Quality and Impact of Real-world Data in Chronic Disease Research: Insights from Retrospective Studies Using the Hospital Episode Statistics (HES) Database

RWD139

Evelyn Gomez-Espinosa,[a] Maria Koufopoulou,[a] Kimberly Ruiz[b]

Cencora Global Consulting Services: [a]London, United Kingdom; [b]Conshohocken, PA, United States

Background

Real-world data from administrative medical records have become a vital resource for assessing healthcare quality and outcomes. These data, primarily collected by healthcare providers and insurers for reimbursement purposes, include routine demographic information and clinical details derived from standardized coding systems. Although originally intended for different purposes, the benefits of such datasets are well established; they encompass large populations, are readily accessible, and are cost-effective for acquisition.[1]

The Hospital Episode Statistics (HES) database serves as the main administrative data repository for English hospitals within the National Health Service (NHS) since 1989. Studies sourced by HES data have significantly influenced healthcare delivery in England. Significant changes to HES data processing were introduced in September 2021 to enhance patient record-matching and analytical capabilities and to facilitate direct linkage of patient records, augmenting its analytical potential. However, concerns regarding coding accuracy and data completeness have persisted. While previous systematic reviews have focused mainly on coding accuracy, there has been a lack of comprehensive evaluation regarding the methodological quality of studies employing HES data.[2]

Chronic diseases, such as diabetes, cardiovascular diseases, and respiratory disorders, pose significant challenges to healthcare systems worldwide, leading to substantial morbidity. Healthcare costs represent a major disability burden across 24 Organisation for Economic Co-operation and Development countries, where more than one-third of people aged 16 and older reported living with a longstanding illness or health problem in 2021. Research in this area is critical for developing effective interventions, informing public health strategies, and optimizing healthcare resources.[3]

The objective of the present review was to evaluate the publication landscape and quality of retrospective studies published between 2021 and 2025 that used HES data to assess healthcare outcomes related to chronic diseases. By examining the current state of retrospective studies, this review sought to inform future research and enhance the quality of evidence generated from administrative databases.

Methods

Comprehensive searches were performed in the Embase, Emcare, and MEDLINE databases in June 2025 for retrospective cohort studies published in English since 2021 that included the HES database as a data source. The searches were not limited by therapeutic area or disease chronicity; those data were evaluated and extracted during the review of articles.

Chronic disease was defined as a noncommunicable condition or a disorder of any severity that persists for longer or has long-lasting health effects compared to similar conditions of acute or rapid presentation and requires long-term or continuous treatment.

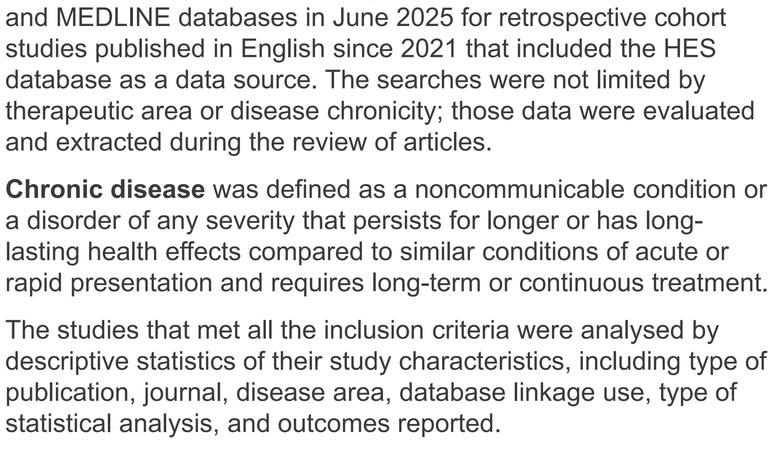
The studies that met all the inclusion criteria were analysed by descriptive statistics of their study characteristics, including type of publication, journal, disease area, database linkage use, type of statistical analysis, and outcomes reported.

Results

Publication trends

After deduplication, 540 articles were reviewed, and a total of 143 publications met the inclusion criteria. There was an average of 33 publications per year since 2021 on retrospective studies of chronic diseases using the HES database as a source (**Figure 1**), highlighting the growing interest in using HES data for healthcare research. Of those publications, 67 were conference abstracts published in journals with an average impact factor of 11.1 (range, 2.3-79.3), and 76 were full articles published in journals with an average impact factor of 4.4 (range, 1.5-11.0).

