Family Practice 2012; **29**:299–314
doi:10.1093/fampra/cmr099
Advance Access published on 3 February 2012

An international comparative family medicine study of the Transition Project data from the Netherlands, Malta and Serbia. Is family medicine an international discipline? Comparing diagnostic odds ratios across populations

Jean K Soler[a,b,*], Inge Okkes[b,c], Sibo Oskam[c], Kees van Boven[d], Predrag Zivotic[e], Milan Jevtic[f], Frank Dobbs[a,b] and Henk Lamberts[b,c] for the Transition Project

[a]Faculty of Life and Health Sciences, University of Ulster, Coleraine, Northern Ireland, [b]Mediterranean Institute of Primary Care, Attard, Malta, [c]Formerly of the Department of General Practice, Academic Medical Centre, University of Amsterdam, the Netherlands, [d]Department of Primary and Community Care, Nijmegen Medical Centre, Radboud University, Nijmegen, the Netherlands, [e]BONEX inzenjering, GandijevaStr. 148a, Belgrade 11070, Serbia and [f]Saga System Integration, 11070 Belgrade, Serbia.
*Correspondence to Jean K Soler, Mediterranean Institute of Primary Care, 19, Triq ir-Rand, Attard ATD1300, Malta; E-mail: jksoler@synapse.net.mt

Received 9 May 2011; Revised 28 September 2011; Accepted 1 October 2011.

Introduction. This is an international study of the epidemiology of family medicine (FM) in three practice populations from the Netherlands, Malta and Serbia. Diagnostic associations between common reasons for encounter (RfEs) and episodes titles are compared and similarities and differences are described and analysed.

Methodology. Participating family doctors (FDs) recorded details of all their patient contacts in an 'episode of care (EoC)' structure using the International Classification of Primary Care (ICPC). RfEs presented by the patient and episode titles (diagnostic labels of EoCs) were classified with ICPC. The relationships between RfEs and episode titles were studied with Bayesian methods.

Results. Distributions of diagnostic odds ratios (ORs) from the three population databases are presented and compared.

Conclusions. ICPC, the RfE and the EoC data model are appropriate tools to study the process of diagnosis in FM. Distributions of diagnostic associations between RfEs and episode titles in the Transition Project international populations show remarkable similarities and congruencies in the process of diagnosis from both the RfE and the episode title perspectives. The congruence of diagnostic associations between populations supports the use of such data from one population to inform diagnostic decisions in another. Differences in the magnitude of such diagnostic associations are significant, and population-specific data are therefore desirable. We propose that both an international (common) and a local (health care system specific) content of FM exist and that the empirical distributions of diagnostic associations presented in this paper are a reflection of both these effects. We also observed that the frequency of exposure to such diagnostic challenges had a strong effect on the confidence intervals of diagnostic ORs reflecting these diagnostic associations. We propose that this constitutes evidence that expertise in FM is associated with frequency of exposure to diagnostic challenges.

Keywords. Diagnosis, electronic medical records, electronic patient records, episode of care, family medicine, general practice, ICPC, international, International Classification of Primary Care, Malta, person-centred care, posterior probability, prior probability, reason for encounter, Serbia, the Netherlands, Transition Project.

Introduction

The development of family medicine (FM, synonymous with general practice) as an academic discipline is supported by the collection of empirical information from clinical practice. The study of the epidemiology of FM using electronic medical record (EMR) databases is a classic example.[1–4] Numerous EMR databases

collected from networks of FM practices are available. Unfortunately, many do not capture the patient's symptoms and complaints and are not structured to reflect the richness and complexity of doctor–patient encounters over time, i.e. data are often not collected with an episode-based model. Consequently, incidence and prevalence rates are inflated since they cannot be precisely corrected for the effect of multiple encounters dealing with one health problem over time, as we have shown in the first article in this series.[3,4]

The use of a classification system with the appropriate granularity and theoretical framework to accurately reduce data complexity is an additional challenge for the quality of such studies.[3] FM is a complex discipline; the reduction of the content of a consultation into one or more medical diagnoses, ignoring the patient's reason for encounter (RfE), is a coarse reduction which lacks important perspectives necessary to fully characterize a population's health care needs.[4] The databases collected as part of the Transition Project[1,4] are an attempt to collect data in a way which addresses many of these challenges. However, practices providing such data are not necessarily representative of all practices within a geographical region or health care system,[1] and inter-doctor and inter-practice variations have effects, which should be considered in analysing such research studies.[5] We have studied inter-doctor variation in another article in this series.[6]

As we have shown previously,[3,4] studies of incidence and prevalence in FM allow the exploration of similarities and differences between FM populations. However, they do not allow study of the core characteristics of the practice of FM, including the process of diagnosis. We hypothesize the existence of both international and local, health care system specific, content of the domain of FM. Distributions of data from EMR studies are a reflection of both effects, besides many other effects.[4] Do studies into the process of diagnosis add further insight into an international core content of the domain of FM?

This study aims to support the development of FM research through the analysis of the epidemiology of FM using a set of longitudinal clinical databases, collected in the Netherlands, Malta and Serbia, since 1985, 2000 and 2003, respectively.[4,7] This paper shall focus on distributions of diagnostic associations between RfEs and diagnoses.

In view of substantial socio-cultural and economic differences between these countries, besides differences in health care delivery, it is to be expected that the distributions of patient needs and disease patterns will be different, and consequently, family doctors' (FDs) process of care and diagnoses too.[4] However, the general content of FM in these countries may be similar, due to the postulated international discipline of FM, with similarities in actual practise across international divides. The study of these similarities and differences reflected in EMR data collected from these practice populations could inform the discussion on an international core content of FM as a clinical and academic discipline and on whether data on diagnostic relationships in one population can be usefully transferred to another population.

Research questions

- What are the quantitative relationships between common RfEs and common diagnoses (episode titles) within episodes of care (EoCs) in routine family practice in practice populations from Malta, the Netherlands and Serbia?
- What are potential explanations for apparent similarities and differences in the approach to the process of diagnosis, within EoCs in these practice populations?
- Notwithstanding such differences, if any, what are the generic similarities in the content and practice of general practice/FM in these practice populations and do these similarities support the existence of an international discipline of FM?

Methodology

Data

The freely available EMR TransHis[7] uses the 'International Classification of Primary Care' (ICPC)[8] to collect data from FDs who participated in the Transition Project and recorded details of all their patient contacts in an 'EoC' structure for a defined period of time [43 577 patient years of observation over 5 years in Malta (2001–2005), 158 370 patient years over 11 years in the Netherlands (1995–2005) and 72 673 patient years during 1 year in Serbia (2003)]. The populations in the Netherlands and Serbia represent registered patient populations (but the Serb FDs should not register children <15 years), while the population in Malta represents patients consulting over a 5-year period.

Data elements

An EoC is defined as a health problem from its first presentation by the patient to the FD, until the completion of the last encounter for it. It encompasses all contact elements related to that health problem. Its name (i.e. the diagnostic label of the EoC) may be modified over time, and in this article, we refer to it as the 'episode title'.[8,9] A symptom label can, and indeed should, be used as an episode title when appropriate,[3,8] and this is termed a 'symptom diagnosis'.

The RfE(s) is defined as an agreed statement of the reason(s) why a person enters the health care system, representing the demand for care by that person. The

RfE should be recognized by the patient as an acceptable description of the demand for care.[8,9] Doctors recording data for the Transition Project were trained to record RfEs according to the definitions above and the ICPC book,[8] reflecting the patient's symptoms and requests as they expressed them. Symptoms elicited during history taking (i.e. the history of the presenting complaint) were recorded in a separate cell in the EMR Transhis and were not used for the analyses in this study. RfEs presented by the patient and diagnostic labels for each EoC were classified with ICPC-2-E (ICPC-1 in the Netherlands). These data were recorded in an episode-oriented model using the EMR TransHis in various versions.[7]

ICPC has a biaxial structure, with 17 chapters on one axis and 7 components on the other.

Chapters are based on body systems, with an additional chapter for psychological problems and one for social problems.[8] Each chapter is identified by a single alphabetic code, which is the first character of all rubrics belonging to that chapter. Each chapter is divided into seven components, identified by a range of two digit numeric codes. Component 1 codes symptoms and complaints, while Component 7 codes diseases. A RfE can be either a symptom (Component 1) or a disease (Component 7) when a patient presents with the RfE such as 'doctor, I have migraine'. Conversely, an EoC may have a disease-label diagnostic title or it may be labelled with a Component 1 'symptom' diagnosis, such as when the FD cannot be more precise than label an EoC with the title 'shortness of breath'. Components 2–6 deal with interventions and can be used to code an RfE, which is presented as a request for an 'intervention'.[8]

Data analysis

The relationships between RfEs and episode titles were studied using Bayesian probability analysis to calculate the 'posterior (post-test) odds' of an episode title given a RfE, at the start of a new EoC. The post-test odds of an episode title, given the presence of a specific RfE, is given by the pre-test odds (related to the prevalence of the EoC in the population under study) multiplied by the 'likelihood ratio' (LR). The diagnostic odds ratio (OR) is derived from the positive and negative LRs (see Table 1 below). The diagnostic relationships between RfEs and episode titles are thus summarized by such diagnostic ORs, and these numbers are reported in the tables. The ORs were calculated in a similar way to the method by Okkes *et al.*,[9] representing the odds of disease against no disease over odds of RfE present against absent (Table 1). We modified the method slightly so as to calculate odds within an EoC rather than patient years of observation. This has the advantage of estimating LRs and ORs for a new problem at the beginning of an EoC.

