Original article

Validity for Use in Research on Vascular Diseases of the SIDIAP (Information System for the Development of Research in Primary Care): the EMMA Study

Rafel Ramos,[a,b,c,*] Elisabet Balló,[a,d] Jaume Marrugat,[e] Roberto Elosua,[e,f] Joan Sala,[b,c,g] Maria Grau,[e] Joan Vila,[e,f] Bonaventura Bolíbar,[h] Maria García-Gil,[a,c] Ruth Martí,[a] Francesc Fina,[h] Eduardo Hermosilla,[h] Magdalena Rosell,[h] Miguel Angel Muñoz,[i] Daniel Prieto-Alhambra,[h] and Miquel Quesada[a,b,c]

[a] *Unidad de Investigación en Atención Primaria de Girona, IDIAP Jordi Gol, Institut Català de la Salut, Girona, Spain*
[b] *Instituto de Investigación Biomédica de Girona (IdIBGi), Girona, Spain*
[c] *Departamento de Ciencias Médicas, Facultad de Medicina, Universitat de Girona, Girona, Spain*
[d] *Programa de Doctorado en Salud Pública y Metodología de la Investigación Biomédica, Universitat Autònoma de Barcelona, Barcelona, Spain*
[e] *Grupo de Epidemiología y Genética Cardiovascular, Programa de Investigación en Procesos Inflamatorios y Cardiovasculares, IMIM, Barcelona, Spain*
[f] *CIBER de Epidemiología y Salud Pública, Barcelona, Spain*
[g] *Unidad Coronaria, Servicio de Cardiología, Hospital Josep Trueta, Institut Català de la Salut, Girona, Spain*
[h] *Institut d'Investigació en Atenció Primària, IDIAP Jordi Gol, Barcelona, Spain*
[i] *Unidad de Investigación en Atención Primaria de Barcelona, IDIAP Jordi Gol, Institut Català de la Salut, Barcelona, Spain*

ABSTRACT

Introduction and objectives: Information in primary care databases can be useful in research, but the validity of these data needs to be evaluated. We sought to analyze the validity of the data used in the EMMA study based on data from the Information System for the Development of Research in Primary Care.
Methods: We compared the prevalence of cardiovascular risk factors observed in EMMA—hypertension, diabetes, hypercholesterolemia (and its treatments), obesity, and smoking—with equivalent data from the Registre Gironí del Cor (REGICOR), a population-based study that uses standardized methodology, in 2000. We also compared the incidence rates of vascular diseases and its association with these risk factors in a 5-year follow-up.
Results: We analyzed data from 34 823 participants included in EMMA and 2540 REGICOR2000 study participants aged 35 to 74. The prevalence of risk factors did not differ significantly between the 2 studies, except for the prevalence of former smokers in men, which was higher in REGICOR2000 (24.7% [95% confidence interval, 23.9%-25.5%] vs 30.1% [95% confidence interval, 27.1%-33.1%]), and the proportion of patients with lipid-lowering and antihypertensive therapy, which was higher in EMMA (46.9% vs 32.7% and 8.7% vs 6.3%, respectively). There were no differences between the 2 studies when comparing the incidence of vascular diseases (2.1% in both studies in men and 1.18% [95% confidence interval, 0.7%-1.7%] in REGICOR2000 vs 0.75% [95% confidence interval, 0.64%-0.87%] in EMMA in women) and its association with risk factors.
Conclusions: The prevalence of cardiovascular risk factors and their association with the incidence of vascular disease observed in the EMMA study are consistent with those observed in an epidemiological population-based study with a standardized methodology.

© 2011 Sociedad Española de Cardiología. Published by Elsevier España, S.L. All rights reserved.

Validez del Sistema de Información para el Desarrollo de la Investigación en Atención Primaria (SIDIAP) en el estudio de enfermedades vasculares: estudio EMMA

RESUMEN

Introducción y objetivos: Las bases de datos clínicas de atención primaria ofrecen un gran potencial para la investigación. Nuestro objetivo es analizar la validez de los datos del estudio EMMA, basado en el Sistema de Información para el Desarrollo de Investigación en Atención Primaria.
Métodos: Se compararon las prevalencias de los factores de riesgo cardiovascular —hipertensión, diabetes mellitus, hipercolesterolemia (y sus tratamientos), obesidad y tabaquismo— observadas en el EMMA con datos equivalentes del estudio Registre Gironí del Cor (REGICOR) en el año 2000. También se compararon la incidencia de enfermedad vascular y su asociación con dichos factores de riesgo en un seguimiento a 5 años.
Resultados: Se analizaron datos de 34.823 participantes de 35 a 74 años incluidos en el EMMA y 2.540 participantes del estudio REGICOR2000. Las prevalencias de los factores de riesgo no difirieron significativamente entre los dos estudios, excepto para la prevalencia de exfumadores varones, que fue

* Corresponding author: Unidad de Investigación en Atención Primaria de Girona, IDIAP Jordi Gol, Institut Català de la Salut, Maluquer Salvador 11, 17003 Girona, Spain.
E-mail address: rramos.girona.ics@gencat.cat (R. Ramos).

superior en el REGICOR2000 (el 24,7% [intervalo de confianza del 95%, 23,9-25,5%] frente al 30,1% [intervalo de confianza del 95%, 27,1-33,1%]) y la proporción de hipertensos tratados y pacientes con tratamiento hipolipemiante, que fueron superiores en el EMMA (el 46,9 frente al 32,7% y el 8,7 frente al 6,3%, respectivamente). No se observaron diferencias entre uno y otro estudio en la comparación de la incidencia de enfermedad vascular (varones, el 2,1% en ambos estudios; mujeres, el 1,18% [intervalo de confianza del 95%, 0,7-1,7%] en el REGICOR2000 frente al 0,75% [intervalo de confianza del 95%, 0,64-0,87%] en el EMMA) ni en su asociación con los factores de riesgo.
Conclusiones: Las prevalencias de los factores de riesgo cardiovascular y su relación con la incidencia de enfermedades vasculares observada en el EMMA concuerdan con las observadas en un estudio epidemiológico estandarizado de base poblacional.

