The PRONE score: an algorithm for predicting doctors’ risks of formal patient complaints using routinely collected administrative data

Matthew J Spittal,1 Marie M Bismark,1 David M Studdert2,3

ABSTRACT
Background Medico-legal agencies—such as malpractice insurers, medical boards and complaints bodies—are mostly passive regulators; they react to episodes of substandard care, rather than intervening to prevent them. At least part of the explanation for this reactive role lies in the widely recognised difficulty of making robust predictions about medico-legal risk at the individual clinician level. We aimed to develop a simple, reliable scoring system for predicting Australian doctors’ risks of becoming the subject of repeated patient complaints.

Methods Using routinely collected administrative data, we constructed a national sample of 13 849 formal complaints against 8424 doctors. The complaints were lodged by patients with state health service commissions in Australia over a 12-year period. We used multivariate logistic regression analysis to identify predictors of subsequent complaints, defined as another complaint occurring within 2 years of an index complaint. Model estimates were then used to derive a simple predictive algorithm, designed for application at the doctor level.

Results The PRONE (Predicted Risk Of New Event) score is a 22-point scoring system that indicates a doctor’s future complaint risk based on four variables: a doctor’s specialty and sex, the number of previous complaints and the time since the last complaint. The PRONE score performed well in predicting subsequent complaints, exhibiting strong validity and reliability and reasonable goodness of fit (c-statistic=0.70).

Conclusions The PRONE score appears to be a valid method for assessing individual doctors’ risks of attracting recurrent complaints. Regulators could harness such information to target quality improvement interventions, and prevent substandard care and patient dissatisfaction. The approach we describe should be replicable in other agencies that handle large numbers of patient complaints or malpractice claims.

INTRODUCTION
Medico-legal agencies—such as malpractice insurers, medical boards and complaints handling bodies—are essentially reactive regulators: they deal with the aftermath of care that has gone badly. This posture has confined them largely to the sidelines of the patient safety movement. Clinicians eye medico-legal processes with scepticism and fear, and sometimes with outright disdain. The ex post nature of these processes, coupled with their focus on provider fault, is derided as antithetical to quality improvement efforts focused on prevention and systemic causes of harm.

Part of the explanation for medico-legal agencies’ limited role lies in their inability to make reliable predictions about which clinicians will experience complaints or claims. The conventional wisdom is that these events cannot be predicted at the individual practitioner level with acceptable levels of accuracy. Numerous studies have tried,2–12 with limited success.

In a recent national study13 of formal patient complaints against Australian doctors lodged with state regulators, we found that 3% of doctors (or 18% of doctors with at least one complaint) accounted for half of all complaints lodged over a 10 year period. We proposed and tested a new method for identifying doctors at high risk of incurring repeated medico-legal events. Among doctors who had already incurred one or more complaints, we found that risks of incurring more complaints in the near
term varied dramatically—from less than 10% to more than 80%—depending on certain observable characteristics, or risk factors. However, one limitation of the method we deployed, recurrent event survival analysis, is its complexity. Few health regulators are likely to have the technical capacity to replicate this approach, much less incorporate it into their front-line case management activities. Relatedly, although survival analysis accounts well for changing baseline risk over time, it does not generally provide estimates of these risks that can be easily integrated into a risk algorithm.

In this study, we extended our analysis of patient complaints to create a simple predictive algorithm. Our immediate objective was to provide health complaints commissions in Australia with a tool that would allow them to reliably estimate practitioners’ future risk of complaints in ways that could support proactive intervention. More generally, we sought to demonstrate the feasibility of using routinely collected administrative data to construct a risk calculator for predicting medicolegal events.

