

Online Research @ Cardiff

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/106941/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Ingram, John R., Jenkins-Jones, S., Knipe, D. W., Morgan, C. L. I., Cannings-John, Rebecca and Piguët, Vincent 2018. Population-based Clinical Practice Research Datalink study using algorithm modelling to identify the true burden of hidradenitis suppurativa. *British Journal of Dermatology* 178 (4) , pp. 917-924. 10.1111/bjd.16101 file

Publishers page: <http://dx.doi.org/10.1111/bjd.16101>
<<http://dx.doi.org/10.1111/bjd.16101>>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies.

See

<http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



Article type : Original Article

Population-based Clinical Practice Research Datalink study using
algorithm modelling to identify the true burden of hidradenitis
suppurativa

J.R. Ingram*,¹ S. Jenkins-Jones,² D.W. Knipe,³ C.L.I. Morgan,² R. Cannings-John,⁴ V. Piguet^{1,5,6}

Current Author Affiliations:

1. Department of Dermatology & Academic Wound Healing, Division of Infection and Immunity, Cardiff University, Cardiff CF14 4XN, UK.
2. Pharmatelligence, Cardiff CF14 3QX, UK.
3. School of Social and Community Medicine, University of Bristol, Bristol BS8 2PS, UK.
4. Centre for Trials Research, College of Biomedical and Life Sciences, Cardiff University, Cardiff CF14 4YS, UK.
5. Division of Dermatology, Women's College Hospital, Toronto, Canada.
6. Department of Medicine, University of Toronto, Toronto, Canada.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/bjd.16101

This article is protected by copyright. All rights reserved.

Correspondence to: John R Ingram, 3rd floor Glamorgan House, University Hospital of Wales, Heath Park, Cardiff, CF14 4XN, UK. Email: ingramjr@cardiff.ac.uk

Running head: Population-based CPRD study of hidradenitis suppurativa prevalence

Funding / Support: Dr Ingram was supported by a Health Fellowship (HF 14-08) from Health and Care Research Wales (HCRW). The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication. The views expressed in this publication are those of the authors and not necessarily those of HCRW.

Conflict of Interest Disclosures: Dr Ingram was a local principal investigator for a HS observational study sponsored by Abbvie and received a speaker fee from UCB Pharma for a presentation to raise awareness of HS. He received travel expenses to attend a HS meeting sponsored by Abbvie but his speaker fee was donated directly to charity. Ms Jenkins-Jones and Mr Morgan work for Pharmatelligence, a research consultancy receiving funding from pharmaceutical companies. Prof Piguet undertakes personal advisory work with Pfizer, AbbVie, Janssen, Novartis, Almirall and Celgene and has received departmental support from AbbVie, Almirall, Alliance, Beiersdorf UK Ltd, Biotest, Celgene, Galderma, Genus Pharma, Janssen, LEO Pharma, Meda, MSD, Novartis, Pfizer, Sinclair Pharma, Spirit Pharmaceuticals, Stiefel, Sumed, and TyPham. Authors not named here have disclosed no conflicts of interest.

What's already known about this topic?

Previous population-based studies suggest that the prevalence of hidradenitis suppurativa (HS) may be as low as 0.05% but did not quantify undiagnosed cases.

What does this study add?

Analysis of 4.3 million research-standard records in the UK Clinical Practice Research Datalink demonstrates a HS prevalence of 0.77% using the most stringent disease definition (95% CI 0.76 to 0.78%), one-third being undiagnosed cases, rising to 1.19% (95% CI 1.18 to 1.20%) if probable cases are included.

Summary

Background

Epidemiology data regarding hidradenitis suppurativa (HS) are conflicting and prevalence estimates vary 80-fold, from 0.05% in a population-based study, to 4%.

Objectives

To assess the hypothesis that previous population-based studies under-estimated true HS prevalence by missing undiagnosed cases.

Methods

We performed a population-based observational and case-control study using the UK Clinical Practice Research Datalink (CPRD) linked to Hospital Episode Statistics data. Physician-diagnosed cases in CPRD were identified from specific Read codes. Algorithms identified unrecognised 'proxy' cases, with at least five Read code records for boils in flexural skin sites. Validation of proxy cases

was undertaken with General Practitioner questionnaires to confirm criteria-diagnosed cases. A case-control study assessed disease associations.

Results

On 30 June 2013, 23,353 physician-diagnosed HS cases were documented in 4,364,308 research-standard records. 68,890 proxy cases were identified, reduced to 10,146 criteria-diagnosed cases after validation, extrapolated from 107 completed questionnaires (61% return rate). Overall point prevalence was 0.77% (95% CI 0.76% to 0.78%). An additional 18,417 cases had a history of 1-4 flexural skin boils.

In physician-diagnosed cases, ORs for current smoker and obesity (BMI>30) were 3.61 (95% CI 3.44 to 3.79) and 3.29 (95% CI 3.14 to 3.45). HS was associated with type 2 diabetes, Crohn's disease, hyperlipidaemia, acne and depression and not associated with ulcerative colitis or polycystic ovary syndrome.

