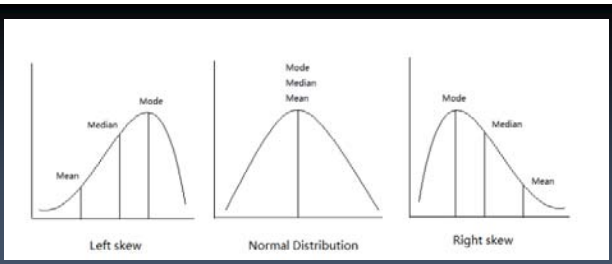


An Introduction to Useful Statistics When Clinicians are Reading the Medical Literature

Alan Barkun, MD,CM, FRCP(C), FACP, FACG, AGAF, MSc (Clinical Epidemiology)
Chairholder, Douglas G. Kinnear Chair in Gastroenterology
Professor of Medicine
McGill University and the McGill University Health Centre
Montreal, QC

Aim / Disclaimer

- I am NOT a statistician!
- I did not really want to present this talk to you but was coerced into doing so (although I did volunteer for it)
- My aim is NOT to make statisticians out of you (especially since I am not one myself)

My aim is to:

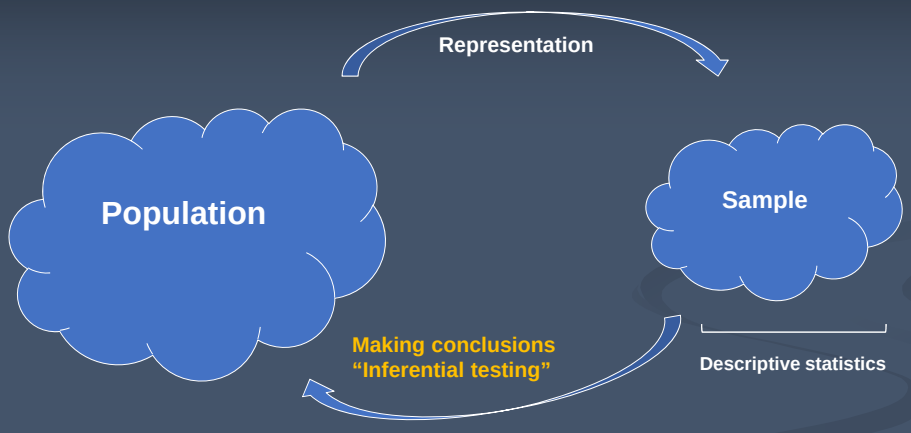
HELP YOU MAKE SENSE OF THE EVER-INCREASING VOLUME OF PUBLISHED LITERATURE AND SEEMINGLY COMPLEX NATURE OF THE STATISTICS THAT ARE USED TO UNDERSTAND RESULTS

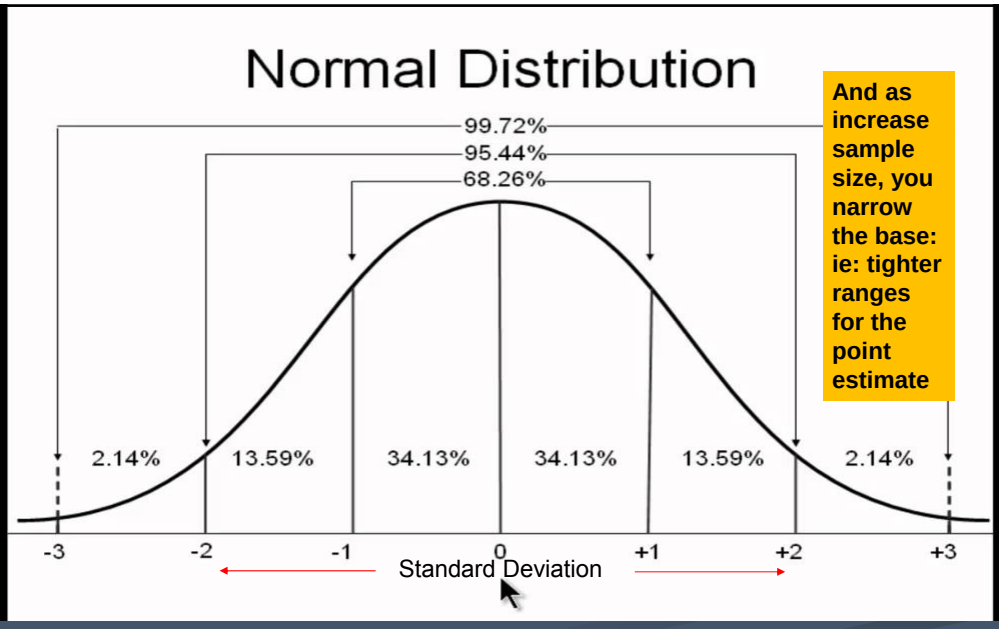
- I will stick to a few selected concepts -

Outline

- The role of statistics: Inferential testing and sample distributions; choosing the correct inferential test
- Hypothesis testing: Significance, statistical power, types I and II errors
- Probabilities vs Odds ratios
- Absolute and relative risks; number needed to treat/harm/screen
- Diagnostic testing, ROC analysis
- Confounding and adjusting for confounding
- Meta-analyses

