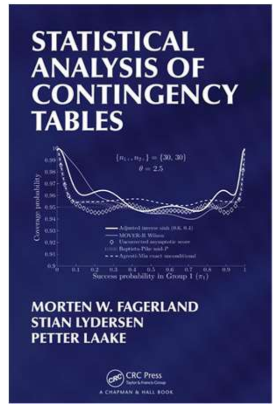
NTNU – Trondheim
Norwegian University of Science and Technology

Contingency tables:
How to choose appropriate methods for analysis

by
Stian Lydersen
7th Nordic-Baltic Biometric Conference
Vilnius 3-5 June 2019
(revised 19 August 2019)



Fagerland, M. W., Lydersen, S., & Laake, P.:
"Statistical Analysis of Contingency Tables."
Chapman and Hall/CRC, 2017.

www.contingencytables.com

A contingency table, or a cross-tabulation:
A display of the observed counts of categorical variables.

Table with r rows and c columns:

1x2
1xc
2x2
Ordered rx2
Ordered 2xc
Unordered rxc
Singly ordered rxc
Doubly ordered rxc
Paired 2x2
Paired cxc
… and higher ordered tables

Examples

TABLE 2.2
Male and female births among Indian immigrants in Norway, 1987–1996 (Singh et al., 2010)

Birth order	Males	Females	Total
1st	250	283	533
2nd	204	208	412
3rd	103	64	167
4th	33	12	45

We shall regard each line (birth order) as a 1x2 contingency table, assuming a binomial distribution
General population: Pr(male offspring) = 0.513

TABLE 4.3
Treatment of epinephrine in children with cardiac arrest (Perondi et al., 2004)

	Survival at 24h		
Treatment	Yes	No	Total
Standard dose	7 (21%)	27 (79%)	34*
High dose	1 (2.9%)	33 (97%)	34*
Total	8 (12%)	60 (88%)	68*

*Fixed by design

The number of successes n_{i1} in row number i is assumed bin(n_{i+}, π_i)

A cross-over RCT: The number of subjects with satisfactory relief of symptoms at least one week, Ligaarden et al. (2010).

	Placebo period		
LpMF1298 period	no	yes	Total
no	4	7	11
yes	1	4	5
Total	5	11	16

Effect measure (parameter of interest), examples:

- 1x2 table:
  - Success probability π in the binomial distribution
- 2x2 table, two independent binomials (and several other settings):
  - Difference between probabilities (risk difference, absolute risk reduction, attributable risk) $\Delta = \pi_1 - \pi_2.$
  - Ratio of probabilities (risk ratio, relative risk): $\phi = \frac{\pi_1}{\pi_2}$
  - Odds ratio $\theta = \frac{\pi_1/(1-\pi_1)}{\pi_2/(1-\pi_2)}$

8

Evaluation criteria for confidence intervals:

1. Coverage probability: Ought to not dip (much) below the nominal coverage (usually 95%). This is the primary criterion.
2. Interval width: Among intervals with similar coverage, we prefer the narrower.
3. Interval location: For a (1−α) confidence interval, we prefer the left and right non-coverage to be near α/2.

9

Evaluation criteria for hypothesis tests:

1. The actual significance level (ASL) should ideally equal the nominal significance level (usually 5%). If the ASL level is lower, say 2% or 3%, the test is conservative. If the ASL is higher, the test is liberal.
2. Among tests with acceptable ASL, we prefer the one with highest power.

10

Some methods for statistical inference:

- Wald
- Likelihood ratio
- Score

11

Let θ be the parameter of interest, maximum likelihood estimate $\hat{\theta}$, standard error $SE(\widehat{\theta})$ be its standard error, $H_0 : \theta = \theta_0$.

The Wald statistic

$$Z_{\text{Wald}} = (\hat{\theta} - \theta_0)/\widehat{\text{SE}}(\hat{\theta})$$

is asymptotically standard normally distributed under H_0 .

Let l_0 and l_1 be the maximized likelihood under the null and alternative hypotheses, and L_0 and L_1 their logarithms.

