- その他
  - 他の病院団体とのかからみはあるか。
  - 今は公表している指標や方法などで統一性がないので課題になっていくと思う。

  （CI の要望や病院名の略称、公表に関する広報への意見は出なかった）

- 患者満足度・職員意識調査の説明と質疑応答（今中）

以上

QIP ワークショップ「臨床指標の公表」その 2：議論概要

2010 年 11 月 27 日

2010 年 11 月 27 日（土）　10：30〜12：30　京都大学東京オフィス

参加者：

QIP 参加病院関係者、スタッフ：今中（説明）、猪飼（説明）、田中、宇川、國澤（書記）

内容抄録

・資料に沿って、今中、猪飼より説明

指標 1）2）

→ 当院での高齢化が進んでおり、平均年齢や、標準偏差など付け加えられないか。グラフでなくとも、表の方などに。高齢者の割合など。

→ **指標の項目によっては付けているものもある。今後の検討課題の一つとする。**

→ 資料自体はどのようなものを入手されているのか。退院処方以外の入院処方もわかるのか。

→ **様式 1 と EF ファイルを説明。**

→ 転機は、死亡や短時間の入院も分母に含めているのか。

→ **短期間でも使われていると考えている処方なので含めている。**

→ 木村先生が、βブロッカーの、心機能如何での比較トライアルを計画されている。日本人では必ずしも必要ない薬剤かもしれないが。

→ **今後、心機能によって層別化する必要があるかもしれない。**

→ βブロッカー、ARB などは、両方とも使うことは少なく、どちらか一方だけとかでだしていないか。

→ **海外のガイドラインと日本のエビデンスとの相違は念頭に入れなくてはならない。いずれにせよ、どちらかというのは検討予定とする。**

指標 6）

→ 選出コードはこれだけでよいか。

→ **脳梗塞後遺症などほかのコードは別にある。**

→ なにか、このコードだけでは大雑把で不十分な気がする。急性期の他にもフラグができるとよいのではないか。エダラボン？や手術など。

→ **薬剤は、その使用頻度自体にばらつきがみられ頻度も低く、症例が少なくなる。絞り込みは難しい。**

指標 7）

→ 開頭も含むのか。開頭はほとんどしない。穿頭のみがほとんど。穿頭だけでも良いと思う。

→ 亜急性になると、患者の状態がばらつき、患者の質が不均質と思える。

→ **検討課題とする。**

指標 18）

→ 公表する意義がわからない。患者に公表する意味がない。

→ 率の高いところは保険病名では。施設ごとのチェックが必要。

→ T コードにカテーテル敗血症がある。それらを区別できないと、他の原因での敗血症を拾っている数値になるかもしれない。それらが混ざると、全体の信頼性が低下する。

→ **DPC からの測定結果と、その他の結果との比較検討など、他の意見を受ける機会にもしたい。学会で発表する方がふさわしいのかもしれない。他に、公表することで、病名をつける作業の質が上がる可能性を期待したい。**

指標 19）

→ 分母に悪性腫瘍以外を含んでいるの（外傷とか）のはおかしい。

→ 退院後に良悪がわかることがあり、データに反映されない。

→ たとえばがんセンターと一般病院では、様子が違って当然だろう

→ **C コード、D(上皮内がん)、K で限定したほうがよいかもしれないので、検討する。**

指標 20）

→ 緩和ケアはのぞいているのか。除かないと、地域のすべての死亡を引き受けているなどという病院もある。

→ むしろ、指標というよりは、このように、いろいろな病院がこの QIP に参加しているということを示す結果になっているように思える。

→ 4 日以上の根拠は？

→ **48 時間以上という既存の指標に合わせている。**

指標 21、22）

→ 様式 1 の緊急入院は緊急入院加算がとれるもののみがそこに区分されているのではないか。

→ **2009 年は OK。2010 年 7 月から。定義表示を書き換える必要はあり。**

指標 23）

→ 良いと思う。

全般）

→ 公表の意義として「認知される」とはどのような意味なのか。

→ **公表していることが、責任と自信、透明化など、対外的にポジティブなメッセージになる。**

→ 現段階の項目はよいと思うが、今後増えるのであればそのたびにこのような検討が必要と思う。

→ 他のプロジェクトと比較して（もしくは混在していて）効率が悪いと思う。

→ **指標一つ一つに公表の承諾をもらっている。**

→ **それぞれのプロジェクトが異なる指標を異なる方法で出しているため、今後全体の調整が必要となることが予想される。年度内に他の団体との調整を検討したい。**

→ 指標選択の基準は？出しやすさなどか。

→ **今回は、エビデンスが明確で、プロセスを扱い、改善が可能なもの、そして算出しやすいものを選択している。重要度はあまり考慮されていない。**

→ 最終的には、死亡率とか合併症とか（アウトカム）を出すようになるのだろうが、逐一確認検討が必要である。また、公表している病院としていない病院が逆に目立ってしまうのではないか。

→ **目立つほど少なくないのが現状である。非公表とするデメリットは事実上ないと思われるが明確にはわからない。**

→ アウトカム指標（合併症など）は、リスク調整などをしたいと思うのだが、調整は不十分性を内在しているので当面は実名公表にはふさわしくないと考えている。

→ **アウトカムは匿名を基本として、フィードバック方法・出し方は、改善・改訂していきたい。**

→ **アウトカム指標など、一般人やマスコミなど本当はこちらが知りたいのだと思うのだが、数値に飛びつく傾向があるため、十分な説明も必要。**

→ 実名公表のデメリットとして、批判や攻撃を受けたり、ガイドラインは従うべきものではないのに、裁判などが絡むと、ガイドラインに従っていない程度の低い病院という印象をうけたりするのか。

→ **批判や攻撃については海外も含め文献事例で問題視されていないようだが今後も留意していく。院内や患者への説明責任を果たしていく役割は大きくなると思われる。**

→ グラフ（評価）のどのあたりの病院が同意しているのか。良いところだけとか、不均等になっていないか。

→ **全体的に分散していると思われるが、指標によって順位が違う。ある領域で低い値が**

出ている病院も、公表に同意されている。また、限定的な公表同意も一部である。

→ 病院内での意見では、指標によって公表したりしなかったりを分けるのはよくなさそう。ただ、外科などは、抗生剤の使用方法の独自のポリシーがあるらしい。

→ 当院は都心だが、婦人科の撤退があって指標が出せない。もっと普遍的な疾患（誤嚥性肺炎など）を使用した指標がほしい。そして、層別化などもしてもらいたい。

→ **QIP は層別化、リスク調整も従来より行ってきている。病院種別などもある。**

→ 各項目を継続していくのか、新しい項目を作るのか。

→ **今後も継続するつもりである。指標は増やし、既存のものも改訂はするが、毎年同じ項目を出して経年に比較ができるようにしていきたい。それらの数値が改善されていることを示すのも目標の一つである。改訂された数値は、過去にさかのぼって再計算する。また、指標は順次増やしていく。**

患者満足度等）

患者満足度・職員意識調査の説明と質疑応答

An outcome prediction model for adult intensive care

Takeshi Umegaki, Miho Sekimoto,
Kenshi Hayashida and Yuichi Imanaka

ABSTRACT

Objective: To develop a prediction model of 28-day mortality in adult intensive care units using administrative data.

Design, setting and participants: We obtained data from 33 ICUs in Japan on all adult patients discharged from ICUs in 2007. Three predictive models were developed using (i) the five variables of the Critical Care Outcome Prediction Equation (COPE) model (age, unplanned admission, mechanical ventilation, hospital category and primary diagnosis) (the C model); (ii) 11 variables, including the COPE variables and six additional variables (sex, reason for ICU entry, time between hospital admission and ICU entry, use of fresh frozen plasma or a platelet preparation, dialysis, and use of pressors/vasoconstrictors (the P^+ model); and (iii) ten of the 11 variables, excluding primary diagnosis (the P^- model). Data for 6758 patients were stratified at the hospital level and randomly divided into test and validation datasets. Using the test dataset, five, 10 or nine variables were subjected to multiple logistic regression analysis (sex was excluded [$P > 0.05$]).

Main outcome measure: Mortality at 28 days after the first ICU day.

Results: Areas under the Receiver Operating Characteristic curve (AUROCs) for the test dataset in the C, P^+ and P^- models were 0.84, 0.89 and 0.87, respectively. Predicted mortality for the validation dataset gave Hosmer–Lemeshow χ^2 values of 12.91 ($P = 0.12$), 10.76 ($P = 0.22$) and 13.52 ($P = 0.1$), respectively, and AUROCs of 0.84, 0.89 and 0.90, respectively.

Conclusions: Our P^- model is robust and does not depend on disease identification. This is an advantage, as errors can arise in coding of primary diagnoses. Our model may facilitate mortality prediction based on administrative data collected on ICU patients.

Crit Care Resusc 2010; 12: 96–103

Concerns about quality of care and patient safety have increased the importance of monitoring of intensive care units in health care organisations. Performance measures for intensive care have been developed in response to increased demands to improve the quality of care.[1-3] Most studies have included mortality as an indicator of outcome, but mortality has varied between ICUs because of differences in the nature and severity of illness.[4-6] Several ICU risk-adjustment models[1] have been developed to compare mortality between institutions, including the Acute Physiology and Chronic Health Evaluation (APACHE) score, the Mortality Prediction Model (MPM), and the Simplified Acute Physiology Score (SAPS). Render and colleagues have proposed an automated ICU risk-adjustment tool.[7]

Severity scores have been constructed from demographics, physiological data and clinical diagnosis, and their validity has been confirmed in large-scale studies.[8-10] However, it is difficult to compare mortality rates between different ICUs based on the data available. Recently, the Critical Care Outcome Prediction Equation (COPE) model was proposed as a hospital mortality prediction model using administrative data. This model was constructed using five variables (age, unplanned admission, mechanical ventilation, hospital category and primary diagnosis). It showed that mortality could be well predicted from this model (area under the Receiver Operating Characteristic curve [AUROC] = 0.83–0.84).[11] Administrative data have the advantage of being available in a standardised format, which facilitates data collection from a large population and enables large-scale studies.