Abbreviations

EMMA: Atherosclerotic Disease Monitoring Study
HDL-C: high-density lipoprotein cholesterol
ICD: International Classification of Diseases
LDL-C: low-density lipoprotein cholesterol
REGICOR: Girona Heart Registry
SIDIAP: Information System for the Development of Research in Primary Care

INTRODUCTION

Primary care records are an important resource for use in epidemiological studies, assessment of treatment effectiveness, and pharmacovigilance studies.[1–5] Their value stems from the characteristics of the primary care services themselves, which have a wide population coverage, provide regular follow-up, and function as a point of access to the healthcare system for all types of complaints. The usefulness of the records, however, is determined by the reliability and validity of the data they contain.[6,7] Comparison of rates is a validation method involving comparison of prevalences, frequencies of risk factors, or incidences of diseases with those obtained in reference population-based studies using a standardized methodology.[8]

The Information System for the Development of Research in Primary Care (SIDIAP) was designed to provide a valid, reliable database of selected information obtained from the clinical records of patients registered with primary care centers belonging to the Catalan Public Health Service (*Institut Català de la Salut*, ICS) for use in biomedical research.[9] Research into vascular diseases based on the SIDIAP database is currently being developed through the Atherosclerotic Disease Monitoring Study (*Estación de Monitorización de enferMedades Arterioscleróticas*, EMMA).

The aim of this study was to analyze the validity of the SIDIAP database used in the EMMA study by: *a)* comparison of data from a subgroup of participants in the REGICOR2000 study[10] with those obtained in the EMMA study; *b)* comparison of prevalences for the main cardiovascular risk factors and frequencies of their treatment, and *c)* comparison of the incidence of vascular disease and its association with risk factors observed in both studies over a follow-up period of 5 years.

METHODS

As a reference, we used the REGICOR2000 study,[10] an epidemiological, population-based study that used a standardized methodology.

REGICOR2000 Study

Reference Population and Baseline Data

The cohort comprises 3056 individuals recruited in 2000 through a process of random selection based on a census. Sampling was undertaken in 2 steps: first, the populations were randomly selected and then the same number of men and women were randomly recruited in 10-year age groups. Participation was greater than 71%.[10] In this validation study, we included the 2540 patients aged between 35 and 75 years, since this range coincides with that of the EMMA study.

Measures were obtained using standardized procedures.[11] An adapted questionnaire was completed on history and treatment of diabetes mellitus, hypertension, smoking, and dyslipidemia.[11] Blood pressure was measured after 5 min' rest with a regularly calibrated automatic aneroid sphygmomanometer. Two measurements were taken 10 min apart. A standardized smoking questionnaire was used.[11] Weight was measured using a precision scale and the body mass index (BMI) calculated by dividing weight in kilograms by the square of the height in meters.

A blood sample was taken after 10 to 14 h of fasting and concentrations of the following components were analyzed in a central laboratory: glucose, total cholesterol and triglycerides (Roche Diagnostics, Basel, Switzerland), high-density lipoprotein cholesterol (HDL-C) (Boehringer-Mannheim, Germany), and low-density lipoprotein cholesterol (LDL-C) using a direct method with a selective detergent.

The following risk factors were assessed: *a)* hypertension, defined as participants diagnosed with or treated for hypertension or who had a systolic blood pressure $\geq$140 mmHg or a diastolic blood pressure $\geq$90 mmHg; *b)* diabetes mellitus, defined as participants diagnosed with or treated for the condition or who had a serum glucose concentration $\geq$126 mg/dL; *c)* hypercholesterolemia, defined as participants diagnosed with or treated for the condition or who had total cholesterol levels $\geq$250 mg/dL; *d)* obesity, defined as a BMI $\geq$30, and *e)* current smokers (participants who reported smoking at least 1 cigarette per day), exsmokers (greater than 1 year without smoking), and nonsmokers.

Follow-up and Events of Interest

Telephone follow-up was undertaken to determine the vital status of the individuals and question them on the occurrence of cardiovascular events. In addition, participants were cross-referenced with the Catalan death registry (*Registre de Mortalitat de Catalunya*) to identify patients who had died. Angina, myocardial infarction, stroke, and peripheral artery disease were considered events of interest. In cases with suspicion of events having occurred, the hospital clinical history was reviewed. Classification of each event was carried out by a committee after review of the different sources of information using standardized diagnostic criteria. The criteria applied in the MONICA study

(undertaken by the World Health Organization) were used for the diagnosis of myocardial infarction.[11] Cases were classified according to symptoms, electrocardiography (ECG) findings, and markers of myocardial necrosis. Fatal cases were classified according to the information on the death certificate or the results of autopsy (ICD9 category 410; ICD10 categories I21-I22). Angina (ICD9 categories 411.1 and 413; ICD10 category I20) was only recorded in the event of ECG changes or a positive stress test. Stroke included those patients with a primary diagnosis in the hospital clinical history of hemorrhagic stroke, ischemic stroke, or stroke of unknown etiology (ICD9 categories 433.X1, 434.X1, and 438; ICD10 category I63). Subarachnoid hemorrhage and transient ischemic attack were excluded.

The presence of peripheral artery disease (ICD9 categories 411.1 and 413; ICD10 categories I73, I73.8, and I73.9) was recorded when a diagnostic arteriogram or Doppler scan (ankle-to-arm ratio <0.9) was present or in the case of lower-limb amputation, ulcers, or gangrene due to ischemia.

A combined variable was defined for the presence of cardiovascular disease during follow-up that included the first event of those defined above (angina, myocardial infarction, stroke, and peripheral artery disease).

The EMMA Study

Reference Population and Baseline Data

The reference population for the EMMA study comprised all individuals aged between 35 and 74 years who were registered at one of the 23 ICS primary health care centers in Girona, Spain between 1998 and 2002 (197 620 patients registered in 2000). The health centers are made up of basic care units (BCU), each with a physician and a nurse who are responsible for the same group of patients. All patients registered with the BCU for whom clinical records were of sufficient quality were included in the study. Quality of records was assessed based on 5 indicators: *a)* population coverage (defined as the proportion of registered patients who have attended the unit in 1 year) >70%; *b)* proportion of registered patients who had attended the unit without any diagnosis recorded in the clinical history <5%; *c)* mean number of diagnoses recorded for each registered patient >4 (50th percentile of the distribution of the mean number of diagnoses per physician); *d)* prevalence of smokers >20% (50th percentile of the distribution of the prevalence of smoking in the basic care units), and *e)* prevalence of heart failure >1.7% (50th percentile of the distribution of prevalence among the basic care units). Among the patients assigned to these units, those for whom complete information was available on the study variables were included in the analysis.

