Critical Review Form

Clinical Prediction or Decision Rule

Prospective evaluation of a clinical decision guideline to diagnose spinal epidural abscess in patients who present to the emergency department with spine pain, J *Neurosurg Spine* 2011; 14:765-770

<u>Objectives:</u> "In this study, we prospectively evaluated this decision guideline in a cohort of patients who presented to the ED with spine pain. In addition, the diagnostic test characteristics of the ESR and CRP level in patients at risk for SEA (spinal epidural abscess) were determined". (p. 765)

<u>Methods:</u> This study had mixed methods. One component of the study was a prospective convenience sampling (16-hours/day) for a 9 month period at one ED with 45000 annual visits utilizing undergraduate research associates for all patients experiencing spine pain. The elements of the prospective data collection instrument are not described in the manuscript, nor are the timing of data collection and definitive diagnostic testing.

The second component of the study was a retrospective chart review for 9 years before and 5 years after a decision guideline was implemented in the hospital for SEA (1992-2005). The decision guideline is illustrated below.

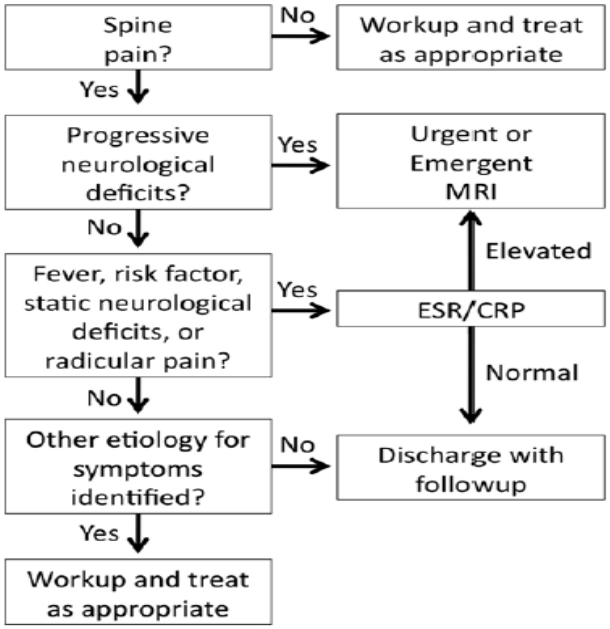


FIG. 1. Decision guideline for the ED diagnosis of spinal epidural abscess.

The authors do not detail whether a standardized data abstraction form was used, who abstracted the data or how SEA cases were identified retrospectively. During the 9 month period of prospective data collection in 2003 research associates also reminded clinicians of the decision guide which had also been presented to residents and faculty through a series of education sessions and intranet postings. SEA risk factors for ED spine pain patients included diabetes, IVDA, chronic kidney or liver disease, indwelling spine hardware, recent spine procedure, recent spine fracture, indwelling vascular catheter, immunocompromised, or other site of infection. Unfortunately, the authors do not provide standardized definitions for these risk factors nor do they state how these risk factors were identified by whom or when.

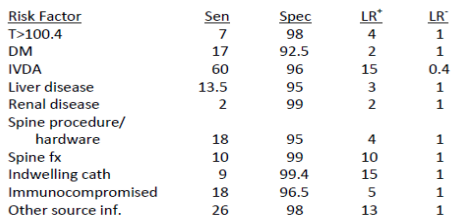
The primary outcome was the presence of a diagnostic delay defined as either multiple ED visits or admission to a nonsurgical service without diagnosis of SEA with subsequent SEA diagnosis during that admission. Physician adherence to the guideline was also assessed. A secondary outcome was an assessment of sensitivity and specificity for ESR and CRP using the prospective cohort of 9 months.

Guide		Comments
I.	**Is this a newly derived instrument (level IV)?**	
A.	Was validation restricted to the retrospective use of statistical techniques on the original database? (If so, this is a Level IV rule & is not ready for clinical application).	The SEA decision guide was not constructed using Stiell's principles. Were NSG, EM, and Radiology involved in deriving the instrument? Was any CART-modeling used? How reliable are clinicians at defining each variable? Therefore, this is a Level IV CDR.
II.	**Has the instrument been validated? (Level II or III). If so, consider the following:**	
1a	Were all important predictors included in the derivation process?	No. Besides the fact that individual risk factors within the decision guide are not defined and therefore at risk for subjective variability (what constitutes a "fever"? Who is "immunocompromised"? What is "chronic kidney disease"? How "recent" is "recent spine procedure" and does the procedure have to be invasive [chiropractic manipulation]?) the authors neglect important variables such as non-fracture related spine trauma, duration of back pain, alternative explanations for back pain, and clinician gestalt.
1b	Were all important predictors present in significant proportion of the study population?	Uncertain. Except for ESR and CRP the authors do not provide any details about frequency for each variable in the decision tree among those with or without SEA.

1c	Does the rule make clinical sense?	Yes, the decision guideline has content validity and appropriate clinical flow, albeit with important omissions noted in II-1A above.
2	Did validation include prospective studies on several different populations from that used to derive it (II) or was it restricted to a single population (III)?	The rule was not strictly validated over in this population. True validation would entail that clinicians first used the decision guideline recording their impressions for each decision-node before definitive diagnostic testing [MRI]. The outcome assessor who labels SEA as present of absent [Radiology if MRI criterion standard or NSG if operative criterion standard] needs to be blinded to the components and interpretation of the decision guideline. And all enrolled subjects would need to have the same criterion standard applied whether they did or did not have SEA. Therefore, everybody in the study would need to have an MRI which defeats the purpose of the guideline but is the only way to definitively prove that the guideline does not miss SEA cases that may self-resolve. To state that all SEA cases were identified when one only evaluates a portion of the cases with the criterion standard is ascertainment bias and overestimates the sensitivity of CDR.
3	How well did the validation study meet the following criteria?	