The Diagnosis Procedure Combination (DPC) system in Japan was introduced in 2004 and has become the standard method used in the health care financial system. Administrative data in this system include records of patient information and daily medical care. From these data, the types of all tests, medications and procedures and the use of intensive or special care and nursing services can be itemised on a daily basis. Procedures such as mechanical ventilation, renal replacement therapy and the use of vasoactive drugs are closely associated with mortality,[11-16] and their use varies somewhat among intensivists. Therefore, data on these interventions may help to predict mortality.

We used administrative data to develop three 28-day mortality prediction models based on:

- the five variables used in the COPE model (the C model);
- 11 variables: the five COPE model variables and six additional variables (sex, reason for ICU entry, time between hospital admission and ICU entry, use of fresh frozen plasma or a platelet preparation, dialysis, and use of pressors/vasoconstrictors) (the P^+ model); and

Table 1. Candidate variables used to develop the 28-day mortality prediction model

Type	Candidate variables	Category
Demographics	(1) Sex	Male; female
	(2) Age (years)	Continuous variable
Clinical factors	(3) Hospital admission	Scheduled;* emergency
	(4) Hospital category†	Metropolitan = 1; tertiary or regional = 0
	(5) Reason for entering ICU	After scheduled surgery;* after emergency surgery; internal medical disease
	(6) Primary diagnosis on admission	(See Table 2)
Any time during ICU admission	(7) Time between admission and ICU entry (days)	Direct; after 1 day; after 2–4 days;* after > 4 days
	(8) Use of fresh frozen plasma or platelet preparation	Yes = 1; No = 0
	(9) Mechanical ventilation	Used ⩾ 5 hours; used < 5 hours; not used*
	(10) Dialysis	Yes = 1, No = 0
	(11) Pressor/vasoconstrictor	Yes = 1, No = 0

ICU = intensive care unit. * Reference value. † Hospital category for the present study was assumed to be metropolitan.

- ten of the 11 variables (excluding primary diagnosis) (the P^- model).

The aim of our study was to compare the predictive value of the P^- model with that of the models that included primary diagnosis as a variable.

Methods

Data sources and case selection criteria

All data for the study were extracted from the Quality Indicator/Improvement Project (QIP), which collects administrative data and analyses numerical indices of health care process outcomes in Japan. Of the hospitals that voluntarily participate in the QIP, we included 33 acute-care hospitals with ICUs, including surgical ICUs, medical ICUs, and surgical–medical ICUs. These hospitals were relatively large urban teaching hospitals, functioning in a similar manner in provision of cardiac surgery and neurosurgery. The database used in the analysis included all patients aged 20 years or over treated in an ICU at one of the 33 hospitals and discharged between 1 January and 31 December 2007. We were able to identify the time of ICU entry and the dates for the ICU stay based on specific codes in the administrative data. Patients with cardiovascular disease as a primary diagnosis (regardless of internal medical disease) and those who had undergone cardiovascular surgery were excluded from the study, as they were cared for primarily in cardiovascular care units. The data did not indicate whether a patient had been previously hospitalised in another ICU. However, as critical care patients are rarely transferred from one centre to another in Japan, we assumed that patients entering the ICU had not been transferred from another ICU.

Development of the prediction model and potential risk factors

Data for 6758 patients were stratified at the hospital level and randomly divided into test and validation datasets. Using the test dataset, five, 10 or nine variables were subjected to multiple logistic regression analysis (sex was excluded, as it was not significantly associated with mortality in the univariate analysis). Hospitals were stratified based on the number of beds, and hospitals were paired based on a similar number of beds. Test and validation datasets were established that contained similar numbers of hospitals, hospitals of similar sizes, and similar numbers of patients. The primary measure was defined as outcome 28 days after the first ICU day. A survivor who was discharged from hospital within 28 days was defined as a survivor at 28 days after the first ICU day. The mortality prediction model was constructed using the test dataset and evaluated using the validation dataset. Coefficients obtained from the test dataset were applied to cases in the validation dataset to calculate the predicted mortality.

Model development was based on up to 10 variables (Table 1), including the five variables in the COPE model.[11] Age was defined as a continuous variable. In defining the reason for ICU entry, patients who underwent surgery on the first ICU day or earlier were considered to be surgical patients. Among surgical patients, those who underwent surgery on the day of hospital admission or the following day were defined as "emergency" surgery cases, and those who did not have emergency surgery were defined as "scheduled" surgery cases. All other patients were considered to be internal medical patients. To define admission categories, items in the administrative data pertaining to the course of admission were used. The "emergency"

admission category indicates hospital admission after transport by ambulance or an unexpected admission. For the time between admission and ICU entry (in days), we referred to the Project IMPACT study.[7] As Japan is a comparatively small country and development of access to hospitals has occurred through medical care policy, the hospital category (as defined in the COPE model) was assumed to be metropolitan. As International classification of diseases, 10th revision (ICD-10) codes rather than ICD-10-AM (Australian modification) codes are used in Japan, we translated ICD-10-AM codes into their nearest equivalent ICD-10 codes (Table 2).

In addition to mechanical ventilation, which is included in the COPE model, dialysis, pressors/vasoconstrictors, and use of fresh frozen plasma or a platelet preparation were considered as life-support interventions. These factors have been found to be significantly associated with prognosis in ICU patients.[12-16] Patients having mechanical ventilation were defined as those requiring the procedure for 5 or more hours after ICU entry. These patients were identified from the corresponding codes. Because the data distinguished between continuous (≥ 5 hours) and temporary (< 5 hours) mechanical ventilation, the patients were divided into two categories. Non-invasive positive pressure ventilation was excluded. Dialysis included continuous renal replacement therapy, intermittent renal replacement therapy, plasma absorption, and plasma exchange, but excluded peritoneal dialysis, as this is rarely used for ICU patients. Pressors/vasoconstrictors included dopamine, dobutamine, noradrenaline (norepinephrine), adrenaline (epinephrine) and vasopressin, but the use of adrenaline in cardiopulmonary resuscitation was excluded. We were unable to identify whether dopamine was given as a renal dose or for cardiovascular support, but as we found no evidence that low-dose (renal-dose) dopamine was used,[17] we assumed that dopamine was used for cardiopulmonary support.

Relationships between individual variables and 28-day mortality were analysed by a χ^2 test using the test dataset. After exclusion of variables with $P > 0.25$, the remaining variables were subjected to multiple logistic regression analyses (stepwise backward selection method). The model was constructed using variables with $P < 0.05$, and the AUROC was calculated.

Prediction model performance

Calibration of the model was evaluated using the Hosmer–Lemeshow χ^2 test. A well calibrated model has a low χ^2 value (< 15.5; df = 8) and a high P value (> 0.05). The discrimination of the model was assessed by the AUROC, for which a value of > 0.80 is favourable.

Table 2. Translation of ICD-10-AM codes to ICD-10 codes in our study

Diagnostic category	ICD-10-AM codes	ICD-10 codes
Anaemia	D5	D5
Aplastic anaemia	D6	D60–61
Bacterial sepsis	A4	A4
Breast cancer	C5	C5
Cardiac arrest	I46	I46
Cardiac arrhythmias	I49	I47–49
Cardiac failure	I22–25	I50
CNS malignancy	C69–72	C69–72
COPD	J40–44	J40–44
Drug poisoning	T36–50	T36–50
Enteritis or colitis	K50–52	K50–52
Environmental disease	T66–79	T66–78
Epilepsy	G4	G40
Fluid and electrolyte disorders	E86–88	E86–88
Fungal sepsis	B30–49	B35–49
GIT investigation	R1	R1
Haemopoietic malignancy	C80–99	C81–96
Haemorrhagic shock	R57–58	R57–58
Head injury	S0	S0
Interstitial lung disease	J8	J8
Intracranial haemorrhage	I60–62	I60–62
Ischaemic bowel	K55	K55
Liver disease	K7	K7
Lower limb trauma	S7	S7
Lung malignancy	C3	C3
Malabsorption	K9	K90
Malignancy – other	D37–49	D37–48
Myocardial ischaemia	I20	I20–25
Other cerebrovascular disease	I65–69	I65–69
Other CNS disease	G9	G9
Other intestinal disease	K63	K63
Pancreatic cancer	C22–26	C25
Penetrating trauma	T15–19	T15–19
Pneumoconiosis	J60–79	J60–70
Pneumonia	J1	J12–18
Protozoal sepsis	B50–64	B50–64
Pulmonary vascular disease	I26–28	I26–28
Renal failure	N1	N17–19
Respiratory failure	J95–99	J96
Secondary malignancy	C76–79	C76–80
Stroke or CVA	I63–64	I63–64
Type 2 diabetes	E11	E11

CNS = central nervous system. COPD = chronic obstructive pulmonary disease. CVA = cerebrovascular accident. GIT = gastrointestinal. ICD-10 = International classification of diseases, 10th revision. ICD-10-AM = ICD-10 (Australian modification).