Data were obtained from the database on sex, age, systolic and diastolic blood pressure, glucose, total cholesterol, triglycerides, HDL-C, LDL-C, weight, and height. For inclusion, participants needed to have had data recorded for the study variables at some point between 1998 and 2002, with no more than 6 months elapsed between the first and last measure, to ensure that the data were contemporaneous. If data were available for more than 1 timepoint during this period, the median of the values was obtained. The last date at which data were obtained on risk factors was considered the date of inclusion for the study participant.

Diagnoses were extracted from electronic records (which follow the recommendations of the clinical practice guidelines of the ICS[12]) with a date prior to the inclusion of each participant. The presence of risk factors was defined according to the following criteria:

- Diabetes mellitus if the patient history included this diagnosis (ICD10 categories E11, E12, E14, and subcategories thereof) or the patient had a fasting glucose concentration ≥126 mg/dL or was receiving glucose-lowering medication.
- Hypertension if this was recorded in the patient history (ICD10 categories I10, I15, and subcategories thereof) or if the patient had a systolic blood pressure ≥140 mmHg, a diastolic blood pressure ≥90 mmHg, or was receiving antihypertensive medication.
- Hypercholesterolemia if this was recorded in the patient history (ICD10 category E78 and its subcategories, except for E78.3 and E78.6) or if the patient had total fasting cholesterol levels >250 mg/dL or was receiving lipid-lowering drugs.
- Smoking if it was recorded in the patient history (ICD10 category F17 for smokers or Z72.0 for exsmokers). Participants diagnosed as smokers who had not been registered with the unit for more than 1 year were also considered exsmokers.
- Obesity if it was recorded in the patient history (ICD10 category E66 and subcategories except for E66.1 and E66.2) or if the patient had a BMI ≥30.

Follow-up and Events of Interest

The sources of information used for follow-up were hospital and emergency department discharge forms, primary care patient records, *Registre de Mortalitat de Catalunya*, and the Spanish registry of myocardial infarctions. The same coding system as employed in REGICOR2000 was used to categorize events. Cases derived solely from emergency department records (mainly angina, transient ischemic attack, or cases of peripheral artery disease) were reviewed to confirm the diagnosis.

Statistical Analysis

Prevalences and frequencies of risk factors, their treatment, and the respective 95% confidence intervals (95%CI), standardized using a direct standardization technique, were calculated according to the age and sex distribution of the world population[13] using the EPIDAT 3.1 program.

The annual cumulative incidence of vascular events (angina, myocardial infarction, stroke, and peripheral artery disease) was calculated for each study, and the hazard ratio (HR) for the variable "source of the data" (REGICOR2000 vs EMMA) was calculated using a Cox proportional hazards model to determine whether the study from which the data were obtained affected the incidence of vascular disease observed. The HRs and their 95%CIs were also calculated for the cumulative incidence of vascular disease at 5 years associated with the main risk factors (hypertension, diabetes mellitus, smoking, and obesity) using a Cox proportional hazards model adjusted for age and sex.

Statistical power: the sample size for male patients, which is the smallest (16 846 individuals in the EMMA study and 1243 in the REGICOR2000 study), provides a statistical power of 87% to detect a significant difference (*P*<.05) of 4.5 percentage units between the 2 studies for a factor that is present in 50% of the individuals (worst-case scenario) and a statistical power of 80% to detect a difference of 1% between the 2 incidences when it was predicted that the 5-year incidence of vascular disease would be 2%.

RESULTS

Out of 212 basic care units, 57 (26.9%) met the data quality criteria. These units were distributed among 16 of the 23 ICS

healthcare centers (69.9%) in Girona. The population of patients aged between 35 and 79 years registered with these units was 59 340, of whom 38 088 (64.2%) had complete information available on the study variables and were included in the EMMA study. In this validation study, the 34 823 participants aged between 35 and 74 years were included, since this is the age range coinciding with the REGICOR2000 study. Figure 1 shows a flow diagram for the inclusion of the participants.

Table 1 shows a comparison of the means and 95% CIs for the risk factors between the 2 studies. In both sexes, the age-adjusted mean for all of the risk factors analyzed in the EMMA study fell within the 95%CI for the estimates obtained in REGICOR2000. A comparison of the age-adjusted prevalences is shown in Figure 2. Again, in both sexes, the prevalence of all risk factors analyzed in the EMMA study fell within the 95%CI of the estimates obtained in REGICOR2000, except for the prevalence of exsmokers in male patients, which was 24.8% in EMMA and 30% in REGICOR2000. There were no differences in the proportion of women receiving treatment, whereas the proportion of men receiving treatment for hypertension or hyperlipidemia was higher in EMMA (46.9% vs 32.7% and 8.7% vs 6.3%, respectively) (Fig. 3). The unadjusted prevalences for the risk factors are shown in Table 2.

