Index

absolute benefit increase (ABI), 506
absolute risk reduction (ARR), 506, 508
acquiescence bias see response acquiescence
activities of daily living (ADL) scales, 7, 31–32, 577
cross-sectional analysis, 325, 326
treatment, 190
question design, 71
scoring, 253
adaptation
to illness, 516, 524, 525
psychosocial, late problems, 15–16
administration, questionnaire, 268–270
cross-sectional analysis, 325, 326
help and proxies, 276–277
named persons responsible, 271–272
post-treatment follow-up, 275
reporting, 443, 444
during therapy, 273–275
time frame, 263–264, 274, 275–276
written guidelines, 272
age
adjusting for, 326, 483–488
co-morbidity and, 333, 334
differential item functioning (DIF) and, 210, 211, 212
normative data by, 480–482
Akaike Information Criterion (AIC), 181
allocation concealment, 530
α see Cronbach’s α: type I error
alpha factoring, 165
Alzheimer’s disease, 114, 115
Alzheimer’s Disease Assessment Scale – Cognitive
subscale (ADAS–Cog), 323–324
analysis of covariance (ANCOVA), 331, 335
analysis of variance (ANOVA), 331–337
Cronbach’s α, 141, 142
intraclass correlation coefficient, 111–112
models, 336–337
multivariate (MANOVA), 388–389, 435
repeated measures, 373–378
standard error of measurement (SEM), 142
anchor items, 229
anchor methods, 482, 492–493
ANOVA see analysis of variance
antihypertensive therapy, 10, 19, 497–498
anxiety, 29, 36, 40
Apgar score, 52, 146
approximate goodness-of-fit indices (AGFIs), 180, 181
area under the curve (AUC), 345–347
cross-sectional analysis, 322
modelling vs., 389–390
reporting, 363–364
sample size estimation, 286
Aristotle, 6, 519
assessment, QoL: see measurement, QoL
association of variables
graphical methods, 340–341, 437, 439
see also correlations
asthma, 28, 85, 276, 506–508
asymptotically distribution-free (ADF) factor
analysis, 170
attrition, 448–449
clinical trials, 262, 265
graphical presentation, 349, 352, 358
missing data, 400–401
reporting, 360–362
sample size estimation, 306
see also compliance, missing data
attrition bias, 531–532
auto-correlation, 368–372
treatment, 371, 372
exchangeable, 372, 373, 375, 376–377, 380–381
independent, 372, 375, 380, 381
matrix, 368–370, 371
multiplicative (time series), 372, 375, 380–381
patterns, 371–372
structure selection, 387
unstructured, 372, 375, 380–381
user-fixed, 372
auto-regression, 377–378
available-case analysis, 395, 424–425
bar charts, 338–339, 436, 437, 438
Barthel Index of Disability (BI), 31–32, 577
baseline
assessment, 273, 385
changes from, 330–331
variables, reporting, 440
INDEX

item bank construction, 224–232
item response theory, 52, 224, 226–228
multidimensional, 238
patient-reported outcomes, 237–238
pros and cons, 233
short-form tests, 239
stopping rules and simulations, 235
testing procedure, 232–234
testing software, 236
computer-assisted tests (CATs), 238–239
concurrent validity, 94
conditional logistic regression, 195
confidence intervals (CI)
common errors, 440
correlation coefficients, 103, 129–130
impact of imputation, 412, 425
intraclass correlation coefficient, 112, 113, 116, 131
longitudinal data, 358, 362–363
meta-analyses, 533, 534, 542
two-group comparisons, 312, 318, 321
confirmatory data analysis, 337–338
confirmatory factor analysis (CFA), 149, 159, 176–178
factor scores, 161
sample size, 162
vs. other approaches, 150
conjoint analysis, 451–453
CONSORT statement, 283, 429, 445, 530
construct validity, 57, 90, 96–103
convergent validity, 98–99, 134
discriminant validity, 99, 134
known-groups validation, 96–98
multi-item scales, 126, 133–138
multitrait-scaling analyses, 135–138
multitrait–multimethod analysis, 100–103, 134
constructs, 36–37
contact persons, 271–272
content validity, 90–93
documentation, 57, 87
instrument development, 62–63, 74
continuous data
analysis methods, 311, 432–433
factor analysis, 168, 171
meta-analysis, 534–536, 538–539
reliability, 110–112
sample size estimation, 289–292
convergent validity, 90, 98–99
causal or formative indicators, 145
multi-item scales, 125–126, 134, 135
multitrait–multimethod analysis, 100–101
coping, 8, 28, 524
corner state utilities, 454
correlation matrix, 150, 151–152, 154
correlations, 125, 127–133
correlation intervals, 129–130
correction for overlap, 133
factor analysis, 150–152
formative or causal items, 144–147
intraclass see intraclass correlation coefficient
limitations of analyses based on, 172–173
Pearson see Pearson correlation coefficient
polychoric, 132–133, 169
range of variables, 128
rank, 131–132
significance tests, 128, 130–131
cost-effectiveness ratio, incremental (ICER), 460
cost-utility ratios, 460–462
costs, computer-adaptive tests, 237
covariance, 152
covariates
cross-sectional data, 324–330
longitudinal data, 385–387
criterion validity, 90, 94–95
Cronbach’s α, 40, 41–42, 139–142
acceptable values, 139
causal or formative indicators, 145, 147
item selection, 84, 139–140, 144
overuse and alternatives, 142
cross-sectional analysis, 309–343, 477
adjusting for covariates, 324–330
analysis of variance, 331–337
changes from baseline, 330–331
data types, 309–311
endpoints, 342–343
graphical summaries, 337–342
non-Normal distributions, 311, 318–321
Normal distributions, 311, 316–318
two-group comparisons, 312–324
cubic spline functions, 383
cultural differences, 6, 310
differential item functioning, 211, 217–218, 219
instrument development, 82, 85–86
instrument suitability, 249
cumulative distribution functions, 323–324
Dartmouth COOP Charts, 13–14
data analysis, 430–436
choice of method, 432–433
errors, 440–442
modelling, 435–436
multiplicity of outcomes, 433–435
