Brief report

Title: Comparability of the Age and Sex Distribution of the UK Clinical Practice Research Datalink and the Total Dutch Population.

Running head: CPRD and The Netherlands: age and sex distribution

Authors

Roy G.P.J. de Jong[1, 2], Arlene M. Gallagher[3, 4], Emily Herrett[3, 5], Ad A.M. Masclee[6], Maryska L.G. Janssen-Heijnen[7, 8], Frank de Vries[4, 9, 10, 11]

Note: RdJ and AG are joint first author.

Affiliations
[1]Department of Internal Medicine, VieCuri Medical Centre, Venlo, The Netherlands;

[2]Department of Internal Medicine, Division of Gastroenterology and Hepatology, GROW School for Oncology and Developmental Biology, Maastricht University Medical Centre+, Maastricht, The Netherlands;

[3]Clinical Practice Research Datalink, Medicines and Healthcare Products Regulatory Agency, London, United Kingdom;

[4]Utrecht Institute for Pharmaceutical Sciences, Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht University, Utrecht, the Netherlands;

[5]Department of Non-communicable Disease Epidemiology, London School of Hygiene & Tropical Medicine, London, United Kingdom;

[6]Department of Internal Medicine, Division of Gastroenterology and Hepatology, NUTRIM School of Nutrition and Translational Research in Metabolism, Maastricht University Medical Centre+, Maastricht, the Netherlands;

[7]Department of Epidemiology, GROW School for Oncology and Developmental Biology, Maastricht University Medical Centre+, Maastricht, The Netherlands;

[8]Department of Clinical Epidemiology, VieCuri Medical Centre, Venlo, the Netherlands;

[9]Department of Clinical Pharmacy and Toxicology, Maastricht University Medical Centre+, Maastricht, the Netherlands;

[10]Department of Health Services Research, CAPHRI School for Public Health and Primary Care, Maastricht University Medical Centre+, Maastricht, the Netherlands;

[11]MRC Life-course Epidemiology Unit, University of Southampton, Southampton, United Kingdom.

Corresponding author

Frank de Vries, Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht 3508 TB, The Netherlands, Telephone number: +31(0)302537324, Fax number: +31(0)302539166, E-mail: f.devries@uu.nl

Keywords: representativeness, database, population-based, epidemiology, pharmaco-epidemiology

Key points:

- The age and sex distribution of the UK CPRD were comparable to the total Dutch population.

- Results from observational studies that have used CPRD data are applicable to the Dutch population.

Conflict of interest: None declared.

Word count: 1668

Prior postings and presentations: none.

ABSTRACT

Purpose: The UK Clinical Practice Research Datalink (CPRD) is increasingly being used by Dutch researchers in epidemiology, and pharmaco-epidemiology. It is however unclear if CPRD is representative of the Dutch population and whether study results would apply to the Dutch population. Therefore, as first step, our objective was to compare the age and sex distribution of the CPRD with the total Dutch population.

Methods: As a measure of representativeness, the age and sex distribution of the CPRD were visually and numerically compared with Dutch census data from the StatLine database of the Dutch National Bureau of Statistics in 2011.

Results: The age distribution of men and women in the CPRD population was comparable to the Dutch male and female population. Differences of more than 10% occurred only occurred in older age categories (75+ in men, and 80+ in women).

Conclusions: Results from observational studies that have used CPRD data are applicable to the Dutch population, and a useful resource for decision making in the Netherlands. Nevertheless, differences in drug exposure likelihood between countries should be kept in mind, as these could still cause variations in the actual population studied, thereby decreasing its generalizability.

INTRODUCTION

The United Kingdom (UK) Clinical Practice Research Datalink (CPRD) is one of the world's largest primary care databases, and is frequently used for post-authorisation safety studies, pharmaco-epidemiology, and disease epidemiology.[1-5] Examples include the evaluation of side effects of dopamine agonists[2] or diabetes drugs[3,4], and the epidemiology of fractures.[5]

From a global perspective, the healthcare system in the Netherlands, a small country not far from the UK, is largely comparable to that of the UK's National Health Service (NHS); i.e. everyone has equal access to medical care regardless of income or socioeconomic status. In both countries, the general practitioner (GP) is the gatekeeper of the public healthcare system, meaning patients cannot refer themselves to secondary or tertiary care without the GP's approval. These conditions are key for conducting population-based pharmaco-epidemiological studies.