Figure 1. Number of retrospective publications on chronic diseases from the HES database per year since 2021



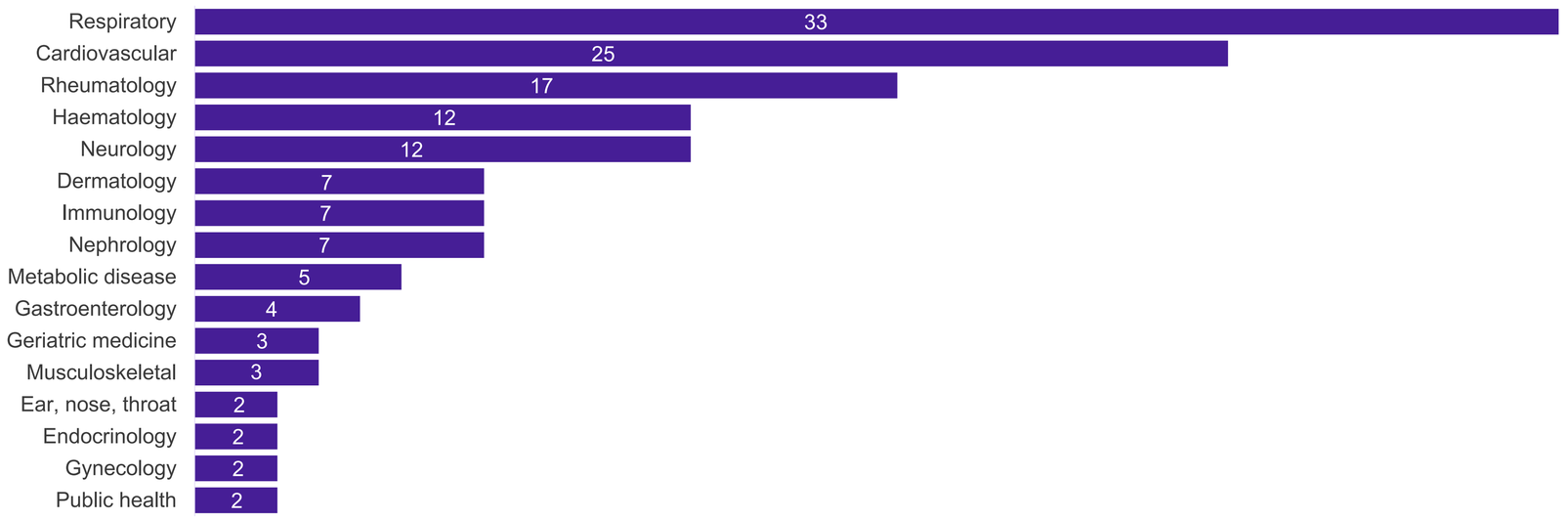
Results (cont.)

Therapeutic areas

A total of 16 different therapeutic areas were studied using HES data (**Figure 2**), encompassing a spectrum of 64 health conditions and diseases. There was a strong emphasis on the following therapeutic research areas:

- Respiratory: 23% (33 publications)
- Cardiovascular: 17% (25 publications)
- Rheumatology: 12% (17 publications)
- Haematology and neurology (each 8%); dermatology, immunology, and nephrology (each 5%)
- Other areas include metabolic disease and gastroenterology (each 3%); geriatric medicine, musculoskeletal conditions (each 2%); and ear/nose/throat conditions and endocrinology, gynaecology, and public health (each 1%)

Figure 2. Number of publications utilizing HES data since 2021 (per therapeutic area)



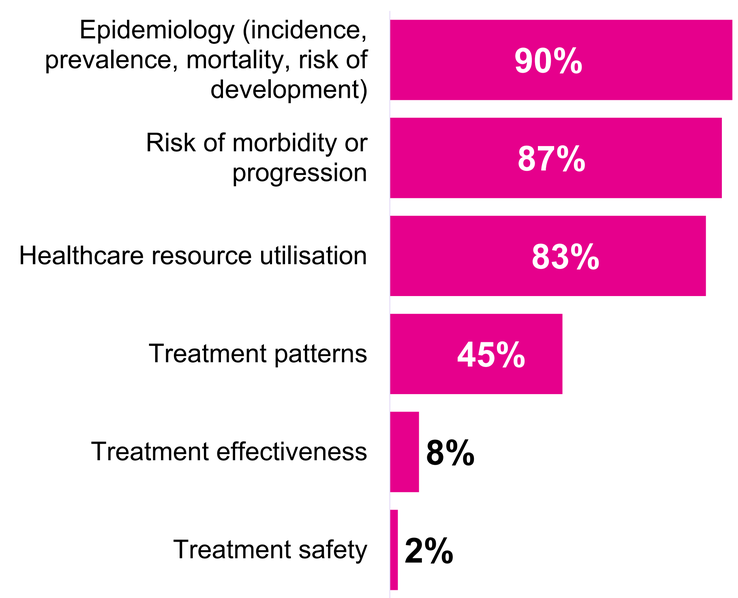
Of the 64 individual conditions identified, the most frequently evaluated included:

- **Chronic obstructive pulmonary disease:** 23 publications
- **Asthma:** 9 publications
- **Chronic kidney disease:** 5 publications
- **Heart failure:** 5 publications
- Other conditions with fewer publications include **hypertrophic cardiomyopathy** and **sickle cell disease** (each with 4 publications)

Outcomes

- Types of outcomes assessed across studies are presented in **Figure 3**.

