

# Risk adjustment of health-care performance measures in a multinational register-based study: A pragmatic approach to a complicated topic

SAGE Open Medicine  
2: 2050312114526589  
© The Author(s) 2014  
Reprints and permissions:  
sagepub.co.uk/journalsPermissions.nav  
DOI: 10.1177/2050312114526589  
smo.sagepub.com  


Tron Anders Moger<sup>1</sup> and Mikko Peltola<sup>2</sup>

## Abstract

**Objectives:** Health-care performance comparisons across countries are gaining popularity. In such comparisons, the risk adjustment methodology plays a key role for meaningful comparisons. However, comparisons may be complicated by the fact that not all participating countries are allowed to share their data across borders, meaning that only simple methods are easily used for the risk adjustment. In this study, we develop a pragmatic approach using patient-level register data from Finland, Hungary, Italy, Norway, and Sweden.

**Methods:** Data on acute myocardial infarction patients were gathered from health-care registers in several countries. In addition to unadjusted estimates, we studied the effects of adjusting for age, gender, and a number of comorbidities. The stability of estimates for 90-day mortality and length of stay of the first hospital episode following diagnosis of acute myocardial infarction is studied graphically, using different choices of reference data. Logistic regression models are used for mortality, and negative binomial models are used for length of stay.

**Results:** Results from the sensitivity analysis show that the various models of risk adjustment give similar results for the countries, with some exceptions for Hungary and Italy. Based on the results, in Finland and Hungary, the 90-day mortality after acute myocardial infarction is higher than in Italy, Norway, and Sweden.

**Conclusion:** Health-care registers give encouraging possibilities to performance measurement and enable the comparison of entire patient populations between countries. Risk adjustment methodology is affected by the availability of data, and thus, the building of risk adjustment methodology must be transparent, especially when doing multinational comparative research. In that case, even basic methods of risk adjustment may still be valuable.

## Keywords

Risk adjustment, hospital discharge data, international comparisons

Date received: 20 December 2013; accepted: 6 February 2014

## Introduction

A major theme in health services research is to develop performance indicators and promote the use of benchmarking information for health policy. In developed countries, there are national benchmarking projects ongoing, but cross-country comparisons on national level are very few. Especially, there is not much information of health system performance between countries based on patient-level data. A central goal of *European Health Care Outcomes, Performance and Efficiency* (EuroHOPE) project is to develop performance indicators and to evaluate the performance of European health-care systems in terms of outcomes, quality, use of resources, and costs.<sup>1</sup> Multinational patient-level studies of health system performance are

hampered most by data availability and the lack of unique patient identifiers.<sup>2</sup> In the EuroHOPE partner countries, linkable patient-level administrative data for use of in- and

<sup>1</sup>Department of Health Management and Health Economics, Institute of Health and Society, University of Oslo, Oslo, Norway

<sup>2</sup>Centre for Health and Social Economics, National Institute for Health and Welfare, Helsinki, Finland

### Corresponding author:

Tron Anders Moger, Department of Health Management and Health Economics, Institute of Health and Society, University of Oslo, P.O. Box 1089 Blindern, 0317 Oslo, Norway.  
Email: tronmo@medisin.uio.no

outpatient hospital services and prescribed medicines, as well as data on mortality, are available for researchers.

One of the challenges when comparing health-care performance measures between countries is to adjust for differences in the patient mix. This is further complicated by the fact that detailed information on the patients may not be available, or variables being very differently defined across countries. EuroHOPE aims to solve this problem by using individual-level register data available for everyone with a specified health problem, which contains detailed information on variables with effect on the health performance measures such as disease-specific comorbidities, number of days in hospital, and medication use prior to the occurrence of the health problem studied.

This study gives a description of the data and methods used to compare health-care performance measures within EuroHOPE. We briefly describe the contents of the data. Then, we discuss the methodological aspects of risk adjustment with regard to how stable the risk-adjusted estimates are depending on what data are used as the reference when calculating them. In multinational studies, it is likely that the effects of risk adjusters differ between countries, but it may be difficult to evaluate the importance of the difference that a single coefficient has on the risk-adjusted value just by looking at, for example, interactions between countries and single-risk adjusters as there are many variables in the models. Changes in other coefficients might offset the effect a single coefficient would have on the risk-adjusted value. Also, in EuroHOPE, some countries (the Netherlands and Scotland) have data sharing restrictions and the whole individual-level data cannot be pooled. We think this would be a problem in other studies using detailed individual-level registry data as well, as several countries are not eager to share their data abroad. This problem limits the choice of methods that can be used for analyzing the data. Most methods require all data to be pooled. Hence, the main aim of this article is to study whether one can get stable results from simple methods of risk adjustment even when the data are complex.

In this study, we used only acute myocardial infarction (AMI) patient data to illustrate the methodology. More comprehensive output will be presented in separate articles focusing on each condition included in EuroHOPE.

## Methods

### Data

A total of seven countries participated in the EuroHOPE project: Finland, Hungary, Italy (the city of Turin), the Netherlands, Norway, Scotland, and Sweden. In this article, we analyze data from all countries except Scotland and the Netherlands (see section “Statistical Methods” below). EuroHOPE applied an episode-based approach to analyzing the performance of countries and regions in the treatment of

certain health problems similarly as done earlier in Finland on national level.<sup>3</sup> The patient populations studied in the project are very low-birth-weight infants and individuals suffering from AMI, cerebral infarction, hip fracture, and breast cancer.