It would be possible to analyse such diagnostic relationships for all possible combinations of episode titles and RfEs. For the purpose of feasibility and clarity, the analysis was limited to the common top 20 distributions of EoCs (prevalent) and RfEs (incident) from all three databases, which we have published

TABLE 1 *Example of a diagnostic decision aid from the Transition Project*

Crosstab of diagnosis against reason for encounter at the start of a new episode of care			
EoCs	With title bronchitis (R78) (%)	With any other episode title (%)	Total
With RfE cough (R05)	4717 (24.4)	14 578 (75.6)	19 295
With any other RfE	1899 (0.6)	316 154 (99.4)	318 053
Total	6616 (2.0)	330 732 (98.0)	337 348
LR+: 16.2	LR–: 0.3	OR: 53.9	
CI: 15.8–16.5	CI: 0.3–0.3	CI: 51.0–57.0	
Sens: 0.71	PV+: 0.24	Pretest odds: 0.02	
Spec: 0.96	PV–: 0.99	Posttest odds: 0.32	

Distributions of new episodes of care (EoC) with, or without, the reason for encounter (RfE) 'cough' (R05) and the episode title 'bronchitis' (R78). The characteristics of a predictor: cough as a predictor of bronchitis at the start of a new EoC (first encounter in a new EoC) in the Netherlands. The proportion of EoCs starting with a diagnosis of bronchitis (code R78) rises from 2.0% [proportion in new EoCs in all (unselected) patients, i.e. the pre-test or prior probability) to 24.4% (observed 'plus expected') in those presenting with the RfE cough (code R05). The LR+ is 16.2 and the LR– is 0.3, with narrow CIs, which exclude unity. The pre-test odds of bronchitis in a new EoC is 0.02, and the post-test odds given the RfE cough is 0.32. The diagnostic OR is 53.9 (numerically equivalent to LR+/LR–) with a narrow CI. LR+, positive likelihood ratio; an expression of the extent to which a symptom increases the probability of a diagnosis. The LR for the existence of the symptom (RfE) is the odds that it will exist in a new EoC with that diagnosis, in contrast to a new EoC without that diagnosis; LR–, negative likelihood ratio; the LR for absence of the symptom (a negative result) is the odds that a test will be negative in a new EoC of that diagnosis, contrasted with an EoC without that diagnosis; PV, predictive value (+positive and –negative); the probability that a new EoC with a positive test (presence of a defined RfE) has the disease (positive predictive value). The probability that a person or a proportion of a population with a negative test does not have the disease is the negative predictive value; OR, diagnostic odds ratio of disease (i.e. odds of episode title bronchitis present against absent) against test (i.e. odds of RfE cough present against absent); the ratio of the probability of occurrence of an event to that of non-occurrence; Sens, sensitivity; a test with high sensitivity detects a high proportion of true cases; Spec, specificity; the specificity is the proportion of truly non-diseased persons who are so identified by the test (synonym: true negative rate); pretest odds, pre-test odds; odds of disease in all new EoCs; posttest odds, post-test odds; odds of disease in the population of new EoCs starting with the RfE cough; CI, 95% confidence interval.

previously.[4] Data from all types of encounters (including telephone encounters and repeat prescriptions) were analysed together, but RfEs presented as requests for interventions (such as a request for a repeat prescription or paperwork) were excluded from the analysis. Thus, only ICPC Component 1 or 7 RfEs [i.e. RfEs expressed as a symptom (Component 1) such as 'cough' or a disease-label (Component 7), such as '(I have) migraine'] were considered for this study.[8]

Clinical and statistical significance

The minimum level of 'clinical significance' for an OR was taken as that which represents a standardized difference of $\geqslant 0.1$ (10%) , which equates to a relative risk of $\geqslant 2$.[10] Since the OR tends to overestimate the relative risk, an arbitrary cut-off level of $\geqslant 3$ (rounded from $\geqslant 2.45$ at two decimal places) for the OR of a positive association and $\leqslant 0.3$ (rounded from $\leqslant 0.34$) for the OR of a negative association were taken as thresholds for the clinical significance of an association. This is intentionally more conservative than the previously published minimally useful values of 2.0 for a positive likelihood ratio (LR+) and 0.5 for a negative likelihood ratio (LR–), respectively.[11] ORs which are outside these limits, as well as calculations based on cells with small numbers, were ignored for the purposes of this study. Confidence intervals (CIs) for these ORs were included to express our confidence limits in generalizing these diagnostic data to a larger population of diagnostic decisions. ORs which are not at least as large as the width of their 95% CI were also ignored as statistically unreliable.[9,12,13] These conservative criteria adjust for the increased chance of describing spurious associations due to the large numbers of statistical tests, or due to small number effects, and for the effect of clustering of data on estimates of variance.[12]

Analysis of similarities

The diagnostic ORs which were both clinically and statistically significant were organized in tables in sets of three, one from each practice population database. The two tables were ordered by the 36 RfEs and 41 episode titles, respectively. Sets of clinically significant ORs were compared, when two or more ORs were significant. Comparisons included direction of the ORs from unity, magnitude of the OR and proportion of sets for a given RfE or episode title which were congruent. Congruency was taken as similarity in direction from unity, i.e. two ORs which are significant and which are both greater than and less than unity. Statistical consistency of ORs was also analysed, two ORs being taken to be consistent when either was within the 95% CI of the other.

Ethical approval

The study does not involve collection of new data. Ethical approval has been applied for locally, when appropriate, for studies in the Netherlands, Serbia and Malta.

Results

We would suggest that a printed copy of all ICPC rubrics and short text labels might be useful while reading the Results and Discussion sections below. Such two-page documents are freely available in many languages from the Wonca website (http://www.globalfamilydoctor.com/wicc/pagers.html).

Exemplar diagnostic association

An example of the analysis of a diagnostic association based on output from the Transition Project is given in Table 1.[9] The characteristics of a predictor are exemplified by testing the predictive power of the RfE cough for the episode title 'bronchitis' at the start of a new EoC (first encounter in a new EoC) in the Dutch database. The proportion of diagnoses of bronchitis (code R78) in new EoCs rises from 2.0% [proportion in new EoCs in all (unselected)] patients: pre-test or prior probability, the 'test' being the RfE presented during a consultation with an FD] to 24.4% ('observed plus expected': post-test probability) in those presenting with the RfE cough (code R05). In a 'population' of EoCs, those new EoCs that start with the RfE cough (R05) have 54 times the odds to have a diagnostic label of bronchitis (R78) of those that do not (OR). This diagnostic OR is used in Tables 2 and 3 to summarize diagnostic relationships. The presence of the symptom cough makes the diagnostic label of bronchitis at the start of a new EoC over 16 times (LR+) more likely than the pre-test odds, and its absence decreases the likelihood to one-third (LR–) of the pre-test odds. The LR+ is high (unity would indicate no association) and the LR– is low, with narrow CIs that exclude unity. The diagnostic OR is consequently high (LR+/LR– = 53.9) with a narrow CI. The pre-test odds is 0.02 and the post-test odds is 0.32, numerically equivalent to the pre-test odds multiplied by the LR+. This exemplar OR fulfils our clinical and statistical significance criteria.

Distributions of diagnostic associations—RfE perspective

The distribution of clinically and statistically significant diagnostic ORs of the 36 RfEs against the 41 episode titles in new EoCs is given in Table 2.[4] For each RfE, episode titles with a significant association are given as an ICPC rubric, along with the diagnostic OR and 95% CI. Each set of numbers in brackets represents the OR for the three populations in turn (given only if they are significant, otherwise left blank '—'). The ORs are calculated at the start of new EoCs as in Table 1.

TABLE 2 *The diagnostic OR of the joint top 20 new reasons for encounter (RfE) against the top 20 episode of care (EoC) counterparts*

