Published in final edited form as:

Br J Dermatol. 2011 March ; 164(3): 602–609. doi:10.1111/j.1365-2133.2010.10134.x.

Validity of The Health Improvement Network (THIN) for the study of psoriasis

N.M. Seminara¹, K. Abuabara^{1,3}, D.B. Shin^{1,2}, S.M. Langan¹, S.E. Kimmel^{2,3}, D. Margolis^{1,2}, A.B. Troxel², and J.M. Gelfand^{1,2}

¹Department of Dermatology, University of Pennsylvania, Philadelphia, PA, United States

²Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania, Philadelphia, PA, United States

³Department of Medicine, University of Pennsylvania, Philadelphia, PA, United States

Abstract

Background—Psoriasis is a common disease frequently studied in large databases. To date the validity of psoriasis information has not been established in The Health Improvement Network (THIN).

Objectives—To investigate the validity of THIN for identifying psoriasis patients and to determine if the database can be used to determine the natural history of disease.

Patients/Methods—First we conducted a cross sectional study to determine if psoriasis prevalence in THIN is similar to expected. Second we created a cohort of 4900 patients, aged 45–65, with a psoriasis diagnostic Read Code and surveyed their GPs to confirm the diagnosis clinically. Third we created models to determine if psoriasis descriptors (extent, severity, duration, and dermatologist confirmation) could be accurately captured from database records.

Results—Psoriasis prevalence was 1.9%, and showed the characteristic age distribution expected. GP questionnaires were received for 4,634 of 4,900 cohort patients (95% response rate), and psoriasis diagnoses were confirmed in 90% of patients. Duration of disease in the database showed substantial agreement with physician query ($\kappa = 0.69$). GPs confirmed that the psoriasis diagnosis was corroborated by a dermatologist in 91% of patients whose database records contained a dermatology referral code associated with a psoriasis code. We achieved good discrimination between patients with and without extensive disease based on the number of psoriasis codes received per year (Area Under Curve, AUC = 0.8).

Conclusions—THIN is a valid data resource for studying psoriasis and can be used to identify characteristics of the disease such as duration and confirmation by a dermatologist.

Keywords

THIN; Validation Study; Psoriasis; Epidemiology

Corresponding Author: Nicole M. Seminara, 399 S. 34th Street, Penn Tower, Suite 1468, Philadelphia, PA 19104, Phone: 860-716-7322, Fax: 215-746-6370, seminara@mail.med.upenn.edu.

Conflicts of Interest: Dr Gelfand has received grants from Amgen, Pfizer, Novartis, and Abbott, and is a consultant for Amgen, Celgene, and Centocor. Dr Margolis is on the data safety monitoring boards for Abbott, Astellas, and Centocor. No other authors have conflicts of interest.

Introduction

Psoriasis is a common, chronic inflammatory disease affecting 1–4% of the adult population^{1–6}. Recently, a variety of studies have used large databases to show that psoriasis patients, especially those with severe disease, are at increased risk of MI, stroke, all-cause mortality, and cardiovascular mortality^{7–15}. Although psoriasis is increasingly studied in administrative and medical records databases, rigorous validation studies to confirm that electronic psoriasis records accurately reflect the patient's clinical state are frequently not performed^{16–18}.

The most extensive psoriasis validation studies were completed in the General Practice Research Database (GPRD). They involved epidemiological confirmation and questionnaires sent to general practitioners (GPs) to demonstrate that psoriasis diagnostic codes accurately reflect the patient's clinical state^{4,5,19–21}.

In recent years a new population-based medical records database, The Health Improvement Network (THIN), has become available. It offers many desirable features including a substantial patient population, patient level laboratory and prescribing information, and a large group of contributing practices that have agreed to answer patient-related queries²². Initial studies in THIN have suggested that it validly captures data on conditions such as peptic ulcer disease, basal cell carcinoma, and death^{23–25}. To our knowledge, studies evaluating the validity of psoriasis in THIN have not been conducted. Therefore, the goals of this study were to validate the accuracy of psoriasis diagnostic codes in THIN, and to determine if disease severity, disease duration, and dermatologist confirmation of psoriasis diagnoses are accurately captured in the database.

