



Published in final edited form as:

*Pharmacoepidemiol Drug Saf.* 2011 September ; 20(9): 948–955. doi:10.1002/pds.2191.

## Temporal and Within Practice Variability in The Health Improvement Network

Kevin Haynes, PharmD, MSCE<sup>1,2,3</sup>, Warren B. Bilker, PhD<sup>1,2,3</sup>, Tom R. TenHave, PhD<sup>1,2</sup>, Brian L. Strom, MD, MPH<sup>1,2,3,4</sup>, and James D. Lewis, MD, MSCE<sup>1,2,3,4</sup>

<sup>1</sup>Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine

<sup>2</sup>Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania School of Medicine

<sup>3</sup>Center for Education and Research in Therapeutics, University of Pennsylvania School of Medicine

<sup>4</sup>Department of Medicine, University of Pennsylvania School of Medicine

### Abstract

**Background**—The Health Improvement Network (THIN) database is a primary care electronic medical record database in the United Kingdom designed for pharmacoepidemiologic research. Matching on practice and calendar year is often used to account for secular trends in time and differences across practices. However, little is known about the consistency within practices across observation years and among practices within a given year, in THIN or other large medical record databases.

**Methods**—We analyzed mortality rates, cancer incidence rates, prescribing rates, and encounter rates across 415 practices from 2000 to 2007 using a practice-year as the unit of observation in separate random and fixed effects longitudinal Poisson regression models. Adjusted models accounted for aggregate practice-level characteristics (smoking, obesity, age, and Vision software experience).

**Results**—In adjusted models, subsequent calendar years were associated with lower reported mortality rates, increasing cancer reporting rates, increasing prescriptions per patient, and decreasing encounters per patient, with a corresponding linear trend ( $p < 0.001$  for all analyses). For calendar year 2007, the ratio of the 75th percentile to the 25th percentile for crude rate of cancer, mortality, prescriptions, and encounters was 1.63, 1.63, 1.45, and 1.42, respectively. Adjusting for practice characteristics reduced the among-practice variation by approximately 40%.

**Conclusions**—THIN data are characterized by secular trends and among-practice variation, both of which should be considered in the design of pharmacoepidemiology studies. Whether these are trends in data quality or true secular trends could not be definitively differentiated.

### Keywords

incidence; epidemiology; databases

---

Corresponding Author: Kevin Haynes, Pharm.D., M.S.C.E., University of Pennsylvania School of Medicine, 729 Blockley Hall, 423 Guardian Dr., Philadelphia, PA 19104-6021, khaynes@mail.med.upenn.edu, Phone: (215) 519-0676, Fax: (215) 573-5315.

**Conflicts of interest:** None of the other authors have any potential conflicts of interest to report.

## Introduction

Medical record databases provide a valuable resource to conduct pharmacoepidemiology studies.<sup>1,2</sup> These resources provide large sample sizes, are relatively inexpensive to use, and results are often widely generalizable. The Health Improvement Network (THIN) database is a primary care electronic medical record database in the United Kingdom designed to be used for pharmacoepidemiologic research.<sup>3,4</sup>

Prior to large-scale use of a data set for research, it is important to assess the quality of the data. Quality in electronic medical record and administrative databases can be conceptualized in terms of the accuracy of the recorded diagnoses and prescriptions (the positive predictive value), completeness of the recorded data (sensitivity), comprehensiveness (what is and is not included), and consistency (whether all data are of the same quality). We and others have previously documented the accuracy of recorded diagnoses for a variety of conditions within THIN.<sup>5-8</sup> We and others have also examined their completeness by comparing recorded rates of death and cancer to that expected based on registry data.<sup>9-11</sup> We have previously demonstrated evidence of consistency of the data by documenting near identical results of epidemiologic analyses of common disease-disease and drug-disease associations across two different categories of THIN practices, those that had previously participated in EPIC's version of the General Practice Database (GPRD) and those that had not.<sup>12</sup>

Because of concerns regarding temporal trends and variability among practices from both practice demographics and practitioner recording of medical events, many investigators using electronic health records such as THIN elect to adjust or match on calendar year and practice.<sup>11,13,14</sup> Alternative approaches which have been used to account for practice variation include random effect models and robust standard errors.<sup>15</sup> However, there are limited data to inform investigators planning studies within THIN about the magnitude of variation in rates of common outcomes across practices and time. In this study, we further assessed the degree of temporal and among-practice variation in reporting of mortality rates, cancer incidence rates, prescribing patterns, and general practitioner (GP) encounters.