Although the Dutch public healthcare system has excellent conditions for establishing a large primary care database for pharmaco-epidemiological research that is comparable to the CPRD in terms of sample size, this has not yet occurred. Smaller primary care databases do exist (e.g. LINH (Netherlands Information Network of General Practice), IPCI (Integrated Primary Care Information), Pharmo Institute), however these are generally restricted by a limited set of medical codes (approximately 1000 different "International Classification of Primary Care" codes versus over 100,000 READ codes in the CPRD), few validation studies, considerable smaller sample size (e.g. 350.000 in LINH and 1.5 million in IPCI versus over 11 million in CPRD in 2011), and limited access to routinely collected lifestyle data, such as tobacco use, alcohol consumption, and socioeconomic status.[6] Furthermore, claims that these data are representative for the total Dutch population are seldom supported by published figures.[7,8] For these reasons, an increasing number of CPRD studies are being conducted by researchers from the Netherlands, and are financially

supported by Dutch universities and funding agencies such as ZonMw and NWO. A recent study showed that the CPRD is representative of the total UK population with respect to age and sex and covers 6.9% of the UK population.[1,9] A wide range of diagnoses in the CPRD have been validated in a number of studies and data quality are further enhanced by NHS annual reward and incentive programme that details GP practice achievement results, the Quality and Outcomes Framework (QOF). The QOF awards GPs for regular recording of detailed data on a wide range of diseases. As a result, the CPRD contains millions of recordings for measurements such as blood pressure, cholesterol values and lung function. In addition, the strength of CPRD's data partially explain why officials of the Food and Drug Administration in the United States perform studies using the CPRD database for drug safety monitoring and regulatory decision making.[10-12]

Although there are many similarities between healthcare systems of the UK and the Netherlands, it is unclear whether results from CPRD studies would apply to the Dutch population. Therefore, our objective is to compare the age and sex distribution of the CPRD to the total Dutch population.

METHODS

Using the same sample of data from the previously published CPRD data resource profile, the age and sex distribution of the CPRD primary care data on 27 March 2011 were visually and numerically compared with UK and Dutch census data in 2011.[1] The CPRD (formerly known as Value Added Medical Products (VAMP), and later General Practice Research Database (GPRD)[13,14]) harnesses data from UK's general practices and produces a primary care dataset since 1987. Through the years, it has become one of the largest databases of longitudinal medical records from primary care in the world, with coverage of over 11.3 million patients from 674 practices. To date, 4.4 million active (alive, currently registered) patients meet quality criteria (approximately 6.9% of the UK population), who are broadly representative of the UK general population in terms of age, sex and ethnicity. For this study, visual comparison was performed by inspecting the overlap between the respective lines in a graph (Figure 1). An additional comparison calculated the differences between proportional distributions of 5-year age groups of CPRD data versus Dutch census data in 2011. We are not aware of any objective methods to define representativeness of patients in a research database compared to a country's total population. We therefore described the absolute and proportional differences between the age and sex distributions of CPRD and Dutch Census data in order to leave this to the reader, and made a subjective decision to consider an age-sex specific difference of <10% representative (Table 1). Numbers for computing sex-stratified age categories of the total Dutch population in March 2011 were obtained from the StatLine database of the Dutch National Bureau of Statistics (www.cbs.nl).

RESULTS

In general, the age distribution of men and women in the CPRD population was comparable to the Dutch male and female population (Figure 1). Overall, the percentage men and women in the CPRD in 2011 was the same as in the Dutch population (49.5% men, 50.5% women). The additional comparison based on calculating the differences between proportional distributions of 5-year age groups, showed that differences of more than 10% occurred only in older age categories, starting from 75+ in men, and 80+ in women (Table 1).

DISCUSSION

In this study, we showed that the age and sex distribution the CPRD was visually and numerically comparable to that of the total Dutch population in 2011.

