Published in final edited form as:

Crit Care Med. 2010 October ; 38(10): 1955–1961. doi:10.1097/CCM.0b013e3181eb924f.

Toward a clinically feasible gene expression-based sub-classification strategy for septic shock: proof of concept

Hector R. Wong, Derek S. Wheeler, Ken Tegtmeyer, Sue E. Poynter, Jennifer M. Kaplan, Ranjit S. Chima, Erika Stalets, Rajit K. Basu, and Lesley A. Doughty

Division of Critical Care Medicine, Cincinnati Children's Hospital Medical Center and Cincinnati Children's Research Foundation, Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH

Abstract

Objective—To develop a clinically feasible stratification strategy for pediatric septic shock using gene expression mosaics and a 100-gene signature representing the first 24 hours of admission to the pediatric intensive care unit.

Design—Prospective observational study involving microarray-based bioinformatics.

Setting—Multiple pediatric intensive care units in the United States.

Patients—Ninety-eight children with septic shock.

Interventions—None other than standard care.

Measurements and Main Results—Patients were classified into 3 previously published, genome-wide, expression-based sub-classes (sub-classes A, B, and C) having clinically relevant phenotypic differences. The class-defining 100-gene signature was depicted for each individual patient using mosaics generated by the Gene Expression Dynamics Inspector (GEDI). Composite mosaics were generated representing the average expression patterns for each of the 3 sub-classes. Nine individual clinicians served as blinded evaluators. Each evaluator was shown the 98 individual patient mosaics and asked to classify each patient into one of the three sub-classes using the composite mosaics as the reference point. The respective sensitivities, specificities, positive predictive values, and negative predictive values of the sub-classification strategy were $\geq 84\%$ across the 3 sub-classes. The classification strategy also generated positive likelihood ratios ≥ 16.8 and negative likelihood ratios ≤ 0.2 across the three sub-classes. The kappa coefficient across all possible inter-evaluator comparisons was 0.81.

Conclusions—We have provided initial evidence (proof of concept) for a clinically feasible and robust stratification strategy for pediatric septic shock based on a 100 gene signature and gene expression mosaics.

Keywords

microarray; gene expression; stratification; staging; septic shock; pediatrics

INTRODUCTION

Human septic shock is a complex syndrome displaying a high degree of heterogeneity [1–3]. Our inability to manage and account for this heterogeneity is a major barrier to the conduct of more effective clinical trials. Indeed, multiple interventional clinical trials have been attempted in human septic shock and the majority of these trials have failed to demonstrate efficacy despite being founded on quality pre-clinical data and sound biological principles. It has been proposed that these repeated failures reflect our relative inability to address the profound heterogeneity of septic shock, rather than intrinsic flaws of the biological principles being tested [1,4]. Robust illness severity scores are currently available for critically ill pediatric and adult populations, but it is clear that these scores are not appropriate tools to stratify individual patients for the purposes of clinical trials [5,6]. Thus, there remains a need to develop more effective, clinical stratification/staging strategies for septic shock to meet the needs of clinical research and individual patient management. Successful achievement of this goal holds the promise of improving our therapeutic approach to human septic shock in an analogous manner to what has been achieved in the oncology field [7–11].

We recently reported the identification of three sub-classes of children with septic shock as a means to more effectively address the intrinsic heterogeneity of this syndrome [12]. The three sub-classes were initially identified based exclusively on genome-wide expression patterns using microarray analysis and unsupervised hierarchical clustering. Importantly, the gene expression patterns represented the first 24 hours of meeting criteria for septic shock, which is an ideal time frame for clinically relevant stratification. *Post-hoc* phenotype analyses revealed that patients in one of the three sub-classes had significantly higher illness severity, organ failure rates, and mortality, compared to the other two subclasses, thus indicating that the expression-based sub-classes are clinically relevant [12]. We further refined our sub-classification strategy to a 100-gene signature. Herein we report our initial attempt to develop a “clinician-friendly” stratification strategy based on these 100 class-predictor genes and the use of gene expression “mosaics”.