## Methods

### Data Source

The THIN database used in this analysis contains data on over 7.5 million patients from over 415 GP offices. The practices participating in THIN receive training on entering quality data and feedback reports. THIN practices use the InPractice Systems Vision software from which THIN generates four types of data files that are available to researchers: patient (with demographic information), therapy (all prescriptions from the GP), medical (signs, symptoms, diagnoses, and procedures), and additional health data (preventative medicine, smoking status, physical examination findings, and laboratory results). Additionally, information on the practices includes the year when the practice started using Vision and the year when the practice reached an acceptable mortality reporting (AMR: a quality indicator for person-time that defines the earliest period of optimal mortality reporting for each practice, developed by the administrators of THIN).<sup>10</sup>

### Cohort Selection Criteria

All patients receiving care from a THIN GP during the period from January 1, 2000 to December 31, 2008 were potentially eligible for inclusion in the study. Patients were excluded if their registration status was something other than an acceptable record (e.g., not permanently registered; out of sequence year of birth or registration date; missing or invalid transfer out date; not male or female).

The start of follow-up for each patient for the mortality, prescription, and encounter analyses was the later of the date of registration of the patient with the practice or January 1, 2000. In the cancer analysis, to avoid categorizing prevalent cancers as incident, we used a start of follow-up date as the later of one year after the registration of the patient with the practice or January 1, 2000. The rationale for this modification in the cancer analysis is that within the GPRD, a similar database, reported cancer incidence rates are highest in the first three months after registration and these rates decline to baseline over the first year of patient specific follow-up.<sup>16</sup>

### Outcome Definitions

We computed practice-specific rates of mortality and cancer for each study year. The denominator for these rates was a count of all patients with active registration time at the practice-year of observation. We did not require patients to have been enrolled for the full year since this would have prevented our analysis of death. Mortality was defined using: 1) the death date in the patient demographic file with a registration status of death, 2) the transfer out date if the patient was known to have died according to registration status but was missing a specific death date and had no records 90 days after the transfer out date, or 3) the date of a Read diagnostic code indicating death with no records 90 days after death diagnosis. Incident cancer was identified as the first occurrence of any Read code selected to be consistent with the diagnoses found in the UK cancer registry, specifically the cancer codes and cancer morphology codes excluding non-melanoma skin cancer.<sup>9</sup>

For each year, we also computed practice-specific prescribing and encounter rates. A practice's average number of prescriptions per patient per year was defined as the number of total prescriptions prescribed (new authorizations or refill authorizations) for the patients for a given practice divided by the number of patients registered with the practice in a given study year. A practice's average number of medical encounters per patient per year was defined as the sum across all patients in a practice of the number of days any patient had a clinical encounter divided by the number of patients registered with the practice in a given study year. A clinical encounter was defined, similar to analysis of consultation rates conducted by the Information Centre,<sup>17</sup> as an entry in the medical file or an entry in the additional health data file consistent with a location code of being seen by a practitioner (GP or practice nurse). Code lists are available upon request.

### Statistical Methods

The unit of analysis was a practice-year, defined as the data from one practice over one calendar year. We excluded practice-year observations if the practice did not have 2 prior years of post-Vision software implementation, the practice had not reached acceptable mortality reporting, the practice contributed only one practice-year of observation, or the practice stopped contributing data. A sensitivity analysis included the practice-years that were excluded from the primary analysis.

Separate statistical analyses were conducted for each outcome: incidence rate of mortality, incidence rate of cancer, average number of prescriptions per patient, and average number of encounters per patient. Variability of rates across practice and calendar year were evaluated with longitudinal Poisson regression adjusted for clustering by practice with either random or fixed effects models,<sup>18,19</sup> depending on the purpose of the analysis. Random effects models used only random intercept. For most analyses, calendar year was treated as a categorical fixed effect variable. We compared this model to the model with a fixed effects linear trend for calendar year to assess presence of curvilinear trends for calendar year. For analyses of change in rates across years, practice was treated as a fixed effect to control for potential confounding of the calendar year-outcome relationship by practice.<sup>18</sup> Accordingly,

the overall test of significance for the hypothesis of among-practice variability was derived from a test of the variance component of the random effect for practice model with a chi-square test with  $\frac{1}{2}$  degree of freedom. Under the random effects model, the empirical Bayes estimates of the practice effects are interpreted relative to the reference practice, which was selected as the practice with a crude rate equal to the mean crude rate among all of the practices for all study years (2000–2008). For a descriptive evaluation of practice variation, we also estimated the magnitude of among-practice variation by examining the ratio of the 75<sup>th</sup> percentile to the 25<sup>th</sup> percentile practice for each outcome in each year.