**METHODS**

*Setting*

Health complaints commissions (Commissions) are statutory agencies established in each of Australia’s six states and two territories. Commissions have responsibility for receiving and resolving patient complaints about the quality of healthcare services. Patients or their advocates must initiate complaints in writing, but the process is free and legal representation is optional.14

Outside the clinic or hospital where care is rendered, Commissions are the primary avenue of redress for patients dissatisfied with the quality of care they have received. Plaintiffs’ lawyers in Australia will rarely take cases unless they have first proceeded through Commission processes (although the vast majority of complaints do not go on to become malpractice claims). At least 10 other Organisation for Economic Co-operation and Development countries—including Austria, Finland, Israel and New Zealand—have similar bodies.15 16 In the UK, the closest analogue is the Parliamentary and Health Service Ombudsman.

Commissions in all Australian states except South Australia participated in the study. In 2011, these seven jurisdictions had 21 million residents and 90% of the nation’s 70 200 practicing doctors.17

*Data*

Between May 2011 and February 2012 we collected data on-site at Commission offices in each participating state and territory. Complaints against doctors were identified by querying the Commissions’ administrative data systems. The filing period of interest spanned 12 years and differed slightly by jurisdiction: 2000–2011 for the Australian Capital Territory, the Northern Territory, Queensland, Tasmania and Victoria; 2000–2010 for Western Australia; and 2006–2011 for New South Wales. (Changes to the data system in New South Wales necessitated a narrower sampling window there.) As described elsewhere,13 we supplemented administrative data from the Commissions with additional doctor-level variables obtained from a database on Australian doctors held by AMPCo Direct, a subsidiary of the Australian Medical Association.

All data were combined in a complaint-level analytical data set, in which multiple complaints against a single doctor could be observed.

*Variables*

Our unit of analysis was the complaint. The primary outcome was the occurrence of a ‘subsequent’ complaint, defined as a complaint that occurred soon after a prior one. Specifically, a complaint against a doctor within 2 years of the most recent complaint against that doctor was coded ‘1’; a complaint that was not followed by another during the 2-year window, was coded ‘0’. To allow sufficient follow-up time, complaints that occurred within 24 months of the end of a jurisdiction’s data period (2011 in all but one jurisdiction) were included in the analysis as subsequent complaints but not as index complaints. We chose a 2-year window based on our discussions with regulators in Australia and New Zealand, who indicated that this is generally the time horizon of interest to them in their efforts to separate complaint-prone practitioners from respondents whose appearance reflects merely ‘baseline’ risks of being complained against.

The set of predictors we examined consisted of variables measured at the doctor level (specialty, age, sex, location of practice) and the complaint level (number of previous complaints, time since previous complaint, complaint issue). Doctor age and all complaint level variables were coded as time-varying variables, meaning their values related to the most recent complaint. Specialty was classified into 13 categories, based on those promulgated by the Medical Board of Australia.18 Doctor age was coded as <35 years or 35–65 years. (Doctors older than 65 years were excluded because of possible bias due to unobserved censoring: an absence of further complaints may have been due to retirement). Principal practice address was classified as urban or rural, based on the location of its postcode within a standard geographical classification system.19 Subsequent complaints were categorised by time intervals (<6 months after the most recent complaint, 6 months to <1 year, 1–2 years).

Complaint issue in the most recent complaint was coded into two categories: complaints in which the primary issue related to clinical care (treatment; diagnosis; medication; hygiene or infection control; discharge or transfer; other clinical care issues) and complaints with other primary issues (communication;
costs or billing, medical records, certificates or reports; access and timelines, sexual contact or relationship; rough or painful treatment; confidentiality or information privacy; breach of conditions; grievance handling; discrimination; other issues).

Statistical model
We used multivariate logistic regression analysis to estimate odds of a subsequent complaint within 2 years. The predictors were number of previous complaints; time since previous complaint; complaint issue; and specialty, age, sex and practice location of the doctor against whom the complaint was made. Cluster-adjusted robust SEs were calculated to account for multiple observations pertaining to single doctors.

