BMJ Open

Predicting risk of hospitalisation: a retrospective population-based analysis in a paediatric population in Emilia-Romagna, Italy

Daniel Z Louis,[1] Clara A Callahan,[1] Mary Robeson,[1] Mengdan Liu,[1] Jacquelyn McRae,[2] Joseph S Gonnella,[1] Marco Lombardi,[3] Vittorio Maio[1,2]

Check for updates

[1]Center for Medical Research in Medical Education and Health Care, Sidney Kimmel Medical College at Thomas Jefferson University, Philadelphia, Pennsylvania, USA
[2]Jefferson College of Population Health, Thomas Jefferson University, Philadelphia, Pennsylvania, USA
[3]Risk Management and Clinical Governance, Parma Local Health Authority, Parma, Italy

Correspondence to
Dr Vittorio Maio;
vittorio.maio@jefferson.edu

ABSTRACT

Objectives Develop predictive models for a paediatric population that provide information for paediatricians and health authorities to identify children at risk of hospitalisation for conditions that may be impacted through improved patient care.

Design Retrospective healthcare utilisation analysis with multivariable logistic regression models.

Data Demographic information linked with utilisation of health services in the years 2006–2014 was used to predict risk of hospitalisation or death in 2015 using a longitudinal administrative database of 527 458 children aged 1–13 years residing in the Regione Emilia-Romagna (RER), Italy, in 2014.

Outcome measures Models designed to predict risk of hospitalisation or death in 2015 for problems that are potentially avoidable were developed and evaluated using the C-statistic, for calibration to assess performance across levels of predicted risk, and in terms of their sensitivity, specificity and positive predictive value.

Results Of the 527 458 children residing in RER in 2014, 6391 children (1.21%) were hospitalised for selected conditions or died in 2015. 49 486 children (9.4%) of the population were classified in the 'At Higher Risk' group using a threshold of predicted risk >2.5%. The observed risk of hospitalisation (5%) for the 'At Higher Risk' group was more than four times higher than the overall population. We observed a C-statistic of 0.78 indicating good model performance. The model was well calibrated across categories of predicted risk.

Conclusions It is feasible to develop a population-based model using a longitudinal administrative database that identifies the risk of hospitalisation for a paediatric population. The results of this model, along with profiles of children identified as high risk, are being provided to the paediatricians and other healthcare professionals providing care to this population to aid in planning for care management and interventions that may reduce their patients' likelihood of a preventable, high-cost hospitalisation.

INTRODUCTION

Healthcare systems have been moving from a passive approach of waiting for and reacting to patients' problems to a more active model

Strengths and limitations of this study

► This study included the entire paediatric population of the Emilia-Romagna Region of Italy, with a total of 527 458 children ages 1–13 years.
► The study used an existing longitudinal administrative healthcare database with both the advantage of much lower cost than new data collection and the disadvantage of gaps and potential errors in administrative data.
► The results of the study are being used to assist paediatricians and health authorities manage high-risk children.

that includes identification of patients at risk, taking the initiative in offering care and actively seeking to avoid recurrence or progression of medical problems. With the ageing of populations worldwide, and high prevalence of chronic diseases, it is not surprising that these efforts have often focused on the elderly. Less attention has been paid to the paediatric population. However, despite the relatively low prevalence of chronic disease in children, there is evidence that children experience preventable hospitalisations.[1] For example, a study of paediatric inpatient claims in the USA estimated that paediatric 'ambulatory care sensitive' conditions accounted for US$4.05 billion in hospital charges and over 1 million hospitalisation days in a 1-year period.[2]

Predictive risk modelling is a tool that can be used to estimate the risk of an outcome within the context of prespecified variables and uncertainty. Predictive risk modelling may offer an opportunity to better understand individuals who may be at higher risk for an undesirable outcome.[3] A number of predictive risk modelling studies have been conducted in paediatrics; however, many of these studies have focused on children

with specific medical problems or use data that are not routinely available in administrative databases.[4–9]

Under the auspices of the Italian National Health Service (NHS), the 21 regional governments are responsible for delivering healthcare through a network of geographically defined Local Health Authorities. Primary care physicians, including paediatricians, work for the Local Health Authorities as independent contractors. Every Italian is expected to enrol with a primary care physician (a paediatrician for those under age 14 years) who serve as the 'gatekeepers' for delivering primary care and coordinating specialty services for their enrolled patients.[10] This focus on primary care is ideal for the development and implementation of a proactive model of healthcare.

To further encourage coordinated care, the Regione Emilia-Romagna (RER) has established Patient-Centered Medical Homes. The identification of patients who would most benefit from outreach efforts is fundamental to achieving the goals of promoting population health and practising proactive medicine. The RER has therefore developed and implemented a population-based model to predict risk of hospitalisation or death for adult residents in the region.[11] The results of the model are presented to physicians in Patient-Centered Medical Homes as patient profiles to support care management and the identification of patients who may benefit from additional outreach such as home healthcare, disease management or case management.

