



# J48-Based Algorithm in Predicting the Success Rate in the Board Examination

Jeffrey A. Clarin

Aklan State University, Philippines, jeffclarin@yahoo.com

## ABSTRACT

J48 algorithm was utilized in this study to predict the performance rate in the Board Examinations for Teachers Education program (LET) of the secondary teacher education graduates. To develop the model, the raw scores in the college entrance examination as well as the general average (GA) in the core subjects, professional subjects, and common subjects were considered. Likewise, the performance in the LET review and in the actual LET Board, whether pass or fail were also taken into consideration. A total of 348 examples was deliberate came from 2012 to 2017 list of graduates. The pruned tree summed up to 16 utilizing 10 leaves. The Kappa value was established to be 0.8195 meaning an almost perfect agreement. In other words, all of the categorized instances by the researcher-developed J48-based academic analytics model had closely matched the actual count data. Hence, predicted failures, could be promptly given appropriate intervention programs by the dean and the professors in order to enhance the scores and GA of students. By so doing, the success rates in the LET will be improved. The J48 algorithm was tested at Aklan State University (ASU) [1] due to the absence of a data-mining system that could predict success in the LET. Therefore, the sets of data used in this study were from ASU specifically in Kalibo Campus. In its entirety, this study might help in determining the success rates of graduates in the LET that could eventually help in the accreditation of teacher education programs.

**Key words:** prediction, analytics, data-mining, J48 algorithm, decision tree, KDD.

## 1. INTRODUCTION

The essential for 21st century abilities prompted the Philippine government to revise its educational learning, covered of 6 years in basic level and 4 years in secondary level. [2]. To warrant that Filipino graduates could have the basic skills and be academically equipped when joining the workforce, Republic Act No. 10533 otherwise known as the Enhanced Basic Education Act of 2013, was signed and

approved [3]. In this law, the Senior High School program was added. As a result of the addition of two more years in basic education, 81,000 qualified teachers were needed by the Department of Education (DepEd) in 2018 [4]. However, in order to be hired, teacher graduates must pass the LET [5].

In the Philippines, the Teacher Education Program were producing many graduates who flanked in the LET [6].

The foregoing prompted the researcher to develop a predictive academic analytics model to help students increase their success rates in the LET. Also, by using the researcher-developed J48-based academic analytics, more HEIs could be presented with simple yet conscientious decisions relative to the identification of teacher education students who are most likely to fail in the LET. This study was reckoned to be timely since the appraisal of students' success rates in the LET could help HEIs demonstrate that they are meeting high standards.

Moreover, academic studies in the country had revealed more LET flankers than passers [7]. Because of this, the researcher finds this study very timely too, due to the current modifications in the teacher education curriculum as an offshoot to the addition of two more years in senior high school program.

## 2. REVIEW OF RELATED LITERATURE

Baradwaj and Pal (2012) utilized data mining to learn the performance academically of students [8]. They collected the attendance, class test, seminar and assignment marks from a data set of 50 MCA (Master of Computer Applications) students of the VBS Purvanchal University, Jaunpur (Uttar Pradesh), India, from 2007 to 2010, in order to predict the student's performance by the end of each semester. They mentioned that at the time of their study, the three widely used decision tree learning algorithms were ID3, ASSISTANT and C4.5. They developed their data mining technique utilizing ID3. They reported that their study would help both students and teachers improve the academic performance. They further reported that their system could identify students who are in need for intervention and therefore help in the

reduction of failures via the application of appropriate action for the next semester examinations.

Mock board exam and general average in the common subjects and core subjects were found to be substantial in forecasting success in the LET. This was established in a study by Tarun, Gerardo, and Tanguilig III (2014) when they applied the PART and JRip classifiers of WEKA [9]. They established that if ever the pre board exam result was below 34% of the overall results, there would be a failure in the LET. Likewise, they forecasted failure in the LET for a fair or fairly good GA in the whole common subjects. Their study was conducted at Isabela State University and their respondents took a review and finally took the LET board examination on September 2013.

In the 2017 study of Dagdag, Sarmiento, and Ibale [10] they disclosed that a college degree program audit status might predict board exam result in major or professional subjects.