We estimated and compared three models: Model 1 used all available predictors; Model 2 focused on a subset of predictors that tend to be the most straightforward for regulators to obtain from routine operational data; and Model 3 had only one predictor, the variable most strongly associated with the outcome. Our goal was to identify the model that best balanced model fit against parsimony. The primary basis for comparing the models’ performance was the c-statistic (or area under the receiver operating characteristic (ROC) curve). We computed c-statistics, adjusting them for optimism to guard against overfitting using a bootstrap sampling approach. We computed the adjusted c-statistic for Model 1 was significantly different from Model 2, and whether Model 2 was significantly different from Model 3. Finally, as a sensitivity analysis, we repeated our analysis using only complaints over clinical care.

Sample characteristics
Table 1 describes characteristics of the doctors and complaints in our study sample. Thirty per cent of the complaints addressed clinical aspects of care, most commonly concerns with treatment (39%), diagnosis (16%) and medications (8%). About a fifth of complaints addressed communication issues, including concerns with the attitude or manner of doctors (13%), and the quality or amount of information provided (6%).

Nearly half of the doctors complained against were general practitioners and 15% were surgeons. Seventy-nine per cent were male and 80% were 35–65 years of age. On average, 398 days (SD 497 days) elapsed between the index and subsequent complaints. There was a trivial amount of missing data (<1%) for all variables except age (14%).

Choice of multivariate prediction model
All of the variables considered in Model 1, except the variable indicating the previous complaint related to clinical care, were statistically significant predictors of subsequent complaints, and this model had a c-statistic of 0.69 after adjusting for optimism (table 2). Model 2 dropped doctor age, the clinical care variable and practice location (a variable many regulators do not routinely collect). All four remaining variables were
statistically significant predictors of subsequent complaints, and discrimination in this model (c-statistic=0.70) was very close to Model 1’s, although there was a statistically significant difference between the c-statistics for the two models (p=0.018).

Model 3 consisted only of the number of prior complaints variable, which was clearly the strongest predictor of subsequent complaints in models 1 and 2. Although this predictor on its own showed reasonable discrimination (c-statistic=0.66), there was strong evidence that its fit was inferior to the other two models (p<0.0001 in both comparisons).

Therefore, considering parsimony, discriminative ability and potential practicality, we selected Model 2 as the basis for the PRONE score algorithm. The far right column in Table 2 shows the PRONE score points assigned to each predictor, based on the corresponding coefficients from Model 2. Other commonly used approaches—for instance, assigning a point for every 100% increase in ORs—produced similar results (data not shown).

To test the robustness of our results, we refit our models using only those complaints that related to clinical care. The c-statistics for these three models were 0.69, 0.70 and 0.65, respectively, and the ORs were similar to those reported above (see Models 4, 5 and 6 in online supplementary appendix table S1).

Performance of PRONE score
A comparison of ROC curves for the PRONE score and the coefficients estimated in Model 2 showed almost perfect overlap, suggesting little if any discrimination was lost in transforming the model’s coefficients to the scoring system (see online supplementary appendix figure S1).

Figure 1 shows, at each level of the PRONE score, the number of complaints predicted by the score alongside the number actually observed. The Hosmer-Lemeshaw χ² statistic was non-significant, χ²(20)=21.83, p=0.35, suggesting that PRONE scores corresponded closely to actual risk. The distribution of observed and expected complaints by PRONE score indicates some skewness, with scores centred around values of 4 and 5.

PRONE score predictions
Table 3 shows the frequency and risk of subsequent complaints within seven PRONE score groups. Risk increased monotonically with PRONE score. For example, among 2000 doctors whose index complaint scored between 0 and 2, 285 of them actually had a subsequent complaint, a 2-year risk of 14.2%. Among 221 doctors with PRONE scores between 15 and 17, 194 had subsequent complaints, a risk of 87.8%.

