

# BMJ Open Validation of an algorithm to determine the primary care treatability of emergency department visits

Molly Moore Jeffery,<sup>1</sup> M Fernanda Bellolio,<sup>2</sup> Julian Wolfson,<sup>3</sup> Jean M Abraham,<sup>4</sup> Bryan E Dowd,<sup>4</sup> Robert L Kane<sup>4</sup>

**To cite:** Jeffery MM, Bellolio MF, Wolfson J, *et al*. Validation of an algorithm to determine the primary care treatability of emergency department visits. *BMJ Open* 2016;**6**:e011739. doi:10.1136/bmjopen-2016-011739

► Prepublication history and additional material is available. To view please visit the journal (<http://dx.doi.org/10.1136/bmjopen-2016-011739>).

Received 3 March 2016  
Revised 18 July 2016  
Accepted 29 July 2016

## ABSTRACT

**Objectives:** We propose a new claims-computable measure of the primary care treatability of emergency department (ED) visits and validate it using a nationally representative sample of Medicare data.

**Study design and setting:** This is a validation study using 2011–2012 Medicare claims data for a nationally representative 5% sample of fee-for-service beneficiaries to compare the new measure's performance to the Ballard variant of the Billings algorithm in predicting hospitalisation and death following an ED visit.

**Outcomes:** Hospitalisation within 1 day or 1 week of an ED visit; death within 1 week or 1 month of an ED visit.

**Results:** The Minnesota algorithm is a strong predictor of hospitalisations and deaths, with performance similar to or better than the most commonly used existing algorithm to assess the severity of ED visits. The Billings/Ballard algorithm is a better predictor of death within 1 week of an ED visit; this finding is entirely driven by a small number of ED visits where patients appear to have been dead on arrival.

**Conclusions:** The procedure-based approach of the Minnesota algorithm allows researchers to use the clinical judgement of the ED physician, who saw the patient to determine the likely severity of each visit. The Minnesota algorithm may thus provide a useful tool for investigating ED use in Medicare beneficiaries.

Older adults are among the most frequent users of emergency departments (EDs).<sup>1</sup> EDs are expensive venues to provide primary care and may have adverse consequences on older patients. Research has identified important gaps between elders' needs and the environment of the ED. A recent review of the literature found several areas where elder ED care sometimes falls short, including inadequate assistance with obtaining meals, using the toilet, and getting around.<sup>2</sup> Moreover, EDs are not designed to provide continuity of care, which has been shown to reduce costs<sup>3</sup> and improve long-term mortality in older adults.<sup>4</sup>

## Strengths and limitations of this study

- This study introduces a measure (the Minnesota algorithm) of the primary care treatability of emergency department (ED) visits that avoids some of the limitations of the Billings/Ballard algorithm currently in use.
- The new measure uses procedures performed by ED physicians to infer their clinical judgement of the severity of each visit.
- The Minnesota algorithm performs similarly to or better than the Billings/Ballard algorithm at predicting hospitalisation and death following an ED visit.
- It is limited to use in data with Current Procedural Terminology (CPT) codes.

It is particularly concerning, then, that Kaskie *et al*<sup>5</sup> found that one-third of ED visits by elderly people are categorised as 'not severe', indicating they may be treatable in other contexts. These visits could represent significant deficits in provision of care meeting patients' mental and physical needs.

Identifying less severe ED visits is a priority for determining the most appropriate setting for medical care. One technique for classifying ED visit severity applies an algorithm developed by Billings *et al*<sup>6</sup> at New York University to categorise visits using the primary diagnosis code recorded in the medical or claims record; see Jones *et al*<sup>7</sup> for a list of recent research studies using the algorithm. The Billings algorithm provides probabilities that an ED visit with a given primary diagnosis is non-emergent, emergent but primary care treatable (PCT), requires ED care but is potentially preventable, or requires ED care and is not potentially preventable. Researchers have expressed concern with aspects of the Billings algorithm. For example, Raven *et al*<sup>8</sup> note the significant differences between pre-



CrossMark

For numbered affiliations see end of article.

## Correspondence to

Dr Molly Moore Jeffery;  
jeffery.molly@mayo.edu

senting symptoms and discharge diagnoses for ED visits; their study found that presenting symptoms for visits rated PCT by the Billings algorithm based on the discharge diagnosis were the same as presenting symptoms for visits rated as ED care needed (EDCN) by the Billings algorithm. This suggests that at the time the patient presents to the ED, there may be little distinguishing patients who need to be seen in the ED from those who do not. Further, for the first 9 years after the Billings algorithm was developed, there was no study testing the validity of the algorithm as applied beyond the original sample.

A 2010 study by Ballard *et al*<sup>9</sup> was the first to validate a revised form of the Billings algorithm. The Ballard version of the algorithm consolidates the four Billings/NYU categories into three: 'non-emergent', 'intermediate' and 'emergent' and assigns a single category to each visit. That study validated the algorithm by estimating the ability of the algorithm to predict hospitalisation on the same day, within 1 day, or within 1 week of the ED visit, and death within 30 days of the ED visit. Results suggested that algorithm is a good predictor of these outcomes. However, the Ballard team used a somewhat unusual population for their validation study. The Medicare sample included only members of Kaiser Permanente Northern California, a managed care plan. Most studies of Medicare-managed care plans have shown favourable risk selection into these private plans<sup>10</sup>—that is, people enrolled in Medicare-managed care plans tend to be healthier than those enrolled in Fee-for-Service (FFS) Medicare. Indeed, the Ballard study sample appears to be significantly healthier than the general Medicare population. The study reports that just 9.2% of people in the Medicare sample had any hospitalisations in 1998. By comparison, the 1998 rate of hospital discharges for Medicare FFS enrollees in California was 31.3 per 100 beneficiaries<sup>11</sup> and the nationwide proportion of Medicare FFS enrollees with at least one hospitalisation in 2000 was 23.2%.<sup>12</sup>

However, the Billings/Ballard algorithm may not be well-suited for application in all contexts. Of concern is its accuracy when applied to populations very different from the one used to create the measure: an all-ages population seen in EDs in the Bronx, New York. They likely differ from a national Medicare population in several ways, including comorbidity profiles and geographic variations in practice.

