Index

AAA. See abdominal aortic aneurysm
abdominal aortic aneurysm (AAA), reliability of
testing for, 37–38, 261
with screening tests, 135–136, 273–274
absolute risk increase (ARI), 198
absolute risk reduction (ARR), 198
confidence intervals and, 231
ACI-TIPI. See Acute Coronary Ischemia – Time
Insensitive Predictive Instrument
Acquired Immune Deficiency Syndrome (AIDS),
dichotomous tests for, 66, 263
acute coronary ischemia, 3
diagnosing of, 3
Acute Coronary Ischemia – Time Insensitive
Predictive Instrument (ACI-TIPI), 175
acute myocardial infraction (AMI), randomized
trials for, 203–204, 282
adjustment from anchor, in probability,
245–246
alternative diagnosis, 227
alternatives to randomized trials
for autism factors, 217, 282–283
for colon cancer, 217–218, 283
for flu vaccines, 218, 283–284
for prostate cancer, 217, 282–283
American Cancer Society, 240
American College of Obstetricians, 240
American College of Radiology, 240
AML. See acute myocardial infraction, randomized
trials for
anchoring bias, 246–247
antidepressants
randomized trials with, confidence intervals in,
237–238, 285
appendicitis
double gold standard bias and, 115, 273
in multiple tests, 180–181, 277–278
reliability of testing for, 34, 256–257
area under an ROC curve (AUROC), 70–72
discrimination in prognostic tests and, 143
ARI. See absolute risk increase
ARR. See absolute risk reduction
as-treated analysis, 189–190
asthma, screening tests for, 135, 273
AUROC. See area under an ROC curve
autism, randomized trials for causes of, 217, 282–283
availability, in probability, 244–245
axillary dissection, 235–236
B. See cost of failing to treat
bacteremia, 6, 45, 69, 71, 73–74, 79–80, 82–84, 87,
165, 167–168, 177, 225, 229, 231
continuous/multilevel testing for, 91–92, 268–269
bacterial meningitis
continuous/multilevel testing for, 91–92, 268–269
randomized trials for, 201–202, 280
bias. See also double gold standard bias; spectrum
bias; verification bias
anchoring, 246–247
cognitive, 246
confirmation, 246
definition of, 244
in diagnostic test studies, 99–107
application of, 109
double gold standard, 101–103
incorporation, 100
overfitting, 106–107
spectrum, 102–106
verification, 100–101
laboratory error and, 227
lead time, 126
length time, 126–127
pseudodisease, 128–129
publication, 148
random error v., 10
result distortion from, 132
in randomized trials, 191
in screening tests, 125–129
slippery linkage, 130–131
spectrum, 102–106
definition of, 102–104
disease prevalence and, 104–105
ESR and, 104
in multiple tests, 162
sensitivity in, 102
specificity in, 102
test in-dependence and, 105–106
for UTIs, 105–106
stage migration, 127–128
sticky diagnosis, 130
bias (cont.)
verification, 99–101
volunteer, 125–126
binding, 187
biopsies, kidney, 2
Bland-Altman plot, 10
variable types and, for reliability, 11
blinding, in prognostic tests, 147
blinding, in randomized trials, 187
blue belts, 231
bone cancers, prognostic tests for, 153–154, 276–277
Boneferroni correction, 226
confidence intervals and, 238, 285
breast cancer, 6, 44–45, 31, 118, 131–132, 144, 149, 192–193, 240, 246
brown belt, 231
Browner, Warren, 224
C.
