

---

UNIVERSITE DE LAUSANNE - FACULTE DE BIOLOGIE ET DE MEDECINE

Département de Médecine,  
Service de Médecine Interne

---

**The Interrater Reliability of the Pulmonary Embolism Severity Index**

THESE

préparée sous la direction du Professeur associé Drahomir Aujesky

et présentée à la Faculté de biologie et de médecine de  
l'Université de Lausanne pour l'obtention du grade de

DOCTEUR EN MEDECINE

par

Montserrat FRAGA CHRISTINET

Médecin diplômée de la Confédération Suisse  
Originaire de Lausanne(VD)

BME 3592

Lausanne

2010

Bibliothèque Universitaire  
de Médecine / BIUM  
CHUV-BH08 - Bugnon 46  
CH-1011 Lausanne

# *Imprimatur*

*Vu le rapport présenté par le jury d'examen, composé de*

*Directeur de thèse*      *Monsieur le Professeur associé Drahomir Aujesky*

*Co-Directeur de thèse*

*Expert*      *Monsieur le Professeur Laurent Nicod*

*Directrice de l'Ecole  
doctorale*      *Madame le Professeur Stephanie Clarke*

*la Commission MD de l'Ecole doctorale autorise l'impression de la thèse de*

*Madame Monserrat Fraga Christinet*

*intitulée*

*The interrater reliability of the pulmonary embolism  
severity index*

*Lausanne, le 19 octobre 2010*

*pour Le Doyen  
de la Faculté de Biologie et de Médecine*



*Madame le Professeur Stephanie Clarke  
Directrice de l'Ecole doctorale*

## Résumé

Le PESI (Pulmonary Embolism Severity Index) est un score clinique pronostique s'appliquant à des patients présentant un diagnostic d'embolie pulmonaire. Notre objectif était de démontrer la reproductibilité de ce score entre différents médecins chez des patients présentant une embolie pulmonaire. Nous avons donc identifié, de façon prospective, des patients présentant une embolie pulmonaire nouvellement diagnostiquée aux urgences d'un Hôpital Universitaire (CHUV, Lausanne). Pour tous ces patients, le médecin assistant en charge ainsi que le chef de clinique superviseur ont individuellement collecté les différentes variables permettant d'établir le score selon le PESI. Ils ont, ensuite, de façon indépendante, classifié les patients dans 5 classes de risque (I-V) ainsi qu'en deux groupes à bas risque versus haut risque, respectivement les classes I-II et les classes III à V.

Nous avons examiné la reproductibilité des données entre deux groupes de médecins (médecins assistants vs chefs de clinique), pour chacune des variables constituant le PESI, pour le score total en points, pour l'attribution aux 5 classes de risque ainsi que pour la classification en deux groupes à haut risque versus bas risque. Cette évaluation de la reproductibilité des résultats obtenus par les différents médecins s'est basée sur le calcul du Kappa (K) ainsi sur les Coefficients de Corrélation Intra-classe (ICC).

Parmi les 48 patients présentant une Embolie Pulmonaire inclus dans notre étude, les coefficients de reproductibilité entre médecins assistants et chefs de clinique étaient supérieurs à 0.60 pour 10 des 11 variables du PESI. La reproductibilité entre les 2 groupes de médecins, pour le total des points, pour l'attribution à une classe de risque I à V, ainsi que pour la classification en bas versus haut risque était presque parfaite.

Nos résultats démontrent la haute reproductibilité du PESI, et appuient donc l'intérêt de son utilisation pour la stratification du risque chez des patients présentant une embolie pulmonaire.

# **The Interrater Reliability of the Pulmonary Embolism Severity Index**

**Running Head:** Reliability of the Pulmonary Embolism Severity Index

Montserrat Fraga, MD

Patrick Taffé, PhD

Marie Méan, MD

Olivier Hugli, MD, MPH

Stéphanie Witzig, RN

Gérard Waeber, MD

Drahomir Aujesky, MD, MSc

**Author affiliations:** Division of General Internal Medicine (Drs Fraga, Méan, Waeber, and Ms. Witzig), Institute of Social and Preventive Medicine (Dr. Taffé), Emergency Department (Dr. Hugli), University of Lausanne, Lausanne; Division of General Internal Medicine (Dr. Aujesky), Bern University Hospital, Bern, Switzerland.