Thresholds for intervention
Decisions regarding suitable score thresholds for intervention involve a trade-off between sensitivity and

---

Table 1  Characteristics of complaints and doctors who were the subject of the complaints

<table>
<thead>
<tr>
<th>Issue in complaints, n=13 849</th>
<th>n*</th>
<th>Per cent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical care</td>
<td>8352</td>
<td>60</td>
</tr>
<tr>
<td>Treatment</td>
<td>5407</td>
<td>39</td>
</tr>
<tr>
<td>Diagnosis</td>
<td>2251</td>
<td>16</td>
</tr>
<tr>
<td>Medication</td>
<td>1083</td>
<td>8</td>
</tr>
<tr>
<td>Hygiene/infection control</td>
<td>104</td>
<td>0.8</td>
</tr>
<tr>
<td>Discharge/transfer</td>
<td>53</td>
<td>0.4</td>
</tr>
<tr>
<td>Other clinical care</td>
<td>81</td>
<td>0.6</td>
</tr>
<tr>
<td>Communication</td>
<td>2909</td>
<td>21</td>
</tr>
<tr>
<td>Attitude or manner</td>
<td>1849</td>
<td>13</td>
</tr>
<tr>
<td>Information</td>
<td>790</td>
<td>6</td>
</tr>
<tr>
<td>Consent</td>
<td>416</td>
<td>3</td>
</tr>
<tr>
<td>Other communication</td>
<td>30</td>
<td>0.2</td>
</tr>
<tr>
<td>Costs or billing</td>
<td>970</td>
<td>7</td>
</tr>
<tr>
<td>Medical records, certificates or reports</td>
<td>891</td>
<td>6</td>
</tr>
<tr>
<td>Access and timeliness</td>
<td>854</td>
<td>6</td>
</tr>
<tr>
<td>Sexual contact or relationship</td>
<td>422</td>
<td>3</td>
</tr>
<tr>
<td>Rough or painful treatment</td>
<td>319</td>
<td>2</td>
</tr>
<tr>
<td>Confidentiality or information privacy</td>
<td>281</td>
<td>2</td>
</tr>
<tr>
<td>Breach of conditions</td>
<td>186</td>
<td>1</td>
</tr>
<tr>
<td>Grievance handling</td>
<td>129</td>
<td>0.9</td>
</tr>
<tr>
<td>Discrimination</td>
<td>54</td>
<td>0.4</td>
</tr>
<tr>
<td>Other</td>
<td>112</td>
<td>0.8</td>
</tr>
</tbody>
</table>

Characteristics of doctors, n=8424

<table>
<thead>
<tr>
<th>Gender</th>
<th>n*</th>
<th>Per cent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>6667</td>
<td>79</td>
</tr>
<tr>
<td>Female</td>
<td>1676</td>
<td>20</td>
</tr>
<tr>
<td>Missing</td>
<td>71</td>
<td>0.8</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22–34 years</td>
<td>464</td>
<td>6</td>
</tr>
<tr>
<td>35–65 years</td>
<td>6756</td>
<td>80</td>
</tr>
<tr>
<td>Missing</td>
<td>1204</td>
<td>14</td>
</tr>
<tr>
<td>Specialty</td>
<td></td>
<td></td>
</tr>
<tr>
<td>General practice</td>
<td>3972</td>
<td>49</td>
</tr>
<tr>
<td>Surgery</td>
<td>1182</td>
<td>15</td>
</tr>
<tr>
<td>Orthopaedic surgery</td>
<td>329</td>
<td>4</td>
</tr>
<tr>
<td>General surgery</td>
<td>296</td>
<td>4</td>
</tr>
<tr>
<td>Plastic surgery</td>
<td>140</td>
<td>2</td>
</tr>
<tr>
<td>Other surgery</td>
<td>417</td>
<td>5</td>
</tr>
<tr>
<td>Internal medicine</td>
<td>934</td>
<td>11</td>
</tr>
<tr>
<td>Obstetrics and gynaecology</td>
<td>416</td>
<td>5</td>
</tr>
<tr>
<td>Psychiatry</td>
<td>504</td>
<td>6</td>
</tr>
<tr>
<td>Anaesthesia</td>
<td>314</td>
<td>4</td>
</tr>
<tr>
<td>Ophthalmology</td>
<td>188</td>
<td>2</td>
</tr>
<tr>
<td>Radiology</td>
<td>144</td>
<td>2</td>
</tr>
<tr>
<td>Dermatology</td>
<td>129</td>
<td>2</td>
</tr>
<tr>
<td>Other specialties</td>
<td>349</td>
<td>4</td>
</tr>
<tr>
<td>Location of practice</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rural</td>
<td>1948</td>
<td>23</td>
</tr>
<tr>
<td>Urban</td>
<td>6378</td>
<td>77</td>
</tr>
</tbody>
</table>

