Decisions making in health care is increasingly informed by the incorporation of knowledge considered within a probabilistic framework.\textsuperscript{23,29} New information derived from the history, physical examination, and other investigations is used to revise prior beliefs about the likelihood of a given diagnosis or outcome by a magnitude proportional to the relative strength of that information.\textsuperscript{10,14} The application of such methods, particularly within a diagnostic context, has been termed \textit{probabilistic reasoning}.\textsuperscript{11,33} Within a probabilistic framework, perfect predictions are not anticipated and error is knowingly accepted.\textsuperscript{12} The goal of probabilistic reasoning is therefore not to predict outcomes with certainty but, rather, to generate predictions that are more often “less wrong” than those generated by other methods.\textsuperscript{20}

Prior beliefs (prevalence of a diagnosis or outcome), also known as “pretest probability,” and new information (outcome of test with known sensitivity and specificity) may be mathematically integrated to produce the quantified probability of a given diagnosis or outcome (known as a “posterior probability” or “posttest probability”) using a well-known application of Bayes’ theorem.\textsuperscript{3}

\[ P(D/O) = \frac{Prevalence \times Sensitivity}{Prevalence \times Sensitivity + (1 - Prevalence) \times (1 - Specificity)} \]

(1) \( P(D/O) \)

Clinical prediction rules (CPRs) are a common application of probabilistic reasoning in health care and function to produce estimates of the likelihood of a target diagnosis, prognosis, or treatment outcome that, in turn, inform clinical decision making.\textsuperscript{16,25} For example, Hicks et al\textsuperscript{17} derived a CPR that functions to identify patients experiencing low back pain with a higher likelihood of achieving success from an 8-week stabilization-exercise program. They found that when 3 or more factors were present at baseline, including a positive prone instability test, aberrant movements, average straight leg raise greater than 91°, and age greater than 40 years, the probability of achieving a successful treatment outcome increased from 33% to 67%. A second example from the physical therapy literature is a CPR designed to identify patients with low back pain more likely to benefit from lumbar pelvic thrust manipulation. Flynn et al\textsuperscript{18} identified that the probability of success from this intervention increased from 45% to 95% when 4 or more factors were present, including

**SYNOPSIS:** Decision making in physical therapy is increasingly informed by evidence in the form of probabilities. Prior beliefs concerning diagnoses, prognoses, and treatment effects are quantitatively revised by the integration of new information derived from the history, physical examination, and other investigations in a well-recognized application of Bayes’ theorem. Clinical prediction rule development studies commonly employ such methodology to produce quantified estimates of the likelihood of patients having certain diagnoses or achieving given outcomes. To date, the physical therapy literature has been limited to the discussion and calculation of the point estimate of such probabilities. The degree of precision associated with the construction of posterior probabilities, which requires consideration of both uncertainty associated with pretest probability and uncertainty associated with test accuracy, remains largely unrecognized and unreported. This paper provides an introduction to the calculation of the uncertainty interval, known as a credible interval, around posterior probability estimates. The method for calculating the credible interval is detailed and illustrated with example data from 2 clinical prediction rule development studies. Two relatively quick and simple methods for approximating the credible interval are also outlined. It is anticipated that knowledge of the credible interval will have practical implications for the incorporation of probabilistic evidence in clinical practice. Consistent with reporting standards for interventional and diagnostic studies, it is equally appropriate that studies reporting posterior probabilities calculate and report the level of precision associated with these point estimates. \textit{J Orthop Sports Phys Ther} 2014;44(2):85-91. \textit{Epub} 30 October 2013. \textit{doi:10.2519/jospt.2014.4877}

**KEY WORDS:** Bayes’ theorem, clinical prediction rules, decision making, probability
duration of symptoms less than 16 days, at least 1 hip with more than 35° of internal rotation, lumbar hypomobility, no symptoms below the knee, and a score of less than 19 on the work subscale of the Fear-Avoidance Beliefs Questionnaire.37

Uncertainty in Clinical Prediction: Credible Intervals
To date, the physical therapy literature has been limited to the discussion and calculation of single values (point estimates) of posterior probabilities. When uncertainty is considered, it is frequently limited to test accuracy (sensitivity, specificity, likelihood ratio), with rare consideration of the precision of the pretest probability. The degree of precision associated with the construction of the posterior probability that is frequently reported in many CPR development studies remains largely unrecognized and unreported. This stands in contrast to the near-uniform reporting of confidence intervals around estimates of treatment effect and diagnostic test accuracy seen in the modern scientific literature.7,38 While it is accepted that confidence intervals reported in interventional studies have important implications for clinical practice,34 no such considerations have been given to the need for uncertainty intervals for posterior probability estimates published in the physical therapy CPR literature.