Inferential testing





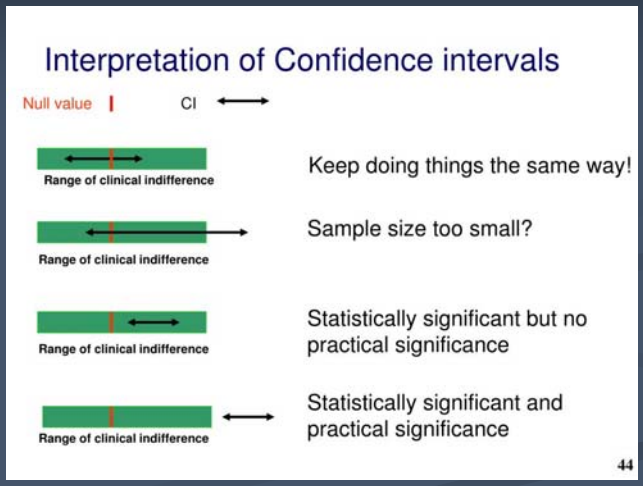
https//www.bing.com/images/search?view=detailV2&ccid=N5YNTOwB&id=E45388EDDB16EFA67AEB6245DFDF3B69D8057891&thid=OIP.N5YNTOwBxrWqWEZd8U7CTAHaEK&mediaurl=https%3a%2f%2fi.ytimg.com%2fvi%2fiiRiOlkLa6A%2fmaxresdefault.jpg&ccdnurl=https%3a%2f%2fth.bing.com%2fth%2fid%2fR37960d4cec01c6b5aa58465df14ec24c%3frik%3dkXgF2Gk7399FYg%26pid%3dImgRaw&exph=720&expw=1280&q=normal+distribution&simid=608012170561724848&ck=9D7D850BAD70876FFC8CB6E61D75E38D&selectedIndex=7&FORM=IRPRST&ajaxhist=0

Confidence Intervals

Comparison of
2 treatments:
Difference in response
rates between Drug X and Drug Y

Nul hypothesis is no
difference between
the 2 Difference = 0

Eg: X 20% (13-32%) and
Y 30% (25-45%),
difference 10% (6-17%)
but clinically significant difference
is also relevant - if for eg 5% or 20%



https://www.bing.com/images/search?view=detailV2&ccid=PMSeZrTb&id=4BA10E6AD661D0820C10242716A3944D8CBC31B5&thid=OIP.PMSeZrTbWjN0a8OoPjr_BwHaFj&mediaurl=https%3a%2f%2fimage1.slideserve.com%2f2735251%2finterpretation-of-confidence-intervals2l.jpg&ccdnurl=https%3a%2f%2fth.bing.com%2fth%2fid%2fR3ca49e66b4db5a33746bc39c3dfaff07%3frik%3dITG8jE2UocYnJA%26pid%3dImgRaw&exph=768&expw=1024&q=confidence+intervals+interpretation&simid=607999401620627305&ck=C3C3D0CDC270D5065AC39510E8C4C58C&selectedIndex=9&FORM=IRPRST&ajaxhist=0

The sample distribution

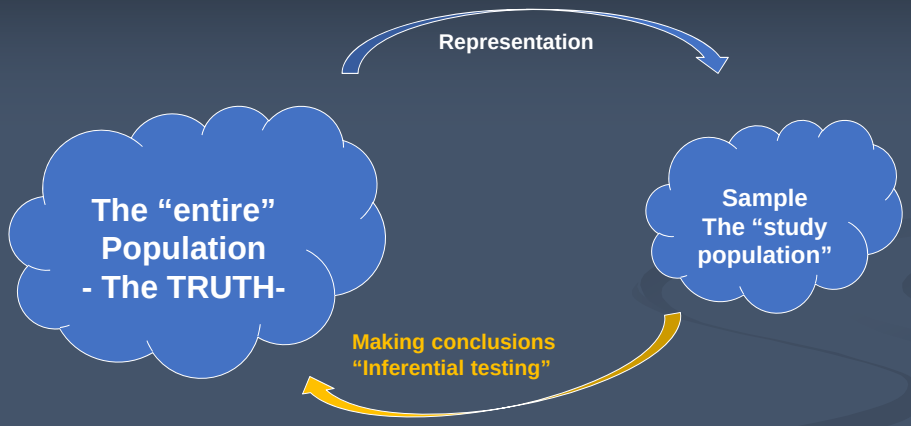
- The sample distribution may be considered as the distribution of the statistic for all possible samples from the same population of a given sample size
- Making assumptions about the "studied population distribution" as a sample of the "whole population", you can make assumptions and adopt certain formulas when performing inferential testing statistics
- This decision also depends on a number of additional factors

Choosing the correct test – Is there a difference

Number of Dependent Variables	Nature of Independent Variables	Nature of Dependent Variable(s)	Test(s)
1	1 IV with 2 or more levels (independent groups)	interval & normal	one-way ANOVA
		ordinal or interval	Kruskal Wallis
		categorical	Chi-square test
	1 IV with 2 levels (dependent/matched groups)	interval & normal	paired t-test
		ordinal or interval	Wilcoxon signed ranks test
		categorical	McNemar
	1 IV with 2 or more levels (dependent/matched groups)	interval & normal	one-way repeated measures ANOVA
		ordinal or interval	Friedman test
		categorical (2 categories)	repeated measures logistic regression

https://stats.idre.ucla.edu/other/mult-pkg/whatstat/

The name of the game: "Inferential testing"



Hypothesis testing

- The statistical practice of hypothesis testing is widespread
- Hypothesis testing:
  - the statement of a null hypothesis (Eg: the study drug is no better than placebo or control drug)
  - the null hypothesis is either true or false
- Making a statistical decision always involves uncertainties, so the risks of making these errors are unavoidable in hypothesis testing
- There are two kinds of errors, which by design cannot be avoided as a result