Conclusions

Contrary to results of previous population-based studies, HS is relatively common, with a UK prevalence of 0.77%, one-third being unrecognised, criteria-diagnosed cases using the most stringent disease definition. If probable cases are included, HS prevalence rises to 1.19%.

Introduction

Hidradenitis suppurativa (HS), also known as acne inversa, is an inflammatory skin disease producing functional impairment similar in magnitude to cardiovascular disease, type 2 diabetes and renal failure, due to pain, purulent discharge and scarring.^{1,2} HS is defined clinically by characteristic lesions occurring chronically in flexural locations.^{3,4} Consensus disease definitions differ slightly in the chronicity element, ranging from a history of two to five boils in the axillae or groins.^{5,6}

The prevalence of HS remains controversial, with estimates ranging from 0.05% from an analysis of patient insurance claims in the USA,⁷ to 4% when young adult women were examined in person.⁸ An accurate prevalence figure is required to quantify the unmet need regarding care of people with HS to assess disease burden and plan health care provision. In particular radical surgery or biological therapies are needed for severe disease which are costly and often require multi-disciplinary teams; for example, in the UK, the National Institute for Health and Care Excellence (NICE) recently approved adalimumab for moderate to severe HS.⁹

One explanation for the 80-fold difference in prevalence estimates currently available is lack of recognition of HS, in the context that average diagnostic delay is seven years.¹⁰ The aim of our study was to provide the most accurate estimate to date of HS prevalence by quantifying physician-diagnosed cases, as well as unrecognised HS cases meeting the most stringent consensus disease definition for HS. A primary care data source was chosen because the majority of HS care takes place within a primary care setting. Our second aim was to assess HS disease associations with the largest case-control study performed to date. Accurate epidemiological data is critical at a time when treatment for HS is dramatically changing and now includes biological therapies.

Patients and methods

Our manuscript was prepared in accordance with RECORD, Reporting of Studies Conducted using Observational Routinely-collected Data¹¹ (for RECORD checklist see Appendix S1; Supporting Information). The Clinical Practice Research Datalink (CPRD) contains routinely recorded, pseudonymised data from participating primary care practices throughout the UK.¹² Most diagnoses made in secondary care are transcribed into CPRD by primary care staff. At January 2015, it contained records from more than 15.6 million patients registered at 684 practices. Patients are labelled by CPRD as being of acceptable research quality if they have been permanently registered at

their general practice and have internally consistent records with regard to age, sex, registration, and event dates. Approximately 75% of contributing practices based in England (representing 60% of the entire data set) participate in a CPRD linkage scheme by which patient-level data are linked to the Hospital Episode Statistics (HES) for England. The NHS number, date of birth, sex and postal code of non-dissenting patients are conveyed from the practice to a trusted third party and matched with equivalent identifiers in the target data set. Only those linkages involving at least two exact matches, which must include an exact match on NHS number, are retained.

The study population was selected from patients of acceptable research quality. Physician-diagnosed cases of HS were captured by two specific diagnostic Read codes, M25y100 (Hidradenitis) and M25y111 (Hidradenitis suppurativa), or, in HES data, by the ICD-10 code L73.2 (Hidradenitis suppurativa). To capture undiagnosed cases, CPRD Read code algorithms were created to identify patients attending primary care for multiple skin boil consultations (Table 1). An algorithm hierarchy was used with different degrees of stringency for each sub-algorithm, based on the HS disease definition (Fig. 1). Some primary care practitioners may not record every skin boil consultation, particularly if the patient also attends with other health issues. To capture these potential cases, treatment codes were used to identify patients receiving multiple short courses of skin-directed antibiotic therapy, such as flucloxacillin, in combination with skin boil Read codes and in the absence of any other indication for the antibiotic, such as eczema, cellulitis or skin ulcers. Prior to validation, these patients are referred to as 'proxy cases.' HES data were not used to identify proxy cases to avoid double-counting boils recorded in both data sources.

Validation

Validation of proxy cases was carried out by sending questionnaires to the primary care physicians of a subset of patients identified from the algorithms. The questionnaire (Table 2) requested confirmation of characteristic skin lesions in flexural locations on at least five separate occasions, in keeping with the most stringent consensus definition of the disease.⁶

A total of 176 questionnaires were sent, with the overall number being determined by our study budget in the context that CPRD charges a fee for each questionnaire issued. The number of questionnaires generated to validate each sub-algorithm was determined by the number of additional proxy cases identified through each, with a minimum of 10 questionnaires being issued for sub-algorithms that captured relatively few cases.

Analysis

Point prevalence on 30 June 2013 was calculated using physician-diagnosed cases and criteria-diagnosed cases as the numerator and the 2013 mid-year population of live patients with research-standard records in CPRD as the denominator. The number of criteria-diagnosed cases was calculated by multiplying the number of proxy cases identified by each sub-algorithm with the proportion of cases validated from the relevant set of GP questionnaires.

Annual incidence rates for physician-diagnosed cases were calculated from 1988, when up-to-research standard practices first contributed to CPRD, until 2013. We used the Brameld-Holman backcasting method, which implements an actuarial retrograde survival curve to derive correction factors for the over-ascertainment of first events for individuals with only a limited duration of CPRD data collection prior to the event.¹³ The Brameld-Holman method accounts for the prevalent pool effect without the need for a clearance period and avoids data being discarded.

