

Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Mangione-Smith R, DeCristofaro AH, Setodji CM, et al. The quality of ambulatory care delivered to children in the United States. *N Engl J Med* 2007;357:1515-23.

Appendix 1

Expert Panel Members

- 1) David A. Bergman, M.D. (Chair)
Lucile Salter Packard Children's Hospital
Palo Alto, California
Clinical Associate Professor
Stanford University
Palo Alto, California
- 2) Yi-Chuan Ching, M.D.
Kaiser Permanente
Honolulu, Hawaii
Associate Professor
University of Hawaii
Honolulu, Hawaii
- 3) John B. Coombs, M.D.
Associate Vice President
Associate Dean
Professor
University of Washington
Seattle, Washington
- 4) Arthur B. Elster, M.D.
American Medical Association, Department of Adolescent Health
Chicago, Illinois
- 5) Caroline Breese Hall, M.D.
Professor
University of Rochester School of Medicine and Dentistry
Rochester, New York
- 6) Katherine S. Lobach, M.D.
New York City Health and Hospitals Corporation
New York, New York
Clinical Professor
Albert Einstein College of Medicine
Bronx, New York
- 7) Thomas K. McInerney, M.D.
Panorama Pediatric Group
Rochester, New York
Clinical Professor, Pediatrics
University of Rochester School of Medicine and Dentistry
Rochester, New York

- 8) Eric M. Wall, M.D., M.P.H.
QualMed Health Plans of Oregon
Portland, Oregon

- 9) Charles J. Wibbelsman, M.D.
Kaiser Foundation Hospital
San Francisco, California
Assistant Clinical Professor of Pediatrics
University of California, San Francisco
San Francisco, California

Appendix 2: Detailed Results for Pediatric Quality Indicators Used in Analysis

Chapter/ Indicator #	Indicator Text	Unit	Type	Function	Modality	Problem	Guideline	# Eligible	# Eligible Events	Mean Score	Level of evidence
Acne											
1	For all patients presenting with acne, location of acne lesions (back, face, neck, chest) should be assessed.	P	C	D	P	U	N	28	28	67.85	III
2	If oral antibiotics are prescribed, papules and/or pustules must be present.	P	C	T	M	M	N	48	56	41.67	III
3	Tetracycline should not be prescribed for adolescents less than 12 years of age.	P	C	T	M	M	N	4	4	100	III
4	Before tetracycline is prescribed, the provider must determine that the patient is not pregnant.	P	C	T	M	M	N	5	5	20	III
5	If isotretinoin is prescribed, there must be evidence of severe acne (papules, pustules, cysts and nodules) and failure of previous therapy.	P	C	T	M	M	N	11	11	18.18	III
6	If isotretinoin is prescribed to post-pubescent girls, a negative pregnancy test should be obtained within two weeks of start of therapy.	P	C	T	L	M	N	2	2	50	III
7	If isotretinoin is prescribed, liver function tests should be performed at least every three months.	P	C	F	L	U	N	8	8	75	III
8	If isotretinoin is prescribed, triglyceride levels should be performed at least every three months.	P	C	F	L	U	N	8	8	62.5	III
ADHD											
1	The health care provider should communicate with school officials to structure school environment.	P	C	T	E	U	N	5	5	60	III
2	If the child has ADHD without a comorbidity, or with oppositional defiant disorder or conduct disorder, or with a learning disorder, and is started on pharmacotherapy, the initial medication choice should be methylphenidate, dextroamphetamine, or pemoline.	P	C	T	M	M	N	1	1	100	III

Chapter/ Indicator #	Indicator Text	Unit	Type	Function	Modality	Problem	Guideline	# Eligible	# Eligible Events	Mean Score	Level of evidence
3	Before a child is started on stimulant medication such as methylphenidate, dextroamphetamine, or pemoline, the health care provider should assess the heart rate of the child.	D	C	T	P	U	N	1	1	0	III
4	Before a child is started on stimulant medication such as methylphenidate, dextroamphetamine, or pemoline, the health care provider should assess the blood pressure of the child.	D	C	T	P	U	N	1	1	100	III
5	If a change in therapy has occurred, the health care provider initiating the change must evaluate the effect of the change within two weeks, either by an office visit or by phone contact	D	C	F	E	U	N	52	58	71.15	III
Adolescent Preventive Services											
1	Between the ages of 11 and 18 years, all adolescents should have an annual visit at which risk assessment/preventive services were provided.	P	P	S	E	U	A	544	819	45.22	III
2	Weight should be measured at least once a year and plotted on a growth chart or be recorded with the age/gender percentile.	P	P	S	P	U	A	544	819	15.44	III
3	Height should be measured at least once a year and plotted on a growth chart or be recorded with the age/gender percentile.	P	P	S	P	U	A	544	819	15.63	III
4	Blood pressure should be measured at least once a year.	P	P	S	P	U	A	544	819	63.24	III
5	If a patient has an elevated blood pressure, at least 2 additional blood pressure readings on 2 separate occasions within 3 months should be recorded in the chart.	P	P	D	P	U	A	19	21	15.79	III
6	All sexually active females or those who present for contraception should have an annual Pap smear within one year of becoming sexually active.	P	P	S	L	U	A	59	74	69.49	II

Chapter/ Indicator #	Indicator Text	Unit	Type	Function	Modality	Problem	Guideline	# Eligible	# Eligible Events	Mean Score	Level of evidence
7	All adolescents with a Pap smear consistent with HPV should have colposcopy performed.	P	P	D	S	U	N	1	1	100	III
8	If abnormal height/weight velocity is found, a follow-up visit should occur.	P	P	F	E	U	N	15	15	86.67	III
Allergic Rhinitis											
1	Treatment for allergic rhinitis should include at least one of the following: recommendation for allergen avoidance, antihistamine, nasal steroids, nasal cromolyn.	D	C	T	M	U	N	157	172	82.8	I-III
2	If topical nasal decongestants are prescribed, duration of treatment should be for no longer than 4 days.	D	C	T	M	O	N	2	2	100	II
Asthma											
1	PEFR should be measured in all patients > 5 years of age with chronic asthma (except for those with only exercise-induced asthma) at least annually in an office visit in which asthma is evaluated.	P	C	F	L	U	A	63	70	17.46	III
2	All patients > 5 years of age with the diagnosis of asthma should have been prescribed a beta2-agonist inhaler for symptomatic relief of exacerbations.	P	C	T	M	U	A	116	140	88.79	III
3	Patients ≤ 5 years of age should be prescribed a nebulizer (for administering asthma medications).	P	C	T	M	U	A	33	36	27.27	III
4	Patients who report using a beta2-agonist inhaler more than 3 times per day on a daily basis should be prescribed a longer acting bronchodilator (theophylline) and/or an anti-inflammatory agent (inhaled corticosteroids, cromolyn).	P	C	T	M	U	A	32	36	43.75	II, III
5	In any patients requiring chronic treatment with oral corticosteroids a trial of inhaled corticosteroids should have been attempted.	P	C	T	M	O	A	23	25	8.7	III

Chapter/ Indicator #	Indicator Text	Unit	Type	Function	Modality	Problem	Guideline	# Eligible	# Eligible Events	Mean Score	Level of evidence
6	Any child with asthma who takes high doses of inhaled corticosteroids should have growth patterns monitored at least annually.	P	C	F	P	U	A	1	1	0	III
7	Patients who require frequent bursts of prednisone (2-3 trials of 5-day therapy with oral corticosteroids after an exacerbation of asthma within the past 6 months) who are not already on inhaled corticosteroids or chromolyn sodium should be started on them.	P	C	T	M	U	A	5	5	80	III
8	All patients > 5 years of age presenting to the physician's office with an asthma exacerbation or historical worsening of asthma symptoms should be evaluated with PEFr or FEV1.	D	C	D	L	U	A	66	67	14.93	III
9	A physical exam of the chest should be performed in all patients presenting with an asthma exacerbation in the physician's office or emergency room.	D	C	D	P	U	A	81	87	97.51	III
10	All patients presenting to the physician's office or emergency department with an FEV1 or PEFr<=70 percent of baseline (or predicted) should be treated with beta2-agonists before discharge. (Note: OP'ed eligibility is patients who present with an exacerbation)	D	C	T	M	U	A	81	87	30.96	III
11	Patients who receive treatment with beta2-agonists for FEV1 or PEFr <70 percent in the physician's office or emergency department (ED) should have an FEV1 or PEFr repeated prior to discharge. (Note: OP'ed eligibility is patients who who receive beta2-agonists for exacerbation)	D	C	T	L	U	A	31	34	32.73	III
12	Patients with an FEV1 or PEFr <=70 percent of baseline (or predicted) after treatment for an asthma exacerbation in the physician's office should be placed on an oral corticosteroid taper. (Note: OP'ed eligibility is patients who who receive >1 beta2-agonists for exacerbation)	D	C	T	M	U	A	13	16	53.13	III

Chapter/ Indicator #	Indicator Text	Unit	Type	Function	Modality	Problem	Guideline	# Eligible	# Eligible Events	Mean Score	Level of evidence
13	Patients who have persistent symptoms, diffuse wheezes on chest auscultation, and a PEFR or FEV1<=40 percent of baseline (or predicted) after treatment with beta2-agonist should be admitted to the hospital.	D	C	T	E	U	A	12	14	35.71	III
14	Patients whose asthma medication is changed (new medication added, current dose decreased/increased) during one visit should have a follow-up within 3 weeks.	P	C	F	E	U	A	35	43	42.86	III
15	Patients on chronic oral corticosteroids should have follow-up visits at least 4 times in a calendar year.	P	C	F	E	U	A	8	8	22.22	III
16	Patients seen in the emergency department with an asthma exacerbation should have a follow-up reassessment within 72 hours.	D	C	F	E	U	A	11	11	21.21	III
17	Patients with a hospitalization for an asthma exacerbation should receive follow-up assessment within 7 days.	D	C	F	E	U	A	11	11	29.76	III
Depression											
1	Once diagnosis of major depression has been made, treatment with anti-depressant medication and/or psychotherapy should begin within 2 weeks.	P	C	T	M	U	A	8	9	87.5	I, II-1, II-2
2	Medication treatment visits or telephone contacts should occur weekly for a minimum of 4 weeks.	P	C	F	E	U	A	6	7	16.67	III
3	A follow-up family meeting should occur within 2 weeks of the initial evaluation.	P	C	T	E	U	N	8	9	12.5	
4	Anti-anxiety agents should NOT be used (except alprazolam) as the sole agent for treatment of depression, unless there is documentation of a comorbid anxiety disorder.	P	C	T	M	M	A	5	5	100	I

Chapter/ Indicator #	Indicator Text	Unit	Type	Function	Modality	Problem	Guideline	# Eligible	# Eligible Events	Mean Score	Level of evidence
5	Persons who have suicidality and either psychosis, current alcohol or drug use, or specific plans to carry out suicide should be hospitalized.	P	C	T	E	U	A	4	4	50	III
6	Persons hospitalized for depression should have follow-up with a mental health specialist or their primary care doctor within one week of discharge.	P	C	F	E	U	A	3	3	33.33	III
Diarrrhea, Acute											
1	The weight should be recorded and, if available, compared to a recent weight obtained prior to the onset of diarrhea.	D	A	D	P	U	N	70	73	81.51	III
2	At least two of the following findings regarding hydration status should be recorded: general condition, appearance of eyes, presence or absence of tears, degree of oral moisture, degree of thirst, degree of skin turgor, or condition of anterior fontanelle.	D	A	D	P	U	N	70	73	21.92	III
3	Severity of dehydration should be determined.	D	A	D	P	U	N	70	73	31.51	III
4	Fecal leukocytes should have been obtained if the child with diarrhea is less than 36 months of age and had fever or blood in the stool.	D	A	D	L	U	N	64	69	0.73	II-2
5	Stool culture should be obtained if the child with diarrhea is less than 36 months of age and had fever or blood in the stool or > 5 fecal leukocytes per high power field.	D	A	D	L	U	N	64	69	9.42	II-2
6	If the child had diarrhea but was not dehydrated, the practitioner should recommend additional fluid intake beyond what is normal for the child.	D	A	T	E	U	N	67	72	19.33	III
7	If the child had mild-moderate dehydration and is able to take oral fluids, oral rehydration therapy should be prescribed.	D	A	T	E	U	N	1	1	100	III