The Billings/Ballard algorithm provides a single set of probabilities that a visit with a given final visit diagnosis was non-emergent, intermediate or emergent. Those probabilities do not vary by patient characteristics like age, comorbidity, frailty or presenting problem. Nor do they vary by social contexts like availability of primary care. The population from the original Billings sample was studied precisely because they lacked sufficient access to primary care. This was expected to affect the severity of presenting problems in the EDs included in that study; one of the study hypotheses was that patients

lacking primary care access would present to the ED with minor problems because they had nowhere else to go for care.

We found considerable evidence that the Billings/Ballard algorithm's diagnosis severities are inaccurate when applied to a Medicare population. Our study sample includes 148 000 ED visits with a primary diagnosis considered 100% PCT<sup>i</sup> in the original Billings study; but in our sample, 49% of these ED visits resulted in a hospital admission, received critical care in the ED or received the most severe level of ED evaluation and management code, indicating a high complexity or high severity problem unlikely to be treatable in primary care.

It would be possible to propose an updated version of the Billings/Ballard approach that adjusted for some patient factors like age and comorbidity to provide a more accurate assessment of visit severity based on diagnoses. This would require an extensive study, including abstraction and review of thousands of medical records of ED visits. Instead, we propose using the clinical judgement of the treating ED physician to determine whether a visit could have been conducted in a primary care setting. We infer physician judgement using the procedures that were billed for the visit, with special attention to the evaluation and management code representing the complexity and severity of the visit.

This study proposes a new measure of primary care treatability of ED visits using procedure codes from administrative data to infer the treating physician's determination of the severity of an ED visit. We use Medicare data to compare the new measure's performance to the Billings/Ballard algorithm in predicting two relevant prognostic outcomes: hospitalisation and death following an ED visit. This approach is consistent with the literature on validation of measures where there is no gold standard definition of the domain being measured.<sup>13</sup>

## METHODS

### Algorithm approach

The Minnesota algorithm categorises the primary care treatability of ED visits using procedure codes. In the USA, most outpatient care is billed using Current Procedural Terminology (CPT) codes, which were developed and are maintained by the American Medical Association (AMA) for the purposes of describing services provided by physicians.<sup>14</sup> Table 1 provides descriptions of evaluation and management (E&M) codes most frequently billed in the ED. The descriptions are taken from the AMA's documentation and Medicare billing manuals.

CPT documentation for E&M codes 99281–99285 includes the physician activities that are required for

<sup>i</sup>Includes Billings categories non-emergent and emergent but primary care treatable.

**Table 1** Key emergency department evaluation and management codes

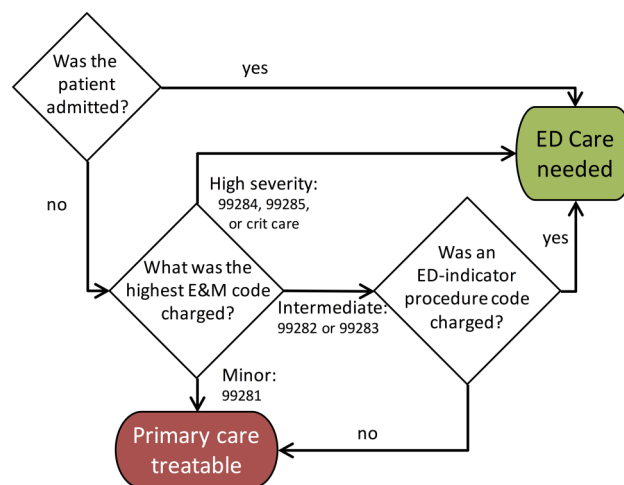
| ED-specific E&M Code   | Requirements for reimbursement  |                          |   | Characteristics  |
|------------------------|---|--------------------------|---|--|
|                        | History   | Examination              | Medical decision-making   | Severity of presenting problem   |
| 99281                  | Problem focused   | Problem focused          | Straightforward   | Usually self-limited or minor  |
| 99282                  | Expanded problem focused  | Expanded problem focused | Low complexity  | Usually low to moderate  |
| 99283                  | Expanded problem focused  | Expanded problem focused | Moderate complexity   | Usually moderate   |
| 99284                  | Detailed  | Detailed                 | Moderate complexity   | Usually high, requiring urgent evaluation by physician, but not posing an immediate significant threat to life or physiological function |
| 99285                  | Comprehensive   | Comprehensive            | High complexity   | Usually high, posing an immediate significant threat to life or physiological function   |
| Critical care<br>99291 | Description<br>Evaluation and management of critically ill or critically injured patient; first 30–74 min |                          | Requirements for reimbursement<br>Constant physician attention; high complexity decision-making to assess, manipulate and support vital system function to prevent or treat single or multiple vital organ system failure.† |  |
| 99292                  | Each additional 30 min  |                          |   |  |

Source: all text taken directly or paraphrased from AMA CPT descriptions except † from CMS Pub 100-4, 12, 30.6.12; note: all ED-specific E&M codes (99281–99285) include counselling and coordination of care ‘consistent with the nature of the problem(s) and/or family’s needs’. AMA, American Medical Society; CMS, Centers for Medicare & Medicaid Services; CPT, Current Procedural Terminology; ED, emergency department; E&M, evaluation and management.

reimbursement of each code, as well as general characteristics of the visits. Reimbursement requirements include how detailed the history and physical examination should be and how complex the medical decision-making for a visit to be appropriately billed under each code. Code documentation also includes a summary statement of the severity of the presenting problem for the visit. The least severe E&M code (99281) is stated to be for a problem that is ‘usually self-limited or minor’. Code 99285, on the other hand, is for problems of high severity ‘posing an immediate threat to life or physiological function’.