The cumulative incidence of vascular disease (angina, myocardial infarction, stroke, and peripheral artery disease) at 5 years (Fig. 4) and the age- and sex-adjusted HRs for the association between cardiovascular disease and hypertension, diabetes mellitus, and smoking (Fig. 5) observed in the EMMA study did not differ from that observed in REGICOR2000. The HR for the variable origin of the data (REGICOR2000 vs EMMA) for the estimated incidence of vascular disease was 1.02 (95%CI, 0.73-1.42; *P*=.91).
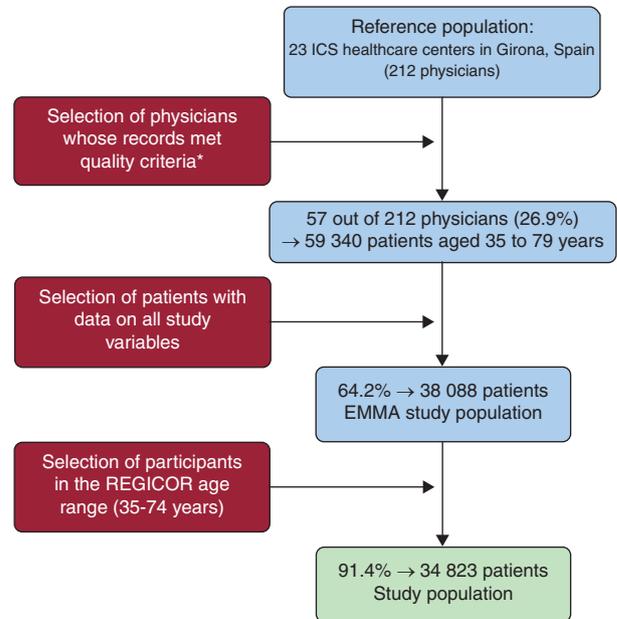

Figure 1. Flow diagram showing the selection of participants for inclusion in the EMMA study ICS, Catalan Public Health Service; REGICOR, Girona Heart Registry. * The following criteria were applied to select professionals with high-quality records: *a)* population coverage (defined as the proportion of registered patients who have attended the unit in 1 year) >70%; *b)* proportion of registered patients who had attended the unit without any diagnosis recorded in the clinical history <5%, and *c)* mean number of diagnoses recorded for each registered patient >4 (50th percentile of the distribution of the mean number of diagnoses per physician), prevalence of smokers >20% (50th percentile of the distribution of the prevalence of smoking in the basic care units), and prevalence of heart failure >1.7% (50th percentile of the distribution of prevalence among the basic care units).
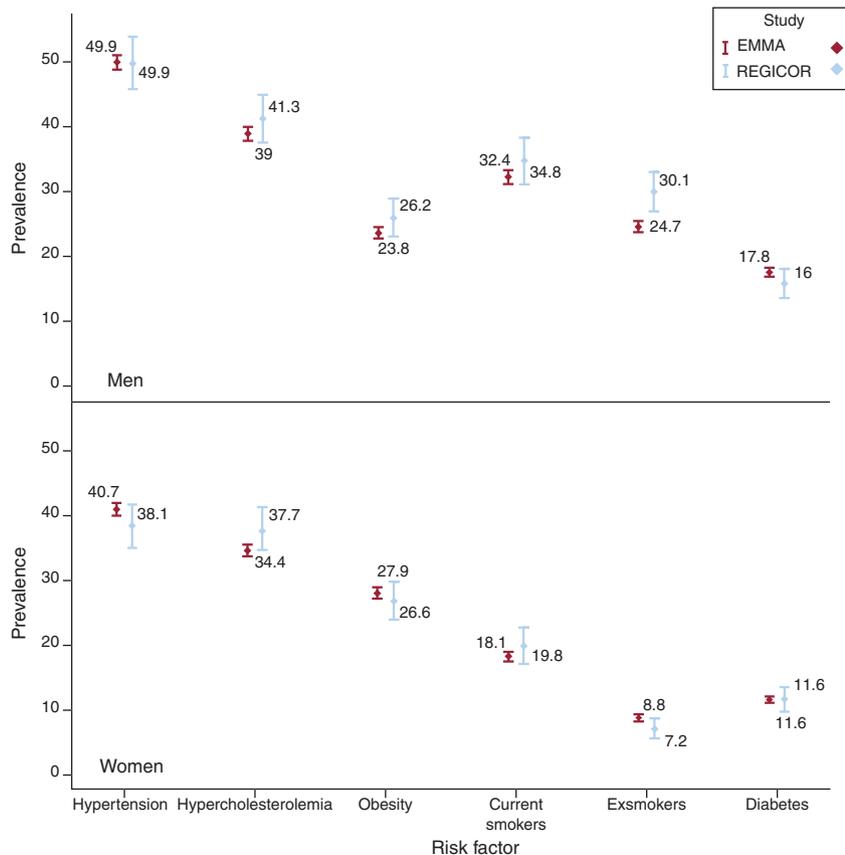

Figure 2. Comparison of age-adjusted prevalences (whiskers show 95% confidence intervals) of hypertension, diabetes mellitus, hypercholesterolemia, obesity, smokers, and nonsmokers in women and men observed between the EMMA and REGICOR2000 studies. REGICOR, Girona Heart Registry.

Table 1
Comparison of the Age-Adjusted Means of Variables Associated With Risk Factors Observed in Men and Women in the EMMA and REGICOR2000 Studies

	EMMA, mean (95%CI)	REGICOR2000, mean (95%CI)
Men, no.	16 846	1243
Systolic arterial pressure, mmHg	133.8 (131.8-135.8)	134.7 (128.1-141.4)
Diastolic blood pressure, mmHg	80.5 (78.9-82)	82.9 (77.6-88.2)
Body mass index, kg/m^2	27.8 (26.9-28.7)	27.8 (24.7-30.8)
Glucose concentration, mg/dL	106.5 (104.7-108.3)	107.7 (101.5-113.9)
Total cholesterol, mg/dL	221.5 (218.9-224.1)	222 (213-230.9)
HDL-C, mg/dL	47.9 (46.8-49.1)	46.3 (42.2-50.4)
LDL-C, mg/dL	149.1 (147-151.3)	151.4 (143.8-158.9)
Triglycerides, mg/dL	121.9 (120-123.9)	125 (118.2-131.7)
Women, no.	17 977	1297
Systolic arterial pressure, mmHg	127.8 (125.9-129.6)	124.6 (118.3-130.8)
Diastolic blood pressure, mmHg	78 (76.6-78.1)	78.1 (73.1-83.1)
Body mass index, kg/m^2	27.6 (26.8-28.5)	27.4 (24.4-30.3)
Glucose concentration, mg/dL	96.6 (95-98.2)	100.2 (94.4-106.1)
Total cholesterol, mg/dL	221.8 (219.4-224.2)	222.7 (214.1-231.3)
HDL-C, mg/dL	57 (55.8-58.2)	55.4 (51-59.8)
LDL-C, mg/dL	145.5 (143.6-147.4)	146.6 (139.4-153.8)
Triglycerides, mg/dL	96.5 (94.9-98)	92.4 (86.9-98)