See cost of treating nondisease
calibration, 31
for low back pain, 141–143
in prognostic tests, 140
cancer screening tests, 101–102
in EBM, 240–241
prognostic, 151–152, 275
Cat Scratch Fever, dichotomous tests for, 65, 262
cardiac arrhythmia suppression trial, 118
categorical variables, 11
inter-observer agreement for, 11–12, 13, 15, 17, 19, 21
CHD. See congenital heart disease, screening tests for chest pain (CP), reliability of testing for, 33–34, 256
CHF. See congestive heart failure, diagnostic test studies for
cost of failing to treat disease (B), 53, 57, 59, 61, 63, 177
cost of test (T), 57–60, 61, 63
cost of treating nondisease (C), 59, 61, 63, 177
CP. See chest pain, reliability of testing for
CT. See computed tomography
custom weights, 20–21
cytomegalovirus (CMV), prognostic tests for, 152, 275
D+/D−. See dichotomous disease state
decision problems, in diagnoses, 5–6
decision rules, 106, 163, 174–179, 249–250
Decision rules, multivariable, 136–185
derivation sets, 178–179
diagnosis, in EBM, 1
of acute coronary ischemia, 3
alternative, 227
with CT, 5
D+/D− and, 4–5
decision problems in, 5–6
purpose of, 1–3
testing in, 4–5
accuracy of, 7
reliability of, 7
usefulness of, 7
value of, 7–8
diagnostic test, 4–5. See also bias; double gold
standard bias; spectrum bias; verification bias
accuracy of, 7
for CHF, studies for, 114–115, 272
for colic, 8–9, 255
dipstick urine testing v. urine microscopy and, 255–256
dipstick urine testing v. urine microscopy and, 255–256
d for small numerators, 233
with UTI studies, 237, 285
confirmation bias, 246
confounding by indication, 206–207
suppression in, 207
congenital heart disease (CHD), screening tests for, 136–137, 275
congestive heart failure (CHF), diagnostic test studies for, 114–115, 272
congenital hypothyroidism, 4
constipation, randomized trials for, 204, 282
continuous measurements, reliability in, 23–28, 29
correlation coefficients and, 25–26
error by magnitude with, 26–28
in test-retest, 23
within-subject standard deviation/repeatability, 23–24
continuous test, 68–85, 88–93. See also receiver operating characteristic curve
ROC curves in, 70–78
for WBC count, 68–69
logistic regression for, 173–175
continuous testing
for bacteremia, 91–92, 268–269
for bacterial meningitis, 91–92, 268–269
for COPD, 90–91, 267
for Down syndrome, 92–93, 269–270
for home pregnancy tests, 88–89, 264–265
in hypothetical trial cases, 91, 267–268
for myeloperoxidase, 89, 265–266
continuous variables, 11
cOPD. See chronic obstructive pulmonary disease,
continuous/multilevel testing for
cost of failing to treat disease (B), 53, 57, 59, 61, 63, 177
cost of test (T), 57–60, 61, 63
cost of treating nondisease (C), 59, 61, 63, 177
for rotavirus, for gastroenteritis, 8, 255
for sexual abuse, in prepubertal girls, 35, 258–259
studies for, 94–109
appraisal of, 95
biases for, 99–107
conclusions in, 97–98
design for, 94–96, 98
outcome variables, 97
predictor variables, 97, 98
research questions, 94, 98
results in, 97, 98
step-by-step appraisal of, 98
subjects in, 96–97, 98
systematic reviews of, 107–109
ROC curve in, 108
usefulness of, 7
value of, 7–8
for vWD, studies for, 111, 270
dichotomous disease state (D+ /D−), 4–5
dichotomous tests, 39–60
accuracy in, 41
benefit/cost quantification in, 41
for Cat Scratch Fever, 65, 262
definitions of, 42, 61
for enplanements, 67, 263
for Gunderschnauzer disease, 65, 262
for HIV/AIDS, 66, 263
information combinations within, 43
likelihood ratios for, 45–47, 49, 51
derivation of, 49–51, 62
multiple, 159–163
logistic regression, 169–171
result combination, 161, 162–163
spectrum bias, 162
negative predictive value and, 40–41
for positive CT scans, 65–66, 262–263
positive predictive value in, 40
prevalence in, 40
sampling schemes for, 41–42
sensitivity, 39–40
specificity, 40
for streptococcal infection, 65, 262
testing thresholds for, 52, 53, 55, 57, 59
treatment thresholds for, 52, 53–55, 57, 59
expected costs of, 53–54
no treat-test, 55
testing of, 53–56
test-treat, 55
2x2 table methods in, 41–46, 49
blank, 44
completed, 45
for influenza tests, 40, 41–49
dipstick urine testing, diagnostic testing studies for,
112, 271
discrete variables, 11
discrimination, in prognostic tests, 141–142
AUROC and, 143
for low back pain, 141–143
diseases
acute coronary ischemia, 3
classification systems for, 2–3
definition of, 3–4
heterogeneous, 158
pre-symptomatic, 3
probability estimates for, 248–249
screening tests for, 117–119