Rustia, Cruz, Burac, and Palaoag [11] studied five models of data mining algorithms in 2018, to find out which of these data mining algorithms could precisely predict student's success in the LET. These were: Logistic Regression, NN, SVM, Naïve Bayes, and C4.5 (J48) Decision Tree. The authors reported that among these models, the C4.5 (J48) accurately predicted success in the LET at a 73.10% accuracy level.

In the 2017 study of Undavia, Patel, and Patel [12], they reported that in order to maintain student growth and good reputation, HEIs had to come up with unique and creative ideas for creating a good image. They also disclosed that HEIs had adopted analytics-based methodologies to be more effective and that it became obligatory for HEIs to adopt analytics approaches with functional and practical values.

Caluza [13] created in 2018 a data mining model that forecasted competence among ICT teachers utilizing J48 algorithm. His model had a cross-validation technique that indicated a 92.78% accuracy with AUC weighted mean that was 92.4% correct. He said that his research could be benchmarked during the development of a software for the determination of competence among ICT teachers.

Al luhaybi, Tucker, and Yousefi (2018), attempted to apply J48 (C4.5) decision tree and Naïve Bayes classification algorithms to evaluate the performance of students in HEIs [14]. They tried to identify key features that affected the prediction process based on a combination of admission information, module-related data and first year final grades of the second year computer science student datasets at Brunel University London for the academic year 2015/16. They established that a predictive model could identify the low,

medium and high risks of failure of the students. They mentioned that this prediction model would help teachers to make appropriate interventions in order to assist high-risk students.

Lumauag [15] created in 2019 aims to develop a decision support model for personnel selection by using the C4.5 Algorithm. The decision support model was implemented by simulating the 110 applicant's record and it was evaluated in terms of accuracy, error rate, precision and recall.

J.Clarin, C.Sta. Romana and L.Feliscuzo (2019, p.193) and (2020, p. 143) [16], [17] had established a predictive analytics system using Decision Tree Algorithm. The main focus of their study was to forecast the performance of graduates in the teacher education program.

The above-cited studies had helped the researcher in the development of the J48-based academic analytics system that could forecast the performance in the LET. Furthermore, this current study had established that academic status in the common education subjects, professional subjects, university admission test results and interview results, might be used to determine the LET performance.

### 3. METHODOLOGY

**This study applies the Knowledge Discovery in Databases (KDD) principle [18]. KDD is a process of identifying valid and understandable patterns in data. According to the KDD principle, the discovered patterns are valid for a new ascertained data. The KDD process is presented in Figure 1.**

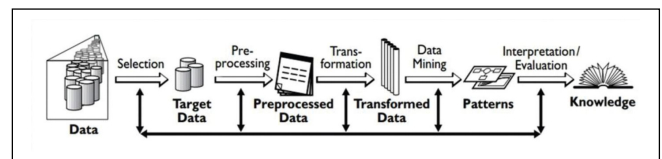


Figure 1: The Process of KDD

In this study, the classification algorithm used was J48-based. [19]. J48 algorithm was utilized to demonstrate and represent the phases when predicting success in the LET. The phases are shown in Figure 2.

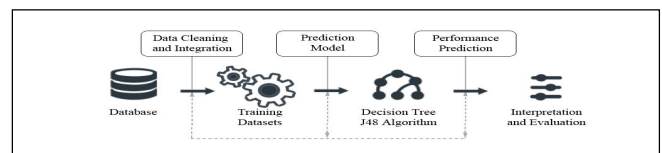


Figure 2: The J48-Based Algorithm

#### 3.1 The J48-Based Algorithm

J48-Based algorithm was applied in this research to develop the decision tree. The data set was divided into training and testing set. These data were classified as: university

admission test result; GA in core subjects, professional subjects and common subjects; participation in the LET review; and the result of LET board exam. The J48-Based algorithm was preferred based on Seringel [20], it can solve almost of the complications and problems in ID3. Moreover, J48 does not over-fit the data so it could create more generalized trees. Also, the J48 algorithm can transform from continuous to nominal attributes.

**Calculating for the global entropy**

The method used for the prediction of ‘will pass’ decision and process for the gain ratios are shown below.