The likelihood ratio statistic

$$T_{\text{LR}} = -2\log\Lambda = -2\log(l_0/l_1) = -2(L_0 - L_1)$$

is asymptotically standard chi-squared distributed under H_0 .

The score function is defined as

$$u(\theta) = \partial L(\theta)/\partial\theta$$

The score statistic in chi-squared form is

$$T_{\text{score}} = \frac{\left[\partial L(\theta)/\partial\theta\right]^2}{-\text{E}\left[\partial^2 L(\theta)/\partial\theta^2\right]}$$

evaluated at θ_0 .

Exact p-value:
Sum of the exact probabilities of possible tables (**x**) that agree less than or equally with the null hypothesis than does the observed table (**n**):

$$\text{exact } P\text{-value} = \sum_{\text{all tables}} I\left[T(\mathbf{x}) \geq T(\mathbf{n})\right] \cdot f(\mathbf{x} \mid H_0).$$

Exact confidence intervals can be obtained by inverting exact tests

The mid-*P* value includes only half the point probability of the tables that agree equally with the null hypothesis as the observed table:

$$\text{mid-}P \text{ value} = \sum_{\text{all tables}} I\left[T(\mathbf{x}) > T(\mathbf{n})\right] \cdot f(\mathbf{x}|H_0) + 0.5 \cdot \sum_{\text{all tables}} I\left[T(\mathbf{x}) = T(\mathbf{n})\right] \cdot f(\mathbf{x}|H_0).$$

Mid-*P* confidence intervals can be obtained by inverting mid-*P* tests

The 1x2 table and the binomial distribution: Confidence intervals

Let $X \sim bin(n,\pi)$, and $\hat{\pi} = X/n$.

The Wald statistic is

$$\frac{\hat{\pi}-\pi}{\widehat{SE}(\hat{\pi})} = \frac{\hat{\pi}-\pi}{\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}}$$

Equating this to $\pm z_{\alpha/2}$ and solving for π gives the Wald confidence interval:

$$\hat{\pi} \pm z_{\alpha/2}\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}$$

The Wilson Score Interval (Wilson, 1927)

The score statistic is

$$\frac{\hat{\pi}-\pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}$$

Equating this to $\pm z_{\alpha/2}$ and solving for π_0 gives the score confidence interval:

$$\frac{2n\hat{\pi}+z^2_{\alpha/2}\pm z_{\alpha/2}\sqrt{z^2_{\alpha/2}+4n\hat{\pi}(1-\hat{\pi})}}{2\left(n+z^2_{\alpha/2}\right)}$$

The Agresti-Coull Interval (1998)

Add two (pseudo-successes) and two (pseudo-failures) to the data, and compute the Wald interval as if the data were X˜ = X + 2 successes in n˜ = n + 4 trials.

This can be viewed as an approximation to the Wilson Score interval.

The Clopper-Pearson exact interval (L,U) is obtained by inverting two one-sided tests, as solutions of:

$$\sum_{i=x}^{n}\binom{n}{i}L^i(1-L)^{n-i}=\alpha/2 \quad \text{and} \quad \sum_{i=0}^{x}\binom{n}{i}U^i(1-U)^{n-i}=\alpha/2$$

It can be expressed as

$$L=B(\alpha/2;x,n-x+1) \quad \text{and} \quad U=B(1-\alpha/2;x+1,n-x)$$

where $B(z;a,b)$ is the lower z-quantile of the beta distribution with parameters a and b.

It is the shortest interval whose left and right non-coverage will be at most $\alpha/2$ for all values of π .

The Blaker (2000) exact interval inverts one two-sided exact test of size α. It does not guarantee that the noncoverage in each tail is limited by α/2, but only that the total non-coverage is limited by α.