Chapter/ Indicator #	Indicator Text	Unit	Type	Function	Modality	Problem	Guideline	# Eligible	# Eligible Events	Mean Score	Level of evidence
8	If while healthy the child was being breast fed, the health care provider should advise the parent to continue breast feeding if the child is able to feed orally.	D	A	T	E	U	N	3	3	0	I
9	If while healthy the child was formula fed or weaned, the health care provider should have instituted refeeding within twenty-four hours of the onset of hydration therapy if able to take oral fluids.	D	A	T	E	U	N	7	7	14.29	I
10	Antimicrobial agents should be used in a child with suspected or culture-proven shigella.	D	A	T	M	U	N	6	6	0	III
11	Antimicrobial agents should be used in a child with Giardia with symptoms of greater than 10-14 days duration and with positive stool ova and parasite examination.	D	A	T	M	U	N	1	1	0	III
12	Antidiarrheal or antimotility medications are not used in treatment of diarrhea in a child.	D	A	T	M	O	N	73	78	93.59	III
Fever											
1	The health care provider should document a temperature measured by oral or rectal thermometry in the medical setting.	D	A	D	P	U	N	148	159	82.06	II-2
2	The assessment should document whether or not the infant appears toxic, i.e., if the infant has signs of sepsis based on: lethargy, poor perfusion, or hypoventilation, hyperventilation, or cyanosis. (<i>Applies to febrile infants 28 to 90 days old</i>)	D	A	D	P	U	A	4	4	50	II-2
3	All children who present with fever > 38°C should receive complete blood count. (<i>Applies to febrile infants 28 to 90 days old</i>)	D	A	D	L	U	A	4	4	25	III
4	If the infant is toxic or at high risk, the infant should be hospitalized. (<i>Applies to febrile infants 28 to 90 days old</i>)	D	A	T	E	U	A	3	3	33.33	III

Chapter/ Indicator #	Indicator Text	Unit	Type	Function	Modality	Problem	Guideline	# Eligible	# Eligible Events	Mean Score	Level of evidence
5	If the infant is toxic or at high risk, the assessment should also include cerebrospinal fluid studies. <i>(Applies to febrile infants 28 to 90 days old)</i>	D	A	D	L	U	A	3	3	33.33	III
6	If the infant is toxic or at high risk, the assessment should also include blood culture. <i>(Applies to febrile infants 28 to 90 days old)</i>	D	A	D	L	U	A	3	3	33.33	III
7	If the infant is toxic or at high risk, the assessment should also include urine culture. <i>(Applies to febrile infants 28 to 90 days old)</i>	D	A	D	L	U	A	3	3	33.33	III
8	If the infant with fever is at low risk and is being managed as an outpatient, the infant's condition should be reassessed in 18-24 hours. <i>(Applies to febrile infants 28 to 90 days old)</i>	D	A	F	E	U	A	1	1	0	III
9	If the infant/child is at high risk for sepsis, the assessment should include a complete blood count and urinalysis. <i>(Applies to febrile infants 3 to 36 months old)</i>	D	A	D	L	U	A	36	38	18.86	III
10	If the infant/child is at high risk for sepsis, cerebrospinal fluid studies (glucose, protein, gram stain, cell count, culture) should be performed. <i>(Applies to febrile infants 3 to 36 months old)</i>	D	A	D	L	U	A	36	38	2.63	III
11	If the infant/child is at high risk for sepsis, a blood culture should be performed. <i>(Applies to febrile infants 3 to 36 months old)</i>	D	A	D	L	U	A	36	38	17.54	III
12	If the infant/child is at high risk for sepsis, a urine culture should be performed. <i>(Applies to febrile infants 3 to 36 months old)</i>	D	A	D	L	U	A	36	38	16.23	III

Chapter/ Indicator #	Indicator Text	Unit	Type	Function	Modality	Problem	Guideline	# Eligible	# Eligible Events	Mean Score	Level of evidence
13	If the infant/child is at low risk for sepsis and the temperature is greater than or equal to 39°C and there is no obvious source of infection, a urine culture should be obtained if the child is a female less than two years of age or a male less than six months of age or an uncircumcised male less than 2 years of age. <i>(Applies to febrile infants 3 to 36 months old)</i>	D	A	D	L	U	A	21	22	13.36	III
14	If the infant/child is at low risk for sepsis and the temperature is greater than or equal to 39°C, a chest radiograph should be obtained in a child with dyspnea, tachypnea, rales, or decreased breath sounds. <i>(Applies to febrile infants 3 to 36 months old)</i>	D	A	D	L	U	A	2	2	0	III
15	If the infant/child at low risk for sepsis with an initial temperature greater than or equal to 39°C is well-appearing at follow-up, no further intervention is required; but, the parent should be advised to return if the fever persists greater than 48 hours. <i>(Applies to febrile infants 3 to 36 months old)</i>	D	A	F	E	O	A	4	4	50	III
Immuniza- tions (Childhood)											
1	All children should have had two OPV/IPV between six weeks and the first birthday.	P	P	T	I	U	A	76	76	77.63	I, III
2	All children should have had three OPV/IPV between six weeks and the second birthday.	P	P	T	I	U	A	155	155	60.65	I, III
3	Children with immunocompromise (hematologic and solid tumors, congenital immunodeficiency, and long-term immunosuppressive therapy) or HIV infection should receive IPV rather than OPV (at the same ages as OPV).	P	P	T	I	M	A	2	2	50	III

Chapter/ Indicator #	Indicator Text	Unit	Type	Function	Modality	Problem	Guideline	# Eligible	# Eligible Events	Mean Score	Level of evidence
4	All children should have had three DTP/DTaP/DT between six weeks and the first birthday.	P	P	T	I	U	A	78	78	71.8	I, III
5	All children should have had four DTP/DTaP/DT between six weeks and the second birthday, with at least six months between the third and fourth dose (the fourth may be DTaP if given after 15 months old).	P	P	T	I	U	A	157	157	46.5	I, III
6	All children should have had two PRP-OMP Hib or three Hib (any combination of formulations) between six weeks and the first birthday.	P	P	T	I	U	A	78	78	62.82	I, III
7	Between the ages of six weeks and 2 years, all children should have had either four Hib vaccinations, or three Hib vaccinations if the first two were PRP-OMB Hib.	P	P	T	I	U	A	157	157	42.68	I, III
8	All children should have had one MMR between their first and second birthdays.	P	P	T	I	U	A	157	157	64.33	I, III
9	All children whose mothers are known to be HBsAg-Negative should have had at least two HBV by the first birthday with at least one month between the first two doses.	P	P	T	I	U	A	155	155	60.65	I, III
10	All children whose mothers are known to be HBsAg-Negative should have had three HBV by the second birthday with at least one month between the first two doses.	P	P	T	I	U	A	255	255	47.45	I, III
11	All children whose mothers are known to be HBsAg-Positive at birth should receive HBIG and HBV by the beginning of the twelfth hour of life.	P	P	T	I	U	A	1	1	0	I, III
12	All children whose mothers are known to be HBsAg-Positive should have had three HBV by the beginning of the ninth month of life.	P	P	T	I	U	A	1	1	100	I, III

Chapter/ Indicator #	Indicator Text	Unit	Type	Function	Modality	Problem	Guideline	# Eligible	# Eligible Events	Mean Score	Level of evidence
13	Children with asthma and other chronic pulmonary diseases, hemodynamically significant cardiac disease, hemoglobinopathies (e.g., sickle cell disease) or undergoing immunosuppressive therapy should receive a yearly influenza vaccine.	P	P	T	I	U	A	9	9	11.11	I, III
14	Each immunization given at that institution should be documented with the manufacturer.	D	P	F	E	U	N	638	638	49.37	III
15	Each immunization given at that institution should be documented with the lot number.	D	P	F	E	U	N	638	638	56.11	III
Otitis Media											
1	Patients with persistent (>3 months duration) bilateral otitis media with effusion should have a hearing evaluation.	D	C	D	L	U	A	1	1	0	II/III
2	All patients with the diagnosis of otitis media with effusion should receive either antibiotics or trial of observation.	D	C	T	M	U	A	22	22	59.09	I, II/III
3	For all patients with a diagnosis of otitis media with effusion, during the initial management period (up to 12 weeks), myringotomy with or without insertion of tympanostomy tubes should not be performed in an otherwise healthy child (no other complications).	D	C	T	S	O	A	15	15	80	II
Prenatal Care											
1	The first prenatal visit should occur in the first trimester.	P	A	T	E	U	N	10	11	50	II-1
2	The physician should make an accurate determination of gestational age.	P	A	D	L	U	N	10	11	70	III
3	Pregnant women should be screened for anemia at the first prenatal visit.	P	A	S	L	U	N	9	10	66.67	III
4	Pregnant women should be re-screened for anemia after 24 weeks.	P	A	S	L	U	N	9	10	88.89	III
5	Women should receive a urine screen at the first prenatal visit.	P	A	S	L	U	N	9	10	77.78	I

Chapter/ Indicator #	Indicator Text	Unit	Type	Function	Modality	Problem	Guideline	# Eligible	# Eligible Events	Mean Score	Level of evidence
6	Women should receive a serologic test for rubella immunity before delivery.	P	A	S	L	U	N	10	11	80	II-2
7	Women should be screened for Hepatitis B surface antigen before delivery.	P	A	S	L	U	N	9	10	88.89	II-2
8	A non-treponemal screening test (e.g., VDRL, RPR) should be performed at the first prenatal visit.	P	A	S	L	U	N	9	10	77.78	II-3
9	A cervical gonorrhea culture should be performed at the first prenatal visit.	P	A	S	L	U	N	9	10	44.44	III
10	Women at high risk (adolescents, unmarried, those with multiple sex partners, low SES, other STD diagnosed) should receive a cervical chlamydia culture or antigen detection	P	A	S	L	U	N	9	10	55.56	III
11	Pregnant women should be offered HIV testing at the first prenatal visit.	P	A	S	L	U	N	9	10	66.67	I/III
12	Women should be offered alpha fetoprotein (AFP) testing between 15 and 20 weeks.	P	A	S	L	U	N	9	10	55.56	II-2
13	Women who are African American or have a family history of sickle cell disease should be offered screening at the first prenatal visit, if status unknown.	P	A	S	L	U	N	2	2	0	II-2
14	Women should receive an Rh factor and antibody screen at the first prenatal visit.	P	A	S	L	U	N	9	10	77.78	II-2
15	Measurements of the symphysis-fundal height should be made at each visit from 20-32 weeks	D	A	S	P	U	N	7	8	85.71	III
16	Blood pressure measurements should be taken at each visit.	D	A	S	P	U	N	9	10	88.89	II-2
17	A one-hour, 50g glucose challenge test should be performed on women with risk factors at 24-28 weeks.	P	A	S	L	U	N	1	2	0	I
18	Women whose symphysis-fundal height is 4 cm less than indicated by their gestational age between 20-32 weeks should have an ultrasound.	P	A	D	L	U	N	1	1	0	III

Chapter/ Indicator #	Indicator Text	Unit	Type	Function	Modality	Problem	Guideline	# Eligible	# Eligible Events	Mean Score	Level of evidence
19	In women without a prior diagnosis of chronic hypertension, who have elevated BPs (systolic > 140mm Hg at 20 weeks or later, or diastolic > 90mm Hg at 20 weeks or later, OR	P	A	D	L	U	N	1	1	100	II-2
20	In women without a prior diagnosis of chronic hypertension, who have elevated BP and either proteinuria (1+ or more) or edema (> trace), PIH diagnosis should be made.	P	A	D	L	U	N	1	1	0	II-2
21	Pregnant women with abnormal glucose challenge tests (>= 140mg/dl or 7.8 mmol/l) should have a 3-hour plasma glucose tolerance test performed.	P	A	D	L	U	N	1	2	100	I
22	Pregnant women with positive urine cultures (> 100,000 bacteria/cc) should receive an appropriate antibiotic.	P	A	T	M	U	N	1	1	100	I
23	If PIH diagnosed and patient is not admitted, bedrest should be recommended and a return visit should occur within 1 week.	P	A	T	E	U	N	1	1	0	II-2
24	If PIH diagnosed and pregnancy is at term (= 37 weeks), either labor should be induced or delivery by cesarean section should take place.	P	A	T	S	U	N	3	3	33.33	II-2
25	Women treated for positive cultures should receive a post-treatment follow-up culture within one month of completing treatment.	P	A	F	L	U	N	1	1	0	I
Upper Respiratory Illness											
1	Antibiotics should only be prescribed in a patient with nasal congestion and pharyngitis if a rapid strep test or throat culture is obtained, or if there is evidence of a diagnosis of other bacterial infections.	D	A	T	M	M	N	239	248	75.69	III
2	Aspirin should not be used in children and teenagers with pharyngitis.	D	A	T	M	M	N	656	707	99.58	III

