ scores on average 40 points lower than those without ID at age 8 and 29 points lower at age 15.	Examined IQ scores at ages 8 and 15 for individuals identified and not identified for inclusion.
Nielsen et al. 2023	Autism; 2–15 years	Identification in any of 3 data sources: Disability services; Hospital admissions; Ambulatory mental health encounters. 15-year data period.	Case identification by source: disability services (87%; 71% unique); hospital admission data (23%; 8% unique); ambulatory mental health data (8%; 3% unique). Cases identified in health data alone more likely to be female, older at first contact, and to live in major cities and less disadvantaged areas.	Comparison with previously published prevalence estimates and trends.

Table 2 (continued)

Study	Disability cohort[1]; age	Method used to construct disability indicator	Findings relevant to method for constructing disability indicator	Validation approach
Nurminen et al. 2022	ID; all ages	Identification in any one of the 5 constituent data sources: Disability Allowance register, Disability pension register, Health Care register, Primary Health Care register, Social Welfare register	–	–
Ouellette-Kuntz et al. 2009	ID; all ages	Identification in any of following data sources: hospital abstracts; physician claims; education; income assistance. 5-year data period.	Case identification by source: hospital abstracts (39%; 29% unique); physician claims (30%; 19% unique); education (45%; 36% unique); income assistance (2%). Hospital abstracts database was the most useful source in identifying cases aged 0–4 years; education database was the most useful source in identifying cases aged 5–19 years.	Comparison with previously published prevalence estimates.
Reppermund et al. 2017	ID; all ages	Identification in any of 3 data sources: Disability services; Admitted hospital; Emergency department. 7-year data period.	Case identification by source: Disability services (82%; 35% unique); Admitted hospital (55%; 12% unique); Emergency department (47%; 0.1% unique).	–
Reppermund et al. 2019	ID; all ages	Identification in any of following data sources: Disability services; Admitted hospital; Emergency department (ED); Ambulatory mental health; Education; Corrective services; Ombudsman; Public guardian. Up to 22-year data period (depending on data source).	Case identification by source: Disability services (64%); Admitted hospital (67%); ED (60%); Ambulatory mental health (22%); Education (24%); Corrective services (3%); Ombudsman (1%); Public guardian (4%). Sociodemographic profile of people with ID in disability, health, mortality, and corrective services data reported.	Comparison with previously published prevalence estimates and trends.
Westerinen et al. 2007	ID; all ages	Identification in any of 8 data sources: 6 financial assistance sources; hospital discharge; social care. 5-year data period for hospital data, 1 year for other sources.	Case identification by source: Child Disability Allowance (10%; 2.8% unique); Disability Pension (64%; 20% unique); Disability Allowance (1.0%; 0.3% unique); Pensioners' Care Allowance (33%; 1.5% unique); Funding of Rehabilitation (10%; 1.1% unique); Preferential Refunding of Long-term Medication (25%; 6.4% unique); Discharge Register of Hospitals (19%; 7.8% unique); Discharge Register of Social Care (29%; 3.8% unique). 44% of cases appeared in only one register. Pattern of case identification by source differed by age group.	Comparison with official administrative prevalence rate.

1. Disability groups in the cohort (DD = developmental disability; DSC = Disability Services Commission; ED = Emergency Department; GP = General Practitioner; ID = Intellectual disability; SDS = Same-day surgery

most useful for identifying children with ID aged 0–4 years and the education data source was the most useful for identifying people with ID aged 5–19 years. Ten of the 19 studies described sociodemographic or disability characteristics of the individuals identified by data source [43, 47, 50, 51, 53–55, 57, 59, 60].

Validation approach

We use the term 'validation' to mean any testing or comparison of the cohort identified by the indicator against a reference cohort or previously published prevalence estimates, to provide insight into the extent to which the derived disability indicator captures the intended target population. Three studies used validation approaches that compared the identified cohort with a target population from an external linked data source [42, 45, 47]. One study tested the sensitivity of multiple algorithms to identify autism in a cohort of children by comparing identified cases with those reported by service providers [45]. The validation demonstrated that false negatives (autistic children identified by service providers but not identified by the derived disability indicator) were more likely to be born outside the province, be diagnosed at an older age, and not have ID or DD. Another study tested their dementia indicator against a reference population of individuals with cognitive impairment identified in a clinical setting and found that false negatives were associated with younger age, lower education, milder impairment, shorter disease duration and lower use of medication [42]. The third used data from a clinical autism database as a reference population to assess the accuracy of each of their algorithms [47]. They investigated neonatal and maternal characteristics and found similar sensitivity across most factors, which included sex, major congenital anomaly, county of residence, maternal age, birth weight and birth order.

