

Research article

Research outputs of England's Hospital Episode Statistics (HES) database: a bibliometric analysis

Zain Chaudhry

Department of Primary Care and Public Health, Imperial College London, London, UK

Fahmida Mannan

Department of Primary Care and Public Health, Imperial College London, London, UK

Angela Gibson-White

Department of Primary Care and Public Health, Imperial College London, London, UK

Usama Syed

Department of Primary Care and Public Health, Imperial College London, London, UK

Azeem Majeed

Department of Primary Care and Public Health, Imperial College London, London, UK

Shirin Ahmed

Kings College London, London, UK

Cite this article: Chaudhry Z, Mannan F, Gibson-White A, Syed U, Majeed A, Ahmed S. Research outputs of England's Hospital Episode Statistics (HES) database: a bibliometric analysis. *J Innov Health Inform.* 2017;24(4):329–333.

<http://dx.doi.org/10.14236/jhi.v24i4.949>

Copyright © 2017 The Author(s). Published by BCS, The Chartered Institute for IT under Creative Commons license <http://creativecommons.org/licenses/by/4.0/>

Author address for correspondence:

Zain Chaudhry

Department of Primary Care and Public Health
Imperial College London, Faculty of Medicine
London W6 8RP, UK

Email: zain.chaudhry10@imperial.ac.uk

Accepted November 2017

ABSTRACT

Background Hospital administrative data, such as those provided by the Hospital Episode Statistics (HES) database in England, are increasingly being used for research and quality improvement. To date, no study has tried to quantify and examine trends in the use of HES for research purposes.

Objective To examine trends in the use of HES data for research.

Methods Publications generated from the use of HES data were extracted from PubMed and analysed. Publications from 1996 to 2014 were then examined further in the Science Citation Index of the Thompson Scientific Institute for Science Information (Web of Science) for details of research specialty area.

Results 520 studies, categorised into 44 specialty areas, were extracted from PubMed. The review showed an increase in publications over the 19-year period with an average of 27 publications per year, however, with the majority of output observed in the latter part of the study period. The highest number of publications was in the Health Statistics specialty area.

Conclusion The use of HES data for research is becoming more common. Increase in publications over time shows that researchers are beginning to take advantage of the potential of HES data. Although HES is a valuable database, concerns exist over the accuracy and completeness of the data entered. Clinicians need to be more engaged with HES for the full potential of this database to be harnessed.

INTRODUCTION

Hospital administrative data, such as that provided by the Hospital Episode Statistics (HES) database in England, are increasingly being used for research and quality improvement. The HES database was developed in 1987 and was designed to collect a detailed record of each episode of care received by a patient, either from National Health Service (NHS) inpatient providers or independent sector providers commissioned by the NHS to deliver care.¹ It has covered inpatient data since 1989, outpatient since 2003 and Accident and Emergency (A&E) since 2007.²

HES data are compiled from the data sent by NHS trusts and foundation trusts, including acute hospital trusts and mental health trusts in England.³ Independent sector organisations also send data to HES for activity commissioned by the NHS England. The data are stored as a large collection of separate records in a secure data warehouse.² Each record relates to a separate episode of care: a period of care for patient under a single consultant at a single hospital.⁴ The Health and Social Care Information Centre (HSCIC) liaises closely with NHS organisations to ensure that data submitted are complete and accurate.³

The HES database houses data containing patient details (such as age, sex, NHS number and the location of residence), administrative details (such as hospital name, consultant, general practitioner referral and emergency/elective admission details and clinical details (such as primary diagnosis, secondary diagnoses and operative procedures).³ HES data are largely designed for non-clinical purposes, including hospital remuneration following the delivery of care to patients. Some other secondary uses of the data available in HES include: monitoring trends and patterns in hospital activity, general medical research and statistical functions, developing and monitoring government policies as well as providing the basis for national health indicators of clinical quality such as hospital readmission rates to examine health outcomes and to improve patient experience.