We found visits in the ED nearly always contained one of the five ED-specific E&M codes described above. In addition, some of the visits included critical care E&M codes 99291 and 99292. These codes indicate critical illness or injury requiring care beyond the most severe ED E&M code. The Centers for Medicare & Medicaid Services (CMS) requirements for billing these codes are included in [table 1](#). Critical care codes can be billed anywhere in the hospital, including intensive care units, coronary care units, and so on. For the purposes of this algorithm, we only used critical care codes that were billed with a place of service or revenue code, indicating they were performed in the ED. These seven codes form the basis for the Minnesota algorithm’s categorisation. [Figure 1](#) presents the approach as a flow chart.

For visits without an inpatient admission, the E&M codes and other procedures associated with the visit are examined for evidence that the patient required ED care. Visits for which the highest E&M code billed was 99281—the least severe code—are coded as PCT. Visits for which the highest E&M code billed was 99284,

**Figure 1** Minnesota algorithm flow chart. ED, emergency department; E&M, evaluation and management.

99285, or critical care codes 99291 and 99292 are coded as EDCN. Visits with an intermediate severity E&M code of 99282 or 99283 are further examined for the presence of procedures that suggest the patient needed to be seen in the ED; we call these procedures ‘ED indicator procedures’. The process of creating this list is further described below. Intermediate severity visits with an ED indicator procedure are categorised as EDCN; visits without one are categorised as PCT.

The ED indicator procedure list includes procedures that take place outside of the ED but were associated with the ED visit, for example, head CT or laboratory tests. Thus, all procedures performed on the same day

as the ED visit must be examined—not just procedures with a place of service in the ED or an ED revenue centre. Furthermore, physicians and hospitals can bill for an ED E&M code, and these codes need not match in severity. In cases where a visit generated multiple E&M codes, the following rules were used to select the code to classify the visit:

1. E&M codes charged by physicians were selected over those charged by institutions.
2. If there were multiple codes charged by physicians or, in the absence of a physician E&M code, by the institution, the most severe code was used.

Visits that resulted in an inpatient admission are determined to have required ED care, regardless of the procedures billed in the ED. This criterion can be omitted if desired by users of the algorithm; it is intended to capture visits where a patient presents to the ED but is immediately admitted, potentially not generating an ED E&M code or generating only a low-severity or intermediate-severity E&M code and no ED indicator procedures. This situation affects a relatively small number of visits; 9% of all ED visits observed did not have one of the 7 E&M codes used in the algorithm. Of those, 31% resulted in hospitalisation. In addition, 0.5% of all ED visits observed would have received a PCT categorisation but for the hospitalisation. That is, these visits either had an ED E&M code of 99281 or had an E&M code of 99282 or 99283 and did not have an ED indicator procedure. Thus, omitting the inpatient criterion changes the visit rating for 3% of all visits, 85% of which are changed from EDCN to unclassified (2.8% of all visits), with the remaining 15% changed from EDCN to PCT (0.5% of all visits). Variants of this approach have

been used before; Davis *et al*<sup>15</sup> categorised visits using the 5 ED-specific E&M codes and the presence or absence of any other billed procedure, and Wolinsky *et al*<sup>16</sup> used just the E&M codes. This study adds to the prior literature by introducing a targeted list of ED-indicator procedures, validating the measure and exploring characteristics of ED visits that are classified differently by the Minnesota algorithm and the most frequently used established algorithm, the Billings/Ballard algorithm.

### Population studied

Two years of claims data (2011–2012) for a nationally representative sample of Medicare beneficiaries were used to develop and validate the new measure.<sup>17</sup> Summary statistics describing the validation sample used in this study are provided in table 2. A cohort flow chart is available in online supplementary file 1.

### ED indicator procedures

To create a list of procedures that indicate that a visit with an intermediate level E&M code (99282–3) was appropriate for the ED, we combined empirical analysis of claims data and physician review. A list of potential indicator procedures to distinguish among moderate severity ED visits was compiled of procedures frequently performed in the ED but not frequently performed in primary care visits. These codes were reviewed by an emergency medicine physician, a geriatrician and an urgent care physician for clinical logic—whether the procedures seem to suggest that the patient required care in an ED. We asked physicians reviewing the list to consider the following criteria to determine whether the

**Table 2** Summary statistics: validation sample

|  | Mean      | SD    | Minimum | Maximum |
|--|-----------|-------|---------|---------|
| <b>Beneficiary characteristics</b>     |           |       |         |         |
| Age (in 2011)*                         | 67.75     | 14.51 | –1      | 98      |
| <b>Race</b>                            |           |       |         |         |
| White                                  | 82.4%     |       |         |         |
| Black                                  | 12.1%     |       |         |         |
| Other                                  | 5.6%      |       |         |         |
| Female                                 | 57.3%     |       |         |         |
| Medicaid†                              | 27.9%     |       |         |         |
| HCC community score                    | 2.51      | 1.66  | 0.12    | 18.28   |
| Died during study period               | 18.7%     |       |         |         |
| ED visits in sample                    | 2.96      | 4.03  | 1       | 351     |
| Months in sample                       | 21.86     | 5.21  | 1       | 24      |
| <b>Visit-level outcome frequencies</b> |           |       |         |         |
| Death within 7 days of visit           | 1.9%      |       |         |         |
| Death within 1 month of visit          | 4.6%      |       |         |         |
| Hospitalisation within 1 day of visit  | 32.5%     |       |         |         |
| Hospitalisation within 1 week of visit | 36.3%     |       |         |         |
| Total ED visits                        | 2 448 112 |       |         |         |
| Total beneficiaries                    | 827 844   |       |         |         |

\*Three beneficiaries born in 2012 have an age of –1 in 2011.