Two of the nineteen studies used other validation approaches [48, 54]. One study conducted clinical case-note review for a small portion (3.4%) of the sample of autistic youth along with cross-validation of co-existing cases (0.5% of cases) in a national twin study [48]. In the study that used an eight-item indicator, two of the eight criteria for ID identification were an IQ score of less than 70 at age 8 years and an IQ score of less than 70 at age 15 years [54]. To assess the validity of their indicator, the authors compared mean IQ scores at ages 8 and 15 years for those with ID indicated by each source of information, and also compared mean IQ scores for the cohort identified as having ID by the indicator with those not identified by the indicator [54].

Thirteen studies compared estimated prevalence based on the identified cohort to previously published prevalence estimates and trends [42–44, 46, 48–52, 55, 57, 59, 60]. Nine of these studies discussed reasons that their prevalence estimates may have differed from previously published estimates [42, 43, 48, 49, 51, 52, 55, 57, 60]. For example, one study suggested that variation between their estimated prevalence of dementia and that reported in a previous study may have resulted from their use of a comparatively lower age cut-off, shorter observation period, and use of a greater number of administrative databases for dementia cohort identification [49].

In addition to comparison with previously published estimates, one study examined the cohort demographic profile of autistic individuals, making comparison with previously published patterns across gender, ethnicity, socioeconomic status, urbanicity, and rates of co-occurring conditions [43]. The estimated prevalence of autism in this study was lower than previously reported estimates. The authors suggested this may be due to a large portion of the cohort being identified in disability support data, as the cohort identified through this source may be those with more severe and complex needs. Their demographic comparisons, however, demonstrated consistency between their estimated relative rates of autism across gender, ethnicity and urbanicity with those previously published. They also found that their estimated rates of co-occurring conditions were similar to those previously reported, with the exception of anxiety, which was underestimated in this study, which the authors suggested may have been due to variation in clinician-based diagnostic definitions [43].

Three of the 19 studies did not report any type of validation [53, 56, 58].

Discussion

In this review, we set out to scope the literature describing methods for constructing disability indicators from linked administrative data sources, with a focus on approaches used to assess the validity of the indicators and to understand their strengths and limitations. Validation, to determine the extent to which a derived disability indicator captures its intended target population, is important to inform appropriate use of the indicator for specific purposes and caveats that may need to be applied when reporting findings.

Characteristics of published methods for constructing disability indicators

An indicator constructed using linked administrative data can only identify individuals who have had contact with the services or programs to which the data relate [52]. As such, the selection of data sources will influence how comprehensively the indicator identifies the target population and which subgroups are more or less comprehensively represented. Indeed, Lin et al. [52] discuss two important considerations when using linked data: "First, is an understanding of the various real-world

pathways by which individuals […] do (or do not) enter into, travel through, and exit services and supports. Second, is an understanding of how these pathways are reflected (or missed) in different administrative data".

In this scoping review, health data sources were used in over 90% of the studies, with 28% of studies using only health data for indicator construction. Studies that reported on demographic profiles of the cohort members identified in each data source demonstrated that different subgroups of the target population were captured from different sources. For example, Lin et al. [52] reported that people with intellectual or developmental disability identified by health data sources compared to disability income support data sources differed in terms of age, neighbourhood affluence, psychiatric comorbidity, receipt of income support, and use of health services. Furthermore, the utility of different data sources is likely to differ depending on the target population, including according to disability group and the target age range. Though most studies in our review relied on health data sources, in their analysis of autistic individuals, Nielsen et al. [55] reported that 71% of the cohort was uniquely identified in disability services data, with only 8% in hospital data and 3% in ambulatory mental health data. Ouellette-Kuntz et al. [57] found important differences by age group, with hospital records most useful for identifying children with intellectual disability aged 0 to 4 years and education data most useful for children aged 5 to 19 years.

These findings suggest that an indicator drawing upon data sources from different sectors is likely to more comprehensively capture the target population. Brown et al. [20] drew a similar conclusion from their systematic review of methods for identifying reproductive-aged women with physical and sensory disabilities in administrative health data, suggesting that sources that include functional assessments (e.g., for income support programs or other services) together with administrative health data may provide a more complete picture of disability status than approaches using only health data. However, it is likely that the contributions of different data sources will vary depending on the target population of interest and the patterns of contact individuals have with different services and programs in a given jurisdiction. Therefore, it will generally be valuable to examine the characteristics of cohort members identified in different data sources in order to optimise the indicator construction method and gain insights into strengths and limitations that should be considered when using the indicator.