Rank	Code	Label	EoCs with OR <0.3 or >3.0
1	R05	Cough	U71 (0 [0–0];-;-), **R74 (13.7 [13.1–14.3]; 4.3 [4.1–4.5]; 11.6 [10.6–12.7]), R05 (122.2 [114.2–130.8]; 153.3 [112.0–209.7]; 46.4 [29.6–72.7]) R78 (53.9 [51.0–57.0]; 17.1 [15.4–19.0]; 19.3 [17.4–21.5]),** A97 (0.1 [0–0.1];-;-), R96 (13.6 [12.0–15.3]; 20.6 [16.8–25.3];-). A04 (0.1 [0.1–0.2];-;-) R75 (3.1 [2.9–3.3];-; 3.0 [2.2–4.0]), A85 (0.2 [0.2–0.3]; 0.2 [0.2–0.4];-), D73 (-;0 [0–0.1];-), **R77 (35.4 [32.4–38.7]; 13.4 [11.4–15.7]; 16.6 [13.3–20.8]) R80 (9.1 [8.2–10.1]; 4.8 [4.2–5.4]; 6.1 [3.8–9.6]),** R79 (-;-;3.2 [2.4–4.3]), R29 (-;12.0 [10.5–13.7];-)
2	A03	Fever	**R74 (5.9 [5.6–6.4]; 2.8 [2.6–3.0]; 5.3 [4.2–6.6]), R78 (7.6 [7.1–8.2]; 3.1 [2.8–3.5]; 5.5 [4.3–7.1]),** A97 (0.1 [0.1–0.1];-;-), A85 (0.1 [0.1–0.2];-;-), D73 (4.3 [3.8–4.9];-;-) **R76 (12.9 [11.7–14.2]; 13.2 [11.8–14.7]; 6.2 [4.6–8.5]),** R77 (2.7 [2.3–3.2];-;-), R80 (33.3 [30.0–36.9]; 6.9 [6.1–7.8];-), R29 (-;6.8 [6.0–7.7];-)
3	S06	Local redness/ erythema/rash	**S88 (32.3 [30.4–34.3]; 51.1 [40.5–64.5]; 29.0 [18.4–45.8]),** A97 (0.1 [0.1–0.1];-;-), S74 (13.2 [12.4–14.2]; 40.4 [32.5–50.4];-)
4	R21	Symptom/complaint throat	A98 (-;0 [0–0];-), **R74 (12.6 [11.8–13.3]; 11.9 [11.3–12.6]; 15.9 [14.2–17.8]),** A97 (0.2 [0.1–0.2]; 0.1 [0.1–0.1];-), A85 (0.2 [0.2–0.3];-;-) D73 (-;0.1 [0.1–0.2];-) **R76 (181.7 [164.3–200.9]; 18.6 [16.6–21.0]; 14.9[12.9–17.3]),** R77 (9.7 [8.7–10.8];-; 5.8 [4.3–7.9]), R80 (3.4 [2.8–4.2];-;-), R29 (-;6.6 [5.8–7.4];-)
5	A04	General weakness/tiredness	A98 (-;0.2 [0.1–0.4];-), P06 (0.3 [0.2–0.5];-;-), R05 (0.2 [0.2–0.3];-;-), A04 (887.4 [802.2–981.6];-;-), P76 (4.8 [4.1–5.7]; 5.9 [4.0–8.7];-), A85 (-;3.3 [2.3–4.6];-) P01 (-;3.5 [2.2–5.5];-), K77 (2.6 [1.8–3.7];-;-), R80 (2.7 [2.2–3.4]; 3.0 [2.3–4.0];-),
6	S04	Local swelling/papule/lump/mass	A85 (0.1 [0.1–0.2];-;-)
7	L03	Low back complaint excl radiation	**L03 (1055.8 [972.1–1146.8]; 282.6 [191.1–417.8]; 27.8 [18.8–41.2]), L86 (20.9 [19.0–23.1]; 64.4 [48.6–85.2]; 28.4 [22.6–35.7]),** L18 (2.6 [2.0–3.4]; 22.8 [18.4–28.1];-) L84 (18.9 [13.9–25.9];-; 28.0[22.4–34.9])
8	D06	Other localized abdominal pain	**U71 (2.7 [2.4–3.0]; 9.5 [7.8–11.5]; 3.0 [1.9–4.5]),** L03 (0.3 [0.2–0.4];-;-), A97 (0.3 [0.2–0.4];-;-), **D87 (6.8 [5.6–8.1]; 21.2 [16.9–26.4]; 17.9 [12.8–25.1])** D73 (7.0 [6.2–8.0]; 3.8 [3.4–4.3];-)
9	H01	Ear pain/earache	R78 (0.3 [0.2–0.4];-;-), A97 (0.3 [0.2–0.4];-;-), H81 (3.7 [3.3–4.1]; 12.6 [10.1–15.8];-)
10	N01	Headache (excl N02 N89 R09)	A98 (-;0.2 [0.1–0.3];-), R05 (0.2 [0.1–0.3];-;-), A97 (0.3 [0.2–0.4];-;-), R97 (-;4.2 [3.1–5.8];-), P76 (-;2.5 [1.6–4.1];-), **R75 (10.9 [10.0–11.8]; 16.8 [14.2–19.8]; 4.6 [2.9–7.3])** P01 (-;3.0 [1.9–4.5];-), R80 (5.3 [4.4–6.4];-;-), L83 (4.5 [3.0–6.7];-; 3.0 [2–1–4.4])
11	R02	Shortness of breath dyspnoea	R74 (2.7 [2.4–3.1]; 0.3 [0.2–0.5];-), **R78 (20.2 [18.8–21.6]; 5.9 [4.5–7.6]; 6.5 [4.5–9.2]),** A97 (0.3 [0.2–0.5];-;-), R96 (36.0 [31.6–40.9]; 41.8 [33.0–53.0];-) K74 (7.1 [5.2–9.8];-;-), R77 (7.3 [6.3–8.4];-;-), K77 (64.1 [54.1–75.9]; 75.0 [47.8–117.6];-), R79 (-;-;20.6 [13.2–32.1])
12	L08	Shoulder symptoms/complaints	L18 (9.7 [8.1–11.6]; 18.9 [15.4–23.1];-), L83 (13.1 [9.9–17.2];-; 20.7 [13.8–31.1])
13	L17	Foot and toe symptoms/complaints	S74 (-;8.4 [5.7–12.3];-)
14	L15	Knee symptoms/complaints	
15	S02	Pruritus	**S88 (26.7 [24.8–28.7]; 35.2 [27.0–45.8]; 107.1 [83.4–137.6]), S74 (8.2 [7.4–9.0]; 6.2 [4.0–9.4]; 15.7 [9.8–25.0]),** A85 (-;4.2 [2.9–6.1];-)
16	R74	URI (head cold)	R74 (45.5 [42.7–48.5]; 22.3 [16.4–30.3];-), R78 (5.3 [4.8–5.9];-;-), R97 (7.4 [6.1–8.9];-;-), R75 (21.6 [20.0–23.3];-;-), R77 (3.0 [2.4–3.8];-;-)
17	L01	Neck symptom/complaint excl headache	R74 (0.2 [0.1–0.3]; 0.2 [0.2–0.4];-), L18 (16.1 [13.8–18.9]; 21.7 [18.5–25.4];-), **L83 (58.7 [47.8–72.0]; 410.0 [268.1–627.0]; 63.8 [53.4–76.2]),** L84 (-;-;3.1 [2.3–4.3])
18	L14	Leg/thigh symptoms/complaints	A97 (0.2 [0.1–0.3];-;-), **L86 (41.1 [37.2–45.3]; 30.1 [22.7–40.0]; 30.9 [22.4–42.7]),** L18 (6.7 [5.4–8.3]; 11.8 [9.7–14.3];-)
19	H02	Hearing complaints (excl H84)	**H81 (68.5 [64.1–73.3]; 97.4 [71.9–131.8]; 177.5 [117.9–267.1])**
20	L04	Chest symptoms/complaints	R74 (-;0.1 [0.1–0.2];-), P01 (-;5.1 [3.4–7.7];-), L86 (-;-;6.9 [4.6–10.4]), K74 (4.1 [2.6–6.5];-;-), L18 (4.1 [3.1–5.4]; 24.3 [21.1–28.0];-), L84 (-;-;4.2 [2.7–6.7])
21	N17	Vertigo/dizziness	A98 (-;0.2 [0.2–0.4];-), K86 (3.0 [2.1–4.4];-;-), R74 (0.3 [0.2–0.5]; 0.2 [0.2–0.3];-), A04 (2.7 [2.3–3.2];-;-), A85 (-;3.4 [2.4–4.7];-), L83 (-;-;4.2 [2.7–6.7])
22	D11	Diarrhoea	R74 (0.3 [0.2–0.5]; 0.1 [0–0.1];-), A85 (3.5 [3.0–4.1];-;-), D87 (-;-;2.9 [1.8–4.7]), **D73 (163.6 [149.5–179.0]; 131.4 [115.1–149.9]; 74.2 [58.1–94.8])**
26	U02	Urinary frequency/urgency	**U71 (54.4 [50.3–58.9]; 221.3 [173.1–283.0]; 83.0[64.3–107.1]),** T90 (4.3 [2.8–6.5];-;-)
27	U01	Dysuria/painful urination	**U71 (109.4 [100.7–119.0]; 471.1 [375.6–590.9]; 35.4 [29.0–43.1])**

TABLE 2 *Continued*

Rank	Code	Label	EoCs with OR <0.3 or >3.0
28	D09	Nausea	R74 (-;0.1 [0.1–0.2];-), A85 (7.8 [6.9–8.9]; 2.7 [1.9–3.8];-), **D87 (23.5 [20.0–27.7]; 13.8 [10.6–17.9]; 20.4 [14.9–28.1])** **D73 (25.6 [22.7–28.8]; 21.6 [19.2–24.4]; 25.6 [17.9–36.7])**
29	L02	Back symptom/complaint	L03 (3.9 [3.3–4.5]; 12.9 [8.2–20.1];-), **L86 (5.0 [4.0–6.4]; 23.0 [18.0–29.4]; 15.4 [12.6–19.0]),** L18 (13.7 [11.2–16.7]; 31.6 [27.8–35.8];-), L83 (-;-;5.8 [4.4–7.7]), L84 (-;-;32.3 [27.0–38.7])
34	D10	Vomiting	R74 (-;0.2 [0.2–0.3];-), R78 (-;0.3 [0.2–0.4];-), A85 (2.5 [2.0–3.1];-;-), D87 (9.1 [7.1–11.6]; 12.1 [9.5–15.3];-), **D73 (109.5 [99.3–120.8]; 51.8 [46.5–57.7]; 36.9[24.4–55.9])**
35	D01	Abdominal pain/cramps general	R74 (-;0.1 [0.1–0.2];-), **D87 (8.7 [6.8–11.2]; 3.8 [2.6–5.6]; 9.2 [7.0–12.0]), D73 (20.6 [18.0–23.5]; 25.2 [22.4–28.4]; 12.1 [8.8–16.5]),** D85 (-;-;10.9 [8.5–14.0])
36	P01	Feeling anxious/nervous/tense	P06 (2.8 [2.1–3.7];-;-), P76 (14.9 [12.4–17.7]; 44.6 [33.2–59.7];-), **P01 (459.7 [413.0–511.6]; 160.8 [125.5–206.1]; 184.9 [143.9–237.7])** **P74 (52.2 [40.5–67.2]; 73.2 [52.2–102.7]; 29.6 [24.4–36.0])**
39	F02	Red eye	
41	D02	Abdominal pain epigastric	A85 (3.4 [2.8–4.2];-;-), **D87 (146.9 [129.4–166.7]; 14.4 [8.1–25.7]; 23.2 [18.4–29.2]), D73 (6.1 [4.9–7.7]; 9.5 [6.9–13.1]; 15.3 [11.2–20.9])** D85 (120.6 [79.2–183.6];-; 10.1 [7.7–13.3])
53	R07	Sneezing/nasal congestion	**R74 (14.1 [12.5–16.0]; 14.9 [14.0–15.8]; 7.3 [4.8–11.2]),** R97 (92.1 [80.4–105.6]; 14.2 [11.7–17.3];-), R75 (6.6 [5.5–7.9]; 14.2 [12.3–16.5];-), D73 (-;0.1 [0–0.1];-), R76 (-;0.3 [0.3–0.4];-), R80 (-;3.5 [3.1–4.0];-)
75	S07	Rash generalized	S88 (2.5 [1.8–3.5]; 26.4 [20.0–34.7];-), S74 (-;4.3 [2.6–6.9];-), A85 (9.3 [7.6–11.4]; 7.0 [5.2–9.4];-)
78	K02	Pressure/tightness of heart	K86 (-;-;10.1 [7.6–13.5]), K85 (-;-;13.1 [9.3–18.5]), K74 (230.6 [186.3–285.5];-; 14.5 [9.0–23.3]), K80 (-;-;12.5 [8.0–19.5])
123	A05	Feeling ill	A04 (3.1 [2.0–4.6];-;-)
129	R01	Pain respiratory system	R74 (4.1 [3.0–5.5];-; 10.6 [9.2–12.1]), R78 (8.9 [7.0–11.4];-; 3.1 [2.5–3.8]), R75 (-;-;7.4 [5.4–10.0]), R76 (-;-;6.4 [5.3–7.9]), R77 (-;36.3 [23.5–55.9]; 5.1 [3.5–7.4]) R80 (13.3 [9.1–19.5];-;-), R79 (-;-;4.9 [3.4–7.1])

For each RfE, the episode titles that had a significant association with it are listed. The OR for each RfE against that episode title is given, along with the 95% CI for each of the three populations in turn. Trends consistent in all three populations highlighted in bold type. Rank, the rank of incidence in the Netherlands;[4] Code, the RfE ICPC rubric; label, the RfE rubric text label (ICPC-2-E labels used); EoCs with OR <0.3 or >3.0, the ICPC rubric (refer to Table 3 or ICPC two pager for text label) for all episode titles with a clinically significant OR for that RfE is given, followed by the OR for each population in turn, in brackets, with the 95% CI in square brackets, i.e. episode title ICPC code (OR in Dutch database [95% CI]; OR in Maltese database [95% CI] and OR in Serb database [95% CI]). Dashes indicate that an OR was not clinically and/or statistically significant.
Note: First cell in Table 2, second cluster of ORs, clarified:
Reason for encounter is cough (R05).
Episode title is upper respiratory tract infection (R74).
Diagnostic odds ratio for each of the three populations in turn are 13.7 (the Netherlands), 4.3 (Malta), and 11.6 (Serbia), with CIs in square brackets. Presentation, on the row of RfE R05: R74(13.7[13.1–14.3]; 4.3[4.1–4.5]; 11.6[10.6–12.7]) is in bold as all three ORs significant. Non-significant ORs represented by a dash '-', but not in this example.

Of 159 sets of significant associations (Table 2), 76 allowed comparisons between databases since two or three ORs were clinically and statistically significant. Seventy-five of these sets of associations were congruent (i.e. all significant ORs were in the same direction, either more than or less than unity) between at least two databases. In only 1 of 76 sets were the ORs not congruent (diverging, in opposite directions from unity).