double gold standard bias, 101–103
for appendicitis, in diagnostic testing of, 115, 273
in cancer screening tests, 101–102
numerical examples of, 103
sensitivity/specificity distortion in, 102
Down syndrome, continuous/multilevel testing for, 92–93, 269–270
EBM. See evidence-based medicine
ectopic pregnancy, diagnostic testing studies for,
112–113, 271
enplanements, dichotomous tests for, 67, 263
equivancy trials, 191
erythrocyte sedimentation rate (ESR), 104
diagnostic testing studies for,
112–113, 271
for influenza tests, 40, 41–49
for Cat Scratch Fever, 65, 262
definitions of, 42, 61
definitions of, 49–51, 62
multiple, 159–163
logistic regression, 169–171
result combination, 161, 162–163
spectrum bias, 162
negative predictive value and, 40–41
for positive CT scans, 65–66, 262–263
positive predictive value in, 40
prevalence in, 40
sampling schemes for, 41–42
sensitivity, 39–40
specificity, 40
for streptococcal infection, 65, 262
testing thresholds for, 52, 53, 55, 57, 59
treatment thresholds for, 52, 53–55, 57, 59
expected costs of, 53–54
no treat-test, 55
testing of, 53–56
test-treat, 55
2x2 table methods in, 41–46, 49
blank, 44
completed, 45
for influenza tests, 40, 41–49
dipstick urine testing, diagnostic testing studies for,
112, 271
discrete variables, 11
discrimination, in prognostic tests, 141–142
AUROC and, 143
for low back pain, 141–143
diseases
acute coronary ischemia, 3
classification systems for, 2–3
definition of, 3–4
heterogeneous, 158
pre-symptomatic, 3
probability estimates for, 248–249
screening tests for, 117–119
double gold standard bias, 101–103
for appendicitis, in diagnostic testing of, 115, 273
in cancer screening tests, 101–102
numerical examples of, 103
sensitivity/specificity distortion in, 102
Down syndrome, continuous/multilevel testing for, 92–93, 269–270
EBM. See evidence-based medicine
ectopic pregnancy, diagnostic testing studies for,
112–113, 271
enplanements, dichotomous tests for, 67, 263
equivancy trials, 191
erythrocyte sedimentation rate (ESR), 104
likelihood ratios for, 104
ESR. See erythrocyte sedimentation rate
evidence-based medicine (EBM), 1
barriers to idealized process of, 243–250
cognitive errors as, 243–249
cognitive errors as, 243–249
oversimplification of diagnostic issues as,
249–250
probability as, 243–249
cancer screening tests in, 240–241
criticisms of, 239–243
limits to clinician autonomy as, 241–243
 nihilism of, as basis for, 240–242
randomized trials and, 239–240
definition of, 1
diagnosis in, 1
future applications of, 251
as malpractice, 242
in media, 240–241, 242
practicality of, 249–250
probability in, 243–249
post-test estimate errors in, 246–249
pre-test estimate errors in, 243–244
PSA in, 242
treatment in, 1
disease classification systems and, 2–3
expected agreements, with kappa statistic,
13–14
balanced/unbalanced, 13–16
calculations of, 13
definition of, 13–14
false-negatives, P-values and, 222–224, 225
false-positives, P-values and, 222–224, 225
focal occult blood screening, 132
flu vaccines, treatment effects on, 218, 283–284
focal segmental glomerulosclerosis, 2
Framingham Risk Score (FRS), prognostic tests and,
152–153, 276
FRS. See Framingham Risk Score, prognostic tests and
gastroenteritis, diagnostic testing for, 8, 255
genetic tests, 148–150
evaluation/interpretation of, 149–150
Glasgow Coma Scale, 11
reliability of testing for, 35–37, 259–261
green belts, 230
Grunderschnauzer disease, dichotomous tests for,
65, 262
GUSTO study, 204
Guyatt, Gordon, 1
hazard ratios, 140, 144
Health Plan Employer Data and Information Set
(HEDIS), 121
Health Professionals study, 212
Healy, Bernadine, 240, 242
HEDIS. See Health Plan Employer Data and Information Set
heterogeneous disease, 158
heterogeneous nondisease, 158
HIV. See human immunodeficiency virus
home pregnancy tests, continuous/multilevel testing for, 88–89, 264–265
human immunodeficiency virus (HIV)
dichotomous tests for, 66, 263
incorporation bias, 100
independence, of tests, 156–158
definition of, 156–157
non-independence v., 157–158
heterogeneous disease and, 158
heterogeneous nondisease and, 158
likelihood ratio slide rule and, 158–159
similarity of measurements in, 157–158
probability and, 157
index testing, 99
influenza testing
treatment thresholds for, 60
2 × 2 table methods for, 40, 41–49
instrumental variables, 207–210
intentionally ordered tests, 225–226
intention-to-treat analysis, 189–190, 209
International Reflux Study in children, 232
inter-rater reliability, 10.