95%CI, 95% confidence interval; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; REGICOR, Girona Heart Registry.
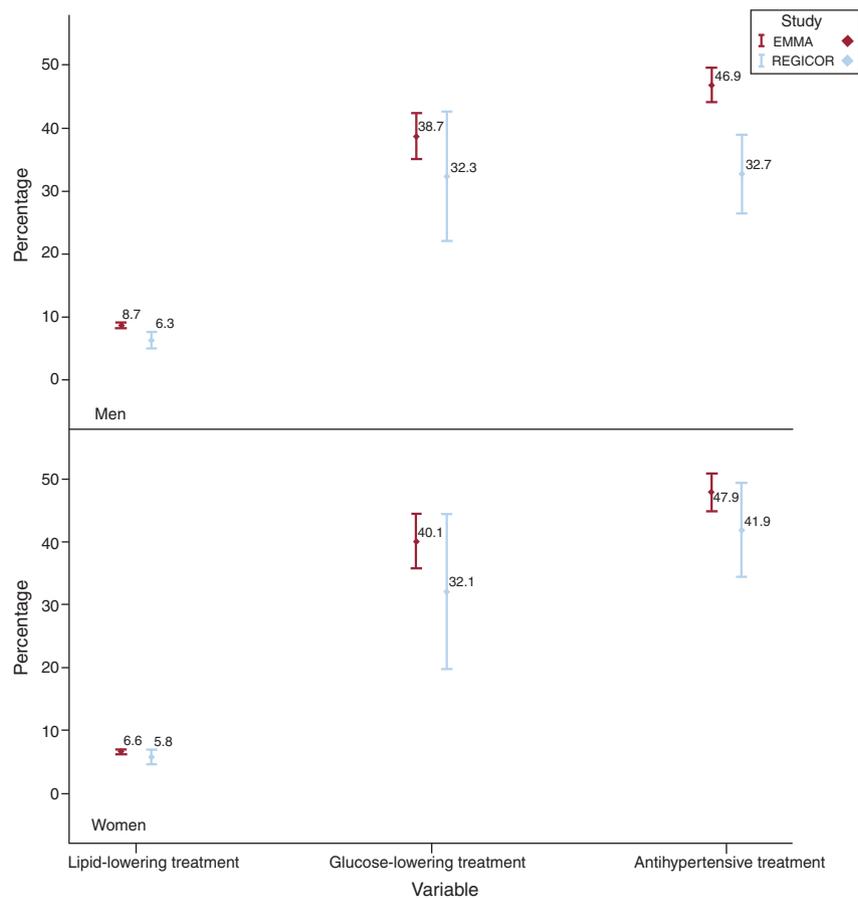

Figure 3. Comparison of the proportion (whiskers show 95% confidence intervals) of men and women receiving treatment for hypertension, diabetes, and hyperlipidemia in the EMMA and REGICOR2000 studies. REGICOR, Girona Heart Registry.

Table 2
Comparison of Unadjusted Prevalences (95% confidence interval) for Risk Factors Observed in Men and Women From the EMMA and REGICOR2000 Studies.

	EMMA	REGICOR2000
Men, no.	16 846	1243
Hypertension	59.9 (59.2-60.6)	54.8 (52-58.4)
Hypercholesterolemia	40.6 (39.9-41.4)	39.1 (36.4-41.8)
Obesity	24.6 (23.9-25.2)	23.8 (21.5-26.3)
Active smokers	26.2 (25.6-26.9)	32.4 (29.8-35.1)
Exsmokers	28.5 (27.8-29.2)	27.7 (22.4-27.2)
Diabetes mellitus	21.7 (21.1-22.3)	17.8 (15.8-20)
Women, no.	17 977	1297
Hypertension	52.2 (51.5-52.9)	44.6 (41.9-47.3)
Hypercholesterolemia	41.3 (41-42)	41.9 (39.3-44.7)
Obesity	31 (30.3-31.7)	29.6 (27.2-32.2)
Active smokers	12.2 (11.7-12.7)	15.8 (13.9-17.9)
Exsmokers	8.3 (7.9-8.7)	6.3 (5.1-7.7)
Diabetes mellitus	15.1 (14.5-15.6)	13.3 (11.6-15.3)

REGICOR, Girona Heart Registry.

DISCUSSION

The prevalence of risk factors, the 5-year incidence of cardiovascular disease, and the HR for the incidence of cardiovascular disease associated with the main risk factors (hypertension, diabetes mellitus, hypercholesterolemia, and smoking) observed in the EMMA study are similar to those observed in REGICOR2000.

Despite differences in the methodology used and the regional variation observed in studies of cardiovascular risk factors,[14,15] the prevalences observed in the EMMA study are consistent with those reported for hypertension,[14,16–18] obesity,[14,19–22] hypercholesterolemia,[17,19] and smoking[14,21] in Spain.

In the case of diabetes mellitus, the prevalence in men was higher than that observed in the ERICE study,[14] although it was similar to those reported by Baena Díez et al.[17] in Catalonia and Rigo et al.[20] in the Balearic Islands. The ERICE study revealed that the Mediterranean region has a high prevalence of diabetes mellitus.

