Verification & Validation Algorithms for Data Used in Critical Care Decision Support Systems

Debra Carlson, B.S.C.I.S., C. Jane Wallace, R.N. M.S.,
Thomas D. East, Ph. D., Alan H. Morris, M.D.
LDS Hospital, Pulmonary Division
8th Avenue C Street
Salt Lake City, Utah 84143

ABSTRACT

A decision support system is only as good as the data generating that decision support system. If the data is incorrect, doesn't relate to the other pieces of data, is missing or is not consistent, the decision support system conclusions may be incorrect and inconsistent. While collecting data from several sites during a multicenter randomized clinical trial, we found that some critical data elements were missing, out of correct ranges, totally illogical, and/or inconsistently recorded. In order to get consistent, correct, and dependable information from a our decision support system, the data elements used in that system had to be checked for completeness, valid values, consistent units of measurement, and relationships to other items. Development of data quality assurance rules and the application of those rules is imperative to using the data to generate daily scores for multiple organ failure, sepsis, and barotrauma.

INTRODUCTION

Data used for on-line, point of care decision support of a clinical process or measuring the outcomes of those processes requires systematic and exhaustive quality assurance. The accuracy of decision support scores, such as Apache II, generated using validated data is questionable at best.

When calculating decision support scores, how do you handle missing data elements, data elements with unbelievable values, data elements that do not match related data items, and unknown calculations from unknown sources? Our goal is to achieve a less than one percent error rate based on data completeness, calculations, valid ranges for values, and validating decision support scores generated by the data. We know of no standards for handling of missing data, out of range values, referential data integrity, and testing of clinical decision support scoring systems. Developing our own scheme for assuring the quality of our data used in calculating decision support scores was our only option. The data had to be complete, within clinically logical ranges, calculated uniformly, and consistent with related data items.

METHODS

The data entry system used is located at remote sites. Data is entered by clinical personnel into the computer. Several limitations exist in the current system. Little or no range checking is done at point of care due to slow system response. Variables we defined as required for describing a patient are allowed to be missing or zero. Forced entry of all values is not feasible because all values are not clinically indicated in every situation. Values related to each other are allowed to be logically incorrect. Calculations may or may not match calculations at other sites due to site specific requirements. In order to compare and relate the data from different sites, we had to make sure that the same set of data quality assurance rules were applied to all data regardless of site. The rules fall into four categories : Data Completeness, Calculations, Data within a Valid Range, and Referential Integrity.

1. Data Completeness: All required data fields contain valid data. In order to assure this when data elements are not available for a particular day we can calculate those that are calculable using standard formulas, fill with normal values, (where there are ranges of normal values, use the midpoint), or carry last value forward where applicable. For some values hospital specific averages are used when no values are entered. Normal values from the ARUP Users Guide January 1994[5].

0195-4210/95/$5.00

Table 1 : A Sample of Missing Data Rules Implemented

Conditions	Actions
Minimum or Maximum temperatures exist in Fahrenheit, but not in Centigrade	Calculate the centigrade temperatures temperature centigrade = (5/9) * (temperature Fahrenheit - 32)
Minimum or Maximum temperatures exist in Centigrade, but not in Fahrenheit	Calculate the Fahrenheit temperatures temperature Fahrenheit = (9/5)* (temperature centigrade + 32)
Minimum daily white blood count (min_wbc) is missing or not between .01 and 500 inclusive	Carry forward previous day's minimum wbc (or use hospital average if there is no minimum white blood count entered)
If Systemic Vascular Resistance (svr) not between 10 and 4000 inclusive or missing	Select minimum svr from the hemodynamics data table for the day of the sepsis barotrauma score. Be sure to use the svr not the svr index
If positive blood culture is missing	Default the positive blood culture to No
Maximum daily creatinine (max_creatinine) is null or marked unavailable (-999)	Set maximum creatinine to previous day's maximum creatinine value or normal if no maximum creatinine for the first scoring day.
Maximum daily bilirubin (max_bilirubin) is null or marked unavailable (-999)	Set maximum bilirubin to previous day's maximum bilirubin value or normal if no maximum bilirubin for the first scoring day.
Minimum daily platelet (min_platelet) is null or marked unavailable (-999)	Set minimum platelet to previous day's minimum platelet value or normal if no minimum platelet value for the first day.
Maximum daily PT (max_pt) is null or marked unavailable	Set maximum pt to previous day's maximum pt value
Maximum daily glascow coma score (max_gcs) is null or marked unavailable	Set maximum gcs to previous day's gcs or to 15 as a normal value if no glascow coma score for the first scoring day.
Maximum daily lipase (max_lipase) is null or marked unavailable (-999)	Set maximum lipase to previous day's maximum lipase value or 14 as normal if no lipase for the first scoring day.
Heparin is null or missing	Default heparin to 'N'