To our knowledge, this is the first study to directly compare the age and sex distribution of the CPRD to the total Dutch population. Apart from LINH (Netherlands Information Network of General Practice), a national data network compiling electronic medical records (EMR) of 92 primary care practices with 211 GPs and over 350,000 patients, we are not aware of any Dutch EMR database with data from GPs, which has published data on its representativeness according to the total Dutch population. The IPCI database is a longitudinal primary care database maintained by the department of Medical Informatics of the Erasmus Medical Centre in Rotterdam. In published papers of studies using the IPCI database it is frequently stated that the database is comparable to the total Dutch population in terms of age and sex.[15,16] However, we could not verify this claim in a (peer-reviewed) publication. In a report of the Netherlands institute for health services research (NIVEL), which maintains the LINH database, it is shown that LINH is generally comparable to the total Dutch population, with a slight underrepresentation of women of 75 years and older.[8] However, age- and sex-stratified proportions were calculated as compared to the total LINH population and not as compared to the total LINH population stratified by sex, making direct comparison with our results difficult.

Our calculations of the differences between proportional distributions of 5-year age groups show that the CPRD is comparable to the Dutch population in terms of age and sex up to age 75 in men and age 80 in women. However, in Figure 1 it can be clearly seen that the age distribution of CPRD and Dutch census data almost overlap in these higher age groups. The large differences for higher age groups seen in Table 2 may be a spurious finding, since the calculations were based on

very small proportions of the total population. Of note, some difference is also seen between CPRD and the UK population in Figure 1. Men and to a lesser extent women aged between 20 and 35 years of age are underrepresented in CPRD, which has been attributed to the fact that these individuals probably do not register with a GP.[9]

Several strengths have to be noted for this study. First, UK CPRD data were compared to data from the StatLine database of the Dutch National Bureau of Statistics, a reliable source of population-based information, regulated by national and European codes and laws.[17] Second, by calculating the differences between proportional distributions in all age groups, we demonstrated that the distributions were not only visually, but also numerically comparable. Lastly, by showing the overall representativeness of the CPRD database of another population besides the UK population, CPRD may be used as rich data source for healthcare policymakers outside the UK.

There are also several limitations to this comparison. First of all, there are no objective methods to define representativeness of patients in a research database compared to a country's total population. To overcome this, we gave the reader insight into the various ways of comparing these data. Furthermore, the CPRD population was compared to the total Dutch population in terms of age and sex only. Therefore, we cannot rule out differences in for instance ethnicity, socioeconomic class, or lifestyle, which may in turn impact disease prevalence, exposure to important risk factors, or the degree of health care seeking behaviour. Based on the report of the OECD health indicators, the UK population has higher rates of tobacco and alcohol consumption, and especially obesity among adults, compared to the Dutch population.[18] Although the results of our study imply that the UK CPRD may also be representative of the total Dutch population, information on a specific population of drug-users is ultimately necessary to know whether a study's results are transferrable to one's own region. As of yet, we have not looked into the

comparability of various subpopulations in CPRD and The Netherlands. In addition, relative risks of disease outcomes found in CPRD could be extrapolated to the Dutch population, incidence rates or absolute risks cannot.

In conclusion, this study showed that the age- and sex distribution of CPRD were generally comparable to that of the total Dutch population. Results from observational studies that have used CPRD data are applicable to the total Dutch population (similar to how relative risks from randomized clinical trials apply to their demarcated population), and a useful resource for decision making in the Netherlands. Nevertheless, differences in drug exposure likelihood between countries should be kept in mind, as these could still cause variations in the actual population studied, thereby decreasing its generalizability. In addition, the results of this study may encourage scientists from other countries with similar healthcare systems to perform comparable studies of CPRD representativeness.

Table 1. Differences between proportional age distributions of CPRD and the total Dutch population.