See also kappa statistic
kappa statistic for,
classifications of, 22
expected agreements with, 13–14
formula for, 13, 14–15
good, 22
for three or more categories, 16–22
measures of, 10
intussusception, diagnostic studies for, 113–114, 272
kappa statistic, 11, 12–22
classifications of, 22
expected agreements with, 13–14
balanced/unbalanced, 15–16
calculations of, 13
definition of, 13–14
formula for, 13, 14–15
good, 22
for three or more categories, 16–22
with custom weights, 20–21
with linear weights, 17–18, 20
with quadratic weights, 21–22
unweighted, 17
linear weights, 17–18, 20
for four categories, 18
Kawasaki disease
multiple tests for, 181–182, 278–279
kidney biopsies, 2
labouratory error, bias and, 227
lead time bias, 126
in lung cancer screening tests, 135, 273
in nephritis screening tests, 134, 273
length time bias, 126–127
in lung cancer screening tests, 135, 273
in posterior fossa medulloblastoma screening tests, 136, 275
letrozole, 192–193, 195
likelihood ratios, 45–47, 49, 51
in CT tests, 66
definitions of, 49
derivation of, 49–51, 62
for ESR, 104
for multilevel tests, 79–81
interval, 79–81
odds ratios v., 170–171
recursive partitioning and, 164–165
slide rule for, 51, 86
for non-independent tests, 158–159
in testing threshold visualization, 56
for WBC count, 79
lipid-lowering agents, treatment effects on, 218, 284
logarithms, 86–87
logistic regression
clinical decision rules with,
cost in, 176–177
test selection in, 178
modeling for, 171–172
with ACI-TIPI, 175
Ottawa Ankle Rule in, 177
with PORT Pneumonia Score, 175, 176
in multiple tests, 168–177
dichotomous, 169–171
odds ratios and, 169–171, 172
for single continuous test, 173–175
low back pain, 141–143
low back pain, prognostic tests for, 152, 275–276
lung cancer, screening tests for, 135, 273
malpractice, EBM as, 242
mammography, 6, 44, 60, 120–122, 130–132, 240, 242, 247
Mayo Lung Study, 129
Menger, Fred, 106
metastatic undifferentiated carcinoma, testing for, 9, 255–256
migraine headaches, randomized trials for, 203, 282
minimal change disease, 2
multilevel testing, 68–85, 88–93
for bacteremia, 91–92, 268–269
for bacterial meningitis, 91–92, 268–269
for COPD, 90–91, 267
for Down syndrome, 92–93, 269–270
for home pregnancy tests, 88–89, 264–265
in hypothetical trial cases, 91, 267–268
likelihood ratios for, 79–81
for myeloperoxidase, 89, 265–266
optimal cutoffs for, 81–82, 83
graphical approach to, 82–84
ROC curves and, 85
probability for, 81
multiple hypotheses
Bonferroni correction and, 226
multiple tests and, 226–227
multiple tests, 156–179.