Each country prepared a data file for each disease, following a disease-specific protocol of inclusion and exclusion criteria.<sup>1</sup> Data were collected from various health registers containing the relevant information with the widest possible coverage on the use of health services of these patients. This included cause-of-death registers; hospital inpatient registers containing length of stay (LOS), comorbidity, and treatment information; and prescription registers containing information on medication use. Data from different registers were merged using unique patient identifiers of patients in each country. For AMI, cases were identified for the year 2007 in all countries, except Norway which used 2009 due to the unavailability of deterministically linkable hospital discharge, prescribed medication, and cause-of-death data before that year. For a more detailed description of the data and methods used in the EuroHOPE project, please see Häkkinen et al.<sup>1</sup>

Each patient had a follow-up of 1 year beginning on the date when the episode started (index admission), excluding patients with AMI admissions in the 365 days before the index admission. In addition, the patients' hospital discharge data and data on purchases of prescribed medicine were collected 1 year back as these were used in defining some of the risk adjustment variables.

Variables used in the risk adjustment include age at index admission, gender, disease-specific comorbidities, and the number of hospital days (for any cause) the year prior to the index admission. The comorbid diseases were specified for each disease group by clinical experts in the field. The comorbid diseases used in the study of AMI are presented in more detail in Häkkinen et al.,<sup>1</sup> including the International Classification of Diseases (ICD; both ICD-10 and ICD-9) codes and the Anatomical Therapeutic Chemical (ATC) classification system codes used for identifying the selected comorbid diseases from the hospital discharge registers and the data on medicine purchases, respectively.

The performance measures were specifically tailored for each disease in EuroHOPE. Main measures used in AMI were all-cause mortality after 30 days, 90 days, and 1 year; LOS for the first hospital episode; all-cause LOS during the first year after the diagnosis; and disease-specific LOS during the first year after the diagnosis. The first hospital episode is defined as continuous hospital inpatient care (overnight stay at home between the hospital stays is allowed), truncated at 365 days if the LOS was longer. For the disease-specific LOS, only days spent in hospital with the particular disease as the main diagnosis are considered. A list of the performance measures and risk adjustment variables used in the study of AMI, with some descriptives, is given in Tables 1 and 2.

**Table 1.** Descriptive statistics on background variables, length of stay, and 90-day mortality for AMI.

Variable	Data											
	Finland		Hungary		Italy		Norway		Sweden		Pooled	
	N (Avg.)	% (SD)	N (Avg.)	% (SD)	N (Avg.)	% (SD)	N (Avg.)	% (SD)	N (Avg.)	% (SD)	N (Avg.)	% (SD)
N	9102	100.0%	14,235	100.0%	1563	100.0%	10,612	100.0%	23,768	100.0%	59,135	100.0%
Male	5404	59.4%	7907	55.5%	1002	64.1%	6562	61.8%	14,251	60.0%	35,126	59.4%
Age (years)	72.6	12.7	66.9	13.1	70.2	13.2	71.8	14.1	73.4	12.8	71.4	13.4
LOS previous year	5.1	10.9	4.8	10.5	1.0	5.0	5.6	13.0	4.6	12.7	4.8	11.8
Age, classified (years)												
18–49	449	4.9%	1429	10.0%	122	7.8%	772	7.3%	1011	4.3%	3783	6.4%
50–54	473	5.2%	1378	9.7%	85	5.4%	649	6.1%	1002	4.2%	3587	6.1%
55–59	716	7.9%	1437	10.1%	157	10.0%	856	8.1%	1696	7.1%	4862	8.2%
60–64	808	8.9%	1601	11.2%	154	9.9%	1134	10.7%	2442	10.3%	6139	10.4%
65–69	883	9.7%	1729	12.1%	171	10.9%	1033	9.7%	2386	10.0%	6202	10.5%
70–74	1161	12.8%	1853	13.0%	208	13.3%	1066	10.0%	2729	11.5%	7017	11.9%
75–79	1429	15.7%	1985	13.9%	234	15.0%	1271	12.0%	3279	13.8%	8198	13.9%
80–84	1526	16.8%	1618	11.4%	218	13.9%	1456	13.7%	3936	16.6%	8755	14.8%
85–89	1098	12.1%	794	5.6%	136	8.7%	1476	13.9%	3416	14.4%	6920	11.7%
90 or older	559	6.1%	266	1.9%	78	5.0%	899	8.5%	1871	7.9%	3672	6.2%
Measures to be risk-adjusted												
90-day mortality	1753	19.3%	3048	21.4%	193	12.3%	1230	11.6%	3541	14.9%	9765	16.5%
LOS of first hospital episode	12.1	14.0	11.9	9.8	11.3	9.5	7.9	5.6	8.5	7.0	9.8	9.2

SD: standard deviation; AMI: acute myocardial infarction; LOS: length of stay.

## Statistical methods

Descriptive statistics of the outcomes and risk adjusters are compared, using measures such as proportions, means, and medians.

The first step of the risk adjustment was to construct a merged database from the countries which were allowed by their national data protection authorities to share data across borders. The countries included Finland, Norway, Sweden, Italy, and Hungary, and pooled data of these are called the reference database. As seen from Table 1, the reference database for AMI includes 59,135 patients in total.

For each response, three different risk-adjusted outputs were produced: adjusted for sex and age only (M1); adjusted for sex, age, LOS previous year, disease-specific comorbidities based on primary and secondary diagnoses the year prior to diagnosis (M2); and M3 identical to M2 except comorbidities were based on both primary and secondary diagnoses and medication purchases the year prior to diagnosis. The reason for using both models M2 and M3 is to compare the effect of a narrow and broad definition of comorbidities. Only comorbidities with prevalence of >1% in all countries based on the definition given for M3 were included as risk adjusters. As seen from Table 2, in the case of AMI, this excludes atherosclerosis, dementia, renal insufficiency, and alcoholism.

Based on the experiences in the PERFORMANCE, Effectiveness and Cost of Treatment episodes (PERFECT) project,<sup>3</sup> the observed/expected approach<sup>4</sup> was used, which roughly corresponds to indirect standardization. Logistic

regression was used for the mortality outcomes, whereas negative binomial regression was used for the LOS measures.

