

Importance of different electronic medical record components for chronic disease identification in a Swiss primary care database: a cross-sectional study

Rahel Meier*, Thomas Grischott*, Yael Rachamin, Levy Jäger, Oliver Senn, Thomas Rosemann, Jakob M. Burgstaller, Stefan Markun

Institute of Primary Care, University Hospital Zurich, University of Zurich, Zurich, Switzerland

* These authors contributed equally to this manuscript

Summary

BACKGROUND: Primary care databases collect electronic medical records with routine data from primary care patients. The identification of chronic diseases in primary care databases often integrates information from various electronic medical record components (EMR-Cs) used by primary care providers. This study aimed to estimate the prevalence of selected chronic conditions using a large Swiss primary care database and to examine the importance of different EMR-Cs for case identification.

METHODS: Cross-sectional study with 120,608 patients of 128 general practitioners in the Swiss FIRE ("Family Medicine Research using Electronic Medical Records") primary care database in 2019. Sufficient criteria on three individual EMR-Cs, namely *medication*, *clinical or laboratory parameters* and *reasons for encounters*, were combined by logical disjunction into definitions of 49 chronic conditions; then prevalence estimates and measures of importance of the individual EMR-Cs for case identification were calculated.

RESULTS: A total of 185,535 cases (i.e. patients with a specific chronic condition) were identified. Prevalence estimates were 27.5% (95% CI: 27.3–27.8%) for hypertension, 13.5% (13.3–13.7%) for dyslipidaemia and 6.6% (6.4–6.7%) for diabetes mellitus. Of all cases, 87.1% (87.0–87.3%) were identified via *medication*, 22.1% (21.9–22.3%) via *clinical or laboratory parameters* and 19.3% (19.1–19.5%) via *reasons for encounters*. The majority (65.4%) of cases were identifiable solely through *medication*. Of the two other EMR-Cs, *clinical or laboratory parameters* was most important for identifying cases of chronic kidney disease, anorexia/bulimia nervosa and obesity whereas *reasons for encounters* was crucial for identifying many low-prevalence diseases as well as cancer, heart disease and osteoarthritis.

CONCLUSIONS: The EMR-C *medication* was most important for chronic disease identification overall, but identification varied strongly by disease. The analysis of the importance of different EMR-Cs for estimating prevalence

revealed strengths and weaknesses of the disease definitions used within the FIRE primary care database. Although prioritising specificity over sensitivity in the EMR-C criteria may have led to underestimation of most prevalences, their sex- and age-specific patterns were consistent with published figures for Swiss general practice.

Introduction

Electronic medical records (EMRs) are increasingly used in primary care [1–2]. They are typically organised into different, often structured, components (EMR-Cs) such as medication data, laboratory values, clinical parameters or coded diagnoses. EMR data from multiple primary health-care providers can be merged into primary care databases, which then offer significant potential for research [3–6]. Large primary care databases have been developed in healthcare systems internationally, including in the UK [7–8], Canada [9], Spain [10] and Italy [11].

In Switzerland, primary care is predominantly provided by general practitioners who work in private practices and bill according to a nationwide, uniform fee-for-service tariff system. Costs exceeding a patient's deductible are covered by compulsory general health insurance. About 70%

LIST OF ABBREVIATIONS

ATC	Anatomical Therapeutic Chemical
c	relative contribution
c _e	relative exclusive contribution
CLP	EMR-C <i>clinical or laboratory parameters</i>
EMR	Electronic Medical Record
EMR-C	Electronic Medical Record Component
FIRE	Family Medicine Research using Electronic Medical Records
GTIN	Global Trade Item Number
ICPC-2	International Classification of Primary Care 2 nd edition
MED	EMR-C <i>medication</i>
p	period prevalence estimate (for year 2019)
PCG	Pharmaceutical Cost Group
RFE	EMR-C <i>reasons for encounters</i>

Thomas Grischott
Institute of Primary Care
University Hospital Zurich
University of Zurich
Pestalozzistrasse 24
CH-8091 Zurich
thomas.grischott[at]usz.ch

of Swiss general practitioners stored their patients' medical records in electronic form in 2019 (82% in 2023) [12], using over 20 different practice information systems [13]. The first primary care database in Switzerland, called FIRE ("Family Medicine Research using Electronic Medical Records"), was established ten years earlier in 2009 [14]. FIRE, which to our knowledge is still the only relevant Swiss primary care database, integrates data from seven different EMR systems used by primary care providers in German-speaking Switzerland. For more information on FIRE, visit www.fireproject.ch/en.

Coded diagnoses are often unavailable in primary care databases. In the Swiss healthcare system in particular, there are no financial incentives for coding diagnoses, and other incentives, such as facilitating access to decision aids by linking them directly to coded diagnoses, are difficult to implement on a larger scale due to the multitude of different practice softwares used. Therefore, coded diagnoses often need to be implemented secondarily in order to realize the primary care database's full potential for research, clinical practice and public health. This can be achieved, for example, using natural language processing, statistical and machine learning methods or rule-based algorithms, which process operationalised diagnostic criteria [15–16]. Rule-based algorithms can yield valid results and have certain advantages over other chronic disease identification techniques – among others, faster implementation and easier interpretability – that make them attractive for many applications and researchers [15, 17]. Previous research on identifying chronic diseases in primary care databases using rule-based algorithms often focused on medication data, but the relevance of exploiting multiple complementary EMR-Cs has been pointed out [18]. Yet there is little quantitative evidence on the importance of different EMR-Cs for chronic disease identification.

The aims of the present study were to analyse period prevalence estimates of 49 chronic conditions in the FIRE primary care database identified by customized rule-based case definitions using three commonly used EMR-Cs available in the FIRE primary care database, and to examine the importance of the different EMR-Cs for case identification using tailored importance metrics.

Methods

Study design, setting and participants

We performed a cross-sectional study using data from the Swiss FIRE primary care database [14]. At the end of 2019, the FIRE primary care database held almost nine million consultation records from over 500 general practitioners with medication prescription data including Anatomical Therapeutic Chemical (ATC) codes [19] and Global Trade Item Numbers (GTIN), clinical parameters and laboratory test results, as well as reasons for encounters coded according to the International Classification of Primary Care, 2nd edition (ICPC-2), a classification method developed by the World Organization of Family Doctors (WONCA) that allows an episodically ordered classification of reasons for encounters, health problems and interventions in primary care encounters [20]. In addition, the FIRE primary care database contains administrative data and demographic information of general prac-

tioners and patients; only year of birth and gender are known of the patients. Patient identification in FIRE is via identification numbers, which are already hashed in the practices before data transmission. A detailed specification of variables contained in the FIRE database has been published [21].

Not all practice softwares export medication data to the FIRE primary care database with date stamps. For this study, we only considered practices where each medication prescribed in, or related to, a consultation in 2019 was exported with both a start date and an end date (or explicitly as a prescription until further notice). From these practices, we included all patients aged 18–99 years with at least one consultation in 2019 ($N = 120,608$) and used all their data available up to their last recorded consultation (i.e. the index consultation) in 2019. For the study flowchart, please see supplementary figure 1.

Studies within the fully anonymised FIRE project do not fall within the scope of the Human Research Act and are exempt from ethics review. Ethical approval was waived by the Ethics Committee of the Canton of Zurich; BASEC No. Req-2017-00797.

Chronic conditions

The selection of chronic diseases to be identified was based on an appraisal of the literature and consideration of their relevance to general practice. Specifically, we combined the list of 75 chronic diseases (considered most relevant in the context of multimorbidity by experts in family medicine) by N'Goran et al. [22] and the list of 40 chronic diseases (considered by clinicians as the most relevant chronic diseases that constitute multimorbidity) by Barnett et al. [23]. The resulting list was reviewed by three authors (RM, TG, SM) for identifiability in the FIRE primary care database, and the selection of chronic diseases for this study was reached by consensus. Selected chronic diseases were further combined into overarching disease complexes based on presumed indistinguishability in the FIRE primary care database (e.g. "asthma" and "chronic obstructive pulmonary disease" were combined into "obstructive lung disease"). This resulted in a selection of 49 chronic diseases or overarching disease complexes – henceforth "chronic conditions" – potentially identifiable in the FIRE primary care database.

Case definition and diagnostic criteria

For these 49 chronic conditions, we specified individual criteria on three commonly used EMR-Cs available in the FIRE primary care database, namely *medication* (MED), *clinical or laboratory parameters* (CLP) and *ICPC-2 coded reasons for encounters* (RFE), each intended to sufficiently characterise the chronic condition in question. We then defined a *case* of a specific chronic condition as a patient with at least one EMR-C criterion for this chronic condition fulfilled. For most high-prevalence chronic conditions like diabetes or hypertension, we relied on criteria already validated in other primary care database research. In some cases however, validated criteria had to be adapted or combined, or new criteria developed, as is often the case in primary care database research [24–25]:

1. MED: Medication criteria used ATC codes or the more specific GTINs and required implementation of the Swiss pharmaceutical cost groups [26] and of all medications' indications as approved by the Swiss Agency for Therapeutic Products (swissmedic) [27]. To reduce false-positives due to episodic rather than chronic prescription, we required a minimal treatment duration of six months to identify a chronic condition based on medication data.
2. CLP: Criteria on *clinical or laboratory parameters* were derived from international guidelines to identify the chronic conditions thyroid disease, obesity, anorexia/bulimia nervosa, dyslipidaemia, diabetes and hypertension [28–33]. Clinical parameters used were blood pressure and BMI, whereas laboratory parameters included, for example, LDL-cholesterol or HbA1c.
3. RFE: The *reasons for encounters* used were those sufficiently specific to allow identification of selected chronic conditions.

We distinguished reversible (e.g. chronic pain or mental disorders) from permanent chronic conditions (e.g. hypertension or dyslipidaemia). In permanent chronic conditions, EMR-C criteria could be met at any time in the patient history, whereas for reversible chronic conditions, they had to be met within no more than six months prior to the index consultation.

The EMR-C criteria for all 49 selected chronic conditions were compiled by RM and independently reviewed by TG and SM. Differences were resolved by consensus. The final criteria for all chronic conditions can be found in supplementary table 1.

Prevalence estimates and EMR-C importance metrics

We estimated the period prevalences p for the year 2019, of specific chronic conditions in the FIRE general practice patient population as the number of cases divided by the total number of patients included, i.e. (using notation from figure 1):

$$p = n/N = (m + l + r + ml + mr + lr + mlr) / \#patients$$

For the analogous prevalence estimate of having *any* chronic condition, multiple cases (of different chronic conditions) with the same patient were counted only once, i.e. as one single case of the collective condition.