Significance value and type I error

- If your results show statistical significance, that means they are very unlikely to occur if the null hypothesis is true
- Alpha (α) is the significance value which is typically set at 0.05, this is the cut off at which we accept or reject the null hypothesis. Making α smaller (α = 0.1) makes it harder to reject the H0
- Interpretation of P<0.05 would be: drug X > drug Y 19 out of 20 times you would run the same study
- In this case, you would reject your null hypothesis; but sometimes, this may actually be a Type I error (find a difference when in fact there is none)

Statistical power and type II error

- If your findings do not show statistical significance, they have a high chance of occurring if the null hypothesis is true
- The statistical power of a study (1-β) is the probability of correctly rejecting the null hypothesis (when the null hypothesis is not true)
- The adopted statistical power is usually 80%or 90%
- Therefore, you fail to reject your null hypothesis; but sometimes, this may be a Type II error - so a 10-20% chance of falsely concluding that Drug B is no different than drug A
- The statistical power increases with effect size and sample size

Type I and Type II errors

STUDY FINDING	TRUTH: Null hypothesis is TRUE	TRUTH: Null hypothesis is FALSE
Reject null hypothesis	Type I Error (False positive)	Correct outcome! (True positive)
Fail to reject null hypothesis	Correct outcome! (True negative)	Type II Error (False negative)

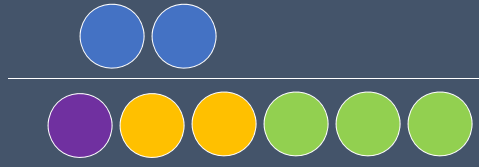
https://www.bing.com/images/search?view=detailV2&ccid=yRBBo%2b%2bP&id=CCB2B4F6BE421A5DA4391354227B531C4C8999&thid=OIP.yRBBo--PyWB62XBo6gsBqAHaD-&mediaurl=https%3a%2f%2fmicrobenotes.com%2fwp-content%2fuploads%2f2020%2f07%2fGraphical-representation-of-type-1-and-type-2-errors.jpg&cdnurl=https%3a%2f%2fth.bing.com%2fth%2fid%2fR6c91041a3ef0fbd6074d97068ea0b01a8%3frik%3dmYIMHBZTeyJUEw%26pid%3dImgRaw&exph=343&expw=640&q=Definition+of+Type+I+Error&simid=608029461347610298&ck=8677CAC4B7DF4F6CD3E09534FB7014CB&selectedIndex=2&FORM=IRPRST&ajaxhist=0

Probability versus Odds

Probability:

The probability is defined as the number of time an event will occurs divided by the number of all possible events

Odds:



The odds are defined as the probability that the event will occur divided by the probability that the event will not occur.

Probability

- **Relative risk / Risk ratio (RR)**

	Cancer	No cancer
Treatment	a	b
Control	c	d

  - The probability of having cancer (event) in the treatment group is a/(a+b)= R1
  - The probability of cancer (event) in the control group is c/(c+d) = R2
  - The ratio of these two probabilities R1/R2 is the relative risk or risk ratio

$$RR = \frac{\text{Risk of event in the Treatment group}}{\text{Risk of event in the Control group}} = \frac{a/(a+b)}{c/(c+d)}$$

Rosner – Fundamental in biostatistics

Odds ratio

- **Odds Ratio (OR)**

	Cancer	No cancer
Treatment	a	b
Control	c	d

  - The odds ratio is the ratio of the odds of an event in the treatment group over the odds of an event in the control group
  - It is equivalent to the probability of an event divided by the probability of a non-event

$$OR = \frac{\text{Odds of event in Treatment group}}{\text{Odds of event in Control group}} = \frac{a/b}{c/d} = ad/bc$$

Rosner – Fundamental in biostatistics

Probability - OR and RR

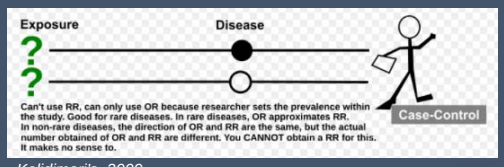
- OR are numerically different from the RR (even if they both compare the same risk between the same group), the relation is nonlinear
- OR and RR are similar when the event is rare in the control group
  - RR=0.15 - the intervention is reduced the risk by 85%
  - OR=0.15 - for every 15 persons who experience the event in the treatment group, 100 subjects will experience the event in the control group

You may also hear about Hazard ratio (HR) which is a measure of an effect of an intervention on an outcome of interest over time. Hazard ratio is reported most commonly in time-to-event analysis or survival analysis

Rosner – Fundamental in biostatistics

Probability - OR and RR

- We use OR in 2 principal situations
  - In case-control studies (subjects with the outcome of interest are matched with a control group who do not) - where the absolute risk (or relative risk) cannot be estimated
  - In logistic regression analyses (models the probability of an event/outcome existing such as success/failure by adjusting for independents variables) where OR are generated as part of the analysis



Kelidimari's, 2009

Mann, Emerg Med J 2003

Absolute Risk vs Relative Risk

https://www.enfic.org/en/understanding-science/article/absolute-vs.-relative-risk-infographic

Number Needed to Treat / Harm

- The Number needed to Treat (NNT) is simply the inverse of the ARR; can be calculated by dividing 100 by the ARR in %
- NNT = 100/ARR
- Note that this is useful if only calculated for a statistically significant difference, and that too has a confidence range
- May be especially useful when explaining to patients

Other closely related entities:

- Number Needed to Harm (NNH) (100/AR increase)
- Number Needed to Screen (NNS) (100/ARR)

PPI side effects… if in fact they are causally related, which most are NOT!...