1. Process for the calculation of global entropy for passed and failed decision:

$$Entropy(Set) = I(Set) = - \sum_{i=1}^n P(value_i) \cdot \log_2(P(value_i))$$

$$Entropy(Decision) = \sum -p(I) \cdot \log_2 p(I)$$

$$= -p(Passed) \cdot \log_2 p(Passed) - p(Failed) \cdot \log_2 p(Failed)$$

$$= -\left(\frac{a}{n}\right) \cdot \log_2 \left(\frac{a}{n}\right) - \left(\frac{b}{n}\right) \cdot \log_2 \left(\frac{b}{n}\right)$$

Where: a = instances for Passed decision  
 b = instances for Failed decision  
 c = instances for values in a nominal attribute  
 n = number of cases being studied

2. Calculation for the gain ratios

$$SplitInfo_A(c) = - \sum_{j=1}^v \frac{c_j}{|n|} \times \log_2 \left( \frac{|c_j|}{|n|} \right)$$

$$Gain(A) = Entropy(Decision) - SplitInfo(A)$$

$$GainRatio(A) = \frac{Gain(A)}{SplitInfo(A)}$$

**3.2 Parameters & Numerical Range Equivalent**

Table 1 displays the parameters set for the training data sets of this study and the numerical range equivalent. This was based on ASU grading system.

**Table 1:** Numerical Range Equivalent and its Parameters

| Entrance Exam Result       |                                     |                      |
|----------------------------|-------------------------------------|----------------------|
| Parameters                 | Grade Ratings /Numerical Equivalent |                      |
| Very Good                  | 90% – 100%                          |                      |
| Good                       | 80% – 89%                           |                      |
| Poor                       | 79% and below                       |                      |
| Academic Performance (GWA) |                                     |                      |
| Parameters                 | Grade Ratings                       | Numerical Equivalent |
| Very Good                  | 90% - 100%                          | 2.0 - 1.0            |
| Good                       | 80% - 89%                           | 3.0 - 2.1            |
| Poor                       | 79% and below                       | 3.1 - 5.0            |

Attributes Used

1. University Admission Test (EntranceExam): Tells the results of the incoming freshmen in the admission test.

2. Core Subjects (Specialization): the GA in all core subjects in the whole program.
3. Professional Subjects (ProfEd): the GA in all professional subjects in the whole program.
4. Common Subjects (GenEd): the GA in all common subjects in the whole program.
5. Board Review (LETReview): The participation during the Board Review either passed or failed (Passed/Failed).
6. LET Board Exam Result (Board Result): The outcomes of graduates during the LET Examination (Passed/Failed).

**3.3 On Establishing the Reliability of the**

**Researcher-developed J48-based Academic Analytics**

The reliability of the researcher-developed J48-based academic-analytics in forecasting the success in the LET was established using Kappa statistic. Its calculation followed the formula given below:

$$K = \frac{Pr(a) - Pr(e)}{1 - Pr(e)}$$

Where: K = stands for kappa  
 Pr(a) = simple agreement among rater  
 Pr(e) = likelihood that agreement is attributable to chance  
 1 = total number of rated items, also called cases

The degree of agreement actually achieved above that of chance is given in the numerator while in the denominator; the degree of agreement obtainable above chance is shown. The Kappa statistic [21] varies from 0 to 1, where:

- 0 = agreement equivalent to chance.
- 0.1 – 0.20 = slight agreement.
- 0.21 – 0.40 = fair agreement.
- 0.41 – 0.60 = moderate agreement.
- 0.61 – 0.80 = substantial agreement.
- 0.81 – 0.99 = near perfect agreement
- 1 = perfect agreement

To interpret the results, if k=1, then there is complete agreement among raters. However, if k≤ 0, then there is no agreement among raters. The reliability process involved the following steps:

- 1) Development of workmanship standards for rating criteria, i.e., very good and good (will pass the LET), and poor (will fail the LET).
- 2) Construction of academic analytics system.
- 3) Training the data sets.
- 4) Establishing system parameters.
- 5) System execution
- 6) Re-evaluation of the entire process.

#### 4. ETHICAL CONSIDERATION

The researcher developed J48-Based algorithm for the purpose and to be used by the management of HEIs. The system will not create a separate access to personal data of the students and teachers. The administrators of the system must keep the confidentiality of the privacy concerns of the students who will identify for intervention programs in order to increase the chances of succeeding in the LET.