Eighteen of 36 RfEs (Table 2) demonstrated associations which were congruent in two or more databases in at least half of the sets for that RfE. In the other 18 cases, only one OR in a set of three was significant in the majority of sets, and as such, no assessment of congruence was possible. Individual examples with positive or negative associations with a number of EoCs included cough (R05), 'nasal snuffles' (R07) and digestive system RfEs ['vomiting' (D10), 'diarrhoea' (D11) and 'nausea' (D09)].

In general, there were greater similarities in the sets of ORs between the Maltese and Dutch databases. However, 'textbook' associations were strong in all three databases, such as the association of the symptoms diarrhoea (D11), vomiting (D10) and nausea (D09) with the episode title 'gastroenteritis' (D73); cough (R05), fever (A03) and shortness of breath (R02) with bronchitis (R78); cough (R05), throat pain (R21) and nasal symptoms (R07) with pharyngitis/upper respiratory tract infection (URTI) (R74); fever (A03) and throat pain (R21) with tonsillitis (R76); localized abdominal pain (D06), 'urinary frequency' (U02) and 'dysuria' (U01) with 'cystitis' (U71); but also cough (R05) and 'fever' (A03) with their respective

TABLE 3 *The diagnostic OR of the joint top 20 episodes of care (EoC) against the top 20 new reason for encounter (RfE) counterparts*

Rank	Code	Label	RfEs with OR <0.3 or >3.0
1	A98	Prevention	R21 (-;0 [0–0];-), A04 (-;0.2 [0.1–0.4];-), N01 (-;0.2 [0.1–0.3];-), N17 (-;0.2 [0.2–0.4];-)
2	W11	Family plan/oral contraceptive	
3	K86	Uncomplicated hypertension	N17 (3.0 [2.1–4.4];-;-), K02 (-;-;10.1 [7.6–13.5])
4	S88	Contact dermatitis/ other eczema	**S06 (32.3 [30.4–34.3]; 51.1 [40.5–64.5]; 29.0 [18.4–45.8]), S02 (26.7 [24.8–28.7]; 35.2 [27.0–45.8]; 107.1 [83.4–137.6]),** S07 (2.5[1.8–3.5]; 26.4 [20.0–34.7];-)
5	P06	Disturbances of sleep/insomnia	A04 (0.3 [0.2–0.5];-;-), P01 (2.8 [2.1–3.7];-;-)
6	U71	Cystitis/other urine infect NOS	R05 (0 [0–0];-;-), **D06 (2.7 [2.4–3.0]; 9.5 [7.8–11.5]; 3.0 [1.9–4.5]), U02 (54.4 [50.3–58.9]; 221.3 [173.1–283.0]; 83.0 [64.3–107.1])** **U01 (109.4 [100.7–119.0]; 471.1 [375.6–590.9]; 35.4 [29.0–43.1])**
7	L03	Low back complaint excl radiation	**L03 (1055.8 [972.1–1146.8]; 282.6 [191.1–417.8]; 27.8 [18.8–41.2]),** D06 (0.3[0.2–0.4];-;-), L02 (3.9 [3.3–4.5]; 12.9 [8.2–20.1];-)
8	R74	URI (head cold)	**R05 (13.7 [13.1–14.3]; 4.3 [4.1–4.5]; 11.6 [10.6–12.7]), A03 (5.9 [5.6–6.4]; 2.8[2.6–3.0]; 5.3 [4.2–6.6]), R21 (12.6 [11.8–13.3]; 11.9 [11.3–12.6]; 15.9 [14.2–17.8])** R02 (2.7 [2.4–3.1]; 0.3 [0.2–0.5];-), R74 (45.5 [42.7–48.5]; 22.3 [16.4–30.3];-), L01 (0.2 [0.1–0.3]; 0.2 [0.2–0.4];-), L04 (-;0.1 [0.1–0.2];-) N17 (0.3 [0.2–0.5]; 0.2 [0.2–0.3];-), D11 (0.3 [0.2–0.5]; 0.1 [0–0.1]);-), D09 (-;0.1 [0.1–0.2];-), D10 (-;0.2 [0.2–0.3];-), D01 (-;0.1 [0.1–0.2];-) **R07 (14.1 [12.5–16.0]; 14.9 [14.0–15.8]; 7.3 [4.8–11.2]),** R01 (4.1 [3.0–5.5];-; 10.6 [9.2–12.1])
9	R05	Cough	**R05 (122.2 [114.2–130.8]; 153.3 [112.0–209.7]; 46.4 [29.6–72.7]),** A04 (0.2 [0.2–0.3];-;-), N01 (0.2 [0.1–0.3];-;-)
10	R78	Acute bronchitis/ bronchiolitis	**R05 (53.9 [51.0–57.0]; 17.1 [15.4–19.0]; 19.3 [17.4–21.5]), A03 (7.6 [7.1–8.2]; 3.1 [2.8–3.5]; 5.5 [4.3–7.1]),** H01 (0.3 [0.2–0.4];-;-) **R02 (20.2 [18.8–21.6]; 5.9 [4.5–7.6]; 6.5 [4.5–9.2]),** R74 (5.3 [4.8–5.9];-;-), D10 (-;0.3 [0.2–0.4];-), R01 (8.9 [7.0–11.4];-; 3.1 [2.5–3.8])
11	A97	No disease	R05 (0.1 [0–0.1];-;-), A03 (0.1 [0.1–0.1];-;-), S06 (0.1 [0.1–0.1];-;-), R21 (0.2 [0.1–0.2]; 0.1 [0.1–0.1];-), D06 (0.3 [0.2–0.4];-;-), H01 (0.3 [0.2–0.4];-;-), N01 (0.3 [0.2–0.4;-;-), R02 (0.3 [0.2–0.5];-;-), L14 (0.2 [0.1–0.3];-;-)
12	S74	Dermatophytosis	S06 (13.2[12.4–14.2]; 40.4 [32.5–50.4];-), L17 (-;8.4 [5.7–12.3];-), **S02 (8.2 [7.4–9.0]; 6.2 [4.0–9.4]; 15.7 [9.8–25.0]),** S07 (-;4.3 [2.6–6.9];-)
13	H81	Excessive ear wax	H01 (3.7 [3.3–4.1]; 12.6 [10.1–15.8];-), **H02 (68.5 [64.1–73.3]; 97.4 [71.9–131.8]; 177.5 [117.9–267.1])**
14	R96	Asthma	R05 (13.6 [12.0–15.3]; 20.6 [16.8–25.3];-), R02 (36.0 [31.6–40.9]; 41.8 [33.0–53.0];-)
15	T93	Lipid metabolism disorder	
16	R97	Hay fever/allergic rhinitis	N01 (-;4.2 [3.1–5.8];-), R74 (7.4 [6.1–8.9];-;-), R07 (92.1 [80.4–105.6]; 14.2 [11.7–17.3];-)
17	A04	General weakness/tiredness	R05 (0.1[0.1–0.2];-;-), A04 (887.4 [802.2–981.6];-;-), N17 (2.7 [2.3–3.2];-;-), A05 (3.1 [2.0–4.6];-;-),
18	P76	Depressive disorder	A04 (4.8 [4.1–5.7]; 5.9 [4.0–8.7];-), N01 (-; 2.5 [1.6–4.1];-), P01 (14.9 [12.4–17.7]; 44.6 [33.2–59.7];-)
19	R75	Sinusitis acute/chronic	R05 (3.1 [2.9–3.3];-; 3.0 [2.2–4.0]), **N01 (10.9 [10.0–11.8]; 16.8 [14.2–19.8]; 4.6 [2.9–7.3]),** R74 (21.6 [20.0–23.3];-;-), R07 (6.6 [5.5–7.9]; 14.2 [12.3–16.5];-) R01 (-;-;7.4 [5.4–10.0])
20	A85	Adverse effect medical agent proper dose	R05 (0.2 [0.2–0.3]; 0.2 [0.2–0.4];-), A03 (0.1 [0.1–0.2];-;-), R21 (0.2 [0.2–0.3];-;-), A04 (-;3.3 [2.3–4.6];-), S04 (0.1 [0.1–0.2];-;-), S02 (-;4.2 [2.9–6.1];-) N17 (-;3.4 [2.4–4.7];-), D11 (3.5 [3.0–4.1];-;-), D09 (7.8 [6.9–8.9]; 2.7 [1.9–3.8];-), D10 (2.5 [2.0–3.1];-;-), D02 (3.4 [2.8–4.2];-;-), S07 (9.3 [7.6–11.4]; 7.0 [5.2–9.4];-)
24	T90	Diabetes non-insulin dependent	U02 (4.3 [2.8–6.5];-;-)
25	P01	Feeling anxious/ nervous/tense	A04 (-;3.5 [2.2–5.5];-), N01 (-;3.0 [1.9–4.5];-), L04 (-;5.1 [3.4–7.7];-), **P01 (459.7 [413.0–511.6]; 160.8 [125.5–206.1]; 184.9 [143.9–237.7])**
27	L86	Back syndrome with radiating pain	**L03 (20.9 [19.0–23.1]; 64.4 [48.6–85.2]; 28.4 [22.6–35.7]), L14 (41.1 [37.2–45.3]; 30.1 [22.7–40.0]; 30.9 [22.4–42.7]),** L04 (-;-;6.9 [4.6–10.4]) **L02 (5.0 [4.0–6.4]; 23.0 [18.0–29.4]; 15.4 [12.6–19.0])**
41	D87	Stomach function disorder	**D06 (6.8 [5.6–8.1]; 21.2 [16.9–26.4]; 17.9 [12.8–25.1]),** D11 (-;-;2.9 [1.8–4.7]), **D09 (23.5 [20.0–27.7]; 13.8 [10.6–17.9]; 20.4 [14.9–28.1])** D10 (9.1 [7.1–11.6]; 12.1 [9.5–15.3];-), **D01 (8.7 [6.8–11.2]; 3.8 [2.6–5.6]; 9.2 [7.0–12.0]), D02 (146.9 [129.4–166.7]; 14.4 [8.1–25.7]; 23.2 [18.4–29.2])**
46	K85	Elevated blood pressure	K02 (-;-;13.1 [9.3–18.5])

