EVIDINNO
EVIDENCE NAVIGATION & SYNTHESIS

MORE COMPLEX THAN IT SEEMS:

DETERMINING **DISEASE INCIDENCE AND PREVALENCE**

Walid Shouman, MSc - Associate Research Scientist
Jean-Paul Collet, MD, PhD - Chief Scientific Officer

Importance of Incidence/Prevalence Systematic Literature Reviews

Systematic literature reviews (SLRs) have become a cornerstone of evidence synthesis in medical research. SLRs aimed at determining incidence and prevalence of a condition have developed considerably in the last 15 years. They provide essential metrics for quantifying and better understanding the disease burden, which represents important elements to establish the possible benefits of a new drug.[1] For drug manufacturers, incidence and prevalence provide critical information regarding the market potential and geographic distribution; it is useful information for planning strategic development and allocating financial investments. Access to incidence and prevalence figures is of paramount importance for developing specific studies aimed at, for instance identifying the unmet needs of patients within healthcare systems, improving disease management, and estimating the market size for planning drug distribution and supporting participation in clinical trials. In case of rare diseases, determining incidence/prevalence is critical information to obtain an Orphan Drug Designation (ODD), which gives multiple advantages such as market exclusivity, tax credits, and reduced regulatory fees.[2, 3, 4]

Disease Frequency Measures

The study of disease occurrence and mortality forms the cornerstone of public health research. Over recent decades, significant advancements have been made in methods for measuring both incidence and prevalence. These metrics are fundamentally based on the relationship between the number of disease cases or deaths (numerator) and the total population at risk (denominator).

Incidence

Incidence refers to the number of new cases of a specific condition occurring within a defined population over a specific time period. It captures the flow of new cases. By focusing on new cases, incidence provides insights into the risk of developing a condition pending valid identification of the population at risk. The measure of incidence is specifically important for epidemiological studies that aim to identify causal factors or to evaluate the effects of preventative measures.[5]

Incidence can be expressed using two related measures:

- Cumulative Incidence (CI): Represents the probability of developing a disease over a specified period.

$$CI = \frac{\text{number of individuals who are diagnosed with the disease during a certain period}}{\text{number of individuals in the population at the beginning of the period}}$$

Example: A population of 1,000 people is being observed over a 5-year period to track the development of diabetes. By the end of the study, 50 individuals have developed the condition. The cumulative incidence is 50/1000 = 5% over 5 years. It means that the risk of developing diabetes in this population over 5 years is 5%.

- Incidence Rate (IR): Reflects the number of new cases per unit of person-time, accounting for the time individuals remains in the study.

$$IR = \frac{\text{number of individuals who are diagnosed with the disease during a certain period}}{\text{sum of person time at risk}}$$

Example: 500 people are monitored for varying lengths of time in a study, resulting in a total of 2,000 person-years of observation. During this time, 25 individuals are diagnosed with lung cancer. The incidence rate is 25/2,000 person-years = 12.5 cases per 1,000 person-years. This measure accounts for differences in the duration of time each person is observed.

Prevalence

Prevalence represents the total number of cases of a particular condition within a defined population at a given point in time (point-prevalence) or over a certain period (period-prevalence). Prevalence captures the overall burden of the disease within a population making it a valuable measure for assessing the extent of healthcare burden.

There are also two types of prevalence [5]:

- Point prevalence which represents the proportion of individuals in a population who have the condition at a specific point in time

$$\text{Point prevalence} = \frac{\text{number of individuals diagnosed with the disease at a specific time}}{\text{number of individuals in the population at that point of time}}$$

- Period prevalence which is the proportion of individuals who have the condition at any time during a specified period.[6,7]

Relationship between incidence and prevalence

Prevalence is affected by the incidence rate and the disease duration. If people live longer with a certain disease, they will remain prevalent for a long period of time.[8] An important condition for this equation to hold is a steady-state condition—meaning that both incidence and disease duration remain relatively constant over time.

$$\text{Prevalence} \approx \text{incidence rate} \times \text{average disease duration}$$

This equation serves as an approximation, but its use in real-world applications depends on a specific set of assumptions:

1- Stable Population:
The population size remains constant, meaning no significant changes due to migration, births, or deaths.