In order to determine the relevance of each EMR-C for case identification, we defined importance metrics as follows (again using notation from figure 1): The *relative contribution* of an EMR-C to case identification of a specific chronic condition was defined as the proportion of cases identified by this EMR-C divided by the total number of cases identified by any EMR-C. An EMR-C's *relative exclusive contribution* to the identification of cases of a specific chronic condition was defined as the proportion of cases identified by this EMR-C but not identified via any other EMR-C, again divided by the total number of cases. Using the EMR-C MED as an example, this is:

$$c = (m + ml + mr + mlr) / n$$

and

$$c_e = m/n$$

Additional analyses

Although redundant in principle, we reported 95% Wilson confidence intervals (CIs) for all p (supplementary table 2) and selected c and c_e as a courtesy to the reader (simultaneous CIs for multinomial proportions where appropriate). For fine-grained comparability with prevalence estimates found in Swiss general practice, and thus for a more thorough validation of our estimates, we also calculated period prevalence estimates stratified for sex \times age groups. Furthermore, we used medians and quartiles as well as the quartile coefficient of dispersion, defined as

$$qcod = (q_{0.75} - q_{0.25}) / (q_{0.75} + q_{0.25})$$

to describe the distributions of the (unstratified) prevalence estimates across general practitioners. The qcod is a robust relative measure of dispersion, useful for comparing the spread of non-normally distributed sets of data that may differ in their medians or units of measurement. Its range and interpretation are similar to those of the better-known coefficient of variation.

All calculations were carried out using R, version 4.0.0 [34]. The Euler diagrams were created using the R library "eulerr", version 6.1.1 [35].

Ethics approval and consent to participate

Studies within the FIRE project do not need ethics approval as they do not fall within the scope of the Human Research Act (waiver granted by the Ethics Committee of the Canton of Zurich; BASEC No. Req-2017-00797).

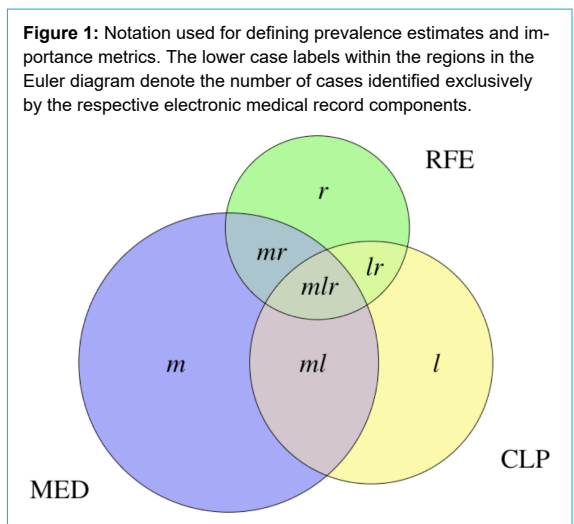
Results

Sample characteristics

We analysed data of $N = 120,608$ patients (52.3% female; median age 52 years [interquartile range (IQR) 36–67]) of 128 general practitioners (35.2%, 52 [43–56]; age data missing for 10 (7.8%) of them) working in 53 different practices. The patients had a median of 4 [2–9] consultations in 2019.

Prevalence estimates and EMR-C importance metrics

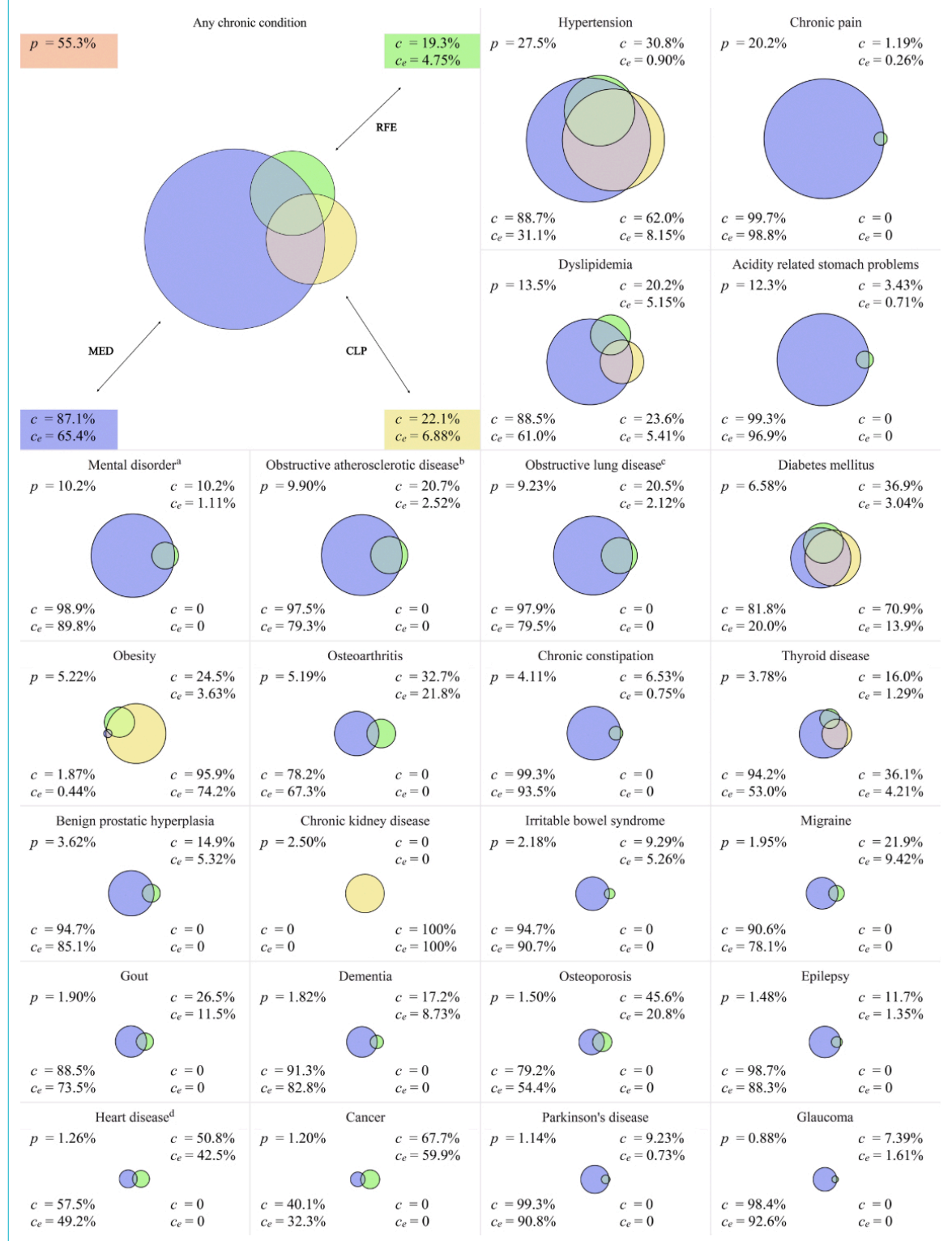
Prevalence estimates are shown in figure 2 for *any* chronic condition and for the 24 most frequently identified chronic



conditions (encompassing 97.0% of all identified cases) together with importance metrics of the three EMR-Cs. Full numerical data for all 49 chronic conditions (including CIs for prevalence estimates) can be found in supplementary table 2. The patients had a mean number of 1.54 chronic

conditions per patient and a mean number of 2.78 chronic conditions per patient with one or more chronic conditions. Of the individual EMR-Cs, MED allowed identification of 161,653 ($c = 87.1\%$; 95% CI 87.0–87.3%) cases, CLP of 40,965 (22.1%; 21.9–22.3%) and RFE of 35,773 (19.3%;

Figure 2: Prevalence estimates (based on $N = 120,608$ patients) and importance metrics, shown in (approximately) area-proportional Euler diagrams, with the areas representing the number of cases identified by the respective electronic medical record component(s). p : prevalence estimate; c : relative contribution; c_e : relative exclusive contribution; MED: medication; CLP: clinical or laboratory parameters; RFE: reasons for encounters. ^a including depression, psychotic disorders and anxiety disorders; ^b including coronary, cerebral and peripheral arteries including arrhythmias and congestive heart disease; ^c including asthma, chronic obstructive pulmonary disease and chronic bronchitis; ^d including arrhythmias and congestive heart disease



19.1–19.5%) cases. Simultaneous identification by all three EMR-Cs was found in 10,143 (5.5%; 5.4–5.6%) cases, all with hypertension, dyslipidaemia, diabetes mellitus, obesity or thyroid disease. Simultaneous identification by exactly two EMR-Cs was found in 32,570 (17.6%; 17.4–17.7%) cases. The identification of the remaining 142,822 (77.0%; 76.8–77.2%) cases depended on one single EMR-C. Exclusive contribution of the EMR-C MED to case identification was highest in chronic pain ($c_e = 98.8\%$ of all cases of chronic pain were identified via MED only; 98.7–98.9%), acidity-related stomach problems (96.6%; 96.3–96.8%) and chronic constipation (93.5%; 92.7–94.1%). Exclusive contribution of the EMR-C CLP was highest in chronic kidney disease (100%; 99.9–100%), anorexia/bulimia nervosa (87.1%; 82.1–90.9%) and obesity (74.2%; 73.1–75.3%). Exclusive contribution of the EMR-C RFE was highest in many low-prevalence chronic conditions and still high in more-prevalent chronic conditions such as cancer (59.9%; 57.3–62.4%), heart disease (42.5%; 40.0–45.0%) and osteoarthritis (21.8%; 20.8–22.9%).

Stratified prevalence estimates and prevalence distributions across general practitioners

Prevalence estimates stratified by sex \times age groups are presented in figure 3 for the 24 most frequently identified chronic conditions, and full numerical data for all 49 chronic conditions can be found in supplementary table 3, which also shows how prevalence estimates of chronic conditions varied across general practitioners. In both sexes, the prevalence estimates of most chronic conditions increased with age, one notable exception being migraine whose prevalence started to decrease in patients in their sixth decade. Distinctly higher prevalence estimates in female patients (compared to their male counterparts) were found for chronic pain, mental disorders, osteoarthritis, thyroid disease, irritable bowel syndrome, migraine, dementia and osteoporosis. Higher prevalence estimates in male patients were observed for hypertension, dyslipidaemia, obstructive atherosclerotic disease, diabetes mellitus, benign prostatic hyperplasia, gout and heart disease. Prevalence estimates of most chronic conditions varied little across general practitioners apart from random fluctuations in low-prevalence chronic conditions and except for chronic kidney disease (median 2.0% [IQR 0.5–3.8%], $q_{cod} = 0.76$), cancer (0.6% [0.3–1.6%], 0.70) and obesity (3.6% [1.9–7.6%], 0.59).