The precision of the posterior probability estimate requires consideration of the uncertainty associated with prevalence, as well as the uncertainty associated with test accuracy.14 The resultant uncertainty interval within a Bayesian framework is known as a credible interval (CrI) and gives the range in which the "true" likelihood of a given diagnosis or outcome lies for a specified probability level.26 Conceptually, it is the probability of a probability. For instance, the range of values expressed in a 95% CrI has a 95% chance of containing the true probability of a given outcome based on the information available.

FIGURE 1 illustrates the relationship between pretest probability and posttest probability for 6 likelihood ratio values. The slope of each curve and, therefore, the influence of a 1% change in pretest probability (x-axis) on posttest probability (y-axis) is defined by the following formula:

\[
\text{Slope} = \frac{\text{Likelihood Ratio}}{(\text{Pretest Probability} \times (\text{Likelihood Ratio} - 1) + 1)^2}
\]

Where the slope of this illustrated relationship is steep, small variations in the pretest probability have a large influence on the posttest probability. Conversely, where the slope is relatively flat, consideration of uncertainty in the pretest probability will only have a relatively small impact on the posterior probability uncertainty interval.4 The degree to which uncertainty in the likelihood ratio impacts the posterior probability is dependent on the magnitude of the pretest probability. Therefore, incorporation of uncertainty of both the pretest probability and the likelihood ratio is required to calculate the degree of uncertainty of the posterior probability.

Calculating the CrI
An appropriate method to calculate the CrI for a posterior probability estimate of a binary outcome is the objective Bayesian method using Monte Carlo simulation. This method takes into consideration uncertainty related to both pretest probability and test accuracy. Mossman and Berger35 identified this approach as having superior performance properties compared to other methods of uncertainty-interval calculation. Detailed below is a step-by-step guide as to how clinicians and researchers can perform this calculation using Microsoft Excel 2010 (Microsoft Corporation, Redmond, WA) and the statistical freeware R (http://www.r-project.org/). Each calculation step is illustrated using 2 separate examples from the physical therapy CPR literature. Of note, we have substituted the use of the Jeffreys prior used in Mossman and Berger’s35 original method with

![FIGURE 1](image-url)
Two-by-Two Table to Obtain Prevalence, Sensitivity, and 1 – Specificity

<table>
<thead>
<tr>
<th>Test Present</th>
<th>Outcome Present</th>
<th>Outcome Not Present</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test positive</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>Test negative</td>
<td>C</td>
<td>D</td>
</tr>
</tbody>
</table>


Table 2

Two-by-Two Contingency Table Derived From the Data by Hicks et al.¹⁷

<table>
<thead>
<tr>
<th>Treatment Successful (50% or Greater Improvement*)</th>
<th>Treatment Not Successful (Less Than 50% Improvement)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPR positive (3 or more variables present)³</td>
<td>10</td>
</tr>
<tr>
<td>CPR negative (fewer than 3 variables present)</td>
<td>5</td>
</tr>
<tr>
<td>CPR positive (4 or more variables present)</td>
<td>8</td>
</tr>
<tr>
<td>CPR negative (fewer than 4 variables present)</td>
<td>31</td>
</tr>
</tbody>
</table>

Abbreviation: CPR, clinical prediction rule.
*Percentage change in baseline and 8-week Oswestry Low Back Pain Disability Questionnaire score.
†Age less than 40 years, average straight leg raise greater than 91°, aberrant movement present, positive prone instability test.

Table 3

Two-by-Two Contingency Table Derived From the Data by Flynn et al.¹³

<table>
<thead>
<tr>
<th>Treatment Successful (Greater Than 50% Improvement*)</th>
<th>Treatment Not Successful (50% or Less Improvement)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPR positive (4 or more variables present)³</td>
<td>20</td>
</tr>
<tr>
<td>CPR negative (fewer than 4 variables present)</td>
<td>2</td>
</tr>
</tbody>
</table>

Abbreviation: CPR, clinical prediction rule.
*Percentage change in Oswestry Low Back Pain Disability Questionnaire score over 3 sessions.
†Symptom duration less than 16 days, at least 1 hip with greater than 35° of internal rotation, hypomobility with lumbar spring testing, no symptoms distal to the knee, Fear-Avoidance Beliefs Questionnaire work subscale score less than 19.