Table 3. Absolute and RRs for Adverse Effects Associated With Long-Term PPIs

Potential Adverse Effect	Relative Risk	Reference for Risk Estimate	Reference for Incidence Estimate	Absolute Excess Risk
Chronic kidney disease[a]	10% to 20% increase	Lazarus et al[48]	Lazarus et al[48]	0.1% to 0.3% per patient/y
Dementia[b]	4% to 80% increase	Haenisch et al[90]	Haenisch et al[90]	.07% to 1.5% per patient/y
Bone fracture[c]	30% to 4-fold increase	Yang et al[27]	Yang et al[27]	0.1% to 0.5% per patient/y
Myocardial infarction	No association in RCTs	—	—	—
Small intestinal bacterial overgrowth	2-fold to 8-fold increase	Lo et al[91]	None available	Unable to calculate
Campylobacter or *Salmonella* infection	2-fold to 6-fold increase	Bavishi et al[26]	Crim et al[92]	.03% to 0.2% per
Spontaneous bacterial peritonitis[d]	50% to 3-fold increase	Xu et al[93]	Fernandez et al[94]	3% to 16% per
Clostridium difficile infection[e]	No risk to 3-fold increase	Furuya et al[95]	Lessa et al[96]	0% to .09% per
Pneumonia	No association in RCTs	—	—	—
Micronutrient deficiencies[f]	60% to 70% increase	Lam et al[97]	Bailey et al[98]	0.3% to 0.4% per patient/y
Gastrointestinal malignancies	No association in RCTs	—	—	—

NNH = 1 in 100-1,500 (need to take PPI for 1 possible s/e)
vs NNT = 1 in 10-20 for benefit in an approved indication

Vaezi, Gastro, 2017

Diagnostic testing

- Diagnostic testing applies to everything a physician does in order to diagnose a disease or make a clinical decision (*i.e.:* diagnosis).
- From a statistical point of view
  - the clinical decision-making process is based on probabilities
  - the goal of a diagnostic test is to move the estimated probability of disease / event toward either end of the probability scale (i.e., “0” when ruling out/ excluding disease, and “1” when ruling in / confirming a disease / event)

Diagnostic testing – 2x2 table

Gold or reference standard

New diagnostic test under study	Disease	No Disease
Test Positive	A True positives	B False positives
Test Negative	C False negatives	D True negatives

The gold standard is the best single test (or a combination of tests) that is considered the current preferred method of diagnosing a particular disease. Gold Standards are used to define true disease status against which the results of a new diagnostic test are compared. A reference standard is the closest gold standard that we have; for example, Colonoscopy is a reference standard (since there is a possibility of missing lesions)

Diagnostic testing – Sensitivity

	Disease	No Disease
Test Positive	A True positives	B False positives
Test Negative	C False negatives	D True negatives

Sensitivity is the probability that an individual with the disease of interest has a positive test (expressed in %)

Sensitivity = a/(a+c)

Diagnostic testing – Specificity

	Disease	No Disease
Test Positive	A True positives	B False positives
Test Negative	C False negatives	D True negatives

Specificity is the probability that an individual without the disease of interest has a negative test (expressed in %)

Specificity = d/(b+d).

Diagnostic testing – Accuracy

	Disease	No Disease
Test Positive	A True positives	B False positives
Test Negative	C False negatives	D True negatives

Accuracy is the probability that the diagnostic test yields the correct determination with regards to presence of the disease

Accuracy= (a+d)/(a+b+c+d)

Diagnostic testing – Positive Predictive Value

	Disease	No Disease
Test Positive	A True positives	B False positives
Test Negative	C False negatives	D True negatives

Positive Predictive Value (PV+) is the probability of disease in an individual with a positive test result

Positive Predictive Value: a/(a+b)

Diagnostic testing – Negative Predictive Value

	Disease	No Disease
Test Positive	A True positives	B False positives
Test Negative	C False negatives	D True negatives

Negative Predictive Value (PV -) is the probability of not having the disease when the test result is negative

Negative Predictive Value : d/(c+d)

Diagnostic testing – Prevalence

	Disease	No Disease
Test Positive	A True positives	B False positives
Test Negative	C False negatives	D True negatives

Prevalence is the probability of having the disease, also called the "prior probability" of having the disease

Prevalence: (a+c)/(a+b+c+d)

FIT test performance

Table 2. Synopsis of Results From Subgroup Analyses Depending on Cutoff Value, Type of FIT and Number of FIT Samples Used for the Diagnosis of Colorectal Cancer or Advanced Neoplasia