Current risk models used in RER focus on the adult population. This paper describes the development of predictive risk models for the paediatric population using the RER's regional longitudinal administrative healthcare database to help identify children who are at risk of hospitalisation for conditions that may be affected through improved patient care.

METHODS

Data source

The RER is a region of northern Italy that lies between the River Po and Apennine Mountains with approximately 4.5 million inhabitants. RER maintains a longitudinal healthcare database for all its residents. The RER database contains patient-level demographic data (age, gender, birth and death dates, location of residence and primary care physician/paediatrician) and utilisation data for inpatient (hospital discharge abstract data with International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) diagnosis and procedure codes and admission/discharge dates), outpatient (laboratory, diagnoses and physician services, and pharmacy claims including WHO Anatomical Therapeutic Chemical (ATC)/Defined Daily Dose (DDD) system codes),[12] specialty (therapeutic procedures, rehabilitation and specialist visits) and emergency room (ER) visits. Inpatient medications are not captured. Patients with disabilities or low family income are eligible for exemption of service copayment for specialty visits and

outpatient prescriptions, which provide some socioeconomic information. Each resident is assigned an anonymous identifier so that utilisation can be tracked over time while maintaining patient privacy.[13]

Study cohort

In Italy, children aged 14 years are required to switch from a paediatrician to a primary care physician; therefore, we limited the study population to children 1–13 years old on 31 December 2014. The study population also was narrowed to exclude children who did not reside in RER for the entire year 2014. The study population was stratified into three age groups: 1–2 years old (on 31 December 2014); 3–5 years old; and 6–13 years old. Children less than 1 year old on 31 December 2014 were not included in the study population due to insufficient data for prediction of outcomes.

Dependent variable

The outcome was defined as the occurrence of hospitalisation that could have potentially been prevented or delayed with appropriate patient care or death by any cause.[11] We developed a list of hospitalisations that are potentially preventable with appropriate patient care using a three-step process. First, we conducted a literature search to evaluate paediatric studies that defined potentially avoidable disease in paediatrics that could require hospitalisation.[14–16] We began with the listing of ICD-9-CM codes for 'pediatric ambulatory care sensitive conditions' identified in Shi and Lu.[15] All hospitalisations in 2013 of children in the target age groups were classified using both ICD-9-CM codes and Disease Staging categories.[17 18] The results were reviewed by the authors of this paper and compared with Shi *et al*'s list. A number of changes were made for this project. For example, the list of immunisation preventable conditions to be included in the dependent variable was expanded to include currently available vaccines. We included additional conditions, such as acute cystitis (ICD-9-CM code of 595.0) and hypoglycaemic coma (ICD-9-CM code of 251.0). Advanced stages of selected medical problems were added where stage one may not be avoidable but advanced stages can potentially be delayed or prevented through timely intervention, for example, stage 2 or 3 appendicitis, stage 2 or 3 sinusitis. While certainly not always preventable, we believed that inclusion of hospitalisations for certain types of trauma and toxicities (eg, acetaminophen toxicity, adverse drug reactions and burns) was appropriate especially for a paediatric population. These changes are summarised in online supplementary appendix 1.

Finally, we used disease staging categories for inclusion of relevant hospitalisations that would have been missed using solely primary ICD-9-CM codes. For example, if a child was hospitalised with a primary diagnosis of respiratory failure with asthma (ICD-9-CM code of 493) as the secondary diagnosis, then the disease staging category of asthma would include that admission that might have been missed by including only primary ICD-9-CM

codes. This is summarised in online supplementary appendix 2.

Children hospitalised for these selected conditions or who died from any cause in 2015 were counted as being positive for the outcome.

Independent variables

A list of predictor variables was developed using the RER administrative data from 2006 to 2014. Independent variables included information such as: demographics, socioeconomic factors, diseases/conditions grouped by aetiology or body systems, mother's medical history and pregnancy/birthing information, ER visits, potentially inappropriate prescriptions and antibiotic usage.

Demographic variables included age on 31 December 2014, gender and citizenship (Italian or non-Italian). Children from low-income families or with disabilities are exempt from copayments for prescriptions and specialty visits. This information was used as a potential predictor variable.