METHODS

Patients and microarray analysis

The 98 patients in the septic shock study cohort and their respective microarray studies have been previously reported with detailed clinical data and confirmation of Institutional Review Board approval [12]. Briefly, we conducted genome-wide expression studies using the Human Genome U133 Plus 2.0 GeneChip (Affymetrix, Santa Clara, CA) and whole blood-derived RNA from 98 individual children meeting published criteria for septic shock [13–16]. All microarray data used in the current report represent the first 24 hours of meeting criteria for septic shock. Using discovery-oriented expression/statistical gene filters and unsupervised hierarchical clustering techniques, we identified 3 sub-classes of children with septic shock based exclusively on their respective genome-wide expression patterns (heretofore referred to as “sub-class A,” “sub-class B,” and “sub-class C”). Sub-class A patients had significantly higher illness severity, organ failure rates, and mortality, compared to sub-class B and C patients [12]. Using K-means clustering and class prediction modeling, we refined the class-defining signature to 100 genes (see *Supplementary Data* for the 100 gene probe list). The microarray-derived expression patterns of these 100 class-defining genes are the basis of the current report.

Generation of gene expression mosaics

Expression mosaics representing the 100 class-defining genes were generated using the Gene Expression Dynamics Inspector (GEDI) platform. GEDI is a publicly available gene expression

analysis program developed by the Ingber Laboratory at Harvard University [17,18]. The signature graphical outputs of GEDI are gene expression mosaics that give microarray data a “face” that is intuitively recognizable via human pattern recognition [19–22]. The underlying algorithm for creating the mosaics is a self-organizing map (SOM). Conventional applications of SOMs involve classifying genes or samples into predetermined numbers of discrete clusters. In contrast, GEDI creates an SOM-based mosaic for every sample. Thus, GEDI is said to be “sample oriented” rather than “gene oriented.” Extensive technical details regarding GEDI can be found at: <http://www.childrenshospital.org/research/ingber/GEDI/gedihome.htm>

Data analysis

The individual patients' microarray expression data for the 100 class-defining genes were uploaded to the GEDI platform and individual gene expression mosaics were generated for each patient. Subsequently, we generated composite mosaics for each of the 3 sub-classes. The respective composite mosaics represent the average expression patterns of the individual patients within a given subclass.

We next conducted internal cross validation procedures to test the ability of the GEDI mosaics to accurately identify the three sub-classes of children with septic shock. Nine individual clinicians participated as evaluators in the cross validation procedures. The nine evaluators are all practicing pediatric critical care medicine specialists. Only one of the nine evaluators has experience interpreting microarray data. Each evaluator was shown the 98 individual patient mosaics, and asked to classify each patient as subclass “A”, “B”, or “C”, based on the individual expression patterns/mosaics and using the composite mosaics as the reference point. The individual evaluators were not provided with any additional prompting or instructions, and were blinded to the original patient sub-classifications.

Individual evaluator responses were catalogued to determine the ability of the individual mosaics to accurately sub-classify patients based on the originally published sub-classifications. When discrepancies occurred for a given patient (i.e. different evaluators classified a given patient into more than one class) the “majority vote” (i.e. ≥ 5 votes for a given class) was considered as the final/consensus classification for that given patient. For example, if 6 of the evaluators classified an individual patient as “subclass B”, and the other 3 evaluators classified the same patient as “subclass C”, then that patient was assigned to subclass B (i.e. the consensus classification). The performance characteristics of the classification strategy (i.e. specificity, sensitivity, etc.) and 95% confidence intervals were based on the consensus classifications of the 9 individual evaluators. We also calculated performance characteristics for each individual evaluator. Performance characteristics were calculated using a web-based clinical calculator (<http://faculty.vassar.edu/lowry/clin1.html>) and the originally published sub-classifications as the gold standard. To further measure inter-rater/evaluator agreement across the 9 evaluators, a kappa coefficient for multiple evaluators was calculated using a web-based calculator (<http://cosmion.net/jeroen/software/kappa/>).

RESULTS

The composite mosaics for the 3 septic shock sub-classes are shown in Figure 1. Examples of individual patient mosaics are provided in Figures 2A through 2C. There was 100% concordance across the 9 evaluators for 62 of the 98 patients (63%). That is, for 63% of the patients all 9 evaluators classified them to the same sub-class. More specifically, there was 100% concordance across the 9 evaluators for 82% of patients in sub-class A, 69% of patients in sub-class B, and 32% of patients in sub-class C. For 80% of the patients in the entire cohort, at least 8 of the 9 evaluators classified them to the same sub-class. The average kappa coefficient across all possible interrater/evaluator comparisons was 0.81 (0 = no agreement; 1 = perfect agreement).