Case-control

Cases and controls were matched in a 1:1 ratio based on age, gender and registration at the same primary care practice. The associations between HS and environmental factors and other medical conditions were estimated using conditional logistic regression techniques. Two separate association analyses were conducted for physician-diagnosed cases compared with matched controls and for proxy cases (including criteria-diagnosed HS cases and those proxy cases who did not meet the validation criteria) compared to a further set of matched controls. Univariable associations were

reported, as well as multivariable analyses that examined combinations of variables demonstrating association in the univariable analysis with a P-value <0.1. In a post-hoc analysis we also tested whether there was evidence that the association between Crohn's disease and HS was altered by smoking status because smoking is known to be a risk factor for both conditions.

IRB Approval

Our study protocol (Appendix S2; see Supporting Information), 15_020R, was prospectively approved by the CPRD Independent Scientific Advisory Committee. No additional Institutional Review Board approval is required for CPRD studies.

Results

Point prevalence

On 30 June 2013 there were 21,575 physician-diagnosed HS patients in the CPRD dataset and a further 1778 from the HES dataset, a total of 23,353 physician-diagnosed cases. Our Read and treatment code algorithms captured a further 68,890 proxy cases in CPRD; the breakdown for each sub-algorithm is shown in Table 3. Of the 176 GP validation questionnaires sent, 107 (61%) were fully completed and returned; completion rates for each sub-algorithm are again in Table 3.

The proportions of questionnaires with 'yes' responses to the first three questions, satisfying the most stringent disease definition of at least five flexural skin boils, are documented in Table 3; from these, 10,146 criteria-diagnosed cases were extrapolated. Table 3 also documents the proportions within each sub-algorithm confirmed to have had 1-4 flexural boils with 'yes' responses to questions one and two only, representing a group of probable HS cases. Overall, only one validation questionnaire confirmed the presence of scarring with a 'yes' answer to question four.

Using the denominator of 4,364,308 research-standard patients in CPRD, the prevalence of physician-diagnosed and criteria-diagnosed HS cases on 30 June 2013 was 7.7 per 1,000 (95% CI 7.6, 7.8). When the 18,417 patients with 1-4 flexural skin boils were included, the upper limit of probable HS prevalence was 11.9 per 1,000 (95% CI 11.8, 12.0).

The demographic profile of physician-diagnosed HS patients is shown in Figure 2. Peak prevalence of 15.1 per 1,000 occurred in the fifth decade of life and the mean female: male ratio across all the age groups was 2.9: 1.

Incidence of physician-diagnosed cases

The Brameld-Holman actuarial retrograde survival curve used for the backcasting correction for incidence rates is in Figure S1 (see Supporting Information), with the resulting correction factors reducing the number of new physician-diagnosed cases prior to 1994. Mean annual incidence rate for physician-diagnosed cases from 1996 to 2013 was 28.3 per 100,000 person-years (Table S1). Prior to 1995, despite the backcasting correction, annual incidence rates became progressively higher, ranging from 35.7 per 100,000 person-years in 1995 to 122.1 in 1988.

Case-control

Descriptive statistics for the occurrence of environmental factors and other medical conditions in cases and controls are in Table 4. Obesity (BMI > 30 kgm⁻²), current smoking and type 2 diabetes mellitus were strongly linked with physician-diagnosed cases of HS, with odds ratios (ORs) of 3.29 (95% CI 3.14, 3.45), 3.61 (95% CI 3.44, 3.79) and 3.39 (3.09, 3.71) respectively in the univariable analysis (Table 4). There was a significant association between HS and Crohn's disease but not ulcerative colitis. The link with Crohn's disease was not stronger for smokers compared with non-smokers (P=0.26 for interaction test between Crohn's disease and never or past/current smoker).

Associations were also found between HS and hyperlipidaemia, pilonidal sinus, acne vulgaris, depression and hypertension, but not with inflammatory arthritis, polycystic ovary syndrome (PCOS),

psoriasis, or Alzheimer's disease (Table 4). Multivariable analysis confirmed the univariable analysis results in all cases (Table S2).

Discussion

Our HS prevalence figure of 0.77% is derived from a population of 4.3 million UK residents with research-standard medical records. Importantly, our study has identified previously undiagnosed HS patients using the most stringent diagnostic criteria available. We have also identified an upper limit of probable HS prevalence of 1.19%, if those with a history of 1-4 flexural skin boils on separate occasions are included. This is in the context that patients may not see their GP for every new skin boil, GPs may not record each skin boil, particularly if the consultation also covers other medical issues, and HS disease definitions vary in the number of flexural skin boils required from 2-5.⁵⁻⁶

Our two prevalence figures span the 0.97% estimate for European prevalence derived from a self-reported questionnaire study of 10,000 members of the French population which is often cited as the most accurate current estimate.¹⁴ In the French study, cases were identified by positive responses to one question within a general skin disease survey and the morphology of skin lesions was not assessed by healthcare professionals.