*Complaint issues sum to more than 100% because some complaints involved multiple issues.
specificity, and are likely to be influenced by the effectiveness, intrusiveness and cost of the intervention. While the exact nature of such interventions is beyond the scope of this paper, we use three hypothetical examples to illustrate the nature of these trade-offs. (We note that some of these interventions may fall outside current statutory powers of Australian Commissioners.)

Table 3 depicts three possible interventions: (1) advising doctors in writing that they are at risk of a future complaint; (2) compelling doctors to undertake a continuing medical education course on a topic that addresses issues commonly arising in their complaint profile; and (3) referral to a regulator (eg, medical board) for further action. For a low cost, relatively unintrusive intervention, such as an informational

<table>
<thead>
<tr>
<th>Complaint number</th>
<th>Model 1 OR (95% CI)</th>
<th>Model 2 OR (95% CI)</th>
<th>Model 3 OR (95% CI)</th>
<th>PRONE score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (ref)</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1.29 (1.11 to 1.49)</td>
<td>1.35 (1.20 to 1.51)</td>
<td>1.82 (1.66 to 2.00)</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>1.85 (1.56 to 2.20)</td>
<td>1.91 (1.65 to 2.22)</td>
<td>2.76 (2.43 to 3.14)</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>2.48 (2.01 to 3.07)</td>
<td>2.64 (2.18 to 3.20)</td>
<td>3.98 (3.35 to 4.72)</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>3.29 (2.51 to 4.31)</td>
<td>3.41 (2.67 to 4.35)</td>
<td>5.36 (4.27 to 6.73)</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>4.35 (3.11 to 6.10)</td>
<td>4.30 (3.15 to 5.87)</td>
<td>6.88 (5.11 to 9.25)</td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>4.76 (3.08 to 7.34)</td>
<td>5.01 (3.35 to 7.49)</td>
<td>8.51 (5.85 to 12.4)</td>
<td>6</td>
</tr>
<tr>
<td>8</td>
<td>4.44 (2.78 to 7.08)</td>
<td>4.79 (3.08 to 7.43)</td>
<td>7.98 (5.23 to 12.2)</td>
<td>6</td>
</tr>
<tr>
<td>9</td>
<td>6.51 (3.38 to 12.53)</td>
<td>6.73 (3.68 to 12.3)</td>
<td>11.1 (6.19 to 19.8)</td>
<td>7</td>
</tr>
<tr>
<td>10+</td>
<td>18.89 (9.76 to 36.56)</td>
<td>18.3 (10.2 to 32.8)</td>
<td>33.8 (19.1 to 59.7)</td>
<td>11</td>
</tr>
</tbody>
</table>

**Doctor’s specialty**

- **Anaesthesia (ref)**
  - 1.00
- **Radiology**
  - 1.00 (0.47 to 2.12)
- **Other specialties**
  - 0.97 (0.63 to 1.49)
- **Internal medicine**
  - 1.40 (1.04 to 1.88)
- **Ophthalmology**
  - 1.36 (0.94 to 1.96)
- **General practice**
  - 1.61 (1.23 to 2.10)
- **Psychiatry**
  - 1.94 (1.44 to 2.62)
- **Orthopaedic surgery**
  - 2.02 (1.49 to 2.74)
- **Other surgery**
  - 2.11 (1.56 to 2.86)
- **General surgery**
  - 2.11 (1.51 to 2.95)
- **Obstetrics and gynaecology**
  - 2.36 (1.73 to 3.23)
- **Dermatology**
  - 2.73 (1.89 to 3.96)
- **Plastic surgery**
  - 3.98 (2.84 to 5.57)