2- Steady-State Condition:
The disease incidence (new cases) and duration (how long people remain affected before recovery or death) are relatively constant over time.

3- Chronic Diseases:
This approximation works best for chronic conditions where individuals live with the disease for a prolonged period (e.g., diabetes, hypertension).

4- Low Incidence, Long Duration:
The equation is most accurate when the disease has a low incidence but a long duration, meaning new cases are relatively rare but affected individuals remain in the population for a long time.

5- No Rapid Cure or High Mortality:
If a disease has a very high fatality rate or short duration (e.g., Ebola), this simple equation does not hold well because the prevalence fluctuates too quickly.

Incidence/Prevalence SLRs vs. General SLRs

While general SLRs and incidence/prevalence SLRs share core principles, their objectives, methodologies, and focus differ in important ways. Below is a detailed exploration of these similarities and differences.

Similarities

1- Structured Approach: Both general SLRs and incidence/prevalence SLRs should follow a structured process. This includes defining a clear research question, developing a protocol, conducting a systematic search, screening studies for inclusion, extracting relevant data, and assessing study quality.

2-Transparency and Reproducibility: Both types of SLR aim to ensure transparency and reproducibility by documenting the review process, including search strategies, inclusion/exclusion criteria, and data extraction methods. This facilitates validation and replication by other researchers.

3- Critical Appraisal: In both types of SLRs, a critical appraisal for the included studies is done to assess their methodological quality and risk of bias. This ensures that findings are based on robust evidence.

- *Incidence/prevalence SLRs are essential for quantifying disease burden, supporting drug development, identifying unmet patient needs, and enabling strategic planning, especially for obtaining benefits like Orphan Drug Designation in rare diseases.*
- *Variability in numerators and denominators, study heterogeneity, and missing data present significant challenges in conducting this type of SLRs. To address these issues, researchers must use standardized methodologies, transparent reporting, and expert guidance from epidemiologists to ensure robust and generalizable findings.*

Differences

1- Objective: Incidence/prevalence SLRs are specifically focused on quantifying the frequency of a condition, events, or outcomes in a defined population, whereas general SLRs often have a broader scope and focus on different interventions, mechanisms, and outcomes.

2- Search strategy: In incidence/prevalence SLRs, the search strategy is designed to identify epidemiological studies, typically focusing on observational designs such as cohort studies, cross-sectional studies, and population-based surveys. Search terms often include keywords like “incidence,” “prevalence,” and “epidemiology.” Meanwhile, general SLRs can include diverse types of studies depending on the research question.[9,10]

3- Inclusion criteria: While the traditional framework for an SLR is based on PICO criteria (Population, Intervention, Comparator, and Outcomes), these criteria do not align with the conduct of SLRs that aim to capture measures of disease frequency. Therefore, inclusion criteria are narrowly defined using CoCoPop criteria (Condition, Context, and Population). Included studies must report explicit measures of incidence or prevalence with a defined numerator (number of cases) and denominator (population at risk). [9,10]

4- Data extraction: While the data extraction for outcomes in general SLRs can be quantitative or qualitative, the outcomes extracted in incidence/prevalence SLRs are highly quantitative focusing mainly on numerical values of incidence and prevalence. [10]

5- Data synthesis and meta-analysis: The synthesis often involves statistical pooling using meta-analysis to calculate weighted averages, heterogeneity assessments, and subgroup analyses to explore variations across populations or settings.[9] On the other hand, general SLRs can involve statistical pooling or narrative description at times depending on the available evidence base.[10]

Challenges of conducting an Incidence/Prevalence SLR

The challenges of conducting incidence/prevalence SLRs can be categorized into general challenges common to all SLRs and specific challenges unique to this type of research. This section outlines these challenges in detail.