Table 3. Demographic data and primary diagnosis for the test and validation datasets

	Test dataset ($n=3505$)	Validation dataset ($n=253$)	P
Number of hospitals	16	17	
Type of hospital			
Teaching hospital	16	17	
University hospital	0	0	
Non-teaching hospital	0	0	
Hospital with provision of cardiac surgery	15	16	0.742
Hospital with provision of neurosurgery	15	16	0.742
Mean number of beds	541.7 (SD, 189.3)	566.7 (SD, 258.4)	0.768
Mean number of ICU beds	7.4 (SD, 4.5)	7.8 (SD, 2.7)	0.802
Mean number of admissions (per year)*	10 767.9 (SD, 5199.7)	118 16.0 (SD, 6937.5)	0.688
Mean number of ICU admissions (per year)*	512.3 (SD, 317.6)	543.2 (SD, 279.6)	0.807
Mean length of stay (days)	13.6 (SD, 1.8)	13.9 (SD, 1.8)	0.721
Mean length of ICU stay (days)	3.6 (SD, 4.8)	4.4 (SD, 7.4)	< 0.001
Hospital mortality (%)	9.6%	13.7%	< 0.001
28-day mortality (%)	7.4%	9.7%	< 0.001
Primary diagnosis on admission	Frequency (%)	Frequency (%)	< 0.001
Infection	6.4%	6.6%	
Toxin	1.2%	0.9%	
Neoplastic†	2.2%	4.2%	
Metabolic†	1.2%	0.6%	
Haematological and immunological	0.8%	0.8%	
Gastrointestinal†	2.9%	2.2%	
Renal	1.5%	1.4%	
Respiratory†	5.9%	6.4%	
Neuromuscular†	1.1%	2.8%	
Other internal medical disease	1.0%	1.1%	
Cerebral surgery†	11.6%	16.7%	
Abdominal surgery†	38.4%	30.6%	
Lung or mediastinal surgery†	9.2%	12.0%	
Orthopaedic surgery†	7.4%	5.2%	
Other surgery†	9.1%	8.5%	

ICU = intensive care unit. * Estimated values (per year) are shown because the information that the Quality Indicator/Improvement Project received from each facility covered the period from 1 January to 31 December or until 30 June. † Significant difference between the two datasets by χ^2 test.

Prediction model validation

The three models were validated as follows. Cross-validation was performed using the validation dataset to demonstrate that the prediction equation obtained from multiple logistic regression analyses of the test dataset had predictive validity. Predicted mortality for the validation dataset was calculated using the coefficients we had derived from the test dataset. The performance of the equation was tested using the Hosmer–Lemeshow χ^2 statistic and the AUROC (95% CI). In the P^- model, a contingency table for different cut-off points was obtained for the validation dataset. Predicted mortality for internal medical disease, emergency surgery, scheduled surgery and sepsis was also examined. All statistical analyses were performed using SPSS software, version 11.0J (SPSS Inc., Chicago, IL, USA).

Comparisons among the three models

The C, P^+ and P^- models were compared using the Hosmer–Lemeshow χ^2 test and the AUROC (95% CI).

Results

Demographic data are summarised in Table 3. Explanatory variables did not differ significantly between the two

Table 4. Frequency and mortality based on individual variables in the test dataset (univariate analysis) ($n = 3505$)

Variable	Frequency (%)	28-day mortality (%)	P*
(1) Sex			0.489
Male	56.3	8.2	
Female	43.7	7.2	
(2) Age (years)			< 0.001
20–44	9.8	3.1	
45–54	9.0	4.1	
55–64	18.8	5	
65–74	26.0	7	
⩾ 75	36.3	10.8	
(3) Admission category			< 0.001
Scheduled	48.8	1.5	
Emergency	51.2	12.9	
(5) Reason for entering ICU			< 0.001
After scheduled surgery	46.4	4.1	
After emergency surgery	29.4	5.9	
Medical disease	24.3	15.4	
(7) Time from admission to ICU entry			< 0.001
Direct	30.5	12.7	
After 1 day	18.2	3.4	
After 2–4 days	25.3	2.8	
After > 4 days	25.9	8.3	
(8) Use of fresh frozen plasma or platelet preparation			< 0.001
Yes	9.2	14.5	
No	90.8	6.6	
(9) Mechanical ventilation			< 0.001
Used ⩾ 5 hours	14.4	25.6	
Used < 5 hours	19.8	5.6	
Not used	65.8	3.0	
(10) Dialysis			< 0.001
Yes	3.7	32.8	
No	96.3	6.4	
(11) Pressors/vasoconstrictors			< 0.001
Yes	41.3	13.3	
No	58.7	3.2	

ICU = intensive care unit. * P value for 28-day mortality.

datasets, except for length of ICU stay and primary diagnosis. Abdominal surgery was the most common type of surgery among surgical patients in both datasets. In patients with medical conditions, the most common reason for hospitalisation was infection. Among surgical patients, rates of cerebral, abdominal, lung, mediastinal and orthopaedic surgery differed significantly between datasets, and similarly, among medical patients, rates of neoplastic, metabolic, gastrointestinal, respiratory and neuromuscular disease differed significantly.

The overall 28-day mortality was 8.5%. In the univariate analysis (Table 4), the strongest association with mortality was found for dialysis (32.8%), followed by mechanical ventilation (⩾ 5 hours) (25.6%). Sex was not significantly associated with mortality ($P = 0.489$). Variables other than sex were subjected to multiple logistic regression analysis.

Coefficients of the variables, odds ratios (ORs), and the final equation for the validation dataset are shown in Table 5. Factors associated with a high risk of death in the C, P^+ and P^- models were haemopoietic malignancy (OR, 23.07 [95% CI, 4.91–108.44]); stroke or cerebrovascular accident (OR, 20.34 [95% CI, 2.34–176.77]); and use of pressors/vasoconstrictors (OR, 7.12 [95% CI, 5.11–9.91]). Hosmer–Lemeshow χ^2 values, P values, and 95% confidence intervals for AUROC values are shown in Table 6. The P^- model showed good calibration for three of four diagnostic groups (being best for internal medical disease [$\chi^2 = 4.00$] and worst for sepsis [$\chi^2 = 17.38$]). We also identified different levels of probability in the validation dataset (Table 7). The discrimination ratio was 91.8% for 50% probability, and the AUROC was 0.87 for the test dataset and 0.90 for the validation dataset.

Discussion

The APACHE score, MPM and SAPS are widely used in intensive care medicine.[18-29] These approaches depend primarily on organ scores that require physiological data. Ohno-Machado and colleagues found that AUROCs for APACHE II, APACHE III, MPM_0 (MPM at admission), MPM_{24} (MPM at 24 hours), MPM II_0, MPM II_{24}, SAPS and SAPS II were all ⩾ 0.80 except for SAPS.[18]

In contrast to these models, Duke and colleagues[11] derived the COPE model using administrative data. This model is favoured because it can predict mortality with relatively few variables, and is currently the only model based on administrative data alone. The COPE model includes mechanical ventilation as intensive-care therapy but does not include other life-support interventions such as dialysis and pressors/vasoconstrictors. However, the Hosmer–Lemeshow χ^2 statistic suggested that calibration of the COPE model was no better than that of APACHE III.

Compared with the COPE model, the P^- model developed in our study is based on prediction of 28-day mortality, rather than hospital mortality, and may serve as a new tool for ICU evaluation based on administrative data. The P^- model also has several other advantages over existing models. First, the variables depend on information that can

Table 5. Coefficients in the C , P^+ , P^- models developed using the test dataset (multivariate analysis) ($n = 3505$)*

Variable	C model		P^+ model		P^- model	
	B	OR (95% CI)	B	OR (95% CI)	B	OR (95% CI)
(2) Age	0.03	1.03 (1.02–1.04)	0.04	1.04 (1.02–1.05)	0.03	1.03 (1.02–1.04)
(3) Admission category (emergency)	1.83	6.26 (4.05–9.67)	1.91	6.74 (4.11–11.08)	1.85	6.38 (3.96–10.30)
(5) Reason for ICU entry						
(i) After emergency surgery			1.02	2.78 (1.37–5.62)	1.06	2.90 (1.47–5.73)
(ii) Medical disease			1.25	3.51 (1.97–6.25)	1.20	3.31 (1.89–5.79)
(7) Time from admission to ICU entry						
(i) Direct			−1.34	0.26 (0.15–0.46)	−1.17	0.31 (0.18–0.54)
(ii) After 1 day			−1.59	0.20 (0.10–0.44)	−1.49	0.22 (0.11–0.47)
(8) Use of fresh frozen plasma or platelet preparation			0.47	1.60 (1.03–2.50)		
(9) Mechanical ventilation (≥ 5 hours)	1.66	5.28 (3.97–7.03)	1.53	4.61 (3.33–6.37)	1.56	4.77 (3.51–6.50)
(10) Dialysis			1.33	3.78 (2.31–6.17)	1.47	4.35 (2.72–6.95)
(11) Pressors/vasoconstrictors			2.07	7.91 (5.62–11.15)	1.96	7.12 (5.11–9.91)
(6) Primary diagnosis on admission						
Haemopoietic malignancy	3.14	23.07 (4.91–108.44)	3.00	20.14 (3.42–118.76)		
Other CNS disease	1.41	4.11 (0.95–17.87)	1.85	6.37 (1.33–30.43)		
Haemorrhagic shock			1.39	4.03 (1.47–11.05)		
Stroke or CVA	1.21	3.37 (1.30–8.73)	3.01	20.34 (2.34–176.78)		
Liver disease	2.01	7.46 (2.72–20.43)	2.08	8.00 (2.52–25.16)		
Intracranial haemorrhage			1.36	3.90 (1.73–8.79)		
Environmental disease	1.70	5.49 (1.33–22.60)	1.60	4.93 (1.13–21.48)		
Lower limb trauma	−2.29	0.10 (0.01–0.74)	−2.28	0.10 (0.01–0.78)		
Renal failure	0.73	2.07 (1.01–4.26)				
Pneumonia	0.64	1.90 (1.05–3.43)				
Constant	−6.14		−8.23		−7.67	

B = β coefficient. CNS = central nervous system. CVA = cerebrovascular accident. ICU = intensive care unit. OR = odds ratio.

