

## Short report

# Generalisability of The Health Improvement Network (THIN) database: demographics, chronic disease prevalence and mortality rates

Betina T Blak MSc PhD  
Health Economist

Mary Thompson PhD  
Research Manager

Hassy Dattani MRPharmS MBA  
Research Director

Alison Bourke BSc MSc MRPharmS  
Managing Director

Cegedim Strategic Data Medical Research Ltd, London, UK

### ABSTRACT

**Introduction** The degree of generalisability of patient databases to the general population is important for interpreting database research. This report describes the representativeness of The Health Improvement Network (THIN), a UK primary care database, of the UK population.

**Methods** Demographics, deprivation (Townsend), Quality and Outcomes Framework (QOF) condition prevalence and deaths from THIN were compared with national statistical and QOF 2006/2007 data.

**Results** Demographics were similar although THIN contained fewer people aged under 25 years. Condition prevalence was comparable, e.g. 3.5% diabetes prevalence in THIN, 3.7% nationally. More

THIN patients lived in the most affluent areas (23.5% in THIN, 20% nationally). Between 1990 and 2009, standardised mortality ratio ranged from 0.81 (95% CI: 0.39–1.49; 1990) to 0.93 (95% CI: 0.48–1.64; 1995). Adjusting for demographics/deprivation, the 2006 THIN death rate was 9.08/1000 population close to the national death rate of 9.4/1000 population.

**Conclusion** THIN is generalisable to the UK for demographics, major condition prevalence and death rates adjusted for demographics and deprivation.

**Keywords:** demography, Great Britain, mortality, prevalence, validation studies, vital statistics

### What this paper adds

- The Health Improvement Network (THIN) UK primary care database is demographically representative of the UK.
- THIN and national crude prevalences of the UK Quality of Outcomes Framework conditions are similar.
- THIN and national death rates are similar when adjusted for demographics and deprivation.

## Introduction

Primary care patient databases consisting of electronic medical records (EMRs) are increasingly used for research.<sup>1</sup> These databases reflect everyday care provided to patients within a sample of practices. Therefore, when databases are used for research, information on the generalisability of the results to the general population is important.<sup>1</sup> The objective of this report is to document the representativeness of observations from The Health Improvement Network (THIN) UK primary care database. This is assessed by comparing observed demographics, chronic condition prevalence, deprivation and deaths with UK national estimates.

## Methods

THIN holds longitudinal anonymised patient EMRs currently collected from 532 general practices across the UK using Vision computer software (INPS, [www.inps4.co.uk/](http://www.inps4.co.uk/)). The database includes more than 10.5 million patients, of whom 3.7 million are currently active (as of end of September 2011). The database holds information on demographics, diagnoses, prescriptions, referrals, hospitalisation, laboratory tests, immunisations, clinical measures taken within the practice and Townsend deprivation.

The following variables were derived from the THIN database:

- distributions of the overall and regional populations, age, gender (from the THIN data set, last data collection January 2011)
- prevalence of the major Quality and Outcomes Framework (QOF) conditions (QOF is a UK national pay for performance scheme to improve the quality of chronic disease management in primary care.
- deprivation (last data collection October 2007)
- death recording (last data collection January 2011).

These were compared descriptively with national estimates of age, gender and death rates and with national QOF 2006/2007 prevalence data.<sup>2,3</sup> To derive the condition prevalence in THIN, Read code lists from the QOF Business Rules<sup>4</sup> were used to make the comparison with the national QOF prevalence more accurate.

The Townsend score is an index of the social and economic deprivation of a locality<sup>5</sup> and has been linked within the general practitioner (GP) software to each patient's postcode. The Townsend score is assigned as a quintile based on census data where the most affluent areas are given a score of 1 and the least

affluent areas are given a score of 5. Therefore, 20% of the overall UK population is assigned to each group.

Age- and gender-standardised mortality rates were derived (indirect standardisation). Standardised mortality ratios (SMRs) were then calculated (observed death rate divided by expected rate) and Byar's approximation formula was applied to calculate 95% confidence intervals (CIs). SMRs were calculated annually from 1990 to 2009.