General Challenges

1- Heterogeneity of studies: Included studies can vary widely in terms of study designs, outcome measurement, and included populations. The definitions of exposures, outcomes, and confounders can vary between studies. For instance, in a SLR aimed at measuring the prevalence of diabetes, some studies might define cases based on fasting blood glucose while others would use HbA1c levels. The quality of the data sources can also differ from one study to another, with direct effects on incidence and prevalence figures. For example, when studying the prevalence of a rare genetic disease such as genetically associated dilated cardiomyopathy (DCM), we found some studies used large databases with genetic information from general populations, while other studies measured number of cases by performing genetic screening on families after detection of an index case with DCM.[11] This heterogeneity can limit the combining of prevalent figures and make comparisons across studies/regions difficult.[12]

2- Risk of biases: The risk of biases in the included studies is challenging in all types of SLRs. Incidence/prevalence SLRs are particularly prone to various types of biases like selection or information biases, and confounding. It is very important to assess the risk of biases in each reviewed study to decide on the inclusion or exclusion of that study.

Unique Challenges

1- Numerator Definition: Variability in case definitions across studies can lead to inconsistencies in the numerator (number of cases). Different diagnostic criteria, self-reported cases, or reliance on ICD codes can impact the accuracy of incidence and prevalence estimates. Different studies identified in the SLR may use different populations which make combining the evidence difficult. For example, when studying the incidence and prevalence of idiopathic pulmonary fibrosis (IPF), we found that some studies group all fibrotic interstitial diseases under a broad "Interstitial Lung Diseases" (ILD) category while others focus strictly on IPF, excluding other forms of ILD, making cross-study comparisons difficult.[13]

2- Denominator Issues: Identifying the appropriate population at risk (denominator) can be challenging. Variability in population characteristics, incomplete population data, or differences in geographic and temporal coverage can lead to biased estimates. For instance, some studies refer to the national population of the country, while others report as denominator the population of the database they used.

3- Timeframe Variability: Differences in the time periods over which studies report data can complicate comparisons and meta-analyses. For example, when studying the birth prevalence of rare genetic disorders, there may be significant variability depending on the year advanced genetic screening technologies were introduced to the market. Before their introduction, the results may suggest a lower birth prevalence.

4- Missing data: Many studies may have incomplete data on case numbers or population sizes, requiring imputation methods that introduce uncertainty into the estimates. When data is missing, researchers often rely on assumptions to impute missing values. For example, they may assume uniform distribution of cases over time, consistent population sizes, or proportionality between subpopulations. This uncertainty can make it difficult to draw firm conclusions, especially if imputed data accounts for a significant proportion of the dataset.

Recommendations for Developing High Quality Incidence/Prevalence SLRs

1- General recommendation for the SLR: The SLR should start with a well-defined protocol that outlines the objectives with precise CoCoPop that informs the inclusion and exclusion criteria, a search strategy, and methods for data extraction and analysis. The search strategy should be comprehensive and the data extraction process should be standardized to ensure accuracy and consistency.[14]

2- Numerator issues: A precise assessment of how new cases are detected should be conducted. This includes employing standardized definitions for identifying new cases to reduce variability and enhance consistency across studies.

3- Denominator issues: A detailed evaluation of the source populations in the included studies is essential. The source population is the group from which new cases arise and within which all cases exist; it represents the denominator for calculating risk and proportions. Clear definitions of the source population should be provided, and potential biases, such as the exclusion of specific subgroups, need to be carefully assessed and reported. Standardization techniques, such as age adjustment, should be applied to ensure comparability across studies.

4- Timeframe variability: Several approaches can be used to address issues related to inconsistent timeframes. For instance, setting a publication date limit for included studies can minimize discrepancies. Additionally, sensitivity analyses can be performed to evaluate the impact of including or excluding studies with mismatched timeframes on the overall estimates.

5- Missing Data: Transparent reporting of how missing data are handled is critical. This includes detailing imputation methods and the assumptions underlying these approaches. Efforts should be made to avoid over-reliance on unvalidated assumptions by seeking supplementary data or excluding studies with substantial missing information.