Figure 3. Distribution of outcomes reported in retrospective studies of chronic diseases using HES data since 2021



Analysis of methodologies

The methodology analysis revealed that up to **91% of studies performed HES data record linkage** based on similar demographic data to increase the strength of the observations, with 80% of them using Clinical Practice Research Datalink (CPRD) databases; for outcomes such as mortality, link HES records were linked to the Office of National Statistics (ONS).

- **CPRD (not specified):** 40% of total publications.
- **CPRD Aurum:** 28%
- **CPRD Gold:** 12%
- **ONS:** 11%
- **Not linked:** 9%

There was a predominant use of descriptive statistics alone (60% of all studies) for data analysis, while 40% of all studies applied more advanced methods of **inferential statistics**.

Discussion and conclusion

This review evaluated the publication landscape and characteristics of retrospective studies, aiming to inform future research efforts and enhance the quality of evidence derived from administrative datasets. Continuing with the trend of increasing volume of publications reported in a 2013 review,[4] the findings of this review highlight a growing, robust, and diverse body of research that is primarily focused on respiratory and cardiovascular conditions and uses administrative databases like HES during the period from 2021 to 2025. This suggests an expanding commitment to understanding chronic diseases and their impact on healthcare systems. Notably, nearly half of the identified studies were published as conference abstracts, often appearing in journals with higher impact factors compared to full articles. This trend suggests an urgency among authors to disseminate findings quickly through reputable sources. However, it is important to note that journal impact factors, while indicative of prestige, do not always reflect the quality or clinical relevance of individual articles.

Most of the identified studies focused on epidemiology outcomes and healthcare resource utilisation; therefore, there remains a significant opportunity to expand research into treatment effectiveness and safety. In term of methodologies, the findings revealed an increased use of database linkage as a critical factor for enhancing data quality and outcome assessments compared to a previous report of only 27% of studies linking the HES data to other sources.[4] Additionally, we noted a predominant reliance on descriptive statistics to report results. While descriptive analyses are helpful for summarising data and identifying trends, they often fail to account for confounding factors or explore relationships and causality, ultimately limiting the strength and applicability of findings. Conducting evaluations with robustly matched cohorts has the potential to provide more meaningful insights, such as risk ratios, odds ratios, and adverse outcomes.

Notably, the recently published ISPOR Suitability of EHR Data (SUITABILITY) Checklist has anticipated the adoption of **advanced epidemiological methods and analytical approaches, along with the integration of new data sources**, as important developments in this field.5 This review contributes to the understanding of the critical role of administrative databases, particularly HES, in advancing research on chronic diseases and their healthcare implications. To further optimize healthcare strategies and improve patient care, future research should prioritize the use of diverse and novel data sources, emphasize methodological rigor, and focus on exploring treatment outcomes in greater depth.

References: 1. Fitzke H, et al. Real-world evidence: state-of-the-art and future perspectives. *J Comp Eff Res*. Apr 2025;14(4):e240130 **2.** Organisation for Economic Co-operation and Development O. *Health at a Glance 2023*: OECD Indicators. 2023. https://doi.org/10.1787/7a7afb35-en **3.** NHS. NHS England: NHS Digital Hospital Episode Statistics (HES). Updated September 2025. Accessed October, 2025. https://digital.nhs.uk/services/hospital-episode-statistics **4.** Sinha S, et al. Studies using English administrative data (Hospital Episode Statistics) to assess health-care outcomes--systematic review and recommendations for reporting. *Eur J Public Health*. Feb 2013;23(1):86-92. **5.** Fleurence RL, et al. Assessing Real-World Data From Electronic Health Records for Health Technology Assessment: The SUITABILITY Checklist: A Good Practices Report of an ISPOR Task Force. *Value in Health*. 2024/06/01/2024;27(6):692-701.

Presented at: ISPOR Annual European Congress
9-12 November 2025 | Glasgow, Scotland, UK
Funded by Cencora

cencora