2. Calculations: Consistant and accurate calculations for derived values. For our own database consistancy, we used formulas in use at LDS Hospital. Patient data were run through the quality control program and compared with either a spreadsheet or native system calculation and a hand calculation. All differences were resolved by looking at the equations used to generate the compared values. Some values will be calculated only by our quality assurance programs. Others will be calculated only if they are missing. Listed in the Table of Equations are only those equations used in Hemodynamics Data Quality Assurance.

Table 2 - Table of Equations

Variable Name	Symbol	Units	Equation	Reference
Body surface area	BSA	m^2	$[\text{weight (kg)}^{0.425}]*[\text{height (cm)}^{0.725}]*71.84+10{,}000$	(1)
Arterial O_2 Content	CaO_2†	ml/dl	$\text{Hgb(gm/ dL)}*1.34(\text{mL/ gm})*(S_aO_2/100)+P_aO_2(\text{mmHg})*.0031$ (ml/mmHg)	(2) . p. 159
Venous O_2 Content	$C_{\bar{v}}O_2$ †	ml/dl	$\text{Hgb(gm/ dL)}*1.34(\text{mL/ gm})*(S_{\bar{v}}O_2/100)+P_aO_2(\text{mmHg})*.0031$ (ml/mmHg)	(2) , p. 160.
End Capillary O_2 Content	$C_{c'}O_2$ †	ml/dl	$(\text{Hgb(gm / dL)}*1.34)+P_AO_2*.0031$	(2) , p.159
Alveolar O_2 Partial Pressure	P_AO_2	mmHg	$[P_B\ (\text{mmHg}) - 47 * \%O_2/100] - [P_aCO_2\ (\text{mmHg}) * 1.25]$	(3) , p. 39
Shunt Fraction	$\dot{Q}_s/\dot{Q}_t$	none (ratio)	$(C_{c'}O_2 - C_aO_2)/(C_{c'}O_2 - C_{\bar{v}}O_2)$	(2) , p. 160
Arterial - Venous O_2 Content Difference	$C_{(a-\bar{v})}O_2$	ml/dl	$C_aO_2 - C_{\bar{v}}O_2$	(2) , p. 160
Oxygen Consumption	$\dot{V}O_2$	ml/min	$C_{(a-\bar{v})}O_2$ (ml/dl) * $\dot{Q}_t$ (L/min) * 10 (dl/L)	(2) , p. 160.
Oxygen Delivery	$\dot{D}O_2$	ml/min	CaO_2 (ml/dl) * $\dot{Q}_t$ (L/min) * 10 (dl/L)	(3) , p. 63-65
Systemic Vascular Resistance	SVR	Wood Units (mmHg/L/min)	[mean blood pressure (BP, mmHg) - right atrial pressure (RA or Central Venous Pressure CVP)]/ $\dot{Q}_t$ (L/min)	(4) , p. 257
Pulmonary Vascular Resistance	PVR	Wood Units (mmHg/L/min)	= [mean pulmonary artery pressure (PA, mmHg) - wedge pressure (Pw, mmHg)]/ $\dot{Q}_t$ (L/min)	(4) , p. 257
Cardiac Index	CI	L/min/m^2	= $\dot{Q}_t$/BSA	(4) , p. 253

†We are using 1.34 instead of 1.39 because COHb or MetHb are not always measured.

3. Data within a Valid Range: Valid data value ranges were defined by physicians analyzing approximately 260 ARDS patients' data for absolute ranges, normal values, and possible values. All values outside of our defined highs and lows are copied to an archive table, set to a null marker value (-999), and marked Quality Assured in a status table. Our ranges are absolute. We do not accept values outside of these ranges. (See Table 3).