Discussion

Case definitions for chronic diseases are often used for research with primary care databases, but little is known about the importance of different EMR-Cs for chronic disease identification. In this study, we implemented case definitions for 49 chronic conditions in a large Swiss primary care database and analysed prevalence estimates and the importance of three commonly used EMR-Cs for case identification. We found that MED was the EMR-C in the FIRE primary care database contributing most to chronic disease identification. CLP and RFE complemented identification of frequent chronic conditions such as chronic kidney disease, cancer, heart disease and obesity. For most chronic conditions, our case definitions yielded prevalence

estimates lower than expected, but observed sex- and age-specific epidemiological patterns were concordant with previous research.

Many chronic diseases can reliably be identified using medication data; consequently exclusively medication-based identification methods have been developed [36–37]. Our results were in line with this; nearly 90% of chronic condition cases could be identified via MED. MED was most important for the identification of chronic pain, acidity-related stomach problems and chronic constipation, but identification of most other chronic conditions strongly depended on MED data too.

Medication-based chronic disease identification, however, may introduce false-positives due to prevention, overtreatment, off-label treatment or as-needed prescriptions in specific cases. For example, primary preventive use of low-dose aspirin is common and may cause overestimation of obstructive arteriosclerotic disease [38]. Overtreatment with proton pump inhibitors is also common in Switzerland and has most likely inflated the prevalence estimate of acidity-related stomach problems in our study [39]. Likewise, off-label prescribing can also cause overestimation of prevalence even if it is medically justified. An obvious example in our study was the misclassification of 0.1% of women as having benign prostatic hyperplasia; the most plausible explanation here would be off-label use of selective α_1 receptor antagonists for kidney stones [40]. Finally, medication prescribed as-needed for conditions with paroxysmal, non-chronic patterns may also lead to an overestimation of chronic conditions. This effect may have contributed to the high prevalence of chronic pain and acidity-related stomach problems, as pain medication and proton pump inhibitors are more likely to be prescribed on an as-needed basis than other drugs.

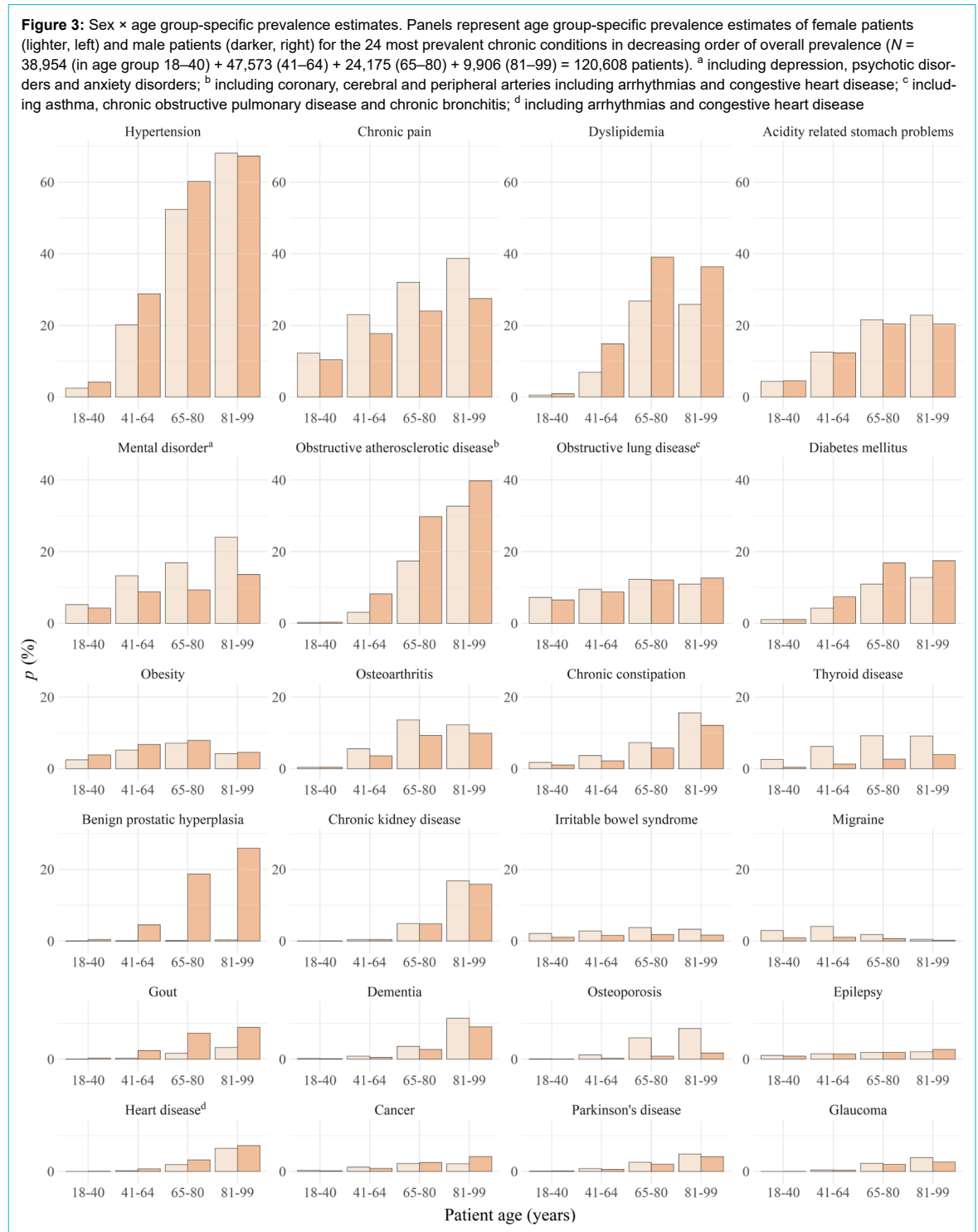
In addition to such false-positives, false-negative identification of chronic diseases based on the EMR-C MED may also occur, especially when drugs used in specific case definitions are typically prescribed in specialised settings only. This constellation occurs with chemotherapeutics or monoclonal antibodies typically prescribed in secondary or tertiary care and explains why this EMR-C is prone to underestimate cancer and autoimmune diseases in primary care. Furthermore, identification of chronic diseases based on medication alone may also fail to detect chronic diseases in mild cases or early stages when non-pharmacological management is still an option (e.g. obesity or early stages of type 2 diabetes mellitus) and thus leads to underestimation of their prevalence. Lastly, patients without medication prescriptions may differ in their socioeconomic status compared to those receiving medication, and case identification depending on the EMR-C MED may be biased accordingly [41, 42].

The value of the EMR-Cs CLP and RFE was limited both in terms of their overall contributions to case identification ($c = 22.1\%$ and 19.3% for CLP and RFE, respectively) as well as their exclusive contributions complementing other EMR-Cs ($c_e = 6.9\%$ and 4.7%). Among all chronic conditions, case identification of chronic kidney disease most strongly depended on laboratory test results ($c = 100\%$ for CLP), as no specific medication could be operationalised, nor is there a sufficiently specific ICPC-2 code for identifying chronic kidney disease.

While considering CLP and RFE may reduce some of the above-mentioned risks of bias linked to medication prescribing, laboratory testing in particular may still be related to disease severity and socioeconomic status, depending on the healthcare setting [42]. Identification based on RFE is likely to be less susceptible to such selection bias. However, our data showed highly incomplete RFE coding by general practitioners in the FIRE network, making this EMR-C per se very unreliable for chronic disease identification. The issue of low coding performance by general practitioners has been documented before; in the context of primary care database research using case definitions

for chronic disease identification, it is most problematic when other EMR-Cs, especially MED, are missing to fill this gap [43–45]. In particular, incomplete coding of cancer and heart disease is well known and confirmed by our low prevalence estimates ($p = 1.2\%$ for cancer, 1.3% for heart disease), and the high relative exclusive contributions of RFE to identifying such cases highlights the problem ($c_e = 59.9\%$ and 42.5%). In the context of RFE coding, the ICPC-3 system should be mentioned, which, in addition to function-related information, also includes conditions missing in ICPC-2, such as chronic kidney disease [46]. It remains to be seen whether the higher level of de-

Figure 3: Sex × age group-specific prevalence estimates. Panels represent age group-specific prevalence estimates of female patients (lighter, left) and male patients (darker, right) for the 24 most prevalent chronic conditions in decreasing order of overall prevalence ($N = 38,954$ (in age group 18–40) + $47,573$ (41–64) + $24,175$ (65–80) + $9,906$ (81–99) = $120,608$ patients). ^a including depression, psychotic disorders and anxiety disorders; ^b including coronary, cerebral and peripheral arteries including arrhythmias and congestive heart disease; ^c including asthma, chronic obstructive pulmonary disease and chronic bronchitis; ^d including arrhythmias and congestive heart disease



tail on the one hand, but the expected higher coding effort on the other, will improve case identification of chronic conditions in primary care databases.

Incomplete documentation is also problematic if vital parameters are used for chronic disease identification: as in other primary care databases, we assume that the body mass index (BMI) may be incompletely documented by general practitioners in the FIRE network, leading to underestimation of both anorexia and obesity whose definitions strongly depend on this measure [47–48].

While a high specificity of our definitions was achieved by stipulating sufficient criteria for each individual EMR-C, sensitivity was increased by logical disjunction over multiple EMR-Cs. However, sensitivity likely remained the most important limitation [17]. To assess its severity, external studies on chronic disease prevalence in Swiss general practice are a useful comparison standard. A recent study by Excoffier et al. surveyed a sample of 118 Swiss general practitioners on chronic diseases of 25 consecutive patients per general practitioner, and produced prevalence estimates for multiple chronic diseases in Swiss general practice [49].