Characteristic	Studies, No.	Participants, No.	Sensitivity, % (95% CI)	Specificity, % (95% CI)	Positive Likelihood Ratio (95% CI)	Negative Likelihood Ratio (95% CI)	Diagnostic Odds Ratio (95% CI)	Positive Predictive Value, %	Negative Predictive Value, %
Colorectal Cancer									
Cut off <15[a]	4	3274	93.0 (63.0-99.0)	91.0 (90.0-92.0)	10.2 (8.1-12.8)	0.08 (0.01-0.53)	130.0 (16.0-1057.0)	6.8	99.9
Cut off 15-25[a]	4	2539	93.0 (73.0-99)	94.0 (91.0-96.0)	15.1 (9.5-23.9)	0.07 (0.02-0.32)	209.0 (36.0-1195.0)	12.3	99.9
Cut off >25[a]	2	1167	NA[b]	NA[b]	NA[b]	NA[b]	NA[b]	NA[b]	NA[b]
Quantitative FIT	6	4218	94.0 (73.0-99.0)	91.0 (89.0-93.0)	10.7 (8.3-14.0)	0.07 (0.01-0.35)	165.0 (25.0-1086.0)	7.8	99.9
Qualitative FIT	1	572	NA[b]	NA[b]	NA[b]	NA[b]	NA[b]	NA[b]	NA[b]
One FIT sample	6	4362	94.0 (39.0-100)	91.0 (90.0-93.0)	11.0 (8.0-15.1)	0.06 (0.00-1.34)	182.0 (6.0-5382.0)	7.8	99.9
Two FIT samples	3	2046	NA[b]	NA[b]	NA[b]	NA[b]	NA[b]	NA[b]	NA[b]
Three FIT samples	2	1428	NA[b]	NA[b]	NA[b]	NA[b]	NA[b]	NA[b]	NA[b]
Advanced Neoplasia									
Cut off <15[a]	7	3909	49.0 (38.0-60.0)	93.0 (90.0-94.0)	6.6 (4.9-8.8)	0.55 (0.45-0.68)	12.0 (8.0-19.0)	44.6	94.1
Cut off 15-25[a]	5	2712	42.0 (32.0-54.0)	97.0 (95.0-98.0)	13.1 (9.2-18.6)	0.6 (0.5-0.72)	22.0 (15.0-31.0)	62.9	93.2
Cut off >25[a]	3	1821	NA[b]	NA[b]	NA[b]	NA[b]	NA[b]	NA[b]	NA[b]
Quantitative FIT	8	4737	47.0 (38.0-56.0)	94.0 (91.0-95.0)	7.3 (5.3-10.2)	0.57 (0.48-0.67)	13.0 (9.0-19.0)	47.9	93.8
Qualitative FIT	4	1467	54.0 (27.0-79.0)	90.0 (87.0-93.0)	5.6 (3.6-8.7)	0.51 (0.27-0.95)	11.0 (4.0-31.0)	28.6	96.3
One FIT sample	11	5776	45.0 (37.0-54.0)	93.0 (90.0-95.0)	6.2 (4.7-8.3)	0.59 (0.51-0.69)	11.0 (7.0-16.0)	42.2	93.7
Two FIT samples	3	2046	NA[b]	NA[b]	NA[b]	NA[b]	NA[b]	NA[b]	NA[b]
Three FIT samples	2	1428	NA[b]	NA[b]	NA[b]	NA[b]	NA[b]	NA[b]	NA[b]

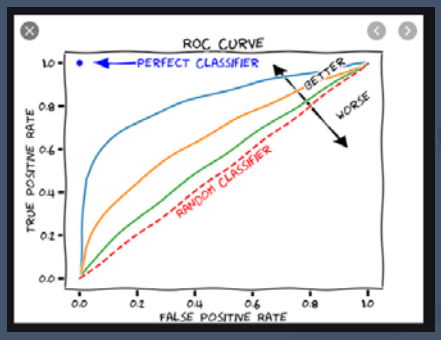
Abbreviations: FIT, fecal immunochemical test; NA, not available.
[a] Cut off value for a positive test result, µg/g.
[b] Insufficient data for pooling results.

Katsoula, JAMA Int Med, 2017

Diagnostic testing – ROC Curve

- A **receiver operating characteristics curve**, or **ROC curve**, is a graphical plot that illustrates the ability of a diagnostic test to discriminate between disease vs no disease according to possible thresholds

sensitivity



1- specificity

https://glassboxmedicine.com

Diagnostic testing ROC analysis - example

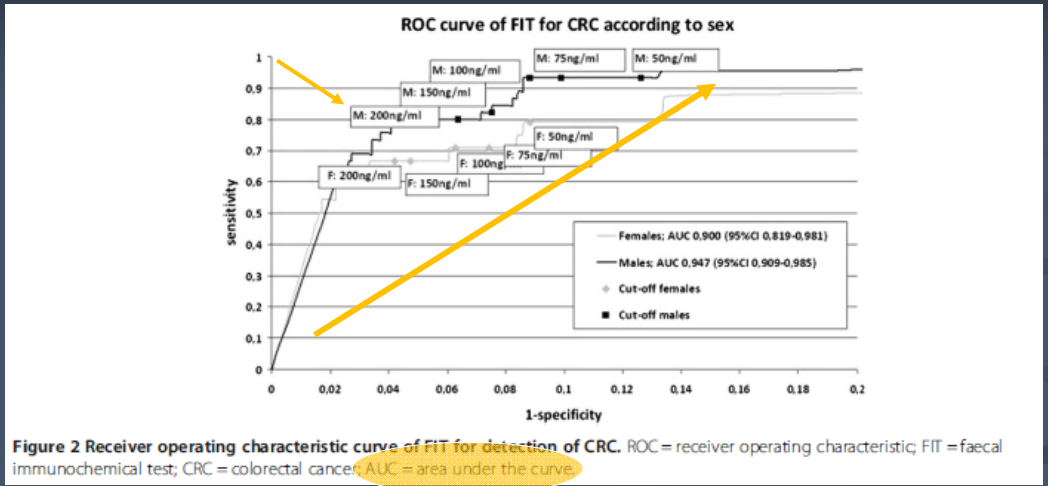
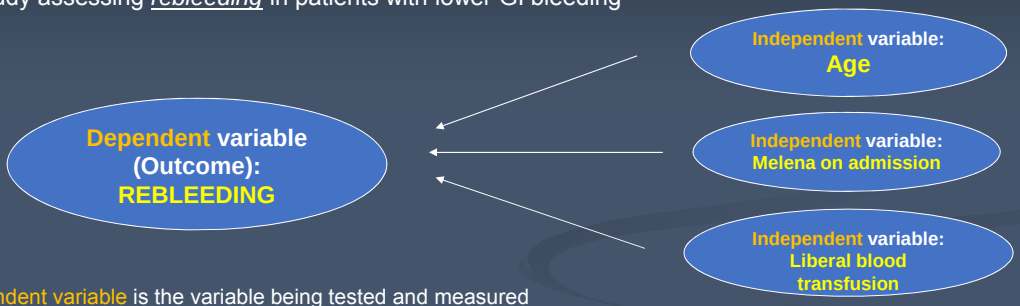


Figure 2 Receiver operating characteristic curve of FIT for detection of CRC. ROC = receiver operating characteristic; FIT = faecal immunochemical test; CRC = colorectal cancer; AUC = area under the curve.