See also independence, of tests; logistic regression
for appendicitis, 180–181, 277–278
derivation sets and, 178–179
dichotomous, 159–163
result combination of, 161, 162–163
spectrum bias in, 162
independence in, 156–158
definition of, 156–157
non-independence v., 157–158
probability and, 157

© Cambridge University Press
Index

for Kawasaki disease, 181–182, 278–279
logistic regression in, 168–177
clinical decision rules with, 175–177
dichotomous, 169–171
modeling for, 171–172
odds ratios and, 169–171, 172
for single continuous test, 173–175
for MI, 184, 279
multiple hypotheses and, 226–227
recursive partitioning in, 163–168
likelihood ratios and, 164–165
Ottawa Ankle Rule and, 163, 165
PROS Febrile Infant Study and, 165
software for, 166–167
treat fewer rule and, 183–184, 279
treat more rule and, 183–184, 279
validation sets and, 178–179
for WBC counts, 183–184, 278–279
randomized trials for, 203–204, 282
multiple tests for, 184, 279
nasal bone, 6, 159–166, 169, 171–172, 178
negative predictive value, 40–41
nephritis, screening tests for, 134, 273
lead time bias in, 134, 273
nephrotic syndrome, 2
New England Journal of Medicine, 90
nominal variables, 11, 68
non-independence, of tests, 157–158
heterogeneous disease and, 158
heterogeneous nondisease and, 158
likelihood ratio slide rule and, 158–159
similarity of measurements in, 157–158
no treat-test threshold, 55
nuchal translucency, 92–93, 160
null hypothesis, 220–221
number needed to harm, 198–199
number needed to treat, 117, 192–195, 197, 231
Nurses’ Health study, 212
odds ratios, 140
likelihood ratios v., 170–171
in logistic regression, for multiple tests, 169–171, 172
in randomized trials, 194
OME. Seoritis media with effusion, randomized trials for
optimal cutoffs, in multilevel testing, 81–82, 83
graphical approach to, 82–84
ROC curves and, 85
dependent variables, 11, 68
continuous, 68
discrete, 68
Glasgow Coma Scale as, 11
otitis media with effusion (OME), randomized trials for, 201, 280
Ottawa Ankle Rule, 163, 165, 177
outcome predictors, 217
outcome variables, 97, 98
Overfitting, 147
overfitting bias, 106–107
partitioning, recursive, 163–168
PCR tests. Sopmerase chain reaction tests
Pediatric Research in Office Settings (PROS) Febrile Infant Study, 165
per-protocol analysis, 189–190
PID. See positive in disease
polymerase chain reaction (PCR) tests, 39
PORT Pneumonia Score, 175
calculation of, 176
positive in disease (PID), 39
positive predictive value, 40
posterior fossa medulloblastoma, screening tests for, 136, 274–275
length time bias in, 275
power, confidence intervals v., 232
predictive value, 40–41
predictor measurement, 212
predictor variables, 97, 98
pre-symptomatic diseases, 3
prevalence, in dichotomous tests, 40
probability, 47–48. See also treatment threshold
probability adjustment from anchor in, 245–246
availability in, 244–245
diseases and, estimates for, 248–249
in EBM, 243–249
post-test estimate errors in, 246–249
pre-test estimate errors in, 243–244
likelihood ratios and, 45–47, 49, 51
in CT tests, 66
definition of, 49
derivation of, 49–51
slide rule for, 51
for multilevel testing, 81
negative predictive value and, 40–41
positive predictive value and, 40
predictive value, 39–43, 45, 90, 96, 111–114, 222–223
propensity scores, 213
P-values and, 222–224, 225
representativeness in, 244
specificity and, 40
test independence and, 157
treatment thresholds and, 52, 53–55, 57
expected cost of, 53–54
formulas for, 62–64
no treat-test, 55
testing of, 55–56
treatment thresholds, 222–223
propensity scores, 213
P-values and, 222–224, 225
representativeness in, 244
specificity and, 40
test independence and, 157
treatment thresholds and, 52, 53–55, 57
expected cost of, 53–54
formulas for, 62–64
no treat-test, 55
testing of, 55–56
treatment thresholds, 222–223
2 x 2 table methods and, 41–46, 49
blank, 44
completed, 45
prognosis, 138
prognostic tests, 138–150
accuracy of, 140
calibration in, 140
prognostic tests (cont.)