## **Corresponding author:**

Drahomir Aujesky, MD, MSc  
Professor of Medicine  
Universitätsklinik für Allgemeine Innere Medizin  
Bern University Hospital  
3010 Bern, Switzerland  
Phone: +41 632 88 84  
Fax: +41 31 632 88 851  
Email: drahomir.aujesky@insel.ch



## SUMMARY

The Pulmonary Embolism Severity Index (PESI) is a validated clinical prognostic model for patients with acute pulmonary embolism (PE). Our goal was to assess the PESI's interrater reliability in patients diagnosed with PE. We prospectively identified consecutive patients diagnosed with PE in the emergency department of a Swiss teaching hospital. For all patients, resident and attending physician raters independently collected the 11 PESI variables. The raters then calculated the PESI total point score and classified patients into one of five PESI risk classes (I-V) and as low (risk classes I/II) versus higher-risk (risk classes III-V). We examined the interrater reliability for each of the 11 PESI variables, the PESI total point score, assignment to each of the five PESI risk classes, and classification of patients as low versus higher-risk using kappa ( $\kappa$ ) and intraclass correlation coefficients (ICC). Among 48 consecutive patients with an objective diagnosis of PE, reliability coefficients between resident and attending physician raters were  $> 0.60$  for 10 of the 11 variables comprising the PESI. The interrater reliability for the PESI total point score (ICC: 0.89, 95% CI: 0.81-0.94), PESI risk class assignment ( $\kappa$ : 0.81, 95% CI: 0.66-0.94), and the classification of patients as low versus higher-risk ( $\kappa$ : 0.92, 95% CI: 0.72-0.98) was near perfect. Our results demonstrate the high reproducibility of the PESI, supporting the use of the PESI for risk stratification of patients with PE.

**Key Words:** interrater reliability, Pulmonary Embolism Severity Index.

The Pulmonary Embolism Severity Index (PESI) is a validated clinical prognostic model to identify patients with acute pulmonary embolism (PE) who are at low-risk of 30-day mortality and other adverse outcomes (1-4). Based on 11 clinical parameters readily available at the time of presentation (Table 1), a total point score is calculated by summing the integer-based prognostic weights of each prognostic variable identified for a given patient with PE. On the basis of the total point score, patients are classified into five risk classes (I-V) of increasing risk of mortality, ranging from 1.1% for patients in class I to 24.5% for patients in class V (1). Patients in risk classes I and II have a very low risk of 30-day mortality and are considered low-risk. In the absence of coexisting illnesses that warrant hospital admission or psychosocial conditions that preclude outpatient care, low-risk patients (risk classes I and II) are potential candidates for outpatient treatment or an abbreviated hospital stay. Patients in risk classes III to V have a substantially higher risk for 30-day mortality and should be treated in the hospital. Thus, in contrast to other clinical prognostic models for PE, the PESI provides clinicians with an easily applied, explicit risk stratification instrument, without any need for imaging studies (e.g., echocardiography, compression ultrasonography) or laboratory tests (5, 6). The prognostic accuracy of the PESI has been validated in multiple retrospective and prospective studies from different settings and countries (1-4).

Current methodological guidelines request that the reproducibility of a prediction rule and its individual predictive variables should be assessed before the rule can be adopted into clinical practice (7). The goal of our study was to prospectively assess the PESI's interrater reliability in patients diagnosed with PE.

## **METHODS**

### **Patient Identification and Eligibility**

We prospectively identified consecutive patients diagnosed with PE in the emergency department (ED) of a Swiss teaching hospital using an electronic patient tracking system. Patients aged  $\geq 18$  years who had an objective diagnosis of acute PE (positive spiral computed tomography, high-probability ventilation-perfusion lung scan, or positive pulmonary angiography) were eligible for the study. There were no exclusion criteria. The study was approved by the institutional review board.