†Operationalised as any months of state coverage buy-in from denominator file.

ED, emergency department; HCC, hierarchical condition category.



procedure should be included as an ED indicator procedure:

1. Performance of the procedure always or nearly always indicates a visit requiring ED care; examples: endotracheal intubation, cardiopulmonary resuscitation (CPR).
2. Procedure is unlikely to be available in a primary care office and performance is likely to be time-sensitive; examples: CTs and MRIs, morphine injections.
3. Procedure is appropriate and widely available in primary care, however performance in the ED indicates the physician believed there may be a severe problem; examples: laboratory tests for troponin or creatine kinase.

Reviewers largely agreed with each other on the inclusion of procedures in the ED indicator list. In the few cases of disagreement, we discussed the procedures and came to an agreement. The reviewers were each separately working off a list of all procedures frequently seen in the ED. There were several hundred procedures on the original list. We grouped them by the frequency with which they appeared in primary care visits. Most of our effort was focused on procedures that were common in the ED and not common in primary care. However, we did review the procedures seen frequently in primary care and included on the final list cardiac enzyme tests. We also added some procedures rarely seen in any context, but logically related to more common procedures that appeared on the list. For example, suturing of very large wounds is rare, but was included on the list because of the inclusion of suturing of less severe wounds. The final list contains 120 procedure codes and is available in online supplementary file 2.

## Outcomes

Two outcomes were used to validate the Minnesota algorithm in comparison with the Billings/Ballard algorithm: death (at 1 week and 1 month following the ED visit) and hospitalisation (at 1 day and 1 week following the ED visit). These outcomes are associated with either severity of illness (resulting in death) or need for more intensive care than can be provided in an outpatient setting (resulting in hospitalisation). The hospitalisation criterion was not used to classify ED visits in the analyses using hospitalisation as an outcome.

## Validation sample

Validation analyses were performed on a subset of observed ED visits: only beneficiaries who were covered by FFS Medicare parts A and B were included in the analysis, to ensure we had as much information as possible on hospitalisation outcomes.

## Statistical methods

Generalised estimating equations were used to estimate ORs for death or hospitalisation after a severely rated ED visit compared with a primary care treatable

visit. The method accounts for correlation due to potentially repeated observations for beneficiaries with multiple ED visits observed over the 2 years of data.

The model estimated is as follows:

$$\text{logit}(\Pr(y_{is} = 1 | \mathbf{x}_i)) = \boldsymbol{\beta} \mathbf{x}_i + \gamma \text{AlgCategory}_{is},$$

$$y_{is} \sim \text{Bernoulli}(\pi_{is})$$

where  $y_{is}$  is a binary variable for whether beneficiary  $i$  died (or was hospitalised) during the relevant time period after ED visit  $s$ ,  $\mathbf{x}_i$  is a vector of person-specific explanatory variables (Medicaid status, hierarchical condition category (HCC) score, race, sex and age) and  $\text{AlgCategory}_{is}$  is a categorical variable indicating which algorithm category was assigned to the ED visit. The omitted reference category was 'non-emergent' for the Billings/Ballard algorithm and 'primary care treatable' for the Minnesota algorithm. An independence working correlation matrix was specified, as was the Huber-White variance-covariance estimator.

Included covariates were available in the data for each beneficiary by year. For example, the comorbidity measure—HCC—was measured for 2011 and 2012. The covariates included in the model are from the year in which in the ED visit took place. Stata V.14 was used to perform all analyses (Stata Statistical Software (program), 14 version. College Station, Texas: StataCorp LP, 2015).

Further details on creation of all variables used in the analysis are included in online supplementary file 3.

## RESULTS

### Classification of visits

**Table 3** compares the results of the Minnesota algorithm to the results of the Billings/Ballard algorithm classifying the total sample of ED visits. The Minnesota algorithm classifies 17.5% of all ED visits observed in the sample as PCT. The Billings/Ballard algorithm classifies a larger proportion of the sample as non-emergent: 28.7%.

Ratings concordance was defined as the Minnesota algorithm rating a visit 'primary care treatable' and the Billings/Ballard rating it 'non-emergent' or the Minnesota algorithm rating it 'ED care needed' and the Billings/Ballard algorithm rating it 'emergent'. In all, two-thirds of ED visits have concordant ratings from the two algorithms: 10.3% are rated 'primary care treatable' and 'non-emergent', while 56.5% are rated 'ED care needed' and 'emergent'.

However, 21% of visits have discordant ratings from the two algorithms: 5.2% are rated 'primary care treatable' by the Minnesota algorithm and 'emergent' by the Billings/Ballard algorithm—for these visits, the Billings/Ballard algorithm gave a more severe rating. In 16.0% of all ED visits, the opposite was true: the Minnesota

**Table 3** Comparison of categorisation of total sample ED visits: Minnesota algorithm and Billings/Ballard algorithm

| MN algorithm category  | Billings/Ballard category |                  |              |                           |                  | Total (%) |
|------------------------|---------------------------|------------------|--------------|---------------------------|------------------|-----------|
|                        | Non-emergent (%)          | Intermediate (%) | Emergent (%) | Special category only (%) | Unclassified (%) |           |
| Primary care treatable | 10.3                      | 0.9              | <b>5.2</b>   | 0.3                       | 0.8              | 17.5      |
| ED care needed         | <b>16.0</b>               | 2.6              | 56.5         | 0.3                       | 0.6              | 76.0      |
| Unclassified           | 2.5                       | 0.2              | 2.8          | 0.1                       | 1.0              | 6.5       |
| Total                  | 28.7                      | 3.7              | 64.4         | 0.7                       | 2.4              | 100       |

N=2 644 545 (total observed sample, not just validation sample).