Although unadjusted analyses addressed our primary aims to determine variability over time and among practices, adjusted analyses were used to assess what proportion of the among-practice variability was explained by selected covariates that describe practice characteristics. The reduction in the variance component for the practice random effect due to the inclusion of practice-level covariates related to outcome was used to estimate the degree of reduction in variance among practices that resulted from accounting for practice level covariates. Potential covariates as a proxy for burden of disease in each practice-year of observation included average age of the practice's patients, proportion of the practice's patients that are female, proportion of newly registered patients, Townsend social deprivation score, proportion of the practice's patients documented to be smokers, and the proportion of the patients with a BMI greater than 30. We also examined covariates that likely reflect the practice's recording patterns (proportion of the practice's patients with documented smoking status and the proportion of patients with a BMI recorded). We also included interaction terms for smoking status (proportion of current smokers) and for BMI status (BMI greater than 30). The interaction term allows for differential effect of the proportion of patients who smoke or are obese in practices with high and low rates of recording smoking and weight, respectively. We also adjusted for years of Vision experience as a proxy for GP practice experience with the information system. We examined the effect on the variance estimate of including each confounder variable separately and of including all potential confounders simultaneously to capture any joint confounding that was not appreciated when only a single variable was included.

Because there was little variation in the practice-specific proportion of female sex across years or among the practices, this variable was not included in the adjusted models. All other potential covariates were treated as fixed effects and included in the final model. Under the models we fitted, this approach may have included covariates not related to the factor of interest (e.g., year or practice) and thus did not satisfy the definition of a confounder.

All statistical tests were performed in STATA 11. The protocol was approved by the Institutional Review Board of the University of Pennsylvania.

## Results

There were 8,708,499 patients in the THIN database, from which we excluded 1,147,987 patients because the quality of their registration status did not meet the criteria for this study. We began with 464 practices over a 9-year period from 2000 to 2009, for a total of 4,176 practice-years. We excluded observations if the practice did not have 2 years of Vision experience (1137 practice-years), had not reached their acceptable mortality reporting date (22 practice-years), stopped contributing data (70 practice-years), or if the practice contributed only one practice-year of observation (10 practice-years). Thus, in the primary analysis, we had 2,937 practice-year observations in 431 practices, with a median of 8 practice-years per practice. The median follow-up time for an individual patient in the study period from 2000–2009 is 5.2 years (interquartile range 2–9 years).

### Relation of calendar year to mortality, cancer, encounter, and prescription rates

The number of deaths per practice per year ranged from 2 to 287, number of cancers per practice per year from 0 to 141, average number of prescriptions per patient per year from 3 to 31, and average number of encounters per patient per year from 0.3 to 12. In the unadjusted models, calendar year treated as a fixed effect categorical variable was associated with lower reported mortality rates (p-value for linear trend:  $p < 0.001$ ), increasing cancer-reporting rates (p-value for linear trend:  $p < 0.001$ ), increasing prescription rates (p-value for linear trend:  $p < 0.001$ ), and increasing visit rates (p-value for linear trend:  $p < 0.001$ ) (Table 1 and Figure 1). We fit a model with an indicator variable for years prior to 2005 vs. 2005 and later which yielded significant associations for mortality (IRR 0.93, 95% CI 0.92, 0.94) and cancer (IRR 1.09, 95% CI 1.08, 1.11). Visually inspecting the data for mortality and cancer incidence, it appears that the greatest changes happened in the period between 2003 and 2006. In contrast, there was a strong linear trend between rate and year in prescribing and encounters across the entire study period with correlation coefficients of 0.99 and 0.97, respectively.

### Among-practice variability in mortality, cancer, encounter, and prescription rates

Among-practice variability was assessed with the variance component of the random intercept for practice, with covariates and year treated as categorical, testing the hypothesis that the variance component was greater than zero. For all outcomes, the variability was greater than zero in unadjusted models ( $p < 0.001$ ). Adjusting for all specified practice-level covariates improved model fit relative to the same random intercept model but without the practice-level covariates, based on the likelihood ratio test ( $p < 0.001$ ). Age of the patients in the practice was the strongest confounder (Table 2). Adjusting for all practice-level covariates reduced the variance component for practice by more than 40% for both mortality and cancer and approximately by 25% for prescriptions, suggesting that some of the among-practice variability could be accounted for by the non-calendar year covariates (Table 2). There was little change in the variance component for practice with encounter as the outcome whether we included only age or all specified covariates (Table 2). In an additional model including only practices with data on the Townsend score, this covariate did not appear to influence among-practice variability (Table 2).

To further describe the different outcomes with respect to the variance among practices, we examined the ratio of the 75<sup>th</sup> percentile practice to the 25<sup>th</sup> percentile practice for each outcome. For calendar year 2007, the ratio of the 75<sup>th</sup> percentile to the 25<sup>th</sup> percentile for crude rate of mortality, cancer, prescriptions, and encounters was 1.63, 1.63, 1.45, and 1.42, respectively. Thus, practices at the 75<sup>th</sup> percentile practice had 63 percent higher rates of mortality and cancer, and over 40% higher rates of prescribing and patient encounters than the 25<sup>th</sup> percentile practice. Figure 1 demonstrates that the interquartile range for each outcome is stable over the study period, confirming the finding that the variance among practices as measured by variance component (Table 2) did not change appreciably during the study period.