HES is not a live system, and it usually takes between 9 and 12 months for a full financial year to be available through online HES or the safe haven extract (secure locations within an organisation accessible via a defined security and administrative protocol, which ensure confidential data transfer).⁵ So, for example, data for the year ending 31 March 2015 would not be available until around December 2015.⁶ Similar databases to HES exist in Wales, Northern Ireland and in Scotland. HES is increasingly being used by researchers due its comprehensive coverage of NHS-funded hospital activity in England. To date, no study has tried to quantify or examine trends in the use of HES for research. The aim of this study was to analyse the research outputs and the growth in the number of publications using the HES database in England.

METHODS

To review the impact of HES data on research, publications generated from the use of HES data from 1996–2014 inclusive

(a 19-year period), were extracted from PubMed and analysed. Extracted publications were examined in the Science Citation Index (SCI) of the Thomson Scientific Institute for Scientific Information (also known as Web of Science) for details of the publication's specialty area as assigned by the Web of Science. Publications over the 19-year period were grouped by specialty area and then tabulated according to the number of publications. The average number of publications per year was calculated, and cumulative growth over the 19-year period was also looked at, as well as the journals that were most published in.

RESULTS

Publication growth

520 studies that used HES data between 1996 and 2014 were extracted from PubMed. As listed in Table 1 and represented in Figure 1, there was an increase in the number of publications over the 19-year period, with the largest growth in the last five years. The average number of publications over the 19-year period was 27 publications per year; however, research output was heterogeneous across the study period with most publications occurring in the latter years (Table 1). The largest number of publications in this period was in 2013 and 2014: 73 and 99 publications, respectively. The 520 publications during the study period were categorised into a total of 46 specialties by SCI.

Table 1 Number of publications using HES data between 1996 and 2014

Year	Publication
1996	1
1997	3
1998	8
1999	2
2000	7
2001	5
2002	10
2003	12
2004	19
2005	12
2006	19
2007	40
2008	27
2009	30
2010	39
2011	45
2012	69
2013	73
2014	99
TOTAL	520

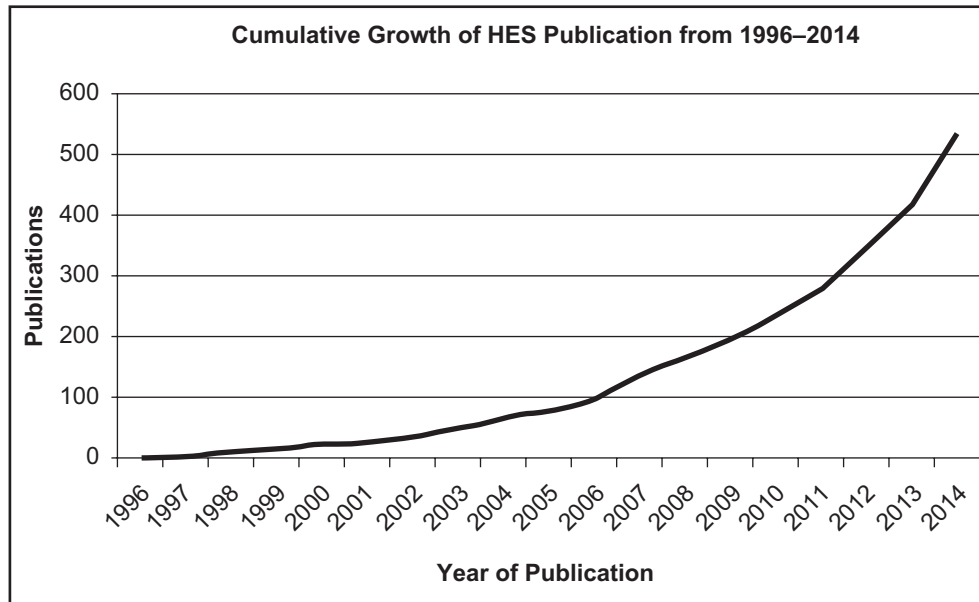


Figure 1 Cumulative growth of HES publications from 1996 to 2014

Specialty areas

As shown in Table 2, the three specialty areas with the highest number of publications were Health Statistics ($n = 50$), Oncology ($n = 40$) and Vascular Surgery ($n = 39$). The least published in specialties with only one publication each were: Anaesthetics, Community Health, Endocrine Surgery, Immunology, Intensive Care, Management, Medical Technology, Musculoskeletal, Pharmacy, Primary Care and Radiology. There was one non-medical specialty area represented under Management ($n = 1$); this was from a publication looking at the management of large-scale change in the NHS.