Turenhout et al., BMC gastro 2014

Logistic Regression model dependent versus independent variable

Example of a study assessing *rebleeding* in patients with lower GI bleeding



A dependent variable is the variable being tested and measured in an experiment/study for a given outcome/endpoint.
You can have outcomes such as mortality, complications, quality of life, satisfaction…

Independent variables are variables that can be changed or controls and are assumed to have a direct effect on the dependent variable (Demographics/labs/interventions)

Kherad et al., APT 2019

Logistic Regression model

- The logit of the multiple logistic regression model is given by the equation:

Dependent variable

Odd of an event

Independent variable

$$\text{Logic Y(x)} = \ln \frac{p}{(1-P)} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p$$

Intercept

Coefficients

… and you can identify which "x"s are statistically significant prognosticators of Y

PREDICTORS OF POOR BOWEL PREPARATION

Variable	Adequate preparation	Inadequate preparation	p-value
Female	54.8%	48.4%	0.13
Age	55.9 ± 12.9	59.8 ± 12.8	<0.01
BMI	27.2 ± 5.7	28.6 ± 7.2	<0.01
Comorbidities			
Myocardial infarction	2.6%	5.1%	0.08
Congestive heart failure	0.6%	0.7%	0.97
Peripheral vascular disease	2.3%	3.9%	0.26
Cerebrovascular disease	1.5%	2.6%	0.30
Dementia	0.3%	0.7%	0.40
Chronic pulmonary disease	5.3%	8.3%	0.12
Connective tissue disease	1.5%	2.6%	0.31
Ulcer disease	3.7%	5.1%	0.38
Mild or moderate liver disease	2.6%	1.9%	0.80
Diabetes	6.8%	18.5%	<0.01
Moderate renal disease	1.2%	1.9%	0.44
Diabetes with end organ damage	0.3%	0.7%	0.40
Any tumor	8.5%	11.5%	0.21
Leukemia	0.1%	0.0%	1.00
Lymphoma	0.7%	1.9%	0.13
Metastaic solid tumor	0.3%	0.6%	0.40
AIDS	0.4%	0.0%	1.00
Neurologic disorder	2.4%	4.5%	0.18
Previous abdominal or pelvic surgery	38.0%	42.7%	0.25
Charlson score			
French or English as a primary language	95.1%	93.6%	0.43
Highest degree of education			
Patient requiring help for bowel preparation instruction at home	0.8%	2.6%	0.06
Irritable bowel syndrome constipation (Rome III)	11.1%	14.8%	0.19
Functional constipation (Rome III)	10.2%	13.0%	0.28
Known IBD	7.5%	6.1%	0.52
Previous colonoscopy	56.6%	57.1%	0.91
Narcotics or chronic laxative or medication induced constipation	11.3%	12.9%	0.56
Indication			
Non screening	35.1%	46.5%	0.02
Screening	40.5%	33.1%	
Surveillance	24.4%	20.4%	
Split dose (vs same-day)	67.0%	68.2%	0.78
High volume	34.2%	28.8%	0.06

Patient Characteristics

Interventions

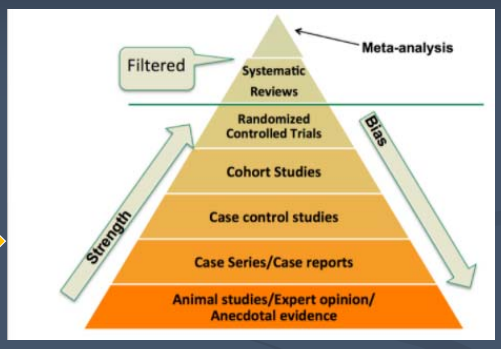
Meta-analysis

- Meta-analysis is the statistical combination of results from two or more separate studies
- Potential advantages of meta-analyses include an improvement in precision (brought about by larger sample sizes)
- Most meta-analysis methods are variations on a weighted average of the effect estimates from the different studies.
- Variation across studies (heterogeneity) must be considered.
- Meta-analyses also have the potential to mislead seriously, particularly if specific study designs, within-study biases, variation across studies, and reporting biases are not carefully considered

Cochrane Handbook

Meta-analysis

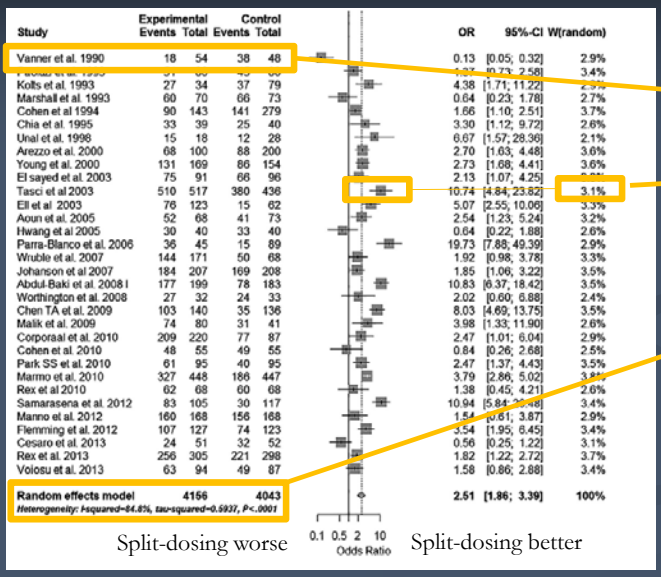
Steps	Steps to conduct a meta-analysis
1	Specify the question to be answered (PICO)
2	Define the inclusion and exclusion criteria
3	Conduct a systematic review of the literature and identify all the relevant citations
4	Data extraction for selected articles
5	Evaluate the risk of bias of studies
6	Conduct statistical analyses
7	Conclude and assess the limit of the meta-analysis