TABLE 3 *Continued*

Rank	Code	Label	RfEs with OR <0.3 or >3.0
48	D73	Gastroenteritis presumed infection	R05 (-;0 [0-0.1];-), A03 (4.3 [3.8–4.9];-;-), R21 (-;0.1 [0.1–0.2];-), D06 (7.0 [6.2–8.0]; 3.8 [3.4–4.3];-), **D11 (163.6 [149.5–179.0]; 131.4 [115.1–149.9]; 74.2 [58.1–94.8])** **D09 (25.6 [22.7–28.8]; 21.6 [19.2–24.4]; 25.6 [17.9–36.7]), D10 (109.5 [99.3–120.8]; 51.8 [46.5–57.7]; 36.9 [24.4–55.9])** **D01 (20.6 [18.0–23.5]; 25.2 [22.4–28.4]; 12.1 [8.8–16.5]), D02 (6.1 [4.9–7.7]; 9.5 [6.9–13.1]; 15.3 [11.2–20.9]),** R07 (-;0.1 [0–0.1];-)
49	R76	Tonsillitis acute	**A03 (12.9 [11.7–14.2]; 13.2 [11.8–14.7]; 6.2 [4.6–8.5]), R21 (181.7 [164.3–200.9]; 18.6 [16.6–21.0]; 14.9 [12.9–17.3]),** R07 (-;0.3 [0.3–0.4];-), R01 (-;-;6.4 [5.3–7.9])
54	K74	Ischaemic heart disease with angina	R02 (7.1 [5.2–9.8];-;-), L04 (4.1 [2.6–6.5];-;-), K02 (230.6 [186.3–285.5];-; 14.5 [9.0–23.3])
55	R77	Laryngitis/tracheitis acute	**R05 (35.4 [32.4–38.7]; 13.4 [11.4–15.7]; 16.6 [13.3–20.8),** A03 (2.7 [2.3–3.2];-;-), R21 (9.7 [8.7–10.8];-; 5.8 [4.3–7.9]), R02 (7.3 [6.3–8.4];-;-), R74 (3.0 [2.4–3.8];-;-) R01 (-;36.3 [23.5–55.9]; 5.1 [3.5–7.4])
59	K77	Heart failure	A04 (2.6 [1.8–3.7];-;-), R02 (64.1 [54.1–75.9]; 75.0 [47.8–117.6];-)
77	L18	Muscle pain	L03 (2.6 [2.0–3.4]; 22.8 [18.4–28.1];-), L08 (9.7 [8.1–11.6]; 18.9 [15.4–23.1];-), L01 (16.1 [13.8–18.9]; 21.7 [18.5–25.4];-), L14 (6.7 [5.4–8.3]; 11.8 [9.7–14.3];-) L04 (4.1 [3.1–5.4]; 24.3 [21.1–28.0];-), L02 (13.7 [11.2–16.7]; 31.6 [27.8–35.8];-)
83	K87	Hypertension complicated	
88	R80	Influenza	**R05 (9.1 [8.2–10.1]; 4.8 [4.2–5.4]; 6.1 [3.8–9.6]),** A03 (33.3 [30.0–36.9]; 6.9 [6.1–7.8];-), R21 (3.4 [2.8–4.2];-;-), A04 (2.7 [2.2–3.4]; 3.0 [2.3–4.0];-), N01 (5.3 [4.4–6.4];-;-) R07 (-;3.5 [3.1–4.0];-), R01 (13.3 [9.1–19.5];-;-)
118	P17	Tobacco abuse	
102	P74	Anxiety disorder/anxiety state	**P01 (52.2 [40.5–67.2]; 73.2 [52.2–102.7]; 29.6 [24.4–36.0])**
128	L83	Neck syndrome	N01 (4.5 [3.0–6.7];-; 3.0 [2.1–4.4]), L08 (13.1 [9.9–17.2];-; 20.7 [13.8–31.1]), **L01 (58.7 [47.8–72.0]; 410.0 [268.1–627.0]; 63.8 [53.4–76.2]),** N17 (-;-;4.2 [2.7–6.7]) L02 (-;-;5.8 [4.4–7.7])
162	L84	Back syndrome without radiating pain	L03 (18.9 [13.9–25.9];-; 28.0 [22.4–34.9]), L01 (-;-;3.1 [2.3–4.3]), L04 (-;-;4.2 [2.7–6.7]), L02 (-;-;32.3 [27.0–38.7])
226	D85	Duodenal ulcer	D01 (-;-;10.9 [8.5–14.0]), D02 (120.6 [79.2–183.6];-; 10.1 [7.7–13.3])
267	K80	Cardiac arrhythmia NOS	K02 (-;-;12.5 [8.0–19.5])
282	R79	Chronic bronchitis	R05 (-;-;3.2 [2.4–4.3]), R02 (-;-;20.6 [13.2–32.1]), R01 (-;-;4.9 [3.4–7.1])
441	R29	Respiratory symptom/ complaint other	R05 (-;12.0 [10.5–13.7];-), A03 (-;6.8 [6.0–7.7];-), R21 (-;6.6 [5.8–7.4];-)

For each episode title, the RfEs that have a significant association with it are listed. The OR for each episode title against the RfE is given, along with the 95% CI, for each of the three populations in turn. Trends consistent in all three populations highlighted in bold type. Rank, the rank of incidence in the Netherlands;[4] Code: the episode title ICPC rubric; label, the episode title rubric text label (ICPC-2-E labels used); RfEs with OR <0.3 or >3.0, the ICPC rubric (refer to Table 2 or ICPC two pager for label) of all RfEs with a clinically significant OR or that EoC is given, followed by the ORs for each population in turn, in brackets, with the 95% CI in square brackets, i.e. RfE ICPC code (OR in Dutch database [95% CI]; OR in Maltese database [95% CI]; OR in Serb database [95% CI]). Dashes indicate that an OR was not clinically and/or statistically significant.

symptom diagnoses (i.e. R05 and A03 as an episode title) at the start of a new EoC. Interesting differences were also found, such as the fact that 'respiratory system pain' (R01) was a predictor of 'tracheitis' (R77) in the Maltese data set, but not in the other two, while it was a predictor for bronchitis (R78) in the Dutch and Serb databases but not in the Maltese.

Distributions of diagnostic associations—episode title perspective

The distribution of clinically and statistically significant diagnostic ORs of the 41 episode titles against the 36 RfEs is given in Table 3.[4] For each episode title, the RfEs with a significant association are given as an ICPC rubric, along with the diagnostic OR and 95% CI. Each set of numbers in brackets represents the OR for the three populations in turn. The ORs are calculated, as in Table 1 at the start of new EoCs.

As expected, 75 of 76 sets of associations with two or more significant ORs were congruent between at least two databases since these are the same data from an episode title perspective. Twenty of 41 episode titles demonstrate sets of associations which were congruent in two or more databases in at least half of the sets. In the majority of the other 21 cases, only one OR in a set of three was significant and as such, no test of congruence was possible. Incongruence was not excluded, but congruence could not be demonstrated on the basis of these data.

There appeared to be greater congruence for diagnoses with classically defined symptom patterns such as the episode title 'eczema', an EoC associated with the RfEs 'local erythema', 'pruritus' and 'generalized rash'; 'cystitis' with the RfEs 'other localized abdominal pain', 'dysuria' and 'frequency'; 'URTI' with 'cough', 'fever', 'throat complaints', 'neck pain',

'dizziness', 'diarrhoea' and 'sneezing', besides the patient presenting with the complaint 'doctor, I have a cold'; bronchitis with cough, fever, shortness of breath and respiratory system pain; 'dermatophytosis' with 'localized rash' and 'pruritus'; 'ear wax' with 'ear pain' and 'hearing complaints'; 'asthma' with 'cough' and 'shortness of breath'; 'depression' with 'lethargy' and 'symptoms of anxiety' (note: 'sadness' was not common enough as an RfE to be included in the common top 20 distribution) and 'sinusitis' with 'cough,' 'headache' and 'nasal symptoms' to list just 9. Congruence for symptom diagnoses as episode titles was less between databases; e.g. for 'family planning,' 'sleep disturbance' and 'cough'; although the symptom diagnosis 'low back pain' as an episode title showed good congruence with the RfEs 'low back pain' and 'back symptoms'.

Distributions of diagnostic associations—all sets

Considering all 159 sets of associations in Tables 2 and 3, it is notable that: (i) 75 sets of associations of 76 where comparisons between two or more ORs were possible, were congruent in at least two databases; (ii) in 36 cases, all three ORs in a set were in the same direction; (iii) in only one case (the association between the RfE 'shortness of breath' (R02) and the episode title 'URTI' (R74)) were the odds higher than unity in one database in the set (2.7 in the Netherlands) and lower than unity in another (0.3 in Malta); (iv) in 35 cases, the ORs in a set were all in the same direction and all greater than 10, indicating a strong association[11] and finally (v) in 19 of 76 sets, the ORs were statistically consistent (one OR lying within the CI of another in the set) between at least two populations.

Discussion

Summary

First research question. This study quantifies, as diagnostic ORs, the relationships between common reasons for encounter and common diagnoses (episode titles) in selected practice populations from Malta, the Netherlands and Serbia. These ORs were remarkably similar between these different FM populations, with many sets of observations being internally congruent, and some statistically consistent. There appeared to be slightly more congruence from the RfE perspective.

Second research question. There are many potential explanations for the observed similarities in the approach to the process of diagnosis in different family practice populations. The existence of an international core content of the process of diagnosis in medicine in general, and FM in particular, is eminent among them and is an appropriate theoretical framework for analysing these observations.

Differences were found in the magnitude of diagnostic relationships (size of diagnostic ORs) rather than their direction. In cases where a diagnostic relationship was significant in one population but was not significant in another, this was mainly due to wide CIs and lack of statistical significance. A larger data set with more observations would probably have found more congruence rather than less.

Third research question. The main conclusion of this study is that diagnostic information from one population, including diagnostic ORs and LRs, can be usefully transferred to other populations. However, the fact that we found important differences in the magnitudes of such diagnostic ORs, which should not be ignored, suggests that this process should be approached with appropriate levels of caution. The availability of more FM data from different populations is therefore highly desirable.

Notwithstanding few observed differences, the remarkable similarities in the content and practice of FM in these practice populations, as reflected in these sets of diagnostic associations, support the existence of an international discipline of FM.

Exemplar diagnostic associations between cough and bronchitis

With a high LR+ and low LR–, and consequently a high diagnostic OR (Table 1), cough is a good predictor for the diagnosis of bronchitis at the beginning of a new EoC. LRs are considered a standard measure of the usefulness of a diagnostic test,[14] and the diagnostic OR is derived from the two LRs (i.e. 53.9, numerically equivalent to LR+/LR–). ORs were used in this article to describe the relationship between a RfE and an episode label at the start of a new EoC. In a companion article, we shall study the contribution of positive and negative LRs to a diagnostic association,[6] but for this study, the OR was the appropriate summary measure of the diagnostic association.[3,9,13]

Distributions of diagnostic associations—RfE perspective

Congruency in sets of diagnostic ORs was remarkable. Diagnostic associations were congruent between two or three populations in at least half of the sets for half of the RfEs (with clinically significant associations, Table 2). In the other sets, congruence could not be assessed since there was only one significant OR. The sets of ORs demonstrate relationships between symptoms (RfEs), and diagnoses (episode titles), and reflect the expertise of clinicians in interpreting diagnostic relationships and in picking up positive and/or negative associations when they, in fact, exist. Congruency in sets of associations between databases reflects common diagnostic concepts. 'Classical' associations between symptoms and disease-label diagnoses, such as

the symptoms and signs of gastroenteritis and respiratory infections, were strongly and consistently reflected in these data. This latter observation not only reflects the effect of pathophysiology and of a common diagnostic concept but also reflects the expertise of the FD in 'picking up' and recording such diagnostic associations.