Our results differed to varying degrees from the prevalences found in this study: prevalence estimates were similar in hypertension, dyslipidaemia and diabetes mellitus (i.e., $p = 27.5\%$ vs 32% , 13.5% vs 12% and 6.6% vs 10% , respectively). Case identification of hypertension and diabetes mellitus depended on definitions that had been validated in other primary care databases and required only minor adaptations in the FIRE primary care database [25, 50]. Substantial underestimation appeared in obesity where case definitions are known to lack sensitivity ($p = 5.2\%$ vs 15%) [51]. On the other hand, we also found several higher prevalence estimates compared to Excoffier et al., most notably of chronic pain (20.2% vs 9%). Interestingly, similar primary care database-based studies also estimated the prevalence of chronic pain to be about 20% , potentially an indication of a systematic error inherent in case definitions for chronic pain [52–54].

Lastly, compared with another study by Tomonaga et al. in Swiss general practice, the prevalence of chronic kidney disease in our study was rather low ($p = 2.5\%$ vs 18%) [55]. Given this discrepancy, our case definition (which depends exclusively on laboratory test results) appears to suffer from poor sensitivity.

While the accuracy of prevalence estimates is questionable for most chronic conditions, our results mirrored known demographic patterns with remarkable precision. First, most prevalence estimates increased with age with the expected exception of migraine which is known to remit in the sixth decade, exactly as seen in our results [56]. And second, all sex differences in our study were consistent with prior knowledge (i.e. higher prevalence among females of chronic pain [57], mental disorders [foremost depression] [58], osteoarthritis [59], thyroid disorders [60], irritable bowel syndrome [61], migraine [62] and osteoporosis [63]; and higher prevalence among males of hypertension [64], dyslipidaemia [65], obstructive arteriosclerotic disease [66], diabetes mellitus [67] and gout [68]).

Assuming that the prevalence of frequent chronic diseases is distributed similarly among general practitioners in real

life, between-general practitioner-dispersion of prevalences as estimated in our study is another method for assessing plausibility of results and understanding methodological implications and risks of bias. In this respect, it is worth noting that the identification of frequent chronic conditions with the highest prevalence variabilities among general practitioners in our study (chronic kidney disease [qcod = 0.76], cancer [0.70] and obesity [0.59]) depended on case definitions that did not – or not predominantly – exploit the EMR-C MED. This observation suggests that well known variabilities between general practitioners in laboratory testing and documentation practice may translate into biased prevalence estimates in primary care database-based research [69–70]. In this context, the EMR-C MED may act as a moderator, provided that there is less dispersion among general practitioners in drug prescribing than in other activities reflected in EMR-Cs. This might not be the case for the prescription of pain medication or proton pump inhibitors, which are themselves subject to significant between-general practitioner variation in Switzerland [71–72]. Therefore, prevalence estimates of chronic pain and acidity-related stomach problems as measured in our study may have suffered from this source of bias.

Strengths and limitations

Our study advances understanding of using multiple EMR-Cs for case identification of chronic diseases in primary care databases. Our detailed analysis of the contributions of different EMR-Cs to case identification and of their relative importances for prevalence estimation may serve researchers as a look-up document for planning, implementing and evaluating similar approaches in other primary care databases, and for the critical appraisal of prevalence estimates gained from primary care databases using similar methods.

The main limitation of this study was the validation of case definitions: since anonymisation and ethical restrictions precluded manual review of the electronic medical records, the accuracy of our case definitions remains speculative. This applies even to case definitions validated in other primary care databases because documentation, data transfer and storage in the FIRE primary care database might differ from those of other primary care databases, and validation studies are not necessarily transferable. In the context of increasingly widespread research with big datasets, methods to explore and unravel specific areas of concern – as demonstrated in this article – are of growing importance. Newly introduced methodology intended to increase accuracy (such as the concept of permanent and reversible chronic diseases) has yet to be validated and can probably be improved. Our case definitions, however, are easy to replicate and implement in other primary care databases, which should facilitate future validation and optimisation.

Conclusions

Our analysis of the relative importances of commonly used EMR-Cs for case identification and of the resulting prevalence estimates has provided new insights into the strengths and weaknesses of case definitions applied to the FIRE primary care database. By far, most cases of chron-

ic conditions were identified via medication data, but clinical or laboratory parameters as well as ICPC-2 coded reasons for encounters were important for the identification of a subset of chronic conditions. The combination of all EMR-Cs produced a spectrum of prevalence estimates of varying concordance with external data but often underestimating presumed chronic condition prevalences in general practice. While sensitivity was the principal limitation to accuracy, our results matched known sex- and age-specific prevalence patterns, which points to the potential of the FIRE primary care database for the study of epidemiological trends of the more common chronic conditions considered in this study.

Availability of data and materials

The dataset analysed in the current study is not publicly available for reasons of data protection but is available from the corresponding author on reasonable request, as is all statistical code.

Acknowledgements

We thank Fabio Valeri for his help in implementing our case definitions in the FIRE primary care database and the FIRE study group of general practitioners for contributing data to this study.

Authors' contributions: RM: conceptualization; methodology; data curation; project administration; writing — original draft; writing — review & editing. TG: conceptualization, methodology; formal analysis and visualization; writing — original draft; writing — review & editing. YR: conceptualization; formal analysis and visualization; data curation; writing — review & editing. LJ: writing — review & editing. OS: writing — review & editing. TR: resources; writing — review & editing. JMB: project administration; writing — review & editing. SM: conceptualization; methodology; writing — original draft; writing — review & editing.

Financial disclosure

This research received no specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Potential competing interests

All authors have completed and submitted the International Committee of Medical Journal Editors form for disclosure of potential conflicts of interest. No potential conflict of interest was disclosed.

References

- Biro SC, Barber DT, Kotecha JA. Trends in the use of electronic medical records. *Can Fam Physician*. 2012 Jan;58(1):e21.
- Djalali S. Wer eHealth sucht, findet einen Haufen Papier [SÄZ]. *Schweiz Arzteztg*. 2015;96(43):1575–8.
- Quan H, Smith M, Bartlett-Esquilant G, Johansen H, Tu K, Lix L; Hypertension Outcome and Surveillance Team. Mining administrative health databases to advance medical science: geographical considerations and untapped potential in Canada. *Can J Cardiol*. 2012;28(2):152–4. <http://dx.doi.org/10.1016/j.cjca.2012.01.005>.
- Cowie MR, Blomster JJ, Curtis LH, Duclaux S, Ford I, Fritz F, et al. Electronic health records to facilitate clinical research. *Clin Res Cardiol*. 2017 Jan;106(1):1–9. <http://dx.doi.org/10.1007/s00392-016-1025-6>.
- Lawrenson R, Williams T, Farmer R. Clinical information for research; the use of general practice databases. *J Public Health Med*. 1999 Sep;21(3):299–304. <http://dx.doi.org/10.1093/pubmed/21.3.299>.
- Klompas M, McVetta J, Lazarus R, Eggleston E, Haney G, Kruskal BA, et al. Integrating clinical practice and public health surveillance using electronic medical record systems. *Am J Prev Med*. 2012 Jun;42(6 Suppl 2):S154–62. <http://dx.doi.org/10.1016/j.amepre.2012.04.005>.
- Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, van Staa T, et al. Data Resource Profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol*. 2015 Jun;44(3):827–36. <http://dx.doi.org/10.1093/ije/dyv098>.
- Blak BT, Thompson M, Dattani H, Bourke A. Generalisability of The Health Improvement Network (THIN) database: demographics, chronic disease prevalence and mortality rates. *Inform Prim Care*. 2011;19(4):251–5.
- Garies S, Birtwhistle R, Drummond N, Queenan J, Williamson T. Data Resource Profile: National electronic medical record data from the Canadian Primary Care Sentinel Surveillance Network (CPCSSN). *Int J Epidemiol*. 2017 Aug;46(4):1091–1092f. <http://dx.doi.org/10.1093/ije/dyw248>.
- Ramos R, Balló E, Marrugat J, Elosua R, Sala J, Grau M, et al. Validity for use in research on vascular diseases of the SIDIAP (Information System for the Development of Research in Primary Care): the EMMA study. *Rev Esp Cardiol (Engl Ed)*. 2012 Jan;65(1):29–37. <http://dx.doi.org/10.1016/j.rec.2011.07.016>.
- Cricelli C, Mazzaglia G, Samani F, Marchi M, Sabatini A, Nardi R, et al. Prevalence estimates for chronic diseases in Italy: exploring the differences between self-report and primary care databases. *J Public Health Med*. 2003 Sep;25(3):254–7. <http://dx.doi.org/10.1093/pubmed/fdg060>.
- Federal Office of Public Health. Swiss primary care doctors give their healthcare system highest marks in international comparison. Bern; 2023.[Available from: <https://www.edi.admin.ch/edi/en/home/dokumentation/medienmitteilungen.html.msg-id-93048.html>].
- FMH Consulting Services AG. Softwarekatalog 2023. Oberkirch; 2023.[Available from: <https://www.fmhservices.ch/softwarekatalog/>].
- Chmiel C, Bhend H, Senn O, Zoller M, Rosemann T; FIRE study-group. The FIRE project: a milestone for research in primary care in Switzerland. *Swiss Med Wkly*. 2011 Jan;140:w13142.
- Shivade C, Raghavan P, Fosler-Lussier E, Embi PJ, Elhadad N, Johnson SB, et al. A review of approaches to identifying patient phenotype cohorts using electronic health records. *J Am Med Inform Assoc*. 2014;21(2):221–30. <http://dx.doi.org/10.1136/amiajnl-2013-001935>.
- Singer A, Yakubovich S, Krocker AL, Dufault B, Duarte R, Katz A. Data quality of electronic medical records in Manitoba: do problem lists accurately reflect chronic disease billing diagnoses? *J Am Med Inform Assoc*. 2016 Nov;23(6):1107–12. <http://dx.doi.org/10.1093/jamia/ocw013>.
- McBrien KA, Souri S, Symonds NE, Rouhi A, Lethebe BC, Williamson TS, et al. Identification of validated case definitions for medical conditions used in primary care electronic medical record databases: a systematic review. *J Am Med Inform Assoc*. 2018 Nov;25(11):1567–78. <http://dx.doi.org/10.1093/jamia/ocy094>.
- Orueta JF, Nuño-Solinis R, Mateos M, Vergara I, Grandes G, Esnaola S. Monitoring the prevalence of chronic conditions: which data should we use? *BMC Health Serv Res*. 2012 Oct;12(1):365. <http://dx.doi.org/10.1186/1472-6963-12-365>.
- WHO Collaborating Centre for Drug Statistics Methodology. Guidelines for ATC classification and DDD assignment. Oslo; 2021.
- Classification Committee of the World Organization of Family Doctors (WICC). ICPC-2: International Classification of Primary Care. Oxford University Press; 1997.
- Valeri F, Burgstaller JM. FIRE5-Project: Institute of Primary Care. Zurich, Switzerland: University of Zurich and University Hospital Zurich; 2023.[Available from: <https://github.com/ihamzurich/FIRE5>].
- N'Goran AA, Blaser J, Deruaz-Luyet A, Senn N, Frey P, Haller DM, et al. From chronic conditions to relevance in multimorbidity: a four-step study in family medicine. *Fam Pract*. 2016 Aug;33(4):439–44. <http://dx.doi.org/10.1093/fampra/cmw030>.
- Barnett K, Mercer SW, Norbury M, Watt G, Wyke S, Guthrie B. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. *Lancet*. 2012 Jul;380(9836):37–43. [http://dx.doi.org/10.1016/S0140-6736\(12\)60240-2](http://dx.doi.org/10.1016/S0140-6736(12)60240-2).
- de Burgos-Lunar C, Salinero-Fort MA, Cárdenas-Valladolid J, Soto-Díaz S, Fuentes-Rodríguez CY, Abánades-Herranz JC, et al. Validation of diabetes mellitus and hypertension diagnosis in computerized medical records in primary health care. *BMC Med Res Methodol*. 2011 Oct;11(1):146. <http://dx.doi.org/10.1186/1471-2288-11-146>.
- Kadhim-Saleh A, Green M, Williamson T, Hunter D, Birtwhistle R. Validation of the diagnostic algorithms for 5 chronic conditions in the Canadian Primary Care Sentinel Surveillance Network (CPCSSN): a Kingston Practice-based Research Network (PBRN) report. *J Am Board Fam Med*. 2013;26(2):159–67. <http://dx.doi.org/10.3122/jabfm.2013.02.120183>.
- Bill M, Meyer D, Telsler H, Stämpfli D, Hersberger K, Schwenkgenks M. Aktualisierung der PCG-Listen für den Schweizer Risikoausgleich 2019 [updated 22.01.2019. Available from: <https://www.bag.admin.ch/bag/de/home/versicherungen/krankenversicherung/krankenversicherung-versicherer-aufsicht/risikoausgleich.html>
- swissmedic. Erweiterte Arzneimittelliste (Listen und Verzeichnisse, 1. Humanarzneimittel) Internet [Available from:

- https://www.swissmedic.ch/swissmedic/de/home/services/listen_neu.html#-257211596
28. Brabant G, Beck-Peccoz P, Jarzab B, Laurberg P, Orgiazzi J, Szabolcs I, et al. Is there a need to redefine the upper normal limit of TSH? *J European Journal of Endocrinology eur j endocrinol.* 2006;154(5):633.
 29. WHO Committee. Physical status: The use and interpretation of anthropometry. 1995.
 30. Cosentino F, Grant PJ, Aboyans V, Bailey CJ, Ceriello A, Delgado V, et al.; ESC Scientific Document Group. 2019 ESC Guidelines on diabetes, pre-diabetes, and cardiovascular diseases developed in collaboration with the EASD. *Eur Heart J.* 2020 Jan;41(2):255–323. <http://dx.doi.org/10.1093/eurheartj/ehz486>.
 31. Mancia G, Fagard R, Narkiewicz K, Redon J, Zanchetti A, Böhm M, et al. 2013 ESH/ESC Guidelines for the management of arterial hypertension: The Task Force for the management of arterial hypertension of the European Society of Hypertension (ESH) and of the European Society of Cardiology (ESC). *Eur Heart J.* 2013 Jul;34(28):2159–219. <http://dx.doi.org/10.1093/eurheartj/ehz151>.
 32. Williams B, Mancia G, Spiering W, Agabiti Rosei E, Azizi M, Burnier M, et al.; ESC Scientific Document Group. 2018 ESC/ESH Guidelines for the management of arterial hypertension. *Eur Heart J.* 2018 Sep;39(33):3021–104. <http://dx.doi.org/10.1093/eurheartj/ehy339>.
 33. Kidney Disease: Improving Global Outcomes (KDIGO) 2012 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease. *Kidney Int Suppl.* 2013;3(1).
 34. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2020.
 35. Larsson J. eulerr: Area-Proportional Euler and Venn Diagrams with Ellipses. R package version 6.1.1 2021. [Available from: <https://CRAN.R-project.org/package=eulerr>].
 36. Chini F, Pezzotti P, Orzella L, Borgia P, Guasticchi G. Can we use the pharmacy data to estimate the prevalence of chronic conditions? a comparison of multiple data sources. *BMC Public Health.* 2011 Sep;11(1):688. <http://dx.doi.org/10.1186/1471-2458-11-688>.
 37. Huber CA, Szucs TD, Rapold R, Reich O. Identifying patients with chronic conditions using pharmacy data in Switzerland: an updated mapping approach to the classification of medications. *BMC Public Health.* 2013 Oct;13(1):1030. <http://dx.doi.org/10.1186/1471-2458-13-1030>.
 38. Kolber M, Sharif N, Marceau R, Szafran O. Family practice physicians' use of acetylsalicylic acid for cardiovascular disease prevention. *Can Fam Physician.* 2013 Jan;59(1):55–61.
 39. Muheim L, Signorell A, Markun S, Chmiel C, Neuner-Jehle S, Blozik E, et al. Potentially inappropriate proton-pump inhibitor prescription in the general population: a claims-based retrospective time trend analysis. *Therap Adv Gastroenterol.* 2021 Apr;14:1756284821998928. <http://dx.doi.org/10.1177/1756284821998928>.
 40. Stewart A, Ferguson C. Towards evidence-based emergency medicine: best BETs from the Manchester Royal Infirmary. BET 4: Alpha blockers v calcium blockers to increase spontaneous passage of renal calculi. *Emerg Med J.* 2013 Feb;30(2):168–9. <http://dx.doi.org/10.1136/emered-2012-202190.5>.
 41. Mayer S, Österle A. Socioeconomic determinants of prescribed and non-prescribed medicine consumption in Austria. *Eur J Public Health.* 2015 Aug;25(4):597–603. <http://dx.doi.org/10.1093/eurpub/cku179>.
 42. File D, Davidovich N, Novack L, Balicer RD. Is socioeconomic status associated with utilization of health care services in a single-payer universal health care system? *Int J Equity Health.* 2014 Nov;13(1):115. <http://dx.doi.org/10.1186/s12939-014-0115-1>.
 43. Sollie A, Roskam J, Sijmons RH, Numans ME, Helsper CW. Do GPs know their patients with cancer? Assessing the quality of cancer registration in Dutch primary care: a cross-sectional validation study. *BMJ Open.* 2016 Sep;6(9):e012669. <http://dx.doi.org/10.1136/bmjopen-2016-012669>.
 44. Jordan K, Porcheret M, Croft P. Quality of morbidity coding in general practice computerized medical records: a systematic review. *Fam Pract.* 2004 Aug;21(4):396–412. <http://dx.doi.org/10.1093/fampra/cmh409>.
 45. Yau MS, Dubreuil M, Li S, Inamdar V, Peloquin C, Felson DT. Validation of knee osteoarthritis case identification algorithms in a large electronic health record database. *Osteoarthr Cartil Open.* 2022 Mar;4(1):100229. <http://dx.doi.org/10.1016/j.ocarto.2021.100229>.
 46. Napel HT, van Boven K, Olagundoye OA, van der Haring E, Verbeke M, Härkönen M, et al. Improving Primary Health Care Data With ICPC-3: From a Medical to a Person-Centered Perspective. *Ann Fam Med.* 2022;20(4):358–61. <http://dx.doi.org/10.1370/afm.2830>.
 47. Rose SA, Turchin A, Grant RW, Meigs JB. Documentation of body mass index and control of associated risk factors in a large primary care network. *BMC Health Serv Res.* 2009 Dec;9(1):236. <http://dx.doi.org/10.1186/1472-6963-9-236>.
 48. Baer HJ, Karson AS, Soukup JR, Williams DH, Bates DW. Documentation and diagnosis of overweight and obesity in electronic health records of adult primary care patients. *JAMA Intern Med.* 2013 Sep;173(17):1648–52. <http://dx.doi.org/10.1001/jamainternmed.2013.7815>.
 49. Excoffier S, Herzig L, N'Goran AA, Déruaz-Luyet A, Haller DM. Prevalence of multimorbidity in general practice: a cross-sectional study within the Swiss Sentinel Surveillance System (Sentinella). *BMJ Open.* 2018 Mar;8(3):e019616. <http://dx.doi.org/10.1136/bmjopen-2017-019616>.
 50. Tu K, Manuel D, Lam K, Kavanagh D, Mitiku TF, Guo H. Diabetics can be identified in an electronic medical record using laboratory tests and prescriptions. *J Clin Epidemiol.* 2011 Apr;64(4):431–5. <http://dx.doi.org/10.1016/j.jclinepi.2010.04.007>.
 51. Mattar A, Carlston D, Sariol G, Yu T, Almustafa A, Melton GB, et al. The prevalence of obesity documentation in Primary Care Electronic Medical Records. Are we acknowledging the problem? *Appl Clin Inform.* 2017 Jan;8(1):67–79.
 52. González-Chica DA, Vanlint S, Hoon E, Stocks N. Epidemiology of arthritis, chronic back pain, gout, osteoporosis, spondyloarthropathies and rheumatoid arthritis among 1.5 million patients in Australian general practice: NPS MedicineWise MedicineInsight dataset. *BMC Musculoskelet Disord.* 2018 Jan;19(1):20. <http://dx.doi.org/10.1186/s12891-018-1941-x>.
 53. Henderson JV, Harrison CM, Britt HC, Bayram CF, Miller GC. Prevalence, causes, severity, impact, and management of chronic pain in Australian general practice patients. *Pain Med.* 2013 Sep;14(9):1346–61. <http://dx.doi.org/10.1111/pme.12195>.
 54. Tian TY, Zlateva I, Anderson DR. Using electronic health records data to identify patients with chronic pain in a primary care setting. *J Am Med Inform Assoc.* 2013 Dec;20 e2:e275–80. <http://dx.doi.org/10.1136/amiajnl-2013-001856>.
 55. Tomonaga Y, Risch L, Szucs TD, Ambühl PM. The prevalence of chronic kidney disease in a primary care setting: a Swiss cross-sectional study. *PLoS one.* 2013;8(7):e67848-e.
 56. Kelman L. Migraine changes with age: IMPACT on migraine classification. *Headache.* 2006;46(7):1161–71. <http://dx.doi.org/10.1111/j.1526-4610.2006.00444.x>.
 57. Bartley EJ, Fillingim RB. Sex differences in pain: a brief review of clinical and experimental findings. *Br J Anaesth.* 2013 Jul;111(1):52–8. <http://dx.doi.org/10.1093/bja/aet127>.
 58. Lim GY, Tam WW, Lu Y, Ho CS, Zhang MW, Ho RC. Prevalence of Depression in the Community from 30 Countries between 1994 and 2014. *Sci Rep.* 2018 Feb;8(1):2861. <http://dx.doi.org/10.1038/s41598-018-21243-x>.
 59. O'Connor MI. Sex differences in osteoarthritis of the hip and knee. *J Am Acad Orthop Surg.* 2007;15 Suppl 1:S22–5. <http://dx.doi.org/10.5435/00124635-200700001-00007>.
 60. Bauer M, Glenn T, Pilhatsch M, Pfennig A, Whybrow PC. Gender differences in thyroid system function: relevance to bipolar disorder and its treatment. *Bipolar Disord.* 2014 Feb;16(1):58–71. <http://dx.doi.org/10.1111/bdi.12150>.
 61. Canavan C, West J, Card T. The epidemiology of irritable bowel syndrome. *Clin Epidemiol.* 2014 Feb;6:71–80.
 62. Victor TW, Hu X, Campbell JC, Buse DC, Lipton RB. Migraine prevalence by age and sex in the United States: a life-span study. *Cephalalgia.* 2010 Sep;30(9):1065–72. <http://dx.doi.org/10.1177/0333102409355601>.
 63. Schuit SC, van der Klift M, Weel AE, de Laet CE, Burger H, Seeman E, et al. Fracture incidence and association with bone mineral density in elderly men and women: the Rotterdam Study. *Bone.* 2004 Jan;34(1):195–202. <http://dx.doi.org/10.1016/j.bone.2003.10.001>.
 64. Yoon SS, Gu Q, Nwankwo T, Wright JD, Hong Y, Burt V. Trends in blood pressure among adults with hypertension: United States, 2003 to 2012. *Hypertension (Dallas, Tex : 1979).* 2015;65(1):54–61.
 65. O'Meara JG, Kardia SL, Armon JJ, Brown CA, Boerwinkle E, Turner ST. Ethnic and sex differences in the prevalence, treatment, and control of dyslipidemia among hypertensive adults in the GENOA study. *Arch Intern Med.* 2004 Jun;164(12):1313–8. <http://dx.doi.org/10.1001/archinte.164.12.1313>.
 66. Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, et al.; Writing Group Members; American Heart Association Statistics Committee; Stroke Statistics Subcommittee. Heart Disease and Stroke Statistics-2016 Update: A Report From the American Heart Association. *Circulation.* 2016 Jan;133(4):e38–360. <http://dx.doi.org/10.1161/CIR.0000000000000350>.
 67. Zhou B, Lu Y, Hajifathalian K, Bentham J, Di Cesare M, Danaei G, et al.; NCD Risk Factor Collaboration (NCD-RisC). Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies

- with 4.4 million participants. *Lancet*. 2016 Apr;387(10027):1513–30. [http://dx.doi.org/10.1016/S0140-6736\(16\)00618-8](http://dx.doi.org/10.1016/S0140-6736(16)00618-8).
68. Harrold LR, Yood RA, Mikuls TR, Andrade SE, Davis J, Fuller J, et al. Sex differences in gout epidemiology: evaluation and treatment. *Ann Rheum Dis*. 2006 Oct;65(10):1368–72. <http://dx.doi.org/10.1136/ard.2006.051649>.
 69. Schumacher LD, Jäger L, Meier R, Rachamin Y, Senn O, Rosemann T, et al. Trends and Between-Physician Variation in Laboratory Testing: A Retrospective Longitudinal Study in General Practice. *J Clin Med*. 2020 Jun;9(6):1787. <http://dx.doi.org/10.3390/jcm9061787>.
 70. Cohen GR, Friedman CP, Ryan AM, Richardson CR, Adler-Milstein J. Variation in Physicians' Electronic Health Record Documentation and Potential Patient Harm from That Variation. *J Gen Intern Med*. 2019 Nov;34(11):2355–67. <http://dx.doi.org/10.1007/s11606-019-05025-3>.
 71. Muheim L, Signorell A, Markun S, Chmiel C, Neuner-Jehle S, Blozik E, et al. Potentially inappropriate proton-pump inhibitor prescription in the general population: a claims-based retrospective time trend analysis. 2021;14:1756284821998928. <http://dx.doi.org/10.1177/1756284821998928>.
 72. Rachamin Y, Jäger L, Meier R, Grischott T, Senn O, Burgstaller JM, et al. Prescription Rates, Polypharmacy and Prescriber Variability in Swiss General Practice-A Cross-Sectional Database Study. *Front Pharmacol*. 2022 Feb;13:832994. <http://dx.doi.org/10.3389/fphar.2022.832994>.

Appendix

Supplementary table 1: Case definition rules

Case definition rules for chronic conditions					
<p>Abbreviations: ICPC-2¹: International Classification of Primary Care 2nd edition ATC²: Anatomical chemical therapeutic codes GTIN³: Global Trade Item Number PCG⁴: Pharmaceutical cost groups GFR: Glomerular filtration rate</p>					
<p>Explanations: <u>Overarching concept</u> determines whether and which chronic conditions have been aggregated in this case definition <u>Disease permanence</u> determines whether a chronic condition is permanent or reversible. Reversible chronic conditions are only identified if according case defining information appeared within 6 months prior to the measurement in this study. Permanent chronic conditions are identified without restriction during the total available observation time.</p>					
<p>References 1) Classification Committee of the World Organization of Family Doctors (WICC). ICPC-2: International Classification of Primary Care (1997). Oxford University Press. 2) WHO Collaborating Centre for Drug Statistics Methodology, Guidelines for ATC classification and DDD assignment, 2021. Oslo. 3) https://www.gs1.org/ 4) Huber CA, et al. Identifying patients with chronic conditions using pharmacy data in Switzerland: an updated mapping approach to the classification of medications. BMC public health. 2013;13(1):1030.</p>					
			EMR-C criteria for case identification		
Number	Chronic Condition (overarching concept)	Disease permanence	Medication (ATC codes or GTIN specified by Swiss PCG)	Clinical or laboratory parameters	Reasons for encounters (ICPC-2 codes)
1	Hypertension	permanent	PCG: Hypertonie	two elevated BP measurements at an interval of ≥ 1 week and ≤ 6 months where elevated = (syst ≥ 140 mmHg or diast ≥ 90 mmHg) or	K85 or K86 or K87

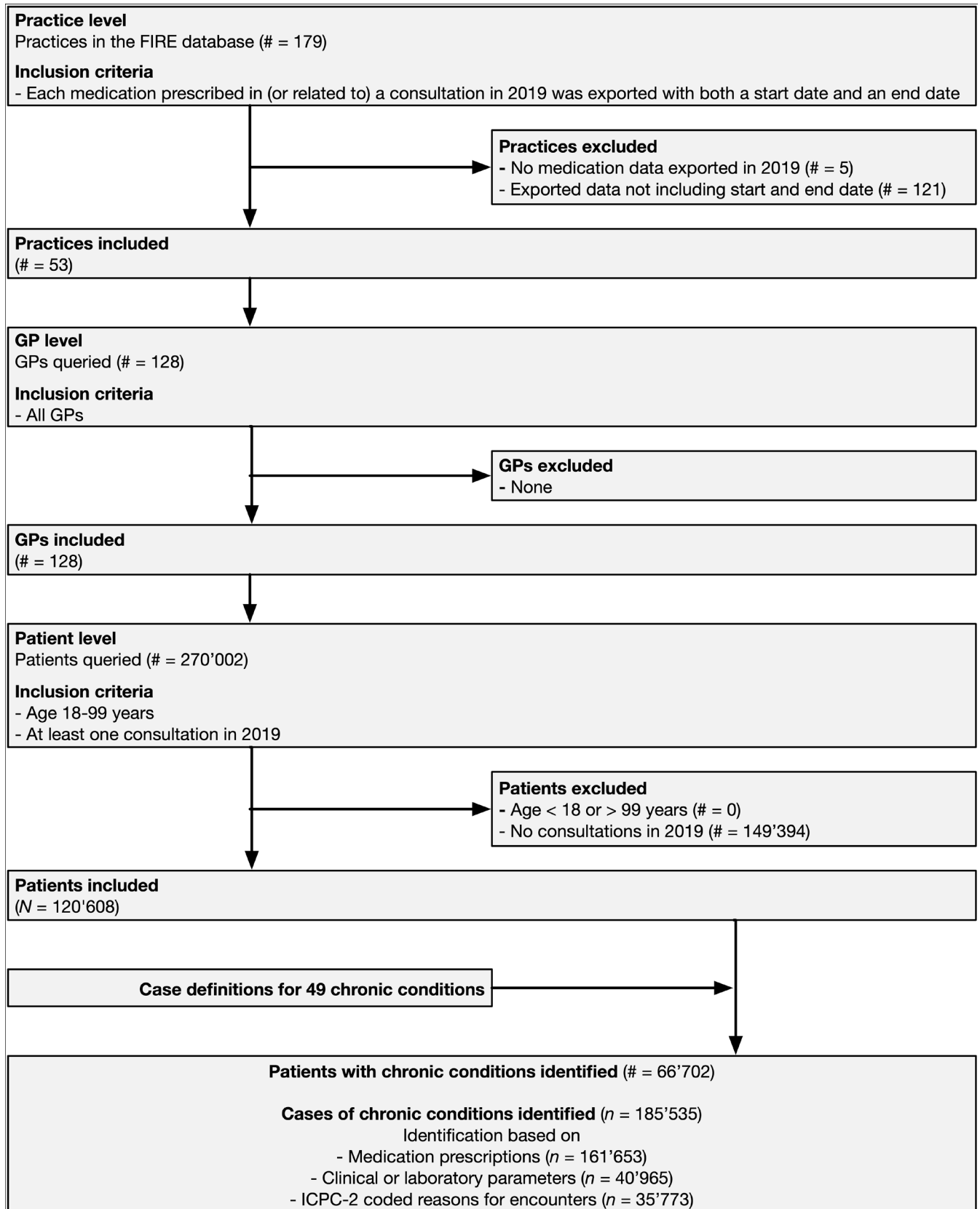
				one extreme BP measurement where extreme = (syst \geq 180 mmHg or diast \geq 110 mmHg)	
2	Chronic pain	reversible	ATC = M01A or N02A or N02B		A01
3	Dyslipidemia	permanent	ATC = C10	two measurements of: triglyceride > 1.7 mmol/l or total cholesterol > 4.9 mmol/l or LDL-cholesterol > 3 mmol/l or (sex=female and HDL-cholesterol \leq 1.2 mmol/l) or (sex=male and HDL-cholesterol \leq 1 mmol/l)	T93
4	Acidity related stomach problems	reversible	ATC = A02A or A02BA or A02BC		D07 or D86 or D03 or D85
5	Mental disorder (including depressive, psychotic, and anxiety disorder)	reversible	ATC = N05A or N05B or N06A		P72 or P73 or P74 or P75 or P79 or P82 or P98
6	Obstructive atherosclerotic disease (including coronary, cerebral or peripheral arteries)	permanent	ATC = B01AC04 or B01AC06 or B01AC07 or B01AC22 or B01AC24 or B01AC25 or B01AC56		K74 or K75 or K76 or K89 or K90 or K91 or K92
7	Obstructive lung disease (including asthma, COPD, and chronic bronchitis)	permanent	ATC = R03		R95 or R96 or R79
8	Diabetes mellitus (including insulin dependent and non insulin dependent)	reversible	ATC = A10 and ATC \neq A10BJ	Diabetes: 2 consecutive HbA1c \geq 6.5 Non-Diabetes: 2 consecutive HbA1c < 6.5 and neither ICPC nor ATC first date of these two is start/stop date.	T89 or T90
9	Osteoarthritis	permanent	ATC = M01AX25 or M09AX01		L89 or L90 or L91