How the definition of the disability target population is operationalised in a derived disability indicator depends on the data items used to construct the indicator, and is therefore constrained by data items available in the included sources [61]. As stated by Leonard et al. [50], this is "an unavoidable weakness of studies using record linkage". Of the included studies, 94% used diagnosis codes for disability identification, with 44% using only diagnosis codes. The implications of this heavy reliance on medical diagnoses to identify people with disability deserve consideration. Conceptual definitions of disability that are aligned with the ICF and the CRPD understand disability in terms of functioning limitations associated with the interaction between a person's health condition/s and contextual factors, including environmental barriers. Though medical diagnoses do not directly provide information about an individual's experience of activity limitations or participation restrictions [20, 62], functioning limitations may be inherent in, or usually associated with, some conditions. For example, diagnostic criteria for autism include reference to limitations in social interaction and communication [63, 64]. However, for many diagnoses, the presence of associated functioning limitations will vary between individuals, influenced by a range of health, personal and environmental factors. The limitations of using diagnosis codes for identifying individuals with disability as conceptualised by the ICF are discussed by Iezzoni [62], in terms of their "questionable accuracy, completeness, clinical scope, and meaningfulness". As such, in order to evaluate the alignment between a derived disability indicator and the target population of interest, studies need to clearly specify the conceptual framework they are adopting and discuss how closely their derived disability indicator aligns with their conceptual approach to disability in light of the disability data items used to construct the derived indicator.

Data items related to eligibility for disability-specific services or payments were used to construct disability indicators in several studies (e.g [29, 34, 39, 57]). Compared with diagnosis data, such data items may more reliably identify individuals with functioning limitations if eligibility criteria require such limitations or associated support needs to be demonstrated. However, changes in eligibility criteria over time, or variation in how criteria are applied in practice, may impact the cohort identified. This issue was discussed by Bowden et al. [43] who reported differing rates of co-occurring intellectual disability among members of the autism spectrum cohort identified through disability support services data before and after changes to eligibility criteria were made.

It is notable that none of the disability indicators in the included studies incorporated 'disability status' data items not linked to eligibility for disability supports, such as the ICF-aligned Standardised Disability Flag developed in Australia for use in administrative data collections as a basis for providing consistent and comparable information about people with disability who access

mainstream services [65]. The introduction of such data items into mainstream data that capture a large portion of the population (e.g., primary health care data) could greatly assist in the construction of comprehensive disability indicators in linked administrative data.

Administrative data are collected for a purpose. Diagnostic codes remain integral to health care systems and so are readily available within health data sources. The inclusion of ICF-aligned data items, such as items relating to activity limitations or participation restrictions, in administrative data sources could provide valuable information to meet the primary administrative needs of services and programs. For example, for disability support programs, data on functioning limitations and associated support needs structured using the activity and participation domains of the ICF could provide insights into the functioning profile of service users and the level of demand for different types of supports [66]. For mainstream services, in sectors such as health and housing, ICF-aligned disability data items could be used to flag service users who may require need for assistance or accommodations to ensure equitable access to services [67]. Valuable secondary use could then be made of such data items in linked administrative data sets to construct disability indicators for a variety of analytical purposes.

The time periods during which cohort members were identified in data sources ranged from one year to 23 years. In some studies (e.g [43]), there was discussion of how the time period covered by the data could affect the ability of the indicator to comprehensively identify individuals in the target population. However, in all studies, there was an implicit assumption that once a person is identified as having disability within administrative data at a given point in time, they can be regarded as having disability at all future points in time. This may be a reasonable assumption if the indicator aims to identify a cohort of people for whom disability is normally life long, for example ID or autism. However, this may not be the case for indicators that aim to capture broad disability cohorts. For many people, the experience of disability is dynamic. Longitudinal studies have demonstrated that individuals' disability status can change over time such that people move into and out of the disability population [68–70]. Therefore, methods for identifying a broad cohort of people with diverse disabilities should consider the data time period used in relation to evidence about how disability status may change over time for cohort members.