Comparison of rates has been widely used to validate the most productive databases employed in research, such as the Database for Pharmacoepidemiological Research in Primary Care (BIFAP) and the QRESEARCH database, where the external validity and representativeness were comparable to those obtained in the EMMA study for hypertension[23] (BIFAP observed a difference close to 1.5% compared with the prevalence observed for both sexes in the national survey), obesity[24] (QRESEARCH observed a difference of 1% compared with the prevalence observed in the Health Survey for England), and diabetes mellitus[23,25] (QRESEARCH observed a difference of 2.9% compared with the prevalence observed in the Royal College of General Practitioners database and BIFAP observed a difference of around 0.2% compared with the prevalence observed in the national healthcare survey).
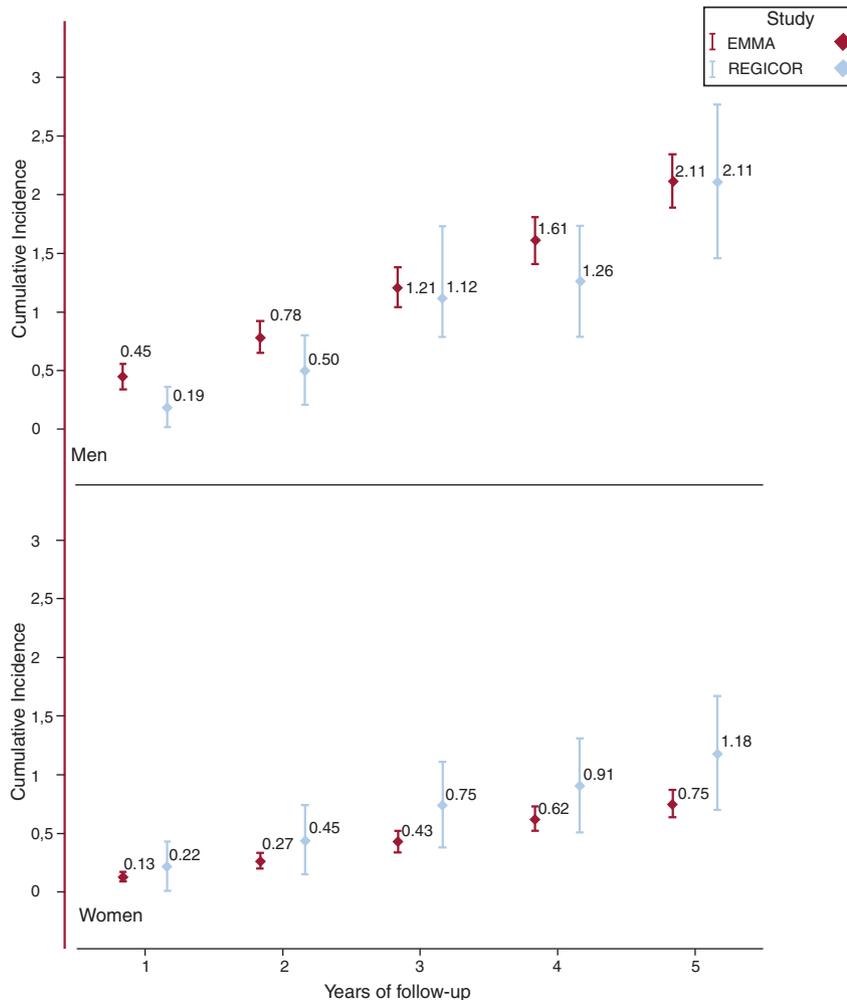

Figure 4. Comparison of the annual cumulative incidence (whiskers show 95% confidence intervals) of cardiovascular disease (angina, myocardial infarction, stroke, and peripheral artery disease) during 5 years of follow-up in the EMMA and REGICOR2000 studies. REGICOR, Girona Heart Registry.
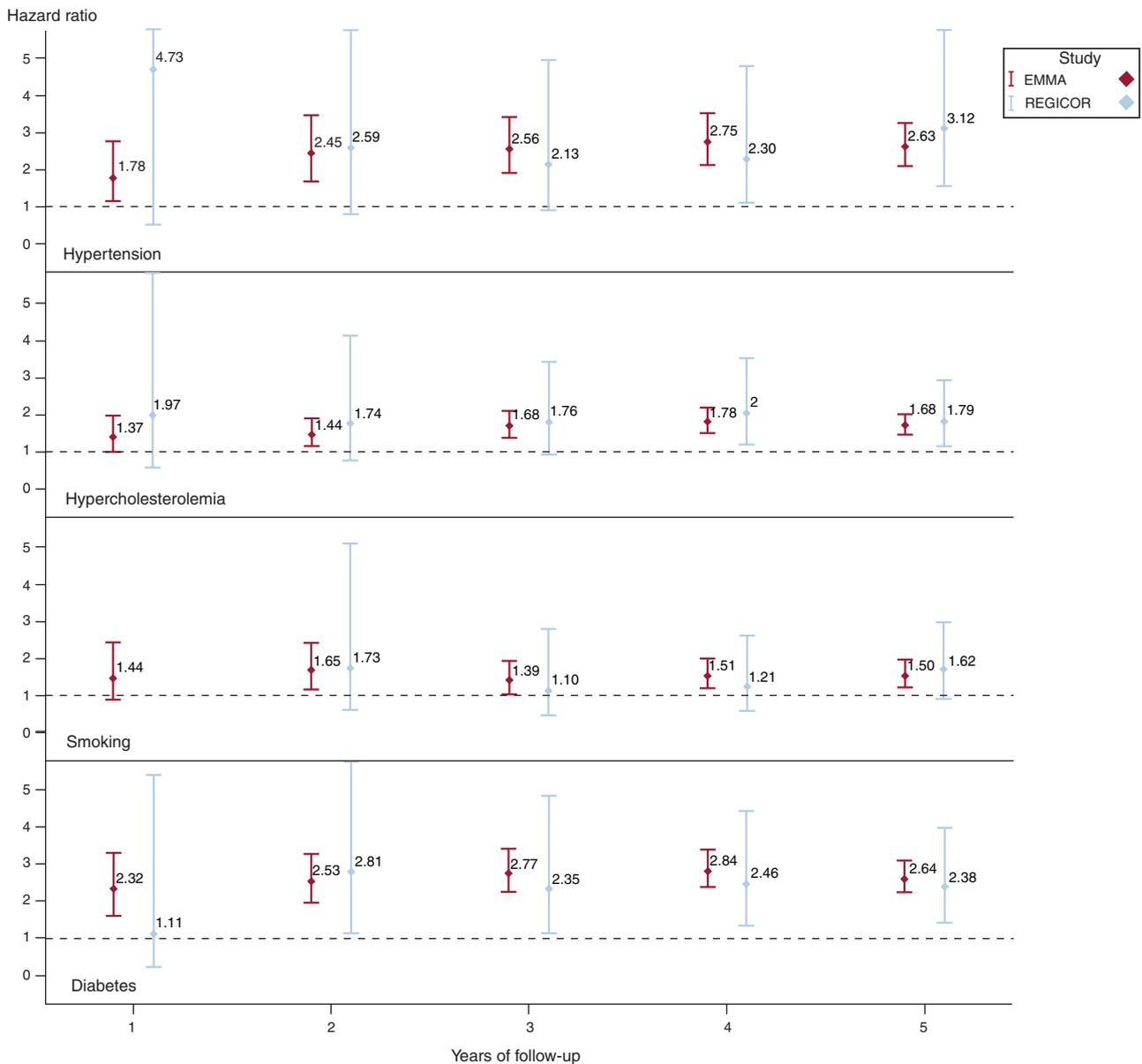
Figure 5. Comparison of the age- and sex-adjusted hazard ratios (whiskers show 95% confidence intervals) for the association with cardiovascular disease of the risk factors hypertension, diabetes mellitus, hypercholesterolemia during 5-year follow-up in the EMMA and REGICOR2000 studies. REGICOR, Girona Heart Registry.