The regression coefficients used to produce the risk-adjusted estimates in each country were based on the reference database. In order to avoid that the relatively large samples from Sweden and Hungary gave a much greater contribution to the estimates compared with the smaller sample from Italy (representative only for the city of Turin), weighted regression was used to ensure equal weight to all five countries as the effect of the risk adjusters might differ between countries. By comparing both the risk-adjusted estimates and the descriptive statistics on background variables and comorbidities in each country, one may get an indication as to why some countries perform better than others. It is also possible to present comparisons of regions between the countries, and an example of such output is given for AMI in Norway.

A sensitivity analysis is presented to study the extent risk-adjusted mortality rates differ depending on whether the two countries with the highest unadjusted mortality (Hungary and Finland) or the three countries with the lowest mortality (Norway, Sweden, and Italy) are used as the reference database. This allows us to study whether results depend on the choice of reference data and, more importantly, whether interaction effects between country and risk adjusters seem to matter in practice. Normally, one would pool data from all countries to use as the reference, but as this is not possible, it is interesting to study the impact different choices of

**Table 2.** Descriptive statistics on comorbidities used in risk adjustment of performance measures for AMI.

Variable	Data											
	Finland		Hungary		Italy		Norway		Sweden		Pooled	
	N	%	N	%	N	%	N	%	N	%	N	%
Comorbidities based on diagnoses and medication during 365 days prior to AMI												
Hypertension	6467	71.1%	12,424	87.3%	1346	86.1%	6676	62.9%	16,293	68.6%	42, 218	71.4%
Coronary artery disease	1146	12.6%	1971	13.8%	119	7.6%	1479	13.9%	2711	11.4%	7426	12.6%
Atrial fibrillation	448	4.9%	498	3.5%	27	1.7%	701	6.6%	1377	5.8%	3051	5.2%
Cardiac insufficiency	538	5.9%	894	6.3%	34	2.2%	643	6.1%	1623	6.8%	3732	6.3%
Diabetes mellitus	2018	22.2%	3618	25.4%	347	22.2%	1636	15.4%	4530	19.1%	11935	20.2%
Atherosclerosis <sup>a</sup>	189	2.1%	1091	7.7%	12	0.8%	193	1.8%	240	1.0%	1725	2.9%
Cancer	255	2.8%	344	2.4%	29	1.9%	235	2.2%	730	3.1%	1588	2.7%
COPD and asthma	1379	15.2%	2507	17.6%	23	1.5%	1762	16.6%	3269	13.8%	8940	15.1%
Dementia <sup>a</sup>	453	5.0%	185	1.3%	9	0.6%	421	4.0%	579	2.4%	1643	2.8%
Depression	1073	11.8%	1053	7.4%	166	10.6%	1330	12.5%	3709	15.6%	7223	12.2%
Parkinson's disease	159	1.7%	290	2.0%	20	1.3%	122	1.1%	484	2.0%	1069	1.8%
Mental disorders	319	3.5%	480	3.4%	26	1.7%	346	3.3%	646	2.7%	1804	3.1%
Renal insufficiency <sup>a</sup>	60	0.7%	345	2.4%	22	1.4%	324	3.1%	413	1.7%	1164	2.0%
Alcoholism <sup>a</sup>	66	0.7%	89	0.6%	1	0.1%	80	0.8%	183	0.8%	419	0.7%
Stroke	232	2.5%	565	4.0%	26	1.7%	375	3.5%	643	2.7%	1841	3.1%
Comorbidities based on diagnoses during 365 days prior to AMI												
Hypertension	570	6.3%	2762	19.4%	89	5.7%	1143	10.8%	2317	9.7%	6881	11.6%
Coronary artery disease	1146	12.6%	1971	13.8%	119	7.6%	1479	13.9%	2711	11.4%	7426	12.6%
Atrial fibrillation	448	4.9%	498	3.5%	27	1.7%	701	6.6%	1377	5.8%	3051	5.2%
Cardiac insufficiency	538	5.9%	894	6.3%	34	2.2%	643	6.1%	1623	6.8%	3732	6.3%
Diabetes mellitus	458	5.0%	1316	9.2%	42	2.7%	619	5.8%	1510	6.4%	3945	6.7%
Atherosclerosis <sup>a</sup>	189	2.1%	1091	7.7%	12	0.8%	193	1.8%	240	1.0%	1725	2.9%
Cancer	234	2.6%	306	2.1%	23	1.5%	217	2.0%	624	2.6%	1404	2.4%
COPD and asthma	260	2.9%	314	2.2%	23	1.5%	539	5.1%	806	3.4%	1942	3.3%
Dementia <sup>a</sup>	203	2.2%	168	1.2%	4	0.3%	372	3.5%	281	1.2%	1028	1.7%
Depression	35	0.4%	162	1.1%	1	0.1%	41	0.4%	135	0.6%	374	0.6%
Parkinson's disease	30	0.3%	64	0.4%	1	0.1%	31	0.3%	62	0.3%	188	0.3%
Mental disorders	24	0.3%	24	0.2%	3	0.2%	12	0.1%	44	0.2%	106	0.2%
Renal insufficiency <sup>a</sup>	60	0.7%	345	2.4%	22	1.4%	324	3.1%	413	1.7%	1164	2.0%
Alcoholism <sup>a</sup>	66	0.7%	89	0.6%	1	0.1%	80	0.8%	183	0.8%	419	0.7%
Stroke	232	2.5%	565	4.0%	26	1.7%	375	3.5%	643	2.7%	1841	3.1%

AMI: acute myocardial infarction; COPD: chronic obstructive pulmonary disease.