* Predicted mortality risk = $e^y / (e^y + 1)$, where $y = [B_{(2)} \times (2)] + [B_{(3)} \times (3)] + [B_{(5\text{-}i)} \times (5\text{-}i)] + [B_{(5\text{-}ii)} \times (5\text{-}ii)] + [B_{(7\text{-}i)} \times (7\text{-}i)] + [B_{(7\text{-}ii)} \times (7\text{-}ii)] + [B_{(8)} \times (8)] + [B_{(9)} \times (9)] + [B_{(10)} \times (10)] + [B_{(11)} \times (11)] + [B_{(6)} \times (6)]$ + constant. Each of the values (3), (5-i), (5-ii), (7-i), (7-ii), (8), (9), (10), (11) and (6) is equal to 1 if the variable is applicable or 0 if the variable is not applicable.

Table 6. Validation of the prediction model

Dataset	Model	No. of patients	28-day mortality	Hosmer–Lemeshow χ^2	*P*	AUROC (95% CI)
Test	P^-	3505	7.4	14.49	0.07	0.87 (0.85–0.90)
Test	P^+	3505	7.4	5.36	0.72	0.89 (0.87–0.91)
Test	C	3505	7.4	20.41	0.01	0.84 (0.82–0.87)
Validation	P^-	3253	9.7	13.52	0.10	0.90 (0.88–0.92)
Validation	P^+	3253	9.7	10.76	0.22	0.89 (0.87–0.90)
Validation	C	3253	9.7	12.91	0.12	0.84 (0.82–0.86)
(Subgroup of validation dataset)						
Internal medical disease	P^-	877	21.7	4.00	0.86	0.85 (0.82–0.88)
Emergency surgery	P^-	854	7.6	14.95	0.06	0.91 (0.88–0.94)
Scheduled surgery	P^-	1522	4.1	11.55	0.17	0.85 (0.81–0.90)
Sepsis	P^-	264	36.0	17.38	0.03	0.82 (0.77–0.89)

Table 7. Contingency table for different levels of cut-off points in the validation dataset in the P⁻ model

Cut-off points			Expected: Survivors	Expected: Non-survivors	DR
20%	Observed	Survivor	2699	237	89.6%
		Non-survivor	100	217	
50%		Survivor	2874	62	91.8%
		Non-survivor	205	112	
70%		Survivor	2926	10	91.4%
		Non-survivor	269	48	

DR = discrimination ratio.

be obtained from administrative data. These variables can be input by doctors and nurses in a timely manner, rather than at or after discharge, which improves the reliability of the data. Moreover, the model uses only eight variables, which facilitates its generalisation and application. Second, the model is independent of the primary diagnosis, which avoids the difficulty of disease identification in critical care patients. Also, coding for primary diagnosis is the basis for reimbursement in the health care system, and this diagnosis may be important for determining illness severity. A disadvantage of this approach is the potential for coding errors, especially in ICU patients.[11]

The 2007 Project IMPACT study[30] used a combination of MPM II_0 to assess clinical performance and a new Weighted Hospital Days scale to assess resource utilisation for ICU benchmarking. Our QIP study and the Project IMPACT study had a similar element of uncertainty regarding the clinical course after discharge, as data collection is difficult after discharge.[31] Thus, although 90-day mortality rate may be a better measure of outcome than 28-day mortality, the latter measure is more accurate because patients are usually discharged after less than 90 days. The COPE model is also a good predictor of hospital mortality. For these reasons, we used 28-day mortality as the endpoint.

There are several limitations to our study. First, we did not compare our model with scoring systems using physiological data, as our data did not include severity scores. Thus we cannot determine whether the accuracy of the model is high or low compared with other systems. Second, the administrative data include information on a calendar-day basis rather than an hourly basis, and thus the first ICU day was defined by a calendar day. This meant we were unable to distinguish between the use of dialysis and pressors/vasoconstrictors before or after ICU entry on the first ICU day. However, these resources are mostly used under monitoring in the ICU. Third, the indications for mechanical ventilation, dialysis and pressors/vasoconstrictors varied among hospitals, which may have produced therapeutic bias. Fourth, the administrative data do not indicate whether renal replacement therapy was given for chronic or acute renal failure or for a non-renal indication; whether mechanical ventilation was used for acute respiratory failure or during the postoperative course; or whether pressors/vasoconstrictors were used to treat hypovolaemic or septic shock. Finally, different admission criteria among ICUs could have produced a selection bias that affected mortality. Our model has a therapeutic bias similar to that of the COPE model, including the use of mechanical ventilation, dialysis, pressors/vasoconstrictors, and the use of fresh frozen plasma or a platelet preparation. However, it is likely that there would have been appropriate selection of these therapies because of common knowledge of guidelines.

Among the variables, sex was not significantly associated with outcome, which is consistent with other scoring systems. Age is an important variable for all scoring systems, and the predictive value of the model can be increased by adding other variables.[31] The COPE model[11] has high discrimination, suggesting that the predictive ability of a model constructed from administrative data is high. The absence of physiological data in administrative data is a disadvantage, as diagnosis is not included in the P⁻ model, but our model has the advantage of using administrative data that is routinely collected on all patients. Comparison of the performance of ICUs is currently being attempted using administrative data, and our model establishes a method for evaluation of illness severity. However, as our study included 9% of hospitals that use the DPC system and did not include university hospitals and non-teaching hospitals, further verification and modification of the model is required in a larger sample of patients and ICUs.

Conclusion

The 28-day mortality prediction model for intensive care (the P⁻ model) proposed in our study is based solely on administrative data, is independent of primary diagnosis, and uses a relatively small number of variables that are easily collected. In addition to the COPE model, this model can be used to evaluate illness severity based on administrative data and may be applicable to critical care studies.

Author details

Takeshi Umegaki, Intensive Care Specialist
Miho Sekimoto, Director
Kenshi Hayashida, Director
Yuichi Imanaka, Project Manager
Department of Healthcare Economics and Quality Management, Kyoto University Graduate School of Medicine, Kyoto, Japan.
Correspondence: imanaka-y@umin.net

References

1 Kuzniewicz MW, Vasilevskis EE, Lane R, et al. Variation in ICU risk-adjusted mortality: impact of methods of assessment and potential confounders. *Chest* 2008; 133: 1319-27.
2 McMillan TR, Hyzy RC. Bringing quality improvement into the intensive care unit. *Crit Care Med* 2007; 35 Suppl: S59-65.
3 Leary T, Ridley S, Burchett K, et al. Assessing critical care unit performance: a global measure using graphical analysis. *Anaesthesia* 2002; 57: 751-5.
4 Knaus WA, Draper EA, Wagner DP, et al. An evaluation of outcome from intensive care in major medical centers. *Ann Intern Med* 1986; 104: 410-8.
5 Knaus WA, Wagner DP, Zimmerman JE, et al. Variations in hospital mortality and length-of-stay from intensive care. *Ann Intern Med* 1993; 118: 753-61.
6 Zimmerman JE, Shortell SM, Rousseau DM, et al. Improving intensive care. *Crit Care Med* 1993; 21: 1443-51.
7 Render ML, Welsh DE, Kollef M, et al. Automated computerized intensive care unit severity of illness measure in the Department of Veterans Affairs: preliminary results. *Crit Care Med* 2000; 28: 3540-6.
8 Beck DH, Smith GB, Pappachan JV, et al. External validation of the SAPS II, APACHE II and APACHE III prognostic models in South England: a multicentre study. *Intensive Care Med* 2003; 29: 249-56.
9 Livingston BM, MacKirdy FN, Howie JC, et al. Assessment of the performance of five intensive care scoring models within a large Scottish database. *Crit Care Med* 2000; 28: 1820-7.
10 Castella X, Artigas A, Bion J, et al. A comparison of severity of illness scoring systems for intensive care unit patients: results of a multicenter, multinational study. The European/North American Severity Study Group. *Crit Care Med* 1995; 23: 1327-35.
11 Duke GJ, Santamaria J, Shann F, et al. Critical care outcome prediction equation (COPE) for adult intensive care. *Crit Care Resusc* 2008; 10: 35-41.
12 Lin YF, Ko WJ, Chu TS, et al. The 90-day mortality and the subsequent renal recovery in critically ill surgical patients requiring acute renal replacement therapy. *Am J Surg* 2009; 198: 325-32.
13 Brar H, Olivier J, Lebrun C, et al. Predictors of mortality in a cohort of intensive care unit patients with acute renal failure receiving continuous renal replacement therapy. *Am J Med Sci* 2008; 335: 342-7.
14 Soubrier S, Leroy O, Devos P, et al. Epidemiology and prognostic factors of critically ill patients treated with hemodiafiltration. *J Crit Care* 2006; 21: 66-72.
15 Pinsky MR. Hemodynamic evaluation and monitoring in the ICU. *Chest* 2007; 132: 2020-9.
16 Lundberg GD. Practice parameter for the use of fresh-frozen plasma, cryoprecipitate, and platelets. *JAMA* 1994; 271: 777-81.
17 Dunning J, Khasati N, Barnard J, Low dose (renal dose) dopamine in the critically ill patient. *Interact Cardiovasc Thorac Surg* 2004; 3: 114-7.
18 Ohno-Machado L, Resnic FS, Matheny ME. Prognosis in critical care. *Annu Rev Biomed Eng* 2006; 8: 567-99.
19 Rosenberg AL. Recent innovations in intensive care unit risk-prediction models. *Curr Opin Crit Care* 2002; 8: 321-30.
20 Afessa B, Gajic O, Keegan MT. Severity of illness and organ failure assessment in adult intensive care units. *Crit Care Clin* 2007; 23: 639-58.
21 Cook DA. Methods to assess performance of models estimating risk of death in intensive care patients: a review. *Anaesth Intensive Care* 2006; 34: 164-75.
22 Junger A, Engel J, Benson M, et al. [Risk predictors, scoring systems and prognostic models in anesthesia and intensive care. Part II: intensive care.] [German] *Anasthesiol Intensivmed Notfallmed Schmerzther* 2002; 37: 591-9.
23 Schusterschitz N, Joannidis M. Predictive capacity of severity scoring systems in the ICU. *Contrib Nephrol* 2007; 156: 92-100.
24 Kramer AA. Predictive mortality models are not like fine wine. *Crit Care* 2005; 9: 636-7.
25 Chen FG, Khoo ST. Critical care medicine: a review of the outcome prediction in critical care. *Ann Acad Med Singapore* 1993; 22: 360-4.
26 Harrison DA, D'Amico G, Singer M. Case mix, outcome, and activity for admissions to UK critical care units with severe acute pancreatitis: a secondary analysis of the ICNARC Case Mix Programme Database. *Crit Care* 2007; 11 Suppl 1: S1.
27 Strand K, Flaatten H. Severity scoring in the ICU: a review. *Acta Anaesthesiol Scand* 2008; 52: 467-78.
28 Capuzzo M, Moreno RP, Le Gall JR. Outcome prediction in critical care: the Simplified Acute Physiology Score models. *Curr Opin Crit Care* 2008; 14: 485-90.
29 Duke GJ, Piercy M, DiGiantomasso D, et al. Comparison of intensive care outcome prediction models based on admission scores with those based on 24-hour data. *Anaesth Intensive Care* 2008; 36: 845-9.
30 Nathanson BH, Higgins TL, Teres D, et al. A revised method to assess intensive care unit clinical performance and resource utilization. *Crit Care Med* 2007; 35: 1853-62.
31 Nagappan R, Parkin G. Geriatric critical care. *Crit Care Clin* 2003; 19: 253-70. ❑