Despite the similarities, there were also some differences. For example, the Dutch FDs demonstrated expertise for the RfE 'tiredness' (A04), with more positive and negative associations with various episode titles when compared to the other populations. The Maltese FDs demonstrated good diagnostic processing of respiratory RfEs, as well as other RfEs often associated with acute viral illness. We found that the Serb FDs exhibited a pattern of high diagnostic ORs for disease-label diagnoses (i.e. diagnoses which are labelled with a Component 7 'disease-label' in ICPC,[3,8] such as 'chronic bronchitis'), reflecting a tendency to avoid using symptom diagnoses. This practice may lead to the medicalization of health problems.[15] Differences from the other two databases also emerged for less classical associations, such as for various symptoms and the diagnosis of a 'drug side effect' (A85). Additional differences between databases were noted; it seems that 'pain of the respiratory system' (R01) was conceptualized more as a predictor for tracheitis (R77) by the Maltese FDs but more for bronchitis (R78) by the Dutch and Serb FDs.

Distributions of diagnostic associations—episode title perspective

As for the distribution of RfEs, the congruency of associations was remarkable. Almost half of the episode titles showed congruency among at least half of comparable sets of associations (Table 3). These tests of association support clinician expertise in interpreting relationships between episode titles and RfEs.

Similarities emerged, as expected from the analysis of the data in Table 2 above, such as the congruence between Dutch and Maltese sets of ORs and the preference of Serb doctors for disease-label diagnoses. However, for a few disease-label (ICPC component 7) episode titles, we could not demonstrate good congruency in associations across the three databases. Such episode titles presented with fewer of the commoner RfEs, and this was reflected in a relative lack of observations and consequently less significant ORs due to wider CIs. Examples include 'prevention' and 'no disease' (such as a consultation for paperwork), 'hypertension,' 'hyperlipidaemia' and 'diabetes,' 'hay fever' and 'side effects of medication'. It may be that the classical symptoms for these episode titles are not frequent enough to be included in the top 20 distributions and so we have not studied them. Additionally, some RfEs may not have been frequent enough to present significant ORs. Alternatively, some of these conditions (such as hypertension, hyperlipidemia and diabetes) may be picked up after requests for blood tests results or a blood pressure exam rather than specific symptoms. We have intentionally excluded such RfEs from this study since they are not in Component 1 or 7 of ICPC. Additionally, it may be that these are to a larger extent 'doctor-driven' EoCs (such as prevention, often started by the doctor rather than the patient), which are not strongly associated with specific symptoms. In any case, we could not demonstrate congruency based on our data rather than making any conclusion that such congruency did not indeed exist.

Distributions of diagnostic associations—all sets

The remarkable observed congruency of diagnostic concepts across populations supports a common theoretical framework and a common core diagnostic process in the discipline of FM. Our observations support an international core content of the domain of FM. Diagnostic ORs derived from one population could and should be used to support diagnostic decisions in another. However, the differences we found would suggest that such data should be made available from as many populations as possible.

Effect of frequency of exposure to diagnostic challenges

With a larger data set or a longer observation period, many more clinically significant ORs would probably have also achieved statistical significance. Relaxing our strict significance limits, such as accepting as significant those ORs with CIs excluding unity, would have had a similar effect. We discuss this further in another article in the series.[6] For example, in the case of the RfE cough (R05) and the EoC 'prevention' (A97), the OR was clinically significant at 0.05 for both the Dutch and the Maltese databases (Table 2). However, the CI for the Maltese observation was two hundredths of a unit wider, just outside our limits for statistical significance. The Serb data set is also characterized by a number of associations, which were similar to the other two datasets, but which were also ignored due to the failure to achieve statistical and/or clinical significance.

Whenever RfEs or EoCs are less prevalent in a database,[4] the CI of a diagnostic OR widens due to the smaller number of observations. The converse is also true, and CIs narrow with larger numbers. The latter case allows an association between a symptom and an episode title in a new EoC to be more precisely defined mathematically, when such a diagnostic relationship indeed exists. Greater exposure to diagnostic challenges increases FDs' experience of these challenges but also provides more data on these associations should they indeed exist, with more precise estimates of an OR and narrower CIs. This mathematical relationship effectively expresses the effect of increased

frequency (of exposure) supporting FDs' clinical acumen in the process of diagnosis, through more precisely defined diagnostic relationships with increased exposure to a specific diagnostic challenge. Common RfEs and common EoCs, with higher prevalence and incidence rates, allow a more precise determination of an OR (narrower CI) and thus allow an association between a symptom and an episode title in a new EoC to be more precisely defined mathematically but only when such a relationship indeed exists.

The diagnostic associations that we present reflect underlying biological and pathophysiological effects, which may be universal. Although the actual diagnostic process is not necessarily universal, we have still observed such associations after they are picked up by the FD during the process of diagnosis. Although it is possible that FDs might learn inappropriate diagnostic processes, which are then empirically observed in a study such as ours, it is unlikely that groups of FDs in different countries are all learning the same 'wrong' behaviour. The appropriateness of these diagnostic associations is also evident from examining them rationally and comparing them with available literature (see below). Indeed, many of them seem very appropriate and logical. Frequency (experience) may indeed be a better teacher than formal training (education) alone, and we present our results as evidence to support this.

Does practice indeed make perfect? The case is strengthened by our finding 'high' diagnostic ORs, reflecting strong diagnostic associations and thus, arguably, better diagnostic acumen, for symptoms of episode titles which are more frequent in a population. Examples include acute respiratory and gastrointestinal disease in Malta and tiredness in the Netherlands (see above). Of course, frequency alone is not enough. Should no diagnostic association exist, it is to be expected that one would not be recorded by a competent FD. In fact, the diagnostic associations we have observed are certainly not casual.

Indeed, when one considers that the CI of a diagnostic OR reflects not only the number of observations of the components of the association (the RfE and the episode title), the effect of exposure, but also the number of observations of the interaction itself (OR of RfE for that episode title). The latter is dependent on the FD's expertise in picking up that association and coding it. Only with experienced FDs frequently exposed to that diagnostic challenge would enough data be recorded to have a reliable estimate of the OR, which then reaches clinical and statistical significance.

These observations support the hypothesis that FDs demonstrate improved diagnostic expertise with commoner EoCs, and frequent presenting RfEs, in contrast to rare conditions. In fact, it is logical that FDs can claim expertise in the care of common conditions but should refer rare diseases to secondary care specialists, who can claim more expertise simply on the basis of seeing more cases. We found that our data evidence similar core diagnostic processes in three databases, for common RfEs and common episode titles, but less so for less common EoCs and RfEs. Thus, we have also replicated this aspect of the phenomenon of frequency in our mathematical model with the OR and its CI.

Comparisons with existing literature

Limited literature is available to directly compare with this international study of the process of diagnosis within EoCs in FM.

In 2002, Okkes *et al.*[13] published a distribution of probabilities of specific diagnoses among family practice patients presenting with common symptoms. The availability of accurate estimates of probabilities for common symptoms and complaints was considered to be of great potential for the development of FM as an academic discipline and to support medical decision making. Two limitations of this study were that the data were only available from Dutch FD practices and that the probabilities were presented as percentages rather than odds of a diagnosis at the start of a new EoC. The paper commented that the stratification of such probabilities by age groups increases their clinical relevance, and in fact, we publish such analyses in a companion paper.[6] Cough and shortness of breath were used as examples in this paper, and the distributions of probable diagnoses overlap considerably with the corresponding data in our article (Table 2),[13] as expected, since both studies have included data from the same database over different time periods. In 2005, the same authors published these Dutch data in electronic form with the Wonca ICPC-2-R book.[9] In the 'EFP' program on the included compact disc,[9] the posterior probabilities are calculated slightly differently from the method in this paper. We have chosen to give probabilities measured at the start of a new EoC rather than for a population of patients presenting to the FD.

In 1987 and 1996, Dobbs published two studies on the occurrence of symptoms and signs in cases of urinary tract infection and streptococcal sore throat, respectively.[16,17] The rates of occurrence of individual symptoms and signs were used to develop a Bayesian scoring system for each condition. Bayesian probability scores (B-scores) were calculated, and these could be corrected for a base prevalence rate (prior probability). The B-score values then used to calculate the probability of a urinary tract infection, or streptococcal throat infection, which would respond to antibiotics in individual patients, with good sensitivity and specificity. The B-scores compare very well with the corresponding ORs for 'cystitis' (U71) and 'tonsillitis' (R76) in Table 3. In the case of cystitis (U71), the finding that frequency, dysuria and urgency are important predictors is confirmed by our data and the same applies for haematuria (data not tabulated; this RfE is

not included in the top 20). However, we could not compare the data for nocturia and 'offensive urine' since these clinical concepts do not have separate ICPC codes. We found the odds for 'nausea' to be low (less than unity) but not with a narrow CI (data not tabulated), and as such, we can only partly corroborate the finding of a negative association reported by Dobbs. In the case of tonsillitis (R76), we can confirm the positive diagnostic relationship with the symptoms of 'sore throat,' 'fever' (Table 3), 'soreness on swallowing' and 'enlarged lymph nodes' (data not tabulated) but did not find a significant relationship with the symptoms 'cough,' 'muscle aches' and 'facial flushing,' in contrast to the findings reported by Dobbs. There are no ICPC codes for 'white tonsils' and 'smelly breath', and as such, we could not directly compare these data.[16,17] Although we have found good agreement with our findings, the studies by Dobbs were not international comparisons and were limited to one diagnosis and a pre-selected limited set of symptoms in a cross-sectional study.

In 1992, van Duijn *et al.*[18] studied the LRs of various signs and symptoms of maxillary sinusitis, the diagnosis being confirmed by ultrasound. The only signs and symptoms with an LR of ⩾2 were 'pain in the teeth', nasal polyp and 'purulent nasal secretions', while the symptom 'URTI' had a LR of 1.4, and maxillary and facial pain had LRs of 1.8 and 1.7, respectively.[18] We found significant ORs for cough, URTI, nasal congestion, headache and respiratory system pain (Table 2), which are quite comparable to van Duijn's findings.

A series of studies in the *Journal of the American Medical Association* (JAMA) have looked at LRs for common diagnoses and selected symptoms.[19–26] In comparison to our study, reported LRs were surprisingly low, and often, few associations were found. For example, a study of the evidence base for the diagnosis of influenza failed to find any symptom or sign with an LR of > 2.0,[21] while on the other hand, we found many symptoms to be predictive of a diagnosis of influenza at the start of a new EoC (cough, fever and tiredness in two or more populations and throat symptoms, headache, nasal snuffles and respiratory system pain in at least one population). The contrast is indeed striking! The JAMA study, however, was not a primary care study. The results of a study of urinary tract infection are more comparable, with the JAMA study finding significant associations with dysuria, frequency and back pain, which is comparable to our findings. However, we found stronger associations with higher ORs, while the JAMA study describes LRs in the range from 1.5 to 1.8.[22] In a study of strep throat,[26] again no symptom predictors were found in the JAMA study, and only the signs from a clinical examination were found to be useful. This is in contrast to our findings and those of Dobbs.[16] We feel that this confirms and in fact underlines the utility of primary FM studies in the field of diagnosis.