<sup>a</sup>Comorbid disease not included in the restricted risk adjustment models (M2 and M3).

reference data have on the risk-adjusted measures. In order to illustrate this, we need access to all data used in the analysis, and Scotland and the Netherlands are hence omitted. The reference data in EuroHOPE will therefore only consist of those countries' data that can be pooled. In case it is impossible to construct a merged database from all participating countries, the approach proposed here is a practical option to study the problem. The data were analyzed using Stata.<sup>5</sup>

## Results

### Examples of risk-adjusted results for AMI

The pooled coefficients from a logistic regression analysis with AMI 90-day mortality as the response and using the

reference database are given in Table 3, left column. The age group 90+ is used as the reference category for age; for all other variables, the coefficients give the effect of scoring on each variable. Area under the curve (AUC) values are above 0.7 for all models M1–M3, with M3 showing the best performance. Most risk adjusters are significant in all models and for all three choices of reference data, but the coefficients can be quite different, as expected.

As seen from Table 1, the unadjusted 90-day mortality proportion varies from 10% in Norway to 21% in Hungary. Figure 1(a) shows the effect of risk adjustment on these proportions, using the pooled regression coefficients from the reference database. The risk adjustment changes the mortality proportions to a limited degree compared with the unadjusted ones, with the exception of

**Table 3.** Coefficients with standard errors for 90-day mortality after AMI based on the reference database and two alternative databases.

Age (years)	Pooled			Finland and Hungary			Italy, Norway, and Sweden		
	M1	M2	M3	M1	M2	M3	M1	M2	M3
	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference
18–49	-3.008***	-2.883***	-2.728***	-3.047***	-2.968***	-2.963***	-3.389***	-3.230***	-2.909***
	0.098	0.099	0.099	0.132	0.133	0.133	0.162	0.163	0.164
50–54	-2.798***	-2.683***	-2.588***	-2.804***	-2.735***	-2.728***	-3.424***	-3.275***	-3.018***
	0.095	0.095	0.096	0.123	0.124	0.124	0.182	0.182	0.183
55–59	-2.527***	-2.432***	-2.344***	-2.426***	-2.377***	-2.379***	-3.106***	-2.970***	-2.737***
	0.075	0.075	0.076	0.104	0.105	0.105	0.128	0.128	0.129
60–64	-2.176***	-2.117***	-2.079***	-2.170***	-2.141***	-2.145***	-2.497***	-2.408***	-2.287***
	0.063	0.064	0.064	0.095	0.096	0.097	0.093	0.094	0.095
65–69	-1.856***	-1.826***	-1.807***	-1.893***	-1.903***	-1.912***	-2.138***	-2.079***	-1.954***
	0.057	0.058	0.059	0.089	0.09	0.09	0.082	0.083	0.085
70–74	-1.523***	-1.517***	-1.520***	-1.569***	-1.581***	-1.596***	-1.791***	-1.772***	-1.685***
	0.051	0.052	0.052	0.082	0.083	0.084	0.071	0.072	0.074
75–79	-1.083***	-1.101***	-1.126***	-1.084***	-1.122***	-1.135***	-1.367***	-1.372***	-1.337***
	0.046	0.046	0.047	0.077	0.078	0.078	0.062	0.063	0.064
80–84	-0.616***	-0.645***	-0.680***	-0.676***	-0.703***	-0.718***	-0.748***	-0.778***	-0.770***
	0.043	0.044	0.044	0.075	0.076	0.077	0.055	0.056	0.057
85–89	-0.340***	-0.371***	-0.406***	-0.417***	-0.437***	-0.442***	-0.366***	-0.400***	-0.427***
	0.044	0.045	0.045	0.08	0.08	0.081	0.054	0.055	0.056
90 or older	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference
Male	0.027	0.024	0.077**	0.069	0.089*	0.086*	0.102**	0.071*	0.138***
	0.025	0.025	0.025	0.036	0.037	0.037	0.035	0.035	0.036
Hypertension		0.011	0.565***		-0.016	-0.147**		-0.085	0.724***
		0.04	0.031		0.058	0.048		0.058	0.043
Coronary artery disease		0.044	-0.02		-0.026	-0.011		0.059	-0.053
		0.038	0.037		0.056	0.053		0.054	0.052
Atrial fibrillation		0.135**	0.124*		0.138	0.15		0.223***	0.187**
		0.05	0.05		0.078	0.078		0.067	0.066
Cardiac insufficiency		0.511***	0.429***		0.451***	0.441***		0.504***	0.418***
		0.047	0.046		0.068	0.068		0.066	0.065
Diabetes mellitus		0.192***	0.319***		0.229***	0.189***		0.088	0.382***
		0.048	0.029		0.068	0.04		0.071	0.043
Cancer		0.720***	0.696***		0.279**	0.298**		1.090***	1.011***
		0.065	0.062		0.1	0.096		0.088	0.084
COPD and asthma		0.183**	0.150***		0.073	0.209***		0.387***	-0.034
		0.06	0.033		0.099	0.045		0.077	0.052

(Continued)

Table 3. (Continued)

	Pooled			Finland and Hungary			Italy, Norway, and Sweden		
	M1	M2	M3	M1	M2	M3	M1	M2	M3
Depression		0.142	0.198***		0.046	0.032		0.176	0.410***
Parkinson's disease		0.146	0.036		0.187	0.057		0.233	0.047
Mental disorders		0.439**	0.352***		0.094	0.215*		0.742**	0.414***
		0.17	0.077		0.227	0.108		0.26	0.111
		0.385	0.761***		0.643	0.716***		-0.238	0.701***
		0.268	0.06		0.331	0.082		0.572	0.09
		0.192**	0.137*		0.048	0.037		0.319***	0.215**
		0.061	0.06		0.09	0.089		0.084	0.083
		0.012***	0.010***		0.018***	0.018***		0.007***	0.005***
LOS in previous year		0.001	0.001		0.002	0.002		0.001	0.001
Constant	-0.461***	-0.640***	-1.176***	-0.022	-0.205**	-0.184*	-0.666***	-0.827***	-1.490***
N	59,135	59,135	59,135	23,192	23,192	23,192	35,943	35,943	35,943
Pseudo R <sup>2</sup>	0.1	0.118	0.135	0.099	0.115	0.12	0.121	0.141	0.166
AIC	46,251.761	45,343.253	44,470.098	21,223.15	20,862.118	20,746.931	23,804.601	23,285.356	22,612.317
BIC	46,350.624	45,549.967	44,676.812	21,311.717	21,047.304	20,932.117	23,897.987	23,480.619	22,807.579
AUC	0.725	0.746	0.762	0.72	0.736	0.742	0.752	0.772	0.793

M1: sex/age adjusted; M2: sex/age/comorbidity without medication adjusted; M3: sex/age/comorbidity with medication adjusted; AMI: acute myocardial infarction; COPD: chronic obstructive pulmonary disease; LOS: length of stay; AIC: Akaike information criterion; BIC: Bayesian information criterion; AUC: area under the curve.