10	Obesity	reversible	ATC = A08AB	one measurement of BMI \geq 30	T82
11	Chronic constipation	reversible	ATC = A06A and ATC \neq A06AX04		D12
12	Thyroid disease	reversible	ATC = H03	two measurements of TSH $<$ 0.4 or TSH $>$ 4	T85 or T86
13	Benign prostatic hyperplasia	permanent	ATC = G04C		Y85
14	Chronic kidney disease	permanent		two decreased GFR measurements at an interval of $>$ 3 months and $<$ 2 years where decreased = (GFR $<$ 60 ml/min)	
15	Irritable bowel syndrome	permanent	ATC = A06AX04 or A03AA05 or (GTIN = 7680478270438 or 7680478270513 or 7680478270780 or 7680006750012 or 7680006770010 or 7680301970115 or 7680301970382)		D93
16	Migraine	permanent	ATC = N02C		N89
17	Gout	permanent	ATC = M04A		T92
18	Dementia	permanent	ATC = N06D		P70
19	Osteoporosis	permanent	ATC = M05BA or M05BB or M05BX06 or A14AB01 and ATC \neq M05BA08		L95
20	Epilepsy	permanent	PCG: Epilepsie		N88
21	Heart disease (including arrhythmias or congestive disease)	permanent	ATC = C01AA05 or C01BC03 or C01BC04 or C01BD01 or C01BD07		K77 or K78

22	Cancer	reversible	PCG: Hormonsensitive Tumore or Krebs or Krebs komplex		A79 or B72 or B73 or B74 or B75 or D74 or D75 or D76 or D77 or D78 or F74 or H75 or K72 or L71 or N74 or N75 or N76 or R84 or R85 or R86 or R92 or S77 or S78 or S79 or T71 or T72 or T73 or U75 or U76 or U77 or U78 or U79 or W72 or W73 or X75 or X76 or X77 or X78 or X79 or X80 or X81 or Y77 or Y78 or Y79
23	Parkinson's disease	permanent	ATC = N04		N87
24	Glaucoma	reversible	ATC = S01E		F93
25	Autoimmune rheumatic diseases (including rheumatoid arthritis)	Permanent	PCG Rheuma and ATC ≠ A07EC01		L88
26	Urinary incontinence	reversible	ATC = G04BD		U04
27	Inflammatory bowel disease	permanent	ATC = A07EC01 or A07EC02 or A07EA06 or GTIN = 7680153030265 and drug name ≠ JORVEZA and ≠ ASAZINE		D94
28	HIV/AIDS	permanent	PCG = HIV/AIDS		B90
29	Liver disease	permanent			D97
30	Multiple sclerosis	permanent	PCG: Krankheiten des Gehirns oder des Rückenmarks: Multiple Sklerose		N86
31	Medication/drug abuse	reversible	ATC = N07BC		P18 or P19
32	Anorexia/bulimia nervosa	permanent		one measurement of BMI ≤ 17 when age ≤ 30	P86
33	Psoriasis	permanent			S91
34	Viral hepatitis	reversible	ATC = J05AB04 or J05AF1[013] or J05AP or L03AB11 or GTIN = 7680548680143 or 7680548680228 or 7680548690104		D72
35	Macular degeneration	permanent	ATC = S01LA		F84
36	Alcohol abuse	reversible	ATC = N07BB and ATC ≠ N07BB04		P15

37	Presbycusis/deafness	permanent			H84 or H86
38	Chronic cutaneous ulcer	reversible			S97
39	Secondary effect of trauma	reversible			A82
40	Bowel incontinence	reversible			D17
41	Retinopathy	permanent			F83
42	Blindness	permanent			F94
43	Pulmonary heart disease	permanent	PCG: Pulmonale (arterielle) Hypertonie		K82
44	Poliomyelitis	permanent			N70
45	Trigeminal neuralgia	reversible			N92
46	Peripheral neuritis/neuropathy	reversible			N94
47	Learning disability	permanent			P24
48	Personality disorder	permanent			P80
49	Intellectual disability	permanent			P85

Supplementary figure 1: Study flowchart



Supplementary table 2: Case numbers, prevalence estimates (based on $N=120'608$ patients), and EMR-C importance metrics

Chronic condition	Case numbers								Prevalence estimate		Importance metrics					
	m	l	r	ml	mr	lr	mlr	n	p (%)	CI (%)	MED		CLP		RFE	
											c (%)	c_e (%)	c (%)	c_e (%)	c (%)	c_e (%)
Any chronic disease	121252	12764	8806	15746	14512	2312	10143	185535	55.3	55.0-55.6	87.1	65.4	22.1	6.88	19.3	4.75
Hypertension	10312	2706	299	9967	2005	745	7161	33195	27.5	27.3-27.8	88.7	31.1	62.0	8.15	30.8	0.90
Chronic pain	24046	0	63	0	227	0	0	24336	20.2	20.0-20.4	99.7	98.8	0	0	1.19	0.26
Dyslipidemia	9941	882	839	2180	1663	158	631	16294	13.5	13.3-13.7	88.5	61.0	23.6	5.41	20.2	5.15
Acidity related stomach problems	14380	0	106	0	405	0	0	14891	12.3	12.2-12.5	99.3	96.6	0	0	3.43	0.71
Mental disorder ^a	11059	0	137	0	1116	0	0	12312	10.2	10.0-10.4	98.9	89.8	0	0	10.2	1.11
Obstructive atherosclerotic disease ^b	9474	0	301	0	2166	0	0	11941	9.90	9.73-10.1	97.5	79.3	0	0	20.7	2.52
Obstructive lung disease ^c	8857	0	236	0	2041	0	0	11134	9.23	9.07-9.40	97.9	79.5	0	0	20.5	2.12
Diabetes mellitus	1585	1102	241	2325	481	105	2098	7937	6.58	6.44-6.72	81.8	20.0	70.9	13.9	36.9	3.04
Obesity	28	4675	229	53	2	1279	35	6301	5.22	5.10-5.35	1.87	0.44	95.9	74.2	24.5	3.63
Osteoarthritis	4213	0	1365	0	680	0	0	6258	5.19	5.07-5.32	78.2	67.3	0	0	32.7	21.8
Chronic constipation	4635	0	37	0	287	0	0	4959	4.11	4.00-4.22	99.3	93.5	0	0	6.53	0.75
Thyroid disease	2414	192	59	1221	437	15	218	4556	3.78	3.67-3.89	94.2	53.0	36.1	4.21	16.0	1.29
Benign prostatic hyperplasia	3716	0	232	0	417	0	0	4365	3.62	3.52-3.73	94.7	85.1	0	0	14.9	5.32
Chronic kidney disease	0	3011	0	0	0	0	0	3011	2.50	2.41-2.59	0	0	100	100	0	0
Irritable bowel syndrome	2382	0	138	0	106	0	0	2626	2.18	2.10-2.26	94.7	90.7	0	0	9.29	5.26
Migraine	1841	0	222	0	293	0	0	2356	1.95	1.88-2.03	90.6	78.1	0	0	21.9	9.42
Gout	1686	0	264	0	345	0	0	2295	1.90	1.83-1.98	88.5	73.5	0	0	26.5	11.5
Dementia	1821	0	192	0	186	0	0	2199	1.82	1.75-1.90	91.3	82.8	0	0	17.2	8.73
Osteoporosis	982	0	376	0	446	0	0	1804	1.50	1.43-1.57	79.2	54.4	0	0	45.6	20.8
Epilepsy	1571	0	24	0	185	0	0	1780	1.48	1.41-1.54	98.7	88.3	0	0	11.7	1.35
Heart disease ^d	746	0	644	0	126	0	0	1516	1.26	1.20-1.32	57.5	49.2	0	0	50.8	42.5
Cancer	468	0	868	0	114	0	0	1450	1.20	1.14-1.26	40.1	32.3	0	0	67.7	59.9
Parkinson's disease	1249	0	10	0	117	0	0	1376	1.14	1.08-1.20	99.3	90.8	0	0	9.23	0.73
Glaucoma	978	0	17	0	61	0	0	1056	0.88	0.83-0.93	98.4	92.6	0	0	7.39	1.61
Autoimmune rheumatic diseases	729	0	130	0	158	0	0	1017	0.84	0.79-0.90	87.2	71.7	0	0	28.3	12.8
Urinary incontinence	818	0	30	0	29	0	0	877	0.73	0.68-0.78	96.6	93.3	0	0	6.73	3.42
Inflammatory bowel disease	595	0	49	0	97	0	0	741	0.61	0.57-0.66	93.4	80.3	0	0	19.7	6.61
HIV/AIDS	320	0	4	0	193	0	0	517	0.43	0.39-0.47	99.2	61.9	0	0	38.1	0.77
Liver disease	0	0	512	0	0	0	0	512	0.42	0.39-0.46	0	0	0	0	100	100
Multiple sclerosis	179	0	41	0	51	0	0	271	0.22	0.20-0.25	84.9	66.1	0	0	33.9	15.1
Medication/drug abuse	104	0	127	0	39	0	0	270	0.22	0.20-0.25	53.0	38.5	0	0	61.5	47.0
Presbycusis/deafness	0	0	245	0	0	0	0	245	0.20	0.18-0.23	0	0	0	0	100	100
Psoriasis	0	0	242	0	0	0	0	242	0.20	0.18-0.23	0	0	0	0	100	100
Anorexia/bulimia nervosa	0	196	19	0	0	10	0	225	0.19	0.16-0.21	0	0	91.6	87.1	12.9	8.44
Alcohol abuse	45	0	104	0	27	0	0	176	0.15	0.13-0.17	40.9	25.6	0	0	74.4	59.1
Macular degeneration	28	0	66	0	2	0	0	96	0.08	0.07-0.10	31.3	29.2	0	0	70.8	68.8
Peripheral neuritis/neuropathy	0	0	86	0	0	0	0	86	0.07	0.06-0.09	0	0	0	0	100	100
Viral hepatitis	26	0	36	0	5	0	0	67	0.06	0.04-0.07	46.3	38.8	0	0	61.2	53.7
Personality disorder	0	0	59	0	0	0	0	59	0.05	0.04-0.06	0	0	0	0	100	100
Pulmonary heart disease	24	0	18	0	5	0	0	47	0.04	0.03-0.05	61.7	51.1	0	0	48.9	38.3
Retinopathy	0	0	39	0	0	0	0	39	0.03	0.02-0.04	0	0	0	0	100	100
Chronic skin ulcer	0	0	32	0	0	0	0	32	0.03	0.02-0.04	0	0	0	0	100	100
Intellectual disability	0	0	21	0	0	0	0	21	0.02	0.01-0.03	0	0	0	0	100	100
Blindness	0	0	13	0	0	0	0	13	0.01	0.01-0.02	0	0	0	0	100	100
Trigeminal neuralgia	0	0	11	0	0	0	0	11	0.01	0.01-0.02	0	0	0	0	100	100
Learning disability	0	0	9	0	0	0	0	9	0.01	0.00-0.01	0	0	0	0	100	100
Secondary effect of trauma	0	0	5	0	0	0	0	5	0.00	0.00-0.01	0	0	0	0	100	100
Poliomyelitis	0	0	5	0	0	0	0	5	0.00	0.00-0.01	0	0	0	0	100	100
Bowel incontinence	0	0	4	0	0	0	0	4	0.00	0.00-0.01	0	0	0	0	100	100