The performance characteristics of the sub-classification strategy, based on consensus classification, are provided in Table 1. Twenty seven of the 28 original sub-class A patients (96%) were correctly classified as sub-class A. Thirty eight of the 45 original sub-class B patients (84%) were correctly classified as sub-class B. Twenty three of the 25 original sub-class C patients (92%) were correctly classified as sub-class C. The respective sensitivities, specificities, positive predictive values, and negative predictive values of the sub-classification strategy were $\geq 84\%$ across the three sub-classes. In addition, the sub-classification strategy yielded positive likelihood ratios ≥ 16.8 and negative likelihood ratios ≤ 0.2 across the three sub-classes.

Table 1 also provides the number of patients incorrectly classified. Two original sub-class B and 2 original sub-class C patients were incorrectly classified as sub-class A patients. Figures 3A and 3B provide the individual mosaics for these incorrectly classified patients. Four original sub-class B patients were incorrectly classified as sub-class C patients. Figure 4 provides the individual mosaics for these incorrectly classified patients. No patient was incorrectly classified as sub-class B. All of the incorrectly classified patients were survivors.

There were 2 patients that did not receive at least 5 “votes” (i.e. the majority vote) for any of the 3 sub-classes and therefore remained unclassified. Figure 5 provides the individual mosaics for these two patients. The patient represented by example 1 was an original sub-class A patient, but received 3 votes for each of the sub-classes. The patient represented by example 2 was an original sub-class B patient, but received 3 votes for sub-class A, 4 votes for sub-class B, and 2 votes for sub-class C. Both of these unclassified patients were survivors.

The performance characteristics for each one of the nine individual evaluators are provided in Tables 2a through 2i, respectively. In addition, the median and intra-quartile performance characteristics of the 9 individual evaluators are provided in Table 3. These data demonstrate similar performance characteristics between the consensus classifications and the individual evaluator classifications.

DISCUSSION

We have provided initial evidence for a clinically feasible and robust stratification strategy for children with septic shock. The assertion of feasibility is supported at two levels. First, the stratification strategy is based on a 100 gene signature using whole blood-derived RNA. With current multiplex technology (e.g. quantitative polymerase chain reaction or digital quantification of mRNA abundance) and a dedicated molecular diagnostics laboratory, it is now feasible to quantify the expression levels of 100 genes within a timeframe that is clinically acceptable [23]. Second, the sub-classification strategy relies on intuitive pattern recognition of GEDI-generated gene expression mosaics. We have demonstrated that clinicians without formal training in bioinformatics and complex gene expression analysis can accurately interpret these mosaics in a relatively reproducible and reliable manner, as objectively measured by a kappa statistic. We would expect that with additional training and experience in the interpretation of these mosaics, that correct response rates could be improved further.

The assertion of robustness is also supported at two levels. First, the data used for sub-classification represent the first 24 hours of meeting criteria for septic shock in the pediatric intensive care unit. This time frame is ideal for the goals of patient stratification, and identifies a sub-class of children (i.e. sub-class A patients) having a substantially higher illness severity and mortality (36% mortality vs. 11% mortality) than the other two sub-classes, with sensitivity and specificity $\geq 94\%$ [12]. The ability to reliably stratify patients within this timeframe could allow for the more selective enrollment of patients into clinical trials and for the selection of individual patients for currently available high risk therapies (e.g. ECMO). In the case of

clinical trial stratification, the strategy is potentially most useful when the test intervention carries more than minimal risk. For example, patients with a low risk of mortality, who otherwise meet clinical entry criteria, could potentially be excluded and thereby not be exposed to the risks of the test agent. Alternatively, patients with a high risk of mortality could be more reliably selected for enrollment. In either case, the net result of this type of approach would be to optimize the risk to benefit ratio of a given experimental therapy.

Second, the performance characteristics of the stratification strategy demonstrate the ability to reliably classify patients into the correct categories with a high level of sensitivity and specificity, and high positive and negative predictive values. In addition, the positive and negative likelihood ratios are well above or below cutoffs, respectively, that are considered to be clinically relevant [24].

