

EPSDT Inter-Rater Reliability Analysis 2004-2005

**Prepared for the Early Periodic Screening, Diagnosis and Treatment Program of the
Maryland Department of Health and Mental Hygiene
by the Center for Health Program Development and Management at UMBC**

Background

The Early and Periodic Screening and Diagnostic Testing (EPSDT) Program is a Medicaid-funded initiative that provides guidance for the preventive, primary care for eligible children. Providers are certified through the Department of Health and Mental Hygiene to provide EPSDT services. As EPSDT-certified providers, and to be eligible for payment for services under this program, providers are required to adhere to a set of age specific-standards that include components related to history, physical examination, diagnostic tests, and age-appropriate anticipatory guidance, including the documentation of these activities.

To assure compliance with the programs requirements, a group of nurses are charged with completing annual record reviews in the practices of certified providers. Upwards of 3,000 records are reviewed each year. The quality of the nurse reviews is monitored by assessing the consistency of the nurses' ratings of providers through an inter-rater reliability evaluation. The Center for Health Program Development and Management was asked to conduct this evaluation.

Methodology

Up to thirty records were used for each analysis, selected by having each nurse acquire a record from each of five provider practices within the nurses' respective jurisdictions. The copied charts were brought to the EPSDT offices, where each of the five nurses reviewed all thirty records. Three age groups are included in this analysis: 4-5-month-olds, 2-year-olds, and 14-15-year-olds.

Using the "Healthy Kids Program Performance Improvement Guide for Clinical Record Reviews," each nurse reviewed the records for "complete," "incomplete," or "missing" indications of rendered care, and entered a score of 2, 1, or 0, respectively, for each element directly into a database on a laptop computer.

The charts were each reviewed for age-specific elements that fall into five categories: health and developmental history, comprehensive unclothed physical exam, laboratory tests, immunizations, and health education. A percentage and Kappa score was calculated for each pair of nurses, for each element, averaged for each category, and overall for each age group. It is important to note that paired-inter-rater reliability does not assess the accuracy of the nurses' evaluation, but rather the extent to which they agree in their record reviews. A known

standard was not used against which the nurses' accuracy could be tested. Therefore, variation among the nurses could reflect variation in the skill levels of nurses, but not indicate which nurse is better than others. What can be detected are which elements and which nurse pairs have better agreement.

High levels of agreement could mean that both nurses in a given pair are very good, or equally not as good at performing the record reviews. Low agreement could indicate that nurses are very uneven in their abilities to review records, or that documentation of the review element is difficult to locate in the medical record. For example, graphed height and weight are very standardized; almost all practices use formal height and weight charts that are often color coded (blue for boys and pink for girls) so that they are highly visible and easier to find. On the other hand, assessed hearing in an infant is probably not consistently done among providers and the notation of it is not generally standardized, so that finding it in the record is difficult. It would be expected that elements that nurses find and record as such are definitely present. However, recording that an element is missing does not necessarily mean that it was not there, but rather, that it was not found.

Assessing inter-rater reliability looks at the extent to which items are:

1. Found by both nurses—positive, definite identification **→ AGREEMENT**
2. Found by neither nurse **→ AGREEMENT**
3. Found by one nurse—negative or false negative finding (the element was not there or not identified) **→ NO AGREEMENT**

This analysis was performed using Excel to compute arithmetic percentages, and SPSS to calculate Cohen's Kappa, a measure of agreement between pairs, which incorporates an adjustment for chance agreement.

Cohen's Kappa

Cohen's Kappa is widely used in health services research and is specifically suited to studies of inter-rater agreement. The score ranges from -1.0 to 1.0, where -1.0 indicates no agreement and 1.0 indicates perfect agreement. Kappa scores are considered to indicate "substantial agreement" when they are 0.61-0.80, and "nearly perfect agreement" from 0.81-1.00. (See Table 1.)

Table 1: The interpretation of the Kappa Score

Kappa Score	Extent of Agreement
-1.0 to 0	No agreement or poor agreement
0 to .20	Slight agreement
.21 to .40	Fair agreement
.41-.60	Moderate agreement
.61-.80	Substantial agreement
.81-1.0	Nearly perfect agreement

Percentage of Agreement

Calculation of the percent of times out of a total number of assessments that raters apply the same score, gives a simple arithmetic valuation of extent of agreement. While the percent provides a more readily understood conceptualization of degree of agreement, it may be less precise than statistical measures methods that can account for a degree of chance, such as the Kappa. The percentage of agreement can be calculated by comparing each rater with each of the other raters, or by comparing each rater against a known or expert standard. In this case, the 5 nurse reviewers were paired with each of other four, creating ten unique pairs.

Age-Specific Analyses: 4-5 Month Olds, 2 Year Olds, and 14-15 Year Olds

4-5-Month-Old Analysis Using Cohen's Kappa

Each of the five nurses was paired with her four colleagues, creating a total of 260 pairs (10 pairs and 26 scoring elements for each pair), and 104 pairs for each nurse (four pairings with 26 scoring elements each). Of the 260 pairs, 166 (64 percent) were 0.61 or more, showing substantial or nearly perfect agreement.

Seven of 26 elements had all nurse pairs scoring greater than 0.61, including:

- Recorded peri-natal history
- Documentation of a minimum of five systems
- Measured head circumference
- Graphed head circumference
- Measured height
- Measured weight
- Conducted second hereditary/metabolic screening at 2-4 weeks

The remaining 19 elements had 7 or fewer pairs (out of 10 pairs) with Kappa values of 0.61 or better. (See Table 2.)

Table 2: Number of Nurse-pairs with Kappa Scores Greater than or Equal to .61 (out of 10 pairs)

	ELEMENT	Number of Nurse-Pairs
Health and Developmental History	Recorded Health History	3
	Recorded Family History	5
	Recorded Peri-natal History	10
	Recorded Psycho-social History	2
	Recorded Immunization History	6
	Recorded Developmental Assessment/History	7
Comprehensive Unclothed Physical Exam	Documentation of Minimum 5 Systems	10
	Assessed Vision	3
	Assessed Hearing	1
	Assessed Nutritional Status	6
	Conducted Oral Screening	3
	Measured Height	10
	Graphed Height	3
	Measured Weight	10
	Graphed Weight	6
	Measured Head Circumference	10
	Graphed Head Circumference	10
Laboratory	Conducted Second Hereditary/Metabolic Screening at 2-4 Weeks	10
Immunizations	Hepatitis B Vaccine(s) Per Schedule	1
	DTP/DTAP(DT Vaccine(s) Per Schedule	10
	Hib Vaccine(s) Per Schedule	6
	Polio Vaccine(s) Per Schedule	3
	Assessed if immunizations UTD	3
Health Education	Specified Requirements for Return Visit	8
	Provided Ed/Referral for Identified Problems/Tests	5
	Provided Age Appropriate Guidance	7

Sixteen elements have Kappa scores greater than 0.61, and 10 elements have scores below 0.61, falling in the poor to moderate agreement range. (See Figures 1 and 2.)