**Time since previous complaint**

- **1–2 years (ref)**
  - 1.00
- **6 months to 1 year**
  - 1.23 (1.04 to 1.47)
- **Less than 6 months**
  - 1.68 (1.44 to 1.95)

**Doctor’s sex**

- **Female (ref)**
  - 1.00
- **Male**
  - 1.45 (1.27 to 1.66)

**Doctor’s age**

- **22–34 years (ref)**
  - 1.00
- **35–65 years**
  - 1.41 (1.10 to 1.82)

**Location of practice**

- **Urban (ref)**
  - 1.00
- **Rural**
  - 1.12 (1.01 to 1.24)

**Complaint issue**

- **Other issue (ref.)**
  - 1.00
- **Clinical care**
  - 1.02 (0.9 to 1.16)

**C-statistic (adjusted for optimism)**

- 0.69
- 0.70
- 0.66
Predicted Risk Of New Event (PRONE) score groups


Figure 1 Calibration curves for 22-point PRONE (Predicted Risk Of New Event) score.

letter, a reasonably low threshold for intervention is appropriate. This will ensure that most doctors at risk of a subsequent complaint will receive the intervention (high sensitivity), although many letter recipients will not actually go on to incur further complaints (low specificity). A cut point of 3 or more on the PRONE score, which has sensitivity of 94% and a specificity of 19%, may be a reasonable threshold for this type of intervention.

A much sterner intervention, such as referral to a medical board for assessment and possible further action, would impose a significant burden on affected doctors, and would also generate substantial costs for regulators. For such interventions, there should be high confidence that the practitioner will in fact incur additional complaints in the near term (high specificity), even though setting tolerances in that way will mean that a non-trivial number of doctors who will incur additional complaints will fall below the threshold (low sensitivity). A PRONE score of 12 or more has specificity of 99% and sensitivity of 13%.

Requiring doctors at relatively high risk of additional complaints to undertake a relevant continuing medical education course is an example of intervention that is moderately intrusive. The PRONE score threshold triggering such an intervention should seek to strike a balance between sensitivity and specificity. A cut point of 5 or more, for example, would achieve sensitivity of 69% and specificity of 58%.

DISCUSSION

The PRONE score exhibited strong predictive properties and has considerable potential as a tool for determining the likelihood that doctors named in complaints will reappear as the subjects of additional complaints in the near future. The precise contours of the algorithm we report map directly to complaints lodged with health service commissions in Australia, not other settings. However, the evidence presented of the feasibility and potential value of this approach has wider implications. It is ripe for testing and possible replication in other places and agencies, including liability insurers and medical boards.

Previous attempts4–8 to predict medicolegal risk have substantial limitations and, to the best of our knowledge, none, with the exception of the Patient at Risk Score (PARS) score developed by Hickson et al,8 has attempted to convert findings into an user-friendly scoring system to guide interventions. An impressive feature of the PARS system is that it links scores to a suite of doctor-focused interventions aimed at preventing recurrence. However, the PARS predictive algorithm is based on patient complaints accumulated in a handful of hospitals, not a population-wide sample of complaints. Three of the four variables used in PARS are similar to PRONE—namely, doctor’s sex, specialty group (surgeons vs non-surgeons) and complaint count. PARS also has a variable for clinical activity, which PRONE does not have. Beyond that, it is very difficult to compare PRONE with PARS because details of the structure and performance of the identification algorithm used in PARS have not been published.

This study extends a recent analysis of patient complaints against Australian doctors. Our earlier analysis showed that the incidence of complaints was extremely skewed, and ‘frequent flyers’ accounted for a very substantial proportion of all complaints; it also demonstrated the feasibility of identifying those frequent flyers early in the trajectory of their complaints profile. This study takes an important next step: converting such predictive modelling into a simple risk scoring system that is amenable to routine use by regulators.