  discrimination in, 141–142
  quantification of, 140–143
  blinding in, 147
  for bone cancers, 153–154, 276–277
  for cancer, 151–152, 275
  for CMV, 152, 275
  diagnostic v., 138–139
  effects of treatment in, 146
  follow-up losses in, 146–147
  FRS and, 152–153, 276
  genetic, 148–150
  evaluation/interpretation of, 149–150
  hazard ratios in, 140, 144
  for low back pain, 152, 275–276
  multiple/composite outcomes in, 147
  new information quantification in, 148
  outcome variables of, 138
  overfitting in, 147
  publication bias in, 148
  rate ratios in, 144
  risk factors v., 139–140
  risk ratios in, 140, 144
  sample selection in, 146
  sample size in, 147–148
  for TP53 gene mutations, 154, 277
  value assessment for, 144–145
  propensity scores, 212–216
  analysis of, 215–216
  probability in, 213

PROS Febrile Infant Study. See Pediatric Research in Office Settings Febrile Infant Study

prostate cancer, treatment effects on, 217, 282–283
prostate-specific antigen (PSA), 5
in EBM, 242
PSA. See prostate-specific antigen
pseudodisease, 128–129. See also double gold standard bias
in Mayo Lung Study, 129
types of, 129
publication bias, 148
P-values, 222–233, 234, 236, 238
as conditional probability, 222–224, 225
confidence intervals and, 229–232
ARR and, 231
blue belt, 231
brown belt, 231
green belt, 230
white belt, 230
yellow belt, 230
definition of, 221–222
diagnostic test analogies with, 224, 225
false-positive/false-negative confusion with, 222–224
intentionally ordered tests and, 225–226
quadratic weights, 21–22
random error, bias v., 10
randomized trials
additional outcome measurements in, 211–212
for AMI, 203–204, 282
analysis of, 189–191
as-treated, 189–190
intention-to-treat, 189–190, 209
multiple comparisons in, 190–191
per-protocol, 189–190
subgroup, 190
between v. within-group comparisons in, 191
for bacterial meningitis, 201–202, 280
bias directions in, 191
binding in, 187
for chronic constipation, 204, 282
for chronic hepatitis, 202–203, 281
composite endpoints in, 188
conduct of, 187–189
confidence intervals, with antidepressants, 237–238, 285
confounding by indication in, 206–207
suppression in, 207
critical appraisal of, 187, 189
design of, 187–189
for EBM, 239–240
effect size measurement in, 192
equivalency trials and, 191
follow-up losses in, 189
GUSTO study, 204
Health Professionals study, 212
instrumental variables v., 207–210
for migraine headaches, 203, 282
number needed to harm and, 198–199
Nurses’ Health study, 212
for OME, 201, 280
predictor measurement in, 212
propensity scores v., 212–216
analysis of, 215–216
probability in, 213
purpose of, 186–187
of screening tests, 124–125, 130–131
total v. cause-specific mortality in, 130–131
set-up for, 186
surrogate outcomes in, 188
treatment cost in, 194–198
bad outcome v. traditional, 196
uncertainty about patient disease and, 197–198
treatment effect quantification with, 186–199
alternatives to, 206–216
continuous outcome variables in, 191–192
dichotomous outcome variables in, 192–194
effect size inflation in, with odds ratio, 194
relative v. absolute measures of, 192–193
unrelated variable measurements in, 211
rate ratios, 144
ratios. See hazard ratios; likelihood ratios; odds ratios; rate ratios; risk ratios
receiver operating characteristic (ROC) curve, 70–78
approach to, 74–78
AUROC and, 70–72
in diagnostic testing, systematic reviews of, 108
information in, 78
optimal cutoffs and, 85
in signal detection theory, 70
SROC, 108–109
WBC counts and, 74
Wilcoxon Rank Sum test and, 77
recursive partitioning, 163–168
likelihood ratios and, 164–165
Ottawa Ankle Rule and, 163, 165
PROS Febrile Infant Study and, 165
software for, 166–167
reliability, in testing, 10–32.