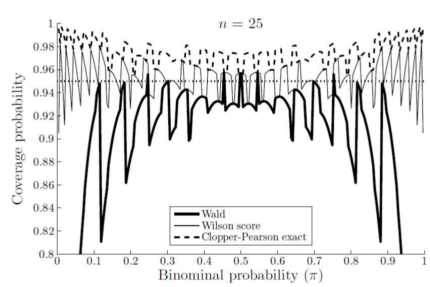


FIGURE 2.2
Coverage probabilities of three confidence intervals for the binomial probability

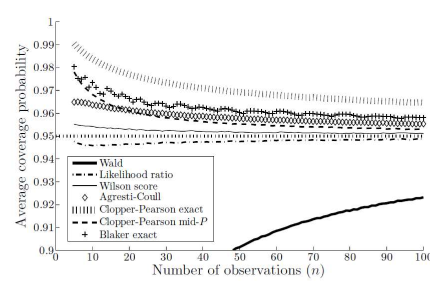


FIGURE 2.10
Average coverage probabilities of seven confidence intervals for the binomial probability as functions of the number of observations

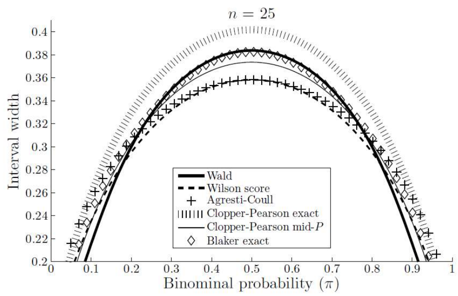


FIGURE 2.11
Expected width of six confidence intervals for the binomial probability

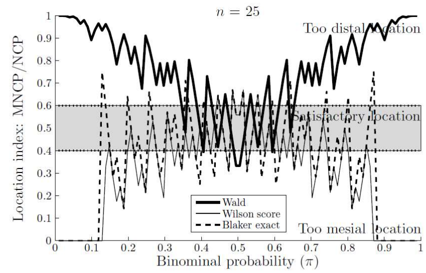


FIGURE 2.12
Location, as measured by the MNCP/NCP index, of three confidence intervals for the binomial probability

TABLE 2.7
Recommended confidence intervals (CIs) and tests for the binomial parameter

Analysis	Recommended methods	Sample sizes
CIs for the binomial parameter	Wilson score*	all
	Blaker exact	small/medium
	Clopper-Pearson mid-P	medium
	Agresti-Coull*	medium/large
Tests for the binomial parameter	Score*	all
	Blaker exact	small/medium
	Mid-P binomial	medium

*These methods have closed-form expression

The Wald interval and the Wald test are not recommended!

TABLE 2.2
Male and female births among Indian immigrants in Norway, 1987–1996 (Singh et al., 2010)

Birth order	Males	Females	Total
1st	250	283	533
2nd	204	208	412
3rd	103	64	167
4th	33	12	45

We shall regard each line (birth order) as a 1x2 contingency table, assuming a binomial distribution
General population: Pr(male offspring) = 0.513

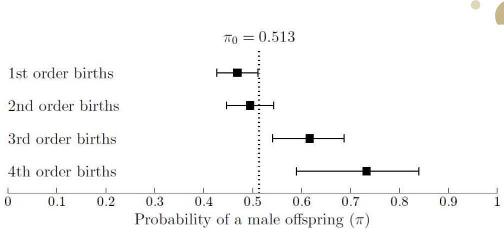


FIGURE 2.1
95% Wilson score confidence intervals based on the data in Table 2.2

The 2x2 table: Hypothesis tests for association

TABLE 4.3
Treatment of epinephrine in children with cardiac arrest (Perondi et al., 2004)

	Survival at 24h		
Treatment	**Yes**	**No**	**Total**
Standard dose	7 (21%)	27 (79%)	34*
High dose	1 (2.9%)	33 (97%)	34*
Total	8 (12%)	60 (88%)	68*

*Fixed by design

The number of successes n_{i1} in row number i is assumed bin(n_{i+} , π_i)

TABLE 4.1
The observed counts of a 2 × 2 table

	Success	**Failure**	**Total**
Group 1	n_{11}	n_{12}	n_{1+}
Group 2	n_{21}	n_{22}	n_{2+}
Total	n_{+1}	n_{+2}	N

Additional notation:
$\mathbf{n} = \{n_{11}, n_{12}, n_{21}, n_{22}\}$: the observed table
$\mathbf{x} = \{x_{11}, x_{12}, x_{21}, x_{22}\}$: any possible table

The Pearson chi squared test statistic is

$$T_{Pearson}(\mathbf{n}) = \sum_{i,j} \frac{(n_{ij} - m_{ij})^2}{m_{ij}}$$

where $m_{ij} = n_{i+}n_{+j} / N$ are the expected counts under H_0.