Figure 1: Elements with an Average Kappa Above 0.61 Indicating Substantial to Nearly Perfect Agreement (16 out of 26)

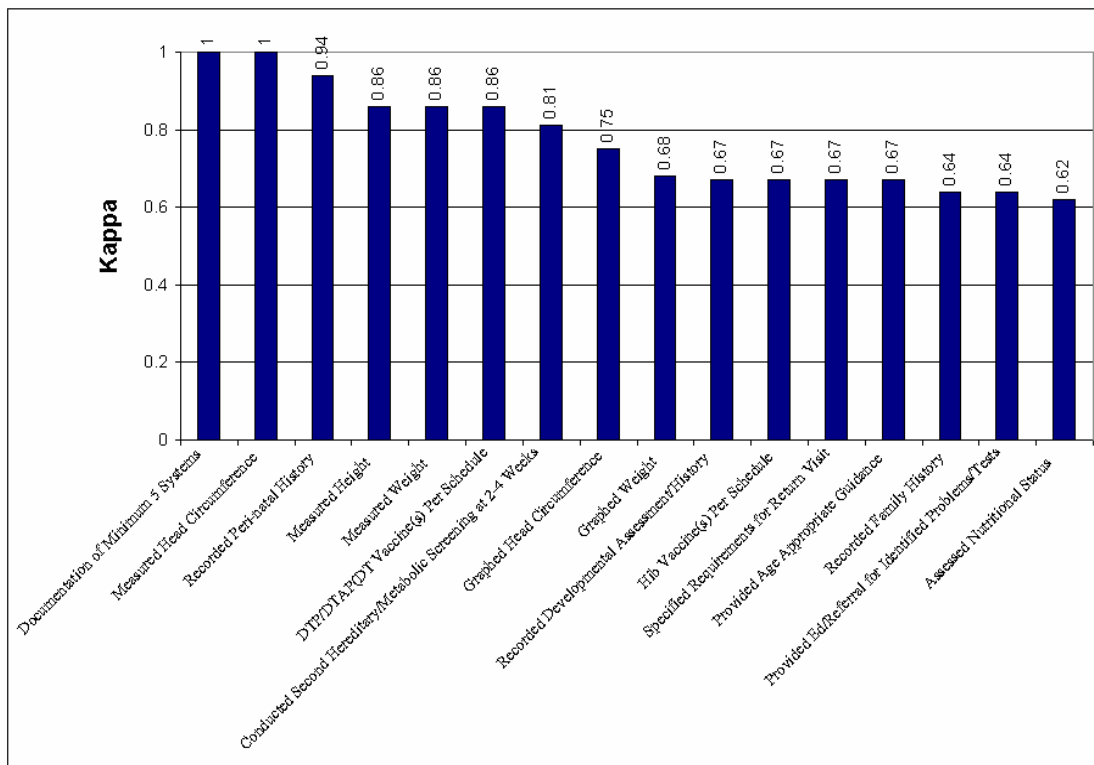


Figure 2: Elements with an Average Kappa Below 0.61 Indicating Fair to Moderate Agreement

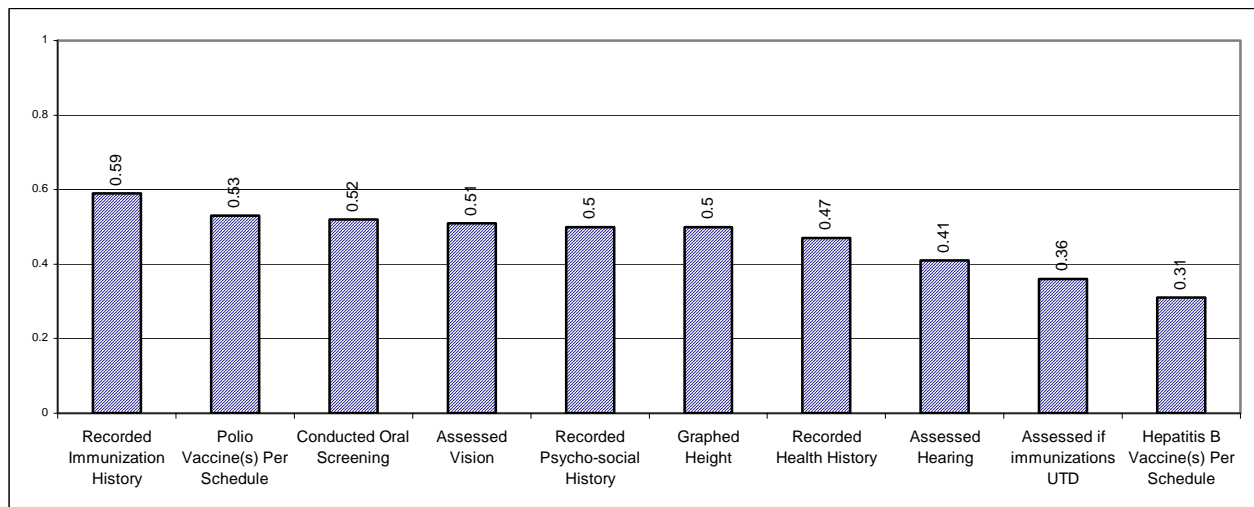


Table 3: 4 – 5-Month-Old Inter-Rater Reliability Using Cohen's Kappa*

ELEMENT	B/A	B/C	B/E	B/D	A/C	A/E	A/D	C/E	C/D	E/D	Average	Category Average
Health and Developmental History												0.64
Recorded Health History	0.31	0.79	0.31	0.53	0.27	0.72	0.42	0.27	0.64	0.42	0.47	
Recorded Family History	0.80	0.89	0.65	0.50	0.91	0.61	0.38	0.68	0.43	0.50	0.64	
Recorded Peri-natal History	1.00	1.00	0.84	1.00	1.00	0.84	1.00	0.84	1.00	0.84	0.94	
Recorded Psycho-social History	0.57	0.49	0.55	0.46	0.64	0.67	0.48	0.54	0.33	0.31	0.50	
Recorded Immunization History	0.65	1.00	1.00	0.20	0.65	0.65	0.38	1.00	0.20	0.20	0.59	
Recorded Developmental Assessment/History	0.54	0.46	1.00	0.79	0.79	0.54	0.72	0.46	0.63	0.79	0.67	
Comprehensive Unclothed Physical Exam												0.70
Documentation of Minimum 5 Systems	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
Assessed Vision	0.32	0.47	0.31	0.31	0.37	0.54	0.54	0.64	0.64	1.00	0.51	
Assessed Hearing	0.49	0.36	0.25	0.31	0.28	0.43	0.31	0.39	0.46	0.79	0.41	
Assessed Nutritional Status	0.64	0.79	0.54	0.64	0.69	0.62	0.44	0.45	0.53	0.87	0.62	
Conducted Oral Screening	0.27	1.00	0.40	0.40	0.27	0.63	0.63	0.40	0.40	0.82	0.52	
Measured Height	0.65	1.00	1.00	1.00	0.65	0.65	0.65	1.00	1.00	1.00	0.86	
Graphed Height	0.35	0.84	0.43	0.63	0.27	0.53	0.35	0.35	0.52	0.71	0.50	
Measured Weight	0.65	1.00	1.00	1.00	0.65	0.65	0.65	1.00	1.00	1.00	0.86	
Graphed Weight	0.65	0.78	0.46	0.65	0.47	0.65	1.00	0.35	0.47	0.65	0.61	
Measured Head Circumference	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
Graphed Head Circumference	0.67	0.82	0.63	0.63	0.71	0.71	0.71	0.83	0.83	1.00	0.75	
Laboratory Tests												0.81
Conducted Second Hereditary/Metabolic Screening at 2-4 Weeks	0.85	0.78	0.78	0.78	0.79	0.80	0.93	0.80	0.73	0.80	0.81	
Immunizations												0.55
Hepatitis B Vaccine(s) Per Schedule	0.35	0.26	0.26	0.63	0.10	0.35	0.35	0.26	0.26	0.26	0.31	
DTP/DTAP(DT Vaccine(s) Per Schedule	0.65	1.00	1.00	1.00	0.65	0.65	0.65	1.00	1.00	1.00	0.86	
Hib Vaccine(s) Per Schedule	0.65	1.00	0.47	1.00	0.65	0.35	0.65	0.47	1.00	0.47	0.67	
Polio Vaccine(s) Per Schedule	0.47	1.00	0.20	1.00	0.47	0.30	0.47	0.20	1.00	0.20	0.53	
Assessed if immunizations UTD	0.65	0.17	1.00	0.10	0.10	0.65	0.19	0.17	0.41	0.10	0.36	
Health Education												0.66
Specified Requirements for Return Visit	0.75	0.63	0.89	0.75	0.42	0.63	0.44	0.73	0.62	0.88	0.67	
Provided Ed/Referral for Identified Problems/Tests	0.79	0.79	0.53	0.45	1.00	0.64	0.54	0.64	0.54	0.50	0.64	
Provided Age Appropriate Guidance	0.59	0.67	0.82	0.70	0.50	0.61	0.51	0.77	0.67	0.91	0.67	
# of Kappas >=.61	16	20	14	17	15	19	12	14	14	15	17	
% of Kappas >= 0.61	62	77	54	65	58	73	46	54	54	58	65	