Chapter/ Indicator #	Indicator Text	Unit	Type	Function	Modality	Problem	Guideline	# Eligible	# Eligible Events	Mean Score	Level of evidence
3	Treatment for acute sinusitis should be with antibiotics for at least 10 days.	D	A	T	M	U	N	10	11	46.97	III
4	In the absence of symptoms of allergic rhinitis (thin, watery rhinorrhea, and sneezing), antihistamines should not be prescribed for acute sinusitis.	D	A	T	M	M	N	14	15	33.33	III
5	In the absence of symptoms of allergic rhinitis (thin, watery rhinorrhea, and sneezing), antihistamines should not be prescribed for chronic sinusitis.	D	A	T	M	M	N	1	1	0	III
Urinary Tract Infection											
1	If an infant or child presents with symptoms/signs, either a urine culture should be performed or a urinalysis should be performed; if urinalysis is positive, a urine culture should be performed.	D	A	D	L	U	N	82	91	67.58	III
2	In order to diagnose UTI, a positive culture from one of the following methods of urine collection is necessary: bladder tap, catheterization, or clean catch.	D	A	D	L	U	N	28	31	9.14	III
3	Children with UTI and systemic symptoms such as hypotension, poor perfusion, anorexia, or emesis, should be treated initially with parenteral antibiotics.	D	A	T	M	U	N	4	4	0	III
4	A child with four UTIs in a single year should receive prophylactic antibiotics for at least six months.	P	A	T	M	U	N	1	1	0	III
5	Any boy less than 10 years old with a first UTI or with systemic symptoms (and/or who has not had the following studies before) should have a VCUG and one of the following within three months of diagnosis: RUS, IVP, or nuclear medicine renal scan.	P	A	F	L	U	N	1	1	0	III

Chapter/ Indicator #	Indicator Text	Unit	Type	Function	Modality	Problem	Guideline	# Eligible	# Eligible Events	Mean Score	Level of evidence
6	Any girl less than 3 years old with a first UTI or less than 10 years old with systemic symptoms (and/or who has not had the following studies before) should have a VCUG or IC and one of the following within three months of diagnosis RUS, IVP, or nuclear medicine renal scan.	P	A	F	L	U	N	6	6	0	III
Vaginitis and STDs											
1	In adolescent girls who are having vaginal intercourse and who present with a complaint of vaginal discharge, the practitioner should perform a speculum exam to determine the source of discharge.	D	A	D	P	U	N	15	19	87.71	III
2	At a minimum, the following tests should be performed on the vaginal discharge: normal saline wet mount for clue cells and trichomads; KOH wet mount for yeast hyphae.	D	A	D	L	U	N	21	27	30.25	III
3	Routine testing for gonorrhea and chlamydia (culture and antigen detection, respectively) should be performed with the routine pelvic exam in all adolescent girls.	D	P	S	L	U	N	53	65	41.54	III
4	If a patient has symptoms of urethritis, he should be tested for both chlamydia and gonorrhea or receive proper treatment for both.	D	A	T	M	U	N	3	4	43.4	III
5	If a patient presents with any STD, HIV testing should be offered.	D	P	S	L	U	A	6	10	0	III
6	If a patient presents with any sexually transmitted disease (gonorrhea, chlamydia, herpes, chancroid, syphilis) a non-treponemal test (VDRL or RPR) for syphilis should be obtained.	D	A	S	L	U	A	6	10	20	III
7	Treatment for bacterial vaginosis should be with metronidazole (orally or vaginally) or clindamycin (orally or vaginally).	D	A	T	M	U	A	5	5	80	I

Chapter/ Indicator #	Indicator Text	Unit	Type	Function	Modality	Problem	Guideline	# Eligible	# Eligible Events	Mean Score	Level of evidence
8	Treatment for T. vaginalis should be with oral metronidazole in the absence of allergy to metronidazole.	D	A	T	M	U	A	2	4	75	I
9	Treatment for non-recurrent (three or fewer episodes in previous year) yeast vaginitis should be with topical "azole" preparations (e.g., clotrimazole, butoconazole, etc.) or fluconazole.	D	A	T	M	U	A	8	8	70.83	I
10	All women treated for gonorrhea should also be treated for chlamydia.	D	A	T	M	U	A	1	1	100	II-2, III
11	Total antibiotic therapy for PID should be for no less than 10 days (inpatient, if applicable, plus outpatient).	D	A	T	M	U	A	2	2	0	III
12	Sexual partners of patients with new diagnoses of T. vaginalis, gonorrhea, chlamydia, chancroid, and primary or secondary syphilis should be referred for treatment.	D	A	F	E	U	A	5	9	33.33	III
13	Patients receiving outpatient therapy for PID should receive a follow-up visit within 72 hours of diagnosis.	D	A	F	E	U	A	2	2	0	III
14	All patients being treated for PID should have a microbiological re-examination (i.e., cultures) within 10 days of completing therapy.	D	A	F	L	U	A	2	2	0	III
15	Patients with a sexually transmitted disease should have a follow-up visit within 4 weeks of the diagnosis.	D	A	F	E	U	N	6	10	30	III
Well Child Care											
1	Before discharge from the hospital, the newborn should be weighed.	P	P	S	P	U	A	47	47	100	III
2	The child's weight should be taken at least four times between the end of the first week and the end of the first year of life. This information must either be plotted on a growth curve or be recorded with the age/gender percentile.	P	P	S	P	U	A	73	73	64.38	III

Chapter/ Indicator #	Indicator Text	Unit	Type	Function	Modality	Problem	Guideline	# Eligible	# Eligible Events	Mean Score	Level of evidence
3	The child's weight should be taken at least twice during the second year of life. This information must either be plotted on a growth curve or be recorded with the age/gender percentile.	P	P	S	P	U	A	65	65	60	III
4	The child's weight should be taken at least once a year during the third-sixth year of life. This information must either be plotted on a growth curve or be recorded with the age/gender percentile.	P	P	S	P	U	A	418	418	30.62	III
5	The child's weight should be taken at least once every other year during the seventh-twelfth year of life. This information must either be plotted on a growth curve or be recorded with the age/gender percentile.	P	P	S	P	U	A	418	418	12.22	III
6	Before discharge from the hospital, the newborn's height should be taken.	P	P	S	P	U	A	47	47	100	III
7	The child's height should be documented at least four times between the end of the first week and the end of the first year of life. This information must either be plotted on a growth curve or be recorded with the age/gender percentile.	P	P	S	P	U	A	73	73	63.01	III
8	The child's height should be measured at least twice during the second year of life. This information must either be plotted on a growth curve or be recorded with the age/gender percentile.	P	P	S	P	U	A	65	65	60	III
9	The child's height should be measured at least once a year during the third-sixth year of life. This information must either be plotted on a growth curve or be recorded with the age/gender percentile.	P	P	S	P	U	A	418	418	30.14	III

Chapter/ Indicator #	Indicator Text	Unit	Type	Function	Modality	Problem	Guideline	# Eligible	# Eligible Events	Mean Score	Level of evidence
10	The child's height should be measured at least once every other year during the seventh-twelfth year of life. This information must either be plotted on a growth curve or be recorded with the age/gender percentile.	P	P	S	P	U	A	418	418	12.44	III
11	Before discharge from the hospital, the newborn's head circumference should be measured .	P	P	S	P	U	A	47	47	97.87	III
12	The child's head circumference should be measured at least four times between the end of the first week and the end of the first year of life. This information must either be plotted on a growth curve or be recorded with the age/gender percentile.	P	P	S	P	U	A	73	73	53.43	III
13	The child's head circumference should be measured at least twice during the second year of life. This information must either be plotted on a growth curve or be recorded with the age/gender percentile.	P	P	S	P	U	A	65	65	52.31	III
14	Heart rate should be assessed during the first 24 hours of life.	P	P	S	P	U	A	52	52	96.15	III
15	Respiratory rate should be assessed during the first 24 hours of life.	P	P	S	P	U	N	52	52	96.15	III
16	A heart examination (in addition to heart rate) should be performed during the first 24 hours of life.	P	P	S	P	U	A	52	52	96.15	III
17	Femoral pulses should be examined during the first 24 hours of life.	P	P	S	P	U	A	52	52	84.62	III
18	An abdominal examination should be performed during the first 24 hours of life.	P	P	S	P	U	A	52	52	96.15	III
19	A hip examination should be performed at least once during the first two months of life.	P	P	S	P	U	A	115	115	65.22	III
20	A heart examination should be performed at least once between 72 hours and two months of life.	P	P	S	P	U	A	115	115	66.96	III

Chapter/ Indicator #	Indicator Text	Unit	Type	Function	Modality	Problem	Guideline	# Eligible	# Eligible Events	Mean Score	Level of evidence
21	A lung examination should be performed at least once during the first two months of life.	P	P	S	P	U	A	115	115	67.83	III
22	An abdominal examination should be performed at least once between the end of the first week and two months of life.	P	P	S	P	U	A	115	115	67.83	III
23	Examination of the femoral pulses should be performed between the end of the first week and two months of life.	P	P	S	P	U	A	115	115	63.48	III
24	A neurologic examination (e.g., tone, reflexes) should be performed at least once during the first two months of life.	P	P	S	P	U	A	115	115	66.09	III
25	A genital examination (e.g., descended testes) should be performed at least once during the first 2 months of life.	P	P	S	P	U	A	115	115	63.48	III
26	A physical examination should be documented at least twice during the second six months of life.	P	P	S	P	U	A	107	107	34.58	III
27	A physical examination should be documented at least twice during the second year of life.	P	P	S	P	U	A	65	65	41.54	III
28	A physical examination should be documented at least once a year during the third-sixth year of life.	P	P	S	P	U	A	418	418	43.3	III
29	A physical examination should be performed at least once every other year during the seventh-eleventh year of life.	P	P	S	P	U	A	378	378	14.02	III
30	If the child fails to hear the stimulus at two frequencies in one ear further evaluation/work-up should be ordered within one month.	P	A	F	L	U	N	16	16	81.25	III
31	Screening for congenital hypothyroidism should have been done by the seventh day of life.	P	P	S	L	U	A	52	52	69.23	III
32	Screening for phenylketonuria should have been done after 24 hours of age and before two weeks of age.	P	P	S	L	U	A	52	52	75	III

Chapter/ Indicator #	Indicator Text	Unit	Type	Function	Modality	Problem	Guideline	# Eligible	# Eligible Events	Mean Score	Level of evidence
33	A hemoglobin or hematocrit should be checked by the end of the 18th month of life.	P	P	S	L	U	A	32	32	37.5	III

LEGEND:

Unit: P=patient; D=dyad; E=episode

Type: P=preventive; A=acute; C=chronic

Function: S=screening; D=diagnosis; T=treatment; F=follow-up

Modality: H=history; P=physical exam; C=counseling/education; E=encounter or other intervention; M=medication; L=laboratory/radiology; I=immunization; S=surgery.

Problem: U=underuse; O=overuse

Guideline: A=adhered to an existing guideline; N=indicator was not guideline-based

Number eligible: the number of persons in the study who were eligible for the indicator

Number eligibility events: the weighted denominator for the indicator

Mean Score: the mean performance on the indicator

Level of Evidence: I: Randomized Controlled Trial; II-1: Nonrandomized controlled trials; II-2: Cohort or case analysis; II-3: Multiple time series; III: Opinions or descriptive studies

REFERENCE LIST FOR GUIDELINES USED IN THE DEVELOPMENT OF PEDIATRIC QA TOOLS INDICATORS

1. U.S. Preventive Services Task Force. 1989. Guide to Clinical Preventive Services: An Assessment of the Effectiveness of 169 Interventions. Baltimore, MD: Williams and Wilkins.
2. American Academy of Pediatrics, Committee on Psychosocial Aspects of Child and Family Health, 1985-1988. 1988. Guidelines for Health Supervision II, Second ed. Elk Grove Village, Illinois: American Academy of Pediatrics.
3. American Academy of Family Physicians. 1994. Adolescent Health Care. 1993-1994 Compendium of AAFP Positions on Selected Health Issues. American Academy of Family Physicians, Kansas City, MO.
4. Elster, AB, and NJ Kuznets. 1994. AMA Guidelines for Adolescent Preventive Services (GAPS): Recommendations and Rationale. Chicago, IL: American Medical Association.