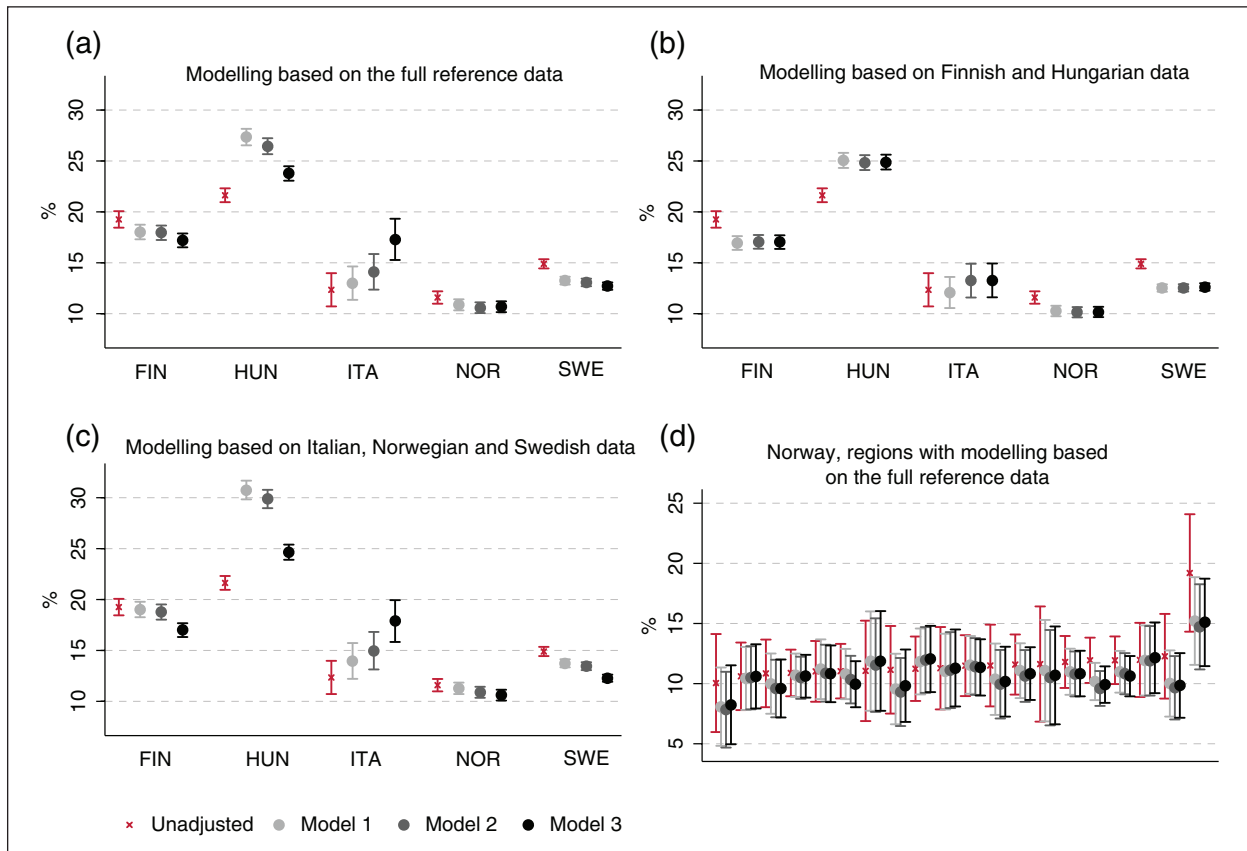
Pseudo R<sup>2</sup>, AIC, BIC, and AUC values are included at the end for comparison of explanatory power between the models.

\*Significant at 5% level.

\*\*Significant at 1% level.

\*\*\*Significant at 0.1% level.





**Figure 1.** Unadjusted and risk-adjusted 90-day mortality after AMI in five countries: (a) 90-day mortality proportions with 95% confidence intervals for each country, full data used as reference in adjustment; (b) 90-day mortality proportions for each country, data of Finland and Hungary used as reference in adjustment; (c) 90-day mortality proportions for each country, data of Norway, Sweden, and Italy used as reference in adjustment; and (d) regional 90-day mortality proportions in Norway. M1: sex/age adjusted; M2: sex/age/comorbidity without medication adjusted; M3: sex/age/comorbidity with medication adjusted; AMI: acute myocardial infarction.

Hungary. The effect here is that Hungary has younger patients, who from the regression output in Table 3 are expected to have lower mortality; thus, when adjusting for age and sex only, this causes the mortality proportion for Hungary to increase in M1. However, Hungary also has quite a lot more comorbidities than the other countries, most of which have an increasing effect on mortality, so adjusting for these in M2 and M3 causes the mortality proportion to become closer to the unadjusted value.