See also kappa statistic; variable types, reliability and for AAA, 37–38, 261
for appendicitis, 34, 256–257
bias/random error, 10
Bland-Altman plot for, 10
variable types and, 11
coefficient of variation and, 10
of continuous measurements, 23–28, 29
average standard deviations and, 24
correlation coefficients and, 25–26
error by magnitude with, 26–28
in test-retest, 23
within-subject standard deviation/repeatedly, 23–24
for CP, 33–34, 256
for diagnosing sexual abuse, in prepubertal girls, 35, 258–259
for Glasgow Coma Scale scores, 35–37, 259–261
inter-rater, 10
kappa statistic for, 11, 12–22
measures of, 10
intra-rater, 10
literature studies of, 31–32
method comparison for, 28–31
calibration for, 31
standard deviation and, 10
variable types and, 11
Bland-Altman plot and, 11
categorical, 11
continuous, 11
discrete, 11
kappa statistic and, 12–22
nominal, 11, 68
ordinal, 11, 68
representativeness, in probability, 244
research questions, in diagnostic test studies, 94, 98
results, in diagnostic test studies, 97, 98
risk ratios, 140, 144
ROC curve. See receiver operating characteristic curve
rotavirus testing, for gastroenteritis, 8, 255
sample selections
consecutive, 101
in prognostic tests, 146
selective, 101
sample size, in prognostic tests, 147–148
San Francisco Chronicle, 193
screening tests, 116–132. See also pseudodisease; randomized trials, of screening tests
for AAA, 135–136, 273–274
for asthma, 134, 273
biases in,
lead time, 126
length time, 126–127
pseudodisease, 128–129
stage migration, 127–128
volunteer, 125–126
cardiac arrhythmia suppression trial, 118
for colon cancer, 134, 273
definitions of, 116–117
excessive, 119–122, 123
HEDIS and, 121
false results from, 121–123
fetal occult blood, 132
goals of, 117
harms from, 119, 120–121
for lung cancer, 135, 273
for nephritis, 134, 273
for posterior fossa medulloblastoma, 136, 274–275
public support for, 121
randomized trials of, 124–125, 130–131
total v. cause-specific mortality in, 130–131
studies of, 123–129, 131
appraisals of, 123–125, 127, 129, 131
observational, 125–129
types of, 117–119
for presymptomatic disease, 118
for risk factors, 118–119
for unrecognized symptomatic disease, 117–118
underutilization of, 123
selective sample, 101
sensitivity, 39–40
PID and, 39
sentinel-node biopsy, 235–236
sexual abuse in prepubertal girls, reliability of testing for, 35, 258–259
signal detection theory, 70
slide rule, for likelihood ratios, 51, 86
for nonindependent tests, 158–159
in testing threshold visualization, 56
slippery linkage bias, 130–131
small numerators, 233
specificity, 40
spectrum bias, 102–106
definition of, 102–104
disease prevalence and, 104–105
ESR and, 104
in multiple tests, 162
sensitivity in, 102
specificity in, 102
test nondependence and, 105–106
for UTIs, 105–106
SROC. See summary receiver operating characteristic curve
stage migration bias, 127–128
survival rates and, 128
standard deviation, reliability and, 10
statistical significance testing, 120–221
sticky diagnosis bias, 130
strep throat, 65
dichotomous tests for, 65, 262
study design, for diagnostic test studies, 94–96, 98
subjects, in diagnostic test studies, 96–97, 98
summary receiver operating characteristic (SROC) curve, 108–109
suppression, 207
surrogate outcomes, 188
T. See cost of test
tests. See also continuous testing; diagnostic testing; dichotomous tests; multiple tests; prognostic tests; randomized trials; screening tests
cancer screening, 101–102
continuous, 68–85, 88–93
dichotomous, 68–70
cost, T, 52
diagnostic, 4–5
accuracy of, 7
prognostic v., 138–139
P-values and, analogies with, 224–225
reliability of, 7
studies of, 94–109
<table>
<thead>
<tr>
<th>Index</th>
</tr>
</thead>
</table>
| tests (cont.)  