*Shaded cells have Kappa's greater than 0.61, in the good to very good agreement range. A=Grant B=Barnes C=Reese-Carter D=Watson E=Richter

Overall Avg.
Kappa .67

Analysis of 4-5 Month Olds Using Percentage of Agreement

Percentage

Overall, there was a very high rate of agreement considering the percentage scores. All five categories of analysis were well above the programmatic threshold of 80 percent. (See Table 1.)

Table 4: Overall Agreement Among EPSDT Nurse Reviewers

Category	Overall Percent of Agreement
Health and Developmental History	89
Comprehensive Unclothed Physical Exam	93
Laboratory Tests	91
Immunizations	90
Health Education	90

Only 2 of the 26 elements had average scores below 80 percent. Among the twenty-six elements reviewed, 8 contained pairs of reviewers with scores below 80 percent. For each pair of reviewers, there were 5 or fewer times out of a possible 30 that the agreement fell below 80 percent. See Table 3.

Table 5: 4-5-Month-Old Inter-Rater Reliability Using Percent of Agreement and Standard Deviation

ELEMENT	A/B	B/C	B/E	B/D	A/C	A/E	A/D	E/C	C/D	E/D	Average	Category Average
												89
Recorded Health History	73	97	73	90	73	87	77	73	93	77	82	
Recorded Family History	93	97	90	87	97	87	80	90	83	87	89	
Recorded Peri-natal History	100	100	93	100	100	97	100	97	100	97	99	
Recorded Psycho-social History	83	80	80	80	83	83	77	77	70	67	78	
Recorded Immunization History	97	100	100	80	97	97	83	100	80	80	92	
Recorded Developmental Assessment/History	90	87	100	97	93	90	93	87	90	97	93	
												93
Documentation of Minimum 5 Systems	100	100	100	100	100	100	100	100	100	100	100	
Assessed Vision	73	80	77	77	83	90	90	93	93	100	86	
Assessed Hearing	77	73	73	77	70	80	77	83	87	97	80	
Assessed Nutritional Status	93	97	90	93	93	90	87	87	90	97	92	
Conducted Oral Screening	73	100	83	83	73	83	83	83	83	93	84	
Measured Height	97	100	100	100	97	97	97	100	100	100	99	
Graphed Height	90	97	87	93	87	90	90	83	90	93	90	
Measured Weight	97	100	100	100	97	97	97	100	100	100	99	
Graphed Weight	97	97	93	97	93	97	100	90	93	97	96	
Measured Head Circumference	100	100	100	100	100	100	100	100	100	100	100	
Graphed Head Circumference	90	93	87	87	90	90	90	93	93	100	92	
												91
Conducted Second Hereditary/Metabolic Screening at 2-4 Weeks	93	90	90	90	90	90	97	90	87	90	91	
												90
Hepatitis B Vaccine(s) Per Schedule	90	77	87	93	73	90	90	77	77	87	84	
DTP/DTAP(DT Vaccine(s) Per Schedule	97	100	100	100	97	97	97	100	100	100	99	
Hib Vaccine(s) Per Schedule	97	100	93	100	97	90	97	93	100	93	96	
Polio Vaccine(s) Per Schedule	93	100	80	100	93	80	93	80	100	80	90	
Assessed if immunizations UTD	97	77	100	63	73	97	67	77	73	63	79	
												90
Specified Requirements for Return Visit	93	87	97	93	80	90	87	90	87	97	90	Overall Ave Overall Agreement for Age Group 91
Provided Ed/Referral for Identified Problems/Tests	97	97	90	87	100	93	90	93	90	87	93	
Provided Age Appropriate Guidance	83	87	93	90	77	83	80	90	87	97	87	
# of elements >=80%	22	23	23	23	21	26	22	22	23	24	24	
% of elements >=80%	85	88	88	88	81	100	85	85	88	92	92	

14-15-Year-Old Analysis Using Percent of Agreement

Percentage of agreement was calculated for each nurse pair, for each element. A level of 80 percent agreement was selected as the target level for EPSDT inter-rater agreement.

- 21 out of 29 elements had average percentage agreement scores greater than 80 percent

Table 6: Percentage of Agreement: 14-15-Year-Olds

Elements With Avg. Percent of Agreement \geq 80 Percent	Elements With Avg. Percent of Agreement \leq 80 Percent
Recorded Health History	Recorded Family History
Recorded Immunization History	Recorded Psycho-social History
Recorded Substance Abuse Assessment	Recorded Developmental Assess/History
Documentation of Minimum 5 Systems	Assessed Nutritional Status
Assessed Vision	Recorded Cholesterol Risk Assess per Schedule
Assessed Hearing	Assess if Immunizations UTD
Conducted Oral Screening	Provided Ed/Referral for Identification Problems/Tests
Measured and Graphed Height and Weight (4)	
Measured Blood Pressure	
Recorded Tb Risk Assessment per Schedule	
Recorded STD/HIV Risk Assessment per Schedule	
Hepatitis B Vaccine(s) Per Schedule	
TD Vaccine(s) Per Schedule	
MMR Vaccine(s) Per Schedule	
Varicella Vaccine(s) per Schedule	
Specified Requirements for Return Visit	
Provided Age-Appropriate Guidance	
Advised Dental Visit	

- Among the nurse pairs:
 - Seven elements had 0-4 pairs out of 10 with scores \geq 80 percent
 - Eight elements had 5-8 pairs out of 10 with scores \geq 80 percent
 - 14 elements had 9-10 pairs out of 10 with scores \geq 80 percent

14-15-Year-Olds Analysis Using Cohen's Kappa

Cohen's Kappa is considered to indicate substantial to nearly perfect agreement with scores greater than 0.61. In this analysis:

- Nurse pairs had CK ≥ 0.61 from 52 percent to 76 percent of the time through thirty reviewed charts (five nurses were paired to create ten unique pairs)
- 18 of 29 elements had average CK scores ≥ 0.61 and 11 had scores ≤ 0.61 :

**Table 7: Kappa Score Above and Lower than 0.61
Indicating “Substantial” or Better Agreement**

Elements With Avg. CK ≥ 0.61	Elements With Avg. CK ≤ 0.61
Recorded immunization	Recorded health history
Recorded Substance Abuse Assessment	Recorded Psycho-social History
Documentation of Minimum 5 Systems	Assesses Nutritional Status
Assessed Hearing	Assessed if immunizations UTD
Measured and Graphed Height	Provided Ed/Referral for Identified Problems/Tests
Measured and Graphed Weight	Recorded Family History
Measured Blood Pressure	Recorded Developmental Assess/History
Recorded Tb Risk Assessment per Schedule	Recorded Mental Health Assess
Recorded STD/HIV Risk Assessment per schedule	Assessed Vision
Hepatitis B Vaccine(s) Per Schedule	Conducted Oral Screening
TD Vaccine(s) Per Schedule	Recorded Cholesterol Risk Assess.
MMR Vaccine(s) Per Schedule	
Varicella Vaccine(s) Per Schedule	
Specified Requirements for Return Visit	
Provided Age Appropriate Guidal	
Advised Dental Visit	