One of the largest population-based HS investigations performed prior to our study used data from the Rochester Epidemiology Project (REP), which maintains a centralised electronic medical record database for the 144,000 people living in Olmsted County Minnesota, USA.^{15,16} Identification of HS cases used codes including synonyms of 'hidradenitis', as well as the terms 'infection, sweat gland or inflammation, sweat gland'. Validation by chart review identified 178 HS cases, giving a prevalence of 0.13%. More recently, interrogation of pooled insurance and self-pay databases covering 48 million people in the USA using ICD-9 coding for HS gave a point prevalence of 0.10%.¹⁷ In both studies, identification of undiagnosed cases by considering multiple flexural skin boil presentations

was not attempted. Part of the explanation for the lower USA prevalence figures compared with our data is failure to capture undiagnosed cases. In addition, recognition and coding of HS by health care professionals will vary from region to region and variations in health systems may influence whether people with HS persevere with seeking medical attention for their condition, an issue requiring further investigation.

Our finding of peak HS prevalence in women in their fourth and fifth decades of life is in agreement with the French self-reported questionnaire study, in which the female: male ratio was 2.7:1 and the mean age of patients was 43.2 years.¹⁴ Our peak prevalence finding of 1.51% in women in their fifth decade is lower than the 4% prevalence found by examination of a group of female healthcare professionals and those attending for benign skin lesions in Denmark, but the study had a relatively low sample size of 100, in which there were four cases.⁸

Our mean annual incidence of physician-diagnosed HS cases of 28.3 per 100,000 person-years from 1996 to 2013 is more than four times higher than the 6.0 per 100,000 person-years figure found using REP data for Olmsted County Minnesota.¹⁶ The difference is in keeping with the higher prevalence figure for physician-diagnosed cases found in our study. The more recent study of the USA population using pooled insurance and self-pay databases found a HS incidence of 11.4 (95% CI 11.1 to 11.8) per 100,000 population, double the incidence of the REP-derived estimate.¹⁸ Our calculated incidence rates were much higher in the first few years of available CPRD data, despite the backcasting correction to negate any prevalent pool effect. The observation is likely explained by bulk computerisation of patient paper records during this period, producing an artefact in this result.

Our case-control study is entirely consistent with other, smaller studies previously performed. A recent systematic review of nine relevant studies containing 6174 patients and 24,993 controls confirmed an association between HS and obesity (OR 3.45, 95% CI 2.20, 5.38), as well as an association with diabetes (OR 2.85, 95% CI 1.34, 6.08), current smoking (OR 4.34, 95% CI 2.48, 7.60) and past smoking (OR 6.34, 95% CI 2.41, 16.68).¹⁹ Our finding that HS is associated with Crohn's

disease but not ulcerative colitis is in agreement with other studies, including a cross-sectional investigation of 3,267 HS patients²⁰ and, in addition, we showed that the association between HS and Crohn's disease is not modified by smoking status. The higher rates of smoking, type 2 diabetes mellitus, hypertension and hyperlipidaemia confirmed by our study are in keeping with the nearly doubled risk of cardiovascular-associated death in Danish HS patients compared with controls (adjusted incidence rate ratio 1.95, 95% CI 1.42, 2.67).²¹

The strongest association demonstrated by our case-control study was between HS and pilonidal sinus, with an odds ratio of 5.61. The result is expected because both conditions are members of the 'follicular occlusion tetrad' and there is ongoing debate about whether pilonidal sinus may be a phenotypic variant of HS. We confirmed a link between HS and depression, with an odds ratio of 1.55, which is in keeping with a chronic, painful and socially isolating disease and in agreement with a study of psychiatric comorbidities in 3207 HS patients (adjusted OR for depression 1.7, 95% CI 1.3-2.1).²² However, we did not assess the timing of disease associations in our study and so cannot comment further on possible causation.

Debate continues regarding a possible link between HS and inflammatory arthritis. A syndrome known as PAPASH (pyoderma gangrenosum, acne, psoriasis, arthritis and suppurative hidradenitis) has been proposed but we did not find evidence of an association between HS and inflammatory arthritis or psoriasis.²³ Our finding of no link between HS and PCOS disagrees with a case-control study based on 2292 patients in the USA with at least one billing code for HS, which found an odds ratio of 13.7 (95% CI 4.00, 47.3).²⁴ However, the study found a HS prevalence of only 0.08%, reflecting limited case-capture from billing codes, and so case ascertainment bias may have affected the results. We specifically checked for any link between HS and Alzheimer's disease because of reports of loss of function gamma secretase gene mutations in some Han Chinese HS patients and a small proportion of European families with HS.^{25,26} Missense gamma secretase gene mutations are implicated in familial Alzheimer's disease.²⁷ We found no association between HS and Alzheimer's

disease, in agreement with a recent case-control study from Denmark,²⁸ although it is possible that our relatively young cohort of patients might subsequently develop dementia later in life.

Considering our study limitations, some relevant data is missing within CPRD and it is likely that these will not be missing at random because cases are likely to have greater completeness due to their condition, increasing GP contacts and providing more opportunity for the data items to be recorded. Coding imperfections may affect results, mitigated by our validation of undiagnosed cases using the GP questionnaire. Study budget limitations, rather than statistical considerations, restricted the number of validation questionnaires that were sent and so a sampling error is possible.