Methods²⁶

Study Design

This validation study involved three discrete stages. First, to measure convergent validity (i.e. degree to which two measures of the same construct yield similar findings) we conducted a cross sectional study to determine if psoriasis prevalence in the THIN population is similar to expected. Second, to measure criterion validity (i.e. extent to which a measure agrees with a gold standard) we surveyed GPs to confirm that patients with an electronic psoriasis code in fact had psoriasis clinically^{16,18}. This survey was designed by dermatologists and epidemiologists with the guidance of UK-based general practitioners. Third we attempted to derive information about the patient's psoriasis (i.e. extent, severity, duration, and confirmation of the diagnosis by a dermatologist) from database codes and compared these to GP surveys. This study was approved by the University of Pennsylvania institutional review board and the Cambridgeshire Research Ethics Committee and was funded by the National Heart Lung and Blood Institute of the NIH.

Data Source

THIN is an electronic medical records database that includes more than 7.5 million people, and broadly represents 4.6% of the UK population across more than 400 practices. It contains patient demographics, medical diagnoses, laboratory results, and prescriptions as recorded by GPs, who serve as the primary point of medical contact in the UK. THIN utilizes monetary incentives, quality targets, and training to ensure accurate and complete records which should include information from hospital and specialty care²². It is estimated that approximately half of the practices in THIN are also participating in the GPRD database²⁷.

Study Population and Definition of Psoriasis

The entire THIN population was included in the prevalence study. Patients were considered to have psoriasis if they had ever received an electronic psoriasis Read Code. THIN utilizes more than 100,000 unique hierarchical alpha-numeric Read Codes to capture medical information²². Four researchers, two of whom were dermatologists, identified 26 psoriasis Read Codes by employing a keyword search and then examining affiliated codes in hierarchical proximity. The list of codes along with the percent of patients assigned each code is available online in Table 5.

In order to validate the psoriasis diagnostic codes and test the psoriasis descriptor models (extent, severity etc), we created a cohort of psoriasis patients in THIN and gathered additional information about them via a GP questionnaire. Patients were eligible for the cohort if they were 45–64 years of age; this age range was selected as this cohort will be followed prospectively for cardiovascular outcomes. They were further required to be alive at the time of sampling, have received at least one psoriasis Read Code in the two years prior to sampling, and be registered to a practice with an Additional Information Services (AIS) contract. AIS practices agree to complete questionnaires in exchange for compensation. More than half (55%, n=228) of all THIN practices belonged to AIS at the time of sampling. For these portions of the study, patients were defined as having psoriasis if their GP confirmed the diagnosis.

To ensure data integrity, we calculated the proportion of internally inconsistent survey responses from each practice, as determined by comparing two questions about Body Surface Area (BSA) involved. If more than a quarter of a practice's responses were inconsistent the practice was excluded from the analysis.

Sampling

There are 7,520,293 patients registered in THIN of whom 6,985,418 had at least one recorded visit to their GP. We randomly sampled 5000 of the eligible 5008 patients who met the cohort criteria above. The sample size of 5000 was selected to ensure appropriate power to detect cardiovascular outcomes in subsequent work. A pilot group of 100 randomly selected patients were excluded from this study due to different selection criteria. Between June and September of 2009 a questionnaire was mailed to the remaining 4900 patients' GPs to confirm their psoriasis diagnosis and describe the disease. Practices that did not respond were reminded with follow up phone calls and letters. Survey collection closed in June 2010. The questionnaire is available online.

Statistical Methods

The first stage of the study involved analysing patient demographics and psoriasis prevalence to determine if our findings are consistent with previous epidemiological research. Prevalence was calculated by dividing the number of psoriasis patients in each age group by the total number of THIN patients in that age group. A sensitivity analysis involved excluding patients who had never received a medical code, and thus may have been registered but never seen by a practice.

Subsequently, electronic psoriasis diagnostic codes were validated by comparing them to GP survey responses which were utilised as a gold standard. Basic demographics were summarised for patients with confirmed and refuted diagnoses to determine if there were any systematic differences between the groups. The positive predictive value (PPV, the probability of having psoriasis confirmed by questionnaire given the presence of an electronic psoriasis code) was calculated. We then determined if we could discriminate between patients with and without GP-confirmed psoriasis using Receiver Operating

Characteristic curves (ROC, described below) based upon the number of psoriasis diagnoses and/or treatments. For illustration purposes the PPV, sensitivity (the percent of GP-confirmed psoriasis patients the test captures), specificity (the percent of GP-denied patients the test excludes), and the correct classification rate (probability that the classification result [i.e., psoriasis or no psoriasis] matches the GP designation) are shown for the first three cut points of the model with the highest value for the Area Under the ROC Curve (AUC).