Thirty of the 36 included studies constructed disability indicators to identify people with intellectual and/or developmental disability (including autism). Few studies aimed to identify a broad population of people with disability, with indicators encompassing diverse disability types used in only three of the studies. Thus, our findings point to a paucity of evidence on methods to construct broad disability indicators that aim to identify people with diverse disabilities. Constructing an indicator to capture a broad disability population is likely to pose greater challenges than indicators for specific types of disability, such as ID and autism, because of the greater diversity in terms of patterns of contact with services and limited data items available that can reliably be used to identify members of the target population.

Validation of disability indicators

Of those studies with a primary aim of creating a disability cohort or estimating disability prevalence (Table 2) [42–60], many evaluated the derived disability indicators by comparing prevalence rates estimated using their indicator with previously published estimates. In some cases, the authors discussed possible reasons for differences between their prevalence estimates and those previously published, and sometimes more detailed comparisons were made, for instance comparing age and gender profiles. However, only three studies conducted validation testing that compared the identified cohort with a target population to explore the extent to which the methods used were successful in identifying the target population, which is a critical step in understanding the limitations of disability indicators derived from administrative data. In discussing the value of using administrative data to study people with disability, Iezzoni [62] cautions that findings must be interpreted with an understanding of the limitations of the methods and data sources used to identify study populations.

Validation techniques may include exploration of the indicator's ability to correctly identify individuals with disability (sensitivity) and individuals without disability (specificity) in comparison to a reference cohort, examination of false positives and false negatives, and comparison of the sociodemographic and disability characteristics of the identified cohort with known characteristics of the target population. Some studies, for example Lin et al. [51], acknowledged the lack of external validation of their indicator as a key limitation, however most studies did not address this limitation.

Often, the ability to conduct validation testing will be limited by the lack of suitable reference data about the target population. Reference data might include sources such as population survey, census or registry data considered to comprehensively identify the target population. However, clear definition of the target population is an essential prerequisite for validation of an indicator. Some of the included studies, for example Calver et al. [27], referred to definitions of disability in legislation, or in policy and program documentation. However, many studies did not clearly define their target population, or discuss the extent to which the operational definition

manifested in their disability indicator aligned with this. Correspondingly, none of the studies considered in any depth the potential implications of the limitations of the disability indicators in terms of their use in analyses to inform policy, program development or resource allocation decisions.

Limitations

We may not have identified all relevant studies. In particular, as our database search spanned the period 2010 to 2023, studies published prior to 2010 that were not picked up through searching the reference lists of included articles may have been missed.

Half of the included studies focused on reporting findings on the outcomes explored, with little information on the methods used to construct disability indicators. This limited our ability to investigate and discuss findings related to methods used to construct and validate derived disability indicators across all of the included studies.

Conclusions

A key finding from this scoping review is the paucity of evidence on methods to construct broad disability indicators in linked administrative data that aim to identify people with diverse disabilities, despite the urgent need to generate robust evidence to tackle disability-related health inequities. A substantial number of studies reported methods for constructing indicators to identify specific types of disability, particularly intellectual and/or developmental disability, though often these studies did not report detailed findings concerning the contributions of different data sources to the cohort identified or test the validity of the derived indicators. There is a high level of reliance on diagnostic information for constructing disability indicators, which likely reflects a lack of other types of disability-relevant information, particularly information about activity limitations and participation restrictions, in the source data. Including disability identifiers aligned with the ICF conceptualisation of disability in a wider range of administrative data sources, such as education and primary care data, would enable more comprehensive identification of the population with disability.

Few studies conducted validation of their disability cohort construction. This is a crucial step which should be implemented when developing algorithms to identify people with disability. At a minimum, comparison should be made against existing prevalence estimates. However, we recommend that, where possible, an additional step should be taken to link records in the constructed cohort to a data source containing high-quality disability information, such as from a population survey. This enables quantitative comparative analysis to gain understanding of how well the target population is represented, and importantly to identify which subpopulations may be excluded or underrepresented in the disability cohort.

Linked data approaches hold great potential to examine health and other outcomes for people with disability, to investigate contact with services, and to study the prevalence and characteristics of particular cohorts of people with disability, overcoming limitations of population surveys and (unlinked) administrative data collections. Using linked administrative data can enable more sophisticated causal analyses to generate policy-relevant insights concerning the drivers of inequities in health and the social determinants of health for people with disability. Importantly, linked data approaches could assist countries in meeting their obligations under the CRPD to monitor and report on progress towards achieving the rights of people with disability, provided that reliable disability indicators can be derived. To fully realise this potential, it is crucial to conduct validation testing, to understand the extent to which derived disability indicators capture the target population, determine limitations of the indicators, and identify caveats that should apply to their use for specific analytic purposes.