The overall fit of models treating practice as a random effect to the models with practice treated as a fixed effect was not different for any of the outcomes ( $p = 1.00$ ). Given that treating practice as a fixed effect controls for all practice-level unmeasured confounders whereas treating practice as random does not,<sup>18</sup> the lack of any difference in model fit between these two models suggests that there is little unmeasured practice-level confounding for the associations that we studied.

The sensitivity analysis includes all practice-year observations following Vision conversion with no exclusion for AMR date. The results were unchanged from the primary analysis (data not shown).

## Discussion

We found subsequent calendar years were associated with lower reported mortality rates, increasing cancer reporting rates, increasing prescriptions per patient, and increasing encounters per patient. Our findings also demonstrated evidence of among-practice variability within a given calendar year for the same measures, and that some of our observed practice-level covariates may contribute to the variability among practices. In a sensitivity analysis, inclusion of practice-year observations from before the practice reached their AMR and practice-year observations with less than 2 years of Vision experience did not appreciably alter the results. The implication of these results is that, when using THIN data, investigators may need to consider secular trends and among-practice variability in pharmacoepidemiology analyses. While bias resulting from failure to account for among practice variability may often be small, the among practice variability can be accounted for through a variety of methods, such as matching or through the use of random effect models, as presented in this paper.

We observed that the variability among practices was reduced by inclusion of covariates describing the practice characteristics. Although not previously studied in THIN, other investigators have observed similar findings in similar populations. For example, using the GPRD, patient level morbidity as measured by the Johns Hopkins adjusted clinical group case mix system explained 57% of the variation in prescribing among practices,<sup>20</sup> 30% of the variation in referral practices,<sup>21</sup> and 2% of the variation in home visits.<sup>22</sup> Importantly, even adjustment at the patient level as done in these prior studies failed to fully account for among-practice variability. Variability in disease-specific outcomes has also been noted in a UK study of 78 participating practices in an audit-based educational program in primary care.<sup>23</sup>

We and other investigators have also observed secular trends for outcomes within primary care practices. The current findings for cancer rates are similar to those of our previous study.<sup>9</sup> Our data on secular trends in mortality, prescribing, and encounter rates are all generally consistent with public health data from the United Kingdom indicating declining rates of mortality and increasing rates of prescribing and encounters over the same time period.<sup>17,24,25</sup> As shown before for mortality<sup>10</sup> and cancer<sup>9</sup> our analysis showed an observed difference before and after 2004. Although not the specific purpose of this study, this finding supports the generalizability of THIN data to the broader United Kingdom population.

There are several potential limitations to this study. We did not require patients to have been enrolled for the full year, as this would have prevented our analysis of death (i.e., registration ends when the patient dies). One hypothesis for the observed among-practice variability is that practices with high turnover rates have fewer encounters per patient per year and fewer prescriptions per patient per year. To test this, we included a variable for the proportion of patients that were new to the practice each year. In univariate analysis, inclusion of this variable reduced the variance component for prescriptions per year by 4.1% and for encounters per year by 1.2%. However, adding this variable to the fully adjusted models had relatively little effect on the variance estimate (data not shown).

As one of our primary aims was to assess among-practice variability, the unadjusted analyses provide important insight to the magnitude of this variability. We used adjusted

analyses to determine how much of the variability could be explained as a result of confounding. We selected a limited number of practice characteristics for inclusion in the adjusted models, although of course there could be other practice characteristics that were not captured. Nonetheless, after adjusting for covariates we observed a reduction in among-practice variability of approximately 40%, most of which was due to adjustment for mean age of the patients in the practice. Other practice-level factors that may influence rates of encounters and prescribing such as GP knowledge and experience and patient expectations are not easily captured in the electronic health records of THIN, and as such were not adjusted for.<sup>26–28</sup> In addition, we did not examine patient level factors, such as measuring burden of disease by examining established comorbidity indices. Rather, our intent was to identify practice-level factors that are readily summarized across the practice-year observations. As such, it is important to interpret the clinical implications of these data with the appropriate cautions regarding the potential for residual confounding.

A challenge in pharmacoepidemiology research is to quantify the magnitude of among-practice variability in a manner that is easily interpretable. To facilitate this, we calculated the ratio of the 75<sup>th</sup> to 25<sup>th</sup> percentile observations. This metric demonstrated that there was approximately a 30% to 60% difference among the high and low rates for these measures which was generally stable across calendar periods for most of the outcomes we examined. Examining trends in among-practice variability based on the estimated random effect variance provides the same conclusion (Table 2).