Table 3 - Range Check Value List

Variable	Units	Low	High	Variable	Units	Low	High
Aa Gradient	mmHg	5	560/670	**pHa**		6.6	7.8
Amylase	u/dL	0	1000	**pHv**		6.6	7.8
AvDO2	mL/dL	1	15	**Platelet**	#/mm^3	10,000	500,000
Bands	%	0	100	**Potassium, serum**	mEq/L	1	8
Base Excess	mEq/L	-50	50	**Ppeak**	cmH2O	15	120
Bicarbonate, HCO3	mEq/L	5	60	**Pplat**	cmH2O	15	120
Bilirubin	mg/dL	0	50	**Pressure Support Level**	cm H2O	0	100
BUN (Serum Urea)	mg/dL	0	200	**PT**	sec	0	50
CaO2	mL/dL	4	30	**PT INR**		0.7	15
Cardiac Output	L/min	1	20	**PTT**	sec	0	150
Creatinine, serum	mMol/L	0	25	**PvCO2**	mmHg	10	200
Cth (Thoracic Compliance)	mL/cmH2O	1	100	**PvO2**	mmHg	10	100
CvO2	mL/dL	3	25	**Resp Rate**	1/min	2	80
CVP	mmHg	-5	60	**SaO2**	%	20	100
DBP	mmHg	15	200	**SBP**	mmHg	20	350
FiO2	%	21	100	**SIMV Rate (Setting)**	1/min	0	80
Glascow Coma Score		3	15	**Sodium, serum**	mEq/L	95	190
Heart Rate	1/min	20	300	**SvO2**	%	10	95
Height (cm)	cm	90	250	**SVR (dynes)**	dyne sec cm^-5	10	4000
Hematocrit	%	9	75	**SVR (Wood Units)**	mmHg/L/min	1	40
Hemoglobin	gm/dL	2	25	**Temperature Centigrade**	Degrees	25	45
I:E Ratio		1:5	4:1	**Temperature Fahrenheit**	Degrees	77	113
Lipase	u/dL	0	1500	**Total Calories**	kcal/day	0	5000
MBP	mmHg	20	180	**Total Carbohydrates**	gm/day	0	1000
Mean Airway Pressure	cm/H2O	5	100	**Total Fat**	gm/day	0	500
PaCO2	mmHg	10	200	**Total Protein**	gm/day	0	200
PaDiastolic	mmHg	0	80	**Urinary Output**	L/day	0	24
PaMean	mmHg	5	80	**VE Measured**	L/min	0	25
PaO2	mmHg	10	500	**VE Setting**	L/min	0	25
PaO2/FiO2 Ratio		10	200	**VO2**	mL/min	50	1500
PaSystolic	mmHg	5	100	**VT**	mL	0	1500
Peak Flow	L/min	5	120	**VT (during CPAP mode)**	mL	0	800
PEEP Setting	cmH2O	0	60	**WBC**	1000/mm^3	0.01	500
PEEP (Patient)	cm H2O	0	60	**Wedge (PAOP, Pw)**	mmHg	-2	60
				Weight (kg)	kg	20	200

4. Referential Integrity : - Certain data elements are related and must be referentially correct. For Example, peak pressure cannot be less than plateau pressure. Hospital admission dates cannot be in the future. Discharge dates cannot be prior to admission dates. (See Table 4).

Table 4 - Sample of Referential Integrity Rules Applied

Peak Pressure less than Plateau Pressure	Null Plateau Pressure
Hospital Discharge prior to Hospital Admit Date	Do not allow entry into system
Age less than 12 years	Do not allow entry into system.
Measured Ventilatory Rate less than Ventilatory Rate Setting	Null Measured Ventilatory Rate
If Ventilatory Rate Setting & Tidal Volume Setting equal 0 and Ventiltory RateMeasured and Tidal Volume Measured greater than 0 and Ventilation Mode is not CPAP and Ventilation Mode prior to and after current measurement are CPAP	Change Ventilation Mode to CPAP