The low prevalence of exsmokers observed in the EMMA study can be explained by the fact that this variable was not collected exhaustively during digitalization of patient records when the patient had stopped smoking many years earlier or reported being a nonsmoker. The larger proportion of men receiving antihypertensive therapy in EMMA may be partly due to the very small percentage of young men in REGICOR2000 who were treated and to the fact that, following adjustment for age, the proportion of treated patients went from 45% (unadjusted percentage) to 32% (adjusted percentage) whereas in EMMA the rate went from 49% to 46% following adjustment for age. The prevalence of hypertension in EMMA is similar to that observed in other studies.[26]

The absence of significant differences in the comparison of the HRs for the different risk factors for the incidence of cardiovascular diseases also indicates a high internal validity and accuracy of the diagnosis of the risk factors recorded. The observed HRs do not differ in magnitude from those reported in other studies.[27–30]

Limitations and Characteristics of the Study

Although the validation method applied in this study has been used in other consolidated databases,[7,8] the main limitation of this study is the impossibility of guaranteeing the quality of the coding, since the information recorded for each patient was not compared individually. Nevertheless, the agreement observed in the comparison of the HRs of the different risk factors for the incidence of cardiovascular diseases confirms the high level of validity of the database. Although a trend was observed toward higher rates of treatment in the EMMA study, this was only statistically significant for hypertension and hypercholesterolemia in men; nevertheless, the magnitude of the difference was notable for the other factors analyzed. This finding is consistent with the fact that participants in the EMMA study were seen in a clinical setting and therefore more likely to be medicated, and it may also be explained in part by the selection of patients with complete information for the study variables in EMMA.

Another limitation relates to the choice of REGICOR2000 as a reference. It remains possible that for certain measures, such as risk factors that require repeated measures, epidemiological studies in which data are obtained from single measures may incorrectly classify a proportion of patients.

A key element in obtaining good internal validity is the selection of high-quality records, as in the EMMA study. This selection is designed to minimize possible biases in the recording of information that could occur in patient records compared with traditional epidemiological studies. In the case of the EMMA study, these criteria were arbitrarily chosen based on the consensus of the investigators and, although the results confirm their validity, they nevertheless represent a possible selection bias. Consequently, the external validity of the EMMA study can only be extrapolated to the population of patients treated by those physicians whose records met the quality criteria applied for the database.

In general, clinical databases used as tools for research have the following strengths: sample size, representativeness of the participants, minimization of recall bias, an almost complete medical history including comorbidity data, and long periods of follow-up. These characteristics make it possible to study rare symptoms or diseases, to analyze very specific populations, and to observe infrequent events or effects that occur after a long delay. They also make it possible to carry out pharmacovigilance studies or analyses of treatment effectiveness, as has been the case with the BIFAP database,[31] a pioneer in this area in Spain.

EMMA is the first cohort study derived from the SIDIAP database. This database offers considerable advantages for use in epidemiological studies and can offer reliable results without substantial financial investment and in a shorter time than traditional epidemiological studies. Its use can therefore complement traditional epidemiological studies. Further validation studies are now required to confirm the potential of the SIDIAP database.

CONCLUSIONS

The prevalence of traditional risk factors for cardiovascular disease and their relationship with the incidence of vascular diseases observed in the EMMA study are consistent with those observed in a population-based epidemiological study with a well-documented standardized methodology. These results indicate a high level of validity and good representativeness of the population in the EMMA study and the SIDIAP database for use in epidemiological studies of cardiovascular disease.

ACKNOWLEDGMENTS

FUNDING

CONFLICTS OF INTEREST

None declared.