Case numbers m , l , r , ml , mr , lr , mlr and prevalence p , relative contribution c and relative exclusive contribution c_e as defined in the methods section (Figure 1). CI : 95% Wilson confidence interval.

^a including depression, psychotic disorders, and anxiety disorders

^b including coronary, cerebral, and peripheral arteries

^c including asthma, chronic obstructive pulmonary disease, and chronic bronchitis

^d including arrhythmias and congestive heart disease

Supplementary table 3: Prevalence estimates by sex and age group (based on N =120'608 patients), and variability of overall prevalence estimates over GPs

Chronic condition	Prevalence estimates p (%)								Variability over GPs (of unstratified prevalences in %)			
	18-40 years		41-64 years		65-80 years		81-99 years		q _{0.25}	median	q _{0.75}	qcod
	f	m	f	m	f	m	f	m				
All chronic diseases	32.4	29.8	56.5	58.3	77.8	80.5	82.2	82.2	47.2	56.4	65.5	0.16
Hypertension	2.46	4.17	20.1	28.8	52.4	60.2	68.1	67.3	19.6	27.7	33.7	0.27
Chronic pain	12.3	10.4	23.0	17.7	32.1	24.0	38.7	27.5	15.0	19.2	26.7	0.28
Dyslipidemia	0.50	0.92	6.90	14.9	26.8	39.0	25.8	36.4	8.24	12.1	17.2	0.35
Acidity related stomach problems	4.37	4.53	12.5	12.3	21.6	20.4	22.9	20.4	8.75	11.9	15.5	0.28
Mental disorder ^a	5.26	4.25	13.3	8.81	16.9	9.30	24.0	13.6	7.84	10.0	12.4	0.23
Obstructive atherosclerotic disease ^b	0.25	0.29	3.09	8.24	17.4	29.8	32.7	39.8	5.89	8.84	12.9	0.37
Obstructive lung disease ^c	7.23	6.51	9.52	8.74	12.3	12.1	11.0	12.7	6.21	8.42	11.5	0.30
Diabetes mellitus	1.03	1.07	4.20	7.40	10.9	16.8	12.8	17.5	4.19	5.93	7.91	0.31
Obesity	2.47	3.84	5.19	6.76	7.10	7.93	4.28	4.57	1.95	3.59	7.60	0.59
Osteoarthritis	0.39	0.42	5.57	3.59	13.6	9.31	12.3	9.90	2.15	3.44	7.04	0.53
Chronic constipation	1.75	1.08	3.69	2.19	7.30	5.73	15.7	12.1	2.44	3.94	5.51	0.39
Thyroid disease	2.55	0.42	6.20	1.26	9.25	2.63	9.14	3.91	2.89	3.81	4.63	0.23
Benign prostatic hyperplasia	0.05	0.41	0.09	4.52	0.10	18.7	0.31	25.9	1.81	2.83	4.62	0.44
Chronic kidney disease	0.03	0.06	0.41	0.40	4.88	4.80	16.8	15.9	0.53	2.04	3.84	0.76
Irritable bowel syndrome	2.10	1.08	2.81	1.55	3.80	1.84	3.31	1.68	0.84	1.66	2.97	0.56
Migraine	2.96	0.90	4.09	1.08	1.84	0.71	0.49	0.21	1.06	1.93	2.60	0.42
Gout	0.04	0.30	0.28	2.42	1.67	7.27	3.28	8.87	0.91	1.52	2.45	0.46
Dementia	0.20	0.13	0.84	0.51	3.60	2.73	11.5	9.03	0.73	1.45	2.79	0.59
Osteoporosis	0.07	0.04	1.16	0.27	5.98	0.85	8.61	1.71	0.54	1.04	2.03	0.58
Epilepsy	1.07	0.87	1.52	1.47	1.88	1.91	2.08	2.73	0.90	1.47	2.05	0.39
Heart disease ^d	0.01	0.05	0.21	0.71	1.92	3.23	6.45	7.27	0.35	0.72	1.31	0.58
Cancer	0.30	0.20	1.18	0.86	2.26	2.52	2.15	4.20	0.28	0.60	1.56	0.70
Parkinson's disease	0.08	0.13	0.81	0.57	2.54	2.01	4.87	4.17	0.55	1.00	1.48	0.46
Glaucoma	0.02	0.04	0.43	0.36	2.32	1.97	3.84	2.65	0.35	0.81	1.28	0.57
Autoimmune rheumatic diseases	0.39	0.12	1.08	0.59	1.86	1.26	1.57	1.05	0.41	0.69	1.11	0.46
Urinary incontinence	0.15	0.15	0.54	0.35	1.58	1.29	2.77	2.31	0.33	0.61	0.92	0.47
Inflammatory bowel disease	0.36	0.40	0.66	0.76	0.88	0.62	0.95	0.45	0.29	0.54	0.85	0.49
HIV/AIDS	0.15	0.38	0.48	1.08	0.04	0.31	0.02	0.13	0	0.08	0.18	1
Liver disease	0.06	0.28	0.40	0.71	0.45	0.78	0.34	0.45	0	0	0.24	1
Multiple sclerosis	0.31	0.10	0.49	0.17	0.17	0.05	0.07	0.05	0	0.17	0.33	1
Medication/drug abuse	0.11	0.27	0.26	0.46	0.10	0.08	0.03	0.05	0	0.09	0.20	1
Presbycusis/deafness	0.03	-	0.07	0.07	0.35	0.55	0.85	1.18	0	0	0	-
Psoriasis	0.10	0.14	0.17	0.28	0.25	0.37	0.13	0.18	0	0	0.14	1
Anorexia/bulimia nervosa	0.69	0.38	0.04	0.00	0.02	-	-	-	0	0.09	0.26	1
Alcohol abuse	0.01	0.04	0.17	0.32	0.09	0.30	-	0.10	0	0	0.26	1
Macular degeneration	-	-	0.03	0.00	0.11	0.12	0.75	0.39	0	0	0.09	1
Peripheral neuritis/neuropathy	0.02	0.02	0.05	0.05	0.09	0.22	0.13	0.26	0	0	0	-
Viral hepatitis	0.02	0.06	0.06	0.12	0.01	0.07	-	-	0	0	0.07	1
Personality disorder	0.05	0.04	0.07	0.06	0.02	0.05	0.05	-	0	0	0	-
Pulmonary heart disease	0.00	0.01	0.02	0.01	0.12	0.10	0.10	0.08	0	0	0	-
Retinopathy	0.00	0.01	0.01	0.03	0.03	0.09	0.13	0.16	0	0	0	-
Chronic skin ulcer	-	0.01	0.00	0.01	0.04	0.06	0.16	0.10	0	0	0	-
Intellectual disability	0.01	0.03	0.01	0.03	0.02	0.01	-	0.03	0	0	0	-
Blindness	0.00	-	-	-	0.03	0.02	0.08	0.03	0	0	0	-
Trigeminal neuralgia	-	0.01	0.01	0.00	0.01	0.02	0.02	0.05	0	0	0	-
Learning disability	0.01	0.02	0.00	0.01	-	-	-	-	0	0	0	-
Secondary effect of trauma	-	-	0.00	0.01	0.01	-	-	-	0	0	0	-
Poliomyelitis	-	-	-	-	0.02	0.01	0.02	0.03	0	0	0	-
Bowel incontinence	-	-	-	-	0.01	-	0.05	-	0	0	0	-

^{a-d} as in Supplementary table 2