Other studies of the process of diagnosis in FM using different methodologies, such as case studies presented to FDs, questionnaires to FDs or patients, rating scales or simulated patients, have also typically focussed on one disease or condition at a time. Some studies have used routinely coded EMR data to look at the process of diagnosis, but few of these have collected data on RfEs presented by patients. In many studies, the collection of symptom data has been performed with self-administered questionnaires before the patient entered the consultation room and not with data on symptom(s) actually presented to the FD during usual care. Diagnostic associations in the vignettes were not based on FM data, but rather derived from all available literature, including expert consensus.[27–36] Other studies have tried to improve FDs' diagnostic acumen and adherence to guidelines, for a particular 'under-diagnosed' condition, without collecting any empirical evidence on the quality of routine diagnoses in day-to-day practice.[27–36] None of these studies, except the Transition Project studies, have collected data on all RfEs and EoCs coded during routine care, structured data within EoCs and studied the process of diagnosis in an international comparison.

One finding of this study is the association between exposure and expertise. Does practice indeed make perfect? We found evidence for this in systematic reviews of the process of diagnosis, which found that volume is associated with better outcomes across various medical disciplines.[37,38] We also found evidence of the development of more flexible, and improved, diagnostic approaches (such as 'intuition', or 'pattern recognition') with experience; expertise was also associated with better discrimination between lower and higher risk patients.[39] In fact, previous studies of decision making in medicine have found that an increase in doctors' expertise, and experience, is associated with a preference for different diagnostic decision making processes (such as pattern recognition), and this was in turn associated with much higher odds of making a correct diagnosis.[40] One recent study of diagnosis in FM did not find that more experienced FDs make less misdiagnoses,[36] but this study was limited to uncommon or difficult diagnoses. This is an important limitation since difficult diagnoses are associated with misdiagnoses.[40] Some studies, additionally, tend to be limited to studying self-reported responses to artificial case vignettes and not real-life performance.[36,40]

A systematic review of the relationship between clinical experience and quality of health care, including some studies from FM, reported that half of the studies retrieved did find decreasing trends in knowledge performance and adherence to guidelines as clinicians' age. A fifth of studies found this for only some

outcomes studied, while another fifth found no association. A major limitation of this review was that it was biased towards studies of knowledge assessments and adherence to guidelines, rather than primary outcomes of care processes, and was thus not assessing actual doctor performance in practice. Furthermore, the minority of papers that did indeed assess patient outcomes were disease-specific studies of hospital care (the largest studied myocardial infarction outcomes in hospital).[41]

Different theories of decision making are supported by empirical evidence but present disparate views of medical decision making.[39,42] However, there is good evidence that doctors' approach to diagnosis changes, and improves with experience, even though disease-specific knowledge and quality of care may slowly deteriorate with age.[39] To some degree, the jury is still out on whether practice makes perfect, but there is evidence in the literature that it does, and our findings add evidence to support this.

Evidence based on family practice to support FM practice

The practice of evidence-based medicine requires information on the predictive power of symptoms and signs in the domain. Intuitive and pattern-matching diagnostic processes have been shown to perform well, especially with experienced FDs. However, such an approach has been criticized as not being formally evidence-informed. This argument constitutes a paradox: if evidence to inform practice is not based on evidence from practice, what should it be based on? We defend the study of diagnosis in FM, based on observed data, as an entirely appropriate source of prior probabilities, likelihood functions, and therefore posterior probabilities for diagnoses in FM.

The differences we discovered are arguably due to many effects, including differences in primary health care systems, care delivery, FD training and socio-cultural effects, but also direct FD experience based on frequency of exposure to RfEs and EoCs, and their relationships, and the prevalence and incidence of health problems.[4] However, the similarities in the distributions of diagnostic relationships in these three databases cannot be explained away as simply due to these effects or simply due to the biology or pathophysiology of disease. It appears evident that a core diagnostic process underpins these observations in different populations. This observation supports the internationality of the discipline of FM, at least with respect to the core process of making a diagnosis.

Limitations

The components of the observed variation in associations between populations are complex, being composed of multiple interacting effects (e.g. age, sex, geographical location, culture, socio-economic status, co-morbidity, inter-doctor variation, changing evidence over time, etc.) as discussed above. It is not possible to tease out these effects in a complex adaptive system such as family practice. As such, our responses to the second and third research questions are constrained by the complexity of the system.

The ORs presented represent statistics calculated from a number of practices, which are not corrected for the effect of clustering; the criteria for considering an OR as significant, both clinically and statistically, were thus tightened (see below) to avoid type 1 error. Another publication in the series assesses the effect of inter-doctor and inter-practice variation, to estimate the actual strength of the cluster effect.[6] However, there is evidence that the influence of the health care system on the 'task profile' of FDs seems to have a larger effect on utilization than on distributions of EoCs and patient needs for care (RfEs).[43] In fact, we have found differences in the associations for some episode titles, such as 'prevention' or 'no disease' which may be better understood by analysing distributions of requests for, and provision of, specific medical interventions rather than symptom RfEs.

FDs are often selected to participate in EMR research after they have accepted to record such data carefully, and in depth, over time. Thus, such FDs are often not representative of all FDs in a national system, but rather tend to collect data at a higher level of detail and accuracy than their colleagues, and may have an incentive to do so (financial or academic). The differences in the practice populations (e.g. no registered population in Malta) and the sizes of the databases (number of patients and observation periods) are also unavoidable due to different research circumstances. Consequently, the analysis of such data exhibits many of the qualities and limitations of both qualitative and quantitative research methodologies, sacrificing some generalizability for increasing depth and accepting inherent biases, which cannot be adjusted for without introducing new systematic error.

This paper stimulates the discussion of whether a sample of practices, or a practice population, such as those we have studied, allows one to generalize findings to a regional or national level. The populations we have studied seem comparable, especially with respect to the distributions of diagnostic relationships. Previous literature, including another paper in this series, suggests that the variation between practices in 1 year is less than between years for one practice and that variation impacts differently on different data (such as ICPC chapters, types of interventions, prescriptions).[5,6,43] As such, the issue of inter-doctor variation may be smaller than was previously considered.

The positive and negative LRs themselves would give different perspectives on these diagnostic relationships, but ORs were used as a summary measure for this study. The analysis of the diagnostic relationships was performed one-way (uni-variate). A multi-variate

analysis would theoretically allow the analysis of probabilities, given multiple independent variables, including multiple RfEs. However, a fundamental assumption of such analyses is that variables entered into the model must be independent, normally distributed and have an equivalent standard deviation.[12] This assumption does not hold for multiple RfEs in an encounter since symptoms tend to cluster. We hope to improve on this methodology by correcting for conditional dependence using latent class analysis in a future study.

The comparisons between 41 prevalent EoCs against 36 incident RfEs presents a challenge in the potential risk of describing spurious associations due to the number of statistical tests performed. However, by choosing to accept as clinically significant only those associations with an odds of three or more (or 0.3 or less), and ignoring those observations which are smaller than the width of their 95% CI (a stricter criterion than accepting as significant those whose CI excludes unity), the chance of such spurious associations is compensated for.[12] It is hoped that the problem of the relative lack of power to identify associations with less common RfEs and episode titles will be addressed over time as the databases become larger.

The diagnostic ORs and models presented in this paper are limited in that they represent an analysis of diagnostic associations at the start of an EoC (first encounter for a new episode) and do not take into account that the diagnosis may have changed later on during the episode. Such data are captured in the Transition Project and will be the subject of a planned future study. Furthermore, it is quite possible to miss rare, but important, diagnostic associations due to their infrequent nature, and the wide CI, for such an association. These data guide but do not replace the expertise of an experienced FD.

This study focussed on the relationship between one RfE and one episode title and did not look at the effect of age, sex, doctor, past history and other known predictors of the prior probability of disease in one individual. A companion paper looks at these effects in more detail.[6]

Strengths

In other specialities, the diagnosis is often inappropriately given a disease-label, even if it does not entirely fit the criteria. This is an anomaly often found with the application of such classifications, which are not primary care oriented and which do not facilitate labelling symptom diagnoses, such as the International Classification of Disease (ICD).[3] With ICPC, the availability of the symptom diagnosis keeps other disease-label diagnostic classes (concepts) clean and allows appropriate handling of diagnostic uncertainty. The similarities we have observed are therefore reinforced in their validity.

The data analysed in this research project allow the precise description of relationships between diagnoses and RfEs and the calculation of 'prior' and 'posterior (post-test) probabilities' for a diagnosis. This allows for the study of such probabilities and functions (prior probability, likelihood function, posterior probability) for practically all combinations of RfEs and episode labels coded in ICPC. This is the first such research project to publish such data within an international comparison. These data are of value for decision support systems to support diagnosis in primary care. Such decision support systems[9] are currently available to doctors participating in the Transition Project, and they have been used for clinical care, for research and for educational purposes.

Congruency between diagnostic ORs in different populations was assessed conservatively. Even with these strict criteria, we could demonstrate congruency in around half of the listed RfEs and EoCs, for half or more of the sets of ORs. Our conclusions are strengthened by the clear trends in the data.

Implications of the study

This study reports on an original comparative analysis of the relationships between RfEs and episode titles during routine FM care of practice populations from three countries. The applicability of such data, analysis and methodology is broad and direct, informing clinical practice, education and training, quality assurance and research in FM.

A key feature of this study is the publication of ORs of episode titles, given a RfE at the start of a new EoC. Such data support the interpretation of the predictive power of signs and symptoms in FM, and this has direct application to the area of evidence-based decision support systems driven by the RfE. Our study explores a developing area and should inform future research in the core discipline of diagnosis in FM. The study contributes evidence to support the positive effect of experience on expertise and to support the existence of international common core diagnostic concepts amongst FDs in different countries.

Conclusions

The ICPC, the RfE and the EoC allow the collection of precise data which describe the international epidemiology of FM in depth, including the study of diagnostic associations between RfEs and diagnoses/episode titles, from both the RfE and the episode title perspective.

The congruence of diagnostic associations between populations supports the use of such data from one population in another. However, we found differences in the magnitude of diagnostic associations even though

the associations were similar in direction. Evidently, it is desirable to have diagnostic data from the population they are to be used on for decision support.

Observed differences in distributions of diagnostic associations between RfEs and episode titles in the Transition Project populations could be due to cultural, social, intellectual, religious, spiritual, legal, health care system effects, besides inter-doctor variation and differences in the incidence and prevalence of illness. Evidently, however, there are remarkable similarities and congruencies in the process of diagnosis between these three populations, which are stronger than the above effects.