Percentages may not total 100% due to rounding.

Discordant algorithm ratings in bold (algorithms disagree on visit severity).

ED, emergency department; MN, Minnesota.

algorithm rated the visits as more severe than the Billings/Ballard algorithm.

The five most common primary diagnoses for visits with a more severe Billings/Ballard rating were chest pain not otherwise specified (NOS), shortness of breath, lumbago, sciatica and sprain of the shoulder/arm NOS. The most common primary diagnoses for discordant visits where the Billings/Ballard algorithm was less severe than the Minnesota algorithm were headache, abdominal pain of unspecified site, abdominal pain of other specified site, dizziness and giddiness, urinary tract infection NOS, and malaise and fatigue NEC.

### Validation of algorithm

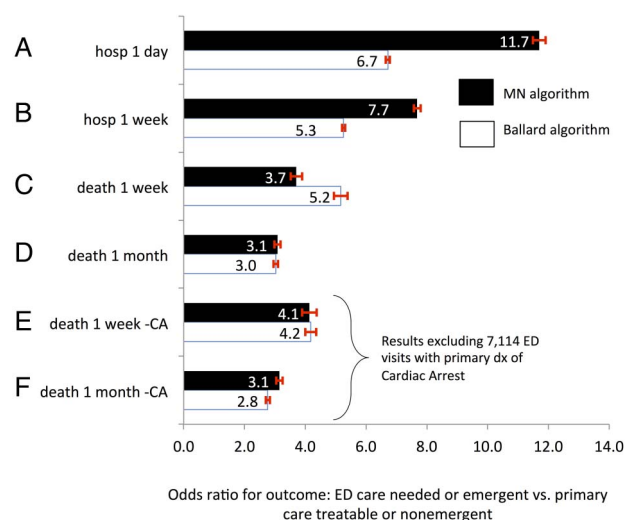
Full results of all validation analyses are presented in online supplementary file 3. A summary of key analyses is provided in figure 2A–F

### Hospitalisation outcomes

The predictive validity of the algorithms was tested using individual outcomes. For analyses of hospitalisation outcomes, the Minnesota algorithm was applied without the inpatient criterion, as described on page 9; that is, the first decision node in the algorithm flow chart was ignored and only E&M codes and ED indicator procedures were used to classify visits. ED visits that appear in the inpatient file but do not have any ED-specific E&M codes are unclassified and separated into their own category to avoid mixing them with visits that had neither an E&M code nor an inpatient admission (presumably a very different group of visits).

Figure 2A presents the analysis of hospitalisation within 1 day of an ED visit from the Minnesota algorithm (black) and the Billings/Ballard algorithm (white). Results are given as ORs, with the reference category being the least severe category for the algorithms: for the Minnesota algorithm, it is the ‘primary care treatable’ category, and ‘non-emergent’ for the Billings/Ballard algorithm.

The odds of being hospitalised within 1 day of an ED visit are 11.70 times higher when a visit is rated ‘ED care



Confidence intervals indicated as error bars.

N=2,448,112 ED visits (2,440,998 excluding Cardiac Arrest visits)

Validation analyses adjusted for comorbidities using HCC score, race, sex, age, and Medicaid status.

OR for Ballard is odds of outcome after Emergent ED visit over odds after Non-Emergent visit. OR for MN alg is ED care needed over Primary care treatable.

**Figure 2** Predictive validity of Minnesota Algorithm and Billings/Ballard algorithm. CA, cardiac arrest; ED, emergency department; HCC, hierarchical condition category; MN, Minnesota.

needed’ by the Minnesota algorithm compared with visits rated ‘primary care treatable’ by the Minnesota algorithm. For the Billings/Ballard algorithm the odds of hospitalisation within 1 day of an ED visit are 6.72 times higher for visits rated ‘emergent’, compared with visits rated ‘non-emergent’ by that algorithm.

Results for hospitalisation within 1 week of an ED visit (figure 2B) are similar to those for hospitalisation within 1 day of a visit.

### Death outcomes

The predictive validity of the algorithm was also tested using death at 1 week and 1 month after each ED visit as the outcome of interest. Results are compared with the Billings/Ballard algorithm in figure 2C, D.

Figure 2C gives the results for the analysis of death within 1 week of an ED visit. A person with a visit rated

'ED care needed' by the Minnesota algorithm faces 3.70 times the odds of death within 1 week compared with a person with a visit rated 'primary care treatable'. Using the Billings/Ballard algorithm, a person with a visit rated 'emergent' has 5.16 times the odds of death within a week of the visit compared with a person with a visit rated 'non-emergent'.

The analysis of death within 1 month of an ED visit shows no difference in performance between the two algorithms (figure 2D). The odds of death within 1 month of an ED visit are 3.08 times higher for a person with a visit rated 'ED care needed' compared with one rated 'primary care treatable' by the Minnesota algorithm, and the odds are 3.03 times higher for a person with a visit rated 'emergent' compared with one rated 'non-emergent' by the Billings/Ballard algorithm.