Age- and gender-standardised death rates were calculated for each Townsend quintile and each year between 2000 and 2006 to give one death rate per Townsend quintile per year (direct standardisation). After evaluating the distribution of the Townsend quintiles in THIN, the standardised death rates were adjusted by quintile to account for deprivation. The average of the death rates by Townsend quintile and year constituted the annual age-, gender- and deprivation-adjusted THIN death rate. Death rates for THIN, with and without deprivation adjustment, were compared with UK death rates.

Only data recorded after each practice's Acceptable Mortality Reporting (AMR) year were analysed,<sup>6</sup> except for death rates which were also analysed without applying the AMR year, i.e. all years were included. Analyses were conducted using SAS version 9.1 and Microsoft Office Excel 2003.

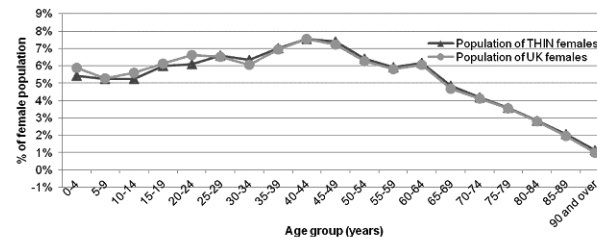
## Results

In 2009, the UK population was approximately 61.7 million.<sup>2</sup> The 3 710 794 actively registered patients in THIN at mid-year 2009 covered approximately 6.0% of the UK population. By region, THIN covered 8.9% of the population in the south of England, 4.6% of London, 4.8% of the east and Midlands, 4.2% of the north of England, 8.5% of Scotland, 7.4% of Wales and 6.6% of Northern Ireland. The THIN population and the overall UK population were similarly distributed across age and gender although THIN contains slightly fewer people aged under 25 years and the male population matches slightly less well than the female population (Figure 1).

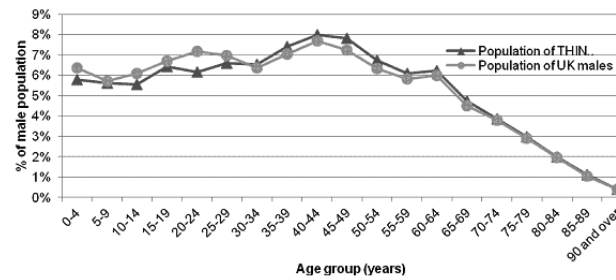
Of the active THIN patients with a Townsend deprivation score (93% of patients), 23.5% belonged to the most affluent group (quintile one), 21.3% belonged to quintile two, 21.0% belonged to quintile three, 19.7% belonged to quintile four and 14.6% belonged to the least affluent group (quintile five).

The QOF condition crude prevalence within THIN was also similar to the national estimates; for example, the crude prevalence of diabetes was 3.5% in THIN compared with 3.7% nationally (Table 1). The greatest

Female



Male



**Figure 1** Comparison of The Health Improvement Network (THIN) population and UK population in 2009 according to age and gender

**Table 1** QOF<sup>a</sup> condition crude prevalences in THIN compared with UK national QOF data

QOF conditions	THIN (%)	UK national QOF (%)
Atrial fibrillation	1.4 (1.4–1.4) <sup>b</sup>	1.3
Asthma	6.0 (5.9–6.0)	5.8
Cancer	0.9 (0.9–0.9)	0.9
Coronary heart disease	3.9 (3.9–3.9)	3.7
Chronic kidney disease	2.5 (2.5–2.5)	2.3
COPD	1.6 (1.6–1.6)	1.5
Dementia	0.5 (0.5–0.5)	0.4
Diabetes	3.5 (3.5–3.5)	3.7
Epilepsy	0.6 (0.6–0.6)	0.6
Heart failure	0.9 (0.9–0.9)	0.8
Hypertension	12.7 (12.6–12.7)	12.6
Learning disability	0.3 (0.3–0.3)	0.3
Mental health	0.7 (0.7–0.8)	0.7
Obesity	8.3 (8.3–8.3)	7.5
Palliative care	0.1 (0.1–0.1)	0.1
Stroke/TIA	1.9 (1.9–1.9)	1.7
Hypothyroidism	2.7 (2.7–2.8)	2.6

<sup>a</sup>Data from the Quality and Outcomes Framework (QOF) year 2006/2007 were used and data from The Health Improvement Network (THIN) were derived for this 2006/2007 cross-section in time. <sup>b</sup>95% confidence intervals are given in parentheses. COPD, chronic obstructive pulmonary disease; TIA, transient ischaemic attack.

difference was in obesity where the crude prevalence in THIN was 8.3% compared with 7.5% nationally.