Example 1. Based on the published results of Hicks et al.⁷ for determining the likelihood of success for a stabilization exercise program for low back pain for when 3 or more variables were present at baseline, a 2-by-2 contingency table may be derived (Table 2). The following calculations may be derived from this table:

\( P_0 = x_0/n_0 = 18/54 (33\%) \)
\( P_1 = x_1/n_1 = 10/18 (56\%) \)
\( P_2 = x_2/n_2 = 5/36 (14\%) \)

Example 2. Using the results of Flynn et al.¹³ for identifying the likelihood of treatment success from lumbopelvic manipulation for low back pain for when 4 or more variables are present, a 2-by-2 contingency table may be derived (Table 3) that enables the following calculations of prevalence \( (P_0) \), sensitivity \( (P_1) \), and 1 – specificity \( (P_2) \):

\( P_0 = x_0/n_0 = 32/71 (45\%) \)
\( P_1 = x_1/n_1 = 20/32 (63\%) \)
\( P_2 = x_2/n_2 = 1/39 (3\%) \)

Step 2 The beta distribution is a type of continuous probability distribution that is employed in Bayesian analysis. The shape parameters \( a \) and \( b \) that will be used to construct beta distributions \( (\beta(a, b)) \) for prevalence, sensitivity, and 1 – specificity are calculated using the following: \( \beta(a, b) = \beta(x + 1, n – x + 1) \).

Example 1. Continuing our example for stabilization exercises for low back pain:

Prevalence: \( \beta(a, b) = \beta(18 + 1, 54 – 18 + 1) = \beta(19, 37) \)
Sensitivity: \( \beta(a, b) = \beta(10 + 1, 18 – 10 + 1) = \beta(11, 9) \)
1 – specificity: \( \beta(a, b) = \beta(5 + 1, 36 – 5 + 1) = \beta(6, 32) \)

Example 2. Using the data for lumbopelvic thrust manipulation for low back pain:

Prevalence: \( \beta(a, b) = \beta(32 + 1, 71 – 32 + 1) = \beta(33, 40) \)
Sensitivity: \( \beta(a, b) = \beta(20 + 1, 32 – 20 + 1) = \beta(21, 13) \)
1 – specificity: \( \beta(a, b) = \beta(1 + 1, 39 – 1 + 1) = \beta(2, 39) \)

Step 3 Use a random number generator to select a large number of values \( N \) (Mossman and Berger¹⁸ suggest 10,000) from each beta distribution. This can be performed using the statistical freeware R, with the following command to facilitate simple importing into Microsoft Excel (replace \( a \) and \( b \) with the relevant shape parameters calculated in step 2): write.csv(rbeta(10000, \( a \), \( b \)),).
drawn values for prevalence ($P_0$), sensitivity ($P_1$), and $1 - \text{specificity}$ ($P_2$) using the following formula:

$$\frac{P_0 \times P_1}{P_0 \times P_1 + (1 - P_0) \times P_2}$$

Note that using the point estimates of $P_0$, $P_1$, and $P_2$ in the above will provide the point estimate of the posterior probability.

- Example 1: For the stabilization-exercise example:

$$\frac{0.33 \times 0.56}{0.33 \times 0.56 + (1 - 0.33) \times 0.14} = 67\%$$

- Example 2: For the lumbopelvic thrust manipulation example:

$$\frac{0.45 \times 0.63}{0.45 \times 0.63 + (1 - 0.45) \times 0.03} = 95\%$$

**Step 5** Next, sort the resultant array of values in ascending order.

**Step 6** To determine the lower and upper boundaries of a 2-tailed 95% CrI, find the values corresponding to $N \times 0.025$ and $N \times 0.975$, respectively, in the sorted array.

- Example 1. For the stabilization-exercise example, the 95% CrI is 41% to 85%.
- Example 2. For the lumbopelvic thrust manipulation example, the 95% CrI is 77% to 99%.

Thus, for the stabilization-exercise example data on the likelihood of success of this intervention for low back pain, the corresponding 2-tailed 95% CrI is 41% to 85%. This means that, given the data, we can be 95% certain that the true probability of success from this exercise program for a patient presenting with 3 or more of the identified baseline predictors is between 41% and 85%. For the lumbopelvic thrust manipulation CPR example, the 95% CrI corresponding to positive status on this rule is 77% to 99%.