We did not validate physician-diagnosed HS cases because false positives are unlikely in the context that HS is a clinically diagnosed, characteristic skin disease with a low chance of being confused with other skin conditions. Review of medical records for patients who had received at least two ICD-9 705.83 codes for 'hidradenitis' in the Massachusetts General Hospital database was previously conducted, in the context that the 705.83 code includes neutrophilic eccrine hidradenitis, and recurrent palmoplantar hidradenitis, as well as HS.²⁹ Confirmation of HS was obtained in 89.6% of cases, while inclusion of the terms "hidradenitis" or "HS" in the medical record gave positive predictive values of 99.6% (95% CI 98.9-99.9%) and 100% (95% CI 95.8-100%) respectively for identification of chart-verified HS. To minimise any risk of false positives we used only the Read codes for 'hidradenitis' and 'hidradenitis suppurativa' in our search for physician-diagnosed cases in CPRD and used the specific ICD-10 code for 'hidradenitis suppurativa', L73.2, when searching HES data.

Validation in person would have been ideal but is not possible within the ethical parameters of CPRD. Validation using a GP questionnaire to check for the consensus definition of five flexural skin boils would be prone to false negatives. This is because five skin boil consultations may not be recorded as separate entries in the medical record in those with a known diagnosis. For the same

reason, we did not analyse the performance of our algorithms in detecting physician-diagnosed cases. Descriptive statistics comparing controls with proxy cases and diagnosed cases support a conclusion that the proxy group contains mainly HS patients as well as some individuals who do not have HS, while the physician-diagnosed HS group is representative of a HS population. For example, smoking is a known HS risk factor and our physician-diagnosed group contained 51% active smokers, compared with 40% in the proxy group and 24% in controls.

In conclusion, analysis of the 4.3 million UK people with research-standard medical records in CPRD has permitted the most comprehensive estimate of HS prevalence so far obtained, identifying physician-diagnosed and criteria-diagnosed cases in primary care. Prevalence of HS in the UK is 0.77% using the most stringent validation criteria of five flexural skin boils, rising to a maximum possible prevalence of 1.19% if patients with a history of 1-4 flexural skin boils are included. Our results demonstrate that HS is a common condition, highlighting the relative lack of HS clinical trial evidence to date. For example, the recent HS Cochrane review included only 12 randomised controlled trials, involving only 615 participants, one seventh of the number of trials and participants included in the updated Cochrane review of vitiligo, a condition with a similar prevalence.^{30,31} Availability of an accurate prevalence figure is one of the key foundations to drive further research in the field of HS.

References

1. Riis PT, Vinding GR, Ring HC, Jemec GB. Disutility in Patients with Hidradenitis Suppurativa: a Cross-sectional Study Using EuroQoL-5D. *Acta Derm Venereol* 2016; **96**: 222-6.
2. Møller AH, Erntoft S, Vinding GR, Jemec GB. A systematic literature review to compare quality of life in psoriasis with other chronic diseases using EQ-5D-derived utility values. *Patient Relat Outcome Meas* 2015; **6**: 167-77.

- Accepted Article
3. Jemec GB. Clinical practice. Hidradenitis suppurativa. *N Engl J Med* 2012; **366**: 158–164.
 4. Ingram JR. Hidradenitis suppurativa: an update. *Clin Med (Lond)* 2016; **16**: 70-3.
 5. Zouboulis CC, Desai N, Emtestam L, et al. European S1 guideline for the treatment of hidradenitis suppurativa/acne inversa. *J Eur Acad Dermatol Venereol* 2015; **29**: 619-44.
 6. von der Werth JM, Williams HC, Raeburn JA. The clinical genetics of hidradenitis suppurativa revisited. *Br J Dermatol* 2000; **142**: 947–53.
 7. Cosmatos I, Matcho A, Weinstein R, et al. Analysis of patient claims data to determine the prevalence of hidradenitis suppurativa in the United States. *J Am Acad Dermatol* 2013; **68**: 412-9.
 8. Jemec GB. The symptomatology of hidradenitis suppurativa in women. *Br J Dermatol* 1988; **119**: 345-50.
 9. National Institute for Health and Care Excellence. Adalimumab for treating moderate to severe hidradenitis suppurativa: technology appraisal guidance [TA392]. <https://www.nice.org.uk/guidance/ta392>. Accessed 23 February 2017.
 10. Saunte DM, Boer J, Stratigos A, et al. Diagnostic delay in hidradenitis suppurativa is a global problem. *Br J Dermatol* 2015; **173**: 1546-9.
 11. Benchimol EI, Smeeth L, Guttman A, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. *PLoS Med* 2015; **12**: e1001885.
 12. Herrett E, Gallagher AM, Bhaskaran K, et al. Data resource profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol* 2015; **44**: 827–36.
 13. Brameld KJ, Holman CD, Lawrence DM, Hobbs MS. Improved methods for estimating incidence from linked hospital morbidity data. *Int J Epidemiol* 2003; **32**: 617-24.
 14. Revuz JE, Canoui-Poitaine F, Wolkenstein P, et al. Prevalence and factors associated with hidradenitis suppurativa: Results from two case-control studies. *J Am Acad Dermatol* 2008; **59**: 596-601.