Without applying AMR, between 1990 and 1997 the SMRs ranged from 0.43 (95% CI: 0.14–1.02; 1990) to 0.72 (95% CI: 0.32–1.39; 1997); between 1998 and 2009 the SMRs ranged from 0.77 (95% CI: 0.35–1.44; 1998) to 0.88 (95% CI: 0.43–1.62; 2000). An SMR of 1 would indicate that the THIN death rate was identical to the national rate. Applying the AMR year, between 1990 and 1997 SMRs ranged from 0.81 (95% CI: 0.39–1.49; 1990) to 0.93 (95% CI: 0.48–1.64; 1995); between 1998 and 2009 the SMRs ranged from 0.86 (95% CI: 0.38–1.67; 2009) to 0.91 (95% CI: 0.45–1.65; 2000); between 1990 and 2009 the observed death rate in THIN ranged from 8.29/1000 population (2009) to 11.62 (1995).

Evaluating the THIN death rates by Townsend quintile, in 2000, the death rate for patients within quintile one (most affluent) was 8.34/1000 population and for patients within quintile five (least affluent) it was 13.06/1000 population. In 2006, these figures were 7.33 and 11.02 deaths/1000 population, respectively. Within each year, the death rate steadily increased with decreasing affluence. In 2006, the THIN age- and gender-adjusted death rate was 8.72/1000 and the age-, gender- and deprivation-adjusted rate was 9.08/1000 population. Adjusting for deprivation therefore resulted in a rate that was more similar to the national rate of 9.4 deaths/1000 population.<sup>2</sup> In the other study years, i.e. 2000–2005, the deprivation-adjusted death rate was consistently higher than the rate without deprivation adjustment in THIN; the percentage difference ranged from 3.1 to 4.4 across the study years. However, the deprivation-adjusted death rates in THIN were consistently slightly lower than national rates.

## Discussion

In addition to THIN, several databases are available for research, e.g. the UK database QRESEARCH<sup>7</sup> and General Practice Research Database,<sup>1</sup> Longitudinal Patient Data from other European countries,<sup>1,8</sup> the Saskatchewan database from Canada<sup>1</sup> and US medical claims databases, e.g. the Kaiser Permanente National Research Database.<sup>1</sup> Other studies have investigated the usefulness of primary care data for research and found them valuable.<sup>9,10</sup>

THIN currently covers approximately 6.0% of the overall UK population. The regional coverage in the dataset ranges from 4.2% in the north of England to 8.9% in the south. Although this is governed by the number of practices using Vision software within each region, THIN is demographically representative of the UK population. The THIN major chronic condition

crude prevalences are similar to, although generally slightly higher than, national estimates with the greatest difference in obesity. Differences in crude condition estimates are likely to be due to factors that are not controlled for, such as age and gender, but could also represent differences between general practice software systems and in GPs' recording practices. Observed death rates with and without applying AMR year were closer to national estimates in the later years than in the earlier years.

THIN has a higher proportion of patients living in the most affluent areas than the national average and there was a notable increase in mortality by decreasing affluence, which is supported by previous studies.<sup>11</sup> The imbalanced deprivation distribution of THIN meant that adjustment for deprivation resulted in estimates closer to national death rates. Deprivation therefore partially explains the lower THIN death rates and adjusting for social deprivation at the patient level is therefore important when estimating mortality and morbidity rates.

## Conclusion

THIN is generalisable to the UK in terms of demographics and crude prevalences of major conditions. THIN and national death rates are similar when adjusted for demographics and deprivation. THIN is therefore a useful data source for conducting research and deriving information on UK health care although investigations into study-specific variables are recommended before initiating a study.