- Some elements were scored more successfully (had better agreement) than others. The greater number of the nurse pairs with scores ≥ 0.61 the better the inter-rater reliability. Of the 29 elements:
 - 10 had no pairs with scores ≤ 0.61
 - 6 had 1-3 out of 10 pairs with scores ≤ 0.61
 - 8 had 4-6 out of 10 pairs with scores ≤ 0.61
 - 5 had 9-10 out of 10 pairs with scores ≤ 0.61

Table 8: 14 - 15 Year Old Inter-Rater Reliability Using Percent of Agreement and Standard Deviation

ELEMENT	A/B	B/C	B/E	B/D	A/C	A/E	A/D	E/C	C/D	E/D	Average	Category
Health and Developmental History	79											
Recorded Health History	80	90	90	80	83	77	73	87	80	77	82	
Recorded Family History	70	80	60	73	80	77	73	70	77	77	74	
Recorded Psycho-social History	67	57	63	60	70	77	77	73	73	83	70	
Recorded Immunization History	97	83	77	90	87	80	80	73	77	70	81	
Recorded Developmental Assessment/History	77	83	73	80	83	70	83	70	83	77	78	
Recorded Mental Health Assessment	77	83	80	73	90	83	57	80	60	70	75	
Recorded Substance Abuse Assessment	87	93	93	93	87	93	93	93	93	93	92	
Comprehensive Unclothed Physical Exam	89											
Documentation of Minimum 5 Systems	97	97	93	93	100	97	97	97	97	100	97	
Assessed Vision	83	80	83	90	70	87	87	73	77	93	82	
Assessed Hearing	87	80	83	87	73	93	100	87	73	93	86	
Assessed Nutritional Status	47	83	73	63	47	57	80	70	67	77	66	
Conducted Oral Screening	77	97	77	77	80	93	90	80	80	90	84	
Measured Height	100	100	100	100	100	100	100	100	100	100	100	
Graphed Height	90	87	83	83	90	87	93	90	83	87	87	
Measured Weight	100	100	100	100	100	100	100	100	100	100	100	
Graphed Weight	90	87	90	83	90	93	93	97	83	87	89	
Measured Blood Pressure	100	100	97	97	100	97	97	97	97	100	98	
Laboratory Tests	83											
Recorded Tb Risk Assessment per Schedule	83	90	90	93	87	87	90	93	90	90	89	
Recorded Cholesterol Risk Assessment per Schedule	73	90	80	63	70	80	73	77	60	83	75	
Recorded STD/HIV Risk Assessment per Schedule	87	93	93	87	80	87	80	87	87	80	86	
Immunizations	82											
Hepatitis B Vaccine(s) Per Schedule	77	73	80	90	83	87	87	80	73	90	82	
TD Vaccine(s) Per Schedule	83	87	80	83	87	80	90	73	80	73	82	
MMR Vaccine(s) per Schedule	97	93	93	100	93	93	97	93	93	93	95	
Varicella Vaccine(s) Per Schedule	83	90	90	93	83	83	83	93	83	90	87	
Assessed if immunizations UTD	67	63	67	60	70	53	77	70	60	60	65	
Health Education	85											
Specified Requirements for Return Visit	87	80	80	77	83	83	80	83	90	87	83	
Provided Ed/Referral for Identified Problems/Tests	57	87	80	80	63	67	60	73	87	73	73	
Provided Age Appropriate Guidance	93	87	90	100	93	90	93	83	87	90	91	
Advised Dental Visit	90	93	100	97	87	90	90	93	90	97	93	
% elements >80% Agreement for Each Pair	66	90	76	72	76	76	76	66	72	69	72	Overall Average for Age Group
												84

Figure 3: Average Percentage of Agreement for 14-15 Year Olds EPSDT Inter-rater Reliability