This initial proof-of-concept study has important limitations from the standpoint of validation. We have not prospectively validated the existence of these three sub-classes with regard to the expression patterns and the related clinical phenotypes. Our original study that led to the identification of the 3 putative classes was conducted in a rigorous manner, but has not yet been prospectively validated in an independent cohort of patients. Accordingly, the veracity of these sub-classes, while biologically plausible, remains to be formally demonstrated. Related to this limitation, the current study was conducted using an internal cross validation approach based on the same patients as the original study. This 100 gene- and mosaic-based approach also requires formal prospective validation in a separate cohort of patients.

These limitations notwithstanding, we have taken an important first step in bringing genomic-based data closer to the bedside of critically ill children for the purpose of clinically relevant stratification: proof-of-concept. A major barrier to the application of genomic data in the field of critical care medicine has been the translation, interpretation, and acceptance of these complex data by clinicians. We propose that the mosaic-based strategy described in this report may potentially allow for the use of genomic data by clinicians at the bedside of critically ill children. Validation studies to directly test this proposal are currently in progress.

Acknowledgments

Contributing investigators and centers for genomic database that generated the original sub-classification report [12]: Thomas P. Shanley (C.S. Mott Children's Hospital at the University of Michigan, Ann Arbor, Michigan); Natalie Cvijanovich (Children's Hospital and Research Center Oakland, Oakland, CA); Richard Lin (The Children's Hospital of Philadelphia, Philadelphia, PA); Geoffrey L. Allen (Children's Mercy Hospital, Kansas City, MO); Neal J. Thomas (Penn State Children's Hospital, Hershey, PA); Douglas F. Willson (University of Virginia, Charlottesville, VA); Robert J. Freishtat (Children's National Medical Center, Washington, D.C.); Nick Anas (Children's Hospital of Orange County, Orange, CA); Keith Meyer (Miami Children's Hospital, Miami, FL); and Paul Checchia (St. Louis Children's Hospital, St. Louis, MO)

Supported by grants from the National Institutes of Health: R01GM064619 and RC1HL100474

REFERENCES

1. Marshall JC. Sepsis: rethinking the approach to clinical research. *J Leukoc Biol* 2008;83(3):471–482. [PubMed: 18171697]
2. Wheeler DS, Zingarelli B, Wheeler WJ, Wong HR. Novel pharmacologic approaches to the management of sepsis: targeting the host inflammatory response. *Recent Pat Inflamm Allergy Drug Discov* 2009;3(2):96–112. [PubMed: 19519586]
3. Wong HR, Cvijanovich N, Wheeler DS, Bigham MT, Monaco M, Odoms K, Macias WL, Williams MD. Interleukin-8 as a stratification tool for interventional trials involving pediatric septic shock. *Am J Respir Crit Care Med* 2008;178(3):276–282. [PubMed: 18511707]
4. Marshall JC, Reinhart K. Biomarkers of sepsis. *Crit Care Med* 2009;37(7):2290–2298. [PubMed: 19487943]