Age group	Total				Men				Women			
	CPRD	NL Census	Δ^a	$\Delta\%^b$	CPRD	NL Census	Δ^a	$\Delta\%^b$	CPRD	NL Census	Δ^a	$\Delta\%^b$
0-4	4.99	5.54	-0.55	-9.94	5.15	5.72	-0.57	-9.98	4.82	5.35	-0.53	-9.90
5-9	5.61	5.89	-0.28	-4.75	5.80	6.09	-0.29	-4.74	5.43	5.70	-0.27	-4.77
10-14	5.68	6.00	-0.32	-5.38	5.87	6.20	-0.33	-5.32	5.49	5.80	-0.31	-5.43
15-19	5.99	6.03	-0.04	-0.65	6.18	6.23	-0.05	-0.80	5.80	5.83	-0.03	-0.50
20-24	5.83	6.23	-0.40	-6.44	5.98	6.35	-0.37	-5.89	5.68	6.10	-0.42	-7.00
25-29	6.32	6.02	0.30	4.96	6.34	6.13	0.21	3.41	6.31	5.92	0.39	6.54
30-34	6.51	6.03	0.47	7.83	6.57	6.11	0.46	7.65	6.44	5.96	0.48	8.01
35-39	6.69	6.67	0.01	0.20	6.84	6.74	0.10	1.62	6.53	6.61	-0.08	-1.22
40-44	7.59	7.77	-0.17	-2.22	7.80	7.91	-0.11	-1.42	7.39	7.62	-0.23	-3.03
45-49	7.77	7.79	-0.02	-0.31	7.97	7.95	0.02	0.22	7.58	7.64	-0.06	-0.86
50-54	6.95	7.19	-0.24	-3.41	7.12	7.30	-0.18	-2.44	6.78	7.09	-0.31	-4.38
55-59	6.06	6.56	-0.49	-7.54	6.14	6.64	-0.50	-7.62	5.99	6.47	-0.48	-7.46
60-64	6.15	6.64	-0.48	-7.24	6.16	6.72	-0.56	-8.30	6.15	6.55	-0.40	-6.18
65-69	5.30	4.79	0.51	**10.64**	5.23	4.79	0.44	9.18	5.37	4.79	0.58	**12.08**
70-74	4.03	3.83	0.20	5.27	3.87	3.67	0.20	5.31	4.19	3.98	0.21	5.23
75-79	3.36	3.00	0.36	**11.90**	3.09	2.66	0.43	**15.98**	3.62	3.33	0.29	8.71
80-84	2.56	2.17	0.39	**18.00**	2.15	1.70	0.45	**26.28**	2.97	2.64	0.33	**12.80**
85-89	1.64	1.28	0.36	**28.27**	1.20	0.82	0.38	**47.05**	2.07	1.73	0.34	**19.62**
90+	0.98	0.57	0.41	**70.52**	0.55	0.27	0.28	**102.20**	1.40	0.87	0.53	**60.92**