5. Green, M. 1994. Bright Futures: Guidelines for Health Supervision of Infants, Children, and Adolescents. Arlington, VA: National Center for Education in Maternal and Child Health.
6. National Asthma Education Program. August 1991. Guidelines for the Diagnosis and Management of Asthma. U.S. Department of Health and Human Services, Hyattsville, MD.
7. Depression Guideline Panel. April 1993. Depression in Primary Care: Volume 1. Detection and Diagnosis. Clinical Practice Guideline, Number 5. AHCPR Publication No. 93-0550. Rockville, MD: U.S. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research.
8. Depression Guideline Panel. April 1993. Depression in Primary Care: Volume 2. Treatment of Major Depression. Clinical Practice Guideline, Number 5. AHCPR Publication No. 93-0551. Rockville, MD: U.S. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research.
9. American Academy of Pediatrics, Committee on Children With Disabilities. May 1994. Screening Infants and Young Children for Developmental Disabilities. Pediatrics 93 (5): 863-5.
10. Baraff LJ, JW Bass, GR Fleisher, et al. 1993. Practice Guideline for the Management of Infants and Children 0 to 36 Months of Age with Fever without a Source. Pediatrics 92 (1): 1-12.
11. Centers for Disease Control. 8 August 1991. Recommendations of the Advisory Committee on Immunization Practices (ACIP): Diphtheria, Tetanus, and Pertussis: Recommendations for Vaccine Use and Other Preventive Measures. Morbidity and Mortality Weekly Report 40 (RR-10): 1-28.
12. Centers for Disease Control. 7 April 1989. Recommendations of the Advisory Committee on Immunization Practices (ACIP): General Recommendations on Immunization. Morbidity and Mortality Weekly Report 38 (13): 205-27.

13. Centers for Disease Control. 9 June 1989. Recommendations of the Advisory Committee on Immunization Practices (ACIP): Mumps Prevention. Morbidity and Mortality Weekly Report 38 (22): 388-400.
14. Centers for Disease Control. 9 February 1990. Recommendations of the Advisory Committee on Immunization Practices (ACIP): Protection Against Viral Hepatitis. Morbidity and Mortality Weekly Report 39 (RR-2): 1-26.
15. Centers for Disease Control. 23 November 1990. Recommendations of the Advisory Committee on Immunization Practices (ACIP): Rubella Prevention. Morbidity and Mortality Weekly Report 39 (RR-15): 1-18.
16. Centers for Disease Control. 11 January 1991. Recommendations of the Advisory Committee on Immunization Practices (ACIP): Haemophilus B Conjugate Vaccines for Prevention of Haemophilus Influenzae Type B Disease among Infants and Children Two Months of Age and Older. Morbidity and Mortality Weekly Report 40 (RR-1): 1-7.
17. Centers for Disease Control. 24 May 1991. Recommendations of the Advisory Committee on Immunization Practices (ACIP): Prevention and Control of Influenza. Morbidity and Mortality Weekly Report 40 (RR-6): 1-15.
18. Centers for Disease Control. 9 April 1993. Recommendations of the Advisory Committee on Immunization Practices (ACIP): Use of Vaccines and Immune Globulins for Persons with Altered Immunocompetence. Morbidity and Mortality Weekly Report 42 (RR-4): 1-18.
19. Centers for Disease Control. 28 January 1994. Recommendations of the Advisory Committee on Immunization Practices (ACIP). Morbidity and Mortality Weekly Report 43 (RR-1): 1-38.
20. Centers for Disease Control. 6 January 1995. Recommended Childhood Immunization Schedule--United States, January 1995. Morbidity and Mortality Weekly Report 43 (51 & 52): 959-61.
21. Stool, SE, AO Berg, S Berman, et al. July 1994. Otitis Media with Effusion in Young Children: Clinical Practice Guideline. Agency for Health Care Policy and Research, Public Health Service, U.S. Department of Health and Human Services, No. 12. Rockville, MD.

22. Centers for Disease Control. 24 September 1993. Sexually Transmitted Diseases Treatment Guidelines. Morbidity and Mortality Weekly Report 42 (RR-14): 1-102.

23. Centers for Disease Control. 26 April 1991. Pelvic Inflammatory Disease: Guidelines for Prevention and Management. Morbidity and Mortality Weekly Report 40 (RR-5): 1-25.

24. Centers for Disease Control. May 1991. Sexually Transmitted Diseases: Clinical Practice Guidelines. Atlanta, GA: U.S. Department of Health and Human Services.

Appendix 3

Non-Respondent Subgroups

	Starting Sample	Phone Interview Non-Respondents	Phone Interview Non-Response Rate (%)	Written Consent Non-Respondents*	Overall Non-Response Rate (%)
Total - Overall	3698	847	22.9	2162	58.5
<i>Gender</i>					
Male	1887	425	22.5	1117	59.2
Female	1811	422	23.3	1045	57.7
<i>Age</i>					
0-5 years	1103	251	22.8	621	56.3
6 - 10 years	961	209	21.7	564	58.7
11 - 15 years	1008	238	23.6	626	62.1
16 - 17 years	496	149	30	351	70.8
18+ years	130	0	0	0	0
<i>Ethnicity</i>					
White	2655	586	22.1	1473	55.5
African American	450	88	19.6	289	64.2
Native American/Asian/Other	127	34	26.8	95	74.8
Hispanic	466	139	29.8	305	65.5
<i>Household Income</i>					
<= \$15000 per year	483	82	17	273	56.5
\$15,000 - \$50,000 per year	1411	280	19.8	811	57.5
>= \$50000 per year	1804	485	26.9	1078	59.8
<i>Utilization (during two years prior to survey)</i>					
No Hospital stays	3456	795	23	2036	58.9
1 Hospital stay	202	43	21.3	105	52
>= 2 Hospital stays	40	9	22.5	21	52.5
No physician visits	539	149	27.6	391	72.5
1-2 physician visits	1420	347	24.4	837	58.9
3-4 physician visits	875	184	21	470	53.7
>=5 physician visits	864	167	19.3	464	53.7
<i>Parent/guardian-reported health status</i>					
Excellent health	2164	518	23.9	1318	60.9
Very good health	1029	222	21.6	579	56.3
Good health	361	81	22.4	193	53.5
Fair/Poor health	144	26	18.1	72	50

*This group represents 277 parents who provided written consent to review their child's medical records but for whom we received no medical records and 1,885 parents who refused to provide written consent.

TECHNICAL APPENDIX

The Quality of Ambulatory Care Delivered to Children in the United States

Elizabeth A. McGlynn, Ph.D., Rita Mangione-Smith, M.D., M.P.H., John Adams, Ph.D.

The methods used in this article are built on decades of quality of care research conducted at RAND and other academic institutions. This technical appendix was written to provide more detail to interested readers about the methods used in this study and the prior work on which those methods are based. We refer to the measurement system we developed for this study and the related study of adults¹ as RAND's Quality Assessment Tools or QA Tools.

OVERVIEW OF CONCEPTUAL APPROACH

Recent years have seen the proliferation of numerous (and often contradictory) standards of quality for the process of health care. Medical societies, governmental agencies, health plans, and academic centers have all been active in developing mostly condition- or procedure-specific measures. Sometimes these measures narrow the focus even further to care occurring in specific clinical sites such as the emergency room.

This approach to the development of quality measures has several problems. First, condition-specific measures lead to condition-specific responses on the part of providers, perhaps at the expense of care for other conditions. For example, knowing that they will be measured on the quality of immunizations may cause plans and providers to skimp (intentionally or unintentionally) on the unmeasured quality of care for developmental screening. Second, if providers know which conditions are being evaluated, they may alter documentation or other practices to "game the system" and improve the assessment of their performance on the quality measure without actually improving the quality of care. Finally, policymakers and consumers may find summary measures of quality more useful than discrete ones in evaluating choices among health plans or providers. Many patients are looking for a doctor who can meet a range of health care needs. Health plans typically contract with physicians for a range of services rather than for a specific condition. In both of these instances, a broader measure of quality is more consistent with the information being sought.

One possible solution to this dilemma is to develop a more comprehensive measure of quality. A global quality measure that incorporates well-developed, condition-specific measures into an integrated system of quality indicators would be broad enough to preclude shifting of resources or "gaming" to improve quality scores, and would be more useful to consumers and purchasers because it is designed to produce aggregate scores. Such a system could stimulate quality improvement by both producing overall quality scores as well as scores specific to conditions, functions or sites. We built on work done by Nutting and colleaguesⁱⁱ and Lawther and colleagues.ⁱⁱⁱ Our conceptual framework guided the selection of a representative and comprehensive set of indicators. Our taxonomy of quality (Figure 1) has three dimensions: type of care, function of medicine, and modality.

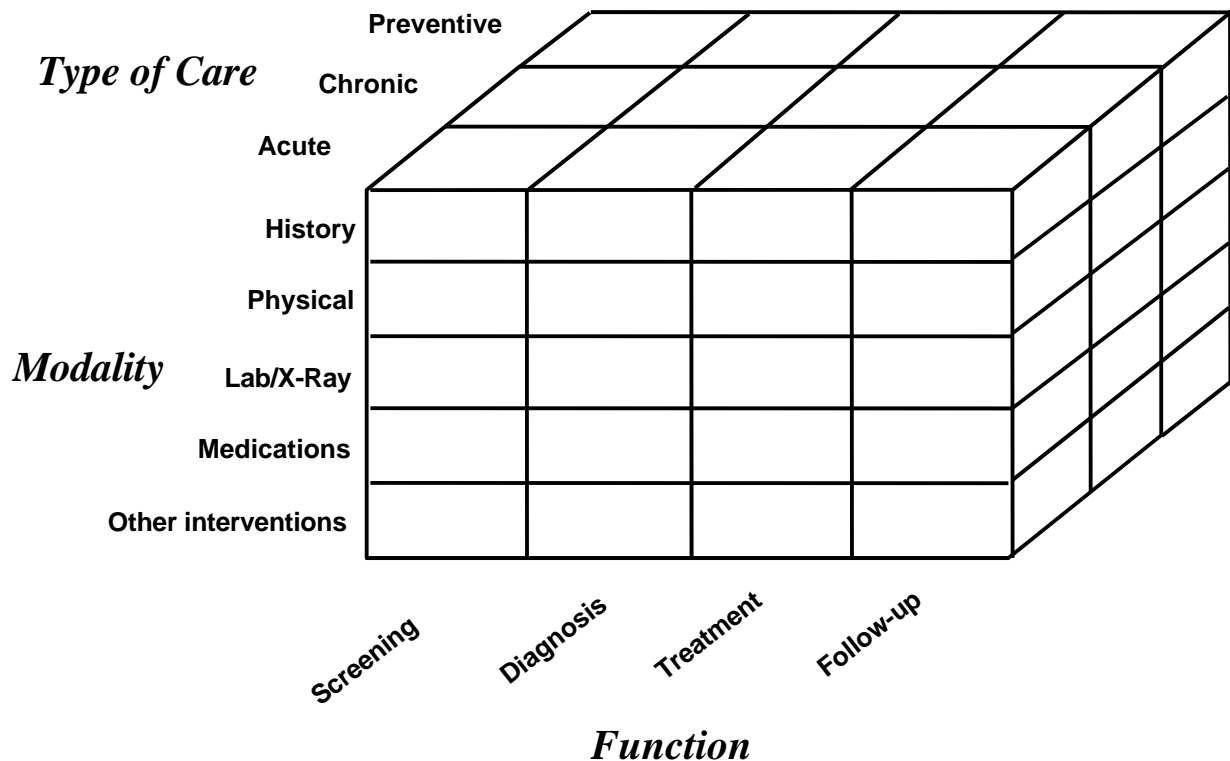


Figure 1. Conceptual Framework for RAND's QA Tools System

Type of care: Quality of care may vary greatly across types of care (acute, chronic, preventive) even among similar providers.^{iv,v} An acute condition like urinary tract infection requires a different approach to both the delivery and measurement of process quality than a chronic condition like asthma. High quality preventive care requires planning to ensure that services are delivered at recommended intervals. The management of chronic conditions requires proactive management to prevent or limit the development of complications. Care for acute conditions must be responsive.

We deliberately included conditions in our system representing acute and chronic problems as well as preventive care for the population of interest. The conditions had to be relatively common, associated with significant morbidity and/or healthcare utilization and amenable to effective interventions. We selected conditions representing acute and chronic problems in both physical and mental health so that the system would be reasonably comprehensive.

Function: The continuum of health care includes screening (for early detection of disease), appropriate diagnosis, effective treatment, and adequate follow-up or monitoring. Many quality measurement systems emphasize preventive care and screening because easily obtainable administrative data suffice to measure performance. Yet diagnosis or treatment of the same condition may have stronger links to the health outcomes of interest. And ongoing monitoring is necessary to ensure that a course of treatment is having the intended effect. Because we intended this system to be

comprehensive, we deliberately included indicators of quality across the continuum of health care services.

Modality: Modalities are the mechanisms through which screening, diagnosis, treatment and follow-up are provided for preventive, acute, and chronic care and include: history, physical examination, laboratory, radiology, medication, counseling and education, surgery, hospital admissions, follow-up contacts in person or by telephone, referrals, and so on. Although we did not deliberately seek to include indicators that covered all of these modalities, one can assess the comprehensiveness of the system by examining the degree to which multiple modalities are included. Many experts have suggested that we will have to redesign the health care delivery system to improve quality. The modalities reflect one important dimension of the delivery system structure and categorizing indicators by modality may point to system changes that could improve care across types of care and functions.