We propose that both an international (common) and a local (health care system specific) content of FM exist, and the empirical distributions of diagnostic associations that are presented in this paper are a reflection of both these effects. In this respect, the comparative study of distributions of diagnostic ORs is a superior reduction of FM epidemiology than distributions of incidences and prevalences of illness and disease.

We also observed that the frequency of exposure to diagnostic challenges had a strong effect on the CIs of diagnostic ORs reflecting these diagnostic associations. We propose that this constitutes evidence that expertise in FM is associated with frequency of exposure to diagnostic challenges.

Acknowledgements

Author contributions: JKS developed the research methodology, collected data (from Malta), analysed data and was the leading author of the manuscript; IO developed the research methodology, analysed data and participated in writing the manuscript; SO developed the research methodology and analysed data; KvB developed the research methodology and collected data; PZ collected data; MJ collected data; FD developed the research methodology and participated in writing the manuscript; HL developed the research methodology, collected data and analysed data, participated in writing the manuscript until his death in late 2008. The authors would like to thank the associate editor, Prof. Martin Dawes, and the three reviewers who helped improve this paper with their detailed review and feedback. This study would not have been possible without the participation of the Transition Project doctors. From the Netherlands: C van Boven, MD, PhD, Franeker; P H Dijksterhuis, MD, PhD, Wirdum and Olst; A Groen, MD, Amstelveen; J de Haan, MD, Franeker; A M Honselaar-De Groot, MD, Amstelveen; D Janssen, MD, Franeker; T A L Polman, MD, Franeker; G O Polderman, MD, Amstelveen; K E I Stolp, MD, Amstelveen; N Valken, MD, Wirdum; M T M Veltman, MD, PhD (deceased), Amstelveen; M Woerdeman, MD, Amstelveen. From Malta: Francis Paul Calleja, MD, Birkirkara; Carmen Sammut, MD, Siggiewi; Mario R Sammut, MD, MSc, Siggiewi; Daniel Sammut, MD, Zabbar; David Sammut, MD, Zabbar; Jason Bonnici, MD, Zabbar; John Buhagiar, MD, Zabbar; Andrew Baldacchino, MD, Zabbar. From Serbia: the FDs in the region of Kraljevo, part of the International Committee of the Red Cross project.

Declaration

Funding: The European Union Financial Protocol 7 project 'TRANSFoRm' (Grant number FP7 247787 www.transformproject.eu) supported part of the protected time of the authors in performing this study, through its partner the Mediterranean Institute of Primary Care (www.mipc.org.mt).
Ethical approval: The study does not involve collection of new data. Ethical approval has been applied for locally, when appropriate, for studies in the Netherlands, Serbia and Malta.
Conflict of interest: none.

References

[1] Okkes IM, Polderman GO, Fryer GE *et al*. The role of family practice in different health care systems. A comparison of reasons for encounter, diagnoses, and interventions in primary care populations in the Netherlands, Japan, Poland, and the United States. *J Fam Pract* 2002; **51:** 72.

[2] Hofmans-Okkes IM, Lamberts H. The International Classification of Primary Care (ICPC): new applications in research and computer-based primary care information system. *Fam Pract* 1996; **13:** 294–302.

[3] Soler JK, Okkes I, Lamberts H, Wood M. The coming of age of ICPC: celebrating the 21st birthday of the International Classification of Primary Care. *Fam Pract* 2008; **25:** 312–7.

[4] Soler JK, Okkes I, Oskam S *et al*. An international comparative family medicine study of the Transition Project data from the Netherlands, Malta and Serbia. Is family medicine an international discipline? Comparing incidence and prevalence rates of reasons for encounter and diagnostic titles of episodes of care across populations. *Fam Pract* 2012; **29:** 283–98.

[5] Marinus AMF. [Inter-doctor variation in family medicine]. [In Dutch, with a summary in English]. PhD Thesis. University of Amsterdam, the Netherlands, 1993.

[6] Soler JK, Okkes I, Oskam S *et al*. An international comparative family medicine study of the Transition Project data from the Netherlands, Malta, Japan and Serbia. An analysis of diagnostic odds ratios aggregated across age bands, years of observation and individual practices. *Fam Pract* 2012; **29:** 315–31.

[7] The Transition Project. http://www.transitieproject.nl (accessed on 1 November 2009).

[8] Wonca International Classification Committee. *ICPC-2: International Classification of Primary care*, 2nd edn. Prepared by the International Classification Committee of WONCA (WICC). Oxford, UK: Oxford University Press, 1998.

[9] Okkes IM, Oskam SK, Van Boven K, Lamberts H. EFP. Episodes of care in family practice. Epidemiological data based on the routine use of the International Classification of Primary Care (ICPC) in the Transition Project of the Academic Medical Center/University of Amsterdam. (1985–2003). In: Okkes IM, Oskam SK, Lamberts H (eds). *ICPC in the Amsterdam Transition Project*, CD-Rom. Amsterdam, the Netherlands: Academic Medical Center/University of Amsterdam, Department of Family Medicine, 2005.

[10] Feinstein AR. Indexes of contrast and quantitative significance for comparisons of two groups. *Stat Med* 1999; **18:** 2557–81.
[11] Jaeschke R, Guyatt GH, Sackett DL. User's guides to the medical literature III. How to use an article about a diagnostic test. Are the results of the study valid? *JAMA* 1994; **271:** 703–7.
[12] Altman DG, Machin D, Bryant T, Gardner MJ. *Statistics with Confidence. Confidence Intervals and Statistical Guidelines.* BMJ Books, 2000.
[13] Okkes IM, Oskam SK, Lamberts H. The probability of specific diagnoses for patients presenting with common symptoms to Dutch family physicians. *J Fam Pract* 2002; **51:** 31–6.
[14] Guyatt G, Rennie D (Eds). Users' Guides to the Medical Literature. Chicago, IL: American Medical Association, 2002. pp. 121–140.
[15] Soler JK, Okkes I. Reasons for encounter and symptom diagnoses: a superior description of patients' problems in contrast to Medically Unexplained Symptoms (MUS). *Fam Pract* 2012; **29:** 272–82.
[16] Dobbs F. A scoring system for predicting group A streptococcal throat infection. *Br J Gen Pract* 1996; **46:** 461–4.
[17] Dobbs F, Fleming DM. A simple scoring system for evaluating symptoms, history and urine dipstick testing in the diagnosis of urinary tract infection. *J R Coll Gen Pract* 1987; **37:** 100–4.
[18] Van Duijn NP, Brouwer HJ, Lamberts H. Use of symptoms and signs to diagnose maxillary sinusitis in general practice: comparison with ultrasonography. *BMJ* 1992; **305:** 684–7.
[19] Metlay JP, Kapoor WN, Fine MJ. Does this patient have community-acquired pneumonia? Diagnosing pneumonia by history and physical examination. *JAMA* 1997; **278:** 1440–5.
[20] Bundy DG, Byerley JS, Liles A *et al*. Does this child have appendicitis? *JAMA* 2007; **298:** 438–51.
[21] Cass SA, Vollenweider MA, Hornung CA, Simel DL, McKinney WP. Does this patient have influenza? *JAMA* 2005; **293:** 987–97.
[22] Bent S, Nallamothu BK, Simel DL, Fihn SD, Saint S. Does this woman have an acute uncomplicated urinary tract infection? *JAMA* 2002; **287:** 2701–10.
[23] Shaikh N, Morone NE, Lopez J *et al*. Does this child have a urinary tract infection? *JAMA* 2007; **298:** 2895–904.
[24] Margaretten ME, Kohlwes J, Moore D, Bent S. Does this adult patient have septic arthritis? *JAMA* 2007; **297:** 1478–88.
[25] Trowbridge RL, Rutkowski NK, Shojania KG. Does this patient have acute cholecystitis? *JAMA* 2003; **289:** 80–6.
[26] Ebell MH, Smith MA, Barry HC, Ives K, Carey M. The rational clinical examination. Does this patient have strep throat? *JAMA* 2000; **284:** 2912–8.
[27] Armstrong D, Earnshaw G. What constructs do GPs use when diagnosing psychological problems? *Br J Gen Pract* 2004; **54:** 580–3.
[28] Brooks LR, LeBlanc VR, Norman GR. On the difficulty of noticing obvious features in patient appearance. *Psychol Sci* 2000; **11:** 112–7.
[29] Fink W, Lipatov V, Martin Konitzer M. Diagnoses by general practitioners: accuracy and reliability. *Int J Forecast* 2009; **25:** 784–93.
[30] Goldberg D, Steele JJ, Johnson A, Smith C. Ability of primary care physicians to make accurate ratings of psychiatric symptoms. *Arch Gen Psychiatry* 1982; **39:** 829–33.
[31] Heintze C, Matysiak-Klose D, Krohn T *et al*. Diagnostic work-up of rectal bleeding in general practice. *Br J Gen Pract* 2005; **55:** 14–9.
[32] Hjortdahl P, Laerum E, Mowinckel P. Clinical assessment of pharyngitis in general practice. *Scand J Prim Health Care* 1988; **6:** 219–23.
[33] Jensen PM, Lous J. Criteria, performance and diagnostic problems in diagnosing acute otitis media. *Fam Pract* 1999; **16:** 262–8.
[34] Van Den Hoogen HM, Koes BW, Van Eijk JT, Bouter LM. On the accuracy of history, physical examination, and erythrocyte sedimentation rate in diagnosing low back pain in general practice. A criteria-based review of the literature. *Spine* 1995; **20:** 318–27.
[35] Ward DG, Halpin DMG, Seamark DA. How accurate is a diagnosis of asthma in a general practice database? A review of patients' notes and questionnaire-reported symptoms. *Br J Gen Pract* 2004; **54:** 753–8.
[36] Kostopoulou O. Diagnosis of difficult cases in primary care. *J Health Serv Res Policy* 2010; **15 (suppl 1):** 71–4.
[37] Halm EA, Lee C, Chassin MR. Is volume related to outcome in health care? A systematic review and methodological critique of the literature. *Ann Int Med* 2002; **137:** 511–20.
[38] Davoli M, Amato L, Minozzi S *et al*. [Volume and health outcomes: an overview of systematic reviews]. *Epidemiol Prev* 2005; **29 (3–4 suppl):** 3–63.
[39] Reyna VF. A theory of medical decision making and health: fuzzy trace theory. *Med Decis Making* 2008; **28:** 850–65.
[40] Coderre S, Mandin H, Harasym PH, Fick GH. Diagnostic reasoning strategies and diagnostic success. *Med Educ* 2003; **37:** 695–703.
[41] Choudhry NH, Fletcher RH, Soumerai SB. Systematic review: the relationship between clinical experience and quality of health care. *Ann Intern Med* 2005; **142:** 260–73.
[42] Reyna VF. Theories of medical decision making and health: an evidence-based approach. *Med Decis Making* 2008; **28:** 829–33.
[43] Boerma WGW. *Profiles of General Practice in Europe. An International Study of Variation in the Tasks of General Practitioners.* Utrecht, The Netherlands: NIVEL, 2003.