For those more experienced in using R, steps 3 to 6 of the method outlined above can be performed entirely within R using the syntax commands listed below. Computationally, this method is very quick and permits the use of an extremely large value for N (e.g., 1 million) to produce even more precise estimates of the posterior uncertainty interval, as the calculation remains in vector form and does not require exporting to a spreadsheet.

A useful adjunct to this method is that the mode of this distribution will correspond to the point estimate of the posterior probability. The intervals generated by this method for both the stabilization-exercise example data and the lumbopelvic thrust manipulation example data are almost identical (less than 1% different) to those generated by the objective Bayesian method using Monte Carlo simulation.

In the instance of zero false positives, a 1-sided interval appears preferable, such that the point estimate (100%) will be included within the uncertainty interval. To construct a 1-tailed 95% CrI in Microsoft Excel in the case of zero false positives, the following formula may be applied:

$$\text{Lower boundary: } \beta \text{.inv}(0.025, \text{true positive } + 1, \text{false positive } + 1)$$

$$\text{Upper boundary: } \beta \text{.inv}(0.975, \text{true positive } + 1, \text{false positive } + 1)$$

**Quick and Simple Approximations**

In situations where it may be appropriate to gather prevalence, sensitivity, and specificity data from within a single study (this excludes case-control designs), 2 methods may provide quick and relatively simple approximations of the uncertainty interval of the posterior probability that may be sufficiently accurate for clinical purposes in the majority of instances.

Tuyl\textsuperscript{42} proposed that adequate intervals may be constructed by using the beta(true positive + 1, false positive + 1) distribution. To calculate the 95% interval using this method, the following formulas may be used in Microsoft Excel:

Lower boundary: =beta.inv(0.05, true positive + 1, false positive + 1)

Upper boundary: =beta.inv(0.95, true positive + 1, false positive + 1)

An alternative method is to calculate the binomial proportion confidence interval of the positive predictive value: [True Positives/(True Positives + False Positives)]. The following command in the statistical freeware R will produce the relevant 95% interval using the Clopper-Pearson method:\textsuperscript{43}: 

```r
Lower boundary: =beta.inv(0.05, True Positive + 1, 1)
Upper boundary: =1
```
binom.test(True Positives, True Positives + False Positives)

This method produces uncertainty intervals that tend to be more conservative than the aforementioned methods\textsuperscript{28} but may nevertheless be adequate for clinical decision-making purposes. For the stabilization-exercise example data, this method provides an interval of 38\% to 88\%, and for the thrust manipulation example data, an interval of 76\% to 100\%.

A Brief Discussion on Selected Alternative Methods

It may seem intuitive to many to calculate the boundaries of the uncertainty interval of the posterior probability by combining the lower boundary of the pretest probability confidence interval with the lower boundary of the likelihood ratio confidence interval, and repeating this procedure for the upper boundaries. This method, however, produces uncertainty intervals that are generally much more conservative than the aforementioned procedures. For the stabilization-exercise example data, this method produces an interval of 30\% to 90\%, and for the lumbar pelvic thrust manipulation example data, an interval of 63\% to 100\%. Given the computational steps involved in completing this method, it is also unlikely to be considered simpler than the approximation methods previously outlined.

As already discussed, in some instances, uncertainty in the pretest probability will only have a very small influence on the uncertainty of the posttest probability. In such cases, simply combining the point estimate of the pretest probability with the upper and lower boundaries of the likelihood ratio confidence interval will produce an uncertainty interval similar to that produced by the methods already outlined. For the stabilization-exercise example data, this method produces an interval of 42\% to 83\%, and for the lumbar pelvic thrust manipulation example data, an interval of 74\% to 99\%. This method, however, may only give a suitable approximation when small variations in the pretest probability do not greatly influence the posttest probability across any portion of the pretest probability uncertainty interval. The influence of the pretest probability on the postprobability at a given likelihood ratio value may be calculated at its upper and lower uncertainty intervals using formula 2 presented above. Alternatively, it can be approximated by considering the slope of the corresponding parts of the likelihood ratio curves presented in FIGURE 1. Given the calculations required to check the appropriateness of this approximation method, it will be quicker and simpler in most cases to use the methods previously recommended.