The analysis of death within 1 week of an ED visit is the only analysis in which the Billings/Ballard algorithm was a better predictor of the outcome. We explored this finding further and found that the higher OR for the Billings/Ballard algorithm compared with the Minnesota algorithm is driven by a single diagnosis code: cardiac arrest. Figure 3 presents the Minnesota algorithm classifications for visits with and without a primary diagnosis of cardiac arrest. Out of 2.4 million ED visits included in the analysis, only 0.29% (N=7114) have a primary diagnosis code of cardiac arrest. The Minnesota algorithm categorises 6% of these visits as PCT—that is, the E&M code assigned for the visit was 99281 or it was 99282/99283 and none of the ED indicator procedures appeared in the claims related to the visit. The ED indicator procedure list includes CPR and cardioversion, which we would expect to see in cases of cardiac arrest if physicians in the ED believed the person was at all resuscitable; the absence of these procedures suggests that the patients may have arrived in the ED dead and beyond help or with a do not resuscitate order. In support of this hypothesis, our records indicate death within 1 day for 98.5% of people with a

visit rated PCT and a primary diagnosis of cardiac arrest. This is a higher proportion of deaths than cardiac arrest visits rated EDCN by the Minnesota algorithm (87.1%), potentially suggesting that physicians were noting either the futility or the patients' wishes against further rescue attempts in those visits rated PCT by the Minnesota algorithm.

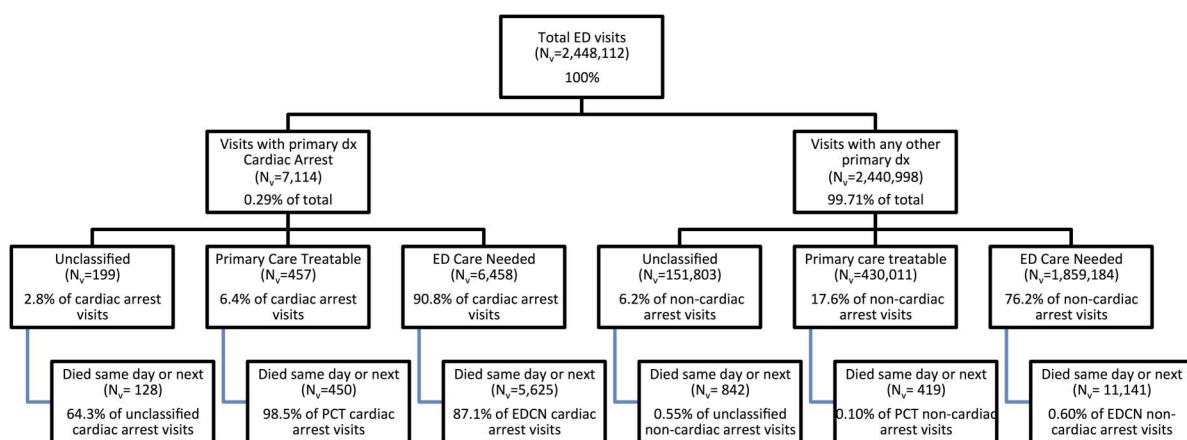
When 7114 ED visits with cardiac arrest, as the primary diagnosis code, are excluded from the analysis, the Billings/Ballard algorithm is no longer a better predictor of death than the Minnesota algorithm (figure 2E); the ORs for death within 1 week of the ED visit are statistically equal across the two algorithms. The OR for death within 1 month (figure 2F) is slightly higher for the Minnesota algorithm.

### Sensitivity analyses

Because the HCC measure was developed to predict costs rather than usage, we performed sensitivity analyses using a combination of the Charlson and Elixhauser comorbidity measures.<sup>18</sup> There was no change in inference; the direction of the ORs do not change, and the relative magnitudes of the ORs for the Billings/Ballard versus the Minnesota algorithm do not change.

The Medicare population includes subpopulations that may be different in their healthcare use, including dual eligibles and people who qualify for Medicare coverage due to disability or end-stage renal disease (ESRD). We ran all analyses on the following subpopulations: Medicaid eligible, not Medicaid eligible, eligible due to age and eligible due to disability or ESRD. Again, inference was not changed.

Comorbidity scores (ie, HCC scores in the main analysis) were calculated for the same year as the ED visit, meaning that an individual ED visit's HCC score could include illnesses that didn't manifest until after the ED visit. To address this concern, we repeated the analysis on 2012 ED visits by people for whom we had 2011 comorbidity scores, using the prior year score in the equation; the analysis included 1.2 million ED visits by



**Figure 3** Minnesota algorithm classification of ED visits for cardiac arrest with mortality rates. ED, emergency department; EDCN, emergency department care needed; PCT, primary care treatable.

547 124 beneficiaries. Results did not change substantially. Results for all sensitivity analyses are available on request.

## DISCUSSION

The Billings/Ballard algorithm and the Minnesota algorithm produced concordant ratings for the majority of ED visits. We don't have sufficient information available in claims records to determine which algorithm is correct in cases where they disagree. However, the list of diagnoses associated with discordantly rated visits may provide some insight.

Looking just at visits that were rated more severely by the Billings/Ballard algorithm than the Minnesota algorithm, the two most common diagnoses seem quite severe: chest pain and shortness of breath. Although we can't say for certain why the physicians who saw these patients did a very modest workup when the primary diagnosis seems potentially severe, we know that in 93% of these discordant visits with chest pain or shortness of breath as the primary diagnosis, the ED-specific E&M code was 99282 or 99283, so the PCT rating for the visit was based on the lack of any ED indicator procedures. That is, in 93% of these visits, the patient was discharged without receiving laboratory tests for creatine kinase, troponin, myoglobin, or D-dimer, or blood gases. The other 7% of these visits had an E&M code of 99281, which designates a very minor problem and workup. This suggests that the physicians attending these patients did not have a high suspicion of severe causes of chest pain like heart attack or hypoxaemia.