All meta-analyses should be registered in Prospero
https://www.crd.york.ac.uk/prospero/

Cochrane Handbook

Meta analysis: Colonoscopy preparations



Data extracted from the 2x2 table (dichotomized outcome) for each study

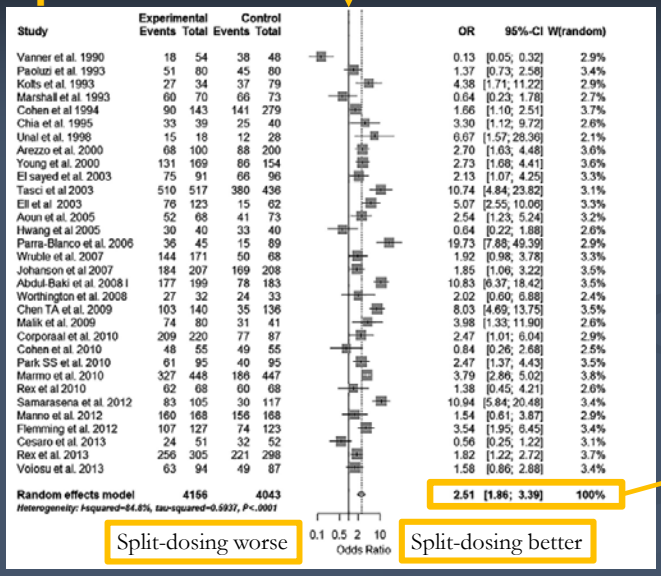
Weight of each individual study (also related to size of box)

If the p-value<0.10, the test is considered to be heterogeneous (variation in study outcomes between studies), a random effect model is needed. Otherwise, a fixed effect model will be preferred

Martel et al, Gastro 2015

Colonoscopy preps

Line of unity (OR=1); if overlapped study result is not significant



Study	Experimental Events	Experimental Total	Control Events	Control Total	OR	95%-CI	W(random)
Vanner et al. 1990	18	54	38	48	0.13	[0.05; 0.32]	2.9%
Paoluzi et al. 1993	51	80	45	80	1.37	[0.73; 2.58]	3.4%
Kolts et al. 1993	27	34	37	79	4.38	[1.71; 11.22]	2.9%
Marshall et al. 1993	60	70	66	73	0.64	[0.23; 1.78]	2.7%
Cohen et al 1994	90	143	141	279	1.66	[1.10; 2.51]	3.7%
Chia et al. 1995	33	39	25	40	3.30	[1.12; 9.72]	2.6%
Unal et al. 1998	15	18	12	28	6.67	[1.57; 28.36]	2.1%
Arezzo et al. 2000	68	100	88	200	2.70	[1.63; 4.48]	3.6%
Young et al. 2000	131	169	86	154	2.73	[1.68; 4.41]	3.6%
El sayed et al. 2003	75	91	66	96	2.13	[1.07; 4.25]	3.3%
Tasci et al 2003	510	517	380	436	10.74	[4.84; 23.82]	3.1%
Ell et al 2003	76	123	15	62	5.07	[2.55; 10.06]	3.3%
Aoun et al. 2005	52	68	41	73	2.54	[1.23; 5.24]	3.2%
Hwang et al 2005	30	40	33	40	0.64	[0.22; 1.88]	2.6%
Parra-Blanco et al. 2006	36	45	15	89	19.73	[7.88; 49.39]	2.9%
Wruble et al. 2007	144	171	50	68	1.92	[0.98; 3.78]	3.3%
Johanson et al 2007	184	207	169	208	1.85	[1.06; 3.22]	3.5%
Abdul-Baki et al. 2008 I	177	199	78	183	10.83	[6.37; 18.42]	3.5%
Worthington et al. 2008	27	32	24	33	2.02	[0.60; 6.88]	2.4%
Chen TA et al. 2009	103	140	35	136	8.03	[4.69; 13.75]	3.5%
Malik et al. 2009	74	80	31	41	3.98	[1.33; 11.90]	2.6%
Corporaal et al. 2010	209	220	77	87	2.47	[1.01; 6.04]	2.9%
Cohen et al. 2010	48	55	49	55	0.84	[0.26; 2.68]	2.5%
Park SS et al. 2010	61	95	40	95	2.47	[1.37; 4.43]	3.5%
Marmo et al. 2010	327	448	186	447	3.79	[2.86; 5.02]	3.8%
Rex et al 2010	62	68	60	68	1.38	[0.45; 4.21]	2.6%
Samarasena et al. 2012	83	105	30	117	10.94	[5.84; 20.48]	3.4%
Manno et al. 2012	160	168	156	168	1.54	[0.61; 3.87]	2.9%
Flemming et al. 2012	107	127	74	123	3.54	[1.95; 6.45]	3.4%
Cesaro et al. 2013	24	51	32	52	0.56	[0.25; 1.22]	3.1%
Rex et al. 2013	256	305	221	298	1.82	[1.22; 2.72]	3.7%
Volosu et al. 2013	63	94	49	87	1.58	[0.86; 2.88]	3.4%
Random effects model	**4156**		**4043**		**2.51**	**[1.86; 3.39]**	**100%**