Future Directions

The additional knowledge of the CrI surrounding a posterior probability point estimate has important implications concerning the appropriate clinical application of probabilistic evidence into practice. Rather than sole reliance on the point estimate, consideration of its precision, as highlighted by the corresponding CrI, provides substantially greater depth of information, which will assist decision making. Analogous to the use of confidence intervals reported for treatment effect sizes,\textsuperscript{34} the CrI of a posterior probability may be used to inform clinical decision making by considering its position relative to a “threshold level of certainty”\textsuperscript{33} required by a clinician to make a decision, such as commencing a particular treatment program or confirming or negating a diagnostic hypothesis. Such thresholds are dependent on the relative risks and benefits associated with that decision. In the context of treatment decision making, a clinician may have a relatively low threshold of certainty required for a treatment that is low cost and low risk, and may have a higher threshold of certainty required for a treatment that is associated with a higher risk of an adverse outcome and is more expensive and time consuming.

In the stabilization-exercise CPR example provided, it is plausible that the lower boundary of the uncertainty interval (41\%) may be below a “treatment threshold”\textsuperscript{38} many clinicians and patients would consider sufficient for an 8-week treatment program that requires 16 supervised sessions and daily home exercises. Within a probabilistic framework, further confirmatory evidence from the history and physical examination would be required for clinicians and patients to have confidence in the benefit that may be achieved from this intervention. To their credit, Hicks et al\textsuperscript{39} explicitly stated that the results of their study about the likelihood of success from a program of stabilization exercises represent the preliminary step in the development of a CPR, and they have not recommended that the derived tool be implemented in clinical practice. We have used the data from their derivation study to illustrate how clinical decisions informed by posterior probabilities may be influenced by the additional knowledge of the CrI.

In our second CPR example, concerning the probability of treatment success from lumbar pelvic thrust manipulation, the published results of Flynn et al\textsuperscript{13} enable the calculation of a CrI of 77\% to 99\%. In contrast to the stabilization-exercise program example, it is plausible that many clinicians and patients may perceive that the lower boundary of this uncertainty interval (77\%) is above a suitable treatment threshold\textsuperscript{38} for this intervention, given its relatively low cost, short time frame, and low risk of serious adverse events.

The primary method of calculating the CrI outlined in this paper (objective Bayesian method using Monte Carlo simulation) enables the incorporation of data from more than 1 study. Given that several independent studies are required in the development of a CPR,\textsuperscript{23} meta-analysis of such data may be helpful in the construction of more precise estimates of the posterior probability interval. For example, a recent systematic review\textsuperscript{50} of a CPR designed to help identify patients at risk of falling used a meta-analysis of the included studies to calculate more
precise estimates of the tool's sensitivity and specificity. These data were then used to help calculate more precise estimates of the posttest probability of a fall in patients who were either positive or negative on that CPR. To the best of our knowledge, published guidelines specific to the meta-analysis of CPR studies do not yet exist; however, many considerations concerning the appropriateness of pooling CPR accuracy data may be plausibly generalized from guidelines on the meta-analysis of diagnostic tests. Such considerations include the methodological rigor of included studies, as well as between-study variations in study design, study participants, CPR application, and the assessment of the reference standard. 6,19,20,24,30

CONCLUSION

This commentary focuses on the precision of posterior probability estimates in CPR development studies. This is a necessary, but not sufficient, consideration in the application of such evidence into clinical practice. Other factors beyond the scope of this commentary, including study design, methodological quality, validation, and impact analysis, require equally due consideration and have been well described in the existing physical therapy literature. 3,8,15,20-22,27 Consistent with reporting standards for intervention 26 and diagnostic 28 studies, we believe that it is equally appropriate that studies reporting posterior probabilities, as commonly practiced in CPR development studies, calculate and report the level of precision associated with these point estimates.

REFERENCES

GO GREEN By Opting Out of the Print Journal

JOSPT subscribers and APTA members of the Orthopaedic and Sports Physical Therapy Sections can help the environment by “opting out” of receiving JOSPT in print each month as follows. If you are:

- **A JOSPT subscriber:** Email your request to jospt@jospt.org or call the JOSPT office toll-free at 1-877-766-3450 and provide your name and subscriber number.

- **APTA Orthopaedic or Sports Section member:** Go to http://www.apta.org/, log in, and select My Profile. Next click on Email Management/GoGreen. Toward the bottom of the list, you will find the Publications options and may opt out of receiving the print JOSPT. Please save this preference.

Subscribers and members alike will continue to have access to JOSPT online and can retrieve current and archived issues anytime and anywhere you have Internet access.