### **Prospective Data Collection**

For all eligible patients, the ED residents and attending physicians in charge of the patient independently filled out a standardized data collection form. Collected data included the rater's age, gender, and years of clinical experience. Based on the patient's medical chart, the raters also recorded the diagnostic procedure (positive spiral computed tomography, high-probability ventilation-perfusion scan, or positive pulmonary angiography) that led to the diagnosis of PE and all 11 patient parameters comprising the PESI (age in years, gender, presence of cancer, heart failure, or chronic lung disease, pulse  $\geq 110$ /minute, systolic blood pressure  $< 100$  mm Hg, respiratory rate  $\geq 30$ /minute, temperature  $< 36^{\circ}\text{C}$ , altered mental status, and arterial oxygen saturation  $< 90\%$ ). Parameters that were not documented in the medical chart were assumed to be normal by the raters, a strategy successfully used in the derivation and validation of the PESI (1, 2, 4). The raters then calculated the PESI total score by summing the patient's age in years and the points for each predictor and classified patients into one of five PESI risk classes (I-V) and as low (risk classes



I and II) versus higher-risk (risk classes III-V). The raters were blinded to each other's assessment.

## **Statistical Analyses**

We compared baseline characteristics of resident and attending physician raters using Fisher's exact tests for categorical variables and Wilcoxon rank-sum tests for continuous variables. We estimated the interrater reliability between resident and attending physician raters for each of the 10 individual categorical PESI variables and patient classification as low (risk classes I–II) versus higher-risk (risk classes III–V) using the unweighted kappa ( $\kappa$ ) coefficient (8). We estimated the interrater reliability for the five PESI risk classes using the weighted  $\kappa$  coefficient (with weights  $1-|i-j|/(k-1)$ , where  $i$  and  $j$  index the rows and columns of the ratings by the two raters and  $k$  is the maximum number of possible ratings) (9). We also estimated interrater agreement for categorical variables using total percent agreement. To assess the interrater reliability for continuous variables such as patient age and the PESI total point score, we calculated the intraclass correlation coefficient (ICC) (10). We classified interrater reliability based on the magnitude of the reliability coefficients as follows: 0.0 to 0.19, poor; 0.20 to 0.39, fair; 0.40 to 0.59, moderate; 0.60 to 0.79, substantial; and 0.80 to 1.0, almost perfect (11). We used linear mixed regression to examine the association between the PESI total point score and rater characteristics (resident vs attending physician, gender, and years of clinical experience).

Assuming that 50-70% of patients are classified as higher-risk and a minimally acceptable  $\kappa$  of 0.4, we calculated that a sample size of 48 subjects with PE with two raters per subject would achieve 80% power to detect a  $\kappa$  of 0.8 using a two-tailed alpha level of 0.05 (12). All analyses were performed using Stata 10.2.



## RESULTS

We identified 48 consecutive patients who presented with an objective diagnosis of acute PE at the ED between October 2, 2009, and April 7, 2010. PE was diagnosed based on a positive spiral computed tomography in all cases. The median age of our study population was 65 years (interquartile range: 48 to 77 years); 56% were male. Overall, 26 non-unique resident and 25 attending physician raters participated in this study. Attending physician raters were significantly older (median, 35 vs 31 years;  $P = 0.002$ ), had a longer professional experience (median, 8 vs 4 years;  $P < 0.001$ ), and were somewhat more likely to be male than resident physician raters (50 vs 32%;  $P = 0.25$ ). All resident and attending physicians completed the data collection forms.

Table 2 shows patient age and the prevalence of the individual PESI variables based on the resident and attending physician raters and interrater reliability and total percent agreement for these variables. Reliability coefficients ranged from 0.46 for respiratory rate  $\geq 30$ /minute to 1.0 for age, gender, and chronic lung disease. Using conventional definitions, five variables had a near perfect interrater reliability, five variables had a substantial interrater reliability, and one variable had a moderate interrater reliability. Total percent agreement varied between 92% for respiratory rate  $\geq 30$ /minute and 100% for gender and chronic lung disease.