We mapped diseases defined primarily by the affected body system with the exceptions of cancer, genetic conditions and trauma, which were based on aetiology[11] using 2014 hospital discharge data, outpatient prescription information and specialty visit claims. A total of 24 groups were defined. Disease staging diagnostic categories was used to map hospital admissions to the 24 body system/aetiology groups[17] (see first column of table 1). Patients with cardiovascular diseases, chronic respiratory diseases, diabetes mellitus, epilepsy and disorders of the thyroid were identified using the ATC Classification System codes from outpatient prescriptions.[19] Specialty visit records were also used for identifying medical conditions of some body systems. For example, if a child was admitted to the hospital for type 1 diabetes mellitus, or visited an endocrinologist, or had filled a prescription for insulin injection(s) (ATC code of A10AB), this patient would be identified as having an endocrine diagnosis in 2014.

Severity level codes (critical (C), acute (A), urgent but deferred (U) and not urgent (N)) are assigned to individuals on discharge from the emergency department. We excluded ER visits that resulted in a hospital admission because diagnosis information was captured by hospital discharge data with more accurate information. We believe more frequent or severe ER visits may indicate a poor outcome; therefore, number of ER visits by severity level was calculated for each patient.

There is evidence that the risks outweigh the benefits for certain medication usage in the paediatric population.[20] For example, certain mood-altering medications such as, citalopram, sertraline, fluvoxamine and any tricyclic antidepressants are not recommended in children of any age. Some medications can be harmful within specified ages. For example, loperamide is not indicated for children under 3 years old. For children who filled an outpatient prescription in 2014, we calculated their age at dispensation date and amount of medications they had filled, in order to identify patients with potentially

inappropriate prescriptions in 2014. The number of antibiotic prescriptions used in 2014 was estimated since high utilisation of antibiotics has been linked to decreased gut microflora, decreased immune function and resistant strains of bacteria.[21]

For children ages 1–5 years, the models considered problems identified at birth as potential predictors using hospital discharge abstract data. About 86% of the newborns were healthy, with no serious medical problems noted on their birth records. Infants with diagnostic categories of premature birth with low birth weight, full-term infants with abnormal birth weight, premature with very low birth weight or extremely low birth weight, were classified as abnormal birth weight; all other conditions were considered as a group. The mothers' delivery information, such as age at delivery, C-section and parity, were identified based on the mothers' hospitalisation records and linked to children. Information about deliveries that occurred outside hospitals could not be captured.

Children ages 1–5 years old were also linked with information regarding their mothers' medical history and drug use during pregnancy. There is evidence on the association between prenatal (up to 270 days before delivery) exposure to antibiotics and the development of asthma.[22] We estimated the total exposure to any antibiotics during the prenatal period using the mother outpatient prescription claims. We included two categories of mother's potentially inappropriate drug use, class D (potential risks outweigh the benefits) and X (contraindicated during pregnancy), since these drugs may be linked to harm to children. Mothers' 3-year medical history before delivery was retrieved for identifying certain conditions such as abortion, diabetes and psychological condition. For about 22% of children, we were not able to establish the mother–baby linkage.

We developed history variables with up to 5 years of data (pharmacy, specialty, hospital admission and ER visit) for children in age strata 3–5 years old and 6–13 years old. Children who had conditions in any year from 2009 to 2013 were flagged as having a utilisation history.

Modelling

Logistic regression was used to estimate predicted probabilities for the occurrence of an inpatient hospital stay for the selected conditions, or death from any cause, for the individual patients. Since age and gender may be strongly correlated with children's risk, we fit a total of six multivariable logistic regression models: female and male by age groups (1–2, 3–5 and 6–13 years old). All models were developed using SAS V.9.3 statistical software.

Model validation

The predicted accuracy of the modelling was evaluated using C-statistics (the area under the receiver operating characteristics curve), comparing the results of the 'predicted' to the 'observed' outcomes in 2015. We stratified patients into risk strata based on the predicted risk of hospitalisation or death. 'At higher risk' was defined as children with