- Accepted Article
15. Shahi V, Alikhan A, Vazquez BG, et al. Prevalence of hidradenitis suppurativa: a population-based study in Olmsted County, Minnesota. *Dermatology* 2014; **229**: 154-8.
 16. Vazquez BG, Alikhan A, Weaver AL, et al. Incidence of hidradenitis suppurativa and associated factors: a population-based study of Olmsted County, Minnesota. *J Invest Dermatol* 2013; **133**: 97–103.
 17. Garg A, Kirby JS, Lavian J, et al. Sex- and Age-Adjusted Population Analysis of Prevalence Estimates for Hidradenitis Suppurativa in the United States. *JAMA Dermatol* 2017; doi: 10.1001/jamadermatol.2017.0201. [Epub ahead of print].
 18. Garg A, Lavian J, Lin G, et al. Incidence of hidradenitis suppurativa in the United States: A sex- and age-adjusted population analysis. *J Am Acad Dermatol* 2017; **77**: 118-122.
 19. Tzellos T, Zouboulis CC, Gulliver W, et al. Cardiovascular disease risk factors in patients with hidradenitis suppurativa: a systematic review and meta-analysis of observational studies. *Br J Dermatol* 2015; **173**: 1142-55.
 20. Shalom G, Freud T, Ben Yakov G, et al. Hidradenitis Suppurativa and Inflammatory Bowel Disease: A Cross-Sectional Study of 3,207 Patients. *J Invest Dermatol* 2016; **136**: 1716-8.
 21. Egeberg A, Gislason GH, Hansen PR. Risk of Major Adverse Cardiovascular Events and All-Cause Mortality in Patients With Hidradenitis Suppurativa. *JAMA Dermatol* 2016; **152**: 429-34.
 22. Shavit E, Dreier J, Freud T, et al. Psychiatric comorbidities in 3207 patients with hidradenitis suppurativa. *J Eur Acad Dermatol Venereol* 2015; **29**: 371-6.
 23. Garzorz N, Papanagiotou V, Atenhan A, et al. Pyoderma gangrenosum, acne, psoriasis, arthritis and suppurative hidradenitis (PAPASH)-syndrome: a new entity within the spectrum of autoinflammatory syndromes? *J Eur Acad Dermatol Venereol* 2016; **30**: 141-3.
 24. Shlyankevich J, Chen AJ, Kim GE, Kimball AB. Hidradenitis suppurativa is a systemic disease with substantial comorbidity burden: a chart-verified case-control analysis. *J Am Acad Dermatol* 2014; **71**: 1144-50.

25. Wang B, Yang W, Wen W et al. Gamma-secretase gene mutations in familial acne inversa. *Science* 2010; **330**: 1065.
26. Pink AE, Simpson MA, Desai N, et al. Mutations in the γ -Secretase Genes NCSTN, PSENEN, and PSEN1 Underlie Rare Forms of Hidradenitis Suppurativa (Acne Inversa). *J Invest Dermatol* 2012; **132**: 2459-61.
27. Kelleher RJ 3rd, Shen J. Genetics. Gamma-secretase and human disease. *Science* 2010; **330**: 1055-6.
28. Theut Riis P, Egeberg A, Gislason GH, Jemec GB. Hidradenitis suppurativa patients have no increased risk of Alzheimer's disease. *Br J Dermatol* 2016 doi: 10.1111/bjd.15064.[Epub ahead of print].
29. Kim GE, Shlyankevich J, Kimball AB. The validity of the diagnostic code for hidradenitis suppurativa in an electronic database. *Br J Dermatol* 2014; **171**: 338-42.
30. Ingram JR, Desai N, Kai AC, et al. Interventions for hidradenitis suppurativa. *Cochrane Database Syst Rev* 2015: CD010081.
31. Whitton ME, Pinart M, Batchelor J, et al. Interventions for vitiligo. *Cochrane Database Syst Rev* 2015: CD003263.

Table 1 Description of algorithms used to identify proxy hidradenitis suppurativa cases in CPRD

Sub-algorithm	Description
1a	≥ 5 boils, furuncles, carbuncles or abscesses in flexural skin sites
1b	1-4 boils in flexural sites & ≥ 5 boils in total
1c	≥ 5 boils in unspecified skin sites
2a	“Multiple boils” & ≥ 1 flexural boil
2b	“Multiple boils” & no flexural lesion specified
3	“Drainage or Incision of boil of skin” & ≥ 5 boils in total
4	Surgical excision / laser destruction of flexural skin
5a	≥ 5 short courses of oral flucloxacillin / erythromycin / clarithromycin & ≥ 1 flexural boil, excluding eczema / skin ulcer / cellulitis
5b	≥ 5 short courses of oral flucloxacillin / erythromycin / clarithromycin & ≥ 1 boil in unspecified skin site, excluding eczema / skin ulcer / cellulitis

Table 2 Validation questionnaire sent to the primary care physicians of a subset of proxy HS cases