The interrater reliability between resident and attending physician raters for PESI point score calculation (ICC: 0.89, 95% confidence interval [CI]: 0.81 to 0.94) and PESI risk class assignment was near perfect (weighted  $\kappa$ : 0.81, 95% CI: 0.66 to 0.94). Compared to the attending physicians, the residents classified two patients into a higher PESI risk class and five patients into a lower PESI risk class (total percent agreement: 85%) (Table 3). None of the rater characteristics (resident vs attending

physician, gender, and years of clinical experience) was significantly associated with the PESI total point score.

The interrater reliability for the classification as low versus higher-risk was also near perfect (unweighted  $\kappa$ : 0.92, 95% CI: 0.72 to 0.98). The residents and attending physicians disagreed only on two patients (total percent agreement: 96%) (Table 4). In one case, the resident failed to include the patient's age in the PESI total point score, which led to the erroneous classification of the patient as low-risk. In the second case, the resident and attending physician disagreed as to whether the patient had a respiratory rate  $\geq 30$ /minute.

## DISCUSSION

This study documents near perfect interrater reliability between resident and attending physician raters for PESI risk stratification and the identification of low versus higher-risk patients. The interrater reliability for most individual variables that comprise the PESI was substantial or near perfect, with reliability coefficients  $> 0.6$  for 10 of 11 PESI variables. Prior retrospective and prospective validation studies have convincingly demonstrated that the PESI accurately identifies patients with PE who are at low risk of death and other adverse outcomes and who are potential candidates of outpatient treatment or an abbreviated hospital stay (1-4). The present study demonstrates the PESI's high reproducibility among physician raters with varying levels of clinical experience, supporting the use of the PESI to guide the initial admission decision.

Our results are consistent with a retrospective study that found a near perfect interrater reliability between two unique physician raters for PESI point score calculation (ICC: 0.82) (13). The interrater reliability for PESI risk class assignment and the classification of patients as low versus higher-risk was lower in this study, with a  $\kappa$  value of 0.57 and 0.69, respectively (13). While the  $\kappa$  values for the individual variables that comprise the PESI were not reported in this study, some disagreement in the ascertainment of vital signs occurred in 64% of cases (13). Our findings indicate that the PESI's interrater reliability may be better when the PESI is prospectively applied by physicians while being actively involved in the care of patients with PE.

Methodological guidelines for the development of clinical prediction rules recommend the assessment of the interrater reliability of the rule itself as well as of the individual predictor variables (7). According to these guidelines, predictor



variables with reliability coefficients  $< 0.6$  are considered unreliable and should not be included in a prediction rule (7). Although a good interrater reliability is a crucial component of a clinical prediction rule's internal validity, only 3% of studies on the development of clinical prediction rules reported any reliability measures (7).

Although one of 11 variables included in the PESI had an interrater reliability below 0.6 (respiratory rate  $\geq 30$ /minute), the clinically more relevant interrater reliability for assignment to one of five PESI risk classes and classification of patients as low versus higher-risk remained near perfect, with reliability coefficients  $> 0.80$ . In comparison, in a prior study examining the interrater reliability of the Wells and Charlotte score, two prediction rules that estimate the clinical pre-test probability of PE, individual predictor variables had reliability coefficients ranging from 0.5 to 0.94 (14). The reliability coefficients for classifying patients as low, intermediate, or high-risk of having PE (Wells score) or as "safe" versus "unsafe" (Charlotte score) were 0.54 and 0.56, respectively (14), although other studies reported somewhat higher values (15, 16). The moderate reliability coefficient for respiratory rate  $\geq 30$ /minute ( $\kappa$ : 0.46), an easily available and explicit variable, is somewhat surprising. A recently developed simplified version of the PESI that does not comprise respiratory rate and altered mental status as predictor variables may potentially have an even higher reproducibility (17).

Our study has several strengths. First, we prospectively enrolled unselected, consecutive patients with PE, making a selection bias unlikely. Second, ratings were performed prospectively by resident and attending physicians who were actively involved in the care of the patients with PE, which closely simulates real-life application of the PESI in the ED. Third, raters were strictly blinded to each others' assessment, which reduced the risk of any cross-contamination between raters and

preserved independence between ratings. Finally, the fact that we used two large groups of non-unique resident and attending physician raters rather than two unique raters to assess interrater reliability is likely to increase the generalizability of our results.