5. Vincent JL, Opal SM, Marshall JC. Ten reasons why we should NOT use severity scores as entry criteria for clinical trials or in our treatment decisions. *Crit Care Med* 38(1):283–287. [PubMed: 19730252]
6. Marcin JP, Pollack MM. Review of the acuity scoring systems for the pediatric intensive care unit and their use in quality improvement. *J Intensive Care Med* 2007;22(3):131–140. [PubMed: 17562737]
7. Osuchowski MF, Connett J, Welch K, Granger J, Remick DG. Stratification is the key: inflammatory biomarkers accurately direct immunomodulatory therapy in experimental sepsis. *Crit Care Med* 2009;37(5):1567–1573. [PubMed: 19325479]
8. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 1999;286(5439):531–537. [PubMed: 10521349]
9. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, Schreiber GJ, Peterse JL, Roberts C, Marton MJ, et al. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 2002;347(25):1999–2009. [PubMed: 12490681]
10. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* 2001;98(19):10869–10874. [PubMed: 11553815]
11. Rosenwald A, Wright G, Chan WC, Connors JM, Campo E, Fisher RI, Gascoyne RD, Muller-Hermelink HK, Smeland EB, Giltner JM, et al. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *N Engl J Med* 2002;346(25):1937–1947. [PubMed: 12075054]
12. Wong HR, Cvijanovich N, Lin R, Allen GL, Thomas NJ, Willson DF, Freishtat RJ, Anas N, Meyer K, Checchia PA, et al. Identification of pediatric septic shock subclasses based on genome-wide expression profiling. *BMC Med* 2009;7:34. [PubMed: 19624809]
13. Goldstein B, Giroir B, Randolph A. International pediatric sepsis consensus conference: definitions for sepsis and organ dysfunction in pediatrics. *Pediatr Crit Care Med* 2005;6(1):2–8. [PubMed: 15636651]
14. Wong HR, Shanley TP, Sakthivel B, Cvijanovich N, Lin R, Allen GL, Thomas NJ, Doctor A, Kalyanaraman M, Tofil NM, et al. Genome level expression profiles in pediatric septic shock indicate a role for altered zinc homeostasis in poor outcome. *Physiol Genomics*. 2007
15. Wong HR, Cvijanovich N, Allen GL, Lin R, Anas N, Meyer K, Freishtat RJ, Monaco M, Odoms K, Sakthivel B, et al. Genomic expression profiling across the pediatric systemic inflammatory response syndrome, sepsis, and septic shock spectrum. *Crit Care Med* 2009;37(5):1558–1566. [PubMed: 19325468]
16. Shanley TP, Cvijanovich N, Lin R, Allen GL, Thomas NJ, Doctor A, Kalyanaraman M, Tofil NM, Penfil S, Monaco M, et al. Genome-level longitudinal expression of signaling pathways and gene networks in pediatric septic shock. *Mol Med* 2007;13(9–10):495–508. [PubMed: 17932561]
17. Eichler GS, Huang S, Ingber DE. Gene Expression Dynamics Inspector (GEDI): for integrative analysis of expression profiles. *Bioinformatics* 2003;19(17):2321–2322. [PubMed: 14630665]
18. Guo Y, Eichler GS, Feng Y, Ingber DE, Huang S. Towards a holistic, yet gene-centered analysis of gene expression profiles: a case study of human lung cancers. *J Biomed Biotechnol* 2006;2006(5):69141. [PubMed: 17489018]
19. Tsigelny IF, Kouznetsova VL, Sweeney DE, Wu W, Bush KT, Nigam SK. Analysis of metagene portraits reveals distinct transitions during kidney organogenesis. *Sci Signal* 2008;1(49):ra16. [PubMed: 19066399]
20. Camphausen K, Purow B, Sproull M, Scott T, Ozawa T, Deen DF, Tofil PJ. Influence of in vivo growth on human glioma cell line gene expression: convergent profiles under orthotopic conditions. *Proc Natl Acad Sci U S A* 2005;102(23):8287–8292. [PubMed: 15928080]
21. Huang S, Eichler G, Bar-Yam Y, Ingber DE. Cell fates as high-dimensional attractor states of a complex gene regulatory network. *Phys Rev Lett* 2005;94(12):128701. [PubMed: 15903968]
22. Bjorkbacka H, Fitzgerald KA, Huet F, Li X, Gregory JA, Lee MA, Ordija CM, Dowley NE, Golenbock DT, Freeman MW. The induction of macrophage gene expression by LPS predominantly utilizes Myd88-independent signaling cascades. *Physiol Genomics* 2004;19(3):319–330. [PubMed: 15367722]

23. Fortina P, Surrey S. Digital mRNA profiling. *Nat Biotechnol* 2008;26(3):293–294. [PubMed: 18327237]
24. Puntmann VO. How-to guide on biomarkers: biomarker definitions, validation and applications with examples from cardiovascular disease. *Postgrad Med J* 2009;85(1008):538–545. [PubMed: 19789193]

Figure 1. Composite mosaics for sub-classes A, B, and C. The composite mosaics represent the average expression patterns of patients in each of the 3 respective sub-classes and served as the reference point for the cross validation procedures.

Figure 2.

Examples of individual mosaics for patients in sub-classes A, B, or C. The evaluators viewed 98 individual patient mosaics and classified each patient into one of the three sub-classes using the composite mosaics in Figure 1 as the reference point.

Figure 3.