Yates's correction

$$T_{YatesCC}(\mathbf{n}) = \sum_{i,j} \frac{(|n_{ij} - m_{ij}| - 1/2)^2}{m_{ij}}$$

is a historic curiosity and should never be used!

Exact p-value: In general

$$\text{exact } P\text{-value} = \sum_{\text{all tables}} I[T(\mathbf{x}) \geq T(\mathbf{n})] \cdot f(\mathbf{x} \mid H_0).$$

Fixed row sums (two binomials),
common success probability $\pi = \pi_1 = \pi_2$ under H_0:

$$f(\mathbf{x} \mid \pi, \mathbf{n}_+) = \binom{n_{1+}}{x_{11}} \binom{n_{2+}}{x_{21}} \pi^{x_{11}+x_{21}} (1-\pi)^{N-x_{11}-x_{21}}$$

The exact P-value depends on the unknown π!

Possible solutions:

1. Exact conditional test (Fisher exact test)
2. Mid-P test (quasi-exact)
3. Exact unconditional test

Possible counts x_{ij} if the marginals are fixed (or conditioned on)

TABLE 4.1
The observed counts of a 2 × 2 table

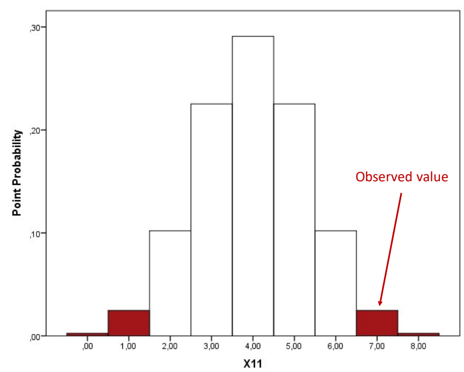
	Success	**Failure**	**Total**
Group 1	x_{11}		n_{1+}
Group 2			n_{2+}
Total	n_{+1}	n_{+2}	N

Additional notation:
$\mathbf{n} = \{n_{11}, n_{12}, n_{21}, n_{22}\}$: the observed table
$\mathbf{x} = \{x_{11}, x_{12}, x_{21}, x_{22}\}$: any possible table

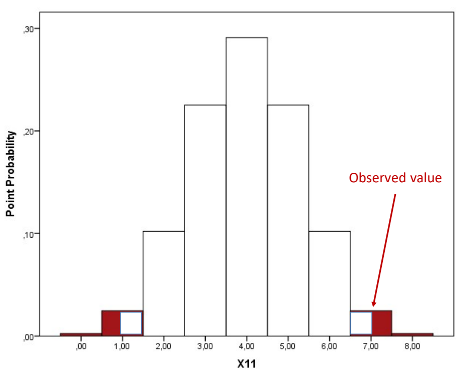
The conditional distribution of x_{11} under H_0 is hypergeometric and does not depend on any unknown parameter!

$$f(x_{11} \mid n_{1+}, n_{2+}, n_{+1}, n_{+2}) = \frac{\binom{n_{1+}}{x_{11}}\binom{n_{2+}}{n_{+1} - x_{11}}}{\binom{N}{n_{+1}}}$$

P-value = P(Observed or more extreme)



Mid-P value = 0.5 P(Observed) + P(more extreme)



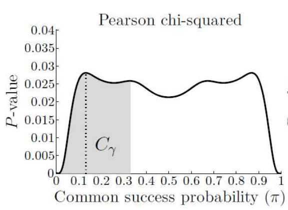
$$\text{exact } P\text{-value} = \sum_{\text{all tables}} I\big[T(\mathbf{x}) \geq T(\mathbf{n})\big] \cdot f(\mathbf{x} \mid H_0).$$