Our approach differs from previous attempts to predict medicolegal risk in two main respects: inclusion of time-varying predictors (such as a continuously changing measure of prior events) and a focus on practitioners who have already experienced at least one complaint. These distinctive aspects of our analytical design boost our ability to make stronger risk predictions than earlier studies have.

How might the PRONE score be used in practice? We envision two ways it could be incorporated into

Table 3 Frequency and risk of complaint within 2 years, by Predicted Risk Of New Event (PRONE) score groups

<table>
<thead>
<tr>
<th>PRONE Groups</th>
<th>Number of doctors in group</th>
<th>Total complaints in group</th>
<th>Risk of subsequent complaint (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–2</td>
<td>2000</td>
<td>285</td>
<td>14.2 (12.7 to 15.9)</td>
</tr>
<tr>
<td>3–5</td>
<td>7144</td>
<td>1868</td>
<td>26.1 (25.1 to 27.2)</td>
</tr>
<tr>
<td>6–8</td>
<td>2474</td>
<td>1034</td>
<td>41.8 (39.8 to 43.8)</td>
</tr>
<tr>
<td>9–11</td>
<td>1057</td>
<td>620</td>
<td>58.7 (55.6 to 61.6)</td>
</tr>
<tr>
<td>12–14</td>
<td>353</td>
<td>263</td>
<td>74.5 (69.6 to 79.0)</td>
</tr>
<tr>
<td>15–17</td>
<td>221</td>
<td>194</td>
<td>87.8 (82.7 to 91.8)</td>
</tr>
<tr>
<td>18–21</td>
<td>149</td>
<td>137</td>
<td>91.9 (86.4 to 95.8)</td>
</tr>
</tbody>
</table>
the complaint-handling process. First, because the score can be recalculated each time a new complaint is lodged, it could be useful for ‘red-flagging’ cases for a deeper file review—for example, reviewing all previous complaints against a practitioner to ascertain if there are any troubling patterns. Second, the score could be useful for ‘tiering’ interventions. For example, a low PRONE score may suggest that minimal action is required beyond resolution of the immediate complaint, while a high score may prompt a regulator to consider whether more active intervention is needed to guard against the risk of future harm.

The strengths of the PRONE score are that it is simple and relies on information that most complaint-handling or claim-handling bodies collect routinely. Its predictive properties, based on the risk of a subsequent complaint within each score band, are fairly good. Although the c-statistic (0.70) for the multivariate model from which the scoring system is derived indicates only moderate discrimination, our various tests of PRONE score performance were encouraging; in particular, we observed few ‘false positives’ among doctors who score in the upper reaches of the scale (>15), which is where the scoring system performs best.

However, our study also has several limitations. First, factors others than those we considered predict complaints. For example, patient and doctor characteristics, aspects of the doctor-patient relationship, and the system in which the doctor works are all known to affect patient dissatisfaction and subsequent complaining and claiming behaviour. However, the difficulty regulators face in measuring such factors across an entire case load makes them poor candidates for a risk calculator intended for routine use.

Nonetheless, recognition that unobserved variables may influence a practitioner’s complaint risk is crucial at the point of intervention. Environmental or system-related factors—such as solo practice, poor information systems or especially challenging patient populations—are particularly powerful examples of confounding factors. The relationship between individual-related and system-related causes of poor quality care is complex. But the coexistence of observed individual factors and unobserved systemic factors does not negate the value of the PRONE score. Even in situations where a practitioner’s outlier status is explained primarily by systemic causes, the score may be an effective way of spotlighting practice environments of concern. Thus, the score is best understood as a method of identifying practitioners whose behaviour and event history warrants special attention and further investigation, rather than a direct determinant of the type of action needed to improve quality.

Third, risk factors acting in concert (ie, interaction effects) may increase or decrease risk of subsequent complaints. Inclusion of interactions may increase the score’s predictive power, but at the expense of simplicity, which could act as a barrier to adoption by regulators. In exploratory analyses we found only one significant interaction (between complaint number and specialty). It was a poor candidate for inclusion in the score however, because it would substantially increase the number of parameters without improving discrimination.