Heterogeneity: I-squared=64.8%, tau-squared=0.5937, P<.0001

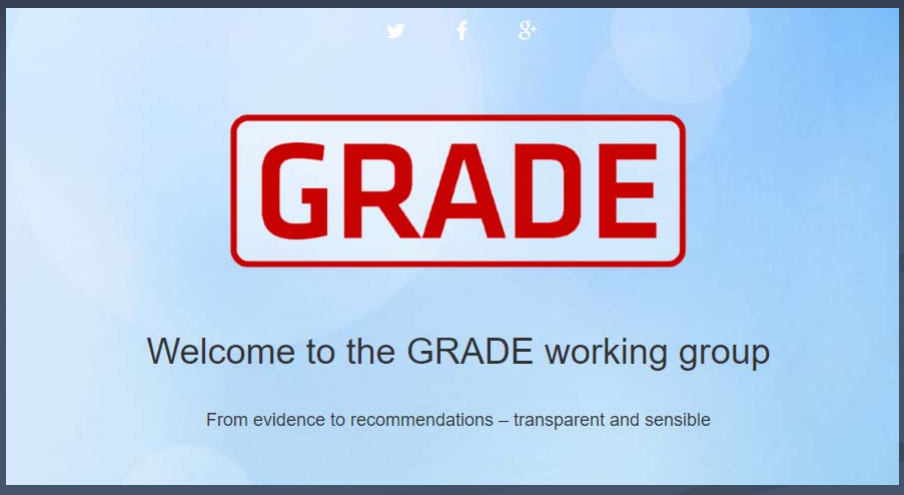
0.1 0.5 2 10
Odds Ratio

Split-dosing worse

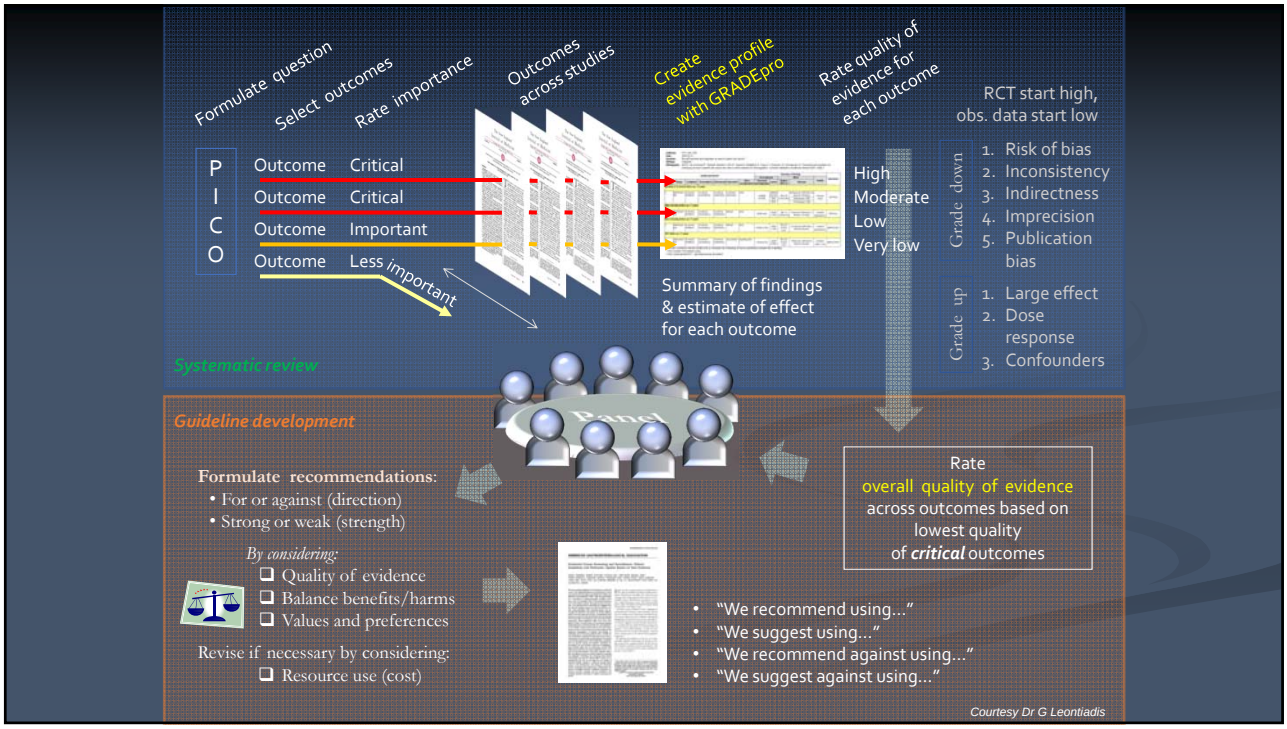
Split-dosing better

Overall OR and 95%CI;it is significant as does not overlap OR=1

Martel et al, Gastro 2015



https://www.gradeworkinggroup.org/



Conclusion

- Inferential testing with assumptions about the sample distributions; choosing the correct inferential test
- Hypothesis testing: significance, statistical power, types I/II errors
- Probabilities vs Odds ratios; absolute/relative risks; NNT/H/S Diagnostic testing, ROC analysis
- Confounding and adjusting for confounding
- Meta-analyses
- HOPE THIS HELPS MAKE SENSE OF SOME OF YOUR READINGS!