Question	Domain	Wording	Response
1	Characteristic lesions	Has your patient experienced painful, erythematous papules, nodules or abscesses of the skin (boils)?	Yes/ no
2	Flexural sites	Were these lesions predominantly located in flexural skin sites, specifically the axillae, groins, buttocks, perineum and/or inframammary regions?	Yes/ no
3	Chronicity	Has your patient experienced at least five of these lesions in flexural skin sites since the age of 10?	Yes/ no
4	Scarring	Does your patient have any “rope-like” scars, atrophic (depressed) scars or sinus tracts in affected sites?	Yes/ no/ unknown

Table 3 Sub-algorithm validation from completed questionnaires returned by primary care practitioners

Sub-algorithm	Questionnaires sent	Questionnaires completed	Proxy cases prior to validation	Proportion validated with ≥ 5 flexural boils	Extrapolated number of criteria-diagnosed cases ^a	Proportion validated with 1-4 flexural boils	Extrapolated additional probable cases ^b
1a	10	6	650	0.67	435	0	0
1b	10	6	1908	0.33	629	0.17	325
1c	10	7	3120	0.57	1778	0.29	905
2a	10	3	2083	0.33	687	0.34	708
2b	59	38	27,492	0.13	3573	0.29	7973
3	0	0	0	0	0	0	0
4	10	5	594	0	0	0.20	118
5a	14	10	7050	0.10	705	0.60	4229
5b	53	32	25,993	0.09	2339	0.16	4159
Total	176	107	68,890		10,146		18,417

(a) 'Criteria-diagnosed cases' are defined as those confirmed by their primary care physician to have a history of at least five flexural skin boils.

'Additional probable cases' are defined as those confirmed by their primary care physician to have a history of between one and four flexural skin boils.

Table 4 Case-control study univariable analysis

	Control n (%)	Proxy case n (%)	Physician-diagnosed case n (%)	Proxy cases		Physician-diagnosed cases of hidradenitis suppurativa	
				OR (95% CI)	P-value	OR (95% CI)	P-value
Obesity (BMI>30)	19862 (21.2)	29916 (42.8)	11145 (46.4)	2.77 (2.70, 2.85)	<0.001	3.29 (3.14, 3.45)	<0.001
Type 2 diabetes mellitus	17689 (4.8)	30845 (9.0)	2167 (9.0)	3.32 (3.18, 3.47)	<0.001	3.39 (3.09, 3.71)	<0.001
Smoking status							
Never	48431 (51.6)	28751 (41.2)	7828 (32.6)	1		1	
Past	12671 (13.5)	10523 (15.1)	3260 (13.6)	1.44 (1.39, 1.49)	<0.001	1.93 (1.81, 2.05)	<0.001
Current	22712 (24.2)	27748 (39.7)	12347 (51.4)	2.18 (2.13, 2.24)	<0.001	3.61 (3.44, 3.79)	<0.001
Crohn's disease	708 (0.2)	310 (0.4)	122 (0.5)	2.21 (1.81, 2.69)	<0.001	2.65 (1.89, 3.72)	<0.001
Ulcerative colitis	521 (0.1)	78 (0.1)	42 (0.2)	0.82 (0.61, 1.11)	0.2	1.24 (0.79, 1.94)	0.36
Pilonidal sinus	293 (0.3)	887 (1.3)	444 (1.9)	4.21 (3.62, 4.89)	<0.001	5.61 (4.41, 7.12)	<0.001
Acne vulgaris	5828 (6.2)	7858 (11.3)	2885 (12.0)	2.17 (2.09, 2.27)	<0.001	1.77 (1.66, 1.88)	<0.001
Depression	56050 (15.3)	16968 (24.3)	6138 (26.3)	1.65 (1.60, 1.69)	<0.001	1.69 (1.62, 1.77)	<0.001
Hyperlipidaemia	10871 (3.0)	3344 (4.8)	791 (3.3)	1.59 (1.50, 1.69)	<0.001	1.79 (1.59, 2.02)	<0.001
Hypertension	33457 (9.1)	8596 (12.3)	1882 (7.8)	1.40 (1.35, 1.45)	<0.001	1.42 (1.32, 1.53)	<0.001
Inflammatory arthritis	3000 (0.8)	634 (0.9)	201 (0.8)	1.06 (0.94, 1.18)	0.35	1.06 (0.87, 1.30)	0.54
PCOS	377 (0.4)	252 (0.4)	181 (0.8)	1.12 (0.94, 1.34)	0.22	1.19 (0.96, 1.49)	0.11
Psoriasis	3444 (0.9)	671 (1.0)	247 (1.0)	1.03 (0.92, 1.15)	0.62	1.09 (0.91, 1.31)	0.33
Alzheimer's disease	529 (0.1)	87 (0.1)	16 (0.01)	0.95 (0.71, 1.27)	0.71	1.60 (0.73, 3.53)	0.24

Based on 93,869 control individuals, 69,842 proxy cases including criteria-diagnosed hidradenitis suppurativa (HS) cases and those who did not meet the validation criteria, and 24,027 physician-diagnosed cases of HS. Smoking status was missing from a total of 10,055 records, including 2,820 proxy cases and 592 physician-diagnosed cases. PCOS = polycystic ovary syndrome.