These three dimensions combine to form the cubical model shown in Figure 1. The figure represents the health care system (or the part of the system one is evaluating). The cube contains all the transactions that constitute recommended care processes for children. Each cell represents the intersection of a specific type of care, function and modality. For example, a cell might depict indicators of follow-up care for an acute condition (urinary tract infection in a young infant) requiring a follow-up radiology test (a VCUG). A potential indicator in this category might evaluate whether an infant diagnosed with a UTI received a follow-up VCUG. Assigning potential indicators to individual cells in the framework allows for a flexible indicator development process and graphically ensures that the system as a whole is comprehensive. The indicators can be viewed as a sample of what is contained in the cube. That sample provides an estimate of the overall quality delivered. Sections of the cube can be used to characterize quality on a particular dimension.

Although not depicted in Figure 1, our conceptual model of the scope of the quality indicator system will also take several other considerations into account. First, our conceptual model of quality must take into account changes in the ways and places health care is delivered. The care for a condition that once took place solely within the walls of the hospital may now start with office based diagnostic tests, proceed to a brief hospital stay for a surgical procedure, and end with outpatient follow-up care. We selected indicators to cover the spectrum of ambulatory sites that have traditionally received less attention in quality assessment but reflect where the majority of pediatric care occurs. Developing an indicator system that allows care to be evaluated wherever it is delivered will enable comparisons of quality to be made across organizations or entities with different structures or utilization patterns. Second, some quality indicators will vary by patient characteristics, such as age or developmental stage whereas others will vary by clinical characteristics such as severity. Third, quality measurement systems work at different levels of aggregate analysis. Some systems measure quality of care for individual patient-provider encounters, others profile physicians, still others compare plans or systems of care. While the quality standards may remain the same across levels of aggregation, both the attribution of responsibility and the necessary sampling frame will depend upon the intended level of analysis. The QA Tools system is designed to apply to any level of analysis; the specific application (i.e., the question to be answered) will require that different choices be made about the sampling frame and sample sizes. Fourth, our conceptual framework must take available data sources into account. No one source is likely to suffice in measuring overall quality. Because we wanted to develop a

system that was comprehensive and clinically detailed, we designed the system to use medical records as the primary data source. Ultimately, the optimum solution may blend lower-cost, routinely collected administrative data with more-expensive-to-collect, but more clinically detailed medical record data sources and patient surveys. Recent advances in data systems have already made it possible to merge disparate automated data systems ranging from laboratory test results to pharmacy records. Our conceptual framework can be adapted to reflect medical informatics advances over time and could guide the development of automated clinical data systems. Eventually, our system could be expanded to integrate patient-centered measures based on routine surveys.

PROCESS USED TO SELECT INDICATORS

Selecting Clinical Areas for Indicator Development

RAND staff used national data sources to identify the leading causes of mortality, morbidity, and functional limitation for different age groups in the population. The principal data sources for this review were Vital Statistics, the National Health Interview Survey (NHIS), the National Ambulatory Medical Care Survey (NAMCS), and the National Hospital Discharge Survey (NHDS). Because the conditions were originally selected in the mid-1990s, we show here the statistics from the time frame during which those selections were undertaken. Mortality rates for 1992 for children, adolescents and young adults by cause of death are shown in Table 1.^{vi} Table 2 shows the number of visits for children with conditions that were initially included in the QA Tools set in different age ranges from the 1993 NAMCS.^{vii} Activity limitations for children fall into four major areas: learning, communication, mobility, and self-care. Asthma is the most common reason for physical activity limitations among children.^{viii} ^{ix} Most of the areas listed here remain among the leading causes of mortality, morbidity, and functional limitation.

This quality assessment system was intended to encompass a substantial portion of ambulatory care. In order to estimate the percentage of ambulatory care visits covered by this system, we aggregated applicable ICD-9 codes into the clinical areas for which we developed quality indicators. We then calculated the number of visits for each condition in the 1993 National Ambulatory Medical Care Survey (NAMCS). Aggregating ICD-9 codes into the clinical areas covered by this system was an imprecise task, requiring a rather broad definition of what is “included” in each clinical area. The conditions applicable to children and adolescents cover 55 percent of the reasons for ambulatory visits in this population.

Table 1
Leading Causes of Death for Children, Adolescents & Young Adults, 1992
(Rates Per 100,000 Population)

Cause of Death	Ages 1-4	Ages 5-14	Ages 14-24
All causes	43.6	22.5	95.6
Unintentional injuries	15.9	9.3	37.8
Homicide	2.8	1.6	22.2
Suicide	*	0.9	13.0
Birth defects	5.5	1.2	1.2
Cancer	3.1	3.0	5.0
Heart disease	1.8	0.8	2.7
HIV	1.0	0.3	1.6
Asthma	*	0.3	0.5

*Unreliable data, no estimate

Table 2
Frequency of Ambulatory Visits for Conditions Included in QA Tools
as Reported in NAMCS, 1993

Clinical Area	Age Range (in years)	Frequency of Visits	Percentage of Total
Acne	11-18	151	2.7
Adolescent preventive services	11-18	115	2.1
Allergic rhinitis	2-18	344	6.2
Asthma	0-18	333	6.0
Attention deficit/hyperactivity disorder	3-18	175	3.1
Depression	11-18	48	0.9
Developmental screening	0-12	95	1.7
Diabetes mellitus	11-18	2	0.0
Diarrheal disease, acute	0-2	26	0.5
Family planning/contraception	11-18	6	0.1
Fever without known source	0-2	14	0.3
Headache	11-18	28	0.5
Immunizations	0-18	19	0.3
Medication allergies	0-18	0	0.0
Otitis media	1-3	247	4.4
Sickle cell disease	0-5	1	0.0
Tuberculosis	0-18	3	0.1
Upper respiratory tract infections	2-18	740	13.2
Urinary tract infections	0-12	28	0.5
Vaginitis and sexually transmitted diseases	12-18	63	1.1
Well child care	0-12	578	10.3
Visits with multiple codes	0-18	58	1.0
Total Included in QA Tools		3,074	55.0

Conducting Literature Reviews

The literature reviews were conducted in 1995 for the child and adolescent conditions. The literature reviews were done by a team of physician investigators, many with clinical expertise in the conditions selected for this project. The literature review process began with a MEDLINE search for clinical review articles and guidelines published in the last five years on each topic. Based on an initial inspection of the review articles for areas that were seemingly most appropriate for indicator development, each clinician investigator requested further MEDLINE searches for publications of clinical trials, guidelines and other supportive evidence or consensus statements in these specific areas. Each investigator then drafted a review of the literature for his or her topic area, focusing on important areas for quality measurement (as opposed to a clinical review of the literature, which would focus on clinical management) and drafted potential indicators. Each review and set of indicators was reviewed by a senior physician co-investigator (Dr. Mark Schuster) and went through several revisions before being finalized. On a few occasions, when questions remained even after detailed literature review, we requested that a clinical leader in the field read and comment on the draft review and indicators.

DEVELOPING QUALITY INDICATORS

The method we used to develop quality indicators is an adaptation of the RAND-UCLA modified Delphi method that was originally designed to select and evaluate clinical indications for appropriate or inappropriate uses of diagnostic and treatment procedures.^x
^{xi xii xiii xiv xv} This method is effective for combining information in the scientific literature with expert judgment. It has been used to develop quality indicators for adults,^{xvi} the vulnerable elderly,^{xvii} in the United States and the U.K.^{xviii} This method of selecting indicators is reliable and has been shown to have content, construct, and predictive validity in other applications.^{xix xx xxi} Other researchers following similar steps have demonstrated the reliability of the approach.^{xxii} In this application, staff developed clinical indicators and a panel of experts was convened to evaluate the validity and feasibility of those indicators (as compared to evaluating the appropriateness of performing a specific procedure for a patient with a particular clinical scenario). This process is described in more detail below.

In each clinical area, investigators developed indicators defining the explicit criteria by which quality of care would be evaluated. These indicators focus on technical processes of care for the various conditions and are organized by function: screening, diagnosis, treatment and follow-up. While we developed indicators across the continuum of management for each condition, we did not attempt to cover every important area nor all possible clinical circumstances. The indicators were designed to apply to the average patient with the specified condition who is seeing the average physician. We included indicators that represented different types of quality problems including overuse (providing care that is not needed), underuse (failing to provide care that is needed), and misuse (providing care that has a high probability of resulting in harm).

Our approach makes a distinction between indicators of quality care and guidelines (see Table 3). While guidelines are intended to be comprehensive in scope, indicators are meant to apply to specific clinical circumstances where there is believed to be a strong link between a measurable health care process and patient outcomes. Indicators are not intended to measure all possible care for a condition. Furthermore, guidelines are applied prospectively at the individual patient level, while indicators are applied retrospectively and scored at an aggregate level. Finally, to form useful measures

of quality based on the specific attributes of a data source, indicators must be written precisely.

Table 3
A Comparison of Key Characteristics of Clinical Guidelines and Quality Indicators

Clinical Guidelines	Quality Indicators
Comprehensive: Cover virtually all aspects of care for a condition.	Targeted: Apply to specific clinical circumstances where there is evidence of a process-outcome link.
Prescriptive: Intended to influence provider behavior prospectively at the individual patient level.	Observational: Measure provider behavior at an aggregate level; generally applied retrospectively.
Flexible: Intentionally leave room for clinical judgment and interpretation.	Precise: Precise language that can be applied systematically to medical records data to ensure comparability.

For each indicator we developed a table that showed the indicator statement, the population to whom the indicator applied, a "grade" for the strength of the evidence that supports each indicator, the specific literature used to support each indicator, a statement of the health benefits of complying with each indicator, and comments to further explain the purpose or reasoning behind each indicator. The grades assigned to the scientific evidence are those used by the Canadian Task Force on Preventive Health Services:

- I Randomized controlled trials
- II-1 Nonrandomized controlled trials
- II-2 Cohort or case analysis
- II-3 Multiple time series
- III Opinions or descriptive studies

Panel Selection

We requested nominations for potential participants in the expert panel from the relevant specialty societies: the American Academy of Pediatrics, Society for Adolescent Medicine, Ambulatory Pediatrics Association, and the American Academy of Family Physicians. Each nominee was sent a letter summarizing the purpose of the project and requesting that a curriculum vitae and calendar with available dates be returned if the candidate was interested in participating. We received positive responses from 74% of potential panelists. In selecting from this pool we sought to ensure that the panel was diverse with respect to type of practice (academic, private practice, managed care organizational practice), geographic location, gender, and specialty. The quality of the recommended panelists was excellent and availability of individuals on the selected dates became one of the main constraints. Multidisciplinary panels, in our experience, produce more balanced sets of indicators than single specialty panels.^{xxiii xxiv xxv}

Initial Panel Ratings

A draft of the literature reviews for all clinical chapters, proposed indicators and evidence supporting those indicators, and indicator rating sheets were sent to all

members of the expert panel several weeks before the panel meeting. Each panelist was asked to read the literature reviews and rate each indicator on a nine point scale for two dimensions: validity and feasibility.

The panelists were instructed that a quality indicator should be considered valid if:

1. There is adequate scientific evidence or professional consensus to support the indicator;
2. There are identifiable health benefits to patients who receive care specified by the indicator;
3. Based on the panelists' professional experience, physicians with significantly higher rates of adherence to the indicator would be considered higher quality providers;
4. The majority of factors that determine adherence to an indicator are under the control of the physician (or are subject to influence by the physician, such as encouraging breastfeeding).

Ratings of 1 to 3 mean that the indicator is not a valid criterion for evaluating quality; ratings of 4 to 6 mean that the indicator is an uncertain or equivocal criterion for evaluating quality; and ratings of 7 to 9 mean that the indicator is clearly a valid criterion for evaluating quality.

Panelists were also instructed that a quality indicator should be considered feasible if:

1. The information necessary to determine adherence is likely to be found in an average medical record, or failure to document such information is itself a marker for poor quality;
2. Estimates of adherence to the indicator based on medical record data are likely to be reliable and unbiased.

Ratings of 1 to 3 mean that it is not feasible to find the information necessary to score the indicator in the average chart; ratings of 4 to 6 mean that there will be considerable variability in the feasibility of finding the necessary information to score the indicator; and ratings of 7 to 9 mean that it is clearly feasible to find the information necessary to score the indicator.