repeated measurements, 435
reporting, 444–445
simple comparisons, 433
data saturation, 66, 67–68
death, attrition due to see attrition
debriefing questionnaires, 75, 81
decision-making, medical, 16–17
degrees of freedom, effect of imputation, 425
dementia, 245
dentures, 519, 525
depression, 29, 36, 40
computer-adaptive test, 235
meta-analysis, 533, 539, 543–545
multi-item scales, 72
descriptive data analysis, 337, 338, 436
detection bias, 531
development, questionnaire, 57–87
defining target population, 58–59
evaluating adequacy, 248
factor analysis, 160
test response theory, 209
phase 1: generating QoL issues, 61–68
phase 2: developing items, 68–74
phase 3: pre-testing, 74–80
phase 4: field-testing, 80–86
phases, 59–61
qualitative methods see qualitative methods,
instrument development
differential item functioning (DIF), 85, 210–222
age-related, 210, 211, 212
computer-adaptive testing and, 238
item response theory, 215
logistic regression, 215–218, 220
Mantel–Haenszel test, 213–214
pseudo, 217–218
quantifying, 219
sample size, 218–219, 220
tips for exploring, 219–221
uniform/non-uniform, 216
difficulty
item, 191–192, 197
test, 190
difficulty-factors, 182–183
dimensionality
computer-adaptive tests, 226–228, 237, 238
Cronbach’s $\alpha$ and, 140–141
evaluating, 125, 134, 138, 139
see also unidimensionality
dimensionless numbers, 505
dimensions, QoL, 5, 37
disability-adjusted life years (DALY), 473
disability paradox, 517–518
disability scales, 31–32, 577
discounting, QALYs, 460
discrete choice experiments (DCE), 451–453
discriminant validity, 90, 99
causal or formative indicators, 145
multi-item scales, 126, 134, 135
multitrait–multimethod analysis, 100–101
discrimination
item, 193, 197
test, 194
discriminative scales, 54–55, 118
disease-specific instruments, 24–28, 559–571
finding suitable, 244, 245–246
mapping utilities across, 458–459
disutility, 453
divergent validity see discriminant validity
domain-specific instruments, 28–32, 573–577
finding suitable, 245–246
downward comparisons, 515
dropouts, 400, 409
see also attrition; compliance; missing data
dummy variables, 334–335, 387, 414
dysphagia, 109–110, 348, 349, 383

Edmonton Symptom Assessment Scale (ESAS), 69
educational tests
computer-adaptive tests, 219–221, 237
item response theory, 51–52, 190, 197
effect indicators see reflective indicators
effect size (ES), 500–505
meta-analyses, 505, 534–537, 542–543
paired data, 502, 535
sample size estimation, 284–290, 305, 499
scale validation, 118, 122, 123
two independent groups, 500–502
see also standardised response mean
effect-size index (Cohen), 499, 503
effect size statistics, 498–505
differential item functioning, 215–216, 220
standardised, 498, 534, 535
eigenvalues, 155–156
greater than one rule, 156, 163, 164
EM algorithm, 420–421
embarrassing questions, 74, 76–77, 402, 403–404
dependence, 515
end points, 443
cross-sectional analysis, 342–343
meta-analyses, 528
multiple see multiple endpoints
EORTC (European Organisation for Research and Treatment of Cancer), 25–26, 32
guidelines for developing questionnaires, 59, 66–67, 75, 87
EORTC QLQ-C30, 25, 559–560
auto-correlation, 369–370
clinical interpretation, 490–491, 494–495
corrections for overlap, 133
Cronbach’s $\alpha$, 141
design, 36, 68–69, 71, 72, 86
differential item functioning, 212, 214, 216–218, 219
disease- or treatment-specific modules, 25–26
differences of use, 15–16, 17
FACT-G vs., 26, 86
graphical summaries, 338–339, 340, 341, 342
longitudinal data, 349
mapping studies, 458–459
missing data, 398–399, 404, 405–406, 407, 408, 411
normalise data, 480–482, 485–487, 488
QLQ-CR29 module, 146
QLQ-ELD14 module, 26, 561
QLQ-LMC21 module, 135–136, 137, 138, 141, 143–144
sample size estimation, 285–286
scoring, 251, 255
selection, 245, 246–247
structural equation modelling, 177, 181, 182
validation, 41, 93, 101–102
epilepsy, 27–28, 65
EQ-5D see EuroQol
equamax rotation, 166
equivalence trials, 260
equivalent-forms reliability, 104, 116–117
error component, 336
error terms, 49, 336, 376
erythropoietin, 246
essentially tau-equivalent tests, 50
estimation, 126
European Organisation for Research and Treatment of Cancer see EORTC
EuroQol (EQ-5D), 23, 555–556
design, 5, 7, 70
quality-adjusted survival, 455, 458–459, 466
selection, 246
youth version (EQ-5D-Y), 106–107
evaluative scales, 54–55, 118
existential beliefs, 8, 28
expectations model, Calman’s, 8
expected mean scores, 483
expert judgement, 494–495
exploratory data analysis, 337, 338, 436
exploratory factor analysis (EFA), 149, 154–175
difficulty-factors, 182–183
factor rotation, 165–166
factor scores, 161
formative or causal models, 47, 51, 173–175, 186–187
historical perspective, 159
structural equation modelling vs., 176
vs. other approaches, 150

F-distribution, 584
F-statistic (F-ratio), 119, 373
t-test, Fisher's, 331, 333, 334
face validity, 74, 91–92
five-and-twenty, 266–267, 532

factor(s), 37, 154
loadings, 156, 252
number of, 163–164
rotation see rotation, factor
factor analysis, 50–51, 149–188
assumptions, 167–171
asymptotically distribution-free (ADF), 170
categorical data, 168–169, 171
choices and decisions, 161–167
confirmatory see confirmatory factor analysis
correlation patterns, 150–152
covariance, 152
dimensionality, 139
exploratory see exploratory factor analysis
formative indicators, 51
historical perspective, 159
illustrative example, 154–157
method of estimation, 164–165
multitrait scaling and, 138
oblique axes, 166–167
orthogonal rotation, 165–166
path diagrams, 152–154
QoL research, 171–175
sample size, 162–163
uses, 159–161