We hypothesized that among-practice variation would decrease over time as practices became more accustomed to using Vision software and benefited from ongoing clinical quality improvement programs such as the implementation of Quality Outcomes Framework (QoF) (NHS Information Centre. Quality and outcomes framework [www.qof.ic.nhs.uk](http://www.qof.ic.nhs.uk)) measures. For example, practice recording patterns may become more uniform as the GPs familiarize themselves with the software management system, receive feedback reports, and respond to changes in the QoF. That we did not observe these trends could be in part attributable to our design. We limited our analysis to practices with two years of Vision software experience to ensure that all practices had ample time to become familiar with using the software. We also excluded practice-years prior to the practice specific AMR date. However, neither of these is likely to have contributed substantially to these results as including these practice-year observations in the sensitivity analysis did not appreciably change our results.

Likewise, some of the variation among practices could reflect variable data quality. We observed a limited number of practices having high or low rates of events under study, denoted as outliers on the box plots in Figure 1. THIN provides research files for researchers to develop suitable practice level criteria to select only practices with the most complete data on selected variables of interest. Presumably, matching on practice minimizes some of the bias that could result from variable data quality, although there was little evidence of such bias, as we did not observe improved model fit with inclusion of practice as a fixed effect for any of our outcomes, which controls for such bias at the practice level.

Another potential limitation to our study is that coding schemes change over time.<sup>29</sup> This could affect the analysis of changes over time. This would not be expected to affect the analysis of among-practice variability as these changes would affect all practices.

Although size is a major strength of studies in THIN and other large databases, it also is a potential limitation for interpreting test results. Because of the large size of the data, some of our observations could be statistically significant, while not being clinically significant. For

example, we observe significant variation among practices, yet as noted above, adjustment for practice did not improve the model fit at least for the outcomes we examined.

To what extent these results extrapolate to other primary care databases, including those inside and outside of the United Kingdom, or to administrative claims data is unknown. Similarly, whether these data are generalizable to cohorts with specific diseases or exposures remains to be determined. Additional study of other primary care databases and other outcomes will be important to ensure consistency across practices and calendar time.

In conclusion, THIN data are characterized by significant secular trends and among-practice variation that should be considered in the design of pharmacoepidemiology studies. Empiric data documenting the impact of matching on practice and/or calendar year or implementing different statistical methods to account for this would be useful since there are often studies where matching results in considerable logistic challenges or reduction in sample size. Until such data are available, matching and/or adjusting for practice and calendar time remains a prudent approach in pharmacoepidemiology studies.

## Acknowledgments

This study was supported in part by an Agency for Healthcare Research and Quality (AHRQ) Centers for Education and Research on Therapeutics cooperative agreement (grant #HS10399), an NIH Clinical and Translational Science Award (grant #UL1-RR024134), and by a grant from CSD EPIC in London, United Kingdom. The authors designed and implemented the study. CSD EPIC in London, United Kingdom was allowed to provide non-binding comments on the study design and manuscript. The final opinions expressed are those of the authors. Dr. Lewis has served as an unpaid member of an advisory board for CSD THIN. The authors would like to thank Andrew Maguire MSc BSc FSS, Mary Thompson PhD, BSc, Charles Wentworth III, and Gillian Hall Ph.D. for helpful discussions.

## List of abbreviations and acronyms

<b>AMR</b>	Acceptable Mortality Reporting
<b>GPRD</b>	General Practice Research Database
<b>GP</b>	General Practitioner
<b>SIR</b>	Standardized Incidence Ratios
<b>THIN</b>	The Health Improvement Network
<b>UK</b>	United Kingdom

## References

1. Tannen RL, Weiner MG, Xie D. Replicated studies of two randomized trials of angiotensin-converting enzyme inhibitors: further empiric validation of the 'prior event rate ratio' to adjust for unmeasured confounding by indication. *Pharmacoepidemiol Drug Saf.* 2008; 17(7):671–85. [PubMed: 18327852]
2. Weiner MG, Xie D, Tannen RL. Replication of the Scandinavian Simvastatin Survival Study using a primary care medical record database prompted exploration of a new method to address unmeasured confounding. *Pharmacoepidemiol Drug Saf.* 2008; 17(7):661–70. [PubMed: 18327857]
3. Marston L, Carpenter JR, Walters KR, Morris RW, Nazareth I, Petersen I. Issues in multiple imputation of missing data for large general practice clinical databases. *Pharmacoepidemiol Drug Saf.*
4. Bourke A, Dattani H, Robinson M. Feasibility study and methodology to create a quality-evaluated database of primary care data. *Inform Prim Care.* 2004; 12(3):171–7. [PubMed: 15606990]
5. Hall GC. Validation of death and suicide recording on the THIN UK primary care database. *Pharmacoepidemiol Drug Saf.* 2009; 18(2):120–31. [PubMed: 19109801]