Finally, we sought to determine if attributes of the patient's psoriasis (i.e., extent, severity, duration, and diagnosis corroboration via dermatology consultation) could be determined from the database. THIN does not contain information directly referring to these attributes so potential surrogate markers were investigated.

If the GP confirmed the psoriasis diagnosis, s/he was asked about the BSA the disease typically involves ($\leq 2\%$, 3–10%, or $> 10\%$). Demographics and descriptive data were compared for patients with extensive ($> 10\%$ BSA) and non-extensive disease ($< 10\%$ BSA). Multiple algorithms were developed to predict extensive disease. Models with the highest AUCs are presented, and cut points are shown at the 10th, 25th, 50th, 75th, 90th and 95th percentile of all GP responses for the model with the highest AUC.

An ROC curve is constructed by plotting the true positive rate of the test on the y-axis as a function of the false positive rate on the x-axis for each possible cutoff value of a diagnostic test (e.g., 0 codes, 1 code, 2 codes etc). The AUC is the probability that the test will rank a randomly chosen psoriasis patient with extensive disease higher than a randomly chosen psoriasis patient without extensive disease. A value of 0.5 would indicate that the test has no predictive power and assigns patients at random, while a value of 1.0 would indicate that the test is always correct^{28,29}.

GPs were also asked if the diagnosis was corroborated by a dermatologist. THIN physicians are expected to record hospital referrals as a dated NHS specialty code. We calculated descriptive statistics (PPV, negative predictive value-NPV, sensitivity, and specificity) to determine whether dermatology NHS codes, alone or in conjunction with psoriasis codes, could predict dermatology confirmation.

To determine agreement between duration of disease from the database, as determined by the first psoriasis code (which GPs can backdate to reflect the time of disease onset), and duration of disease according to the GP questionnaire, a Cohen's kappa statistic was calculated. The kappa statistic is a measure of agreement; it ranges from 1 (perfect agreement) to 0 (agreement expected from chance alone). Duration of disease was broken into 10 year increments and linear weights were applied. Ten years was selected because it was the smallest increment that the GP questionnaire responses could be grouped into to allow the same number of years in each category.

Descriptive statistics, and ROC curves were calculated using SAS (Version 9.2, SAS Institute Inc, Cary, NC). ROC curves were graphed using STATA (Version 10, Stata Corporation, College Station, TX). Statistical significance was determined using two-sided p-values at the 0.05 level (95% confidence level). Missing responses on the surveys were excluded from the analysis.

Results

Table 1 describes the prevalence and demographics of psoriasis patients within the THIN database, and compares them to the THIN population as a whole. Prevalence rates are shown overall, by decade, and by sex. Overall prevalence was found to be 1.9% (2.2% in adults 20 years or older) and did not differ substantially by sex. The prevalence is low in the

pediatric population, increases sharply in young adulthood, gradually increases until late adulthood and then decreases in older individuals. In a sensitivity analysis, patients who had never received a medical code were removed from the calculation (534,875 patients). Prevalence rates increased only slightly, with overall prevalence becoming 2.0%. The frequency of various subtypes of psoriasis based on Read Codes is shown in Table 5 (available online). Psoriasis unspecified (55.5%) or NOS (58.7%) was most common (note patients could receive more than one code and therefore the sum is greater than 100%). Variants of psoriasis were coded less frequently and include guttate (7.4%), scalp (6.8%), pustular (2.8%) palmer or plantar (0.8% and 0.5% respectively) and erythrodermic (0.3%).

At the close of the survey collection period we had received 4,634 of the 4,900 surveys yielding a response rate of 95%. One practice (39 surveys, 1% of the sample) was removed from the survey-based analysis based on a high proportion (52%) of internally inconsistent surveys. The demographics of the sampled patients and survey respondents were nearly identical. Amongst those with returned surveys, 51% (n= 2361) of patients were male and the median age was 55 years (IQR: 49.9–60.9).