factor score indeterminacy, 161
factor scores, 161
faking good/bad, 515
false negative, 284
false positive, 284, 433
fatigue, 12, 54
area under the curve, 363–364
cross-sectional analysis, 334
gender
adjusting for, 326, 483–488
index, 380–381
interpretation, 494
normative data by, 480–481
generic instruments, 20–24
finding suitable, 244, 245–246
Glass's Δ, 535, 536
global questions, 5, 35, 37–38, 56
interpreting results, 475–476
multitrait scaling, 40–42
see also single-item scales
global ratings of change, 489–492
goodness-of-ﬁt, 199
Goodness of Fit Index (GFI), 180, 181
item response theory, 191, 193
item response theory models, 202–205
goodness-of-fit test, chi-square, 178–179
g graded response model (GRM), 197, 198
graphical presentations, 436–439
common errors, 440–442
cross-sectional data, 337–342
logistic item response models, 199–202
longitudinal data, 348–358, 367
showing variability, 357–358
groin hernia repair, 537–538, 540, 541, 544
guidelines
expert judgement for, 494–495
questionnaire administration, 272
reporting, 442–445
Guttman scales, 71–72, 210
HADS see Hospital Anxiety and Depression Scale
halo effects, 515
Hamilton Rating Scale for Depression (HRSD), 533, 544–545
happiness, 6
hazard ratio, 297
Headache Impact Test (HIT), 216, 226, 228, 231–232
health economics, 256, 456–457, 458–462
health outcomes assessment, 3
health-related quality of life (HRQoL), 3, 5, 9
health status measures, 4, 7, 20
Health Utilities Index (HUI-2 and HUI-3), 454, 455, 458
health, WHO definition, 4, 6
healthy-years equivalent (HYE), 472–473
Hedge’s adjusted g, 535, 536
help, with completing forms, 276–277
heterogeneity
sample see sample selection/heterogeneity study, meta-analysis, 542–544
hierarchical scales, 71–72, 406
historical development, QoL concepts, 6–9
HIV Overview of Problems – Evaluation System (HOPES), 99–100
Holm step-down procedure, 434
home parenteral nutrition, 146
homogeneity
sample see sample selection/heterogeneity scale, 134, 139
study, meta-analysis, 542–544
test, 314
horizontal mean imputation, 413–414, 415
Hospital Anxiety and Depression Scale (HADS), 11, 29, 573
clinical interpretation, 500–501
correlation patterns, 150–152
criterion validity, 95
cross-sectional analysis, 313–317, 321, 329
factor analysis, 154–161, 164, 167–169, 173
indicator variables, 44
item response theory, 202–203
longitudinal data, 349, 352–353, 355–356, 357
multi-item scales, 125
non-Normal distribution, 290, 302, 319
path diagrams, 153, 154
reliability, 40
response shift, 518
sample size estimation, 287–288, 290, 293–298, 301, 307
wording of questions, 74
hot deck imputation, 419–420
Huntingdon’s disease (HD), 115–116
Hypericum see St John’s wort
hypertension, 10, 19, 476
$I^2$ statistic, 543–544
identity link, 388
impact of illness (or treatment), 8
imputation, 358, 395–396, 410–426
assessing study quality, 532
degrees of freedom, 425
deterministic methods, 415
incorporating variability, 415–421
missing forms, 410–425
missing items, 405–407
multiple, 421–422, 423
sensitivity analysis, 426
simple methods, 410–415
incremental cost-effectiveness ratio (ICER), 460
index, 39, 45
development, 73
score, 253
indicator variables, 42–48, 55–56
scoring, 252
individual patient data (IPD) meta-analysis, 541
information functions, 200
information leaflets, patient, 269, 278–279
informative censoring, 408
informative missing data, 394, 402
instrumental activities of daily living (IADL) scales, 32
instruments, 3, 19–32
adding ad-hoc items, 249–250
administration see administration, questionnaire battery, 39
content and presentation, 246–247
developing new see development, questionnaire
disease-specific see disease-specific instruments
domain-specific see domain-specific instruments
generic see generic instruments
historical development, 7–9
identifying suitable, 244–247
profile, 23, 39, 447–448
scoring, 250–256
selection, 243–244, 247–249, 430
see also measurement scales
intelligence, 38, 49, 159
inter-rater reliability, 104, 114–116
internal consistency/reliability, 47
causal or formative indicators, 145
multi-item scales, 125, 139–142
single-item scales, 104
see also Cronbach’s $\alpha$
interpretation, clinical, 475–509
interval scales, 69
interviews
instrument development, 63–64
patient, 63
specialist, 62–63
intraclass correlation coefficient (ICC), 110–112
inter-rater reliability, 114, 115–116
multi-item scales, 131
sample size, 112, 113
test–retest reliability, 113, 114
variants, 124
inverse variance weighted method, 538–539
item(s), 35–36
ad hoc, adding, 249–250
anchor, 229
coverage, 47, 91
distribution of responses, 82–83
linking, 228–229
list construction, 59, 68–74
missing see missing items
not applicable, 408
reduction, 83–85, 226–228
relevance, 91
sequencing, 237
see also questions
item bank, 224–232
item calibration, 228
item evaluation and reduction, 226–228
item linking, 228–229
stages of development, 227
test equating, 230–231
test information, 231–232
item bias, 190, 211
item calibration, 225, 228
concurrent, 229
separate, 229
item characteristic curves (ICC), 191–193, 194
assessing goodness of fit, 199–202
differential item functioning, 215
item difficulty, 191–192, 197
computer-adaptive tests, 237
item discrimination, 193, 197
item information curves, 201–202
item information functions, 200
item location, 192
item-misfit indexes, 204, 208
item non-response, 400
item pool see item bank
item response theory (IRT), 51–52, 189–210, 221–222
based scoring, 252, 255–256
computer-adaptive tests, 52, 224, 226–228
detecting redundant items, 84
differential item functioning, 215