Table 1 Study population 2014

	Total population 527 458		At higher risk Risk >2.5% 49 486		Higher than average Risk >1.2%–2.5% 99 714		Lower than average Risk ≤1.2% 378 258	
	Number	%	Number	%	Number	%	Number	%
Gender								
Female	255 875	48.5	20 315	41.1	43 030	43.2	192 530	50.9
Male	271 583	51.5	29 171	58.9	56 684	56.8	185 728	49.1
Age group (years)								
1–2	78 051	14.8	18 112	36.6	44 084	44.2	15 855	4.2
3–5	125 459	23.8	20 180	40.8	35 543	35.6	69 736	18.4
6–13	323 948	61.4	11 194	22.6	20 087	20.1	292 667	77.4
Selected condition/body system								
Cancer	1138	0.2	477	1.0	252	0.3	409	0.1
Cardiovascular	1624	0.3	653	1.3	211	0.2	760	0.2
Dental conditions	442	0.1	138	0.3	109	0.1	195	0.1
Endocrine	6458	1.2	1276	2.6	1074	1.1	4108	1.1
Ear, nose, throat	31 919	6.1	21 664	43.8	7376	7.4	2879	0.8
Eye	821	0.2	165	0.3	145	0.1	511	0.1
Genetic conditions	274	0.1	188	0.4	29	0.0	57	0.0
Gastrointestinal	7380	1.4	2724	5.5	1578	1.6	3078	0.8
Genitourinary	3389	0.6	987	2.0	836	0.8	1566	0.4
Obstetric and gynaecologic conditions	128	0.0	17	0.0	19	0.0	92	0.0
Haematological	1114	0.2	596	1.2	247	0.2	271	0.1
Hepatobiliary	245	0.0	82	0.2	39	0.0	124	0.0
Immunological disease	199	0.0	80	0.2	45	0.0	74	0.0
Infectious disease	869	0.2	596	1.2	160	0.2	113	0.0
Male genital	1329	0.3	179	0.4	209	0.2	941	0.2
Musculoskeletal	3817	0.7	664	1.3	453	0.5	2700	0.7
Neurological diseases	3738	0.7	2123	4.3	912	0.9	703	0.2
Nutrition	924	0.2	446	0.9	201	0.2	277	0.1
Other conditions	1703	0.3	1150	2.3	247	0.2	306	0.1
Neonatal conditions	186	0.0	111	0.2	50	0.1	25	0.0
Psychological	854	0.2	388	0.8	141	0.1	325	0.1
Respiratory	20 450	3.9	7285	14.7	5886	5.9	7279	1.9
Skin	39 344	7.5	5809	11.7	7461	7.5	26 074	6.9
Trauma	737	0.1	177	0.4	167	0.2	393	0.1
Emergency room visits based on severity level								
Critical	182	0.0	117	0.2	35	0.0	30	0.0
Acute	15 029	2.8	5219	10.5	3915	3.9	5895	1.6
Urgent but could be deferred	118 372	22.4	26 945	54.5	33 241	33.3	58 186	15.4
Not urgent	45 336	8.6	11 216	22.7	13 080	13.1	21 040	5.6
Inappropriate medication prescriptions[20]	8077	1.5	2376	4.8	3090	3.1	2611	0.7
Antibiotic use								
1	114 421	21.7	8248	16.7	24 544	24.6	81 629	21.6

Continued

Table 1   Continued

| | Total population | | At higher risk | | Higher than average | | Lower than average | |
| | 527 458 | | Risk >2.5%
49 486 | | Risk >1.2%–2.5%
99 714 | | Risk ≤1.2%
378 258 | |
	Number	%	Number	%	Number	%	Number	%
2	63 151	12.0	9359	18.9	19 035	19.1	34 757	9.2
3+	76 878	14.6	25 587	51.7	29 144	29.2	22 147	5.9
Non-Italian citizen	90 760	17.2	8975	18.1	18 390	18.4	63 395	16.8
Copay exempted based on family income/employment status	244 911	46.4	37 502	75.8	64 776	65.0	142 633	37.7
Copay exempted based on disabled status	6173	1.2	2029	4.1	1321	1.3	2823	0.7
Specialty visits in paediatrics								
Medical	12 642	2.4	3987	8.1	2735	2.7	5920	1.6
Surgical	8982	1.7	2060	4.2	2294	2.3	4628	1.2

a predicted risk greater than 2.5%. 'Higher than average' was defined as children with a predicted risk of hospitalisation or death between the mean rate and 2.5%. The rest of population was grouped into 'Lower than average'. These risk strata were defined to yield a manageable number of patients to review for the typical paediatric panel of approximately 800 patients. Calibration of the model across these risk groups was assessed by comparing observed to predicted rates among the risk groups. We also report the sensitivity, specificity and positive predictive value (PPV) for the defined risk group cut-offs.

RESULTS
Characterisation of risk groups
A total of 568 117 children ages 1–13 years resided in RER in 2014. We excluded from our analysis 40 659 children (7.2%) who did not reside in RER for the entire year, resulting in a population of 527 458 children. Of those, 6391 children (1.21%) were hospitalised for selected conditions or died in 2015. Table 1 displays the distribution of gender, age category, presence of selected chronic conditions, ER visits, selected prescription drug usage, copay exemption for income or disability and specialty visits for the eligible RER residents as of 31 December 2014.