To validate the electronic psoriasis Read Codes, the questionnaire asked GPs to confirm or refute the psoriasis diagnosis. The diagnosis of psoriasis was confirmed in 90% (n = 4543, 95% CI 89–91%) of patients who met the criteria for entry into our cohort. Table 2A compares patients with confirmed and refuted diagnoses, and describes the best models for predicting an accurate psoriasis diagnosis. Patient demographics did not differ significantly between the confirmed and refuted groups. The model producing the highest AUC was the number of electronic psoriasis diagnostic codes (AUC=0.75). On average, patients with a confirmed diagnosis had 6 psoriasis codes while those with a refuted diagnosis had only 2. Patients whose diagnosis was confirmed also received significantly more treatments specific for psoriasis (8 vs 2); however, incorporating psoriasis treatments did not improve the AUC or PPV of the models.

Table 2B shows descriptive statistics of the best model at several cut points along the ROC. One or more electronic psoriasis codes yields a PPV of 90%. Requiring more than one code increases the PPV to 95% but will exclude 26% of patients with a valid diagnosis.

If the GP confirmed the psoriasis diagnosis s/he was further asked to identify the percent of the patient's body surface area the psoriasis typically involves. In our cohort, 12% (n = 478, 95% CI = 11–13%) had extensive disease covering more than 10% of their body. Table 3A compares the extensive and non-extensive psoriasis patients. Patients with extensive disease were slightly younger and more likely to be male.

We developed models to determine if electronic codes could reliably identify patients with extensive disease. Though the AUCs of the models presented were similar, the model producing the highest AUC was the number of electronic psoriasis codes (treatment or diagnostic) the patient received per year of psoriasis since their prospective data collection began in THIN (marked with a ‘***’ in table 3A and shown in Fig. 1). On average, patients with extensive disease received 16 codes per year while those with less involved disease received only 3.

Table 3B shows attributes of the model at several cut points. Even with the strictest test criteria we were only able to achieve a PPV of 45%. Though this cut point correctly classified 87% of patients, the sensitivity was only 18.9%. Less strict cut points were able to achieve a substantially higher sensitivity at the expense of a lower PPV.

Models were also developed to determine if disease severity (as defined by skin disease that would require a systemic agent to achieve clearance per the GP's opinion) could be

predicted from the database. Results were similar to the results for extensive disease and are not shown.

If the GP confirmed the psoriasis diagnosis s/he was asked if the diagnosis was also corroborated by a dermatologist. In our cohort, 46% ($n = 1,816$, 95% CI 44–47%) of psoriasis patients had their diagnosis corroborated. We tested whether a dermatology NHS code in the database could be used as a surrogate marker to indicate that a dermatologist confirmed the psoriasis diagnosis. Results are shown in Table 4. More than three-quarters (77%) of psoriasis patients with a dermatology NHS code had their diagnosis corroborated, and 75% of those without a dermatology NHS code did not have their diagnosis corroborated. If we further require the NHS code to be entered in conjunction with a psoriasis code the PPV increases to 91%, but only a third of patients with a dermatologist confirmation are captured.

Finally, GPs were queried regarding the number of years the patient has had psoriasis. This response was compared to the number of years since their first psoriasis code in the database. When the duration of disease was broken into 10-year increments the linearly weighted Cohen's kappa was 0.69 (95% CI 0.67–0.71, 0.62 without weighting). The physician survey and database date matched exactly in 76.5% of cases.

Discussion

The data from this study suggest that THIN is a valid data resource for the study of psoriasis. The epidemiology of psoriasis in THIN is similar to findings from previous population-based studies providing a measure of convergent validity^{1–6}. Moreover, the PPV for one or more psoriasis Read Codes was 90%, demonstrating criterion validity. Of note, the PPV of psoriasis codes in THIN is similar to that observed in validation studies of psoriasis in GPRD^{5,19,21}. Moreover, given that THIN and GPRD cover similar populations and have similar methods for capturing electronic medical data, it is likely that the results of this study generalize to GPRD.