INVESTIGATION REPORT J Jpn Soc Intensive Care Med 2010;17:555～559.

Current anticoagulation therapy for sepsis-induced disseminated intravascular coagulation in Japan: results of a multicenter study using administrative data

Takeshi Umegaki, Miho Sekimoto, Hiroshi Ikai, Yuichi Imanaka †

Abstract: Objective: Disseminated intravascular coagulation (DIC) is a serious complication associated with various underlying disorders, including sepsis. The aim of the current study was to investigate the status of therapy for patients with sepsis-induced DIC and to examine the association between 28-day mortality and use of anticoagulants. **Methods:** A multicenter cross-sectional study was performed from January 1, 2007 to December 31, 2008 in 45 ICUs in Japan. Using administrative data, 579 cases of sepsis-induced DIC were identified among patients who were admitted to an ICU, and these cases were used to assess the status of DIC therapy. The 28-day mortality was adjusted for the Critical care Outcome Prediction Equation (COPE) score, the Charlson comorbidity index and patient age, and associations with anticoagulants were then examined. **Results:** Protease inhibitors were used in 413 cases (71.3%), and antithrombin, unfractionated heparin, and low molecular weight heparin/danaparoid were used in 313 (54.1%), 385 (66.5%) and 201 (34.7%) cases, respectively. The overall 28-day mortality was 37%. In a Cox proportional hazards regression model, the hazard ratio (HR) of unfractionated heparin was 1.41, with a significant adverse effect on mortality ($P=0.02$). In a similar analysis, the HRs for protease inhibitors, antithrombin and low molecular weight heparin/danaparoid were 0.86, 0.90 and 0.88, respectively. These agents showed a tendency to reduce 28-day mortality, but the effect was not significant. **Conclusions:** A review of administrative data revealed that protease inhibitors were most frequently used in DIC anticoagulation therapy in ICUs in Japan. Unfractionated heparin was the only therapy to have a significant adverse effect on mortality.

Key words: ①disseminated intravascular coagulation, ②anticoagulants, ③multicenter study

Introduction

Disseminated intravascular coagulation (DIC) is a complex, acquired life-threatening disorder characterized by massive systemic intravascular coagulation that leads to widespread deposition of fibrin in the circulation [1]. Anticoagulation therapy has been suggested to be beneficial for DIC [2], and a study on guidelines for this therapy was published in 2007 in Japan [3]. The aims of the current retrospective study were to investigate the status of therapy for patients with sepsis-induced DIC in ICUs in Japan, and to evaluate the association between use of anticoagulants and 28-day mortality.

This article is featured in "HIGHLIGHTS IN THIS ISSUE". Please see the issue of J Jpn Soc Intensive Care Med 2010; 17: 474-476.

Methods

Cases and selection criteria

Data were obtained from the Quality Indicator/Improvement Project (QIP), in which detailed administrative claim data were collected from institutions. In the QIP, these data were analyzed to obtain numerical indices for the healthcare process, patient outcomes, and management efficiency to provide feedback to establishments that participate voluntarily in the project. In the current study, administrative data were surveyed from 45 acute care hospitals with an ICU. These data provided information on the characteristics of patients and daily medical care, thereby permitting collection of data for a large population. The following selection criteria were

Department of Healthcare Economics and Quality Management, Kyoto University Graduate School of Medicine
Yoshida Konoe-cho, Sakyo-ku, Kyoto, Kyoto 606-8501, Japan
† Corresponding author

Submitted for publication May 22, 2009
Accepted for publication February 17, 2010

Table 1 ICD-10 codes utilized for identification of septic cases

Condition	Code
Salmonella septicaemia	A02.1
Septicaemic plague	A20.7
Anthrax septicaemia	A22.7
Erysipelothrix septicaemia	A26.7
Listerial septicaemia	A32.7
Streptococcal septicaemia	A40
Other septicaemia	A41
Actinomycotic septicaemia	A42.7
Disseminated herpesviral disease	B00.7
Candidal septicaemia	B37.7
Disseminated coccidioidomycosis	B38.7
Disseminated histoplasmosis capsulati	B39.3
Disseminated blastomycosis	B40.7
Disseminated paracoccidioidomycosis	B41.7
Disseminated sporotrichosis	B42.7
Disseminated aspergillosis	B44.7
Disseminated cryptococcosis	B45.7
Disseminated mucormycosis	B46.4
Puerperal sepsis	O85

ICD-10, International Classification of Diseases, 10th version.

Table 2 Hospital and patient backgrounds

Hospital background	
Number of hospitals	45
Number of patients	579
Number of beds in hospital	404.8 ± 230.2
Patient background	
Age	71.1 ± 13.2
Gender (male %)	57.3
Hospital admission course (Emergency %)	86.2
Charlson comorbidity index	Frequency (%)
0	45.6
1	23.7
2	19.3
3	6.0
4	2.6
5	2.4
≥6	0.3
Length of ICU stay (days)	7.7 ± 7.1
Length of hospital stay (days)	35.8 ± 27.4
Expected mortality (%)	22.8 ± 17.8
28-day mortality (%)	37

used in the study: (1) primary diagnosis of sepsis (Table 1); (2) secondary diagnosis of DIC [code D65 in the International Classification of Diseases, 10th version (ICD-10)]; (3) age ≥ 20 years old; and (4) entry into the ICU and hospital discharge from January 1, 2007 to December 31, 2008.

We excluded patients who were not treated with anticoagulation agents and those who died on the day on which anticoagulant therapy was initiated. Cases that required dialysis were also excluded, since anticoagulant agents such as protease inhibitors and unfractionated heparin are also administered in renal replacement therapy.

Practice patterns in DIC anticoagulant therapy

Anticoagulation therapy was defined as the use of a protease inhibitor (gabexate mesylate or nafamostat mesylate), antithrombin, unfractionated heparin, or low molecular weight heparin/danaparoid. Low molecular weight heparin and danaparoid sodium were included in one category because danaparoid sodium has stronger anti-Xa activity than unfractionated heparin. In this study, the therapeutic dose of unfractionated heparin for treatment of DIC was defined as at least 10,000 U/day. Use of a low dose of unfractionated heparin was excluded due to the difficulty of distinguishing usage of this dose between arterial monitoring and DIC therapy.

Statistical analysis

All statistical analyses were performed using SPSS software for Windows (Dr. SPSS-II, SPSS Japan Inc.). For calculation of the 28-day mortality, the onset of DIC was defined as the day on which anticoagulant therapy was started, since the true onset time could not be established from the administrative data. The association between anticoagulants and 28-day mortality was examined using a Cox proportional hazards regression model adjusted for the critical care outcome prediction equation (COPE) score [4], Charlson comorbidity index [5] and patient age. A P value below 0.05 was considered as significant. The COPE model is a risk-adjustment tool for use in ICUs that is constructed from 5 variables (age, unplanned admission, mechanical ventilation, hospital category, and admission diagnosis). This model provides the only prognostic severity score that can be obtained from administrative data alone, without use of physiological data. Previous work has shown that hospital mortality in critical care patients can be predicted using the COPE model [receiver operating characteristics area under the curve (ROC AUC) = 0.83–0.84] [4]. The Charlson comorbidity index, on which a higher score indicates greater severity, is a useful tool for measurement of the comorbid status or case mix in health care databases and has been adapted for use with ICD-10 data [6].

Results

Background

Data were examined for 37,456 patients who were treated in 45 ICUs. Among these patients, 724 (1.9%) cases of sepsis-induced DIC were identified, but 145 (20%) fulfilled the exclusion criteria, leaving 579 (80%) that were included in our analysis. The patient and hospital backgrounds are shown in Table 2. The mean

Table 3 Use of anticoagulant agents in cases of DIC

Anticoagulant agent (alone or in combination)	Number of cases (%)
Protease inhibitor	413 (71.3)
Antithrombin	313 (54.1)
Unfractionated heparin	385 (66.5)
Low molecular weight heparin/danaparoid	201 (34.7)

n=579 (45 ICUs).