The nine point scale has been used for more than two decades at RAND in developing explicit indicators for evaluating appropriateness and quality.^{xxvi} Essentially these methods require panelists to place indicators into one of three categories (valid criterion for quality, equivocal criterion for quality, invalid criterion for quality) and each category can be rated on a three point scale to allow panelists some variation within category. The scale is ordinal so that a 9 is better than an 8 and so on. Because quantities (e.g., risk-benefit ratios) are not assigned to each number on the scale, the difference between an 8 and a 9 is not necessarily the same as the difference between a 5 and a 6. Explicit ratings are used because in small groups some members tend to dominate the discussion and this can lead to a decision that does not reflect the sense of the group.^{xxvii}

Figure 2 Sample Panelist Summary Rating Sheet

Panelists individually rated the indicators before the panel meeting. They returned their first round ratings to RAND staff and we summarized the first round results. Panelists were encouraged to comment on all of the indicators. We conducted one-on-one discussions with panelists after they had returned their ratings to better understand

panelist A; round 1; page 1

Chapter 1 ASTHMA	Validity	Feasibility
---------------------	----------	-------------

DIAGNOSIS

3. Spirometry should be measured in patients with chronic asthma who are > 5 years old at least every 2 years.

1	2	3	1	1	3	4	2												
1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9	(1- 2)	
									^									^	

TREATMENT

7. Patients requiring chronic treatment with systemic corticosteroids during any 12 month period should have been prescribed inhaled corticosteroids during the same 12 month period.

1	6	2	2	3	4														
1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9	(3- 4)	
									^									^	

10. All patients seen for an acute asthma exacerbation should be evaluated with a complete history including all of the following:

a. time of onset

2	2	3	2	1	1	3													
1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9	(5- 6)	
									^									^	

b. all current medications

4	1	4	3	1	5														
1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9	(7- 8)	
									^									^	

c. prior hospitalizations and emergency department visits for asthma

5	1	3	5	1	3														
1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9	(9-10)	
									^									^	

d. prior episodes of respiratory insufficiency due to asthma

1	1	3	2	2	1	2	3	1	2										
1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9	(11-12)	
									^									^	

Scales: 1 = low validity or feasibility; 9 = high validity or feasibility

reasons for disagreement. In general we find two main reasons for disagreement among panelists: unclear indicator language or unclear scientific evidence. If an indicator is poorly written, panelists can disagree because they are thinking of different patient scenarios. In these circumstances, staff worked with panelists to revise the indicators to more clearly state the intent. If an indicator is based on unclear or mixed scientific evidence it is difficult to construct an acceptable quality indicator. Based on panelists' suggestions, we modified some indicators before the panel meeting. In some instances, when one or more panelists felt strongly that we had not included an indicator in an important area, we incorporated panelists' suggestions for additional indicators prior to the meeting.

Panel Meeting

Prior to the panel meeting each panelist received a summary of the first round of ratings. The individualized panelist summary was a printout of the rating sheet with the distribution of all panelist ratings displayed above the rating line. Below the rating line a caret (^) marked the panelist's own rating in the first round. For example, for Indicator 7 in Figure 2, two panelists rated validity as "9," six panelists rated validity as "7," and one panelist gave a rating of "6." The panelist whose rating sheet is shown here rated the indicator a "9" for validity. Indicators with multiple subparts (e.g., Indicator 10 in Figure 2) received a separate rating for each part. The summaries were sent out about a week before the panel meeting to allow the panelists time to compare their ratings to those of their colleagues. Prior to the panel meeting, the research staff also contacted each panelist individually to discuss sources of agreement and disagreement identified from the initial ratings.

The panel met face-to-face in a two day meeting held in the Fall of 1995 to provide final ratings for the indicators. After a brief introduction to the study and its goals, panelists reviewed and revised the definitions of key terms, discussed, and then re-rated each of the indicators. Discussions were organized by clinical chapter and were designed to ensure that the panelists had a common understanding of the content and application of each indicator. The panel chair, Dr. David Bergman, led the discussion of the indicators. The principal investigators and the physician(s) who had developed the indicators participated in the discussion in order to answer questions about the literature supporting particular indicators and to comment on how modifications might affect the ability to operationalize the indicators. All literature that had been reviewed and additional literature recommended by the panelists prior to the meeting was available for inspection in the meeting (i.e., to resolve differences in panelists recall about the findings of certain studies).

Analysis of the Indicator Ratings

The median was used to measure the central tendency for the nine panelists' ratings, and the mean absolute deviation from the median to measure the dispersion of the ratings. The final disposition of each indicator is based on its median validity and feasibility scores. To be included in the final set, an indicator had to have a median rating of 7 to 9 on validity and 4 to 9 on feasibility (Table 4).

Table 4
Criteria for Indicator Inclusion

Median Validity Rating	Median Feasibility Rating	Indicator Disposition
1-3	1-3	Exclude
1-3	4-6	Exclude
1-3	7-9	Exclude
4-6	1-3	Exclude
4-6	4-6	Exclude
4-6	7-9	Exclude
7-9	1-3	Exclude
7-9	4-6	Include
7-9	7-9	Include

The only exception to this rule is when there is substantial disagreement among panelists on validity; these indicators are also excluded. To determine agreement and disagreement among panelists, we adopted a statistical definition that could be applied regardless of the number of panelist ratings available. This approach frames the definitions of “agreement” and “disagreement” in terms of hypotheses about the distribution of ratings in a hypothetical population of repeated ratings by similarly selected panelists. For agreement we test the hypothesis that 80 percent of the hypothetical population of repeated ratings are within the same region (1-3, 4-6, 7-9) as the observed median rating. If we are unable to reject that hypothesis on a binomial test at the 0.33 level, we say that the indicator is rated “with agreement.” For nine ratings, this definition of agreement requires that no more than two of the ratings be outside the three-point region that contains the median.

For disagreement, we test the hypothesis that 90 percent of the hypothetical population of repeated ratings are within one of two extra-wide regions (1-6 or 4-9). If we have to reject the hypothesis on a binomial test at the 0.10 level, we conclude that the indicator is rated “with disagreement.” For nine ratings, this definition of disagreement is satisfied when three or more ratings are in the 1 to 3 region and three or more are in the 7 to 9 region. Finally, if the ratings cannot be classified as “with agreement” or “with disagreement,” then they are classified as “indeterminate.”

The literature reviews, final panel ratings, and summary of the disposition of staff recommended indicators are available on RAND’s website.^{xxviii} The original QA Tools system for children and adolescents contained 453 indicators in 22 clinical areas. Some clinical areas were dropped from the current study because in a random sample of children, few or no children were eligible for indicators in the chapter (e.g., diabetes, sickle cell, tuberculosis). Other clinical chapters are not included in the main results reported in the current study because manuscript reviewers were concerned that the indicators were subject to significant under-reporting bias in medical records (e.g., developmental screening).

Keeping Indicators Up-to-Date

A common question when examining quality of care research is whether the indicators used reflect the most current standards of practice. Since the research reported here

(and in most similar studies) is based on a retrospective review of medical records, the critical concern is whether the indicators we used were consistent with standards during the time period of review (in this case 1996-2000). We reviewed the indicators to ensure that no major changes had occurred that would make an indicator invalid for any or all of the time period studied.

After the study reported here was completed, RAND staff updated the indicator sets. This process of updating is something that would be necessary to do routinely and systematically to ensure that any indicator system remains consistent with best evidence and that the quality indicators being used are appropriate to the time frame of the study. In the 2004 update, 74% of indicators used in this study were retained in their original form, 19.4% were retained with minor revisions, 4.5% underwent major revisions, and 2.1% were dropped. Thus, despite concerns about the degree to which science is changing the standards by which care is delivered, we have found a high level of consistency over time in the quality of care indicators that are appropriate. This finding is consistent with what we have found in previous work.

DEVELOPING SPECIFICATIONS FOR INDICATORS

There are two different approaches to evaluating quality of care in medical records: implicit review and explicit review. Under implicit review, medical records are reviewed by physicians who give either an overall assessment of quality or who rate quality for different aspects of care (e.g., diagnostic process, treatment of a specific condition).^{xxix xxx xxxi} Under explicit review, detailed information is collected from medical records to determine whether care processes were delivered consistent with a standard set by the study team.^{xxxii} We use the explicit approach in the QA Tools system.

The process of designing a measurement strategy to ensure that data elements necessary for determining eligibility for an indicator and scoring of that indicator are collected efficiently and reliably is called specifications development. The indicators in the QA Tools system can be thought of as a research question that we must endeavor to answer from data that are available in medical records (e.g., how often do children with persistent asthma receive appropriate medical therapy?). We describe here our approach to developing specifications for indicators.

Indicator-Level Analysis

We begin by examining each indicator to determine the criteria for eligibility (i.e., how will we identify those to whom the indicator should be applied). Most indicators apply to a subset of persons in a sample. For example, eligibility for preventive care indicators, such as immunizations, is generally determined by age. Eligibility for an acute care problem, such as an upper respiratory infection is often based on presenting symptoms, reason for the visit, or a diagnosis. Eligibility for a chronic care indicator is usually based on the presence of a diagnosis in the chart. The second step is to determine what clinical process is consistent with the intent of the indicator. In the QA Tools system we are deliberately evaluating whether the physician (or other health care professional) took actions that are consistent with an effort to deliver indicated care. For example, if an indicator requires that a laboratory test be done, we will look for whether the test was ordered (a note), done (a result), or offered and refused (a note). This approach in most cases separates what the physician intended to do from whether the patient followed through. Third, we write the scoring rules that determine whether or not the indicator has

been met. Finally, we identify the data elements (referred to as variables) that are necessary to determine eligibility, assess adherence, and score the indicator. The data element may be expressed as a question to be answered by the abstractor, or it may be a derived variable that is a combination of multiple data elements.

Take for example the asthma indicator shown in the box. The purpose of this indicator is to ensure that providers limit the use of systemic (oral) steroids in patients with asthma because the side effects associated with these agents are significantly worse than the side effects of inhaled preparations.

In any patient requiring chronic treatment with oral corticosteroids, a trial of inhaled corticosteroids should have been attempted.

A patient is eligible for this indicator if he/she has a diagnosis of asthma and has received chronic treatment with systemic corticosteroids. We have to create an operational definition of “chronic” treatment with this medication. Defining clinical terms involves a combination of looking at the literature (including the National Heart Lung and Blood Institute (NHLBI) Guidelines in this instance),^{xxxiii} consulting with the research project’s clinical team (both nurses and physicians), and if necessary consulting with an expert in the clinical area. Using these methods we defined chronic as 30 or more days of treatment with oral corticosteroids or 2 or more prescriptions for a corticosteroid taper within a 3 month period. Among those receiving chronic treatment with oral corticosteroids we then look to see if they received a prescription for inhaled corticosteroids within the same 12 month time window. This in turn requires that the dates for the beginning and end of the 12 month time window be specified. For example, should the 12 month clock start ticking from the time of the first prescription for oral corticosteroids and search forward for inhaled corticosteroids (i.e., before it is known that the patient will require chronic treatment) or should the clock start ticking at the point that the chronic definition is met and search backwards (i.e., to see whether steps were taken to avoid chronic use of systemic steroids)? The challenge is balancing the intent of the panel or guideline writers with concrete instructions that result in comparable data being collected. We chose to look back from the point at which criteria for chronic use were met (2 tapers within a 3 month period, 30 or more days continuous use) to see whether inhaled corticosteroids were prescribed. To score this indicator we need both the pattern of prescriptions (type and number of days) but also the dates associated with each prescription and drug classifications. We typically collect all dates from all providers for processes of interest (e.g., laboratory orders, prescriptions, visits) so that we can analytically construct each patient’s profile across all providers. This also minimizes the calculations that abstractors are asked to do (e.g., whether a prescription falls within a certain time frame). Abstractors may be asked to verify whether an action was taken in response to a particular clinical situation or to provide other information that sets the context for interpreting the pattern of care that is observed.

This analysis is done for each indicator in the system. The process was overseen by a project manager and was closely supervised by physicians who managed the literature reviews and panel meetings and the principal investigator. The oversight ensures that similar situations encountered in indicators are treated in the same way so that the system is consistent across indicators (within the bounds of clinical intent). Occasionally in doing the indicator analysis we identified indicators that could not be reliably specified. In such instances the indicator was dropped from the system. In the

pediatric system, this most frequently occurred with indicators added by the panel because the language and review process was not as extensive as for indicators originally proposed by staff.