6. Lo Re V 3rd, Haynes K, Forde KA, Localio AR, Schinnar R, Lewis JD. Validity of The Health Improvement Network (THIN) for epidemiologic studies of hepatitis C virus infection. *Pharmacoepidemiol Drug Saf.* 2009; 18(9):807–14. [PubMed: 19551699]
7. Ruigomez A, Martin-Merino E, Rodriguez LA. Validation of ischemic cerebrovascular diagnoses in the health improvement network (THIN). *Pharmacoepidemiol Drug Saf.*
8. 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(10):900–9. [PubMed: 19623573]
9. Haynes K, Forde KA, Schinnar R, Wong P, Strom BL, Lewis JD. Cancer incidence in The Health Improvement Network. *Pharmacoepidemiol Drug Saf.* 2009; 18(8):730–6. [PubMed: 19479713]
10. Maguire A, Blak BT, Thompson M. The importance of defining periods of complete mortality reporting for research using automated data from primary care. *Pharmacoepidemiol Drug Saf.* 2009; 18(1):76–83. [PubMed: 19065600]
11. Bhayat F, Das-Gupta E, Smith C, McKeever T, Hubbard R. The incidence of and mortality from leukaemias in the UK: a general population-based study. *BMC Cancer.* 2009; 9:252. [PubMed: 19630999]
12. Lewis JD, Brensinger C. Agreement between GPRD smoking data: a survey of general practitioners and a population-based survey. *Pharmacoepidemiol Drug Saf.* 2004; 13(7):437–41. [PubMed: 15269927]
13. Myles PR, Hubbard RB, McKeever TM, Pogson Z, Smith CJ, Gibson JE. Risk of community-acquired pneumonia and the use of statins, ace inhibitors and gastric acid suppressants: a population-based case-control study. *Pharmacoepidemiol Drug Saf.* 2009; 18(4):269–75. [PubMed: 19235776]
14. Bhayat F, Das-Gupta E, Smith C, Hubbard R. NSAID use and risk of leukaemia: a population-based case-control study. *Pharmacoepidemiol Drug Saf.* 2009; 18(9):833–6. [PubMed: 19551758]
15. Rait G, Walters K, Griffin M, Buszewicz M, Petersen I, Nazareth I. Recent trends in the incidence of recorded depression in primary care. *Br J Psychiatry.* 2009; 195(6):520–4. [PubMed: 19949202]
16. Lewis JD, Bilker WB, Weinstein RB, Strom BL. The relationship between time since registration and measured incidence rates in the General Practice Research Database. *Pharmacoepidemiol Drug Saf.* 2005; 14(7):443–51. [PubMed: 15898131]
17. Hippisley-Cox, J.; Vinogradova, Y. Trends in Consultation Rates in General Practice 1995 to 2008: Analysis of the QResearchR database. The Health and Social Care Information Centre; 2009.
18. Gardiner JC, Luo Z, Roman LA. Fixed effects, random effects and GEE: what are the differences? *Stat Med.* 2009; 28(2):221–39. [PubMed: 19012297]
19. Gibbons RD, Hur K, Bhaumik DK, Bell CC. Profiling of county-level foster care placements using random-effects Poisson regression models. *Health Serv Outcomes Res Method.* 2007; 7:97–108.
20. Omar RZ, O'Sullivan C, Petersen I, Islam A, Majeed A. A model based on age, sex, and morbidity to explain variation in UK general practice prescribing: cohort study. *BMJ.* 2008; 337:a238. [PubMed: 18625598]
21. Sullivan CO, Omar RZ, Ambler G, Majeed A. Case-mix and variation in specialist referrals in general practice. *Br J Gen Pract.* 2005; 55(516):529–33. [PubMed: 16004738]
22. Sullivan CO, Omar RZ, Forrest CB, Majeed A. Adjusting for case mix and social class in examining variation in home visits between practices. *Fam Pract.* 2004; 21(4):355–63. [PubMed: 15249522]
23. de Lusignan S, Chan T, Wood O, Hague N, Valentin T, Van Vlymen J. Quality and variability of osteoporosis data in general practice computer records: implications for disease registers. *Public Health.* 2005; 119(9):771–80. [PubMed: 15893348]
24. Dunnell K. Ageing and mortality in the UK--national statistician's annual article on the population. *Popul Trends.* 2008; (134):6–23. [PubMed: 19172922]
25. The NHS Information Centre PSU. Prescriptions Dispensed in the Community Statistics for 1998 to 2008: England. The Health and Social Care Information Centre; 2009.