Table 4 Relationship between anticoagulant agents and adjusted 28-day mortality (Cox proportional hazards regression model)

Anticoagulant agent	HR	95% CI	P value
Unfractionated heparin	1.41	1.06–1.87	0.02*
Protease inhibitor	0.86	0.63–1.16	0.32
Antithrombin	0.90	0.67–1.20	0.46
Low molecular weight heparin/danaparoid	0.88	0.59–1.31	0.53

*: $P < 0.05$.
CI, confidence interval; HR, hazard ratio.

age of the patients was approximately 71 years old and the expected mortality in the COPE model was 22.8%. More than half of the patients had a Charlson comorbidity index of 0 or 1. The mean length of ICU stay was 7.7 days and the mean length of hospital stay was 35.8 days.

DIC anticoagulation therapy

The frequencies of use of anticoagulation therapies are displayed in Table 3. A protease inhibitor was used in 413 of the 579 cases (71.3 %), and antithrombin, unfractionated heparin, and low molecular weight heparin/danaparoid were used in 313 (54.1%), 385 (66.5%) and 201 (34.7%) cases, respectively. These interventions were combined in various ways in many cases.

Cox proportional hazards regression model

The 28-day mortality was 37%. After adjustment for the COPE score, Charlson comorbidity index and patient age, unfractionated heparin had a significant adverse effect on the 28-day mortality [hazard ratio (HR) =1.41, P=0.02]. In contrast, protease inhibitors, antithrombin and low molecular weight heparin/danaparoid showed a tendency to reduce the 28-day mortality, but this effect was not significant (Table 4).

Discussion

Wada [3] suggested guidelines for DIC treatment in 2007 to facilitate evidence-based practice of anticoagulation therapy. In these guidelines, antithrombin was recommended over protease inhibitors and low molecular weight heparin/danaparoid, but our results suggest that protease inhibitors are used more frequently than antithrombin in current practice. We also found that low molecular weight heparin/danaparoid is used less commonly than other anticoagulant agents. Thus, there is a discrepancy between the guidelines and current practice, which may be due to the historical development of DIC anticoagulant agents in Japan and the absence of a high level of evidence in support of anticoagulant therapy for DIC [7]. Protease inhibitors were developed in the 1980s in Japan and unfractionated heparin has traditionally been used as an anticoagulant. In contrast, use of low molecular weight heparin/danaparoid has been relatively uncommon.

Evidence for use of unfractionated heparin for DIC has not been obtained, but this agent has traditionally been administered as a conventional indication [3]. We found that administration of unfractionated heparin is less frequent than that of protease inhibitors, and more frequent than that of antithrombin in Japan. Furthermore, unfractionated heparin had a significant adverse effect on mortality, consistent with the lower level of evidence in the guidelines. This unfavorable result with unfractionated heparin may be caused by complications of hemorrhage or use in cases with severe bleeding. However, the presence of thrombosis often requires use of unfractionated heparin based on risks and benefits, and therefore use of this agent is likely to continue in DIC cases. Previous trials of anticoagulants for DIC have often used unfractionated heparin as a control [8),9)]. An appropriate control group for a clinical study should be based on standard or usual practice, but this concept is difficult to define [10]. Our results suggest that future trials should consider using protease inhibitors or antithrombin as the control, since most physicians in multiple centers in Japan prefer these therapies for DIC. These findings are also consistent with the guidelines.

There is little evidence that shows that anticoagulation agents significantly reduce mortality [2),11)–14)]. Our analysis revealed a small trend for reduction of the 28-day mortality with use of these agents, but the effect

was not significant, as also found in previous studies. A large-scale trial in cases with severe sepsis supported a slight, but not significant, benefit of low-dose heparin for reducing 28-day mortality, but underscored the importance of heparin withdrawal in cases that involved DIC and abnormal coagulation [14]. In this study, we were unable to assess the effect of low-dose heparin since this treatment was excluded from our analysis.

The difference between the expected mortality and the 28-day mortality may have been due to the longer period of 28 days used in our study, in contrast to the period of hospitalization used in other studies. Expected mortality in the COPE model obtained from ICU data in Australia and New Zealand was defined as hospital mortality, but acute care hospitals in Japan traditionally include acute care, sub-acute care, and nursing-home care. Therefore, the length of the hospital stay in Japan is generally reported to be longer than that in other developed nations [15]. However, the length of hospital stay for patients with septic DIC may be less than 28 days in other countries, and this might account for the difference between the expected and actual 28-day mortality in our study.

There are several limitations in the present study. First, most importantly, the administrative data did not include physiological data, the International Society on Thrombosis and Haemostasis (ISTH) DIC score, the Japanese Association for Acute Medicine (JAAM) DIC score, the DIC score developed by the Ad Hoc Group of the Japanese Ministry of Health and Welfare (JMHW), or a severity score for intensive care based on APACHE II or SAPS II. These scores reflect the severity of DIC, are used to predict patient outcomes [16]~[18], and can lead to early diagnosis and improved treatment when applied appropriately in clinical management of DIC [19]. This lack of information on DIC scores is the main disadvantage of using administrative data. However, evaluation of the severity of critical care patients using the COPE model eliminated this concern, and we also adjusted for comorbidity using the Charlson comorbidity index. Second, while we had detailed data for anticoagulant use, we were unable to assess the protocol in each ICU. Therefore, we were unable to evaluate the association between mortality and each practice pattern, despite the large apparent variation in practice. Finally, the study was not performed as a randomized controlled study or a prospective study, suggesting that the results might provide a lower level of evidence. However, it was offset by the advantage of data collection from a large population with reduced effort and time.

Conclusion

In summary, current treatment for sepsis-induced DIC in Japan commonly includes use of protease inhibitors. A tendency for improvement of outcome was found with use of protease inhibitors, antithrombin, and low molecular weight heparin/danaparoid, whereas unfractionated heparin had a significant adverse effect on 28-day mortality.

The work in the manuscript was presented at the 36th Annual Meeting of the Japanese Society of Intensive Care Medicine in Osaka, 2009.

Acknowledgements

The authors are grateful for the voluntary participation of the hospitals in the Quality Indicator/Improvement Project (QIP). The hospitals also participated in or cooperated with the Diagnosis Procedure Combination (DPC) system in Japan. The datasets in QIP are managed under the informational security system of ISO27001. The Institutional Review Board of the Faculty of Medicine at the Graduate School of Medicine of Kyoto University approved QIP and this study.

Conflict of Interest and Declarations

All authors declare no conflicts of interest. The design, data collection and analysis, and writing of the manuscript were performed by all four authors. The corresponding author takes full responsibility for the validity of the data.

References

1) Kruger D. Acute systemic disseminated intravascular coagulation: managing a complex medical condition. JAAPA 2006;19:28-32.
2) Wiedermann CJ. Clinical review: molecular mechanisms underlying the role of antithrombin in sepsis. Crit Care 2006;10:209-17.
3) Wada H. Study on guidelines for the treatment of DIC. Rinsho Ketsueki 2007;48:377-85.
4) Duke GJ, Santamaria J, Shann F, et al. Critical care outcome prediction equation (COPE) for adult intensive care. Crit Care Resusc 2008;10:41.
5) Charlson ME, Pompei P, Ales KL, et al. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 1987;40:373-83.
6) Sundararajan V, Henderson T, Perry C, et al. New ICD-10 version of the Charlson comorbidity index predicted in-hospital mortality. J Clin Epidemiol 2004;57:1288-94.
7) Wada H, Asakura H, Okamoto K, et al; Japanese Society of Thrombosis Hemostasis/DIC subcommittee. Expert consensus for the treatment of disseminated intravascular coagulation in Japan. Thromb Res 2010;125:6-11.
8) Aramoto H, Saito H, Shigematsu H, et al. Synthetic protease inhibitors in the treatment of disseminated intravascular coagulation. Nippon Rinsho 1993;51:93-8.
9) Sakuragawa N, Hasegawa H, Maki M, et al. Clinical evaluation of low-molecular-weight heparin (FR-860) on disseminated intravascular coagulation (DIC)--a multicenter co-operative double-blind trial in comparison with heparin. Thromb Res 1993;72:475-500.
10) Brower RG, Matthay M, Schoenfeld D. Meta-analysis of acute lung injury and acute respiratory distress syndrome trials. Am J Respir Crit Care Med 2002;166:1515-7.
11) Sakuragawa N, Hasegawa H, Maki M, et al. Clinical evaluation of low molecular weight heparin (FR-860) on disseminated intravascular coagulation (DIC)--a multicenter cooperative double-blind trial in comparison with heparin--. Rinsho Iyaku 1992;8:423-53.
12) Warren BL, Eid A, Singer P, et al; KyberSept Trial Study Group. Caring for the critically ill patient. High-dose antithrombin III in severe sepsis: a randomized controlled trial. JAMA 2001;286:1869-78.

13) Nishiyama T, Matsukawa T, Hanaoka K. Is protease inhibitor a choice for the treatment of pre- or mild disseminated intravascular coagulation? Crit Care Med 2000;28:1419-22.
14) Levi M. Disseminated intravascular coagulation. Crit Care Med 2007;35:2191-5.
15) Ikegami N, Campbell JC. Medical care in Japan. N Engl J Med 1995;333:1295-9.
16) Gando S, Saitoh D, Ogura H, et al. Natural history of disseminated intravascular coagulation diagnosed based on the newly established diagnostic criteria for critically ill cases: Results of a multicenter, prospective survey. Crit Care Med 2008;36:145-50.
17) Bakhtiari K, Meijers JCM, de Jonge E, et al. Prospective validation of the International Society of Thrombosis and Haemostasis scoring system for disseminated intravascular coagulation. Crit Care Med 2004;32:2416-21.
18) Angstwurm MW, Dempfle CE, Spannagl M. New disseminated intravascular coagulation score: A useful tool to predict mortality in comparison with Acute Physiology and Chronic Health Evaluation II and Logistic Organ Dysfunction scores. Crit Care Med 2006;34:314-20.
19) Gando S, Iba T, Eguchi Y, et al. A multicenter, prospective validation of disseminated intravascular coagulation diagnostic criteria for critically ill cases: Comparing current criteria. Crit Care Med 2006;34:625-31.