Journals

The three principal journals in which HES data were published were: The British Medical Journal (9.2% and $n = 48$), The British Journal of Surgery (5.6% and $n = 29$) and The Journal of Public Health (3.7% and $n = 19$). Most journals, however, published just one article, with 111 examples of this in our review.

DISCUSSION

HES is primarily an administrative database that collects detailed records of each episode of hospital care received by a patient in England's NHS. Its uses include supporting payment to hospitals, benchmarking performances between trusts and as a source of data for researchers. This review looked at the research output of HES over the past 19 years. Our analysis shows that in the last 19 years there has been a slow but consistent growth in the use of HES data for research, with an output of just one publication in 1996 increasing significantly to total 520 publications by 2014.

This review demonstrates that researchers from a variety of specialties are drawing on HES data for use in their research, indicating its perceived value. However, of the 46 specialty areas represented, the most published in specialty

(Health Statistics) produced only 50 publications over a 19-year period, indicating the HES database is a somewhat untapped resource. The British Medical Journal, British Journal of Surgery and the Journal of Public Health published the largest proportion of the HES data publications, but the majority of journals appeared to have published one or two articles. 111 journals published just one article perhaps suggesting that HES data are relatively unknown and subsequently underused in most specialties.

There are some known limitations with this study such as exclusion of some HES publications. As only publications extracted from PubMed were used in this study, there could be other HES publications listed in non-PubMed indexed journals that may have been overlooked. Indeed, much high impact, grey literature such as that output from the Nuffield Trust and The King's Fund are not indexed on PubMed and, therefore, would be missed by this analysis. Some of the articles extracted could have also been published in non-peer reviewed journals. Using publications extracted from PubMed alone, however, is still sufficient to indicate the steady growth in the use of HES data for research. Significant growth over the last five years is also indicative of the fact that, as technology advances and the ability to share, store and transfer data grows, this will lead to a rise in popularity of this method of research. It seems that the use of large clinical databases for research is becoming increasingly common worldwide.⁷

With HES now able to link to other data such as the Office for National Statistics (ONS) mortality data,⁸ this gives even more of an opportunity to researchers because the linkage captures deaths of people in the HES database who died outside of hospital. HES is now also linked to Patient Reported Outcome Measures (PROMS), which contains data from questionnaires completed by patients before and after hip replacement, knee replacement, groin hernia and varicose vein surgery.⁹ HES data are also now linked to the patient records in the Clinical Practice Research Database (CPRD) that pulls in data from primary care, extending the scope of data available to researchers.¹⁰

Table 2 Journal specialties ranked by number of publications between 1996 and 2014

Specialty	Number of publications
Health Statistics	50
Oncology	40
Vascular Surgery	39
Infectious Diseases	37
Orthopaedic Surgery	31
Obstetrics and Gynaecology	28
Colorectal Surgery	27
Paediatrics	22
Psychiatry	21
Gastroenterology	20
Urology	20
Public Health	18
Ophthalmology	17
Cardiology	16
Neurology	15
Diabetes	11
Emergency Medicine	11
Upper GI Surgery	10
Respiratory	8
Cardiothoracic Surgery	7
Clinical Pharmacology	6
ENT Surgery	6
Nephrology	6
Breast Surgery	5
Dentistry	5
Health Economics	5
Sociology	5
Bariatric Surgery	4
Haematology	4
Maxillofacial Surgery	4
Paediatric Surgery	4
Emergency General Surgery	3
Hepatology	3
Clinical Toxicology	2
Rheumatology	2
Anaesthetics	1
Community Health	1
Endocrine Surgery	1
Immunology	1
Intensive Care	1
Management	1
Medical Technology	1
Musculoskeletal	1
Pharmacy	1
Primary Care	1
Radiology	1