The indicator analysis also identified terms that required definition and situations that required instructions for the abstractor. This information was made available to the abstractor through access to “help” windows that contained definitions, synonyms, or instructions about where to find data or about exclusion criteria. For example, in a question asking about whether a child had a temperature > 38.5°C, the abstractor was instructed to document the highest temperature recorded by the parent prior to the visit, whether an anti-pyretic had been administered by the parent, and the timing of the administration. All temperatures taken in the office were recorded as well. A question about whether a bacterial diagnosis was made would include the various terms an abstractor might encounter in the record (e.g., sinusitis, acute otitis media). A question about growth parameters (height, weight, percentiles) being assessed might point the abstractor to the growth chart in the front of the chart as well as provider notes during visits. The help windows are designed to enhance inter-rater reliability and to increase the efficiency of data collection by maintaining within the data collection tool environment all needed information. The help lists were updated throughout the abstraction process to include additional notation encountered in the field.

Module-Level Analysis

Once all of the indicators in a module (i.e., clinical chapter, such as asthma) were analyzed, the module level analysis was conducted. The purpose of the analysis was to remove unnecessary duplication of data elements, develop branching logic to make data collection more relevant (i.e., the abstractor only sees questions that are relevant to the chart being abstracted), and to organize data collection so that it reflected the organization of most charts (e.g., laboratory test results are often kept together in a separate part of the chart).

The eligibility questions were organized at the beginning of the module. For example, for acute problems, we required that the patient be seen for the condition of interest during the study period. Alternatively, particularly for patients with chronic diseases, we looked for a history of a particular condition whether or not the patient was seen or diagnosed with the condition during the study period. The latter approach increased the likelihood that we could identify underuse of needed medical care.

Once a patient was determined to be eligible for a particular module, data collection was organized around the data elements necessary to determine eligibility for the specific indicators within the module and to determine whether the indicated care was received. For example, in the asthma module, the abstractor was asked questions about each visit the patient made for asthma, including: the purpose of the visit (exacerbation, follow-up, other), whether a beta2 agonist was administered during the visit, and what the peak flow value was if one was taken at the visit. This reflects the effort to organize data collection to match typical charting practices. For example, rather than collecting all peak flow values and then going back and asking the reason for each visit (a variable level method of organizing data collection), all visit-level information is collected at the same time. Other branching logic instructions are included in the table. For example, if a beta agonist was administered at the visit and a peak flow was obtained prior to administration of the medication, a post-medication peak flow was requested. If no beta agonist was

prescribed or if a pre-medication peak flow was not taken, then the abstractor was not asked for the second peak flow.

Another key element is the development of medication lists, which identify, for the abstractor, all medications that qualify for a variable (e.g., beta agonists). The development of these lists required that we identify all medication classes that meet the specifications of the variable and then further identify all specific medications (generic and brand name) within the class. This allowed the abstractor to enter the medication as written in the chart. Using the classification scheme, medications that meet the classification requested for a particular variable can be displayed so that the abstractor can select the one that applies to the question being answered. This relieves the abstractor of having to memorize the uses and types of all possible medications (an impossible task) or to consult a separate reference (e.g., the *Physicians Desk Reference*) to classify a medication (a time consuming endeavor that is unlikely to be reliable).

Worksheet-Level Analysis

The other key step in the specifications process was the design of a front-end mechanism for organizing the chart abstraction process, giving the abstractor a systematic way to begin obtaining information from each chart. This approach provided an additional level of efficiency by identifying data elements that were necessary at every visit regardless of the health problem(s) addressed, data elements used in multiple modules, or data elements most easily obtained while looking at the notes about a visit.

Each documented contact between a patient and a provider was collected along with information that was needed in general or for multiple modules. The figure below is a screen capture from the worksheet portion of the data collection tool. Note that there are several tabs on the worksheet, each for a separate section of the front end data collection. On the contacts tab that is displayed, the abstractor was asked for all contact dates and types (visits, emergency department, hospitalizations, phone calls). At each contact, the abstractor is asked to note whether smoking status was documented (only applicable to adolescents), systolic and diastolic blood pressure, discharge date (hospital stays only), and whether percentile height and weight were recorded at the visit. In the bottom half of the tabs contact sheet information was obtained on all diagnoses recorded at the contact and all medications prescribed or noted by history. Additional tabs structured data collection for screenings and immunizations. The abstractor could at any time during data collection go to the two summary tabs to see a listing of all diagnoses and medications collected for the chart.

Medical Record Data Collection Tool

File Search View Tools Help

1to10

Patient CHT_P5 Provider 1_TRAINING Patient CHT_P5 (Incomplete)

Medical Records

- Not Started
- Incomplete
 - Provider 1_TRAINING
 - Patient CHT_P5
 - Provider AR10177
 - Patient 060220901
- Completed
- Exported
- Archived
- Not Abstracted

Contacts | Screenings | Immunizations | DX Summary | Med Summary

Worksheet is complete

Patient/Provider Contacts

Contact Date	Contact Type	Smoking Status	Systolic BP	Diastolic BP	Discharge Date	Percentile Height	Percentile Weight
01/01/1997	Outpatient Medical	NA	NA	NA	NA	Yes	Yes

Diagnosis

01/01/1997 : Outpatient Medical	
Problem	Type
Asthma	Pre-existir

Medications

01/01/1997 : Outpatient Medical						
Name	Dose	Unit	Route	Frequency	Verify	
Albuterol Sulf	2	inhalation	inhalation	qid	<input checked="" type="checkbox"/>	

Edit Problems Save Demo graphics Main

Worksheet: Contacts 0;-1,True,11/02/1995,11/02/1995,04/20/1987 5/8/2007 6:47 AM

Start | Microsoft Outlook | Gmail - Inbox (26) - Micr... | Medical Record Data ... 6:47 AM

In completing this worksheet, the abstractor obtained an overall view of care received by the patient during the study period. Information obtained in this step was used to create appropriate branching logic to make subsequent data collection more efficient, for example, by using the problems addressed during a contact to determine the modules for which a patient was likely to eligible. Reports could be generated from the worksheet to check for accuracy of abstraction.

SUMMARY OF SPECIFICATIONS DEVELOPMENT

We used an iterative process to develop the data collection strategy for the QA Tools system. We began by creating the analysis plans using “pseudo-code” – phrases describing the data elements needed to determine whether a patient was eligible for or received the indicated care process. We next developed the indicator-level, module-level, and worksheet-level analyses to determine the specific way in which data would be collected. We then went back to the analysis plans to check whether each of the required data elements was represented in the data collection strategy. These iterative steps informed the development of the medical record abstraction software by highlighting needed functionality (e.g., using dates to identify the first occurrence of an event of interest within a condition module).

The development of detailed specifications is critical to ensuring that data are collected with adequate inter-rater reliability and validity. The approach we take is systematic and involves multiple levels of review. We endeavor to minimize judgment and calculations on the part of abstractors except where such judgment is superior to answering a structured set of questions (often true in complex clinical situations). The specifications are deliberately designed to give health care providers the benefit of the doubt. We allow for multiple different ways to document care processes.

RECRUITING THE PARTICIPANTS

In conducting the Community Quality Index (CQI) study, we had an opportunity to collaborate with investigators at the Center for Studying Health System Change (CSHSC) who were already conducting a national study of changes in the organization and financing of health care. This study, known as the Community Tracking Study (CTS), was funded by the Robert Wood Johnson Foundation, and had been underway for two years. The CTS did not originally include a component to evaluate the quality of medical care delivery but the investigators at CSHSC believed that adding such a component would add significantly to their ability to answer questions about the impact of observed changes in health care markets on residents in those markets. Because CSHSC was interested in changes at the market level, the investigators selected a random sample of 60 communities (defined as metropolitan statistical areas) that were representative of the United States population. In 12 communities (also randomly selected from the 60), they deliberately recruited a large enough sample of households to produce community-specific estimates of different aspects of market structure and change (e.g., rates of uninsurance, rates of managed care penetration). The CSHSC recruited participants through a random digit dial telephone survey. The CSHSC used the household as the unit of analysis and collected information from a single adult in the household. The CSHSC obtained information about all members of the household and randomly selected one minor child to include in the study (so all questions about children’s health insurance and utilization were asked about this one child).

We weighed the advantages and disadvantages of collaborating with the CSHSC during the design phase of the CQI. A major disadvantage was that we would start with a sample that had already experienced non-response. A major advantage was that we would have a rich set of variables about demographics, insurance, and utilization that could be combined with the data we collected from medical records to provide an understanding of the relationship between financial aspects of health care delivery and

clinical quality. We decided that the benefits of collaboration outweighed the threats to generalizability.

The Community Tracking Study (CTS) recruited households in 12 metropolitan areas (Boston, MA; Cleveland, OH; Greenville, SC; Indianapolis, IN; Lansing, MI; Little Rock, AK; Miami, FL; Newark, NJ; Orange County, CA; Phoenix, AZ; Seattle, WA; and Syracuse, NY) through a random digit dial telephone survey. We obtained a list of all households that had participated in the second round of the CTS in the 12 communities (referred to by the CTS as the “intensive sites”) and their telephone numbers.

Between October 1998 and August 2000, we telephoned participating households that had a child enrolled in the CTS. The average length of time between the CTS interview with the last household member and the CQI interview was about 6 months. We interviewed the adult in the household most familiar with the child’s medical history to obtain demographic information and both oral and written consent to request copies of the child’s medical records from all providers seen in the two years prior to the date of the interview. The results are based on care delivered between October 1996 and August 2000.

Response Rates

We began with an initial sample of 4,096 children who had participated in the CTS, which had a net response rate of 62.5%. Of these, 398 (10%) were deemed ineligible, primarily because their families had moved. We interviewed the parents of 2,851 (77%) of the 3,698 eligible children and excluded 77 (3%) because they had not seen a healthcare provider in the prior two years. Among the 2,774 children who had at least one visit, we received oral permission to obtain records for 2,415 (87%) and written consent for 1,813 (65%).

Obtaining Records

We obtained a separate written consent form for each provider seen by each child during the study period. RAND subcontracted with vendors that had experience in chart retrieval to obtain the charts. The vendors contacted providers and supplied them with a photocopy of the consent form. RAND, through the vendor, reimbursed providers for the costs of photocopying charts. In cases where the provider did not have the capacity to photocopy charts, the vendors went on site with portable photocopy machines to obtain the copies directly. We received 63% of the medical records for which we had written consent. We obtained at least one medical record for 1,536 of the children for whom we had written consent (85%). Children with at least one medical record were included in the analyses (42% of eligible children).

Recruiting and Training Abstractors

RAND contracted with a local employment agency that specialized in providing registered nurses for a variety of time-limited assignments including chart abstraction. We interviewed and selected those with the most relevant prior experience to participate in the training program. The training program was conducted over 10 days. Abstractors completed assignments each day designed to test their uptake of the material. Some abstractors were asked to (or chose to leave) during the early days of the training because they were not able to perform at an acceptable level. The final test of readiness

was abstraction of a “gold standard” chart. We required 90% agreement on all items in order for abstractors to be selected for the research study. Two abstractors served as supervisors throughout the project and were responsible for ongoing monitoring of the quality of abstraction. Intermediate tests of performance were conducted along with additional training on specific issues that arose during abstraction.

Abstracting Charts

All charts were sent to RAND for abstraction. We developed computer-assisted abstraction software on a Visual Basic platform (version 6.0, Microsoft). The software allowed abstraction to be tailored to the record being reviewed and provided range and consistency checks, calculations (e.g., the presence of fever), and classifications (e.g., drug class) during abstraction. Seven trained registered nurses abstracted from charts the variables necessary to determine eligibility and scoring of indicators. Charts were abstracted separately for each healthcare provider of each child.

To assess inter-rater reliability, we re-abstracted charts from a random sample of 10% of participants (n=160). Average reliability, using the kappa statistic, ranged from substantial to almost perfect at three levels: the presence or absence of a given condition (k=0.89, 95% CI 0.86 to 0.91), the participant’s eligibility for an indicator (k=0.95, 95% CI 0.94 to 0.96), and the score for an indicator (k=0.83, 95% CI 0.80 to 0.85).

ANALYTIC APPROACH

The analyses involved five steps: (1) determining eligibility for and scoring of individual indicators; (2) aggregating indicators by 6 categories; (3) applying sampling weights at the individual patient level; (4) constructing standard errors; (5) preparing results tables.

Determining Eligibility and Scoring of Indicators

For each indicator, algorithms were developed to determine whether each sampled patient was eligible for (yes/no) and received (yes/no or proportion) the process specified by the indicator. For example, in the Fever module, a child was eligible for one indicator if: (1) he/she was at least 28 days and less than 90 days old, and (2) presented with fever of unknown source in the outpatient setting and not subsequently admitted to the hospital, and (3) [a] was febrile by oral, rectal, or axillary measurement or [b] was afebrile but had received antipyretics within 8 hours prior to presentation. Patients eligible for this indicator passed if a complete blood count and urinalysis or urine culture was ordered or obtained on the same date as the presentation.