#### 5. RESULTS

##### 5.1 Decision Rule and Decision Tree

Figure 3 shows the LET performance process of decision tree. Figure 4 presents the full training set or the classifier model highlighting the information that the university admission test/exams, GA of performance in academics, and participation in LET review, whether pass or fail, were substantial predictors of the performance in the LET.

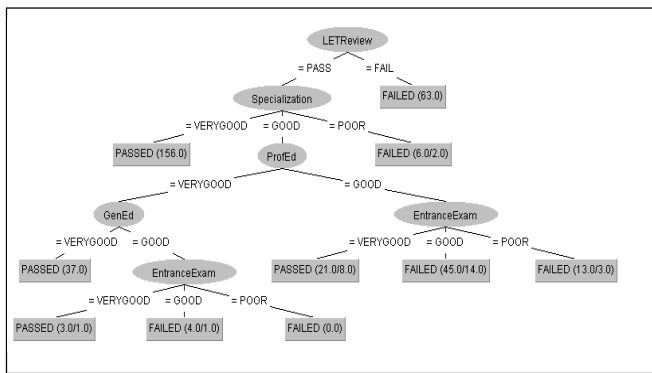


Figure 3: The LET process decision tree.

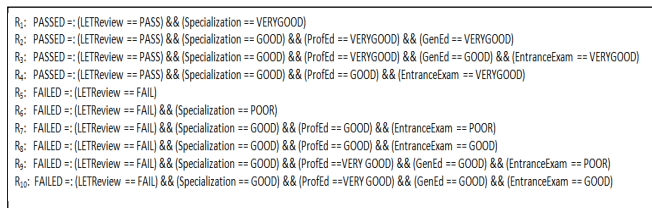


Figure 4: The LET condition decision rule.

##### 5.2 Success result in the LET board examination, when predicted in terms of: will pass or will fail

The assessment of training set summary is presented in Table 2.

Table 2: The Assessment of Training Set Summary

|                                  |           |           |
|----------------------------------|-----------|-----------|
| Correctly Classified Instances   | 319       | 91.6667 % |
| Incorrectly Classified Instances | 29        | 8.3333 %  |
| Kappa statistic                  | 0.8195    |           |
| Mean absolute error              | 0.113     |           |
| Root mean squared error          | 0.2377    |           |
| Relative absolute error          | 24.984 %  |           |
| Root relative squared error      | 49.9991 % |           |
| Total Number of Instances        | 348       |           |

As shown above, of the total 348 cases analyzed in this study, 319 or 91.67% was correctly classified. And, based on the generated Kappa value of 0.8195, which indicated a near perfect agreement, this study established that when the scores and GWA were high, a student would most likely pass the LET. This study also revealed that, should a student receive a score in academic subjects that is below 80% (rated as poor), with a GWA of below 3.0, (rated as poor); there is a strong likelihood that the student would fail the LET. In contrast, when a student gets a score of 90% and above (rated as very good) on the entrance exams, s/he will most likely pass the LET. But, if below 90%, a student is predicted to fail in the LET. Therefore, at this early, the management of teacher education program could already decide, whether to advise the concerned students to shift course or the management could already prepare for an outlined intervention program on areas where students were found to be weak, in order to help the students deter probable failures.

The J48-Based Algorithm Pruned Tree:

##### J48 Pruned Tree

```

LETReview = Pass
| Specialization = VERYGOOD: PASSED (156.0)
| Specialization = GOOD
| | ProfEd = VERYGOOD
| | | GenEd = VERYGOOD: PASSED (37.0)
| | | GenEd = GOOD
| | | | EntranceExam = VERYGOOD: PASSED (3.0/1.0)
| | | | EntranceExam = GOOD: FAILED (4.0/1.0)
| | | | EntranceExam = POOR: FAILED (0.0)
| | | ProfEd = GOOD
| | | | EntranceExam = VERYGOOD: PASSED (21.0/8.0)
| | | | EntranceExam = GOOD: FAILED (45.0/14.0)
| | | | EntranceExam = POOR: FAILED (13.0/3.0)
| | Specialization = POOR: FAILED (6.0/2.0)
LETReview = Fail: FAILED (63.0)
Number of Leaves : 10
Size of the Tree : 16
    
```

##### 5.3 Reliability of the researcher-developed J48-based algorithm, when classified in terms of: reliable or not reliable

With a precision of 0.959 for LET passers and 0.847 for those who failed in the LET, the researcher-developed J48-based academic-analytics system was found to be reliable. This is shown in Table 3.