Table 1 also compares the characteristics of the total selected paediatric population to the subgroups of the population classified by risk categories based on the model results. Forty nine thousand four hundred and eighty-six children (9.4%) of the population were classified in the 'At higher risk' group using a threshold of predicted risk >2.5%. The children predicted to be 'At higher risk' were more likely to be male (58.9%) compared with 51.5% in the total population. The two youngest age strata (1–2 and 3–5 years) had much higher proportions of children identified in the 'At higher risk' group than the children aged 6–13 years. For example, 18 112 (23%)

of the children age 1–2 years were identified in the 'At higher risk' group. This age category includes 36% of the 'At higher risk' children, although it represents 15% of the total paediatric population. Children in the 'At higher risk' category were more likely to have each of the selected conditions. When looking at the highest prevalence conditions, 43.8% of children in the 'At higher risk' category had an ear, nose or throat problem, compared with 6.1% in the overall population, 5.5% had a gastrointestinal problem compared with 1.4% in the overall population, 4.3% had a neurological problem compared with 0.7% in the overall population, 14.7% had a respiratory problem compared with 3.9% in the overall population and 11.7% had a skin problem compared with 7.5% in the overall population.

Children identified as being 'At higher risk' were much more likely to have a history of ER visits and were more likely to have a history of 2, 3 or more antibiotic prescriptions. Overall, 14.6% of children had three or more antibiotic prescriptions, while in the 'At higher risk' category, 51.7% had a history of 3 or more antibiotic prescriptions. Children with exemptions from copayments due to either family income or disability were more likely to be identified as being 'At Higher Risk' as were children with a history of medical or surgical specialty visits.

Table 2 displays information about the delivery (for the children age 1–5 years) and medical history of the mother for those children where we were able to match to their mother's record. First children, children who were delivered by caesarean section and children where an abnormal birth weight or other problems were noted at birth were more likely to be classified in the 'At higher risk' category. If the mother was prescribed a potentially inappropriate drug or an antibiotic during pregnancy, the child was more likely to be classified in the 'At higher risk' category. When examining a 3-year medical history

Table 2 Birthing and medical history of mother*

| | Total population | | At higher risk Risk >2.5% | | Higher than average Risk >1.2%–2.5% | | Lower than average Risk ≤1.2% | |
| | 203 510 | | 38 292 | | 79 627 | | 85 591 | |
	Number	%	Number	%	Number	%	Number	%
Birthing								
Age at delivery (years)†								
24 and less	12 728	6.3	3275	8.6	5651	7.1	3802	4.4
25–34	88 370	43.4	18 227	47.6	35 681	44.8	34 462	40.3
35–39	45 575	22.4	8170	21.3	17 679	22.2	19 726	23.0
40 and over	12 529	6.2	2344	6.1	4528	5.7	5657	6.6
First delivery	99 190	48.7	23 336	60.9	42 662	53.6	33 192	38.8
C-section	48 282	23.7	11 370	29.7	19 480	24.5	17 432	20.4
Baby's birth condition								
Normal newborns	172 497	84.8	30 214	78.9	67 522	84.8	74 761	87.3
Abnormal birth weight	20 128	9.9	4757	12.4	7756	9.7	7615	8.9
Other abnormal birth condition	10 885	5.3	3321	8.7	4349	5.5	3215	3.8
Medical history								
Number of ordinary hospitalisation 1 year before delivery								
1	16 145	7.9	4578	12.0	6856	8.6	4711	5.5
2+	3920	1.9	1670	4.4	1500	1.9	750	0.9
Inappropriate prescription during pregnancy								
Class D	10 594	5.2	2970	7.8	3886	4.9	3738	4.4
Class X	4874	2.4	1086	2.8	1811	2.3	1977	2.3
Antibiotic use during pregnancy	60 679	29.8	14 422	37.7	25 757	32.3	20 500	24.0
3-year history before delivery								
Abortion	19 919	9.8	4970	13.0	8165	10.3	6784	7.9
Asthma	35 590	17.5	9026	23.6	14 894	18.7	11 670	13.6
Bacterial pneumonia	188	0.1	36	0.1	14	0.0	138	0.2
Cardiovascular disease	18 756	9.2	5068	13.2	7742	9.7	5946	6.9
Diabetes	2602	1.3	1003	2.6	1106	1.4	493	0.6
Hypertension	140	0.1	51	0.1	39	0.0	50	0.1
Infection	935	0.5	283	0.7	325	0.4	327	0.4
Psychological condition	9215	4.5	2701	7.1	3709	4.7	2805	3.3

*Information about the delivery was considered only for children 1–5 years old.
†For 22% of children, we were not able to establish the mother–baby linkage.

of the mother, the mother's asthma, cardiovascular disease, diabetes mellitus or mental health problems, or the record of a previous abortion, were all relatively frequent and more prevalent in the mothers of children predicted to be in the 'At higher risk' category.

Calibration
The population was divided into three risk groups based on predicted probability of hospitalisation as defined above. We observed good calibration; each stratum's predicted risks were similar to observed prevalence of hospitalisations

or deaths (figure 1). Individuals, who fell in the 'At higher risk' group, with predicted risk greater than 2.5%, had 2683 predicted events based on the model results, and 2737 observed events. While the overall rate of hospitalisation or death for children ages 1–13 years was 1.21%, the predicted and observed risk of the 'At higher risk' group was over 5%.