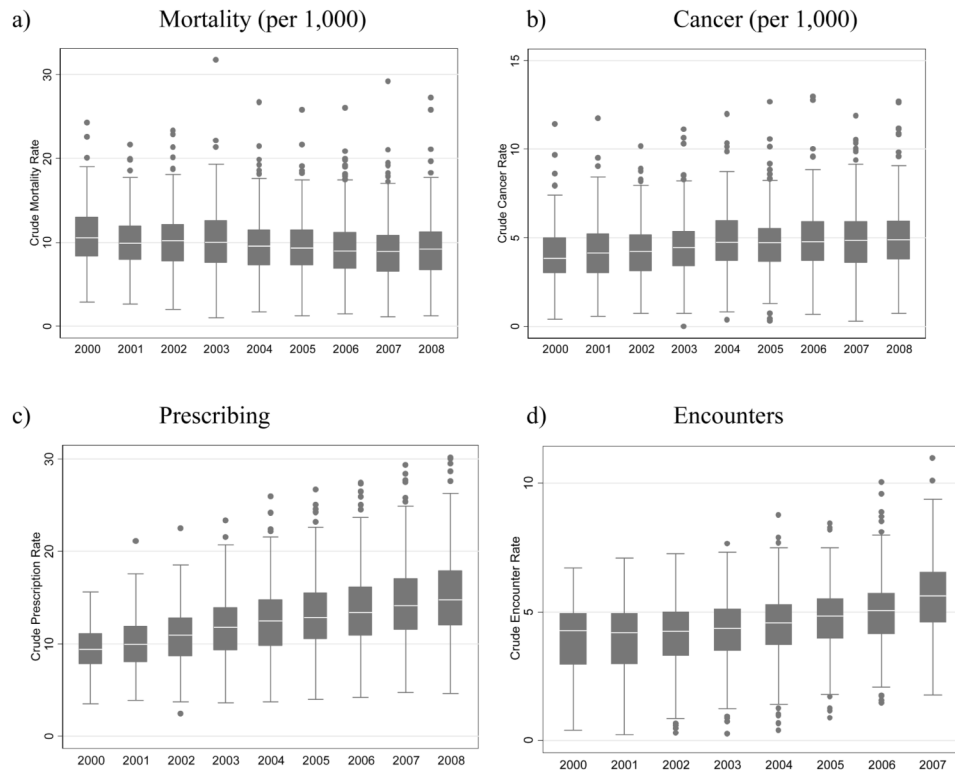
26. Watkins C, Harvey I, Carthy P, Moore L, Robinson E, Brawn R. Attitudes and behaviour of general practitioners and their prescribing costs: a national cross sectional survey. *Qual Saf Health Care*. 2003; 12(1):29–34. [PubMed: 12571342]
27. Ashworth M, Lloyd D, Smith RS, Wagner A, Rowlands G. Social deprivation and statin prescribing: a cross-sectional analysis using data from the new UK general practitioner 'Quality and Outcomes Framework'. *J Public Health (Oxf)*. 2007; 29(1):40–7. [PubMed: 17071815]
28. Britten N, Ukoumunne O. The influence of patients' hopes of receiving a prescription on doctors' perceptions and the decision to prescribe: a questionnaire survey. *BMJ*. 1997; 315(7121):1506–10. [PubMed: 9420493]
29. Gallagher AM, Thomas JM, Hamilton WT, White PD. Incidence of fatigue symptoms and diagnoses presenting in UK primary care from 1990 to 2001. *J R Soc Med*. 2004; 97(12):571–5. [PubMed: 15574853]

### Key Points

THIN is one of several electronic medical record databases from general practices within the United Kingdom that use Vision software. These databases are widely employed in pharmacoepidemiology research.

Within the THIN practices, there have been significant increases in rates of cancer, prescriptions per patient and encounters per patient over the last decade. During the same time period, there were significant declines in mortality rates.

There is significant variability among practices within THIN in terms of mortality, cancer, medication prescribing, and encounter rates of which 40% or less is explained by practice level covariates such as recording of body mass index and smoking and the proportion of patients in a practice with obesity and who are smokers.