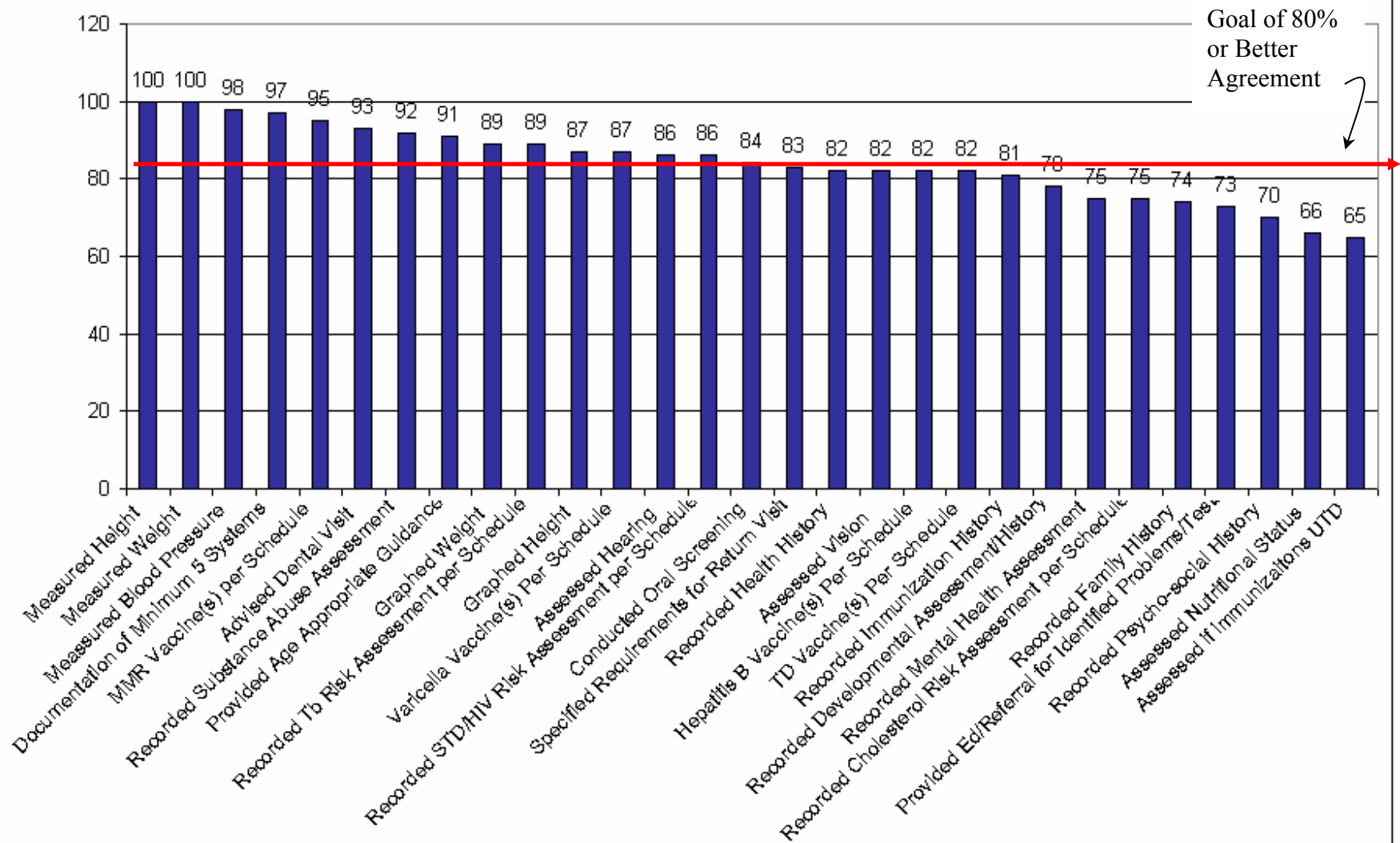
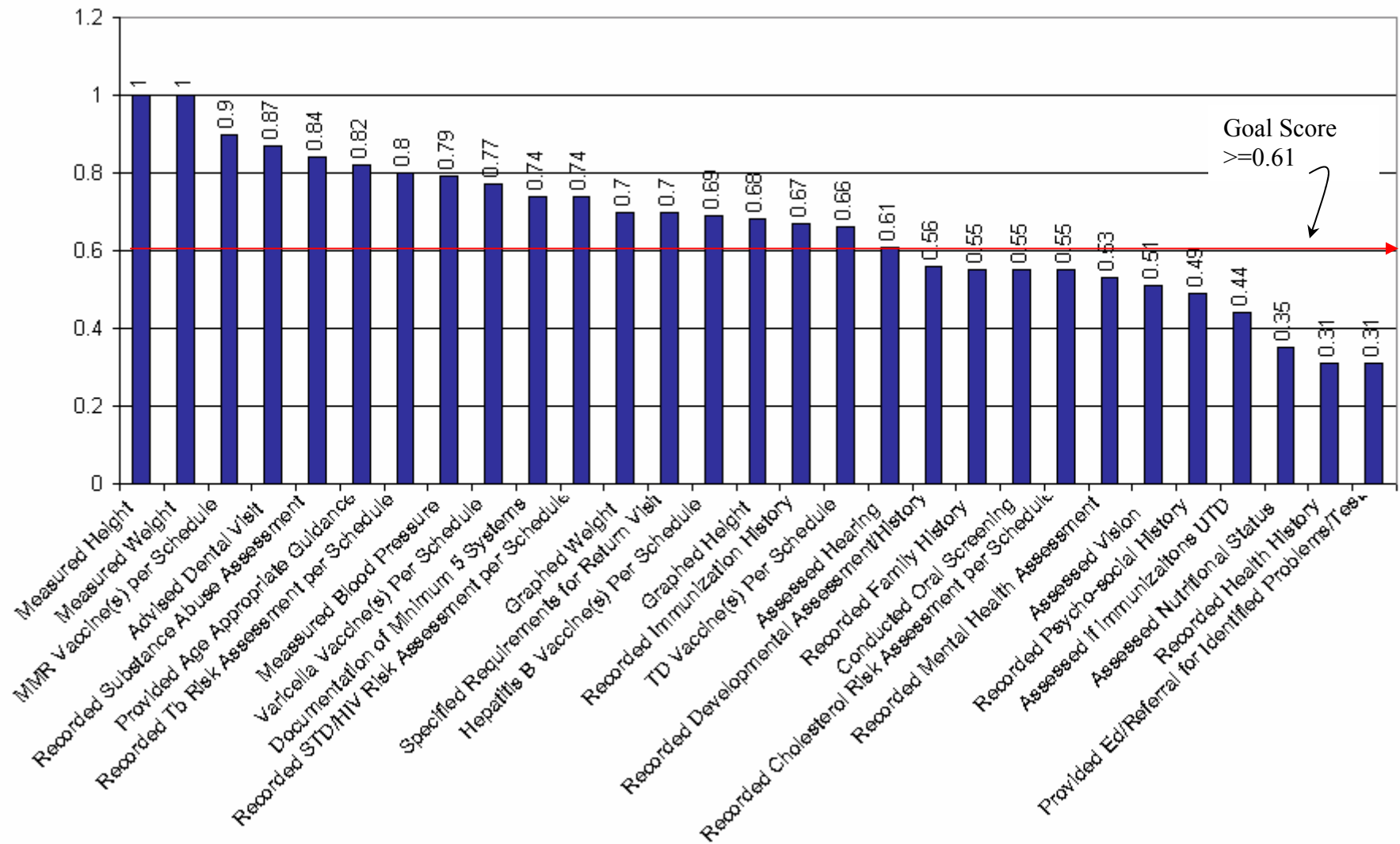


Table 9: 14-15-Year-Old Inter-Rater Reliability Using Cohen's Kappa*

ELEMENT	B/A	B/C	B/E	B/D	A/C	A/E	A/D	C/E	C/D	E/D	Average
Health and Developmental History											
Recorded Health History	0.28	0.46	0.46	0.08	0.48	0.27	0.22	0.45	0.26	0.11	0.31
Recorded Family History	0.46	0.61	0.32	0.53	0.66	0.62	0.56	0.51	0.61	0.63	0.55
Recorded Psycho-social History	0.37	0.19	0.35	0.30	0.51	0.62	0.63	0.57	0.57	0.74	0.49
Recorded Immunization History	0.94	0.69	0.56	0.82	0.75	0.63	0.76	0.51	0.57	0.46	0.67
Recorded Developmental Assessment/History	0.56	0.66	0.50	0.61	0.65	0.43	0.67	0.38	0.64	0.54	0.56
Recorded Mental Health Assessment	0.57	0.67	0.60	0.43	0.70	0.71	0.28	0.63	0.29	0.43	0.53
Recorded Substance Abuse Assessment	0.74	0.86	0.86	0.86	0.75	0.88	0.87	0.87	0.86	0.87	0.84
Comprehensive Unclothed Physical Exam											
Documentation of Minimum 5 Systems	0.65	0.65	0.47	0.47	1.00	0.78	0.78	0.78	0.78	1.00	0.74
Assessed Vision	0.59	0.33	0.51	0.68	0.30	0.70	0.70	0.23	0.28	0.81	0.51
Assessed Hearing	0.67	0.33	0.60	0.67	0.34	0.85	1.00	0.41	0.34	0.85	0.61
Assessed Nutritional Status	0.12	0.45	0.30	0.22	0.17	0.32	0.69	0.33	0.35	0.58	0.35
Conducted Oral Screening	0.31	0.79	0.31	0.31	0.44	0.86	0.78	0.44	0.44	0.78	0.55
Measured Height	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Graphed Height	0.73	0.66	0.59	0.59	0.73	0.66	0.83	0.75	0.59	0.68	0.68
Measured Weight	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Graphed Weight	0.71	0.63	0.71	0.56	0.71	0.79	0.82	0.90	0.56	0.63	0.70
Measured Blood Pressure	1.00	1.00	0.65	0.65	1.00	0.65	0.65	0.65	0.65	1.00	0.79
Laboratory Tests											
Recorded Tb Risk Assessment per Schedule	0.69	0.80	0.80	0.87	0.76	0.76	0.82	0.87	0.81	0.81	0.80
Recorded Cholesterol Risk Assessment per Schedule	0.51	0.81	0.61	0.30	0.47	0.65	0.54	0.57	0.29	0.70	0.55
Recorded STD/HIV Risk Assessment per Schedule	0.75	0.87	0.87	0.74	0.63	0.75	0.62	0.76	0.75	0.62	0.74
Immunizations											
Hepatitis B Vaccine(s) Per Schedule	0.60	0.56	0.62	0.80	0.74	0.78	0.77	0.68	0.57	0.82	0.69
TD Vaccine(s) Per Schedule	0.67	0.72	0.62	0.68	0.75	0.65	0.82	0.51	0.63	0.55	0.66
MMR Vaccine(s) per Schedule	0.94	0.87	0.87	1.00	0.87	0.87	0.94	0.87	0.87	0.87	0.90
Varicella Vaccine(s) Per Schedule	0.70	0.81	0.81	0.87	0.70	0.70	0.71	0.87	0.68	0.81	0.77
Assessed if immunizations UTD	0.50	0.44	0.49	0.38	0.52	0.28	0.61	0.52	0.32	0.32	0.44
Health Education											
Specified Requirements for Return Visit	0.75	0.63	0.63	0.60	0.70	0.70	0.65	0.70	0.83	0.77	0.70
Provided Ed/Referral for Identified Problems/Tests	0.07	0.54	0.33	0.31	0.27	0.27	0.20	0.24	0.60	0.24	0.31
Provided Age Appropriate Guidance	0.87	0.75	0.80	1.00	0.88	0.80	0.87	0.69	0.75	0.80	0.82
Advised Dental Visit	0.82	0.88	1.00	0.94	0.76	0.82	0.83	0.88	0.82	0.94	0.87
% of Nurse Pairs With Kappa >.61	62	72	52	55	69	76	76	55	55	72	.66