As mentioned above, linkage of HES data to other national databases has increased over time. This brings with it a number of advantages by enhancing the potential applications of HES data. For example, linkage with the ONS mortality database enriches its utility for research purposes by providing access to data on cause of death in the community. This continuity of records from inpatient admission to community mortality data could be used to generate survival analyses in addition to answering other research questions. Furthermore, many cohort and longitudinal studies are pursuing linkage to HES, allowing researchers to explore baseline characteristics of a demographic as determinants of hospital admission. For example, the Hertfordshire Cohort Study linked detailed physiological parameters of study participants from a community cohort to HES data, allowing determinants of admission to be established.¹¹ The above is also extending into the discipline of mental health with HES now linked to the Mental Health and Learning Disabilities Data Set, enhancing our understanding of the contact between this specific demographic and acute secondary care.¹²

The comprehensive nature of HES, bestowed by the large amounts of data that it holds, makes it a strong research tool, but its potential can only be harnessed if the quality of the data it holds is complete and accurate. Concerns have been raised about the lack of involvement or engagement of clinicians in the process of data collection. Clinicians who enter the data still need a lot of encouragement to do so. Williams and Mann¹³ in their article 'HES: time for clinicians to get involved', questioned the validity of HES because they believed that there was no uniformity in the quality of data it provided. They concluded that this lack of uniformity was because physicians are not sufficiently engaged in the process of data collection for HES. It does seem that, though, there have been improvements in HES data quality with particular respect to its validity post-accreditation.¹⁴ Another issue with the HES dataset validity arises from clinical coding. Although trained staff in hospitals are very effective at accurately coding and entering information, the information clinicians provide in patient notes and discharge summaries can often be incomplete or unclear for the purposes of coding. This has been cited as a "possible weak link in the data quality chain".¹⁵ Improving HES data quality is being addressed by the Royal College of Physicians, who show commitment to training physicians in order for them to be more engaged with HES and other clinical and administrative databases.¹⁵ Other professional bodies also show a similar commitment, with HSCIC and the Academy of Medical Royal Colleges (AOMRC) creating a joint initiative to help clinicians improve the quality of data entered in HES.¹⁶ Hospital trusts also have a part to play by collaborating with clinicians when there is submission of data to secondary user services for HES; however, clinicians still need regular access to data with an interface that is easy to use.¹⁷

CONCLUSION

The use of HES data for research is increasing. The 520 publications produced within the 19-year period do suggest steady

and consistent growth, but concerns still exist about the accuracy and completeness of the data entered. Clinicians are continuously being urged to get engaged and realise the clinical benefits of HES. The AOMRC is working towards improving the quality of data in HES and, in 2011, set out three major goals towards improving the quality of HES data.¹⁶ If clinicians do get more engaged with HES data and there is more trust in the accuracy of the data, then its use for research is likely to increase further; particularly, as HES becomes linked to additional data sources such as PROMS, CPRD and mortality data and technological advancements streamline the process of extracting data for prospective researchers.