| systematic reviews of, 107–109  
| usefulness of, 7  
| value of, 7–8  
| dichotomous, 39–60  
| accuracy in, 41  
| benefit/cost quantification in, 52–55  
| continuous, 68–70  
| definitions of, 42, 61  
| information combinations within, 43  
| likelihood ratios for, 45–47, 49, 51  
| negative predictive value and, 40–41  
| PCR, 39  
| positive predictive value in, 40  
| prevalence in, 40  
| ROC curves in, 70–78  
| sensitivity in, 39–40  
| specificity in, 40  
| testing thresholds for, 52, 53, 55, 57, 59  
| treatment thresholds for, 52, 53–55, 57, 59  
| 2×2 table methods in, 41–46, 49  
| genetic, 148–150  
| evaluation/interpretation of, 149–150  
| influenza, 40  
| treatment thresholds for, 60  
| 2×2 table methods for, 41–49  
| intentionally ordered, 225–226  
| multilevel, 68–85, 88–93  
| likelihood ratios for, 79–81  
| optimal cutoffs for, 81–83, 88  
| probability for, 81  
| multiple, 156–179  
| derivation sets and, 178–179  
| dichotomous, 159–163  
| imaging finding outcomes in, 177–180  
| independence in, 156–158  
| logistic regression in, 168–177  
| recursive partitioning in, 163–168  
| validation sets and, 178–179  
| prognostic, 138–150  
| accuracy of, 139  
| blinding in, 147  
| diagnostic v., 138–139  
| effects of treatment in, 146  
| follow-up losses in, 146–147  
| genetic, 148–150  
| hazard ratios in, 140, 144  
| multiple/composite outcomes in, 147  
| new information quantification in, 148  
| outcome variables of, 138  
| overfitting in, 147  
| publication bias in, 148  
| rate ratios in, 144  
| risk factors v., 139–140  
| risk ratios in, 140, 144  
| sample selection in, 146  
| sample size in, 147–148  
| value assessment for, 144–145  
| reliability in, 10–32  
| screening, 116–132  
| biases in, 125–129  
| cardiac arrhythmia suppression trial, 118  
| definitions of, 116–117  
| excessive, 119–122, 123  
| false results from, 121–123  
| fecal occult blood, 132  
| goals of, 117  
| harms from, 119, 120–121  
| public support for, 121  
| randomized trials of, 124–125, 130–131  
| studies of, 123–129, 131  
| types of, 117–119  
| underutilization of, 123  
| statistical significance, 220–221  
| test-retest reliability, 23  
| test-treat threshold, 55  
| Time, 242  
| TP53 gene mutations, prognostic tests for, 154, 277  
| treat fewer rule, 183–184, 279  
| treat more rule, 183–184, 279  
| treatment, in EBM, 1  
| disease classification systems and, 2–3  
| treatment threshold probability, 52, 53–55, 57, 59  
| expected cost of, 53–54  
| no treat-test, 55  
| testing of, 53–56  
| formulas for, 62–64  
| for imperfect test, 59–60  
| for influenza tests, 40, 60  
| likelihood ratio slide rule in, 56  
| for perfect test, 58–59  
| visualization of, 56  
| test-treat, 55  
| 2×2 table methods, 41–46, 49  
| blank, 44  
| completed, 45  
| for influenza tests, 40, 41–49  
| trial cases, continuous/multilevel testing in, 91, 267–268  
| trisomy 21. See Down syndrome, continuous/multilevel testing for  
| urinary tract infections (UTIs), spectrum bias and, 105–106  
| confidence intervals and, in studies of, 237, 285  
| VCUG for, 9, 256  
| urine microscopy, diagnostic testing studies for, 112, 271  
| U.S. News and World Report, 240–241  
| U.S. Preventative Health Services Task Force, 123  
| UTIs. See urinary tract infections  
| validation sets, 178–179  
| variable types, reliability and, 11  
| categorical, 11  
| inter-observer agreement for, 11–12, 13, 15, 17, 19, 21  
| continuous, 11  
| discrete, 11  
| instrumental, in randomized trials, 207–210  
| kappa statistic and, 12–22  
| nominal, 11, 68  
| ordinal, 11  
| outcome, 97, 98  
| of prognostic tests, 138  
| predictor, 97, 98  
| VCUG. See voiding cystourethrogram, for UTIs  
| verification bias, 100–101  
| voiding cystourethrogram (VCUG), for UTIs, 9, 256  
| volunteer bias, 125–126
von Willebrand disease (vWD), diagnostic testing studies for, 111, 270
vWD. See von Willebrand disease, diagnostic testing studies for
Walking man, approach to ROC curves, 74–78
WBC. See white blood cell (WBC) counts, multiple testing for
WBC count. See white blood cell count
white belt, 230
white blood cell (WBC) count, 68–69
likelihood ratios for, 79
posterior odds of, 80
post-test odds of, 83
ROC curve and, 74
sensitivity/specificity of, 74
Wilcoxon Rank Sum test, 77
yellow belt, 230