RESULTS

Calculations and programs were extensively tested until all parties were satisfied with the results. Prior to applying the data quality assurance rules, we were never sure that the scores were accurate, statistics were correct, or that we could depend on the descriptive data contained in our database. Most of the referential integrity rules are logical checks applied to insure data validity. Ranges were developed by analyzing patient data highs, lows, averages, standard deviations, normal ranges, and therapeutic limitations to arrive at absolute ranges for data acceptance. Calculations are standard calculations from ventilators and LDS Hospital lab value calculations. Using C with embedded SQL, we programmed the calculations into our data quality assurance programs. Validation of the output from these SQL query programs involved comparing hand calculations, program output, and calculations captured from the on-site computer database. In the case of Systemic Vascular Resistance (SVR) and Pulmonary Vascular Resistance (PVR), there was a problem with the remote site's units of measurement in the calculations, so we decided not to accept any of the remote site's calculations for those values. In order to compare data from the two sites, equivalent units of measurement must be used. We demonstrated less than one percent error rate between the database calculations and the alternate computer calculations done at remote sites for data elements other than SVR and PVR (Table 5). Our scoring systems are acting on accurate date, thus accurately scoring patients, our statistics accurately reflect our patient database, and we can depend on the descriptive data contained in our database.

Table 5 - Sample Calculation Validation Comparison

	Hand Calc.	SQL Query (*hemocqi*) calc.	Alternate Computer Calculation
CaO_2 Hgb= 10.7 gm/dl $SaO_2/100$= 0.869 PaO_2= 65 mmHg	12.4597 + 0.2015 = **12.6612**	**12.65**	**12.6612**
Hgb= 10.9 gm/dl $SaO_2/100$= 0.887 PaO_2= 55 mmHg	12.9555 + 0.1705 = **13.126**	**13.126**	**13.1205**
Hgb= 12.5 gm/dl $SaO_2/100$= 0.91 PaO_2= 62 mmHg	15.2425 + 0.1922 = **15.4347**	**15.4347**	**15.4285**
$C_{\bar{v}}O_2$ Hgb= 12.5 gm/dl $SvO_2/100$= 0.69 PvO_2= 36 mmHg	11.5575+ 0.1116 = **11.6691**	**11.6691**	**11.6655**
$C_{c'}O_2$ P_B = 760 $\%O_2/100$ = 0.40 PaCO2 = 31	246.45 (PAO2) = (760-47)*0.4 - (31 * 1.25) **17.514** ($Cc'O_2$) = 16.75 + 0.7640	**17.514**	**17.512**

Table 5 - Sample Calculation Validation Comparison

	Hand Calc.	SQL Query (*hemocqi*) calc.	Alternate Computer Calculation
Qs/Qt CaO_2= 15.4347 $C\bar{v}\,O2$ = 11.6691 $Cc'O_2$ = 17.514	2.0793/5.8449 = **0.3557**	**0.35575**	**0.35637**
$C(a\text{-}\bar{v})O2$ CaO_2= 15.4347 $C\bar{v}\,O2$ = 11.6691	**3.7656**	**3.7656**	**3.763**
VO2 $C(a\text{-}\bar{v})O2$ = 3.7656 Q t = 9.8	**369.0288**	**369.029**	**368.774**
DO2 CaO_2 = 15.4347 Q t = 9.8	**1512.6006**	**1512.6**	**1511.99**
SVR mean BP = 82 RA = 18 Q t = 9.8	**6.5306**	**6.53061**	No values accepted from remote sites
PVR mean PA = 31 wedge = 18 Q t = 9.8	**1.3265**	**1.32653**	No values accepted from remote sites

CONCLUSION

While the quality of most of the data entered by clinical personnel is good, critical elements for decision support can be incorrect, imprecise, missing, or not belong to the rest of the data set. In order to use the data collected for clinical decision support, data quality assurance rules must be implemented. Range checking, referential integrity checking, standard calculations, and standard units of measurements are crucial when comparing data from vastly different locations. The time to apply these rules is, of course, at the point of care, but, at very least, before the data is used for clinical decision support or data analysis.

References

1. DuBois & DuBois. Archives of Internal Med. 1916;17:863.

2. Morris AH, Kanner RE, Crapo RO, Gardner RM. Clinical pulmonary function testing: A manual of uniform laboratory procedures. (2 ed.) Salt Lake City: Intermountain Thoracic Society, 1984

3. Sykes MK, McNicol MW, Campbell EJM. Ventilation and gas exchange. In: Respiratory Failure. 2 ed. Oxford: Blackwell Scientific Publications, 1976

4. Grossman W, Barry WH. Cardiac catheterization. In: Braunwald E, ed. Heart Disease: A Textbook of Cardiovascular Medicine. Philadelphia: W.B. Saunders, 1988

5. Associated Regional and University Pathologists, Inc., Users Guide January 1994.

Acknowledgments : This work was supported by AHCPR grant #HS06594.