APPENDIX D:

INTERRATOR RELIABILITY

INTERRATER RELIABILITY TESTING (IRR)

The IRR is the foundation for reliability for the WMSN. Reliability refers to the consistency or stability of measurement of the WMSN patient classification from user to user. Reliability is evaluated by having two individuals classify the same patient independently. A comparison of their classifications is then used to compute an index of equivalence or agreement between the classifiers.

The patient classification is completed by two independent raters (a unit nurse and an experienced classifier). Each nurse must classify independently of each other. The classifiers should use the same sources of information for their classification These include:

    a. Unit or hospital - specific SOPs.
    b. Department of nursing administrative procedures.
    c. DA Form 3888 and DA Form 3888-1.
    d. Nursing unit standards of care.
    e. Inpatient history and physical.
    f. DA Form 4677 and DA Form 4678.
    g. DA Form 4700.
    h. SF 511.
    i. SF 510.

Since patient acuity changes can sometimes be difficult to identify, the experienced classifier must classify the patient as closely as possible to the time the unit nurse classified the patient. When a significant amount of time has passed between the time the unit level nurse and the experienced classifier performed the patient classification, the potential for disagreement becomes greater. The experienced classifier must try to classify the patient in the same context as the original classifier. Since patient acuity changes can sometimes be difficult to identify, it is recommended that the raters classify patients within the same timeframe.

The goal of the IRR system is to maintain a minimum of 80 percent agreement. When reliability falls below 80 percent, actions to identify and correct the problems must be initiated. An IRR test will be done every month until a score of 80 percent is obtained. Actions may include staff discussions, inservice classes, 1-to-1 instruction with a specific staff member(s), and the development of new unit SOPs or standards of care to improve documentation.

While achieving an IRR score of at least 80% meets the criteria of the IRR program, nurses should look beyond the score to obtain even greater meaning regarding the strength of their

program. Nurses, especially head nurses, must look to see what amount of agreement was achieved in the selection of the critical indicators. Two separate nurse classifiers who classify the same patient into two separate categories may not have a significant amount of disagreement and yet classify the same patient into different categories. Nurses who classify patients with critical indicators which range on the border of categories may find category disagreements based on the selection or nonselection of one critical indicator. However, because of the numerical width of the specific category bands and the relative stability of the WMSN, the potential for this scenario is not great.

Conversely, nurses should not always be complacent when they have achieved a high degree of concurrence in their IRR score. The width of the category bands creates the potential for classifiers to achieve the same level of categorization on the same patient through the selection of multiple critical indicators. Head Nurses need to review the IRR classifications and assess the amount of agreement that was achieved during the classification of patients. This information will give the head nurse an even greater understanding regarding the relative strength of the IRR system within the unit.

NOTE: The IRR system can only succeed if the unit develops a strong IRR program which is supported and monitored by the nursing leadership. Nurses should note that classifying patients maintains the same level of legal accountability as other professional nursing practices, such as, administering medications. The intentional falsification of patient classification in order to achieve higher degrees of patient acuity is an illegal practice which can have profound ramifications for anyone choosing to perform this practice. (Do not confuse this issue with the honest differences which are expected during the routine classification of patients.)

INTER-RATER RELIABILITY (IRR) TESTING

In order for the patient classification process to generate accurate and useful information, an inter-rater reliability system must be used on a regular basis. Inter-rater reliability refers to the consistency or stability of patient classification from user to user. Reliability is evaluated by having two individuals classify the same patients independently on Patient Acuity Worksheets. This information is then used to compute an index of equivalence or agreement between classifiers.

Policies for Establishing and Maintaining IRR

1. The HN will see that IRR testing is conducted among all the staff nurses assigned to his or her unit. The question to answer is "Does everyone use the WMSN consistently?"

    a. It is recommended that intraunit IRR for all staff nurses on a unit be conducted annually.

    b. When new staff nurses are assigned to a unit, the HN will have one nurse from that unit experienced with the WMSN patient classification system established IRR with the new staff nurse. New staff members will not classify patients until they have established an IRR on the unit.

    c. The HN will periodically conduct IRR when ever he or she believes that there are unit problems with reliability.

2. Quarterly reliability testing is to be done on all nursing units involved in the WMSN by independent, experienced patient classifiers appointed by the nursing administration.

    a. The assignment of a permanent experienced classifier to a unit(s) is recommended. These objective classifiers must be familiar with the unit standards of care, but not assigned to that unit. The experienced classifiers must establish IRR with the unit HNs and among themselves.

    b. A unit may be tested more frequently if required or directed by MTF policy.

    c. Results of the reliability testing will be tabulated and shared with the unit staff and appropriate nurse managers.

    d. IRR testing will take place on different days of the week.

    e. An IRR score of 80 percent is the end goal. When reliability falls below 80 percent, actions to identify and correct the problems must be initiated. An IRR test will be done every month until a score of 80 percent is obtained. Actions may include staff discussions, inservice classes, 1-to-1 instruction with a specific staff member(s), and the development of new unit SOPs or standards of care to improve documentation.