Abbreviations

CRPD	The United Nations Convention on the Rights of Persons with Disabilities
DD	Developmental disability
DSM	Diagnostic and Statistical Manual of Mental Disorders
ED	Emergency department
GP	General practitioner
ID	Intellectual disability
ICD	International Classification of Diseases
ICF	International Classification of Functioning, Disability and Health
SDS	Same-day surgery

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12963-025-00386-w.

Supplementary Material 1

Acknowledgements

Not applicable.

Author contributions

NF and ZA conducted the literature search, screened the articles for eligibility, and performed data extraction for this systematic scoping review. NF, ZA, and SW drafted the manuscript. GB and SB provided critical reviews and feedback on the manuscript, and all authors have read and approved the final version.

Funding

Australian Research Council (IE230100561). National Health and Medical Research Council (2035278, 2010290). Medical Research Future Fund (2024940).

Data availability

No datasets were generated or analysed during the current study.

Declarations

Ethics approval

Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 21 August 2024 / Accepted: 27 May 2025
Published online: 06 June 2025

References

1. United Nations. Convention on the Rights of Persons with Disabilities. New York, USA; 2006.
2. Stein MAL, Janet E. Introductory Note to the Convention on the Rights of Persons with Disabilities; 2023. https://legal.un.org/avl/ha/crpd/crpd.html. Accessed on 19 May 2024.
3. World Health Organization. Global report on health equity for persons with disabilities. Geneva: World Health Organization; 2022.
4. World Health Organization. World bank. World report on disability. Geneva, Switzerland: WHO; 2011.
5. Krahn GL, Walker DK, Correa-De-Araujo R. Persons with disabilities as an unrecognized health disparity population. Am J Public Health. 2015;105:S198–206. https://doi.org/10.2105/AJPH.2014.302182.
6. United Nations. Disability and development report: realizing the sustainable development goals by, for and with persons with disabilities, 2018. New York, USA: UN; 2019.
7. Williamson EJ, McDonald HI, Bhaskaran K, Walker AJ, Bacon S, Davy S, et al. Risks of covid-19 hospital admission and death for people with learning disability: population based cohort study using the opensafely platform. BMJ. 2021;374. https://doi.org/10.1136/bmj.n1592.
8. Krahn G. Where is disability in global public health? Oxford research encyclopedia of global public health. Oxford, United Kingdom: Oxford University Press; 2021. https://doi.org/10.1093/acrefore/9780190632366.013.287.
9. Kavanagh A. Disability and public health research in Australia. Aust N Z J Public Health. 2020. https://doi.org/10.1111/1753-6405.13003.
10. Australian Institute of Health and Welfare. People with disability in Australia 2020. Canberra, Australia: AIHW; 2020.
11. Cantrell A, Croot E, Johnson M, Wong R, Chambers D, Baxter SK, Booth A. Access to primary and community health-care services for people 16 years and over with intellectual disabilities: a mapping and targeted systematic review. 2020; https://doi.org/10.3310/hsdr08050
12. World Health Organization. International classification of functioning, disability and health. Geneva, Switzerland: WHO; 2001.
13. Madden RH, Bundy A. The ICF has made a difference to functioning and disability measurement and statistics. Disabil Rehabil. 2018;41(12):1450–62. https://doi.org/10.1080/09638288.2018.1431812.
14. World Health Organization. Model disability survey (MDS): survey manual. Licence: CC BY-NC-SA 3.0 IGO. Geneva, Switzerland: WHO; 2017.
15. Sabariego C, Fellinghauer C, Lee L, Kamenov K, Posarac A, Bickenbach J, et al. Archives Public Health. 2022;80(1):1–10. https://doi.org/10.1186/s13690-021-00769-z. Generating comprehensive functioning and disability data worldwide: development process, data analyses strategy and reliability of the WHO and World Bank Model Disability Survey.
16. Fortune N, Madden RH, Clifton S. Health and access to health services for people with disability in australia: data and data gaps. Int J Environ Res Public Health. 2021;18(21):11705. https://doi.org/10.3390/ijerph182111705.
17. Smith M, Flack F. Data linkage in australia: the first 50 years. Int J Environ Res Public Health. 2021;18(21):11339. https://doi.org/10.3390/ijerph182111339.
18. Jensen LR. Using data integration to improve health and welfare insights. Int J Environ Res Public Health. 2022;19(2):836. https://doi.org/10.3390/ijerph19020836.
19. Friedman DJ, Parrish RG, Fox MH. A review of global literature on using administrative data to estimate prevalence of intellectual and developmental disabilities. J Policy Pract Intellect Disabil. 2018;15(1):43–62. https://doi.org/10.1111/jppi.12220.