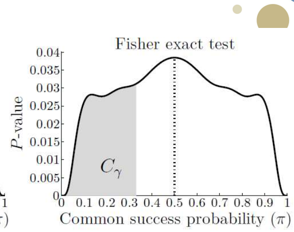
Fixed row sums (two binomials):

$$f(\mathbf{x} \mid \pi, \mathbf{n}_{+}) = \binom{n_{1+}}{x_{11}}\binom{n_{2+}}{x_{21}} \pi^{x_{11}+x_{21}}(1-\pi)^{N-x_{11}-x_{21}}$$

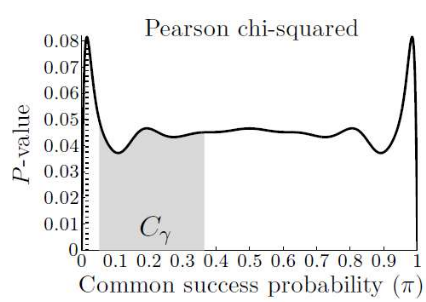
Exact unconditional P-value:

$$P\text{-value} = \max_{0 \leq \pi \leq 1} \left\{ \sum_{\Omega(\mathbf{x}|\mathbf{n}_{+})} I\big[T(\mathbf{x}) \geq T(\mathbf{n})\big] \cdot f(\mathbf{x} \mid \pi, \mathbf{n}_{+}) \right\}$$





Also called:
Unconditional z-pooled
Exact Suissa-Shuster (1985)



Berger and Boos confidence interval method:

$$P\text{-value} = \max_{\pi \in C_\gamma} \left\{ \sum_{\Omega(\mathbf{x}|\mathbf{n}_+)} I[T(\mathbf{x}) \geq T(\mathbf{n})] \cdot f(\mathbf{x} \,|\, \pi, \mathbf{n}_+) \right\} + \gamma$$

γ=0.0001 is generally recommended when α=0.05 or 0.01: (Lydersen, Langaas, Bakke (2012)

TABLE 4.12
Results of exact and mid-P tests on the epinephrine trial data in Table 4.3

Test	Statistic	P-value	$\pi^*_{\max}$
Fisher exact (Fisher-Irwin)	Hypergm[†] distribution	0.0544	n/a
Fisher exact	Pearson chi-squared	0.0544	n/a
Fisher exact	Likelihood ratio	0.0544	n/a
Fisher mid-P	Hypergm[†] distribution	0.0297	n/a
Suissa-Shuster exact uncond.	Pearson chi-squared	0.0281	0.132
Exact unconditional	Likelihood ratio	0.0402	0.076
Exact unconditional	Unpooled Z	0.0281	0.132
Fisher-Boschloo exact uncond.	Fisher exact test	0.0385	0.500

*Value of the common success probability at which the maximum P-value occurred
†Hypergm = Hypergeometric