To assess the effect of heterogeneity in regression coefficients between countries, Figure 1(b) shows the 90-day unadjusted and adjusted mortality proportions when Finland and Hungary are used as the reference data, whereas Figure 1(c) shows the corresponding proportions when Norway, Italy, and Sweden are used as the reference data. There are some differences between the two graphs, but perhaps surprisingly few. We see that M1 gives a higher estimate for the mortality in Hungary if the low-mortality countries Norway, Sweden, and Italy are used as the reference data instead of the full reference. As seen from Table 3, the age effects for

M1 are more protective using the low-mortality country reference compared with the full reference. Hence, this influences Hungary with its young patient population. However, in M2 and M3, some of the comorbidities for which the prevalence is highest in Hungary get greater estimated effects using the low-mortality country reference compared with the full reference, moving the mortality estimates for Hungary to a lower level than the estimates from the full reference. A similar reasoning can be used for Italy and the observed lower mortality in model M3 when Hungary and Finland were used as the reference compared with the other reference data. Effects of comorbidities were smaller when using the Finland/Hungary reference data, so even though the prevalence of the comorbidities was higher in model M3 estimated with this reference data, the adjustment gives a smaller impact on the mortality estimates.

The same approach can also be used to illustrate regional differences in mortality within a country. An example for Norway is shown in Figure 1(d). From the point estimates, it is evident that there is heterogeneity in the 90-day mortality, although the confidence intervals are too wide to give any

significant differences in most cases. The international focus is a key element when looking at regional differences also; otherwise, there would be little point in basing the risk adjustment on pooled regression coefficients over national coefficients.

Another example is to study the LOS of first hospital episode as a measure of performance. As shown in Table 1, the unadjusted averages vary from 8 days in Norway to 12 days in Finland. From the regression output shown for the full reference data in Table 4, it is evident that fewer risk adjusters reach significance, indicating poorer explanatory power than for 90-day mortality. There is, for instance, no clear age trend in the results. The pseudo  $R^2$  values are low.

Looking at the graphs on average LOS of first hospital episode in Figure 2, one can see that when using the full reference database (Figure 2(a)), there is little difference between the unadjusted and risk-adjusted averages. Finally, a graph showing the regional variation in the LOS of first hospital episode in Norway is given in Figure 2(d).

## Discussion

There have been other recent examples of multinational comparisons of health-care quality outcomes with access to individual-level data.<sup>6,7</sup> However, one major complication in EuroHOPE, which could be a general problem in any multinational study, is that not all countries have permission to share data across borders due to confidentiality restrictions. As data cannot be pooled, this limits the number of methods that are possible to use, such as multilevel models, propensity score matching, and other methods.<sup>6,8,9</sup> When it comes to model choice, certain compromises must be made in order for a study like this to be feasible. A model which shows a good fit in one country may not be equally applicable in another. But in order to perform the analysis, a single choice of model has to be made. Also, when the number of responses to risk adjusted in a study is large, it becomes impractical to have different model choices for each response. Hence, not being able to pool all data poses several problems, as the methods used for finding the “best” model have to be simpler than the methods one would ideally wish for. Most covariates are categorized, even at the expense of less discriminating power. If a polynomial or spline was to be fitted for continuous covariates, it would have to be fitted to the data we are able to pool. Then one would have to impose exactly the same fit on the data not part of the pooling in order to get risk-adjusted estimates for those countries. This we thought would be a larger potential source for bias than using a simple categorization. For LOS, we also tried several alternative generalized linear models (GLM) including negative binomial, gamma, and inverse Gaussian models with log and identity link functions, but the negative binomial model showed the best fit.

The methods presented in this article are simple to use. In the risk adjustment, one implicitly assumes that the effects of

**Table 4.** Coefficients with standard errors for length of first hospital episode after AMI based on the reference database and two alternative databases.

Age (years)	Pooled			Finland and Hungary			Italy, Norway, and Sweden		
	M1	M2	M3	M1	M2	M3	M1	M2	M3
	18–49	-0.243*** 0.018	-0.242*** 0.018	-0.239*** 0.018	-0.409*** 0.032	-0.403*** 0.032	-0.396*** 0.032	-0.231*** 0.021	-0.240*** 0.021
50–54	-0.211*** 0.018	-0.212*** 0.018	-0.211*** 0.018	-0.376*** 0.032	-0.372*** 0.032	-0.369*** 0.032	-0.225*** 0.022	-0.236*** 0.022	-0.288*** 0.022
55–59	-0.139*** 0.017	-0.141*** 0.017	-0.142*** 0.017	-0.267*** 0.031	-0.267*** 0.031	-0.268*** 0.031	-0.152*** 0.019	-0.161*** 0.019	-0.210*** 0.019
60–64	-0.094*** 0.016	-0.098*** 0.016	-0.102*** 0.016	-0.186*** 0.03	-0.188*** 0.03	-0.193*** 0.03	-0.128*** 0.018	-0.137*** 0.018	-0.171*** 0.019
65–69	0.013 0.016	0.009 0.016	0.006 0.016	-0.158*** 0.03	-0.164*** 0.03	-0.171*** 0.03	0.043* 0.018	0.036* 0.018	0.002 0.018
70–74	0.077*** 0.015	0.071*** 0.015	0.067*** 0.015	-0.102*** 0.029	-0.111*** 0.029	-0.117*** 0.029	0.111*** 0.018	0.105*** 0.018	0.078*** 0.018
75–79	0.170*** 0.015	0.163*** 0.015	0.158*** 0.015	-0.004 0.028	-0.019 0.028	-0.025 0.028	0.205*** 0.017	0.202*** 0.017	0.182*** 0.017
80–84	0.160*** 0.015	0.156*** 0.015	0.153*** 0.015	0.001 0.029	-0.01 0.028	-0.014 0.028	0.197*** 0.017	0.197*** 0.017	0.185*** 0.017