In addition to validly identifying patients with psoriasis, the database can yield important information about the disease. Comparison of disease duration between the survey and the database yielded a kappa of 0.69, indicating substantial agreement³⁰. This will be valuable for modeling outcomes based on disease duration, though we should expect some misclassification. Electronic codes in THIN can also be used to determine if the psoriasis diagnosis was confirmed by a dermatologist. For example, if a patient has a dermatology referral code in conjunction with a psoriasis code then the PPV is 91% (95% CI: 88–93%) for the patient having their diagnosis confirmed by a dermatologist. This will be valuable for sensitivity analyses where dermatologist confirmation can serve as the ultimate gold standard for an accurate psoriasis diagnosis. Moreover, we were able to achieve good discrimination ($AUC = 0.80$) between those with and without extensive skin disease based on ROC analysis of number of yearly psoriasis-related codes in THIN; however, the prevalence of extensive psoriasis in the general population is low and therefore the PPVs of various cut points of our model are not high enough to serve as an accurate surrogate marker of extensive involvement. Nevertheless, these data can be used to enrich a cohort with patients that have more severe disease.

Particular strengths of this study include its use of a variety of epidemiological methods to examine psoriasis validity. Furthermore this study is broadly representative of psoriasis patients ages 45–64. With nearly 5000 patients sampled, this is, to our knowledge, the largest validation study ever completed for psoriasis, and demonstrates that information above and beyond what is contained in the database can be collected for large cohorts with

an excellent response rate. We have also demonstrated the ability to use electronic codes to measure psoriasis duration and confirmation by a dermatologist.

Limitations

We required patients to have received a psoriasis code within the last two years for inclusion in the cohort. This may have led to under-representation of patients with mild disease who may not have discussed psoriasis with their GP in the last 2 years, or may not have seen their GP in that timeframe. In addition, our cohort was limited to patients between 45–65 years of age so our findings may not generalize to the entire psoriasis population. Moreover, we did not evaluate the NPV of psoriasis in THIN (i.e., the probability that a patient without a psoriasis Read Code does not have psoriasis) as this disease has a low prevalence at baseline.

In all validation studies the possibility of a tarnished gold standard for measuring criterion validity should be addressed. We compared our outcomes to responses from GP surveys which served as our gold standard. Though our survey instrument was designed under the guidance of expert opinion in the fields of epidemiology, dermatology, and primary care medicine in order to ensure face and content validity, it was not formally tested for psychometric properties such as test-retest reliability and criterion validity. Furthermore, compared to dermatologists, GPs are more prone to error when making diagnoses of skin disease though a previous study suggested that GPs from the UK are reasonably accurate when it comes to diagnosing psoriasis³¹. Moreover, our data suggest that coding algorithms can be utilised to identify patients whose psoriasis diagnosis was confirmed by a dermatologist which may allow for a more stringent case definition.

Conclusion

THIN is a valid data resource for the study of psoriasis and can be used to identify characteristics of the disease such as duration and confirmation by a dermatologist. These newly identified and validated characteristics can be utilised in future studies that evaluate the natural history of psoriasis in population based setting.

What is already known about this topic?

- Automated medical record and administrative databases are increasingly used to study psoriasis
- Studies to demonstrate that electronic codes for psoriasis truly reflect the patients' clinical state have been variably performed
- The most extensive validation studies of psoriasis have been performed in the General Practice Research database

What does this study add?

- This is the largest psoriasis validation study ever completed
- The prevalence of psoriasis in THIN is similar to expected and 90% of patients with a diagnostic code for psoriasis had their disease confirmed
- We have demonstrated the ability to use electronic codes to study the natural history of psoriasis (duration, extent and diagnosis confirmation by a dermatologist) which will be important for modeling disease outcomes

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Sponsor/Grant This study was supported by an R01 grant and a Graduate Research Supplement (NMS) grant from the National Heart, Lung, and Blood Institute of the NIH RO1HL089744 (JMG)