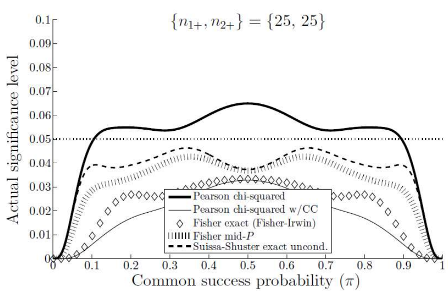


FIGURE 4.4
Actual significance levels of five commonly used tests

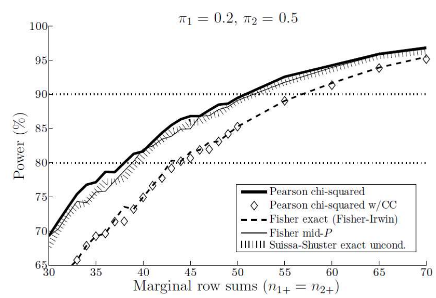


FIGURE 4.10
Power of five commonly used tests

TABLE 4.24
Recommended tests and confidence intervals (CIs) for 2 × 2 tables

Analysis	Recommended methods	Sample sizes
Tests for association	Fisher mid-P*	all
	Suissa-Shuster exact unconditional[†]	small/medium
	Fisher-Boschloo exact uncond.[†]	small/medium
	Pearson chi-squared*	large
CIs for difference between probabilities	Agresti-Min exact unconditional[†]	small/medium
	Agresti-Caffo*	medium/large
	Newcombe hybrid score*	medium/large
	Miettinen-Nurminen asympt. score	medium/large
	Wald*	large
CIs for number needed to treat	The reciprocals of the limits of the recommended intervals for the difference between probabilities	
CIs for ratio of probabilities	Adjusted inverse sinh*	all
	MOVER-R Wilson*	all
	Koopman asymptotic score	all
	Agresti-Min exact unconditional[†]	small/medium
	Katz log*	large
CIs for odds ratio	Adjusted inverse sinh*	all
	MOVER-R Wilson*	all
	Baptista-Pike mid-P	all
	Agresti-Min exact unconditional[†]	small/medium
	Woolf logit*	large

*These methods have closed-form expression
†Preferably with the Berger and Boos procedure ($\gamma = 0.0001$)

The paired 2x2 table: Hypothesis testing

A cross-over RCT: The number of subjects with satisfactory relief of symptoms at least one week, Ligaarden et al. (2010).

	Placebo period		
LpMF1298 period	no	yes	Total
no	4	7	11
yes	1	4	5
Total	5	11	16

TABLE 8.5
The joint probabilities of a paired 2 × 2 table

	Event B		
Event A	Success	Failure	Total
Success	π_{11}	π_{12}	π_{1+}
Failure	π_{21}	π_{22}	π_{2+}
Total	π_{+1}	π_{+2}	1

Additional notation: $\boldsymbol{\pi} = \{\pi_{11}, \pi_{12}, \pi_{21}, \pi_{22}\}$

$H_0 : \pi_{1+} = \pi_{+1}$

TABLE 8.1
The observed counts of a paired 2 × 2 table

	Event B		
Event A	Success	Failure	Total
Success	n_{11}	n_{12}	n_{1+}
Failure	n_{21}	n_{22}	n_{2+}
Total	n_{+1}	n_{+2}	N

Additional notation:
$\mathbf{n} = \{n_{11}, n_{12}, n_{21}, n_{22}\}$: the observed table
$\mathbf{x} = \{x_{11}, x_{12}, x_{21}, x_{22}\}$: any possible table

Conditionally on n_{11} and the number of discordant pairs $n_d = n_{12} + n_{21}$:

$n_{12} \sim bin(n_d, \mu)$ where $\mu = \pi_{12}/(\pi_{12} + \pi_{21})$

Under H_0, $n_{12} \sim bin(n_d, 1/2)$

The McNemar test statistic (1947):

$$Z_{\text{McNemar}}(\mathbf{n}) = \frac{n_{12} - \frac{1}{2}(n_{12} + n_{21})}{\widehat{\text{SE}}_0(n_{12})} = \frac{n_{12} - n_{21}}{\sqrt{n_{12} + n_{21}}}.$$

Versions of the McNemar test:

- Asymptotic
- Asymptotic with continuity correction
- Exact conditional
- Mid-p
- Unconditional
- Unconditional with Berger & Boos confidence interval