[a]Absolute difference, [b]Percentage difference between CPRD versus NL Census data (CPRD minus NL Census, divided by NL Census), with >10% regarded as different. Abbreviations: CPRD: Clinical Practice Research Datalink; NL: Netherlands; Δ difference
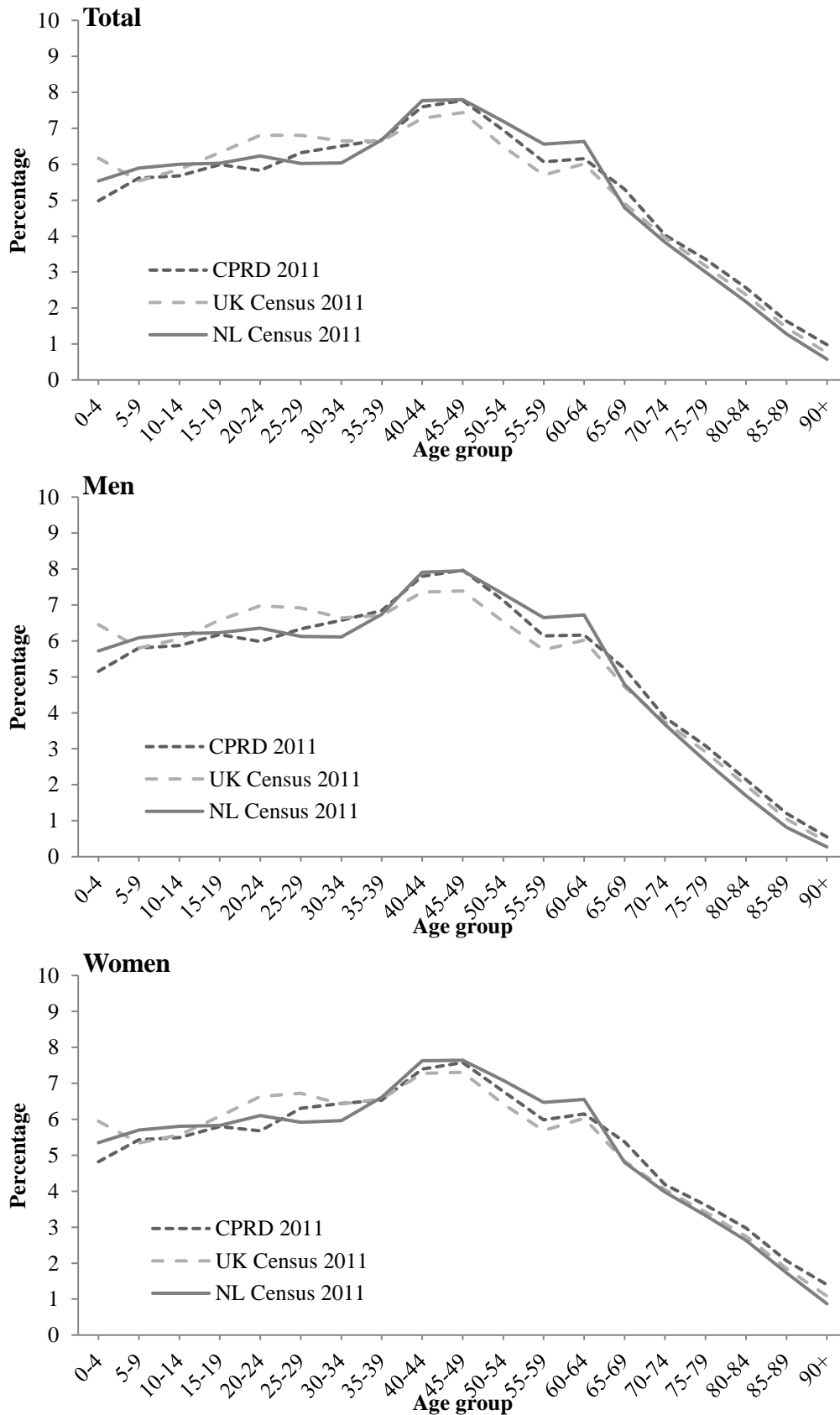
Figure 1. Age distribution of the CPRD primary care data on 27 March 2011 compared with UK and NL Census data from 2011, in both men and women (top panel), men only (middle panel) and women only (bottom panel). These data are based on a one-million patient sample of CPRD. Adapted from Herrett et al. Int. J Epidemiol 2015.

REFERENCES

1.   Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, van Staa T, Smeeth L. Data Resource Profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol* 2015;**44**(3):827-36.

2.   Schade R, Andersohn F, Suissa S, Haverkamp W, Garbe E. Dopamine agonists and the risk of cardiac-valve regurgitation. *N Engl J Med* 2007;**356**(1):29-38.

3.   Knapen LM, van Dalem J, Keulemans YC, van Erp NP, Bazelier MT, De Bruin ML, Leufkens HG, Croes S, Neef C, de Vries F, Driessen JH. Use of Incretin Agents and Risk of Pancreatic Cancer: A Population-Based Cohort Study. *Diabetes Obes Metab* 2015.

4.   Lo Re III V, Haynes K, Ming EE, Wood Ives J, Horne LN, Fortier K, Carbonari DM, Hennessy S, Cardillo S, Reese PP, Reddy KR, Margolis D, Apter A, Kimmel SE, Roy J, Freeman CP, Razzaghi H, Holick CN, Esposito DB, Van Staa TP, Bhullar H, Strom BL. Safety of saxagliptin: rationale for and design of a series of postmarketing observational studies. *Pharmacoepidemiol Drug Saf* 2012;**21**(11):1202-15.