*Shaded cells have Kappa's greater than 0.61, in the good to very good agreement range. A=Grant B=Barnes C=Reese-Carter D=Watson E=Richter

Figure 4: Cohen's Kappa for 14-15 Year Olds EPSDT Inter-rater Reliability



Analysis of 2-Year-Olds Using Percent of Agreement

- Overall, every category of elements averaged above 80 percent
- All pairs of nurses had percentage agreement scores of 80 percent or higher on 23 out of 30 elements
- 6 out of 30 elements had 5 or more nurse pairs with percentages below 80 percent
 - Recorded Family History
 - Recorded Psycho-Social History
 - Assessed Nutritional Status
 - Graphed Weight
 - Anemia Screening per Protocol
- 7 out of 30 elements had average scores below 80 percent
 - Recorded Family History
 - Recorded psycho-Social History
 - Assessed Nutritional Status
 - Graphed weight
 - Anemia Screening per Schedule
 - Blood Lead Test per Schedule
 - Advised Dental Visit

Two-Year-Old Analysis Using Cohen's Kappa

- 12 out of 30 elements have Kappas above the target of .61 among all nurse pairs
- 8 out of 30 elements had 5 or more nurse pairs with Kappas below .61
- Nurse pairs ranged from 18-24 out of 30 elements with Kappas greater than .61
- The immunization category had the best overall number of Kappa agreements at 83 percent; followed by Comprehensive Unclothed Physical Exam, Laboratory Tests, and Health and Developmental History at 68, 64 and 62 percent, respectively. Health Education had an overall average of 57 percent Kappas greater than .61.

ELEMENT		A/B	B/C	B/E	B/D	A/C	A/E	A/D	E/C	C/D	E/D	Average	Category Average
Health and Developmental History													82
	Recorded Health History	61	93	93	96	68	61	64	93	96	96	82	
	Recorded Family History	75	50	57	64	79	75	79	89	93	89	75	
	Recorded Psycho-social History	57	50	57	57	61	71	68	68	64	89	64	
	Recorded Immunization History	93	96	93	93	96	93	93	96	96	93	94	
	Recorded Developmental Assessment/History	96	93	100	96	96	96	100	93	96	96	96	
Comprehensive Unclothed Physical Exam													88
	Documentation of Minimum 5 Systems	100	100	100	100	100	100	100	100	100	100	100	
	Assessed Vision	93	82	86	82	75	86	89	82	79	89	84	
	Assessed Hearing	96	82	82	96	79	86	93	93	86	86	88	
	Assessed Nutritional Status	46	89	75	64	54	46	71	68	68	75	66	
	Conducted Oral Screening	96	100	96	96	96	96	96	96	96	100	97	
	Measured Height	100	100	100	100	100	100	100	100	100	100	100	
	Graphed Height	89	86	89	86	89	93	96	82	93	89	89	
	Measured Weight	100	100	100	100	100	100	100	100	100	100	100	
	Graphed Weight	57	75	75	71	57	61	61	82	93	82	71	
Laboratory Tests													87
	Recorded Tb Risk Assessment per Schedule	100	100	100	100	100	100	100	100	100	100	100	
	Recorded Cholesterol Risk Assessment per Schedule	86	96	86	89	89	93	89	89	93	96	91	
	Anemia Screening Per Schedule	64	68	68	79	82	86	68	75	64	68	72	
	Conducted Lead Risks Schedule	89	96	96	100	86	93	89	93	96	96	93	
	Lead-Blood Test Per Schedule	68	71	75	86	82	93	71	82	68	71	77	
Immunizations													94
	Hepatitis B Vaccine(s) Per Schedule	96	93	96	100	96	93	96	89	93	96	95	
	DTP/DTAP(DT) Vaccine(s) per schedule	96	100	93	96	96	96	100	93	96	96	96	
	Hip Vaccine (s) Per Schedule	89	100	96	93	89	93	96	96	93	96	94	
	Polio Vaccine(s) Per Schedule	100	93	100	100	93	100	100	93	93	100	97	
	MMR Vaccine(s) per Schedule	96	93	96	96	96	100	100	96	96	100	97	
	Varicella Vaccine(s) Per Schedule	96	93	100	96	96	96	93	93	89	96	95	
	Assessed if immunizations UTD	82	93	79	82	82	79	89	79	82	82	83	
Health Education													85
	Specified Requirements for Return Visit	93	93	89	86	93	89	89	89	86	82	89	
	Provided Ed/Referral for Identified Problems/Tests	71	96	93	96	75	75	75	96	100	96	87	
	Provided Age Appropriate Guidance	96	89	96	93	86	93	89	89	86	96	91	
	Advised Dental Visit	57	75	75	71	57	61	61	82	93	82	71	
	% of elements >=80%	70	80	73	80	70	73	70	87	83	90	77	Overall Agreement for Age Group 87
	# of elements 80% agreement or better	21	24	22	24	21	22	21	26	25	27	23	

*The standard deviation is a measure of variation from the mean. The higher the standard deviation the more variation there is between raters. ** *Shaded cells have percent of agreement <0.80, the designated program threshold. A=Grant B=Barnes C=Reese-Carter D=Watson E=Richter