## REFERENCES

- 1 Strom BL. PART IIIb Automated data systems available for pharmacoepidemiology studies. In: Strom BL (ed) *Pharmacoepidemiology* (4e). Wiley: Chichester, 2005, pp. 219–346.
- 2 Office for National Statistics Death Rates. England and Wales: [www.statistics.gov.uk/hub/population/deaths/mortality-rates/index.html](http://www.statistics.gov.uk/hub/population/deaths/mortality-rates/index.html); Scotland: [www.gro-scotland.gov.uk/statistics/theme/vital-events/general/ref-tables/index.html](http://www.gro-scotland.gov.uk/statistics/theme/vital-events/general/ref-tables/index.html); Northern Ireland: [www.nisra.gov.uk/demography/default.asp10.htm](http://www.nisra.gov.uk/demography/default.asp10.htm) (accessed August 2009).
- 3 Quality and Outcomes Framework (QOF) 2006/2007. England: <http://data.gov.uk/dataset/quality-and-outcomes-framework-2006-2007-data-tables>; Scotland: [www.isd.scotland.archive.scot.nhs.uk/isd/6433.html](http://www.isd.scotland.archive.scot.nhs.uk/isd/6433.html); Northern Ireland: [www.dhsspsni.gov.uk/index/hss/gp\\_contracts/gp\\_contract\\_qof/qof\\_data/pc\\_-\\_board\\_qof\\_data\\_2007.htm](http://www.dhsspsni.gov.uk/index/hss/gp_contracts/gp_contract_qof/qof_data/pc_-_board_qof_data_2007.htm); Wales: [www.wales.nhs.uk/sites3/page.cfm?orgid=480&pid=28246](http://www.wales.nhs.uk/sites3/page.cfm?orgid=480&pid=28246) (accessed January 2008).
- 4 QOF Implementation: Business Rules v9.0. [www.pcc.nhs.uk/business-rules-v9-0](http://www.pcc.nhs.uk/business-rules-v9-0) (accessed January 2008).

- 5 Townsend P, Phillimore P and Beattie A. *Health and Deprivation: inequality and the north*. Routledge: London, 1988.
- 6 Maguire A, Blak BT and Thompson M. The importance of defining periods of complete mortality reporting for research using automated data from primary care. *Pharmacoepidemiology and Drug Safety* 2009;18:76–83.
- 7 Hippisley-Cox J, Stables D and Pringle M. QRESEARCH: a new general practice database for research. *Informatics in Primary Care* 2004;12(1):49–50.
- 8 Blak BT, Maguire A, Hards M and Schwalm MS. Cholesterol recording in European primary care data. *Mapi Research Trust. PharmacoEpi&Risk Management Newsletter* 2009;4:5–8.
- 9 Tsang C, Majeed A, Banarsee R, Gnani S and Aylin P. Recording of adverse events in English general practice: analysis of data from electronic patient records. *Informatics in Primary Care* 2010;18(2):117–24.
- 10 de Lusignan S and van Weel C. The use of routinely collected computer data for research in primary care: opportunities and challenges. *Family Practice* 2006;23: 253–63.
- 11 Romeri E, Baker A and Griffiths C. Mortality by deprivation and cause of death in England and Wales, 1999–2003. *Health Statistics Quarterly* 2006;32:19–34.

#### CONFLICT OF INTEREST

All authors were employed by Cegedim Strategic Data Medical Research Ltd during the completion of the report. Cegedim Strategic Data Medical Research Ltd provides access to The Health Improvement Network (THIN) database.

#### STATEMENT ON AUTHOR CONTRIBUTION

All authors have contributed to and agree the submission in its current form.

#### ADDRESS FOR CORRESPONDENCE

Mary Thompson  
Research Manager  
Cegedim Strategic Data Medical Research Ltd  
1 Canal Side Studios  
8–14 St Pancras Way  
London NW1 0QG  
UK  
Tel: +44 (0)207 388 8215  
Fax: +44 (0)207 388 8492  
Email: mary.thompson@cegedim.com

*Accepted February 2012*