Verification Bias in Assessment of the Utility of MRI in the Diagnosis of Cruciate Ligament Tears

Haruo Nishikawa[1]
Yuichi Imanaka
Miho Sekimoto
Hiroshi Ikai

OBJECTIVE. The purpose of this study was to investigate the extent to which verification bias affects the sensitivity and specificity of MRI in the diagnosis of cruciate ligament tears.

MATERIALS AND METHODS. Consecutively registered outpatients who underwent MRI evaluation of the knee were included in the study. The sensitivity and specificity of MRI were calculated for patients whose diagnosis was verified with arthroscopy. For patients who did not undergo arthroscopy, the effect of verification bias was estimated with global sensitivity analysis, a technique of graphic representation of whether a particular combination of sensitivity and specificity estimates is compatible with the observed data.

RESULTS. Among the 356 patients included in the study, 82 patients (23%) had the MRI findings verified at arthroscopy. The sensitivity and specificity of MRI among patients who underwent arthroscopy were 38% and 90%. For patients whose disease status was not verified with arthroscopy, the influence of verification bias was estimated with global sensitivity analysis. The sensitivity of MRI ranged from 3% to 73%, and the specificity from 63% to 98%. The region comprising all possible combinations of sensitivity and specificity had a butterfly shape. The sensitivity and specificity pair estimated from cases verified with arthroscopy was included in this region.

CONCLUSION. Verification bias did not greatly affect assessment of the diagnostic utility of MRI in the evaluation of cruciate ligament tears. The high specificity previously reported for MRI can be considered valid, but the sensitivity may not be as reliable.

MRI has been widely used for screening and is highly regarded as an excellent, cost-effective diagnostic tool that is both noninvasive and accurate [1, 2]. Both the sensitivity and specificity of MRI in the diagnosis of cruciate ligament tears have been reported to be greater than 80%. Concern has been raised, however, about the methods used in studies of the diagnostic accuracy of MRI, including the effect of verification bias [3–7]. Verification bias (also known as workup bias, posttest referral bias, and selection bias) occurs when not all patients are equally likely to have the diagnosis confirmed with a reference standard [4, 6]. In the evaluation of a diagnostic test against a definitive reference standard test, which can be invasive and expensive, not all patients who have negative results of the diagnostic test undergo confirmatory testing with the reference standard. Therefore, patients with verified disease status may not be representative of the population in which the diagnostic test is used. If few cases of negative test results are verified with the reference standard, few false-negative findings will be revealed, leading to overestimation of accuracy.

In the diagnosis of cruciate ligament tear, arthroscopy is currently regarded as the reference standard because of its high reported accuracy, and MRI has been evaluated in reference to arthroscopic results [8–11]. However, because of its invasive nature and the risk of serious complications, not all study participants undergo arthroscopy [12]. Rather, arthroscopy usually is performed for patients with abnormal MRI findings. If patients with arthroscopic verification are more likely to have a cruciate ligament tear than are patients without arthroscopic verification, the difference can lead to verification bias, and the sensitivity and specificity of MRI may be overestimated. In an earlier study [3], we found that verification bias can greatly affect assessment of the diagnostic utility of MRI in the diagnosis of meniscal tear. To our knowledge, however, no study has been conduct-

Keywords: arthroscopy, cruciate ligament tear, diagnosis, MRI, verification bias

DOI:10.2214/AJR.10.4189

Received December 25, 2009; accepted after revision April 23, 2010.

[1]All authors: Department of Healthcare Economics and Quality Management, School of Public Health, Kyoto University Graduate School of Medicine, Yoshida Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan. Address correspondence to Y. Imanaka (imanaka@pbh.Medicinekyoto-u.ac.jp).

WEB
This is a Web exclusive article.

AJR 2010; 195:W357–W364

0361–803X/10/1955–W357

ed to investigate the extent to which verification bias affects the diagnostic performance of MRI in the evaluation of cruciate ligament tear. Because both the cruciate ligament and the meniscus are soft-tissue structures in the knee, verification bias may have considerable influence on the diagnostic accuracy of MRI of cruciate ligament tear. The purpose of this study was to use global sensitivity analysis to investigate the extent to which verification bias affects the diagnostic utility of MRI in the evaluation of cruciate ligament tear. This method, proposed by Kosinski and Barnhart [7], is a robust method of estimating the influence of verification bias [3, 13].

Materials and Methods

Subjects

The study was conducted in an outpatient clinic at a single institution. Included in the study were consecutively registered patients who reported knee pain and visited the hospital for MRI evaluation of the cruciate ligament from April 2006 through July 2008. The patients underwent MRI before arthroscopy. The study plan was announced in a poster in the hospital ward to offer the opportunity to refuse participation. Many of the study patients continued to visit the hospital for follow-up, and none had refused participation as of this writing. This study was performed at the same institution as a study of meniscal tear [3]. The exclusion criteria were previous knee surgery, more than 240 days between MRI and arthroscopy, and poor resolution of MR images. A retrospective chart review was performed to collect information on patient characteristics and clinical findings. Approval for this study was granted by the institutional review board at our institution.

Diagnosis

The same protocol was used for all MRI examinations in this study. All images were obtained with a 1.5-T MRI unit (Excelart with Pianissimo, Toshiba Medical Systems), extremity coil (quadrature coil), and the fast spin-echo method. Sagittal T2-weighted images, sagittal T1-weighted images, sagittal STIR images, sagittal T2*-weighted images, coronal STIR images, coronal T2*-weighted images, and axial T2-weighted images were obtained, each with a scanning time of 2–3 minutes. The parameters for the sagittal T2-weighted images were TR/TE, 3,628/94; field of view, 20 × 20 cm; slice thickness, 3.5 mm; interslice gap, 1.0 mm; matrix size, 224 × 288; flip angle, 90°, 160°; bandwidth, 244 Hz/pixel; echo-train length, 13. The Parameters for the sagittal T1-weighted images were 495/15; field of view, 20 cm × 20 cm; slice thickness, 3.5 mm; interslice gap 1.0 mm; matrix size, 176 × 272; flip angle 90°, 180°; bandwidth, 163 Hz/pixel; echo-train length, 0. Sagittal STIR images were obtained with the following parameters: 5,635/80; inversion time, 130 milliseconds; field of view, 20 × 20 cm; slice thickness 3.5 mm; interslice gap, 1.0 mm; matrix size, 224 × 304; flip angle 90°, 160°; bandwidth, 326 Hz/pixel, echo-train length, 15. The parameters for the sagittal T2*-weighted images were 535/15; field of view, 20 × 20 cm; slice thickness, 3.5 mm; interslice gap, 1.0 mm; matrix size, 160 × 304; flip angle, 25°; bandwidth, 61 Hz/pixel; echo-train length, 0. The parameters for the coronal STIR images were: 5,635/80; inversion time, 130 milliseconds; field of view, 20 × 20 cm; slice thickness, 3.0 mm; interslice gap, 1.0 mm; matrix size, 224 × 272; flip angle, 90°, 160°; bandwidth, 326 Hz/pixel; echo-train length, 15. The parameters for the coronal T2*-weighted images were 535/15; field of view, 20 × 20 cm; slice thickness, 3.0 mm; interslice gap, 1.0 mm; matrix size, 160 × 304; flip angle, 25°; bandwidth, 61 Hz/pixel; echo-train length, 0. The parameters for the axial T2-weighted images were 3,628/94; field of view, 18 × 20 cm; slice thickness, 4.0 mm; interslice gap, 2.0 mm; matrix size, 224 × 400; flip angle, 90°, 160°; bandwidth, 244 Hz/pixel; echo-train length, 13.

One of two radiologists with more than 10 years of experience interpreted all images. Each image was evaluated by either of two radiologists at the hospital. The image review findings were the initial clinical interpretations before arthroscopy. Cruciate ligaments were categorized into two subgroups depending on severity. Partial tear of the cruciate ligament was diagnosed if abnormal signal intensity was found in the ligament or when otherwise intact fibers appeared wavy on sagittal or coronal fast spin-echo images. Complete tear of the cruciate ligament was diagnosed if disruption of all fibers was found or if the ligament was not discernible at all on MR images.

The reference standard used in this study was arthroscopy because of its previously reported accuracy of greater than 95% [8–11]. One orthopedic surgeon with more than 15 years of experience performed arthroscopy in this study.

Analysis

The outcome measure was cruciate ligament tear, classified as anterior or posterior. We compared the results of MRI with those of arthroscopy. For the patients who underwent arthroscopy, we calculated the sensitivity and specificity of MRI. To assess the influence of verification bias, global sensitivity analysis was performed for patients who did not undergo arthroscopy [3, 7]. We simulated the complete range of possible prevalence (0–100%) of cruciate ligament tear for the MRI-positive and MRI-negative subgroups of patients who did not undergo arthroscopy then calculated and graphically plotted the sensitivity and specificity (Appendix 1). This method allowed us to depict all possible combinations of sensitivity and specificity.

We compared sensitivity and specificity in several subgroups of patients. These subgroups were based on the following factors known to influence the diagnostic accuracy of MRI: age (< 45 years, ≥ 45 years), sex (male, female), interval between MRI and arthroscopy (less than the lower quartile of this study population, lower quartile or greater), bundle tear (anterior, posterior), severity of tear (partial, complete). We performed chi-square tests to compare the positivity rate between subgroups of patients. Stata software (version 10, StataCorp) was used for statistical analysis.