Table 3: Accuracy Detailed by Class

|               | TP Rate | FP Rate | Precision | Recall | F-Measure | MCC   | ROC Area | PRC Area | Class  |
|---------------|---------|---------|-----------|--------|-----------|-------|----------|----------|--------|
|               | 0.912   | 0.075   | 0.959     | 0.912  | 0.935     | 0.821 | 0.975    | 0.984    | PASSED |
|               | 0.925   | 0.088   | 0.847     | 0.925  | 0.884     | 0.821 | 0.975    | 0.937    | FAILED |
| Weighted Avg. | 0.917   | 0.079   | 0.920     | 0.917  | 0.917     | 0.821 | 0.975    | 0.967    |        |

Based on the J48 confusion matrix that is presented in Table 4, a 91.23% correctness was produced from the researcher-established J48-based algorithm.



**Table 4:** Confusion Matrix

| Actual Class       | Predicted Class |        |                 |
|--------------------|-----------------|--------|-----------------|
|                    | PASSED          | FAILED | Percent Correct |
| PASSED             | 208             | 20     | 91.23%          |
| FAILED             | 9               | 111    | 92.50%          |
| Overall Percentage | 95.85%          | 84.73% | 91.67%          |

Table 4 also shows that there was an acceptable level of misclassifications in the Passed/Failed categories. This observation is supported by the statistically established Kappa value of 0.82 which meant a near perfect agreement. This is supported by Landis and Koch [22], who stated that a Kappa value of 0.81-1 is interpreted as almost perfect, and also by Fleiss, as cited by McHugh in 2012 [23], who said that a generated Kappa value of 0.81 and above means an almost perfect agreement. In other words, all of the instances classified by the researcher-developed J48-based academic analytics system had closely matched the actual count data, hence, the system was statistically established to be reliable.

## 6 CONCLUSION

The following conclusions were deduced from the findings in this study:

1. The university admission test/exams, GA in core subjects, professional/elective subjects and common subjects, participation in the LET review are important predictors of performance in the LET board examination.
2. The researcher-develop J48-Based algorithm was recognized to be dependable.

## 7 FUTURE WORKS

It remained established in this study that J48-Based algorithm might help the administration in refining the performance in the LET board examination. The researcher will seek approval for the implementation at ASU of this academic analytics system. Furthermore, the researcher will team up with IT experts of other HEIs offering teacher education programs in order to help students increase their success rates in the LET.