The diagnoses associated with visits where the Minnesota algorithm rated visits more severely than the Billings/Ballard algorithm suggests that the Billings/Ballard approach may be problematic, as it rates diagnosis severity without reference to patient age. These visits represent symptoms that may have very different interpretations in an elderly population than a younger one. Many diseases can have atypical presentations in elderly people. As people age, they experience a gradual loss of function that is most apparent in the response to stress (ref. 19, p. 5). While baseline function may be very similar to younger persons', when an organ system is stressed, the elderly person may be less responsive to that stress, resulting in atypical symptoms (or a lack of symptoms) despite serious illness. Thus, the non-specific symptoms of headache, dizziness or malaise may merit a more in-depth workup in an elderly person than a younger one.

We compared the new measure's performance with the Billings/Ballard algorithm in predicting hospitalisation and death following an ED visit. The Minnesota algorithm based on procedure codes and the Billings/Ballard algorithm were able to predict hospitalisations and deaths after ED visits in a Medicare FFS population. The Minnesota algorithm is a better predictor of hospitalisation after an ED visit and a slightly worse predictor

of death after an ED visit than the Billings/Ballard algorithm in analyses that controlled for race, sex and HCC comorbidities.

The Billings/Ballard algorithm's superior ability to predict death in this sample is driven by visits with a primary diagnosis of cardiac arrest, some of which are rated 'primary care treatable' by the Minnesota algorithm. These visits likely include patients who are dead on arrival at the ED and people who have expressed a wish not to be resuscitated, which could result in a low-severity E&M code and the absence of any life-saving measures from the list of ED indicator procedures. Researchers who wish to classify these visits as requiring ED care could consider a hybrid approach that combines the Minnesota algorithm with a flag for the Cardiac Arrest International Classification of Diseases (ICD) code.

Researchers wishing to classify the severity of ED visits observed in claims data will want to consider several factors in choosing between the two algorithms. The Minnesota algorithm uses procedure codes to determine whether an ED visit was primary care treatable, while the Billings/Ballard algorithm uses diagnosis codes. The advantage to the procedure code approach is that it captures the medical care deemed appropriate by a physician who actually saw the patient. However, physicians are often paid based on the amount of care they give; this financial incentive to provide more care could increase the measured severity of the visit.

We used publicly available results from audits of Medicare ED claims by the Office of the Inspector General of the Department of Health and Human Services<sup>20</sup> and the annual Comprehensive Error Rate Testing (CERT) audit programme sponsored by CMS<sup>21</sup> to estimate the impact of improper E&M coding. Our simulation results suggest that if the visits in our sample had been classified according to their 'true' E&M code, the proportion of visits rated 'primary care treatable' would have changed very little: from 17.5% to 19.1% (2.5–97.5 centile range based on 1000 replications: 19.09% to 19.16%). Perhaps more saliently, the Minnesota algorithm is validated here using the actual E&M codes charged and paid and performs quite well. Over time, if coding practices shift significantly, the algorithm may need to be validated again.

The diagnosis-based approach of the Billings/Ballard algorithm avoids problem of basing the rating of a visit on the amount of care received, with the attendant potential for bias from physician incentives. However, diagnoses may be little more objective than procedures; Song *et al*<sup>22</sup> showed that moving from a low-intensity practice area to a high-intensity area doubled the number of diagnoses recorded for Medicare beneficiaries.

Our approach to validating the new measure in the absence of a gold standard definition of the primary care treatability of ED visits follows the advice of Rutjes *et al*<sup>13</sup> to compare the performance of measures in predicting prognostic outcomes associated with the domain being



measured. As a result, we did not attempt to calculate or interpret sensitivity and specificity of the two measures.

Finally, it is important to note that this measure does not claim to determine the appropriateness of an individual ED visit; instead it measures the resources used in a visit. There is no agreement on how to judge which ED visits are inappropriate. The use of expert opinion, self-rating by patients, review of department activities and subsequent admissions have all failed to determine appropriateness when applied to the patient level.<sup>23</sup> When evaluating the appropriateness of an ED visit, one must consider variables not found in claims data, like time of day and primary care booking density. Limited access to primary care, convenience and lack of insurance are possible reasons why patients with PCT concerns present to an ED.<sup>24–27</sup> Focusing on systems issues rather than patient characteristics may be the most productive strategy to improve appropriate use of emergency care.<sup>28</sup>

## CONCLUSIONS

This study introduces a new claims-computable measure of the primary care treatability of ED visits. The Minnesota algorithm shows good validity, as a strong predictor of hospitalisations and deaths. Its performance is similar to or better than the most commonly used existing algorithm to assess the severity of ED visits. The procedure-based approach of the Minnesota algorithm allows researchers to use the clinical judgement of the ED physician who saw the patient to determine the likely severity of each visit, rather than depending on a calculation of average visit severity for a potentially very different population. The Minnesota algorithm may thus provide a useful tool for investigating ED use in Medicare beneficiaries.

### Author affiliations

<sup>1</sup>Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota, USA

<sup>2</sup>Department of Emergency Medicine, Mayo Clinic, Rochester, Minnesota, USA

<sup>3</sup>Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, Minnesota, USA

<sup>4</sup>Division of Health Policy and Management, School of Public Health, University of Minnesota, Minneapolis, Minnesota, USA

**Twitter** Follow M. Fernanda Bellolio at @mfbellolio

**Acknowledgements** The authors thank Frank Wharam for assistance with the ED-indicator procedure list and helpful comments. This research was presented at the AHRQ NRSA conference in San Diego in June 2014 and at AcademyHealth in Minneapolis in June 2015.

**Contributors** All authors contributed to the conception or design of this work, gave final approval of the version to be published and agree to be accountable for all aspects of the work. MMJ was responsible for the acquisition and analysis of the data and the drafting of the work. JMA, BED, JW, MFB and RLK provided critical revision for important intellectual content.

**Funding** Financial support for this research was provided by AHRQ in the form of an NRSA predoctoral training grant, AHRQ T32 HS000036-24 DOWD, BRYAN E (PI).