Figure titles and legends

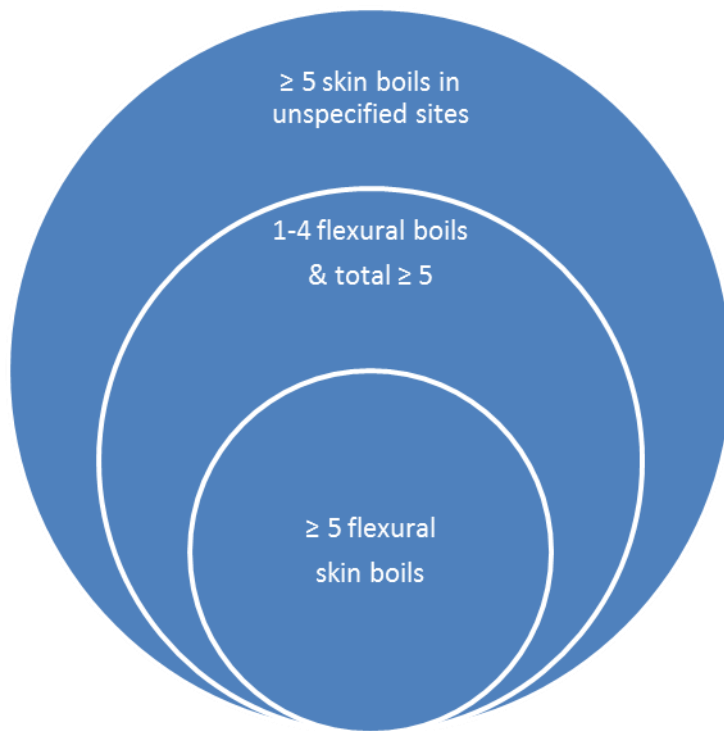
Figure 1: Algorithm hierarchy used to capture unrecognised cases of hidradenitis suppurativa.

Legend: Schematic representation of sub-algorithms used to detect unrecognised hidradenitis suppurativa with differing degrees of stringency.

Figure 2: Demographics of physician-diagnosed hidradenitis suppurativa cases.

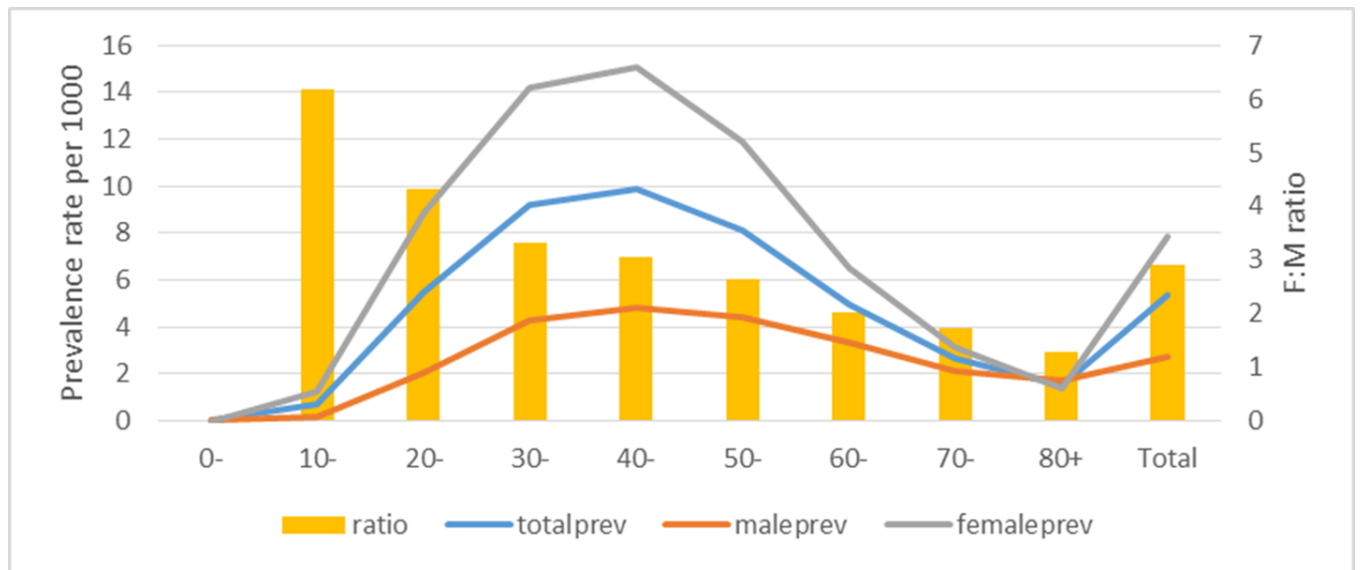
Legend: Line graph of the prevalence rate of physician-diagnosed hidradenitis suppurativa in UK males, females and in total, subdivided by decade of age. Superimposed is a histogram of the female: male ratio for each decade of age.

Figure 1: Algorithm hierarchy used to capture unrecognised cases of hidradenitis suppurativa



Legend: Schematic representation of sub-algorithms used to detect unrecognised hidradenitis suppurativa with differing degrees of stringency.

Figure 2: Demographics of physician-diagnosed hidradenitis suppurativa cases.



Legend: Line graph of the prevalence rate of physician-diagnosed hidradenitis suppurativa in UK males, females and in total, subdivided by decade of age. Superimposed is a histogram of the female: male ratio for each decade of age.