Figure 5: Sorted Average Percentage Scores for 2-Year-Old Analysis

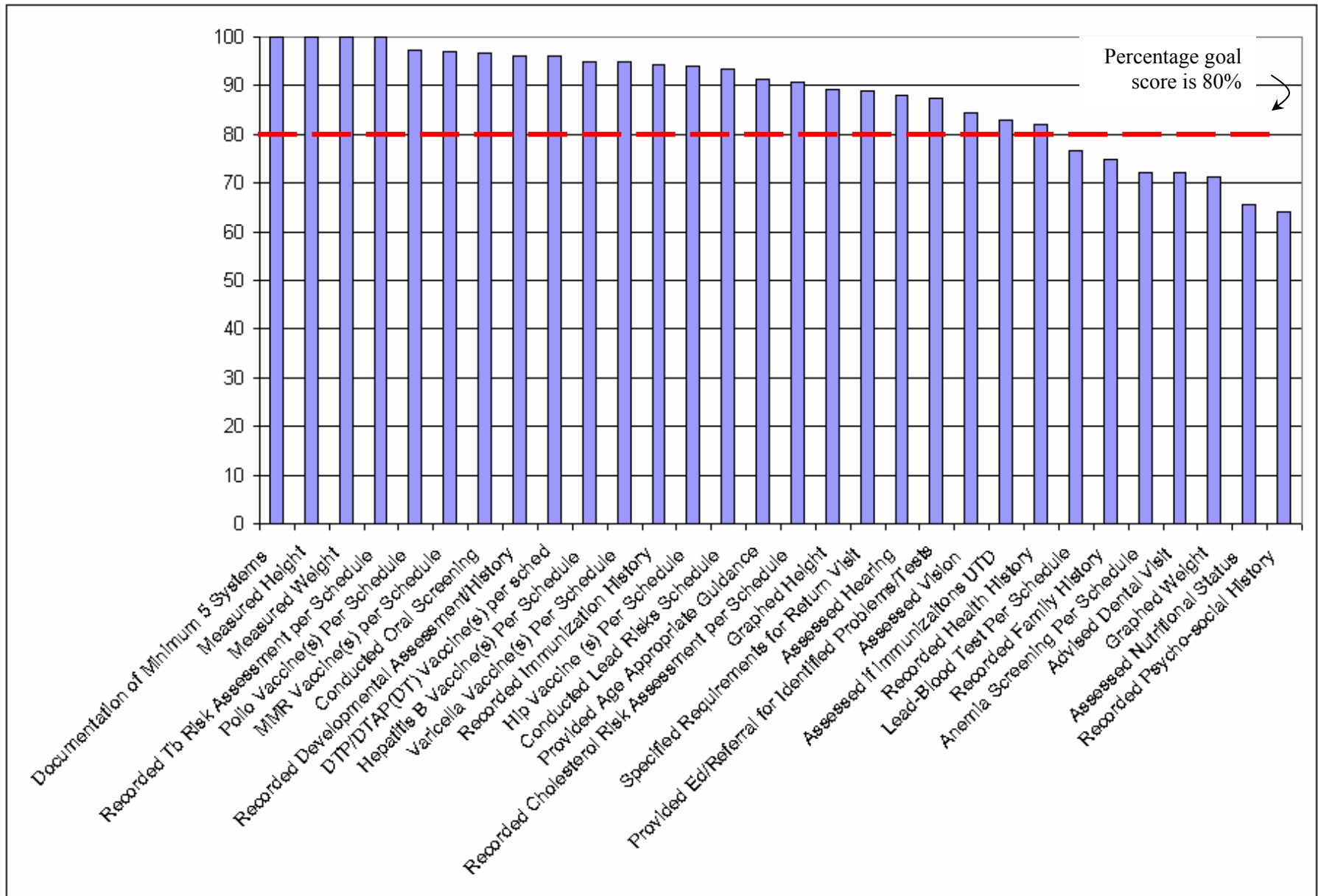
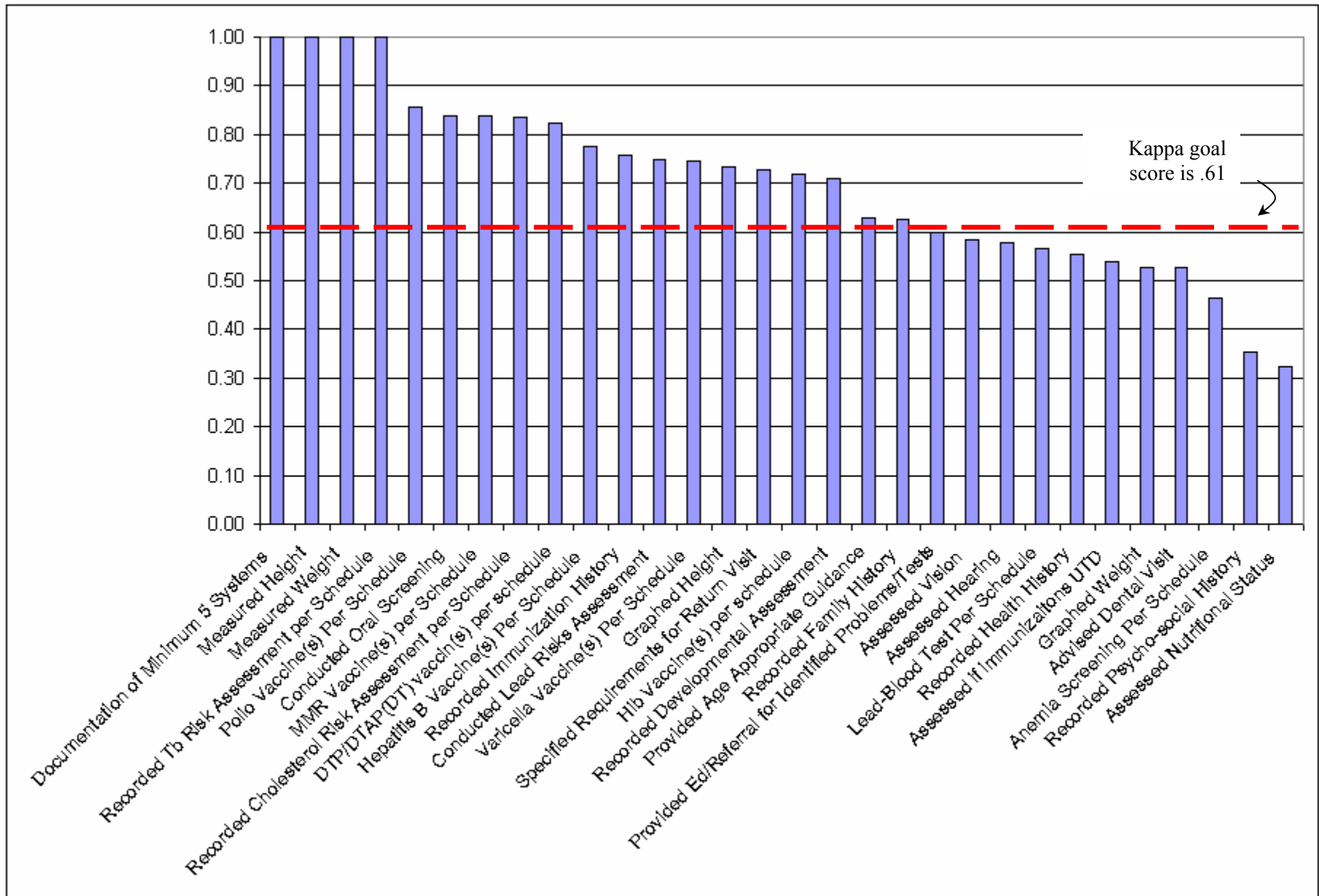