Procedure for Conducting Quarterly IRR: (Note the IRR is currently automated and performs many of the following steps automatically.)

The following guidelines are to be followed by the experienced classifier when conducting an IRR test.

    a. Arrive unannounced on the unit as close to the time classifications are completed as possible.

    b. The number of patients randomly selected for this quarterly IRR is determined based on table 1. For units whose census is less than five, do IRR

testing over several days until five classifications have been tested for the quarter.

_____

Table 1. Sample size selections for quarterly IRR.

 Unit census   Number selected

   <= 5   All Patients
   >= 6 and <= 20   6
   >= 21 and <= 29   7
   >= 30 and <= 39   8
   >= 40   9

_____

     c. Use an accepted method to randomly select the patients to be used. All the patients should have an equal chance of being selected for IRR testing.

        1) One method would be to use a random numbers table such as the random numbers in table 2. Going in any direction from any point on the table produces a random sequence. To use this table, assign each patient on the unit a number, for example for a census of 36, the numbers would be 1 to 36. Go to the random numbers table and select a start point by closing your eyes and letting your finger fall at some point on the table. From this start point, select the first 8 numbers (the needed sample size) listed between 1 and 36. The patients that correspond to those numbers will be in the test sample.

        2) An alternative random selection method is call a systematic sample method. The classifier first chooses a number between one and the total unit census to be the start number. For example, on a unit with 15 patients, select a number between 1 and 15, such as 9. The rater then chooses a number between 2 and 10 to be the selection number, for example 3. Using a patient unit listing, the rater begins at the start number (the ninth listed patient and chooses every third patient listed. If going through the roster or list the first time does not meet the calculated sample size, begin again at the top of the roster or list, selecting every third patient; skipping over those patients already selected.

     d. Classify the selected patients using the source data identified on page two.

     e. Compare classifications between the unit staff and the experienced classifier. Discuss differences with staff members with staff members to determine the nature of the differences.

        1) Discrepancies caused by the intervening time period; that is, changes of orders.

        2) Misunderstandings, misinterpretations, or discrepancies caused by the experienced classifier's lack of knowledge or oversight of the documented source information. Note the use of a critical indicator that is not in accordance with the operational descriptions contained in this reference, the selection of a critical indicator that is not supported by an approved source, and the omission of a critical indicator that is supported by an approved source.

Table 2. Table of random numbers

```
07   84   04   09   42   69   10   41   28   93
07   59   70   03   93   78   06   41   72   93
34   67   93   60   65   04   17   06   98   11
01   63   24   74   18   20   49   28   91   15
48   05   60   34   42   00   70   47   98   68
85   70   55   46   27   40   78   96   41   30
34   76   84   78   91   36   37   43   47   38
47   93   81   01   88   11   31   70   53   24
91   14   66   66   44   47   23   58   07   67
53   15   14   35   11   83   12   63   88   92
90   39   54   41   32   70   17   07   90   53
19   74   90   70   44   97   54   09   60   77
55   08   82   76   91   28   62   44   89   45
01   83   31   72   91   24   02   05   77   36
73   91   48   64   97   25   48   78   39   72
17   64   24   09   56   28   18   96   13   28
27   72   02   33   22   77   36   07   95   95
96   15   32   23   17   57   16   71   04   25
51   79   70   85   18   61   01   87   48   78
44   98   27   35   20   56   18   95   67   56
15   31   01   59   22   91   04   21   49   99
42   53   43   68   99   05   39   54   48   70
35   17   32   21   94   24   84   68   71   67
63   38   65   04   21   84   94   36   88   20
00   98   88   78   69   82   82   89   92   79
23   91   31   52   56   63   29   31   21   86
34   32   02   16   46   04   84   05   02   97
46   21   71   60   73   17   83   60   96   04
78   27   56   66   90   25   93   90   38   46
59   23   60   33   61   58   64   22   18   38
00   15   56   89   67   75   22   83   36   43
19   14   02   88   73   10   64   75   22   66
21   59   28   97   03   09   19   29   20   49
51   98   40   32   98   90   89   34   66   93
82   72   31   86   67   53   04   41   08   43
20   57   84   61   53   23   16   00   53   94
18   75   11   02   75   87   48   89   42   27
31   64   74   87   08   86   74   13   76   72
73   71   67   30   28   43   65   16   51   49
43   19   61   91   27   64   59   22   86   51
29   45   46   09   66   29   30   78   29   58
17   19   14   98   40   49   57   95   91   93
63   95   66   70   53   87   26   01   44   77
89   44   51   79   40   45   10   39   73   28
33   78   56   99   82   65   17   81   74   08
59   68   03   79   00   11   95   55   37   28
85   92   34   31   13   19   39   87   60   57
17   03   90   65   21   27   32   13   76   90
82   73   97   57   25   41   23   10   92   84
87   61   33   08   02   79   18   56   73   52
```

APPENDIX   D - IRR