20. Brown HK, Carty A, Havercamp SM, Parish S, Lunsky Y. Identifying reproductive-aged women with physical and sensory disabilities in administrative health data: a systematic review. Disabil Health J. 2020;13(3):100909. https://doi.org/10.1016/j.dhjo.2020.100909.
21. Munn Z, Peters MD, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. BMC Med Res Methodol. 2018;18:1–7. https://doi.org/10.1186/s12874-018-0611-x.
22. Peters M, Godfrey C, McInerney P, Munn Z, Tricco A, Khalil H. In: Aromataris E, Munn Z, editors. Chapter 11: scoping reviews (2020 version). JBI Manual for Evidence Synthesis: JBI; 2020.
23. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol. 2005;8(1):19–32.
24. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci. 2010;5(1):69. https://doi.org/10.1186/1748-5908-5-69.
25. Brameld K, Spilsbury K, Rosenwax L, Leonard H, Semmens J. Use of health services in the last year of life and cause of death in people with intellectual disability: a retrospective matched cohort study. BMJ Open. 2018;8(2):e020268. https://doi.org/10.1136/bmjopen-2017-020268.
26. Brophy S, Kennedy J, Fernandez-Gutierrez F, John A, Potter R, Linehan C, Kerr M. Characteristics of children prescribed antipsychotics: analysis of routinely collected data. J Child Adolesc Psychopharmacol. 2018;28(3):180–91. https://doi.org/10.1089/cap.2017.0003.
27. Calver J, Balogh R, Rudoler D. Incidence of injury in children and adolescents with intellectual and developmental disability. J Saf Res. 2021;77:56–60. https://doi.org/10.1016/j.jsr.2021.02.003.
28. Deroche CB, McDermott SW, Mann JR, Hardin JW. Colorectal Cancer screening adherence in selected disabilities over 10 years. Am J Prev Med. 2017;52(6):735–41. https://doi.org/10.1016/j.amepre.2017.01.005.
29. Griffith MI, Mann JR, McDermott S. The risk of intellectual disability in children born to mothers with preeclampsia or eclampsia with partial mediation by low birth weight. Hypertens Pregnancy. 2011;30(1):108–15. https://doi.org/10.3109/10641955.2010.507837.
30. Hirvikoski T, Boman M, Tideman M, Lichtenstein P, Butwicka A. Association of intellectual disability with all-cause and cause-specific mortality in Sweden. JAMA Netw Open. 2021;4(6):e2113014–e. https://doi.org/10.1001/jamanetworkopen.2021.13014.
31. Hwang YIJ, Srasuebkul P, Foley K-R, Arnold S, Trollor JN. Mortality and cause of death of Australians on the autism spectrum. Autism Res: Official J Int Soc Autism Res. 2019;12(5):806–15. https://doi.org/10.1002/aur.2086.
32. Lunsky Y, Lin E, Balogh R, Klein-Geltink J, Wilton AS, Kurdyak P. Emergency department visits and use of outpatient physician services by adults with developmental disability and psychiatric disorder. Can J Psychiatry. 2012;57(10):601–7. https://doi.org/10.1177/070674371205701004.
33. Maclean MJ, Sims S, Bower C, Leonard H, Stanley FJ, O'Donnell M. Maltreatment risk among children with disabilities. Pediatrics. 2017;139(4). https://doi.org/10.1542/peds.2016-1817.
34. Mann JR, McDermott SW, Hardin J, Pan C, Zhang Z. Pre-pregnancy body mass index, weight change during pregnancy, and risk of intellectual disability in children. BJOG. 2013;120(3):309–19. https://doi.org/10.1111/1471-0528.12052.
35. Marquis SM, McGrail K, Hayes M. Mental health of parents of children with a developmental disability in British columbia, Canada. J Epidemiol Commun Health. 2020;74(2):173–8. https://doi.org/10.1136/jech-2018-211698.
36. Shooshtari S, Martens PJ, Burchill CA, Dik N, Naghipur S. Prevalence of depression and dementia among adults with developmental disabilities in manitoba, Canada. Int J Family Med. 2011;2011(319574). https://doi.org/10.1155/2011/319574.
37. Su X, Yu Y, Meng L, Duan T, Zhao Y, László KD, et al. Prenatal maternal bereavement and its association with intellectual disability in the offspring. Psychosom Med. 2021;83(8):887. https://doi.org/10.1097/PSY.0000000000000990.
38. Tarasoff LA, Lunsky Y, Chen S, Guttmann A, Havercamp SM, Parish SL, et al. Preconception health characteristics of women with disabilities in ontario: A Population-Based, Cross-Sectional study. J Womens Health (Larchmt). 2020;29(12):1564–75. https://doi.org/10.1089/jwh.2019.8273.
39. Wang Y, McDermott S, Mann JR, Hardin JW. Preventing intellectual disability during pregnancy: what are the potentially high yield targets? J Perinat Med. 2015;44(4):421–32. https://doi.org/10.1515/jpm-2015-0059.
40. Wang H, Lee PMY, Zhang J, Svendsen K, Li F, Li J. Association of intellectual disability with overall and type-specific cardiovascular diseases: a population-based cohort study in Denmark. BMC Med. 2023;21(1):41. https://doi.org/10.1186/s12916-023-02747-4.
41. Weiss JA, Isaacs B, Diepstra H, Wilton AS, Brown HK, McGarry C, Lunsky Y. Health concerns and health service utilization in a population cohort of young adults with autism spectrum disorder. J Autism Dev Disord. 2018;48(1):36–44. https://doi.org/10.1007/s10803-017-3292-0.