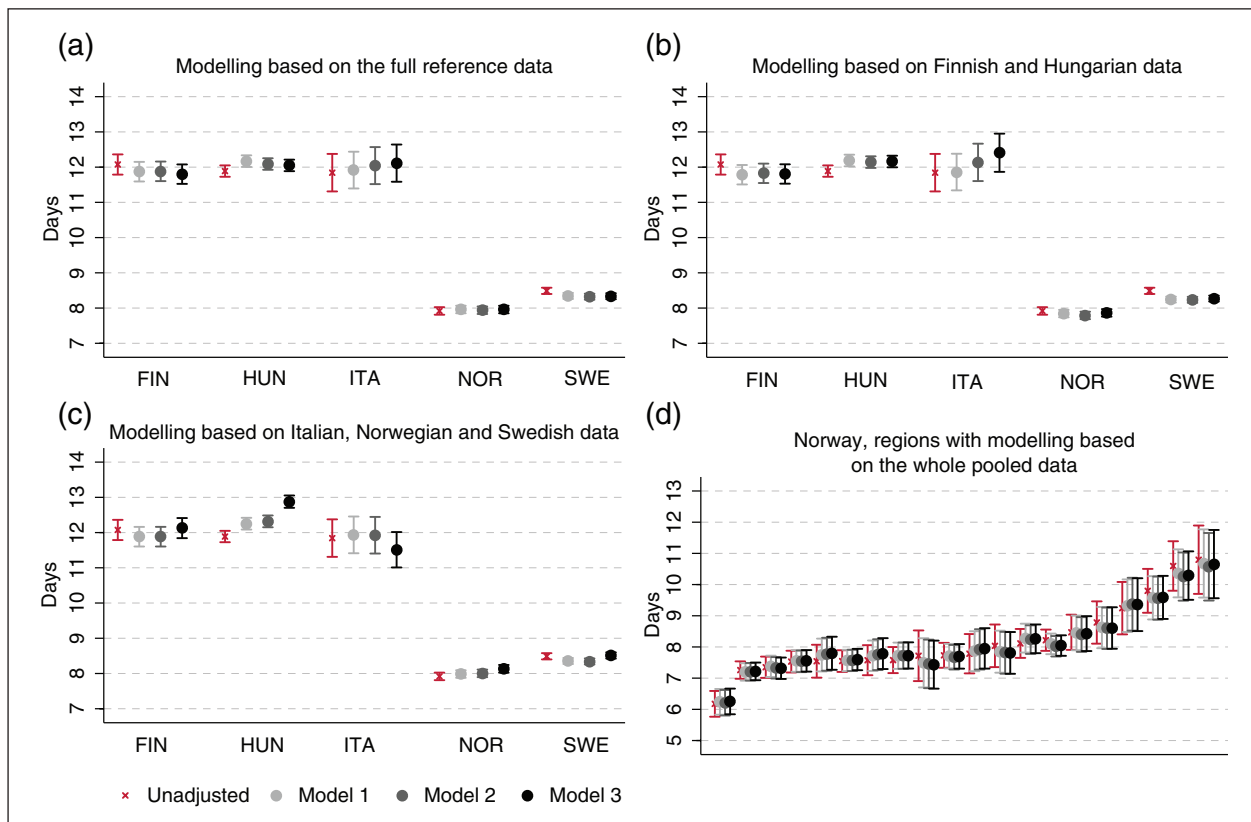


Table 4. (Continued)

	Pooled			Finland and Hungary			Italy, Norway, and Sweden		
	M1	M2	M3	M1	M2	M3	M1	M2	M3
85–89	0.115*** 0.015	0.112*** 0.015	0.111*** 0.015	0.004 0.03	-0.002 0.03	-0.006 0.03	0.150*** 0.017	0.151*** 0.017	0.151*** 0.017
90 or older	-0.084*** 0.006	-0.078*** 0.007	-0.079*** 0.007	-0.081*** 0.011	-0.077*** 0.011	-0.072*** 0.011	-0.051*** 0.008	-0.048*** 0.008	-0.054*** 0.008
Male		0.033*** 0.012	-0.001 0.007	0.031 0.018	0.031 0.018	0.008 0.013	-0.025 0.015	-0.128*** 0.008	-0.128*** 0.008
Hypertension		-0.119*** 0.011	-0.103*** 0.011	-0.118*** 0.018	-0.118*** 0.018	-0.101*** 0.017	-0.127*** 0.014	-0.099*** 0.013	-0.099*** 0.013
Coronary artery disease		-0.019 0.016	-0.013 0.016	0.023 0.027	0.023 0.027	0.023 0.027	0.002 0.019	0.015 0.019	0.015 0.019
Atrial fibrillation		0.036* 0.015	0.042** 0.015	0.038 0.024	0.038 0.024	0.042* 0.024	0.006 0.019	0.042* 0.019	0.042* 0.019
Cardiac insufficiency		0.092*** 0.014	0.105*** 0.008	0.075*** 0.022	0.075*** 0.022	0.097*** 0.012	0.092*** 0.018	0.083*** 0.011	0.083*** 0.011
Diabetes mellitus		-0.003 0.021	-0.004 0.02	-0.071* 0.034	-0.071* 0.034	-0.053 0.033	0.059* 0.026	0.050* 0.025	0.050* 0.025
Cancer		0.033 0.018	0.001 0.009	0.150*** 0.033	0.150*** 0.033	0.032* 0.014	0.009 0.021	-0.074*** 0.012	-0.074*** 0.012
COPD and asthma		0.068 0.042	0.036*** 0.01	0.099 0.058	0.099 0.058	0.042* 0.018	-0.155* 0.064	-0.044*** 0.012	-0.044*** 0.012
Depression		0.038 0.057	-0.042 0.024	0.086 0.08	0.086 0.08	-0.031 0.037	-0.163** 0.082	-0.100** 0.032	-0.100** 0.032
Parkinson's disease		0.300*** 0.071	0.047* 0.019	0.365*** 0.106	0.365*** 0.106	0.029 0.028	0.202* 0.095	0.026 0.025	0.026 0.025
Mental disorders		0.038* 0.019	0.052** 0.018	0.085** 0.029	0.085** 0.029	0.095** 0.029	-0.017 0.024	-0.012 0.024	-0.012 0.024
Stroke		0.003*** 0	0.003*** 0	0.004*** 0.001	0.004*** 0.001	0.005*** 0.001	0.001 0	0.001 0	0.001 0
LOS in previous year		2.366*** 0.013	2.348*** 0.014	2.651*** 0.025	2.651*** 0.025	2.597*** 0.028	2.221*** 0.014	2.231*** 0.014	2.319*** 0.015
Constant	59,135	59,135	59,135	23,192	23,192	23,192	35,943	35,943	35,943
N	381,932	381,686	381,579	160,657	160,472	160,451	218,798	218,691	218,378
Pseudo R <sup>2</sup>	382,040	381,901	381,795	160,754	160,665	160,644	218,900	218,894	218,581