REFERENCES

1. NHS Digital. *Hospital Episode Statistics (HES)*. Available from: <http://content.digital.nhs.uk/hes>. Accessed 20 April 2017.
2. Herbert A, Wijlaars L, Zylbersztejn A, Cromwell D and Hardelid P. Data Resource Profile: Hospital Episode Statistics Admitted Patient Care (HES APC). *International Journal of Epidemiology* 2017;46:1093. doi: 10.1093/ije/dyx015. Accessed 20 April 2017. Available from: <https://doi.org/10.1093/ije/dyx015>.
3. The Health and Social Care Information Centre. *Hospital Outpatient Activity Summary Report 2011–2012*. Available from: <http://www.hscic.gov.uk/catalogue/PUB09379/hosp-outp-acti-11-12-summ-repo-rep.pdf>. Accessed 18 April 2017.
4. NHS Digital. *Hospital Episode Statistics, Admitted Patient Care – England, 2014 – 2015 [NS]*. Available from: <http://content.digital.nhs.uk/catalogue/PUB19124>. Accessed 30 March 2017.
5. NHS Greater Glasgow and Clyde. *About the Safe Haven*. Available from: <http://www.nhsggc.org.uk/about-us/professional-support-sites/nhsggc-safe-haven/about-the-safe-haven/>. Accessed 15 March 2017.
6. Royal College of Physicians and Unit of Health-Care Epidemiology. *HES for Physicians: A Guide to the Use of Information Derived from Hospital Episode Statistics*. 2007. Available from: <http://www.uhce.ox.ac.uk/hesepho/reports/HESForPhysicians.pdf>. Accessed 14 March 2017.
7. Cook JA and Collins GS. The rise of big clinical databases. *The British Journal of Surgery* 2013;102:e93–e101. doi: 10.1002/bjs.9723. Accessed 7 March 2017. Available from: <https://doi.org/10.1002/bjs.9723>.
8. NHS Digital. *Linked HES-ONS Mortality Data*. Available from: <http://content.digital.nhs.uk/article/2677/Linked-HES-ONS-mortality-data>. Accessed 26 February 2017.
9. NHS Digital. *Hospital Episode Statistics Linkage to Patient Reported Outcome Measures*. Available from: <http://www.hscic.gov.uk/hesproms>. Accessed 4 January 2017.
10. Clinical Practice Research Datalink. *CPRD Record Linkage*. Available from: <http://www.cprd.com/recordLinkage/>. Accessed 18 February 2017.
11. Simmonds SJ, Syddall HE, Walsh B, Evandrou M, Dennison EM, Cooper C, et al. Understanding NHS hospital admissions in England: linkage of Hospital Episode Statistics to the Hertfordshire Cohort Study. *Age Ageing* 2014;43:653–60. Available from: doi: 10.1093/ageing/afu020. Accessed 29 March 2017. Available from: <https://doi.org/10.1093/ageing/afu020>.
12. NHS Digital. *Hospital Episode Statistics Linkage to the Mental Health and Learning Disabilities Data Set*. Available from: <http://content.digital.nhs.uk/hesmhlds>. Accessed 14 March 2017.
13. Williams JG and Mann RY. Hospital episode statistics: time for clinicians to get involved? *Clinical Medicine (London)* 2002;2:34–37. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/?term=Williams+JG%2C+Mann+RY,+Hospital+episode+statistics%3A+time+for+clinicians+to+get+involved%3F>. Accessed 16 March 2017. Available from: <https://doi.org/10.7861/clinmedicine.2-1-34>.
14. Thorn JC, Turner E, Hounscome L, Walsh E, Donovan JL, Verne J, et al. Validation of the Hospital Episode Statistics outpatient dataset in England. *Pharmacoeconomics* 2016;34:161–68. Available from: doi: 10.1007/s40273-015-0326-3. Accessed 11 March 2017. Available from: <https://doi.org/10.1007/s40273-015-0326-3>.
15. Royal College of Physicians. *Engaging Clinicians in Improving Data Quality in the NHS; Key Findings and Recommendations from Research Conducted by the Royal College of Physicians' iLab*. 2006. Available from: https://www.rcplondon.ac.uk/sites/default/files/ilab-summary-report_1.pdf. Accessed 12 April 2017.
16. Academy of Medical Royal Colleges. *Hospital Episode Statistics (HES): Improving the Quality and Value of Hospital Data*. 2011. Available from: http://www.hscic.gov.uk/media/1593/Hospital-Episode-Statistics-Improving-the-quality-and-value-of-hospital-data/pdf/HES_-_Improving_the_quality_and_value_of_hospital_data.pdf. Accessed 5 March 2017.
17. Spencer A. Clinicians and trusts have a role in data quality. *British Medical Journal* 2012;344:e2926. doi: 10.1136/bmj.e2926. Available from: <https://doi.org/10.1136/bmj.e2926>.