difficulty factors, 182
Guttman scales vs., 71–72, 210
item characteristic curves, 191–193
logistic models see logistic item response models
main models, 197
model assumptions, 205–208
model fitting: tips, 208
polytomous models, 196
short-form tests, 239
test design and validation, 209
traditional scales vs., 209–210
unidimensionality assumption, 183–184, 206
iteration, 381
Kaplan–Meier survival analysis, 297
Kaplan–Meier survival curves see survival curves
kappa coefficient (κ), 105–107
weighted (κ_Weighted), 108–110, 112
Karnofsky Performance Scale, 7
Katz index, 31
KDQOL (kidney disease quality of life), 121–122, 403
known-groups validity, 96–98, 120
Kruskall–Wallis test, 335–336
labelled categorical scales, 68–69
landmark analysis, 449
language differences, 6, 310
differential item functioning, 211, 216–218, 219
instrument suitability, 249
last value carried forward (LVCF), 410–411, 415
late entry, 400
latent roots, 155–156
latent traits, 37
latent variable model, 38–39
latent variables, 36–37
correlation studies, 134
re-specifying, 48
least squares, unweighted, 165
length, questionnaire, 237
life events, changes in relation to, 496–498
Likert summated scales, 50, 69, 73
scoring, 250–253
linear analogue self-assessment (LASA) methods, 7, 70
linear regression models, 324–326
ANOVA model vs., 337
link function, 388
literature search
instrument development, 61–62
meta-analysis, 528–529
suitable instruments, 244
local independence, 153, 206–207
computer-adaptive tests, 226–228, 237
log odds-ratio (log(OR))
differential item functioning, 215–217, 218, 219, 220
meta-analysis, 535, 537, 541
log relative risk (log(RR)), 535, 537
logarithmic transformation, 290, 302
logistic item response models, 193–196
applying, 197–205
assumptions, 206–207
computer-adaptive testing, 224
fitting, 198–199
goodness-of-fit indices, 202–205
goodness-of-fit methods, 199–201
one-parameter see Rasch one-parameter logistic model
sample size, 205
selection, 197–198
two-parameter, 194, 197
logistic models, longitudinal data, 388
logistic regression
cross-sectional analysis, 328–329
differential item functioning, 215–218, 220
item response theory, 195
logit transformation
cross-sectional analysis, 328
item response theory, 193–194, 195, 199
longitudinal data, 388
longitudinal data, 345–365
area under the curve, 345–347, 363–364, 389–390
auto-correlation, 368–372
between- and within-subject variation, 380
covariates, 385–387
fixed and random effects, 382–384
generalised estimating equations, 380–381
graphical presentations, 348–358, 367
interactions, 378–379
logistic models, 388
MANOVA, 388–389
meta-analysis, 542
missing data, 358, 389
modelling, 367–391
multilevel models, 373, 384–385
paired data, 379–380
repeated measures, 373–387
reporting, 360–364
tabular presentations, 358–360, 367
lung cancer, 11
biased reporting, 514
clinical interpretation, 476, 500–501
compliance, 262, 265, 270, 409
cross-sectional analysis, 313–314
longitudinal data, 348, 351, 352–353, 355–356, 357
\(Q-T\)WiST, 465
sample size estimation, 287–288, 296
manifest variables, 37
Mann–Whitney \(U\)-test
sample size estimation, 294, 295
two-group comparisons, 320
MANOVA (multivariate analysis of variance), 388–389, 435
Mantel–Haenszel test, 213–214, 540
MAP-R program, 138
Markov chain imputation, 415–419
maximum-likelihood (ML) estimation
factor analysis, 163, 165, 168, 171
meta-analysis, 539
McGill Pain Questionnaire (MPQ), 29–30, 574
McNemar test, 293
mean(s), 432
comparing, 316–318, 433
medians vs., 349
sample size estimation, 289–292, 298, 299–300
mean and sigma method, 229
mean imputation
horizontal, 413–414
sample, 411–413, 415
simple, 405–406
measurement bias, 523
measurement, QoL
historical development, 6–9
indications, 17–18
methods, 18–32
reasons for, 9–17
see also administration, questionnaire
measurement scales, 35–56
constructs, 36–37
discriminative, evaluative and predictive, 54–55
indicator variables, 42–48, 55–56
items see item(s)
latent variables, 36–37
multiple item see multi-item scales
precision, 40–41
psychometric vs. clinimetric, 52–53
psychometrics, 48–52
reliability, 40
scope, 42
single item see single-item scales
sufficient and necessary causes, 53–54
validity, 41–42
see also instruments
median, 320, 349, 432
Medical Outcomes Study 36-Item Short Form see
SF-36
medical team, administering QoL forms, 268,
269–270, 271–272
meta-analysis, 527–546
assessing study quality, 529–532
combining studies, 537–542
defining objectives, 528
defining outcomes, 528
forest plots, 542
heterogeneity problem, 542–544
literature searching, 528–529
measures of treatment effect, 505, 534–537
publication bias and funnel plots, 544–545
summarising results, 533
Mini-Mental State Examination (MMSE), 201–202
minimal (clinically) important difference (MID), 488–493
anchoring, 492–493
effect size and, 499
expert judgement, 494–495
minimum detectable change (MDC), 493, 498
minimum-residual factoring, 165
missing at random (MAR), 402–403
imputation methods, 414–415, 426–427
missing completely at random (MCAR), 394, 401–403, 426
missing data, 393–427
area under the curve, 347
attrition vs., 448–449
available-case analysis, 395, 424–425
biases, 393, 396–399, 531–532
complete-case analysis, 394, 424–425, 449
consequences, 265, 396–400
degrees of freedom, 425
generalised estimating equations, 381
ignorable, 402
imputation see imputation
informative, 394, 402
longitudinal data, 358, 389
pattern mixture models, 422–424
reasons for, 81–82, 262, 264–265
reporting, 431–432, 444
sample size effects, 306, 399
selection model, 423
summary measures, 395