The principal investigator, analysts, and the lead physician (Dr. Mangione-Smith) reviewed the patterns of eligibility for each indicator to identify those that were inconsistent with disease prevalence (eligibility). Given our focus on specificity, we would expect to find eligibility rates that were lower than population-based prevalence for a disease, however, if we found eligibility rates that were considerably lower than expected, we investigated whether this was due to problems with data collection, documentation, or programming errors. Similarly, we reviewed all of the scoring results to identify potential problems with data collection, chart documentation, or programming. The principal investigator and lead physician re-reviewed a substantial number of charts to assess the

validity of the eligibility and scoring results. Any problems identified in a particular clinical area that were generalizable to other areas triggered a systematic review across all indicators (e.g., unexpectedly low scores for performing indicated physical exam elements during well child care visits triggered such a review).

Creating Aggregate Scores

We had previously categorized indicators by: (1) level of evidence; (2) type of condition (preventive, acute, chronic); (3) function of medicine (screening, diagnosis, treatment, follow-up); (4) modality of care (history, physical examination, laboratory or radiology test, medication, other); (5) type of quality problem (overuse, underuse, misuse); and (6) clinical area (e.g., asthma).

Each indicator was scored at one of three levels: patient, patient-provider dyad, or episode, depending on the nature of the process being evaluated. The level at which an indicator was scored affected the number of eligibility events (the denominator for an aggregate score). We scored patient-level indicators as passing if at least one of the child's health care providers delivered the indicated care (e.g., immunizations). Dyad level indicators (e.g., duration of nasal decongestants limited to four days), were scored separately for each provider that saw the patient for a care process. Episode level indicators generally require coordination of care across providers (e.g., patients with persistent bilateral otitis media should have a hearing evaluation) allowing for eligibility and scoring to occur with different providers.

To produce composite scores, we divided all instances in which recommended care was delivered for a given eligibility event by the number of eligibility events within the category. This has been called an "opportunity score."^{xxxiv} The results are presented as proportions theoretically ranging in value from 0-100 percent and are interpreted as the proportion of recommended care that was delivered. A simple example is shown in Table 5. In this example there are three patients who contribute a total of 15 eligibility events (eligible count or denominator) to the overall composite score. Each child contributes a different number of events to the denominator reflecting the number of times that child had (or should have had) contact with the health care system for care in one of the measured areas. The successful delivery of care is similarly summed across all patients resulting in a total of 9 for the numerator. This creates a composite score of 9/15 or 60%. The weights in this case are the frequency with which the "system" is required to produce a particular process.

Table 5
Illustration of Approach to Constructing Composite Scores

Patient	Indicator	Eligible Count	Score Count	Patient Average
001	Adolescent well visit	1	0	
001	Asthma controller meds	1	1	
001	Asthma exacerbation mgmt	3	1	.40
002	Well child visits	4	2	.50
003	Otitis media	2	2	
003	Well child visits	4	3	.83
Totals		15	9	
	Composite		60%	
	Average of patient averages			58%

We considered several alternate approaches to scoring which are briefly described here along with the rationale for not selecting the approach. First, we could have constructed a composite that was the average of the patient's averages. This is shown above in the last column where we first create a patient-specific average and then create a simple average of the patients' averages (58%). Although in the simple example the numbers are not very different from one another, the content of the score is focused on the average patient rather than the demand on the system. Second, we considered adding clinical importance weights, which could be included as an additional step in either of these two approaches. We elected not to do this because the outcomes included in our system are heterogeneous (making it difficult to select an appropriate summary weighting factor) and because prior studies had shown that such weights had to be orders of magnitude different from one another to significantly change the summary score.^{xxxv} The approach we used is currently being used by CMS in the Premier Hospital Demonstration Project and has been used by other researchers to create summary scores.

Applying Non-Response/Sampling Weights

The sampling weights reflect the original sampling process of the CTS study. The non-response weights reflect the characteristics of children we ultimately failed to obtain medical record data for relative to the originally sampled children. All of the scores for the whole population and the various subsets are weighted sums of the passed measures divided by similarly constructed weighted sums of the eligibility events.

Estimating Standard Errors

We used the bootstrap method to directly estimate standard errors for all of the aggregate scores.^{xxxvi} The bootstrap method employs repeated resampling from the observed data and uses the change in results from sample to sample to estimate standard errors. A key to proper implementation of the bootstrap method in complex samples is to ensure that the resampling uses the same process as the original sampling. The replicates were analyzed with sampling weights using the same scoring methods as in the original analysis.

Unlike the leading indicator approach, which tends to have larger samples with specific conditions, the QA Tools method includes a broader spectrum of patients. Estimating the correlations, among eligibility events, among passing the indicators and between eligibility and passing, as one would do for a traditional standard error calculation, would be difficult. The bootstrap approach uses the variability from replicate to replicate to directly estimate the standard errors of the scale scores.

CONCLUSION

This technical appendix describes methods that have evolved over the past four decades at RAND and other institutions to evaluate quality of care. At each step of the process, we followed best practices to increase the reliability and validity of the results. This technical appendix provides greater detail about many of those practices so that the interested reader may have a deeper understanding of the procedures we followed.

REFERENCES^{xxxvii}

- ⁱⁱ McGlynn EA, Asch SM, Adams J, Joan Keeseey, et al. The quality of health care delivered to adults in the United States. *New Engl J Med* 2003; 348:2635-45.
- ⁱⁱ Nutting PA, Schoor GI, Burkhalter BR. Assessing the performance of medical care systems: a method and its application. *Med Care* 1981;19:281-296.
- ⁱⁱⁱ Lawthers AG, Palmer RH, Banks N et al. Developing and evaluation performance measures for ambulatory care quality: a preliminary report of the DEMPAQ project. *Jt Comm Qual Improv* 1993;19:552-565.
- ^{iv} Leape LL, LH Hilborne, JS Schwartz, et al. "The Appropriateness of Coronary Artery Bypass Graft Surgery in Academic Medical Centers," *Annals of Internal Medicine*, in press, 1996.
- ^v Kahn KL, EB Keeler, MJ Sherwood, et al. *The Effects of the DRG-Based Prospective Payment System on Quality of Care for Hospitalized Medicare Patients: Final Report*. Publication No. R-3931-HCFA, Santa Monica, CA: RAND, 1992.
- ^{vi} Gardner P, Rosenberg HM. Leading causes of death by age, sex, race, and Hispanic origin: United States, 1992. Prepared in the division of Vital Statistics, Mortality Statistics Branch, Center for Disease Control.
- ^{vii} National Ambulatory Medical Care Survey, 1993 (authors own analysis).
- ^{viii} Taylor WR, Newacheck PW. Impact of childhood asthma on health. *Pediatrics* 1992;90(5):657-662.
- ^{ix} Weitzman M, Gortmaker SL, Sobol AM, Perrin JM. Recent trends in the prevalence and severity of childhood asthma. *JAMA* 1992;268(19):2673-2677.
- ^x Brook RH, Chassin MR, Fink A, Solomon DH, Kosecoff J, Park RE. A method for the detailed assessment of the appropriateness of medical technologies. *Int J Technol Assess Health Care*. 1986;2(1):53-63.
- ^{xi} Park RE, Fink A, Brook RH et al. Physician ratings of appropriate indications for six medical and surgical procedures. *Am J Public Health*. 1986;76(7):766-72.
- ^{xii} Kahn KL, Park RE, Vennes J, Brook RH. Assigning appropriateness ratings for diagnostic upper gastrointestinal endoscopy using two different approaches. *Med Care*. 1992;30(11):1016-28.
- ^{xiii} Kahn KL, Kosecoff J, Chassin MR et al. Measuring the clinical appropriateness of the use of a procedure. Can we do it? *Med Care*. 1988;26(4):415-22.
- ^{xiv} Leape LL, Hilborne LH, Schwartz JS, et al. The appropriateness of coronary artery bypass graft surgery in academic medical centers. Working Group of the Appropriateness Project of the Academic Medical Center Consortium. *Ann Intern Med*. 1996;125(1):8-18.
- ^{xv} Kleinman LC, Kosecoff J, Dubois RW, Brook RH. The medical appropriateness of tympanostomy tubes proposed for children younger than 16 years in the United States. *JAMA*. 1994;271(16):1250-5.

-
- ^{xvi} Malin JL, Asch SM, Kerr EA, McGlynn EA. Evaluating the quality of cancer care: development of cancer quality indicators for a global quality assessment tool. *Cancer* 2000; 88:701-707.
- ^{xvii} Shekelle PG, MacLean CH, Morton SC, Wenger NS. Assessing care of vulnerable elders: methods for developing quality indicators. *Ann Intern Med* 2001;135(8 Pt2):647-52.
- ^{xviii} Marshall MN, Shekelle PG, McGlynn EA, Campbell S, Brook RH, Roland MO. Can health care quality indicators be transferred between countries? *Quality and Safety in Health Care* 2003;12:8-12.
- ^{xix} Hemingway H, Crook AM, Feder G, Banerjee S, Dawson JR, Magee P et al. Underuse of coronary revascularization procedures in patients considered appropriate candidates for revascularization. *N Engl J Med* 2001;344(9):645-654.
- ^{xx} Kravitz RL, Park RE, Kahan JP. Measuring the clinical consistency of panelists' appropriateness ratings: the case of coronary artery bypass surgery. *Health Policy* 1997; 42(2):135-143.
- ^{xxi} Shekelle PG, Chassin MR, Park RE. Assessing the predictive ability of the RAND/UCLA appropriateness method criteria for performing carotid endarterectomy. *Int J Technol Assess Health Care* 1998; 14(4):707-727.
- ^{xxii} Ashton CM, Keykendall DH, Johnson ML et al. A method of developing and weighting explicit process of care criteria for quality assessment. *Med Care* 1994;32(8):755-770.
- ^{xxiii} Leape LL, Park RE, Kahan JP, Brook RH. Group judgments of appropriateness: the effect of panel composition. *Qual Assur Health Care* 1992;4(2):151-159.
- ^{xxiv} Fraser GM, Pilpel D, Kosecoff J, Brook RH. Effect of panel composition on appropriateness ratings. *Int J Qual Health Care* 1994;6(3):251-255.
- ^{xxv} Kahan JP, Park RE, Leape LL, et al. Variations by specialty in physician ratings of the appropriateness and necessity of indications for procedures. *Med Care* 1996;34(6):512-523.
- ^{xxvi} Brook RH. The RAND/UCLA appropriateness method. In: McCormack KA, Moore SR, Siegel RA, editors. *Clinical Practice Guidelines Development: Methodology Perspectives*. Rockville, MD: Agency for Health Care Policy and Research, 1994.
- ^{xxvii} McGlynn EA, Kosecoff J, Brook RH. Format and conduct of consensus development conferences: a multi-nation comparison, in Goodman C and Baratz S (eds), *Improving Consensus Development for Health Technology Assessment*. Washington, D.C.: National Academy Press, 1990.
- ^{xxviii} McGlynn EA, Damberg CL, Kerr EA, Schuster MA (eds). *Quality of Care for Children and Adolescents: A Review of Selected Clinical Conditions and Quality Indicators*. Available at http://www.rand.org/pubs/monograph_reports/MR1283/index.html
- ^{xxix} Weingart SN, Mukamal K, Davis RB et al. Physician-reviewers' perceptions and judgments about quality of care. *Int J Qual Health Care*. 2001 Oct;13(5):357-65.
- ^{xxx} Hayward RA, McMahon LF, Bernard AM. Evaluating the care of general medicine inpatients: how good is implicit review? *Ann Intern Med*. 1993 Apr 1;118(7):550-6.

^{xxxi} Rubenstein LV, Kahn KL, Reinisch EJ, et al. Changes in quality of care for five diseases measured by implicit review, 1981 to 1986. *JAMA*. 1990 Oct 17;264(15):1974-9.

^{xxxii} Kane RL, Rubenstein LZ, Brook RH et al. Utilization review in nursing homes: making implicit level-of-care judgments explicit. *Med Care*. 1981 Jan;19(1):3-13.

^{xxxiii} National Heart, Lung and Blood Institute. *Guidelines for the Diagnosis and Management of Asthma: Expert Panel Report 2*. Bethesda, MD: National Institutes of Health Publication No. 97-4051. July 1997.

^{xxxiv} Landon BE, Normand SL, Lessler A, et al. Quality of care for the treatment of acute medical conditions in hospitals. *Ann Intern Med* 2006;166(22):2511-2517.

^{xxxv} Payne BC. *The Quality of Medical Care: Evaluation and Improvement*. Chicago, IL: Hospital Research and Educational Trust; 1976.

^{xxxvi} Efron B, Tibshirani RJ. *An Introduction to the Bootstrap*. New York, NY: Chapman and Hall, 1993.