Model performance among risk groups
We observed a C-statistic of 0.78 indicating good model performance (table 3). The sensitivity (proportion
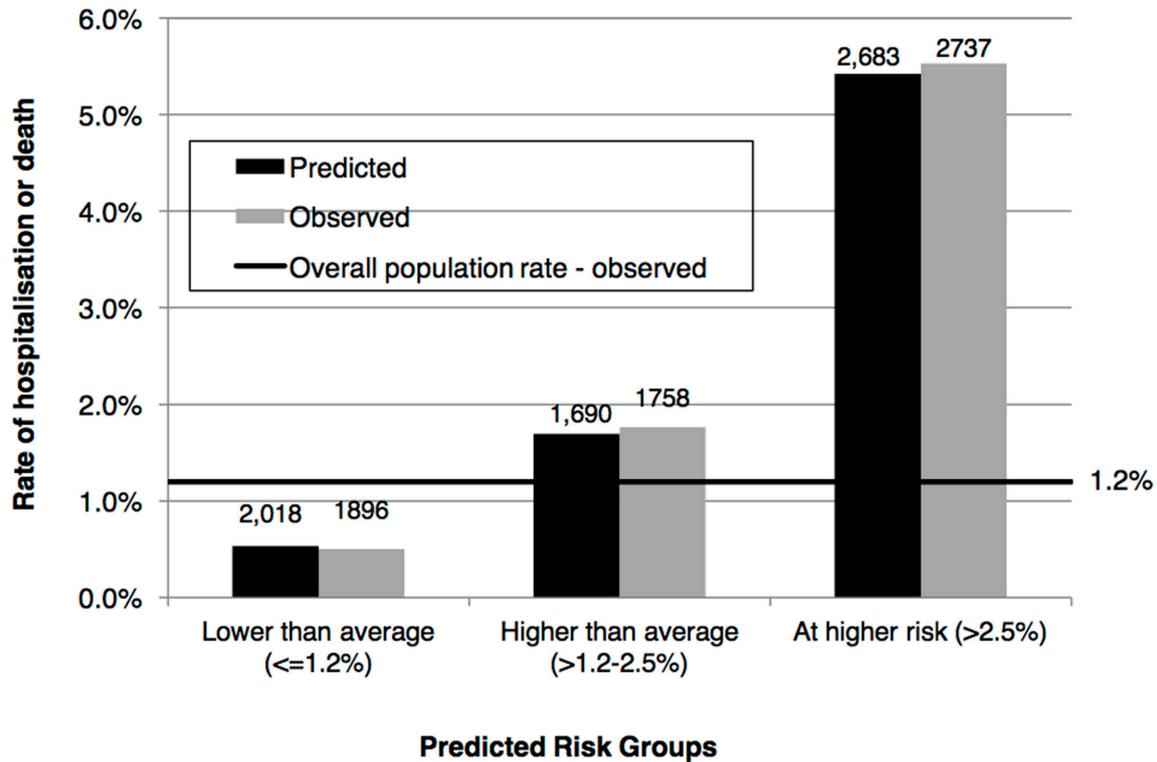
Figure 1 Model calibration: predicted and observed prevalence of hospitalisation or death in 2015 by risk category.

predicted to be at higher risk of those who had an event in 2015) was 0.43 and 0.70 for predicted risk categories of 'At higher risk' and 'Higher than average', respectively (table 4). In other words, among those whom were hospitalised or deceased in 2015, 43% were predicted to have risk greater than 2.5% of hospitalisation or death and 70% have risk higher than average. The specificity (proportion predicted to be at a 'lower' risk of those who did not have an event) was 0.91 and 0.72 for the predicted 'At higher' and 'Higher than average' risk categories, respectively; among those who were not hospitalised and did not die in 2015, 91% were not predicted to be 'At higher risk'. The PPV (proportion with an event of those who were predicted to be at an elevated risk) was 0.06 and 0.03 for the 'At higher' and 'Higher than average' predicted risk categories, respectively. In other words, of those individuals who were estimated to have a >2.5% risk of hospitalisation or death approximately 6% had an event in 2015. (Regression coefficients and significance levels of independent variables for multivariable logistic regression models for each the six age and gender strata are included in online supplementary appendix 3).