REFERENCES

1. GPRD: Much more than the database [webpage]. General Practice Research Database [accessed 2010 Jun 20]. Available from: http://www.gprd.com
2. QRESEARCH specialises in research & analyses using primary care electronic health data [webpage]. QRESEARCH database, The University of Nottingham [accessed 2010 Jun 20]. Available from: http://www.Qresearch.org
3. Salvador Rosa A, Moreno Pérez JC, Sonego D, García Rodríguez LA, De Abajo Iglesias FJ. El proyecto BIFAP: Base de datos para la Investigación Farmacoepidemiológica en Atención Primaria. Aten Primaria. 2002;10:655–61.
4. The finprog study, a Nationwide Breastcancer Database [webpage]. Department of Oncology, Helsinki University Central Hospital [accessed 2010 Jun 20]. Available from: http://www.finprog.org
5. The Health Improvement Network [webpage]. In Practice Systems Ltd, EPIC [accessed 2010 Jun 20]. Available from: http://www.thin-uk.com
6. Thiru K, Hassey A, Sullivan F. Systematic review of scope and quality of electronic patient record data in primary care. BMJ. 2003;326:1070.
7. Khan NF, Harrison SE, Rose PW. Validity of diagnostic coding within the General Practice Research Database: a systematic review. Br J Gen Pract. 2010;60:160–1.
8. Herrett E, Thomas SL, Marieke Schoonen W, Smeeth L, Hall AJ. Validation and validity of diagnoses in The General Practice Research Database: a systematic review. Br J Clin Pharmacol. 2010;69:4–14.
9. Sistema de Información para el Desarrollo de la Investigación en Atención Primaria SIDIAP [webpage]. Instituto de Investigación en Atención Primaria (IDIAP Jordi Gol), Institut Català de la Salut [accessed 2010 Sep 29]. Available from: http://www.idiapjgol.org/sidiap.php
10. Grau M, Subirana I, Elosua R, Solanas P, Ramos R, Masiá R, et al.; on behalf of the REGICOR Investigators. Trends in cardiovascular risk factor prevalence (1995-2000-2005) in northeastern Spain. Eur J Cardiovasc Prev Rehabil. 2007;14:653–9.
11. Manual of The MONICA Project [Internet manual]. Geneva: World Health Organisation; 2000 [accessed 2008 Jan 8]. Available from: http://www.ktl.fi/publications/monica/manual/index.htm
12. Guías de práctica clínica de Atención Primaria [Internet series]. Institut Català de la Salut [accessed 2010 Jun 14]. Available from: http://www.ics.gencat.net/3clics
13. Ahmad OB, Boschi-Pinto C, Lopez AD, Murray CJL, Lozano R, Inoue M. Age standardization of rates: a new WHO standard. Global Programme on Evidence for Health Policy Discussion. Paper No. 31 [Internet series]. Geneva: World Health Organization; 2001 [accessed 2010 Jun 14]. Available from: http://www.who.int/healthinfo/paper31.pdf
14. Gabriel R, Alonso M, Segura A, Tormo MJ, Artigao LM, Banegas JR, et al. Prevalencia, distribución y variabilidad geográfica de los principales factores de riesgo cardiovascular en España. Análisis agrupado de datos individuales de estudios epidemiológicos poblacionales: estudio ERICE. Rev Esp Cardiol. 2008;61:1030–40.
15. Grau M, Elosua R, Cabrera de León A, Guembe MJ, Baena-Díez JM, Vega Alonso T, et al. Factores de riesgo cardiovascular en España en la primera década del siglo XXI: análisis agrupado con datos individuales de 11 estudios de base poblacional, estudio DARIOS. Rev Esp Cardiol. 2011;64:295–304.
16. Wolf-Maier K, Cooper RS, Banegas JR, Giampaoli S, Hense HW, Joffres M, et al. Hypertension prevalence and blood pressure levels in 6 European countries, Canada, and the United States. JAMA. 2003;289:2363–9.
17. Baena-Díez JM, Vidal-Solsona M, Byram AO, González-Casafont I, Ledesma-Ulloa G, Martí-Sans N. Epidemiología de las enfermedades cardiovasculares en atención primaria. Estudio Cohorte Zona Franca de Barcelona. Rev Esp Cardiol. 2010;63:1261–9.
18. Segura A, Rius G. Factores de riesgo cardiovascular en una población rural de Castilla-La Mancha. Rev Esp Cardiol. 1999;52:577–88.
19. Vegazo O, Banegas JR, Civeira F, Serrano Aisa PL, Jiménez FJ, Luengo E; Estudio HISPALIPID. Prevalencia de dislipemia en las consultas ambulatorias del Sistema Nacional de Salud: Estudio HISPALIPID. Med Clin (Barc). 2006;127:331–4.
20. Rigo F, Frontera G, Llobera J, Rodríguez T, Borrás I, Fuentespina E. Prevalencia de factores de riesgo cardiovascular en las Islas Baleares (estudio CORSAIB). Rev Esp Cardiol. 2005;58:1411–9.
21. Huerta JM, Tormo MJ, Egea-Caparrós JM, Ortolá-Devesa JB, Navarro C. Validez del diagnóstico referido de diabetes, hipertensión e hiperlipemia en población adulta española. Resultados del estudio DINO. Rev Esp Cardiol. 2009;62:143–52.
22. García-Ortiz L, Grandes G, Sánchez-Pérez A, Montoya I, Iglesias-Valiente JA, Recio-Rodríguez JI, et al. Efecto en el riesgo cardiovascular de una intervención para la promoción del ejercicio físico en sujetos sedentarios por el médico de familia. Rev Esp Cardiol. 2010;63:1244–52.
23. Estudios de validación de la base de datos BIFAP [webpage]. Proyecto BIFAP, Base de datos para la Investigación Farmacoepidemiológica en Atención Primaria [accessed 2011 Mar 18]. Available from: http://www.bifap.org/actividad-cientifica.php#estudios
24. Hippisley-Cox J, Ronan R. Obesity in the United Kingdom: Analysis of QRESEARCH data [accessed 2010 Jun]. Available from: http://www.qresearch.org/Public_Documents/DataValidation/Obesity%20in%20the%20UK.pdf

25. Hippisley-Cox J, Ronan R. Diabetes in the United Kingdom: Analysis of QRESEARCH data [accessed 2010 Jun]. Available from: http://www.qresearch.org/Public_Documents/DataValidation/Diabetes%20in%20the%20UK%20analysis%20of%20QRESEARCH%20data.pdf

26. Banegas Banegas JR, Villar Alvarez F, Pérez de Andrés C, Jiménez García-Pascual R, Gil López E, Muñiz García J, et al. Estudio epidemiológico de los factores de riesgo cardiovascular en la población española de 35 a 64 años. Rev San Hig Pub. 1993;67:419–55.

27. Prescott E, Hippe M, Schnohr P, Hein HO, Vestbo J. Smoking and risk of myocardial infarction in women and men: longitudinal population study. BMJ. 1998;316:1043–7.

28. Folsom AR, Szklo M, Stevens J, Liao F, Smith R, Eckfeldt JH. A prospective study of coronary heart disease in relation to fasting insulin, glucose and diabetes. The Atherosclerosis Risk in Communities (ARIC) Study. Diabetes Care. 1997;20:935–42.

29. Kromhout D. On the waves of the Seven Countries Study. Eur Heart J. 1999;20:796–802.

30. Kannel WB. Blood pressure as a cardiovascular risk factor: prevention and treatment. JAMA. 1996;275:1571–6.

31. Hernández-Díaz S, García Rodríguez LA. Cardioprotective aspirin users and their excess risk of upper gastrointestinal complications. BMC Medicine. 2006;4:22.