types, 400–403
see also attrition; compliance
missing forms, 400, 408–425
instrument development and, 82
intermittent, 400, 408
reporting, 431–432
statistical methods, 410–425
missing items, 400, 403–408
checking forms for, 272–273
hierarchical scales, 406
instrument development and, 76–77, 81–82
intermittent, 414
methods for dealing with, 404–408
rates, 403–404
reporting, 431
missing not at random (MNAR), 402–403, 408–409
imputation methods, 422–424, 426–427
mixed-effects models, 383
mixed outcomes, meta-analysis, 540–541
modelling methods, 435–436
monotonicity
computer-adaptive tests, 226–228
item response models, 206
mood, effects on responses, 514–515, 516
multi-attribute utility (MAU) measures, 453–454
multi-item scales, 36, 38–39
construct validity, 133–138
correlation-based methods, 127–133
Cronbach’s α and internal consistency, 139–142
developing, 72–73
formative or causal items, 144–147
item reduction, 84
scoring methods, 20, 250–256
significance tests, 126–127
single-item scales vs., 40–42
validation, 125–147
multicentre clinical trials, 259, 410
multidimensional construct, QoL as, 5, 37
Multidimensional Fatigue Inventory (MFI-20), 12, 30–31, 575–576
multilevel models, 373, 384–385
multiple endpoints
clinical trials, 261–262
sample size estimation, 303–305, 306–307
multiple-group comparisons, 432
ANOVA, 333–335
sample size estimation, 303
multiple imputation, 421–422, 423
multiple-indicator multiple cause (MIMIC) models, 51
multiple regression analysis, 326–327
multiple testing, 433–435
differential item functioning, 220
sample size estimation, 303–305
multitrait analysis, 84, 150
multitrait-multimethod analysis (MTMM), 100–103, 134
multitrait-scaling analyses, 135–138
multivariate analysis of variance (MANOVA), 388–389, 435
myeloma, multiple
cross-sectional analysis, 338–339, 340, 341, 342
interpreting results, 490–491
longitudinal data, 359, 361, 363–364
reporting results, 436, 437, 438, 439, 441
myocardial infarction, acute, 13
National Institute for Health and Care Excellence (NICE), 454, 455
nausea and vomiting, 18, 36, 44, 46, 514
necessary causes, 53–54
needs model, 8
nominal data, 309–310
nominal response model (NRM), 197
non-compliance see compliance
non-inferiority studies, sample size estimation, 298–301
non-Normal distributions, 119
adjusting for covariates, 328–329
ANOVA, 335–336
cross-sectional analysis, 311, 318–321
factor analysis, 168, 169–170, 171
sample size estimation, 289, 302
transformation, 290, 302, 321
non-normed fit index (NNFI), 180
norm-based scoring, 20, 253–255
Normal distribution
adjusting for covariates, 324–328
area under the curve, 390
checking for, 335
cross-sectional analysis, 311, 316–318
factor analysis, 168, 169–170, 171
sample size estimation, 289, 302
statistical tables, 579–581
structural equation modelling, 176
summary statistics, 432
T- and Z-scores, 255
Normal plot, 335
normal range, 505–506
norms, population (reference values), 479–488
adjusting for age and gender, 483–488
anchor methods, 482
sample size estimation, 298
Nottingham Health Profile (NHP), 7, 21–22, 265, 551
nuisance variables, 99
null hypothesis, 126–127, 312
number needed to treat (NNT), 506–508
numerical data, 310–311
numerical rating scale (NRS), 41, 69, 71
nurses, administering QoL forms, 268
objectives, new QoL instruments, 58
oblimin, 166
oblique axes, factor analysis, 166–167
observers, assessing QoL, 18–19
odds ratio (OR)
clinical interpretation, 479
cross-sectional analysis, 328–329
log see log odds-ratio
Mantel–Haenszel test, 213, 214
meta-analysis, 537, 540–541
Peto’s, 540
sample size estimation, 292–293, 294–295
oesophageal cancer, 12, 383, 458–459
off-schedule patients, 362
one-parameter logistic model see Rasch one-parameter logistic model
open questions, 64
ordered categorical data, 310
cross-sectional analysis, 314–316
data analysis methods, 432–433
item response theory models, 196, 199–200
reliability, 107–109
sample size estimation, 294–296, 307
ordered categorical scales, 68–69
ordinal outcomes, meta-analysis, 537, 540
ordinal scales, 68–69
orthogonal rotation, 165–166
orthomax rotation, 166
outcome reporting bias see selective reporting bias outliers, 341
Overactive Bladder Symptom Score (OABSS), 123
overlap
correction for, 133, 135
latent variables, 134
Oxford Hip Score, 255

p-values, 127, 284
Bonferroni correction, 434
differential item functioning, 220
multiple comparisons, 433–435
reporting, 440
two-group comparisons, 312, 318
two-tailed or two-sided, 579–580
Paediatric Asthma Quality of Life Questionnaire (PAQLQ), 28, 570–571
Paget’s disease, 520–521
pain
assessment, 29–30, 48, 71
bias and response shift, 512, 513, 516, 520, 523
cultural differences, 217–218
factor analysis, 184–186
impact rating, 495–496
item response theory, 190, 207
longitudinal analysis, 345–346, 347
meta-analyses, 537–538, 540, 541, 544
sample size estimation, 300
scale validation, 114, 120, 130, 146
time-to-event analysis, 322–323
paired data
effect size, 502
longitudinal data, 378–379
meta-analysis, 535
sample size estimation, 291–292, 293–294
standardised response mean, 502
palliative care, 10–12, 260
missing data, 409, 422
non-compliance/attrition, 263, 264–265, 532
response shift, 524
parallel items, 47, 49–50
item response theory for testing, 210
parallel tests, theory of, 49–50
partial correlation, 99, 330
partial credit model (PCM), 197, 198, 202–203
partial-gamma test, 213
path diagrams, 152–154

patient(s)
asking, 18–19, 448
attrition see attrition
clinical trials, 268–269
defining target population, 58–59
help with completing forms, 276–277