DISCUSSION
We have developed a population-based model that identifies risk of hospitalisation for potentially preventable problems in a paediatric population including all children under the age of 14 years living in the RER of Italy. The C-statistic of 0.78 indicates that the model performs well. By comparison, in a study predicting high-cost paediatric patients, Leininger et al reported a C-statistic of 0.73.[9] In their work in predictive risk modelling in the UK, Billings et al reported a C-statistics of 0.685[23] and C-statistics ranging from 0.731 to 0.780.[24] However, neither of these papers focused on a paediatric population. In a project also conducted in the Emilia-Romagna region of Italy but focused on the adult population, Louis et al[11] reported a C-statistic of 0.856. Given the similar organisation of the healthcare system and the similar database used for the adult and paediatric analyses,

Table 3	Observed and predicted events by risk group				
Risk groups (predicted risk range)	N	Average predicted risk, %	Observed prevalence, %	Expected frequency based on predicted risk	Number of observed events
Lower than average (≤1.2%)	378 258	0.5	0.5	2018	1896
Higher than average (>1.2%–2.5%)	99 714	1.7	1.8	1690	1758
At higher risk (>2.5%)	49 486	5.4	5.5	2683	2737
Total	527 458		1.2	6391	6391

Table 4 C-statistic, sensitivity, specificity and PPV

	C-statistic (overall model)=0.78	
	Cut-off points for comparison	
	'At higher risk'* score	'At higher risk'*+'Higher than average'† score
Sensitivity‡	0.43	0.70
Specificity§	0.91	0.72
PPV¶	0.06	0.03
True positives**	2737	4495

*'At higher risk' is defined as patients with a predicted risk of hospitalisation of>2.5%.

†'At higher risk'+'Higher than average', is defined as patients with a predicted risk of hospitalisation of >1.2%.

‡Sensitivity is defined as the proportion of those hospitalised who were predicted to be hospitalised (true positive rate).

§Specificity is the proportion of those not hospitalised who were not predicted to be hospitalised (true negative rate).

¶PPV is the proportion of those predicted to be hospitalised who were actually hospitalised.

**Positive predictives are the number of residents who were predicted to be at risk of hospitalisation at the predicted risk threshold and were actually hospitalised.

PPV, positive predictive value.

we believe that the somewhat lower C-statistic in the paediatric study results form the fact that hospitalisation is less frequent in children.

We believe that the definition of the dependent variable used in our models increases the likelihood that they are identifying patients whose risk may be reduced through proactive care. We have updated previously published criteria to include hospitalisations that may have been prevented by currently available vaccines, and we have used the logic of disease staging to include relevant hospitalisations that would have been missed using solely primary ICD-9-CM codes. Specifics of the selection criteria are available in the supplemental material.

The richness of the administrative data available in the RER allowed for a robust definition of the predictive variables. The RER data allow for the linkage of patients' use of diverse inpatient and outpatient healthcare services over multiple years. In addition, the ability to link child and mother's information allows the models to consider some of the mother's medical history such as the presence of chronic disease and use of prescription drugs in the years prior to birth as well as complications that may have arisen at birth.

There are limitations to our models. The models were developed with administrative data that lack some of the clinical specificity that would be useful in assessing patient risk. Children who have not had the types of encounters included in the RER database would have potentially missing information. The RER database does not have encounter level diagnostic data available documenting visits with the primary care paediatrician. The administrative data have very limited information available about the patient and family socioeconomic status. Our models use

prior utilisation among the predictor variables. With the administrative date, we cannot distinguish appropriate from inappropriate prior utilisation, which may bias our results. Despite their limitations, administrative data have many advantages for a project such as ours. They are relatively inexpensive to analyse and, in the case of the RER, include a large population over multiple years.

While the evidence was mixed, a systematic review suggests that hospitalisations can be prevented in children with medical complexity.[1] The Local Health Authority of Parma has begun working with the primary care paediatricians caring for the patients identified by the models to develop individual 'profiles' of children identified as being at higher risk. Data in the profiles, along with the more detailed information available in the medical record, can be used by the paediatricians to assess what additional intervention, if any, may help to manage the child's risk. For example, review of the profiles of higher risk children can help identify children whose parents might be contacted for a visit if they have not been seen recently. Summaries of prescriptions that have been filled from the profiles can be reviewed for potential over use, under use or inappropriate use of mediation. High-risk children with chronic illness might be referred to a specialist or home healthcare provided.

The RER healthcare system offers several advantages in the goal of reducing potentially preventable hospitalisation. Every child is enrolled with a primary care paediatrician. The population is quite stable allowing for continuity of care. Through the Italian NHS, every child is entitled to healthcare with little or no cost at the point of service. While the primary care paediatricians are paid on a per-capita basis, the RER can negotiate incentive payments and monitor improvements in care that may help to reduce avoidable hospitalisations. If successful, the results of the models can be applied by other Local Health Authorities in the RER, other Italian regions and other countries with similar data availability.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data available. Data used for this research were retrieved from the Regional database of the Emilia-Romagna Region provided through a collaborative agreement between the Regional Health Care and Social Agency, Emilia-Romagna, Italy, the Health Care Authority, Emilia-Romagna, Italy, and Thomas Jefferson University.