5.   van Staa TP, Dennison EM, Leufkens HG, Cooper C. Epidemiology of fractures in England and Wales. *Bone* 2001;**29**(6):517-22.

6.   Requena G, Huerta C, Gardarsdottir H, Logie J, Gonzalez-Gonzalez R, Abbing-Karahagopian V, Miret M, Schneider C, Souverein PC, Webb D, Afonso A, Boudiaf N, Martin E, Oliva B, Alvarez A, De Groot MC, Bate A, Johansson S, Schlienger R, Reynolds R, Klungel OH, de Abajo FJ. Hip/femur fractures associated with the use of benzodiazepines (anxiolytics, hypnotics and related drugs): a methodological approach to assess consistencies across databases from the PROTECT-EU project. *Pharmacoepidemiol Drug Saf* 2015.

7.   Abbing-Karahagopian V, Kurz X, de Vries F, van Staa TP, Alvarez Y, Hesse U, Hasford J, Dijk L, de Abajo FJ, Weil JG, Grimaldi-Bensouda L, Egberts AC, Reynolds RF, Klungel OH. Bridging differences in outcomes of pharmacoepidemiological studies: design and first results of the PROTECT project. *Curr Clin Pharmacol* 2014;**9**(2):130-8.

8.   NIVEL. Representativiteit van LINH. 2010.

9.   Campbell J, Dedman DJ, Eaton SC, Gallagher AM, Williams TJ. Is the GPRD GOLD population comparable to the UK population? *Pharmacoepidemiol Drug Saf* 2013;**22**(Suppl 1):280.

10.   Zhou EH, Gelperin K, Levenson MS, Rose M, Hsueh YH, Graham DJ. Risk of acute myocardial infarction, stroke, or death in patients initiating olmesartan or other angiotensin receptor blockers - a cohort study using the Clinical Practice Research Datalink. *Pharmacoepidemiol Drug Saf* 2014;**23**(4):340-7.

11.   Hammad TA, Graham DJ, Staffa JA, Kornegay CJ, Dal Pan GJ. Onset of acute myocardial infarction after use of non-steroidal anti-inflammatory drugs. *Pharmacoepidemiol Drug Saf* 2008;**17**(4):315-21.

12.   US Food and Drug Administration (FDA). FDA Drug Safety Communication: Possible fracture risk with high dose, long-term use of proton pump inhibitors. 2010.

13.   Kousoulis AA, Rafi I, de Lusignan S. The CPRD and the RCGP: building on research success by enhancing benefits for patients and practices. *Br J Gen Pract* 2015;**65**(631):54-5.

14.   Williams T, van Staa T, Puri S, Eaton S. Recent advances in the utility and use of the General Practice Research Database as an example of a UK Primary Care Data resource. *Ther Adv Drug Saf* 2012;**3**(2):89-99.

15. Noordam R, Aarts N, Verhamme KM, Sturkenboom MC, Stricker BH, Visser LE. Prescription and indication trends of antidepressant drugs in the Netherlands between 1996 and 2012: a dynamic population-based study. *Eur J Clin Pharmacol* 2015;**71**(3):369-75.

16. Masclee GM, Coloma PM, Spaander MC, Kuipers EJ, Sturkenboom MC. NSAIDs, statins, low-dose aspirin and PPIs, and the risk of oesophageal adenocarcinoma among patients with Barrett's oesophagus: a population-based case-control study. *BMJ Open* 2015;**5**(1):e006640.

17. Centraal Bureau voor de Statistiek. CBS Kwaliteitsverklaring. 2014.

18. OECD. Health at a Glance 2015: OECD Indicators. Paris: OECD Publishing, 2015.

FIGURE LEGENDS

Figure 1. Age distribution of the CPRD primary care data on 27 March 2011 compared with UK and NL Census data from 2011, in both men and women (top panel), men only (middle panel) and women only (bottom panel). These data are based on a one-million patient sample of CPRD. Adapted from Herrett et al. Int. J Epidemiol 2015.