M1: sex/age adjusted; M2: sex/age/comorbidity without medication adjusted; M3: sex/age/comorbidity with medication adjusted; AMI: acute myocardial infarction; COPD: chronic obstructive pulmonary disease; LOS: length of stay; AIC: Akaike information criterion; BIC: Bayesian information criterion.

Pseudo R<sup>2</sup>, AIC, and BIC values are included at the end for comparison of explanatory power between the models.  
 \*Significant at 5% level.  
 \*\*Significant at 1% level.  
 \*\*\*Significant at 0.1% level.



**Figure 2.** Unadjusted and risk-adjusted length of stay after AMI in five countries: (a) average length of first hospital episode with 95% confidence intervals for each country, full data used as reference in adjustment; (b) 90-day mortality proportions for each country, data of Finland and Hungary used as reference in adjustment; (c) 90-day mortality proportions for each country, data of Norway, Sweden, and Italy as reference in adjustment; and (d) regional 90-day mortality proportions in Norway. M1: sex/age adjusted; M2: sex/age/comorbidity without medication adjusted; M3: sex/age/comorbidity with medication adjusted; AMI: acute myocardial infarction.

the confounding factors on the response are similar in all countries, which is often not the case. We estimated weighted pooled regression coefficients to be used in the risk adjustment. The weighting would have been unnecessary if we believed that the effect of the risk adjusters would be exactly equal across all countries, as then it would not matter if some country contributed many more cases to the total data than others. Notwithstanding, the point estimates of the regression coefficients used in the risk adjustment would be the same. Equality of coefficients across countries can be checked by studying the interaction effects between the countries and the risk adjusters.<sup>10,11</sup> In large register-based studies like this, many interactions will be statistically significant, and ignoring this in the risk adjustment would lead to the constant risk fallacy<sup>11</sup> potentially causing the standard risk-adjusted estimates to be biased. Although not shown in the results, significant interactions were also the case in EuroHOPE for the data we were able to pool. But again, this is only possible to study thoroughly if one can pool data from all countries included in the study. Also, the impact on the outcome of large differences in the effect of single-risk adjusters between countries is difficult to ascertain, as there are many risk adjusters

working together in the models. In any case, the problem is hard to solve, as to our knowledge there are no ready-made solutions to the problem if there are significant interactions. Different effects of risk adjusters may be due to differences in treatment practices, coding practices, or under-/over-reporting of comorbid diseases. However, the magnitude of the problem can be studied by comparing the risk-adjusted responses using different choices of reference database, like demonstrated in Figures 1 and 2. These figures illustrate that the choice of reference data does not matter too much for AMI; hence, statistically significant interactions may not always be a problem in practice. Thus, even simple methods of risk adjustment may be useful if more advanced methods are difficult or impossible to use.

### Acknowledgements

The authors are grateful to two anonymous reviewers for their comments and constructive input. Any remaining errors and omissions are the authors' responsibility.

### Declaration of conflicting interests

The authors have declared that there is no conflict of interest.

## Funding

This project was undertaken within the European Union 7th Framework Programme European Health Care Outcomes, Performance and Efficiency (EuroHOPE), contract no 241721. Please go to <http://www.eurohope.info> for more details.

## References

1. Häkkinen U, Iversen T, Peltola M, et al. Health care performance comparison using a disease-based approach: the EuroHOPE project. *Health Policy* 2013; 112: 100–109.
2. Klazinga N, Fischer C and ten Asbroek A. Health services research related to performance indicators and benchmarking in Europe. *J Health Serv Res Policy* 2011; 16: 38–47.
3. Häkkinen U. The PERFECT project: measuring performance of health care episodes. *Ann Med* 2011; 43: S1–S3.
4. Ash AS, Schwartz M and Peköz EA. Comparing outcomes across providers. In: Iezzoni LI (ed.) *Risk adjustment for measuring health care outcomes*. 3rd ed. Chicago, IL: Health Administration Press, 2003, pp. 297–333.
5. StataCorp. *Stata statistical software: release 12*. College Station, TX: StataCorp LP.
6. Schreyögg J, Stargardt T and Tiermann O. Costs and quality of hospitals in different health care systems: a multi-level approach with propensity score matching. *Health Econ* 2011; 20: 85–100.
7. Willan AR and Kowgier ME. Cost-effectiveness analysis of a multinational RCT with a binary measure of effectiveness and an interacting covariate. *Health Econ* 2008; 17: 777–791.
8. Normand S-LT, Glickman ME and Gatsonis CA. Statistical methods for profiling providers of medical care: issues and applications. *J Am Stat Assoc* 1997; 92: 803–814.
9. Ash AS, Fienberg SE, Louis TA, et al. Statistical issues in assessing hospital performance. COPSS-CMS White Paper Committee, 2012. Available at: <http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Downloads/Statistical-Issues-in-Assessing-Hospital-Performance.pdf>
10. Mohammed AM, Deeks JJ, Girling A, et al. Evidence of methodological bias in hospital standardised mortality ratios: retrospective database study of English hospitals. *BMJ* 2009; 338: b780.
11. Nicholl J. Case-mix adjustment in non-randomised observational evaluations: the constant risk fallacy. *J Epidemiol Commun H* 2007; 61: 1010–1013.