instrument development, 63–68
preferences, 8, 14–15, 449–453
variability, 505–506
Patient Generated Index (PGI), 8, 23–24, 557
Patient Health Questionnaire 9 (PHQ-9), 29, 313–314
patient information leaflets, 269, 278–279
Patient-Reported Outcome and Quality of Life Instruments Database (PROQOLID), 244

patient-reported outcome measures (PROMs), 3
patient-reported outcomes (PROs), 3–4, 6
Patient-Reported Outcomes Measurement Information System see PROMIS
pattern mixture models, 422–424
Pearson chi-squared (χ²) test, 292
Pearson correlation coefficient (r), 110, 127–131
auto-correlation, 368, 369–370
comparing two, 130–131
confidence intervals, 129–130
factor analysis, 151–152
significance testing, 128
Pediatric Cardiac Quality of Life Inventory (PCQLI), 132

Perceived Adjustment to Chronic Illness Scale (PACIS), 5
percentages, 348–349, 433, 478–479
percentile rank, 253
performance bias, 531
performance status, compliance and, 265
person-misfit indexes, 204, 208
person-reported outcomes, 3
personal well-being, 8
Peto’s odds ratio, 540
pilot study see pre-testing, questionnaire
polychoric correlation, 132–133, 169
polymitous item response models, 196
population norms see norms, population populations
special, 244–245
target see target population
power, 284
loss due to missing data, 399
pre-study calculation, 305, 430
sample size estimation, 285, 289
pre-testing, questionnaire, 60, 74–79
precision, 40–41
predictive instruments, 54–55
predictive validity, 95
preference measures see utilities
preferences, patient, 8, 14–15, 449–453
principal-axes factoring, 165
principal-factor estimates, 163
probability, 127
profile instruments, 23, 39, 447–448
profile plots, 341–342, 440–441
individual patients, 352
reverse, 356–357
summary, 354–356
prognosis, Q-TWISTY and, 4/0
prognostic value, QoL scores, 16–17
promax, 166, 167
PROMIS, 32, 102–103, 130
prophecy formula, Spearman–Browne, 140
proportion of agreement, 105
proportions
binomial, 312–313
sample size estimation, 292–294, 298, 300–301
PROs see patient-reported outcomes
prostate cancer, 411, 416–420
protocols, clinical trial, 259, 261, 270–280, 282
proxy assessment, 3–4, 6
biased, 514, 515
clinical trials, 260–261, 276–277
inter-rater reliability, 114–116
INDEX 623

psychometric scales, clinimetric scales vs, 52–53
psychometrics, 48–52
modern, 49, 51–52
traditional, 48–49, 50–51
PU-QOL (pressure ulcer quality of life), 116–117
publication bias, 544–545
Q-statistic, 543–544
Q-TWiST, 462–472
alternatives to, 472–473
calculation, 465–466
choice of health states, 463
comparing treatments, 466–467
prognosis and variation with time, 470–472
QALY vs., 462
sensitivity analysis, 467–469
survival curves, 463–464
QDIS-CKD (quality-of-life disease impact scale for chronic kidney disease), 121–122
QLQ-C30 see EORTC QLQ-C30
QoL see quality of life
qualitative methods, instrument development, 57–58, 63–68, 86
documentation, 86–87
focus groups, 64–65
interviews, 63–64
sample selection, 66
sample sizes, 66–67
saturation, 66, 67–68
quality-adjusted life years (QALY), 365, 456–462
alternatives to, 472–473
assumptions, 459–460
cost-utility ratios, 460–462
discounting, 460
mapping studies, 458–459
Q-TWiST vs., 462
quality-adjusted survival, 447–474
quality-adjusted time without symptoms and toxicity see Q-TWiST
quality of life (QoL), 3–4
defined, 4–6
dimensions, 5, 37
historical development of concepts, 6–9
impact of state of, 495–496
Quality of Life in Epilepsy Inventory (QOLIE-89), 27–28, 566–569
Quality of Well-Being Scale (QWB), 454, 455
quartimax rotation, 166
questionnaires see instruments
questions
developing, 68–74
embarrassing or problematic, 74, 76–77, 272, 402, 403–404
global see global questions
open and closed, 64
wording, 73–74
see also item(s)
radiotherapy, 12, 15–16, 334, 348, 458–459, 519
random-effects models
longitudinal data, 382–384, 386
meta-analysis, 538, 543, 544
random error term, 40
randomisation, 273, 279–280, 530
randomised controlled trials (RCT), 3
adjusting for covariates, 326, 328, 330
curative treatment, 9–10
impact of response shift, 523–525
missing data, 393–394, 397–399
need for QoL assessment, 17–18
palliative treatment, 10–12
see also clinical trials
randomly parallel tests, 50
rank correlation, 131–132
ranked data see ordered categorical data
Rasch one-parameter logistic model, 194, 195–196, 197, 221
based scoring, 255
sample size, 205
selection, 197, 198
rating scale model (RSM), 197
raw mean difference (MD), 534
raw score, 405
real simulations, 235
recall bias, 512–513
receiver operating characteristic (ROC) curves, 490–491
reference values see norms
reflective (effect) indicators, 42–48
ceiling and floor effects, 82–83
distinction from causal indicators, 46
factor analysis, 174
instrument development, 55–56
item reduction, 84
overlap with causal indicators, 44
reflective model, 43, 47, 48
regression coefficients, 324–325
regression imputation, 406–407, 414–415
regression models
adjusting for covariates, 324–329
ANOVA models and, 336–337
changes from baseline, 330–331
longitudinal data, 376–379
multiple-group comparisons, 334–335
see also logistic regression
rehabilitation programmes, 12–13
reintegration to normal living model, 8
relative change, 118
relative efficiency (RE), 118, 119, 121–122
relative risk (RR), 537
log (log(RR)), 535, 537
relative validity (RV), 118, 119
reliability, 40, 90
acceptable, 112
binary data, 105–107
continuous data, 110–112
equivalent-forms, 104, 116–117
inter-rater, 104, 114–116
internal see internal consistency/reliability
intraclass correlation coefficient (ICC), 110–112