References

1. Kurd SK, Gelfand JM. The prevalence of previously diagnosed and undiagnosed psoriasis in US adults: results from NHANES 2003–2004. *J Am Acad Dermatol*. 2009; 60:218–224. [PubMed: 19022533]
2. Koo J. Population-based epidemiologic study of psoriasis with emphasis on quality of life assessment. *Dermatol Clin*. 1996; 14:485–496. [PubMed: 8818558]
3. Stern RS, Nijsten T, Feldman SR, et al. Psoriasis is common, carries a substantial burden even when not extensive, and is associated with widespread treatment dissatisfaction. *J Invest Dermatol Symp Proc*. 2004; 9:136–139.
4. Gelfand JM, Weinstein R, Porter SB, et al. Prevalence and treatment of psoriasis in the United Kingdom: a population-based study. *Arch Dermatol*. 2005; 141:1537–1541. [PubMed: 16365254]
5. Gelfand JM, Wang X, Qing L, Neimann AL, Weinstein R, Margolis DJ. Epidemiology and Treatment Patterns of Psoriasis in the General Practice Research Database (GPRD). *Pharmacoepidemiol Drug Saf*. 2005; 14:23.
6. Nevitt GJ, Hutchinson PE. Psoriasis in the community: prevalence, severity and patients' beliefs and attitudes towards the disease. *Br J Dermatol*. 1996; 135:533–537. [PubMed: 8915141]
7. Gelfand JM, Neimann AL, Shin DB, et al. Risk of myocardial infarction in patients with psoriasis. *JAMA*. 2006; 296:1735–1741. [PubMed: 17032986]
8. Gelfand JM, Dommasch ED, Shin DB, et al. The risk of stroke in patients with psoriasis. *J Invest Dermatol*. 2009; 129:2411–2418. [PubMed: 19458634]
9. Gelfand JM, Troxel AB, Lewis JD, et al. The risk of mortality in patients with psoriasis: results from a population-based study. *Arch Dermatol*. 2007; 143:1493–1499. [PubMed: 18086997]
10. Mehta NN, Azfar RS, Shin DB, et al. Patients with severe psoriasis are at increased risk of cardiovascular mortality: cohort study using the General Practice Research Database. *Eur Heart J*. 31:1000–1006. [PubMed: 20037179]
11. Cohen AD, Sherf M, Vidavsky L, et al. Association between psoriasis and the metabolic syndrome. A cross-sectional study. *Dermatology*. 2008; 216:152–155. [PubMed: 18216477]
12. Brauchli YB, Jick SS, Miret M, et al. Psoriasis and risk of incident myocardial infarction, stroke or transient ischaemic attack: an inception cohort study with a nested case-control analysis. *Br J Dermatol*. 2009; 160:1048–1056. [PubMed: 19210501]
13. Kimball AB, Robinson D Jr, Wu Y, et al. Cardiovascular disease and risk factors among psoriasis patients in two US healthcare databases, 2001–2002. *Dermatology*. 2008; 217:27–37. [PubMed: 18349542]
14. Prodanovich S, Kirsner RS, Kravetz JD, et al. Association of psoriasis with coronary artery, cerebrovascular, and peripheral vascular diseases and mortality. *Arch Dermatol*. 2009; 145:700–703. [PubMed: 19528427]
15. Abuabara K, Azfar RS, Shin DB, et al. Cause-specific mortality in patients with severe psoriasis: A population-based cohort study in the United Kingdom. *Br J Dermatol*.
16. Strom, BL. *Pharmacoepidemiology*. Fourth Edition edn.. West Sussex, England: John Wiley & Sons, Ltd.; 2005.
17. Gelfand JM, Azfar RS, Mehta NN. Psoriasis and cardiovascular risk: strength in numbers. *J Invest Dermatol*. 130:919–922. [PubMed: 20231829]