Conclusion

Understanding the concepts of incidence and prevalence is essential for addressing diverse objectives, including public health planning, resource allocation, and pharmaceutical development. As the demand for high-quality incidence/prevalence SLRs grows, it is imperative to overcome inherent challenges. Incidence/prevalence SLRs face unique challenges, including variability in numerators and denominators, study heterogeneity, and issues related to missing data. Addressing these challenges requires rigorous planning, the use of standardized methodologies, and transparent reporting to ensure the reliability and generalizability of findings. Additionally, engaging with specialists in epidemiology and biostatistics can provide critical insights and guidance for resolving complex issues and refining methodologies. By systematically addressing these challenges, researchers can generate robust estimates that contribute to evidence-based decision-making in epidemiology and public health.

References

1. Dickersin K. Systematic reviews in epidemiology: why are we so far behind? International Journal of Epidemiology. 2002;31(1):6-12.
2. Fermaglich LJ, Miller KL. A comprehensive study of the rare diseases and conditions targeted by orphan drug designations and approvals over the forty years of the Orphan Drug Act. Orphanet journal of rare diseases. 2023;18(1):163.
3. EMA. Points to consider on the estimation and reporting on the prevalence of a condition for the purpose of orphan designation. 2019; https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/points-consider-estimation-and-reporting-prevalence-condition-orphan-designation_en.pdf, Accessed January 30, 2025.
4. FDA. Designating an Orphan Product: Drugs and Biological Products. 2024; https://www.fda.gov/industry/medical-products-rare-diseases-and-conditions/designating-orphan-product-drugs-and-biological-products, Accessed January 30, 2025.
5. Ahlbom A, Norell S. Introduction to modern epidemiology. Epidemiology Resources; 1990.
6. Henderson KL, Saei A, Freeman R, et al. Intermittent point prevalence surveys on healthcare-associated infections, 2011 and 2016, in England: what are the surveillance and intervention priorities? Journal of Hospital Infection. 2023;140:24-33.
7. Lau DCW, Shaw E, Farris MS, et al. Prevalence of Adult Type 2 Diabetes Mellitus and Related Complications in Alberta, Canada: A Retrospective, Observational Study Using Administrative Data. Canadian Journal of Diabetes. 2024;48(3):155-162.e158.
8. Oregon State University. Foundations of Epidemiology. Measures of Disease Frequency. https://open.oregonstate.education/epidemiology/chapter/measures-of-disease-frequency/, Accessed January 30, 2025.
9. Munn Z, Moola S, Lisy K, Riitano D, Tufanaru C. Methodological guidance for systematic reviews of observational epidemiological studies reporting prevalence and cumulative incidence data. International journal of evidence-based healthcare. 2015;13(3):147-153.
10. Munn Z, Moola S, Lisy K, Riitano D, Tufanaru C. Methodological guidance for systematic reviews of observational epidemiological studies reporting prevalence and cumulative incidence data. JBI Evidence Implementation. 2015;13(3):147-153.
11. Myers MC, Berge A, Zhong Y, et al. Prevalence and incidence of dilated cardiomyopathy in the United States, France, Germany, Italy, Spain, and the United Kingdom: a systematic literature review. Children.11(2.38a):1.19b.
12. Arroyave WD, Mehta SS, Guha N, et al. Challenges and recommendations on the conduct of systematic reviews of observational epidemiologic studies in environmental and occupational health. Journal of exposure science & environmental epidemiology. 2021;31(1):21-30.
13. Golchin N, Lesperance T, Scheuring J, et al. EPH207 Incidence and Prevalence of Idiopathic Pulmonary Fibrosis (IPF): A Systematic Literature Review and Meta-Analysis. Value in Health. 2024;27(12):S259-S260.
14. Borges Migliavaca C, Stein C, Colpani V, et al. How are systematic reviews of prevalence conducted? A methodological study. BMC Medical Research Methodology. 2020;20(1):96.