**Competing interests** MMJ was supported by an AHRQ NRSA training grant.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data sharing statement** No additional data are available.

**Open Access** This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

## REFERENCES

1. National Center for Health Statistics. *Health, United States, 2014: with special feature on adults aged 55–64*. Hyattsville, MD: Centers for Disease Control and Prevention, 2015.
2. Shankar KN, Bhatia BK, Schuur JD. Toward patient-centered care: a systematic review of older adults' views of quality emergency care. *Ann Emerg Med* 2014;63:529–50.e1.
3. Hussey PS, Schneider EC, Rudin RS, *et al*. Continuity and the costs of care for chronic disease. *JAMA Intern Med* 2014;174:742–8.
4. Wolinsky FD, Bentler S, Liu L, *et al*. Continuity of care with a primary care physician and mortality in older adults. *J Gerontol A Biol Sci Med Sci* 2010;65:421–8.
5. Kaskie B, Obrizan M, Jones MP, *et al*. Older adults who persistently present to the emergency department with severe, non-severe, and indeterminate episode patterns. *BMC Geriatr* 2011;11:65.
6. Billings J, Parikh N, Mijanovich T. Emergency department use in New York City: a substitute for primary care? *Issue Brief (Commonw Fund)* 2000:1–5.
7. Jones K, Paxton H, Hagtvedt R, *et al*. An analysis of the New York University Emergency Department Algorithm's suitability for use in gauging changes in ED usage patterns. *Med Care* 2013;51:e41–50.
8. Raven MC, Lowe RA, Maselli J, *et al*. Comparison of presenting complaint vs discharge diagnosis for identifying “nonemergency” emergency department visits. *JAMA* 2013;309:1145–53.
9. Ballard DW, Price M, Fung V, *et al*. Validation of an algorithm for categorizing the severity of hospital emergency department visits. *Med Care* 2010;48:58–63.
10. Mello MM, Stearns SC, Norton EC, *et al*. Understanding biased selection in Medicare HMOs. *Health Serv Res* 2003;38:961–92.
11. Dartmouth Institute for Health Policy & Clinical Practice. The Dartmouth Atlas of Health Care. 2016. <http://www.dartmouthatlas.org>
12. Centers for Medicare & Medicaid Services. CMS Statistics. Department of Health and Human Services, 2003. <https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/CMS-Statistics-Reference-Booklet/Downloads/2003CMSStatistics.pdf#table41> (accessed Aug 2016).
13. Rutjes AWS, Reitsma JB, Coomarasamy A, *et al*. Evaluation of diagnostic tests when there is no gold standard. A review of methods. *Health Technol Assess* 2007;11:iii, ix–51.
14. Centers for Medicare & Medicaid Services. HCPCS Coding Questions. 7/22/2013. [https://www.cms.gov/Medicare/Coding/MedHCPCSGenInfo/HCPCS\\_Coding\\_Questions.html](https://www.cms.gov/Medicare/Coding/MedHCPCSGenInfo/HCPCS_Coding_Questions.html)
15. Davis JW, Fujimoto RY, Chan H, *et al*. Identifying characteristics of patients with low urgency emergency department visits in a managed care setting. *Manag Care* 2010;19:38–44.
16. Wolinsky FD, Liu L, Miller TR, *et al*. Emergency department utilization patterns among older adults. *J Gerontol A Biol Sci Med Sci* 2008;63:204–9.
17. Centers for Medicare & Medicaid Services. Standard Analytic Files—LDS. 2016. <http://www.cms.gov/Research-Statistics-Data-and-Systems/Files-for-Order/LimitedDataSets/StandardAnalyticFiles.html>
18. Gagne JJ, Glynn RJ, Avorn J, *et al*. A combined comorbidity score predicted mortality in elderly patients better than existing scores. *J Clin Epidemiol* 2011;64:749–59.
19. Kane RL, Ouslander JG, Abrass IB, *et al*. *Essentials of clinical geriatrics*. 7th edn. New York: McGraw-Hill Education, 2013.
20. Office of Evaluation and Inspections [OIG OEI]. *Improper payments for evaluation and management services cost Medicare billions in 2010*. Department of Health and Human Services, 2014.
21. Centers for Medicare & Medicaid Services. *Appendices Medicare Fee-for-Service 2011 Improper Payments Report*. 2012. <https://www.cms.gov/Research-Statistics-Data-and-Systems/Monitoring-Programs/Medicare-FFS-Compliance-Programs/CERT/CERT-Reports-Items/Nov2011Appendix.html>
22. Song Y, Skinner J, Bynum J, *et al*. Regional variations in diagnostic practices. *N Engl J Med* 2010;363:45–53.



23. Bezzina AJ, Smith PB, Cromwell D, *et al*. Primary care patients in the emergency department: who are they? A review of the definition of the 'primary care patient' in the emergency department. *Emerg Med Australas* 2005;17:472–9.
24. Begley CE, Vojvodic RW, Seo M, *et al*. Emergency room use and access to primary care: evidence from Houston, Texas. *J Health Care Poor Underserved* 2006;17:610–24.
25. Kellermann AL, Weinick RM. Emergency departments, Medicaid costs, and access to primary care—understanding the link. *N Engl J Med* 2012;366:2141–3.
26. Weinick RM, Burns RM, Mehrotra A. Many emergency department visits could be managed at urgent care centers and retail clinics. *Health Aff (Millwood)* 2010;29:1630–6.
27. Pitts SR, Carrier ER, Rich EC, *et al*. Where Americans get acute care: increasingly, it's not at their doctor's office. *Health Aff (Millwood)* 2010;29:1620–9.
28. Lowe RA, Localio AR, Schwarz DF, *et al*. Association between primary care practice characteristics and emergency department use in a Medicaid managed care organization. *Med Care* 2005;43:792–800.