18. Rawson NS, D'Arcy C. Assessing the validity of diagnostic information in administrative health care utilization data: experience in Saskatchewan. *Pharmacoepidemiol Drug Saf.* 1998; 7:389–398. [PubMed: 15073968]
19. Neimann AL, Shin DB, Wang X, et al. Prevalence of cardiovascular risk factors in patients with psoriasis. *J Am Acad Dermatol.* 2006; 55:829–835. [PubMed: 17052489]
20. Gelfand JM, Berlin J, Van Voorhees A, et al. Lymphoma rates are low but increased in patients with psoriasis: results from a population-based cohort study in the United Kingdom. *Arch Dermatol.* 2003; 139:1425–1429. [PubMed: 14623702]
21. Huerta C, Rivero E, Rodriguez LA. Incidence and risk factors for psoriasis in the general population. *Arch Dermatol.* 2007; 143:1559–1565. [PubMed: 18087008]
22. EPIC. THIN Data from EPIC: A Guide For Researchers. London: 2008.
23. Margulis AV, Garcia Rodriguez LA, Hernandez-Diaz S. Positive predictive value of computerized medical records for uncomplicated and complicated upper gastrointestinal ulcer. *Pharmacoepidemiol Drug Saf.* 2009; 18:900–909. [PubMed: 19623573]
24. Meal A, Leonardi-Bee J, Smith C, et al. Validation of THIN data for non-melanoma skin cancer. *Qual Prim Care.* 2008; 16:49–52. [PubMed: 18700078]
25. Hall GC. Validation of death and suicide recording on the THIN UK primary care database. *Pharmacoepidemiol Drug Saf.* 2009; 18:120–131. [PubMed: 19109801]
26. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Epidemiology.* 2007; 18:800–804. [PubMed: 18049194]
27. Cai B. An Algorithm To Identify Practices Common to Both the GPRD and THIN Database. *Pharmacoepidemiol Drug Saf.* 2010; 19:S86.
28. Choi BC. Slopes of a receiver operating characteristic curve and likelihood ratios for a diagnostic test. *Am J Epidemiol.* 1998; 148:1127–1132. [PubMed: 9850136]
29. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology.* 1982; 143:29–36. [PubMed: 7063747]
30. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977; 33:159–174. [PubMed: 843571]
31. Basarab T, Munn SE, Jones RR. Diagnostic accuracy and appropriateness of general practitioner referrals to a dermatology out-patient clinic. *Br J Dermatol.* 1996; 135:70–73. [PubMed: 8776362]

Figure 1. ROC curve for prediction of extensive psoriasis based on number of psoriasis diagnostic or treatment codes per year with psoriasis in THIN
 Receiver Operator Characteristic (ROC) curve for discriminating between patients with extensive vs non-extensive disease based upon number of psoriasis related codes (treatment or diagnostic) per year of psoriasis prospectively captured in THIN.

Table 1
Prevalence and epidemiology of psoriasis in the THIN database

Age distribution and demographics of all THIN patients and THIN patients with psoriasis. Psoriasis prevalence by age and gender are listed in the final column.

	No. Patients in the THIN database n=7,520,293	No. psoriasis patients in THIN (psoriasis prevalence) n = 140,607
Population distribution by age		
Overall	7,520,293	140,607 (1.9%)
<10yrs	465,482	411 (0.1%)
10–19yrs	794,782	4,663 (0.6%)
20–29yrs	983,090	14,565 (1.5%)
30–39yrs	1,242,769	21,662 (1.7%)
40–49yrs	1,209,776	25,895 (2.1%)
50–59yrs	798,347	20,583 (2.5%)
60–69yrs	676,197	20,139 (3.0%)
70–79yrs	502,172	15,282 (3.0%)
80–89yrs	449,049	11,655 (2.6%)
>90yrs	398,629	5,752 (1.4%)
Adults (≥20 years)	6,260,029	135,533 (2.2%)
Demographics		
Age - Median (IQR)	42.5 (26.5–61.5)	51.5 (37.5–68.5)
Male	3,592,128	66,972 (1.8%)
Female	3,928,165	73,635 (1.9%)

Table 2

2A: Comparison of patients in our cohort whose psoriasis diagnosis was confirmed vs. refuted by general practitioner questionnaire. Models for discriminating between these groups of patients are shown.

2B: Cut points for model predicting if psoriasis diagnosis will be confirmed or refuted based upon number of psoriasis diagnostic codes.

2A: Predictive characteristics of patients with confirmed vs refuted psoriasis diagnoses				
	Confirmed Psoriasis (n = 4068)	Refuted Psoriasis (n = 475)	AUC	P value
Patient demographics				
Age - Mean (SD)	55.3 (5.8)	55.8 (5.8)	-	0.08
Male	2082 (51%)	232 (49%)	-	0.36
Models for predicting psoriasis confirmation				
Number of psoriasis diagnostic codes – Median (IQR) **	6 (2–14)	2 (2–4)	0.75	<0.01
Number of psoriasis treatment codes consistent with psoriasis (topical & systemic) – Median (IQR)	24 (9–65)	12 (5–25)	0.65	<0.01
Number of psoriasis treatment codes specific for psoriasis treatment (i.e. topical vitamin D, topical retinoids, oral retinoids, tar, dithranol, PUVA) – Median (IQR)	8 (2–28)	2 (0–6)	0.72	<0.01
Number of psoriasis related codes (specific treatment or diagnostic codes) – Median (IQR)	15 (6–43)	4 (2–9)	0.74	<0.01