Results

A flow diagram of the study is shown in Figure 1. Of the initial 361 patients, five were excluded from the final analysis because they had undergone knee surgery (three patients), had poor-resolution MR images (one patient), and had more than 240 days between MRI and arthroscopy (one patient). The general characteristics of the 356 patients included in the study (183 male patients, 174 female patients; mean age, 51 years) are shown in Table 1.

Forty-six patients had an abnormal (i.e., positive) test result, and 310 patients had a normal (i.e., negative) test result (Table 2). Only 82 patients (23%) underwent arthroscopy. Among the patients with tears verified

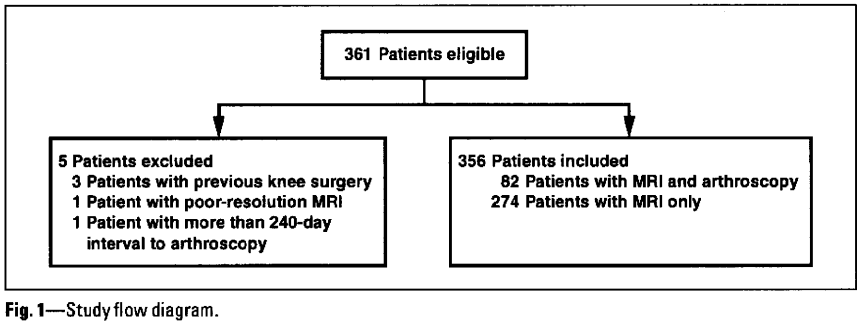

Fig. 1—Study flow diagram.

TABLE 1: Characteristics of Patients Included in Study (n = 356)

Characteristic	Value
Age (y)	
Mean	51
SD	20
Range	7–93
Sex	
Male	182
Female	174
Location of cruciate ligament tear	
Anterior	37
Posterior	9
Severity of cruciate ligament tear	
Partial	33
Complete	13

Note—Except for age, values are number of patients.

TABLE 2: Frequency of Test Results

MRI Result	Verified Cruciate Ligament Tear		Not Verified	Total
	Present	Absent		
Positive	8	6	27	41
Negative	13	55	247	315
Total	21	61	274	356

Note—Values are number of patients.

with arthroscopy, the sensitivity was 38% and the specificity 90%. The graph of the area comprising all possible combinations of sensitivity and specificity based on global sensitivity analysis is shown in Figure 2. This region included the estimated point of sensitivity and specificity (point estimate) calculated from the subgroup of patients with verified disease status (base case). Sensitivity (3–73%) varied to a greater extent than specificity (63–98%).

The subgroup for the second analysis consisted of 82 patients who underwent both MRI and arthroscopy. The characteristics of the 82 patients (48 male patients, 34 female patients; mean age, 52 years) are shown in Table 3. In all cases, the interval between MRI and arthroscopy was less than 6 months. Twenty-four cruciate ligament tears were verified with arthroscopy. Stratified comparisons were made by use of the chi-square test (Table 4). Statistically significant differences in sensitivity were observed for anterior and posterior tear location (35% vs 75%, $p = 0.02$) and in specificity were observed for partial and complete tear severity (90% vs 99%, $p = 0.02$).

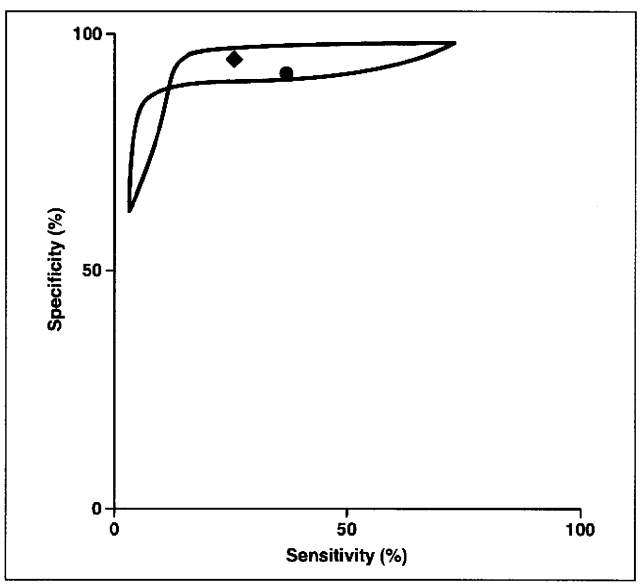

Fig. 2—Plot shows all possible combinations of sensitivity and specificity. Enclosed areas represent all possible combinations of values of sensitivity and specificity estimated with global sensitivity analysis. Circle indicates point estimate of sensitivity and specificity based only on data verified with arthroscopy (base case). Diamond indicates point estimate of sensitivity and specificity based on missing at random assumption.

Discussion

We used global sensitivity analysis to investigate the effect of verification bias on the sensitivity and specificity of MRI in the diagnosis of cruciate ligament tear. Previous studies of the diagnostic utility of MRI in the evaluation of cruciate ligament tear have shown high accuracy, but these studies predominantly included patients whose disease status was confirmed with arthroscopy [14–16]. Several studies have been conducted in attempts to correct for this verification bias, but none of the methods used in those studies successfully eliminate the bias [2, 17]. A simple way to correct for verification bias is to include in the study only patients who undergo both MRI and arthroscopy; the assumption is that patients whose condition is not verified with arthroscopy are as likely as patients whose condition is verified to have a cruciate ligament tear [4, 5]. Alternatively, the condition of patients with negative MRI results can be verified with a different, often less thorough method, such as follow-up imaging or evaluation for physical signs [18]. However, use of such methods cannot exclude the bias completely and can lead to inaccurate conclusions about the diagnostic utility of MRI. Physical signs and the medical history are clinically important but are not sensitive enough to exclude cruciate ligament tear [19].

Global sensitivity analysis is the most robust approach to assessment of the effect of verification bias [3, 7]. This method of analysis simulates the behavior of sensitivity and specificity in that the disease prevalence among patients whose condition is not verified with arthroscopy takes all possible values. Therefore, global sensitivity analysis can be used to determine graphically whether a particular pair of sensitivity and specificity estimates are compatible with observed data. In our study, the region of possible sensitivity and specificity pairs represented a butterfly

TABLE 3: Characteristics of Patients Undergoing Both MRI and Arthroscopy (*n* = 82)

Characteristic	Value
Age (y)	
Mean	52
SD	18
Range	13–79
Sex (no. of patients)	
Male	48
Female	34
Interval to arthroscopic reference test (d)	
Mean	41
SD	38
Range	1–167
Location of cruciate ligament tear (no. of tears)	
Anterior	20
Posterior	4
Severity of cruciate ligament tear (no. of tears)	
Partial	14
Complete	10

TABLE 4: Results of Stratified Comparisons of Sensitivity and Specificity in Five Subgroups of Patients Undergoing Both MRI and Arthroscopy

Characteristic	Sensitivity (%)	*p*[a]	Specificity (%)	*p*[a]
Age (y)		0.40		0.16
<45	79		81	
≥45	87		93	
Sex (no. of patients)		0.07		0.47
Male	92		89	
Female	76		92	
Interval to reference test (d)		0.37		0.64
<16	77		93	
≤16	87		89	
Location of cruciate ligament tear (no. of tears)		0.02		0.14
Anterior	35		90	
Posterior	75		99	
Severity of cruciate ligament tear (no. of tears)		0.48		0.02
Partial	23		90	
Complete	38		99	

[a]Chi-square test.

shape, as depicted in Figure 2. A point estimate within the region of possible sensitivity and specificity pairs indicates that verification bias does not exist or has a small influence. This was the case in our study, suggesting that the base case point estimate is compatible with observed data and thus that verification bias has little effect. Previous studies have shown sensitivity and specificity of more than 80% in the MRI diagnosis of cruciate ligament tear [1, 20, 21]. Although no correction was made for verification bias, indexes from these studies also may be consistent with our data.

One easy method to correct for verification bias is to calculate bias-corrected point estimates based on the missing at random assumption that within each subgroup of patients with positive or negative MRI findings, disease status is independent of whether a patient undergoes arthroscopy [3, 7, 22] (Appendix 2). The sensitivity and specificity of the missing at random estimate were 28% and 94%. This point estimate was within the presumed region and was compatible with the observed data. Compared with this missing at random point estimate, the sensitivity of the base case was lower and the specificity was higher, suggesting that sensitivity was underestimated and specificity was overestimated. Nevertheless, the differences in sensitivity and specificity in the base case point estimate and missing at random point estimate were small, thus verification bias had little effect on the diagnostic utility of MRI in the diagnosis of cruciate ligament tear. Bias correction is a complex field, and various methods have been attempted. The missing at random assumption is certainly not perfect, but the compatibility with the observed data in our study support missing at random as a valid method of estimating actual sensitivity and specificity.

One of the major findings in our study was that the specificity was greater than 85% for most of the presumed range determined with global sensitivity analysis. The presumed range widens as the condition of fewer patients is verified and is particularly wide when more patients have an unverified condition than have a verified condition. In our global sensitivity analysis, specificity varied much less than sensitivity. The high specificity and its narrow presumed range are consistent with the high specificity reported in previous studies.

The sensitivity of MRI in this study was remarkably low compared with values reported in previous studies by other investigators. This low sensitivity may have had a number of explanations. The first is the long interval between MRI and arthroscopy. The sensitivity of MRI for cruciate ligament tear decreases as the interval increases. In our study, more patients had a longer interval between MRI and arthroscopy than had a shorter interval, possibly resulting in lower sensitivity. This finding also may be associated with subsequent spontaneous healing. The second explanation is the possibility of spontaneous healing found in a previous study [23–27]. Spontaneous healing converts a positive arthroscopic result into a negative result. Furthermore, we predict that MRI of a