A, Mosaics for 2 patients in original sub-class B that were incorrectly classified as sub-class A. Six of the 9 evaluators classified example 1 as sub-class A, and 7 of the 9 evaluators classified example 2 as sub-class A. **B**, Mosaics for 2 patients in original sub-class C that were incorrectly classified as sub-class A. All of the 9 evaluators classified example 1 as sub-class A, and 8 of the 9 evaluators classified example 2 as sub-class A.

Figure 4.

Mosaics for 4 patients in original sub-class B that were incorrectly classified as sub-class C. Eight of the 9 evaluators classified example 1 as sub-class C. Seven of the 9 evaluators classified example 2 as sub-class C. Five of the 9 evaluators classified example 3 as sub-class C. Seven of the 9 evaluators classified example 4 as sub-class C.

Figure 5. Mosaics for 2 patients that did not receive at least 5 “votes” for any 1 of the 3 sub-classes and therefore remained unclassified.

Table 1

Performance characteristics of the sub-classification strategy based on the consensus classifications of 9 individual evaluators.

	SUB-CLASS A	SUB-CLASS B	SUB-CLASS C
# in original group	28	45	25
# correctly identified (%)	27 (96)	38 (84)	23 (92)
# incorrectly identified (%)	4 (6)	0	4 (5)
Sensitivity (95% C.I.)	96% (80 – 99)	84% (70 – 93)	92% (72 – 99)
Specificity (95% C.I.)	94% (85 – 98)	100% (92 – 100)	94% (86 – 98)
Positive predictive value (95% C.I.)	87% (69 – 96)	100% (88 – 100)	85% (65 – 95)
Negative predictive value (95% C.I.)	99% (91 – 99)	88% (77 – 95)	97% (89 – 99)
Positive likelihood ratio (95% C.I.)	16.9 (6.5 – 43.8)	INFINITY	16.8 (6.4 – 43.8)
Negative likelihood ratio (95% C.I.)	0.04 (0.01 – 0.26)	0.2 (0.1 – 0.3)	0.08 (0.02 – 0.32)

Table 2a

Performance characteristics of the sub-classification strategy based on evaluator A.

	SUB-CLASS A	SUB-CLASS B	SUB-CLASS C
# in original group	28	45	25
# correctly identified (%)	27 (96)	35 (78)	17 (68)
# incorrectly identified (%)	4 (6)	6 (11)	9 (12)
Sensitivity (95% C.I.)	96% (80 – 99)	78% (62 – 88)	68% (46 – 84)
Specificity (95% C.I.)	94% (85 – 98)	89% (76 – 95)	88% (77 – 94)
Positive predictive value (95% C.I.)	87% (69 – 96)	85% (70 – 94)	65% (44 – 82)
Negative predictive value (95% C.I.)	99% (91 – 99)	82% (70 – 91)	89% (79 – 95)
Positive likelihood ratio (95% C.I.)	16.9 (6.5 – 43.8)	6.9 (3.2 – 14.8)	5.5 (2.8 – 10.8)
Negative likelihood ratio (95% C.I.)	0.04 (0.01 – 0.26)	0.3 (0.1 – 0.4)	0.4 (0.2 – 0.6)

Table 2b

Performance characteristics of the sub-classification strategy based on evaluator B.

	SUB-CLASS A	SUB-CLASS B	SUB-CLASS C
# in original group	28	45	25
# correctly identified (%)	26 (93)	37 (82)	18 (72)
# incorrectly identified (%)	8 (11)	3 (6)	6 (8)
Sensitivity (95% C.I.)	93% (75 – 99)	82% (67 – 91)	72% (50 – 87)
Specificity (95% C.I.)	89% (75 – 99)	94% (83 – 99)	92% (82 – 97)
Positive predictive value (95% C.I.)	76% (58 – 89)	92% (78 – 98)	75% (53 – 89)
Negative predictive value (95% C.I.)	97% (88 – 99)	86% (74 – 93)	91% (81 – 96)
Positive likelihood ratio (95% C.I.)	8.1 (4.2 – 15.7)	14.5 (4.8 – 43.9)	8.8 (3.9 – 19.6)
Negative likelihood ratio (95% C.I.)	0.08 (0.02 – 0.31)	0.19 (0.1 – 0.35)	0.3 (0.2 – 0.6)

Table 2c

Performance characteristics of the sub-classification strategy based on evaluator C.