multi-item scales, 125
ordered categorical data, 107–109
Pearson correlation coefficient, 110
repeatability, 104
sensitivity and, 120
single-item scales, 104–117
test–retest see test–retest reliability
repeatability, 55, 90, 104
repeated assessments
  data analysis, 435
  sample size estimation, 303
  validation, 104–123
repeated-measures ANOVA, 373–380
reporting, 429–445
  compliance, 360–362, 431–432
  data analysis, 430–436
  design issues, 430
  errors, 440–442
  guidelines, 442–445
  longitudinal data, 360–364
  residuals, 180, 336, 337
  auto-correlation, 376, 387
fit, -204
respondent validation, 74
response acquiescence, 277–278, 515
response shift, 511, 512, 516–526
  assessing, 521–522
  impact, 523–526
response thresholds
  disordered, 199
  item characteristic curves, 199–200
  item response theory models, 196, 197
responsiveness, 90, 117–119, 122–123
  different types of scales, 54
  measures of, 118–119, 122
responsiveness statistic, 118
reverse profiles, 356–357
risk difference, 537
Root Mean Square Error of Approximation (RMSEA), 180, 181
rotation, factor, 157–158, 159
  oblique, 166, 167
orthogonal, 165–166
Rotterdam Symptom Checklist (RSCL), 26–27,
  564–565
differential item functioning, 210–211
  factor analysis, 171–172, 173, 174, 175
St John’s wort (Hypericum), 533, 539, 545
  sample, 126, 443
  sample-mean imputation, 411–413, 415
  sample selection/heterogeneity
  correlation studies, 128
  Cronbach’s a and, 141
instrument development, 66, 76, 80–81
  multitrait-scaling analysis, 136–137
sample size, 283–307
  clinical trials, 261, 283–307
  differential item functioning, 218–219, 220
  factor analysis, 162–163
  instrument development, 66–67, 76, 85
  item response theory models, 205
  missing data and, 306, 399
  multitrait-scaling analysis, 136
  reliability studies, 112, 113
  responsiveness studies, 123
sample size estimation, 284–307, 499
  choosing power, 285
  choosing target effect size, 285–288
  choosing type I error, 285
  clinical relevance, 477–478
  compensating for attrition, 306
INDEX
formulae, 288–300
  multiple testing, 303–305
  non-inferiority studies, 298–301
  non-Normal distributions, 302
  pre-study stage, 305
  reference population comparisons, 298
  reporting, 430
  selection of method, 301
  specifying target difference, 305
two-group comparisons, 289–298
  sampling bias, 85
  satisfaction with life, 8, 28
  saturation, data, 66, 67–68
  saturation grid, 68
  scale scores, 20, 39
  linking of see test equating
  precision, computer-adaptive tests, 225, 233, 234
  use of factor analysis, 160–161
  scales, measurement see measurement scales
  scaling errors, 135–136
  scaling success, 135, 137
  scatter plots
  association of variables, 340, 437, 439
  factor analysis, 157
  longitudinal data, 349–352
  Schedule for Evaluation of Individual Quality of Life
    (SEIJoL), 8, 23–24, 519
  scores, scale see scale scores
  scoring, 30, 250–256
  health economics, 256
  IRT-based, 252, 255–256
  norm-based, 20, 253–255
  standard, 20, 250–252
  summed scales, 250–253
  scree plot, 163, 164
  search, literature see literature search
  secondary gain, 517–518
SEIJoL see Schedule for Evaluation of Individual Quality of Life
  selection bias, 530
  selection model, 423
  selective reporting bias, 513–514, 518, 523, 531
  self-esteem, 28
  self-reported health (SRH), 4
  sensitivity, 55, 90, 117, 119–122
  assessment methods, 120–122
  measures of, 118–119
  sensitivity analysis
    imputed data, 426
    Q-TWISt, 467–469
  serial correlation see auto-correlation
  sexual function, questions about
    instrument development, 74, 76–77, 78–79
    missing data, 272, 402, 403–404
  “sexual reorientation” therapy, 525
SF-6D, 23, 455, 457, 458
SF-12, 121–122
SF-36, 15, 22–23, 552–554
  factor analysis, 176
  item response theory, 195–196, 221
  missing data, 424–425
  precision, 41
  sample size estimation, 291–292
  scoring, 254

INDEX

validation, 134
wording of questions, 69, 74
short-form tests, 239
Sickness Impact Profile (SIP), 7, 21, 549–550
side effects
factor analysis, 173–175
interference of, 8
significance tests, 126–127, 284
baseline variables, 440
chi-square test, 178–179
clinical interpretation, 476–477
correlations, 128, 130–131
differential item functioning (DIF) analysis, 220
multiple, 433–435
non-significant, 440
two-group comparisons, 312
single-item scales, 36, 87
multi-item scales vs., 40–42
reliability, 104–117
sensitivity and responsiveness, 117–123
validity, 87–103
single rating, single-item scale, 39
see also global questions
16D (children with epilepsy), 28
skew distributions
ANOVA, 335–336
cross-sectional analysis, 317, 318–319, 322
factor analysis, 169, 170
sample size estimation, 290, 302
social comparison, 515
social desirability, 515
Social Readjustment Rating Scale (SRRS), 497–498
Spearman auto-correlation coefficient, 368, 369
Spearman–Browne prophecy formula, 140
Spearman, Charles, 159
Spearman rank correlation, 131–132
specialist interviews, 62–63
split-plot design, 373
spurious correlation, 99
standard deviation (SD), 152, 432
clinical interpretation, 479, 498–499
effect size and, 118, 500, 502
imputation methods reducing, 412, 414, 425
meta-analysis, 534
standardised response mean and, 502–503
standard error (SE), 127, 129, 498
meta-analysis, 534, 535, 539
two-group comparisons, 312, 317–318
standard error of measurement (SEM), 142
standard gamble (SG), 450–451, 457
standard operating procedures (SOPs), 280–281
standard scoring method, 20, 250–252
standardisation, 498–499
direct, 483
indirect, 483