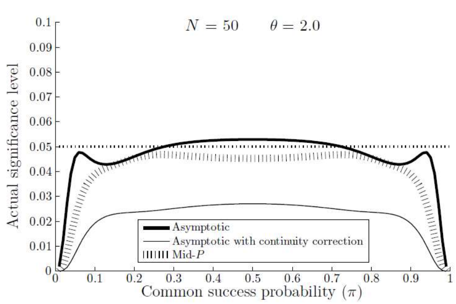


FIGURE 8.2
Actual significance levels of three McNemar tests

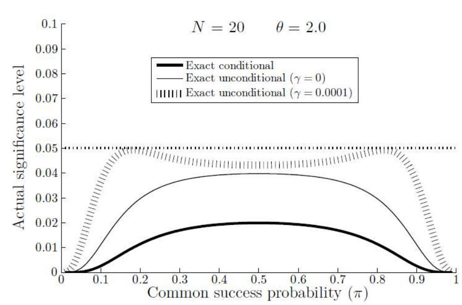


FIGURE 8.3
Actual significance levels of three exact McNemar tests

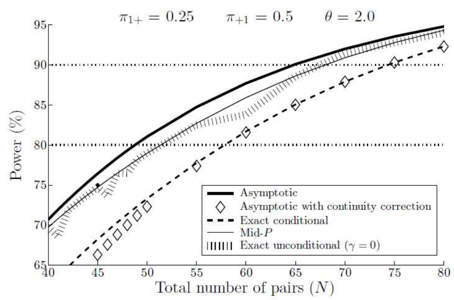


FIGURE 8.6
Power of five McNemar tests

TABLE 8.15
Recommended tests and confidence intervals (CIs) for paired 2 × 2 tables

Analysis	Recommended methods	Sample sizes
Tests for association	McNemar asymptotic*	all
	McNemar mid-P*	all
	McNemar exact unconditional†	small/medium
CIs for difference between probabilities	Wald with Bonett-Price adjust.*	all
	Newcombe square-and-add*	small/medium
	Sidik exact unconditional†	small/medium
CIs for number needed to treat	The reciprocals of the limits of the recommended intervals for the difference between probabilities	
CIs for ratio of probabilities	Bonett-Price hybrid Wilson score*	all
	Tang asymptotic score	all
	MOVER Wilson score*	medium/large
	Wald*	large
CIs for odds ratio	Transformed Wilson score*	all
	Trans. Clopper-Pearson mid-P	all
	Transformed Blaker exact	small/medium
	Wald*	large

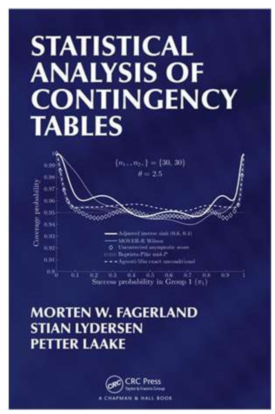
*These methods have closed-form expression
†Preferably with the Berger and Boos procedure ($\gamma = 0.0001$)

!

Summary

- Many different methods are available:
  - Wald, Score, Likelihood ratio, …
  - asymptotic, exact conditional, exact mid p, exact unconditional, …
- Criteria for choice of confidence interval:
  - Coverage
  - Interval width
- Criteria for choice of test:
  - Actual significance level
  - Power
- A method with good properties in one type of contingency table (f.ex independent 2x2) needs not behave well in another type of table (f.ex paired 2x2)

57



Thank you for your attention!

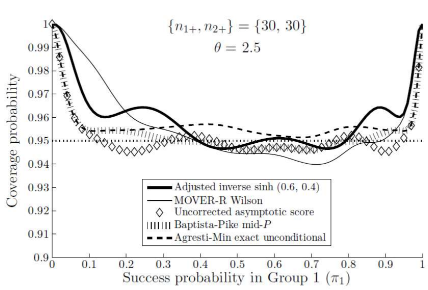


FIGURE 4.42
Coverage probabilities of the best performing intervals for the odds ratio

Fagerland, M. W., Lydersen, S., & Laake, P.: "Statistical Analysis of Contingency Tables." Chapman and Hall/CRC, 2017.

- "This book should be a very useful reference for anyone who wants an overview of the relevant literature (much of it quite recent) or who routinely needs to analyze contingency tables." Alan Agresti.
- "I highly recommended it for masters and doctoral students in statistics … and other fields requiring the analysis of discrete data." Karim F. Hirji
- "I strongly recommend the book both to statisticians and to researchers in health and social disciplines." Robert G. Newcombe
- "… an essential book to own if you analyse low-dimensional contingency tables." John McDonald
- "This book is encyclopaedic in its coverage and would be useful to graduate students and all applied statisticians who are always dealing with contingency tables." Michael J. Campbell

60