Some limitations of our study should be acknowledged. First, our interrater reliability assessment was performed in a relatively small sample of 48 patients with PE and was restricted to residents and attending physicians working in the ED of a research-intensive teaching hospital. Thus, our results are not necessarily applicable to other facilities (e.g., non-teaching hospitals without major research activity) and settings (e.g., general internal medicine services) that are involved in the care of patients presenting with acute PE. Independent confirmation of our results in larger studies and various settings is therefore desirable. Second, participating physicians may have improved their performance of the risk assessment in response to being observed (Hawthorne effect). Thus, we cannot entirely exclude the possibility that the PESI's interrater reliability may be lower outside the setting of a research study.

In conclusion, there was a near perfect interrater reliability between resident and attending physician raters for assignment to PESI risk class and classification of patients as low versus higher-risk. The interrater reliability for the determination of most individual PESI variables was substantial or near perfect. These methodological findings coupled with prior studies demonstrating the accuracy of the PESI to identify low-risk patients with PE support the use of the PESI to guide the initial admission decision.

## **Acknowledgments**

The authors would like to thank all residents and attending physicians from the emergency department at the University Hospital of Lausanne, Switzerland, for their invaluable help with this study. The study was partially supported by a grant from the Swiss National Science Foundation (33CSCO-122659).



## REFERENCES

1. Aujesky D, Obrosky DS, Stone RA, et al. Derivation and validation of a prognostic model for pulmonary embolism. *Am J Respir Crit Care Med* 2005;172:1041-46.
2. Aujesky D, Roy PM, Le Manach CP, et al. Validation of a model to predict adverse outcomes in patients with pulmonary embolism. *Eur Heart J* 2006;27:476-81.
3. Jimenez D, Yusen RD, Otero R, et al. Prognostic models for selecting patients with acute pulmonary embolism for initial outpatient therapy. *Chest* 2007;132:24-30.
4. Donze J, Le Gal G, Fine MJ, et al. Prospective validation of the Pulmonary Embolism Severity Index. A clinical prognostic model for pulmonary embolism. *Thromb Haemost* 2008;100:943-48.
5. Wicki J, Perrier A, Perneger TV, Bounameaux H, Junod AF. Predicting adverse outcome in patients with acute pulmonary embolism: a risk score. *Thromb Haemost* 2000;84:548-52.
6. Sanchez O, Trinquart L, Caille V, et al. Prognostic Factors for Pulmonary Embolism: The PREP Study, A Prospective Multicenter Cohort Study. *Am J Respir Crit Care Med* 2009;181:168-73.
7. Laupacis A, Sekar N, Stiell IG. Clinical prediction rules. A review and suggested modifications of methodological standards. *JAMA* 1997;277:488-94.
8. Cohen JA. Coefficient of agreement for nominal scales. *Ed Psychol Meas* 1960;20:37-46.
9. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull* 1968;70:213-20.

10. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979;86:420-28.
11. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-74.
12. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther* 2005;85:257-68.
13. Chan CM, Woods C, Shorr AF. The validation and reproducibility of the Pulmonary Embolism Severity Index. *J Thromb Haemost* 2010 Apr 16. [Epub ahead of print]
14. Nordenholz KE, Naviaux NW, Stegelmeier K, et al. Pulmonary embolism risk assessment screening tools: the interrater reliability of their criteria. *Am J Emerg Med* 2007;25:285-90.
15. Wolf SJ, McCubbin TR, Feldhaus KM, Faragher JP, Adcock DM. Prospective validation of Wells Criteria in the evaluation of patients with suspected pulmonary embolism. *Ann Emerg Med* 2004;44:503-10.
16. Runyon MS, Webb WB, Jones AE, Kline JA. Comparison of the unstructured clinician estimate of pretest probability for pulmonary embolism to the Canadian score and the Charlotte rule: a prospective observational study. *Acad Emerg Med* 2005;12:587-93.
17. Jiménez D, Aujesky D, Moores L, et al. Simplification of the Pulmonary Embolism Severity Index for prognostication in patients with acute symptomatic pulmonary embolism. *Arch Intern Med* 2010. (in press)