3a	Did the patients represent a wide spectrum of severity of disease?	Yes. (Table 2 page 768) Among SEA patients 17% had DM, 60% IVDA, 13% liver disease, 2% renal disease, 18% spinal procedure or hardware, 9% indwelling vascular catheter, and 18% were immunocompromised. The classic triad was present in 2.3% and 6% were hypotensive (sBP<90 mm Hg). The authors do not detail the duration of symptoms, history of prior back complaints, admission rates, severity of neurological deficits at presentation or discharge.
3b	Was there a blinded assessment of the gold standard?	Uncertain. The criterion standard to establish the diagnosis of SEA is not clearly described (who, what, when, how, blinded to ancillary clinical data?)
3c	Was there an explicit and accurate interpretation of the predictor variables & the actual rule without knowledge of the outcome?	No. Investigators did not provide explicit definitions for the variables of the decision guide in the manuscript and presumably did not do so while educating their clinicians prior to implementation. Each variable is therefore subject to clinician interpretation (see II-IA). The authors did not evaluate the reliability of the decision guide or its components (would two clinicians evaluate the same patient and come to the same diagnostic approach?), but future studies should do so.
3d	Did the results of the assessment of the variables or of the rule influence the decision to perform the gold standard?	Undoubtedly yes, since the criterion standard is MRI (probably) and the decision tree restricts MRI to a subset of patients.
4	How powerful is the rule (in terms of sensitivity & specificity; likelihood ratios; proportions with alternative outcomes; or relative risks or absolute outcome rates)?	• 1019 back pain patients were identified during the 9-month prospective data collection representing 3% of ED visits with a prevalence of SEA of 0.4% (4/1019) during this period. • ED clinicians followed the guideline in 90.9% of cases. • ESR (AUC 0.96) was superior to

		CRP (AUC 0.81) and 100% of SEA patients had EST>20 (vs. 33% of non-SEA patients). • CRP>1.0 in 87% of SEA and 50% non-SEA patients. • The interval LR's for ESR and CRP were not reported by the authors but can be computed from Fig. 4 and Fig. 5. ESR Range / ESR Int. LR / CRP Range / CRP Int. LR 0-20 / 0 / 0-5 / 0.22 40-60 / 0.89 / 5-10 / 0.67 >60 / 20 / >10 / ∞ 20-40 / 0.74 • In the retrospective review 86 pts. had SEA (55 before, 31 after). • Diagnostic delays were observed in 83.6% of pts before guideline implementation vs. 9.7% after guideline implementation. • Motor deficits were present in 81.8% of pts. at the time of diagnosis before guideline vs. 19.4% after guideline. • Risk factors were identified in 100% of SEA pts. vs. 23/3% of those without SEA. • The mean age of SEA pts. was 45-years and 60% were male. • Only 2% of SEA pts. had the classic triad of fever, spine pain, and neuro deficit at initial presentation. • The authors provide sufficient detail in Table 2 to compute the following diagnostic test characteristics:

		
III.	**Has an impact analysis demonstrated change in clinical behavior or patient outcomes as a result of using the instrument? (Level I). If so, consider the following:**	
1	How well did the study guard against bias in terms of differences at the start (concealed randomization, adjustment in analysis) or as the study proceeded (blinding, co-intervention, loss to follow-up)?	No formal impact analysis was performed (cluster randomized trial between centers) but before/after results suggest that the decision aid was associated with reduced diagnostic delays. Other factors that may have also reduced diagnostic delays include increased awareness of SEA 2º educational sessions, Hawthorne effect (awareness that SEA was being investigated), or process changes (more MRI availability, easier access to consultants, etc.).
2	What was the impact on clinician behavior and patient-important outcomes?	No formal impact analysis so not possible to confidently attribute clinician changes to guideline implementation (see III-I) however these observational results are encouraging. No patient-centric outcomes (i.e., functional recovery, long-term morbidity) are reported.

Limitations

1) **Failure to reference or employ Stiell's CDR methods so not confident that all significant predictor variables included in the decision guideline or that clinicians will accept, use, and interpret the guideline reliably and accurately.**

2) **Failure to reference or use established chart review methods.**

3) **Failure to reference or use STARD criteria for assessment of ESR and CRP.**

4) **Failure to clearly define criterion standard for SEA or describe proportions using various criterion standard strategies.**

5) **Failure to report sensitivities, specifications, LR^+, LR^-, or interval LR's in standard fashion.**

6) **Insufficient methods for prospective data collection. Who defined risk factors, when in the course of care, and using what definitions?**

7) **Failure to test for Hawthorne effect with time-stratified analysis.**

Bottom Line

The pre-test probability of SEA is ED patients with back pain is 0.4%. NO risk factors reduce the probability of SEA but a history of IVDA (LR^+ 15), an indwelling vascular catheter (LR^+ 15), and a source of infection (UTI, cellulites, pneumonia, etc., LR^+ 13) all increase the probability of SEA significantly. Only 2% of SEA patients have the classic triad of symptoms. Introduction of a diagnostic guideline was associated with a significant reduction in diagnostic delays in this single hospital retrospective review, but future multisite studies are needed to verify these results and assess clinician reliability and accuracy at interpreting the elements of this guideline.