**Figure 1.** Practice variability in mortality, cancer, prescribing, and encounter rates over the study period.

**Table 1**

Temporal trends for mortality, cancer, prescriptions, and encounter rates

Unadjusted Analysis				
Incident rate ratio				
	Mortality	Cancer	Prescriptions	Encounters
2000	Reference	Reference	Reference	Reference
2001	0.97 (0.95, 1.00)	1.09 (1.05, 1.13)	1.07 (1.05, 1.10)	1.04 (1.01, 1.08)
2002	0.97 (0.95, 0.99)	1.08 (1.05, 1.12)	1.13 (1.11, 1.16)	1.09 (1.05, 1.12)
2003	0.97 (0.95, 0.99)	1.12 (1.08, 1.16)	1.20 (1.18, 1.23)	1.14 (1.10, 1.18)
2004	0.93 (0.91, 0.95)	1.23 (1.19, 1.27)	1.27 (1.25, 1.30)	1.20 (1.16, 1.24)
2005	0.91 (0.89, 0.93)	1.21 (1.16, 1.25)	1.33 (1.31, 1.36)	1.26 (1.22, 1.30)
2006	0.89 (0.87, 0.90)	1.23 (1.19, 1.27)	1.39 (1.36, 1.42)	1.32 (1.28, 1.36)
2007	0.88 (0.86, 0.90)	1.25 (1.21, 1.29)	1.46 (1.43, 1.49)	1.48 (1.43, 1.52)
2008	0.89 (0.88, 0.91)	1.25 (1.21, 1.29)	1.53 (1.50, 1.57)	1.61 (1.56, 1.66)

**Table 2**  
Among-practice variability in mortality, cancer, prescription, and encounter rates

	Unadjusted Analysis Variance Component	Age Adjusted Analysis Variance Component	Adjusted Analysis* Variance Component	Adjusted Analysis+ Variance Component
<b>Mortality</b>				
2000-2001	0.29 (0.26,0.32)	0.19 (0.16, 0.21)	0.17 (0.15,0.19)	0.15 (0.13,0.18)
2002-2003	0.32 (0.29,0.35)	0.21 (0.19, 0.23)	0.19 (0.17,0.21)	0.18 (0.16,0.20)
2004-2005	0.33 (0.30,0.35)	0.21 (0.19, 0.23)	0.18 (0.17,0.20)	0.17 (0.16,0.19)
2006-2008	0.36 (0.33,0.39)	0.25 (0.23, 0.27)	0.21 (0.20,0.23)	0.21 (0.19,0.22)
Overall	0.34 (0.32,0.37)	0.22 (0.21, 0.24)	0.20 (0.19,0.22)	0.19 (0.17,0.20)
<b>Cancer</b>				
2000-2001	0.34 (0.30,0.39)	0.23 (0.20, 0.27)	0.21 (0.18,0.25)	0.21 (0.18,0.25)
2002-2003	0.31 (0.28,0.34)	0.20 (0.18, 0.23)	0.19 (0.17,0.21)	0.18 (0.16,0.21)
2004-2005	0.26 (0.23,0.28)	0.15 (0.13, 0.17)	0.14 (0.12,0.16)	0.14 (0.13,0.16)
2006-2008	0.29 (0.27,0.32)	0.15 (0.14, 0.17)	0.14 (0.12,0.15)	0.14 (0.12,0.16)
Overall	0.29 (0.27,0.32)	0.16 (0.14, 0.17)	0.14 (0.13,0.16)	0.14 (0.13,0.16)
<b>Prescriptions</b>				
2000-2001	0.27 (0.24, 0.30)	0.24 (0.21, 0.27)	0.19 (0.16,0.21)	0.18 (0.16,0.20)
2002-2003	0.27 (0.25, 0.30)	0.24 (0.22, 0.27)	0.19 (0.17,0.21)	0.19 (0.17,0.21)
2004-2005	0.29 (0.27, 0.31)	0.25 (0.23, 0.27)	0.20 (0.18,0.22)	0.20 (0.18,0.21)
2006-2008	0.30 (0.28, 0.32)	0.26 (0.24, 0.28)	0.20 (0.19,0.22)	0.20 (0.19,0.22)
All years	0.30 (0.28, 0.32)	0.26 (0.25, 0.28)	0.24 (0.23,0.26)	0.24 (0.22,0.25)
<b>Encounters</b>				
2000-2001	0.40 (0.35, 0.45)	0.37 (0.33, 0.42)	0.30 (0.26,0.34)	0.28 (0.25,0.33)
2002-2003	0.28 (0.26, 0.32)	0.27 (0.24, 0.30)	0.22 (0.20,0.25)	0.20 (0.18,0.23)
2004-2005	0.22 (0.20, 0.24)	0.20 (0.18,0.23)	0.18 (0.16,0.20)	0.17 (0.15,0.19)
2006-2008	0.21 (0.19, 0.23)	0.20 (0.18, 0.21)	0.18 (0.16,0.19)	0.18 (0.16,0.19)
All years	0.24 (0.22, 0.26)	0.22 (0.21, 0.24)	0.20 (0.19,0.22)	0.19 (0.18,0.21)

\* Adjusted for age, vision experience, smoking status recorded, current smokers, BMI status recorded, number of patients with a BMI over 30 and interaction terms for smoking status recorded with current smokers and BMI status recorded with BMI over 30

+ Adjusted for variables listed above and Townsend score; due to practices missing Townsend values the number of observations is reduced