**Table 1. The Pulmonary Embolism Severity Index (PESI)**

Predictors	Points Assigned
Age, per year	Age, in years
Male gender	+10
Cancer*	+30
Heart failure	+10
Chronic lung disease	+10
Pulse $\geq$ 110/minute	+20
Systolic blood pressure < 100 mm Hg	+30
Respiratory rate $\geq$ 30/minute†	+20
Temperature < 36°C	+20
Altered mental status‡	+60
Arterial oxygen saturation < 90%†	+20

A total point score for a given patient is obtained by summing the patient's age in years and the points for each applicable predictor. Points assignments correspond with the following risk classes:  $\leq$  65 class I; 66-85 class II; 86-105 class III; 106-125 class IV; and > 125 class V. Patients in risk classes I and II are defined as low-risk.

\*Defined as a history of cancer or active cancer.

†Assessed with and without the administration of supplemental oxygen.

‡Defined as disorientation, lethargy, stupor, or coma.



**Table 2. Agreement for the 11 individual PESI variables based on resident and attending physician assessment**

PESI predictor variables	Percent or median (IQR)		Interrater reliability coefficient (95% CI)*	Total % agreement
	Residents	Attending physicians		
Age, years	65 (48-77)	65 (48-77)	1.0	NA
Male gender	56.3	56.3	1.0 (0.85-1.0)	100
Cancer	22.9	20.8	0.94 (0.69-0.99)	98
Heart failure	8.3	4.2	0.65 (0.19-0.90)	96
Chronic lung disease	6.3	6.3	1.0 (0.52-1.0)	100
Pulse $\geq$ 110/minute	16.7	18.8	0.79 (0.47-0.93)	94
Systolic blood pressure < 100 mm Hg	4.2	6.3	0.79 (0.26-0.96)	98
Respiratory rate $\geq$ 30/minute	10.4	6.3	0.46 (0.09-0.78)	92
Temperature < 36°C	12.5	12.5	0.81 (0.44-0.95)	96
Altered mental status	2.1	4.2	0.66 (0.12-0.94)	98
Arterial oxygen saturation < 90%	4.2	6.3	0.79 (0.26-0.96)	98

Abbreviations: IQR = interquartile range; CI = confidence interval; NA = not applicable.

\*We used the unweighted kappa coefficient for categorical variables and the intraclass correlation coefficient for continuous variables.

**Table 3. Agreement of PESI risk class assignment between residents and attending physicians\***

Residents	Attending physicians					Total
	Class I	Class II	Class III	Class IV	Class V	
Class I	11	1	1	0	0	13
Class II	1	9	0	0	0	10
Class III	0	1	10	2	1	14
Class IV	0	0	0	7	0	7
Class V	0	0	0	0	4	4
Total	12	11	11	9	5	48

\*Weighted kappa coefficient: 0.81 (95% CI: 0.66 to 0.94); total percent agreement: 85%.

**Table 4. Agreement of patient classification as low versus higher-risk between residents and attending physicians\***

Residents	Attending physicians		Total
	Low-risk	Higher-risk	
Low-risk	22	1	23
Higher-risk	1	24	25
Total	23	25	48

\*Unweighted kappa coefficient: 0.92 (95% CI: 0.72 to 0.98); total percent agreement: 96%.



## **Additional Table**

### **(1) What is known on this topic**

- The Pulmonary Embolism Severity Index (PESI) is a validated clinical prediction rule for prognosis for patients with pulmonary embolism (PE)
- Current methodological guidelines request that the reproducibility of a prediction rule and its individual predictive variables should be assessed before the rule can be adopted into clinical practice.

### **(2) What this paper adds**

- This prospective study demonstrates that the PESI has an excellent interrater reliability between resident and attending physician raters
- These methodological findings further support the use of the PESI for risk stratification of patients with PE