	SUB-CLASS A	SUB-CLASS B	SUB-CLASS C
# in original group	28	45	25
# correctly identified (%)	26 (93)	35 (78)	20 (80)
# incorrectly identified (%)	7 (10)	3 (6)	7 (10)
Sensitivity (95% C.I.)	93% (75 – 99)	78% (62 – 88)	80% (59 – 92)
Specificity (95% C.I.)	90% (80 – 96)	94% (83 – 98)	90% (81 – 96)
Positive predictive value (95% C.I.)	79% (60 – 90)	92% (77 – 98)	74% (53 – 88)
Negative predictive value (95% C.I.)	97% (88 – 99)	83% (71 – 91)	93% (84 – 97)
Positive likelihood ratio (95% C.I.)	9.3 (4.6 – 18.9)	13.7 (4.5 – 41.7)	8.3 (4.0 – 17.3)
Negative likelihood ratio (95% C.I.)	0.08 (0.02 – 0.3)	0.2 (0.1 – 0.4)	0.8 (0.03 – 0.18)

Table 2d

Performance characteristics of the sub-classification strategy based on evaluator D.

	SUB-CLASS A	SUB-CLASS B	SUB-CLASS C
# in original group	28	45	25
# correctly identified (%)	28 (100)	38 (84)	23 (92)
# incorrectly identified (%)	5 (7)	0 (0)	4 (5)
Sensitivity (95% C.I.)	100% (84 – 99)	84% (70 – 93)	92% (72 – 99)
Specificity (95% C.I.)	93% (83 – 97)	100% (91 – 100)	95% (86 – 98)
Positive predictive value (95% C.I.)	85% (67 – 94)	100% (88 – 100)	85% (65 – 95)
Negative predictive value (95% C.I.)	100% (93 – 100)	88% (77 – 95)	97% (89 – 99)
Positive likelihood ratio (95% C.I.)	14 (6 – 32.6)	Infinity	16.8 (6.4 – 43.8)
Negative likelihood ratio (95% C.I.)	0.0	0.15 (0.08 – 0.31)	0.03 (0.01 – 0.11)

Table 2e

Performance characteristics of the sub-classification strategy based on evaluator E.

	SUB-CLASS A	SUB-CLASS B	SUB-CLASS C
# in original group	28	45	25
# correctly identified (%)	27 (96)	39 (87)	21 (84)
# incorrectly identified (%)	2 (3)	3 (6)	6 (8)
Sensitivity (95% C.I.)	96% (80 – 99)	87% (72 – 94)	84% (63 – 95)
Specificity (95% C.I.)	97% (89 – 99)	94% (83 – 98)	92% (82 – 97)
Positive predictive value (95% C.I.)	93% (76 – 99)	93% (79 – 98)	78% (57 – 91)
Negative predictive value (95% C.I.)	99% (91 – 99)	89% (77 – 95)	94% (85 – 98)
Positive likelihood ratio (95% C.I.)	33.7 (8.6 – 132.5)	15.3 (5.1 – 46.2)	10.2 (4.7 – 22.4)
Negative likelihood ratio (95% C.I.)	0.04 (0.01 – 0.1)	0.14 (0.07 – 0.3)	0.06 (0.02 – 0.15)

Table 2f

Performance characteristics of the sub-classification strategy based on evaluator F.

	SUB-CLASS A	SUB-CLASS B	SUB-CLASS C
# in original group	28	45	25
# correctly identified (%)	26 (93)	40 (89)	23 (92)
# incorrectly identified (%)	3 (4)	1 (2)	5 (7)
Sensitivity (95% C.I.)	93% (75 – 99)	89% (75 – 96)	92% (72 – 99)
Specificity (95% C.I.)	96% (75 – 99)	98% (89 – 99)	93% (84 – 97)
Positive predictive value (95% C.I.)	90% (71 – 97)	98% (86 – 99)	82% (62 – 93)
Negative predictive value (95% C.I.)	97% (89 – 99)	91% (80 – 97)	97% (89 – 99)
Positive likelihood ratio (95% C.I.)	21.7 (7.1 – 65.9)	47.1 (6.7 – 329.2)	13.4 (5.7 – 31.5)
Negative likelihood ratio (95% C.I.)	0.07 (0.02 – 0.28)	0.11 (0.05 – 0.26)	0.09 (0.02 – 0.32)

Table 2g

Performance characteristics of the sub-classification strategy based on evaluator G.