Third, we used head counts of practitioners, not more sophisticated measures of doctors’ exposure to complaint risk, such as volume of patients treated or procedures conducted. Fourth, we used logistic regression to develop our predictive model. This approach allowed us to directly estimate time parameters (eg, time since last complaint). Since the risk of a new complaint decreases over time, this allowed the resultant PRONE score to increase, decrease or remain unchanged at each new calculation of a doctor’s score. Disadvantages of this approach are that it does not fully account for changing baseline risk and the subjectivity associated with choice of time cut points. Survival analysis, which we used in our previous analysis of complaints, deals better with these issues, but has its own disadvantages: it cannot handle time-related predictor variables, risk scores can only increase, and it is a difficult approach to explain to regulators.

Fifth, although our methodology should be generalisable to other medicolegal settings, the extent to which it actually is is unknown. The potential gains from replication in other settings depend on three key factors: (1) the proportion of all events attributable to multievent practitioners; (2) the capacity of available predictors to reliably estimate those events; and (3) the size of the reference population. If the proportion of all events attributable to multievent practitioners is small, or data catchment relates to a relatively small population (eg, a single hospital), this would undermine the usefulness and feasibility of risk prediction. Health regulators and liability insurers interested in developing an approach like the PRONE score should be attentive to these factors.

Finally, even if the PRONE score were to be adopted and deployed effectively, it is insufficient, on its own, to improve the quality and safety of care. It is merely the ‘front end’ of a quality improvement strategy. The scoring system must connect to interventions that work. Exactly what those interventions are, and how they interlock with the scoring system, lie beyond the scope of this study, although table 4 exemplified the types of approaches that regulators are likely to find attractive.

A risk calculator, like the PRONE score, could be deployed retrospectively or prospectively. As part of a general case load review, such an algorithm could be applied to identify practitioners at highest risk of further events and in need of prompt intervention. Another approach would be to incorporate the tool
into day-to-day handling of complaints or claims, giving regulators an ability to observe ascending levels of risk and tailor responses accordingly. The potential for prospective use is particularly novel and exciting because it holds the promise of ushering medicolegal agencies into the prevention business. However, such uses would inevitably raise ethical and legal challenges. The best way to deflect those challenges may be to ensure that, in any attempt to make levels of predicted risk trigger points for intervention, the intrusiveness of the intervention is well matched to the confidence of the prediction.

Twitter Follow Marie Bismark at @mbismark

Acknowledgements The authors thank the health services commissioners and their staff in the Australian Capital Territory, New South Wales, the Northern Territory, Queensland, South Australia, Tasmania, Victoria and Western Australia; their support, assistance and advice made this study possible.

Contributors MJS, MB, and DMS devised the study. MJS and MB collected the data. MJS undertook the analysis and wrote the first draft. MB and DMS contributed to the interpretation of the results and manuscript revision.

Funding The study was funded by an ARC Laureate Fellowship (FL110100102 to DMS) from the Australian Research Council. The research was conducted independently from the funder.

Competing interests None.

Ethics approval Human Research Ethics Committee, University of Melbourne.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement The study data are owned by state health complaint commissions. Permission was granted to use it for purposes of the study, but not for any other use.

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
21 StataCorp. *Stata: Release 13.1*. College Station, TX: StataCorp LP 2013.
Author/s: Spittal, MJ; Bismark, MM; Studdert, DM

Title: The PRONE score: an algorithm for predicting doctors' risks of formal patient complaints using routinely collected administrative data

Date: 2015-06-01

Citation: Spittal, MJ; Bismark, MM; Studdert, DM, The PRONE score: an algorithm for predicting doctors' risks of formal patient complaints using routinely collected administrative data, BMJ QUALITY & SAFETY, 2015, 24 (6), pp. 360 - 368

Persistent Link: http://hdl.handle.net/11343/55302