2B: Predictive value of cut points for number of psoriasis diagnostic codes to identify psoriasis patients with valid diagnoses				
	# of patients who meet the criteria	PPV (% correct classification)	Sensitivity among those with 1+ code (% of confirmed psoriasis patients captured)	Specificity among those with 1+ code (% of pts with refuted psoriasis diagnosis excluded by test)
1+ Code	4543	90% (90%)	100% by definition	N/A
2+ Codes (on different dates)	3175	95% (73%)	74%	67%
3+ Codes (on any date)	3167	95% (73%)	74%	67%

** Model with the highest Area Under the Curve (AUC), elaborated in Table 2B

Table 3

3A: Comparison of patients in our cohort whose psoriasis involves extensive amounts of their body (>10% Body Surface Area) vs those without extensive disease as defined by GP questionnaire. Models for discriminating between these groups of patients are shown.

3B: Cut points for model predicting if psoriasis is extensive or not extensive based upon number of psoriasis related codes (treatment or diagnostic) per year of psoriasis prospectively captured in THIN.

3A: Predictive characteristics of patients with extensive psoriasis ($\geq 10\%$ BSA) vs non-extensive psoriasis ($< 10\%$ BSA)				
	Extensive psoriasis (n = 478)	Non-extensive psoriasis (n = 3421)	AUC	P value
Patient demographics				
Age - Mean (SD)	54.5 (5.9)	55.4 (5.8)	-	<0.01
Male	281 (58.8%)	1,713 (49.9%)	-	<0.01
Models for predicting extensive disease				
Number of psoriasis diagnostic codes – Median (IQR)	16 (8 – 30)	5 (2 – 12)	0.77	<0.01
Number of psoriasis treatment codes consistent with psoriasis (topical & systemic) – Median (IQR)	90 (34–200)	21 (8 – 52)	0.75	<0.01
Number of psoriasis treatment codes specific for psoriasis (i.e. topical vitamin D, topical retinoids, oral retinoids, tar, dithranol, PUVA) – Median (IQR)	46 (14 – 102)	6 (2 – 20)	0.77	<0.01
Number of psoriasis related codes (specific treatment or diagnostic codes) per year in THIN- Median (IQR) **	15.4 (6.4 – 30.8)	3 (1.2–7.6)	0.80	<0.01

3B: Predictive value of cut points for number of psoriasis codes per year to identify patients with extensive disease ($\geq 10\%$ BSA)				
	# of patients who meet the criteria	PPV (% correct classification)	Sensitivity (% of extensive pts captured)	Specificity (% of non-extensive pts excluded)
10 th percentile (0.6 codes per year)	3530	13.3% (21.4%)	98.5%	10.5%
25 th percentile (1.2 codes per year)	3089	15.1% (32.4%)	97.3%	23.3%
50 th percentile (3.6 codes per year)	1976	20.5% (57.8%)	84.7%	54.1%
75 th percentile (10.0 codes per year)	968	32.1% (78.4%)	63.2%	80.5%
90 th percentile (23.8 codes per year)	390	41.0% (85.9%)	33.5%	93.3%
95 th percentile (39.0 codes per year)	199	44.7% (87.2%)	18.6%	96.8%

** Model with the highest AUC, elaborated in Figure 1 and Table 3B

Table 4

Dermatology consultation validation models (n = 3973)

Descriptive statistics for database-based models designed to predict if the patient's psoriasis diagnosis was corroborated by a dermatologist.

	PPV	Correct Classification	NPV	Sensitivity	Specificity
1) NHS specialty codes for dermatology (95% CI)	77% (75–79)	76% (74–77)	75% (73–77)	67% (65–69)	83% (81–84)
2) NHS specialty code in 120 days prior to a psoriasis code (95% CI)	88% (86–90)	74% (73–75)	69% (67–71)	50% (48–52)	94% (93–95)
3) NHS specialty codes in conjunction with psoriasis code (95% CI)	91% (88–93)	68% (67–70)	63% (62–65)	33% (31–36)	97% (96–98)