Table 11: 2-Year-Old Inter-Rater Reliability Using Cohen's Kappa*

ELEMENT	B/A	B/C	B/E	B/D	A/C	A/E	A/D	C/E	C/D	E/D	Average	Category Avg.
												0.60
Recorded Health History	0.19	0.64	0.64	0.79	0.37	0.23	0.29	0.72	0.84	0.84	0.56	
Recorded Family History	0.53	0.36	0.36	0.45	0.68	0.62	0.68	0.84	0.89	0.84	0.63	
Recorded Psycho-social History	0.21	0.08	0.12	0.12	0.37	0.52	0.46	0.46	0.40	0.80	0.35	
Recorded Immunization History	0.72	0.84	0.69	0.72	0.84	0.69	0.72	0.82	0.84	0.69	0.76	
Recorded Developmental Assessment	0.65	0.47	1.00	0.65	0.78	0.65	1.00	0.47	0.78	0.65	0.71	
												0.73
Documentation of Minimum 5 Systems	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
Assessed Vision	0.84	0.48	0.63	0.56	0.37	0.66	0.76	0.44	0.38	0.71	0.58	
Assessed Hearing	0.91	0.38	0.38	0.89	0.33	0.55	0.80	0.64	0.45	0.44	0.58	
Assessed Nutritional Status	0.13	0.54	0.30	0.22	0.26	0.13	0.53	0.23	0.36	0.52	0.32	
Conducted Oral Screening	0.79	1.00	0.79	0.79	0.79	0.82	0.82	0.79	0.79	1.00	0.84	
Measured Height	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
Graphed Height	0.74	0.67	0.76	0.64	0.72	0.82	0.89	0.57	0.80	0.72	0.73	
Measured Weight	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
Graphed Weight	0.28	0.55	0.59	0.51	0.33	0.39	0.40	0.68	0.85	0.69	0.53	
												0.72
Recorded Tb Risk Assessment per Schedule	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
Recorded Cholesterol Risk Assessment per Schedule	0.75	0.93	0.75	0.80	0.81	0.88	0.82	0.81	0.87	0.94	0.84	
Anemia Screening Per Schedule	0.31	0.41	0.37	0.36	0.71	0.75	0.39	0.59	0.35	0.39	0.46	
Conducted Lead Risks Assessment	0.58	0.87	0.84	1.00	0.50	0.69	0.58	0.72	0.87	0.84	0.75	
Lead-Blood Test Per Schedule	0.39	0.48	0.53	0.60	0.71	0.88	0.47	0.71	0.42	0.47	0.57	
												0.76
Hepatitis B Vaccine(s) Per Schedule	0.84	0.64	0.87	1.00	0.79	0.72	0.84	0.54	0.64	0.87	0.78	
DTP/DTAP(DT) vaccine(s) per schedule	0.84	1.00	0.64	0.84	0.84	0.79	1.00	0.64	0.84	0.79	0.82	
Hib Vaccine(s) per schedule	0.53	1.00	0.79	0.64	0.53	0.64	0.84	0.79	0.64	0.79	0.72	
Polio Vaccine(s) Per Schedule	1.00	0.64	1.00	1.00	0.64	1.00	1.00	0.64	0.64	1.00	0.86	
MMR Vaccine(s) per Schedule	0.79	0.64	0.79	0.79	0.79	1.00	1.00	0.79	0.79	1.00	0.84	
Varicella Vaccine(s) Per Schedule	0.79	0.64	1.00	0.79	0.84	0.79	0.64	0.64	0.53	0.79	0.75	
Assessed if immunizations UTD	0.55	0.82	0.44	0.50	0.55	0.43	0.70	0.44	0.50	0.47	0.54	
												0.62
Specified Requirements for Return Visit	0.81	0.82	0.75	0.65	0.81	0.74	0.72	0.75	0.65	0.58	0.73	
Provided Ed/Referral for Identified Problems/Tests	0.24	0.79	0.64	0.79	0.30	0.34	0.30	0.79	1.00	0.79	0.60	
Provided Age Appropriate Guidance	0.84	0.59	0.79	0.64	0.50	0.64	0.53	0.54	0.44	0.79	0.63	
Advised Dental Visit	0.28	0.55	0.59	0.51	0.33	0.39	0.40	0.68	0.85	0.69	0.53	
# of Kappas >.61	18	19	21	21	18	22	20	21	21	24	21	
% of Kappas>.61	60	66	72	72	62	76	69	72	72	83	70	

*Shaded cells have Kappa's greater than 0.61, in the good to very good agreement range. A=Grant B=Barnes C=Reese-Carter D=Watson E=Richter

Figure 6: Sorted Average Kappa Scores for 2-Year-Old Analysis



Overall Findings of 2004-2005 Analysis and Recommendations

It is important that processes that evaluate the quality of care are themselves evaluated and modified as needed to ensure that they in fact are adequate to accomplish the desired goal. To this end, the Maryland Early and Periodic Screening, Diagnosis and Treatment Program staff annually participate in a process to measure inter-rater reliability. The 2004-2005 assessment of inter-rater reliability indicates a high degree of consistency among the nurses for most elements. Sixty-eight out of 85 elements (80 percent) had average scores equal to or greater than the target of 80 percent agreement.

Overall, agreement was better for the younger age groups (4-5 month olds and 2-year olds), and less consistent for adolescents (14-15 year olds) (see Table 12). Only one element (Recorded Psycho-Social History) fell below 80 percent agreement, on average, in all three age groups, and three (Recorded Family History, Assessed Nutritional Status, and Assessed if Immunizations are Up-to-Date) were below 80% in 2 age groups. These four elements could be the focus of a more intensive review to further understand the nature of the nurses' lack of agreement, and to improve agreement scores.

Table 13: Elements with Overall Average Percent Below 80

	4-5 Months (26 elements)	2 Years (30 elements)	14-15 Years (29 elements)
Recorded Family History		x	x
Recorded Psycho-Social History	x	x	x
Recorded Developmental Assessment/History			x
Recorded Mental Health Assessment			x
Assessed Nutritional Status		x	x
Graphed Weight		x	
Recorded Cholesterol Risk Assessment per Schedule	NA		x
Anemia Screening per Schedule	NA	x	NA
Lead-Blood Test per Schedule	NA	x	NA
Assessed if Immunizations are Up-to-Date	x		x
Provided Educational/Referral for Identified Problems/Tests			x
Advised Dental Visit	NA	x	

Across nurse pairs, all achieved 84% or better overall agreement in the 4-5 month review. However, only five pairs achieved better than 80 percent agreement in the 2-year old review, and only 1 pair in the 14-15 year old review. (See Table 14.) A somewhat different picture emerges in considering the paired Kappa scores averaged over age groups. The 2-year old review had all nurse pairs achieving above the target Kappa of 0.61, while the 4-5 month old and 14-15 year old reviews did less well with 4 and 6 pairs, respectively, achieving a Kappa greater than 0.61.

Table 14: Overall Comparison of Percentage of Agreement and Kappa Scores Across Age Groups, by Nurse Pairs

Percent of Average Agreements Above 80 Percent per Nurse Pair

Nurse Pairs	1	2	3	4	5	6	7	8	9	10
4-5-Month-Olds	84	88	88	88	81	100	84	84	88	92
2-Year-Olds	70	83	76	83	72	76	72	90	86	93
14-15-Year-Olds	66	90	76	72	76	76	76	66	72	69

Percent of Average Agreements Above Kappa of 0.61 per Nurse Pair

Nurse Pairs	1	2	3	4	5	6	7	8	9	10
4-5-Month-Olds	62	77	54	65	58	73	46	54	54	58
2-Year-Olds	60	66	72	72	62	76	69	72	72	83
14-15-Year-Olds	62	72	52	55	69	76	76	55	55	72

Recommendations

As a result of this analysis, several observations have been made and a few recommendations have come forward. Firstly, the process of medical record review requires the perspective of a trained health professional who can discern the nuances of medical notation, particularly as it pertains to the free-range (non-formulaic and non-scripted standardized document) notations often favored by medical practitioners, such as SOAP (subjective, objective, assessment, and plan) notes, systems approaches, and others. Coupled with the variability of the organization of the medical record, this process is even more complex.

Also, noteworthy, is an assessment of the guidelines used for the EPSDT chart review, the “Healthy Kids Program Performance Improvement Guide for Clinical Record Reviews” (the Guide). The Guide offers great detail to assist the review nurse’s judgment of whether or

not to score a task as done, however there is a large degree of flexibility and latitude that the Guide affords, as well, which increases the chances that nurses will score differently.

Based on these observations, the following is recommended:

- Implement routine reviews (periodically and not tied to the annual inter-rater analysis, but ongoing throughout the year) of the specific records where percentage of agreement scores and Kappas fall below the target levels for the average element score
- Review the Guide for possible modifications that would reduce the large degree of flexibility, and group the elements into those that have concrete/absolute indicators from those that are more subjective and more prone to nurses' judgment and disagreement. (It is not suggested that critical elements be removed from the program's review requirements. However, that there is subjectivity in some of the elements should be acknowledged and the outcome of the assessment of those elements be considered without detriment to the evaluated entity—organization or provider.)