standardised mean difference (SMD), 534–536, 540–541
standardised response mean (SRM)
clinical interpretation, 500, 502–504
meta-analysis, 535, 536
scale validation, 118, 122, 123
Standardised Root Mean Square Residual (SRMR), 180
standardised score, 405

STATA program, 154, 163
statistical analysis see data analysis
statistical significance, 436, 476–477
tests see significance tests
statistical tables, 579–584
stopping rules, computer-adaptive tests, 235
Stroke Impact Scale (SIS), 503–504
structural equation modelling (SEM), 51, 149, 176–188
approximate goodness-of-fit indices, 180
bifactor analysis, 183–186
chi-square goodness-of-fit test, 178–179
comparative fit of models, 181
difficulty-factors, 182–183
formal or causal relationships, 186–187
vs. other approaches, 150
Student’s t-distribution, 582
Student’s t-test see t-test
subgroup differences, 85, 211
subjective measures, 6
sufficient causes, 53–54, 253
sum-scores, 250–252
standardised, 250–251
weighted, 252–253
summary measures, 4, 73
continuous data, 311
longitudinal data, 346–347, 363, 365
missing data and, 395
selection, 432
summary profiles, 354–356
summated ratings, 73, 250–253
survival, quality-adjusted, 447–474
survival analysis, 297
survival curves, 323, 358, 360–362
missing data, 400–401
Q-TWiST, 463, 464, 465
survival studies, clinical interpretation, 476, 477
survival times, 322
Swal-Qol dysphagia questionnaire, 109–110
symptoms
factor analysis, 173–175, 186–187
improving, 12–13
interference of, 8
item response theory, 207
scoring methods, 252–253
syndromes, 173
t-distribution, Student’s, 582
T-scores, 20, 253–255
t-statistic
paired, 118, 119
squared, 119
t-test
comparing two means, 316, 318
sample size estimation, 286, 287, 289–290
tabular presentations, 358–360, 367, 440
target population
defining, 58–59
instrument development, 63–68
instrument suitability, 248
tau-equivalent tests, 50
terminal missing, 400
test equating (or linking), 225, 230–231
test for heterogeneity chi-square, 543–544
test-for-trend, 314–316, 320
test information functions, 200, 202
  computer-adaptive tests, 225, 231–232
testicular cancer (TC), 61–62, 63, 450
test–retest reliability, 55, 104, 113–114
then-test, 521–522
thought test, 46, 47
threshold utility analysis, 467–469
threshold values, 478–479
thresholds, response see response thresholds
time frame, for form completion, 263–264, 274, 275–276
time-to-event analysis, 322–323
time-to-event data
cross-sectional analysis, 322–323
  sample size estimation, 297–298, 307
time trade-off (TTO), 450
Tourangeau’s cognitive interviewing model, 77
transformation, non-Normal data, 290, 302, 321
transition probabilities, 416–418, 419
transition questions, 489–492
translation, 80
differential item functioning, 211, 216–218, 219
  problems, 82
treatment
comparisons, reporting, 362–364
curative, 9–10
  follow-up assessment, 275
  influence of QoL forms on, 277–278
  palliative see palliative care
  QoL assessment during, 273–275
time interactions, 378–379, 385
treatment effect
cross-sectional analysis, 330–331
  meta-analysis, 505, 534–537
  see also effect size
tuberculosis (TB), 53–54
Tucker–Lewis Index (TLI), 180, 181
TWiST (time without symptoms or toxicity), 463, 465
two-group comparisons, 312–324, 432
ANOVA, 331–333
binomial proportions, 312–313
categorical data, 313–316
  changes from baseline, 330–331
cumulative distribution functions, 323–324
effect size, 500–502
  graphics, 437, 438
  non-Normally distributed data, 318–321
  Normally distributed data, 316–318
  sample size estimation, 289–298
time-to-event data, 322–323
two-parameter logistic item response model, 194, 197
type 1 error ($\alpha$), 284, 285, 289
type 2 error ($\beta$), 284
unidimensionality, 37, 139
  computer-adaptive tests, 226–228
  essential/sufficient, 184
evaluation, 125, 134
  methods assuming, 182, 183–184, 206
  see also dimensionality
unit non-response, 400
unweighted least squares, 165
upward comparisons, 515
urodynamic studies (UDS), 300–301
utilities, 449–453
  mapping across instruments, 458–459
  for traditional instruments, 457–461
  utility-based instruments, 8, 23, 454–455
validation, 87
  by application, 97
evaluating adequacy, 248
  factor analysis, 160
  formative or causal items, 144–147
  item response theory, 209
  multi-item scales, 125–147
  quantitative vs. qualitative, 57–58
  repeated assessments, 104–123
  respondent, 74
  scale alteration and, 143–144
  single-item scales, 87–103
  translated questionnaires, 80
validity, 41–42, 87–103
  see also specific types of validity
variability
graphical display, 357–358, 362
  imputation methods incorporating, 415–421
  patient, 505–506
  scaling for, 498–499
  tabular display, 360
variance, 152, 331
  explained, factor analysis, 155–156
  Mantel–Haenszel test, 213
varimax rotation, 157–158, 165–166, 167
  verbal rating scales (VRS), 68–69, 71
  precision, 41
visual analogue scales (VAS), 7, 69–70, 71
  patient preferences, 449–450
  precision, 41
Visual Function Questionnaire-25 (VFQ-25), 492
voice problems, 92
vomiting see nausea and vomiting
weight of evidence, 284
  weighted mean difference (WMD), 538–539
  weighted sum-scores, 252–253
  weights, 8, 252–253
WHOQOL-HIV-BREF, 97–98
willfulness to pay (WTP), 451
windows (time)
  form completion, 263–264, 274, 362
  graphing longitudinal data, 352–354
World Health Organization (WHO), definition of
  health, 4, 6
yea-saying see response acquiescence
Z-scores, 253–255
  z-statistic, 127, 312, 581
  Z-transformation, 129
Zingiberaceae extracts, 536, 544
Zung self-rating Pain and Distress Scale, 30