42. Bacigalupo I, Lombardo FL, Bargagli AM, Cascini S, Agabiti N, Davoli M, et al. Identification of dementia and MCI cases in health information systems: an Italian validation study. Alzheimer's Dementia: Translational Res Clin Interventions. 2022;8(1):e12327. https://doi.org/10.1002/trc2.12327.
43. Bowden N, Kokaua J, Audas R, Taylor B, Thabrew H, Milne B, et al. Autism spectrum disorder/takiwatanga: an integrated data Infrastructure-based approach to autism spectrum disorder research in new Zealand. Autism. 2020;24(8):2213–27. https://doi.org/10.1177/1362361320939329.
44. Cama AT, Sikivou BT, Keeffe JE. Childhood visual impairment in fiji. Archives of ophthalmology (Chicago, ill: 1960). 2010;128(5):608–12; https://doi.org/10.1001/archophthalmol.2010.54
45. Coo H, Ouellette-Kuntz H, Brownell M, Shooshtari S, Hanlon-Dearman A. Validating an administrative data-based case definition for identifying children and youth with autism spectrum disorder for surveillance purposes. Can J Public Health. 2017;108:e530–8. https://doi.org/10.17269/CJPH.108.5963.
46. Cuypers M, Tobi H, Naaldenberg J, Leusink GL. Linking National public services data to estimate the prevalence of intellectual disabilities in the netherlands: results from an explorative population-based study. Public Health. 2021;195:83–8. https://doi.org/10.1016/j.puhe.2021.04.002.
47. Dodds L, Spencer A, Shea S, Fell D, Armson B, Allen A, Bryson S. Validity of autism diagnoses using administrative health data. Chronic Dis Can. 2009;29(3):102.
48. Idring S, Rai D, Dal H, Dalman C, Sturm H, Zander E, et al. Autism spectrum disorders in the Stockholm youth cohort: design, prevalence and validity. PLoS ONE. 2012;7(7):e41280. https://doi.org/10.1371/journal.pone.0041280.
49. Kosteniuk JG, Morgan DG, O'Connell ME, Kirk A, Crossley M, Teare GF, et al. Incidence and prevalence of dementia in linked administrative health data in saskatchewan, canada: a retrospective cohort study. BMC Geriatr. 2015;15(1):1–16. https://doi.org/10.1186/s12877-015-0075-3.
50. Leonard H, Petterson B, Bower C, Sanders R. Prevalence of intellectual disability in Western Australia. Paediatr Perinat Epidemiol. 2003;17(1):58–67. https://doi.org/10.1046/j.1365-3016.2003.00469.x.
51. Lin E, Balogh R, Cobigo V, Ouellette-Kuntz H, Wilton AS, Lunsky Y. Using administrative health data to identify individuals with intellectual and developmental disabilities: a comparison of algorithms. J Intellect Disabil Res. 2013;57(5):462–77. https://doi.org/10.1111/jir.12002.
52. Lin E, Balogh R, Isaacs B, Ouellette-Kuntz H, Selick A, Wilton AS, et al. Strengths and limitations of health and disability support administrative databases for population-based health research in intellectual and developmental disabilities. J Policy Pract Intellect Disabil. 2014;11(4):235–44.
53. Lloyd M, Dobranowski K, Cheng SY, Temple VA, Foley JT, Lunsky Y, et al. Matching special olympics registration data with administrative health databases: feasibility and health status differences in children and youth with IDD. Disabil Health J. 2022;15(3):101319. https://doi.org/10.1016/j.dhjo.2022.101319.
54. Madley-Dowd P, Thomas R, Boyd A, Zammit S, Heron J, Rai D. Intellectual disability in the children of the Avon longitudinal study of parents and children (ALSPAC). Wellcome Open Res. 2022;7:172. https://doi.org/10.12688/wellcomeopenres.17803.2.