	SUB-CLASS A	SUB-CLASS B	SUB-CLASS C
# in original group	28	45	25
# correctly identified (%)	28 (100)	38 (84)	19 (76)
# incorrectly identified (%)	7 (10)	2 (4)	4 (5)
Sensitivity (95% C.I.)	100% (55 – 100)	84% (70 – 93)	76% (54 – 90)
Specificity (95% C.I.)	90% (80 – 96)	96% (86 – 99)	95% (86 – 98)
Positive predictive value (95% C.I.)	80% (62 – 91)	95% (82 – 99)	83% (60 – 94)
Negative predictive value (95% C.I.)	100% (93 – 100)	88% (76 – 95)	92% (83 – 97)
Positive likelihood ratio (95% C.I.)	10 (4.9 – 20.2)	22.4 (5.7 – 87.6)	13.9 (5.2 – 36.9)
Negative likelihood ratio (95% C.I.)	0	0.14 (0.07 – 0.28)	0.25 (0.13 – 0.51)

Table 2h

Performance characteristics of the sub-classification strategy based on evaluator H.

	SUB-CLASS A	SUB-CLASS B	SUB-CLASS C
# in original group	28	45	25
# correctly identified (%)	26 (93)	41 (91)	21 (84)
# incorrectly identified (%)	5 (7)	3 (6)	2 (3)
Sensitivity (95% C.I.)	93% (75 – 99)	91% (78 – 97)	84% (63 – 95)
Specificity (95% C.I.)	93% (83 – 97)	94% (83 – 99)	97% (90 – 99)
Positive predictive value (95% C.I.)	84% (66 – 94)	93% (80 – 98)	91% (70 – 98)
Negative predictive value (95% C.I.)	97% (89 – 99)	93% (81 – 98)	95% (86 – 98)
Positive likelihood ratio (95% C.I.)	13 (5.5 – 30.4)	16.1 (5.3 – 48.5)	30.7 (7.7 – 121.6)
Negative likelihood ratio (95% C.I.)	0.08 (0.02 – 0.29)	0.09 (0.04 – 0.24)	0.06 (0.02 – 0.15)

Table 2i

Performance characteristics of the sub-classification strategy based on evaluator I.

	SUB-CLASS A	SUB-CLASS B	SUB-CLASS C
# in original group	28	45	25
# correctly identified (%)	25 (89)	40 (89)	20 (80)
# incorrectly identified (%)	3 (4)	4 (8)	6 (8)
Sensitivity (95% C.I.)	89% (71 – 97)	89% (75 – 96)	80% (59 – 92)
Specificity (95% C.I.)	96% (87 – 99)	92% (81 – 98)	92% (82 – 97)
Positive predictive value (95% C.I.)	89% (71 – 97)	91% (77 – 97)	77% (56 – 90)
Negative predictive value (95% C.I.)	96% (87 – 99)	91% (79 – 96)	93% (84 – 97)
Positive likelihood ratio (95% C.I.)	20.8 (6.8 – 63.5)	11.8 (4.6 – 30.4)	9.7 (4.4 – 21.5)
Negative likelihood ratio (95% C.I.)	0.11 (0.04 – 0.33)	0.12 (0.05 – 0.27)	0.2 (0.1 – 0.5)

Table 3

Summary performance characteristics (medians and intra quartile ranges) of the sub-classification strategy based on individual evaluators.

	SUB-CLASS A	SUB-CLASS B	SUB-CLASS C
Sensitivity	93% (93 – 97)	84% (81 – 89)	80% (75 – 86)
Specificity	93% (90 – 96)	94% (93 – 96)	92% (91 – 95)
Positive predictive value	85% (78 – 89)	93% (92 – 96)	78% (75 – 83)
Negative predictive value	97% (97 – 99)	88% (85 – 91)	93% (92 – 95)
Positive likelihood ratio	14 (9.8 – 21)	15.3 (13 – 29)	10.2 (8.7 – 14.6)
Negative likelihood ratio	0.07 (0.03 – 0.08)	0.14 (0.12 – 0.19)	0.2 (0.06 – 0.32)