## REFERENCES

1. Aklan State University. Downloaded from <https://www.asu.edu.ph>
2. McDaid, Emily. (2018 May 15). *Do we need to change the way we teach?* Downloaded from <https://www.siliconrepublic.com/careers/4irc-21st-century-skill-s-education>.
3. Republic Act No. 10533. (2013 May 15) *An act enhancing the philippine basic education system by strengthening its curriculum and increasing the number of years for basic education, appropriating funds therefor and for other purposes.* Downloaded from <https://www.officialgazette.gov.ph/2013/05/15/republic-act-no-10533>.
4. UNTV News. (2017 August 16). *DepEd to address shortage of teachers and lack of classrooms.* Downloaded from <https://www.untvweb.com/news/dep-ed-address-shortage-teachers-lack-classrooms/>.
5. Republic Act No. 7836. (1994). *Philippine teachers professionalization act.* Downloaded from <http://www.chanrobles.com/republicactno7836.htm#>.
6. Baradwaj, B.K., and Pal, S. (2012). **Mining Educational Data to Analyze Students' Performance.** (*IJACSA*) *International Journal of Advanced Computer Science and Applications*, Vol. 2, No. 6. Downloaded from <https://arxiv.org/pdf/1201.3417>
7. David, C.C. and Ducanes, G. (2018 April 18). **Teacher education in the philippines: are we meeting the demand for quantity and quality?** UP CIDS Policy Brief Series 18-002. Downloaded from <https://issuu.com/up.cids/docs/higheredpolicybrief02-06DepEdRegion>.
8. Hemando-Malipot. (2018 June 6). *DepEd to put premium on hiring qualified teachers to teach in public schools.* Downloaded from <https://news.mb.com.ph/2018/06/06/dep-ed-to-put-premium-on-hiring-qualified-teachers-to-teach-in-public-schools/>.
9. Tarun, I.M.; Gerardo, B.D.; and Tanguilig III, B.T. (2014). *Generating licensure examination performance models using part and jrip classifiers: a data mining application in education.* Downloaded from [www.ijcce.org/index.php?m=content&c=index&a=show&catid=41&id](http://www.ijcce.org/index.php?m=content&c=index&a=show&catid=41&id).
10. Dagdag, J.D, Sarmiento, C.S., and Ibale, C.I. (2017). *Examining the factors of licensure examination for teachers performance for program strategy enhancement.* Downloaded from [www.apjmr.com/wp-content/uploads/2017/12/APJMR-2017.5.4.2.05.pdf](http://www.apjmr.com/wp-content/uploads/2017/12/APJMR-2017.5.4.2.05.pdf)
11. Rustia, R.A.; Cruz, M.A.; Burac, M.A.P.; and Palaoag, T.D. (2018). *Predicting student's board examination performance using classification algorithms.* Downloaded from <https://dl.acm.org/citation.cfm?id=3185101>
12. Undavia, J.N., Patel, S., and Patel, A. (2017). *Future trends and scopes of big data analytics in the field of education.* Downloaded from [https://www.researchgate.net/publication/318656487\\_Future\\_trends\\_and\\_scopes\\_of\\_Big\\_Data\\_Analytics\\_in\\_the\\_Field\\_of\\_Education](https://www.researchgate.net/publication/318656487_Future_trends_and_scopes_of_Big_Data_Analytics_in_the_Field_of_Education)
13. Caluza, L.J.B. (2018). **Predicting teachers' ict competence in a philippine university using j48 algorithm.** *Indian Journal of Science and Technology*, Vol 11(7), DOI:10.17485/ijst/2018/v11i7/119062, February 2018.
14. Al luhaybi, M.; Tucker, A.; and Yousefi, L. (2018). *The prediction of student failure using classification*

- methods: a case study.* Downloaded from Computer Science Department, Brunel University, London, UK <https://airccj.org/CSCP/vol8/csit88506.pdf>
15. Lumauag, R. G. **Decision Support System for Personnel Selection.** *International Journal of Recent Technology and Engineering*, 8(1), 177–179. Retrieved from <https://www.ijrte.org/wp-content/uploads/papers/v8i1/A2998058119.pdf>
  16. J. A. Clarin, C. C. Sta. Romana, L. S. Feliscuzo. **Academic Analytics: Applying C4.5 Decision Tree Algorithm in Predicting Success in the Licensure Examination of Graduates.** *2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS)*, (2019, p. 193-197).
  17. J. A. Clarin, C. C. Sta. Romana, L. S. Feliscuzo. **Academic Analytics: Predicting Success in the Licensure Examination of Graduates using CART Decision Tree Algorithm.** *Journal of Advanced Research in Dynamical and Control Systems*, Vol. 12, 01-Special Issue 2020 (p.143-151), DOI: 10.5373/JARDCS/V12SP1/20201057
  18. Fayyad, U., Piatetsky-Shapiro, G., and Smyth P. (1996). **From Data Mining to Knowledge Discovery in Databases**, *AI Magazine Volume 17 Number 3* (© AAAI). Downloaded from <https://studylib.net/doc/13819991/from-data-mining-to-knowledge-discovery-in-databases>.
  19. Serengil, S.I. (2018 May 13). *A step by step C4.5 decision tree example.* Downloaded from <https://sefiks.com/2018/05/13/a-step-by-step-c4-5-decision-tree-example/>.
  20. McHugh, M.L. (2012 Oct 15). *Interrater reliability: the kappa statistic.* Downloaded from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3900052/>.
  21. Landis, J.R. and Koch, G.G. (1997). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 1, 159-174.
  22. McHugh, M.L. (2012) *Interrater reliability: the kappa statistic.* Downloaded from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3900052/> and from NCBI Resources.. *Validity and inter-rater reliability testing of quality assessment instruments.* Downloaded from <https://www.ncbi.nlm.nih.gov/books/NBK92295/table/methods.t2/>.