55. Nielsen TC, Nassar N, Boulton KA, Guastella AJ, Lain SJ. Estimating the prevalence of autism spectrum disorder in new South wales, australia: A data linkage study of three routinely collected datasets. J Autism Dev Disord. 2023. https://doi.org/10.1007/s10803-022-05887-3.
56. Nurminen F, Ratto H, Arvio M, Teittinen A, Vesala HT, Saastamoinen L. Medicine use in people with intellectual disabilities: a Finnish nationwide register study. J Intellect Disabil Res. 2022. https://doi.org/10.1111/jir.12988.
57. Ouellette-Kuntz H, Shooshtari S, Temple B, Brownell M, Burchill C, Yu C, et al. Estimating administrative prevalence of intellectual disabilities in Manitoba. J Dev Disabil. 2009;15(3):69.
58. Reppermund S, Srasuebkul P, Heintze T, Reeve R, Dean K, Emerson E, et al. Cohort profile: a data linkage cohort to examine health service profiles of people with intellectual disability in new South wales, Australia. BMJ Open. 2017;7(4):e015627. https://doi.org/10.1136/bmjopen-2016-015627.
59. Reppermund S, Heintze T, Srasuebkul P, Reeve R, Dean K, Smith M, et al. Health and wellbeing of people with intellectual disability in new South wales, australia: a data linkage cohort. BMJ Open. 2019;9(9):e031624.
60. Westerinen H, Kaski M, Virta L, Almqvist F, Iivanainen M. Prevalence of intellectual disability: a comprehensive study based on National registers. J Intellect Disabil Res. 2007;51(9):715–25. https://doi.org/10.1111/j.1365-2788.2007.00970.x.
61. Royal Commission into, Violence A. Neglect and exploitation of people with disability. Final report, volume 12, beyond the Royal commission. Canberra, Australia: Government of Australia; 2023.
62. Iezzoni LI. Using administrative data to study persons with disabilities. Milbank Q. 2002;80(2):347–79. https://doi.org/10.1111/1468-0009.t01-1-00007.
63. World Health Organization. International statistical classification of diseases and related health problems (11th ed.) 2019. https://icd.who.int/en. Accessed on 15 July 2024.
64. American Psychiatric Association. Diagnostic and statistical manual of mental disorders. (5th ed., text rev.) 2022.
65. Australian Institute of Health and Welfare. Standardised Disability Flag: data collection guide. Cat. no. DAT 6. Canberra; 2016.
66. Australian Institute of Health Welfare. Current and future demand for specialist disability services. Canberra: AIHW. 2007. https://www.aihw.gov.au/reports/disability/demand-for-specialist-disability-services. Accessed on 14 May 2025.
67. Fortune N, Bailie J, Gordon J, Plunkett K, Hargrave J, Madden R, Llewellyn G. Developing self-report disability questions for a voluntary patient registration form for general practice in Australia. Aust N Z J Public Health. 2023;47(2):100032. https://doi.org/10.1016/j.anzjph.2023.100032.
68. Morris SF, Timoney G, Hughes LR. J. Canadian Survey on Disability. The Dynamics of Disability: Progressive, Recurrent or Fluctuating Limitations. Statistics Canada. 2019; Catalogue no. 89-654-X2019002.
69. Shields M, Spittal MJ, Dimov S, Kavanagh A, King TL. Trajectories of disability throughout early life and labor force status as a young adult: results from the longitudinal study of Australian children. Scand J Work Environ Health. 2022;48(2):118–26. https://doi.org/10.5271/sjweh.3994.
70. Burchardt T. The dynamics of being disabled. CASEpaper 36. SRC research centre for analysis of social exclusion. London School of Economics; 2000.

Publisher's note