REFERENCES

1. Coller RJ, Nelson BB, Sklansky DJ, *et al*. Preventing hospitalizations in children with medical complexity: a systematic review. *Pediatrics* 2014;134:e1628–47.
2. Lu S, Kuo DZ. Hospital charges of potentially preventable pediatric hospitalizations. *Acad Pediatr* 2012;12:436–44.
3. Waljee A, Higgins P, Singal A. A primer on predictive models. *Clinical and Translational Gastroenterology* 2013;44:e44.
4. Huang L, Li S, He D, *et al*. A predictive risk model for medical intractability in epilepsy. *Epilepsy Behav* 2014;37:282–6.
5. Cerqueira DC, Soares CM, Silva VR, *et al*. A predictive model of progression of CKD to ESRD in a predialysis pediatric interdisciplinary program. *Clin J Am Soc Nephrol* 2014;9:728–35.
6. Lopes JM, Goulart EM, Siqueira AL, *et al*. Nosocomial infections in Brazilian pediatric patients: using a decision tree to identify high mortality groups. *Braz J Infect Dis* 2009;13:2.
7. Gorelick M, Scribano PV, Stevens MW, *et al*. Predicting need for hospitalization in acute pediatric asthma. *Pediatr Emerg Care* 2008;24:735–44.
8. McAna JF, Crawford AG, Novinger BW, *et al*. A predictive model of hospitalization risk among disabled medicaid enrollees. *Am J Manag Care* 2013;19:e166–74.
9. Leininger LJ, Saloner B, Wherry LR. Predicting high-cost pediatric patients. *Med Care* 2015;53:729–35.
10. Lo Scalzo L, Donatini A, Orzella L, *et al*. *Italy: Health system review*. 11. Copenhagen: World Health Organization, 2009. Report No. 6.
11. Louis DZ, Robeson M, McAna J, *et al*. Predicting risk of hospitalisation or death: a retrospective population-based analysis. *BMJ Open* 2014;4:e005223.
12. WHO Collaborating Centre for Drug Statistics Methodology. *Guidelines for ATC classification and DDD assignment 2013*. Oslo: WHO Collaborating Centre for Drug Statistics Methodology, 2012.
13. Regione Emilia-Romagna Assessorato Politiche per la Salute. *The Emilia-Romagna regional health service and the new welfare system: facilities, expenditure, and activities as of 31.12.2010. Programs, agreements and organisational models*. Bologna: Regione Emilia-Romagna Assessorato Politiche per la Salute, 2012.
14. Flores G, Abreu M, Chaisson CE, *et al*. Keeping children out of hospitals: parents' and physicians' perspectives on how pediatric hospitalizations for ambulatory care-sensitive conditions can be avoided. *Pediatrics* 2003;112:1021–30.
15. Shi L, Lu N. Individual sociodemographic characteristics associated with hospitalization for pediatric ambulatory care sensitive conditions. *J Health Care Poor Underserved* 2000;11:373–84.
16. Agency for Healthcare Research and Quality. *Pediatric quality measures program*. Rockville, MD, 2016. http://www.ahrq.gov/policymakers/chipra/factsheets/fullreports/index.html (accessed Jul 2016).
17. Gonnella JS, Hornbrook MC, Louis DZ. Staging of disease. A case-mix measurement. *JAMA* 1984;251:637–44.
18. In: Gonnella JS, Louis DZ, Gozum MV, Callahan CA, Barnes CA, *et al*. eds. *Disease staging clinical and coded criteria*. 629th Edn. Ann Arbor: Truven Health Analytics, 2012.
19. Maio V, Yuen E, Rabinowitz C, *et al*. Using pharmacy data to identify those with chronic conditions in Emilia Romagna, Italy. *J Health Serv Res Policy* 2005;10:232–8.
20. Prot-Labarthe S, Weil T, Angoulvant F, *et al*. POPI (Pediatrics: Omission of Prescriptions and Inappropriate prescriptions): development of a tool to identify inappropriate prescribing. *PLoS One* 2014;9:e101171.
21. Costelloe C, Metcalfe C, Lovering A, *et al*. Effect of antibiotic prescribing in primary care on antimicrobial resistance in individual patients: systematic review and meta-analysis. *BMJ* 2010;340:c2096.
22. Metsälä J, Lundqvist A, Virta LJ, *et al*. Prenatal and post-natal exposure to antibiotics and risk of asthma in childhood. *Clin Exp Allergy* 2015;45:137–45.
23. Billings J, Dixon J, Mijanovich T, *et al*. Case finding for patients at risk of readmission to hospital: development of algorithm to identify high risk patients. *BMJ* 2006;333:327.
24. Billings J, Georghiou T, Blunt I, *et al*. Choosing a model to predict hospital admission: an observational study of new variants of predictive models for case finding. *BMJ Open* 2013;3:e003352.