Much Ado About Avandia: The meta-analysis of rare events in the service of health policy

Russell Localio

Dept of Biostatistics and Epidemiology

University of Pennsylvania School of Medicine, Philadelphia

John Cornell

Dept of Epidemiology and Biostatistics

University of Texas Health Science Center, San Antonio

Cynthia D. Mulrow

Dept of Medicine

University of Texas Health Science Center, San Antonio

7th International Conference on Health Policy Statistics

Philadelphia, PA January 18, 2008

Avandia (rosiglitazone)--A brief history of safety

(1) August 2004 --Settlement of Spitzer v. Glaxosmithkline (another drug, not Avandia)

GSK agrees to post online the results of its clinical studies (including Avandia)

(2) 2005 -- GSK submits pooled analyses to FDA on risk of Avandia

FDA requires label warning about risk of cardiac events

(3) August 2006 GSK reports to FDA a formal analysis of 42 RCTs and an observational study suggesting a possible 31% increase in risk of adverse cardiovascular events

(4) May 2007 FDA asks for meeting with GSK on risk of cardiovascular events with Avandia

(5) May 2007 – Nissen/Wolski meta analysis appears in N Engl J Med

(6) July 30, 2007 – FDA holds advisory committee meeting focused on issue of myocardial ischemia associated with Avandia

(7) Nissen's study criticized for use of Peto's method (Bracken 2007; Diamond 2007)

(8) Nov 14, 2007 – FDA requires "boxed warning" (FDA's strongest warning) on drug label

(9) Search for “victims”

“Did you or a loved one take Avandia and suffer a heart attack or another adverse cardiovascular event? If so, you have legal rights and are urged to contact us as soon as possible for a FREE CASE REVIEW. We take no legal fees unless compensation is won for you and/or your family.”

http://www.avandialegalrights.com/

Overview of Nissen's 42 trials (using only single outcome of myocardial infarction)

Total numbers of subjects	=	27,500+
Overall event rate among controls	=	0.006
Balance (nc/nt)	=	0.77 (0.24 to 1.98) [mean (min to max)]
Trials with 0 events in one arm	=	30/42(71%)
Trials with 0 events in both arms	=	4/42 (9.5%) ("total zero" trials)
Largest trials size	=	5269
Smallest trial size	=	77

Comparision arm: Placebo, other drugs, combinations

Sources:	(1)	Original FDA submission	5
	(2)	GlaxoSmithKline registry	35
	(3)	DREAM/ADOPT	2

Methods used: Peto

Software: Comprehensive Meta-Analysis v 2.2 (Biostat, Englewood NJ)

Statistical References: Bradburn (2007); Sweeting (2004); Sutton (2002)

Potential statistical “solutions” to the rare events problems:

	Odds Ratio	Risk Differences
Advantages	Good properties of variance estimates for MH odds ratios Ample software for frequentist and Bayesian analyses	All trials are included No need for continuity corrections
Disadvantages	Studies with total zeros are excluded (should they be?)	Effect of treatment more likely to be multiplicative than additive

Simulations on performance of typical methods (Bradburn, Deeks, Berlin, Localio 2007)

Feature	Deeks	Avandia
Baseline risk	0.005	0.006
N trials	19	42
Effect size (OR)	1.33	1.3 to 1.4

Estimate	Method	Bias (% points)
OR	Peto	+1
OR	MH	+1
OR	MH (0.5)	-2
OR	Exact	+1
OR	D&L	-2
RD	MH	1
RD	D&L	-6

D&L= “random effects” DerSimonian and Laird;

MH ( )=Mantel Haenszel (continuity correction)

Exact =StatXact stratified 2x2 tables

Some observations of Deeks et al (2007):

(1) Peto's method works for balanced design, when few trials have zero total events, and event rates are low (<1%)

(2) MH OR and exact methods are preferable when imbalance is present and events rates of 5% to 10%

(3) RD methods use all trials (advantage), but

   a. Very conservative confidence intervals

   b. Additive (risk difference) model might not reflect actual biological effect

(4) D&L method less satisfactory because relies on trial's variance

(5) Pooled 2*2 table (marginal) estimates should be avoided

(6) Standard 0.5 continuity corrections should be replaced by alternatives outlined in Sweeting (2004).

Odds Ratios: Effect of Avandia on risk of myocardial infarction (n=42 trials)

Method	Odds Ratio	Confidence Interval	p-value	Source
Peto	1.43	1.03 to 1.98	0.03	Nissen (2007)
MH fixed (0.5)	1.30	0.96 to 1.75	0.090	Metan (stata)
MH fixed (0.0)	1.45	1.05 to 2.01	0.025	Metan (stata)
Random (D&L) (0.5)	1.31	0.95 to 1.79	0.095	Metan (stata)
Random (D&L) (0.0)	1.31	0.91 to 1.89	0.141	Metan(stata)
Conditional logistic	1.45	1.05 to 2.01	0.025	Clogit (stata)
Exact stratified	1.45	1.03 to 2.04	0.030	StatXact
MH fixed (S,0.01)*	1.45	1.05 to 2.01	0.026	Metabin (R)
Random (D & L) (S,0.01)	1.33	0.93 to 1.91	0.123	Metabin(R)
Random intercept/slope	1.37	0.99 to 1.90	0.059	Xtmelogit
Bayesian (S,0.01) (Warn 2002)	1.46	1.02 to 2.21		Winbugs

*Sweeting, opposite treatment arm correction with total=0.01, to all 42 studies

Comments (and questions) on odds ratios

(1) Software produces variable results on the same data

Some give consistent results

Others give markedly different p-values

(2) Continuity corrections and their sizes matter – 0.5 correction → attenuated ORs

(3) Random effects analyses produce lower and less significant effects (Shuster 2007)

(4) Bayesian analysis gives result similar to exact and fixed effects analyses

Risk Differences: Effect of Avandia on risk of myocardial infarction (n=42 trials)

Method (correction)	RD per 10,000	Confidence Interval per 10,000	p-value	Source
MH fixed (0.0 or 0.01)	21.7	2.5 to 40.9	0.026	Stata metan
MH fixed	17	-4 to 28	0.114	MIX
MH fixed (S, 0.01)	22	4 to 41	0.027	Metabin R*
MH random	18.0	7.3 to 28.7	<0.001	Stata metan
MH random	10	-7 to 26	0.243	MIX
MH random(S, 0.01)	3	-2 to 7	0.244	Metabin R*
Exact	Not reported	- 3.9 to 47.6	0.215	Tian, Cai, Wei (2007) R
Permutation of conditional linear regression (1000 iters)	21.7	7.3 to to 43.8†	0.023	Stata

*Required continuity correction (using Sweeting's method of opposite treatment arm

†Robbins-Monro confidence bounds (Garthwaite 1996) programmed in Stata

Comments (and questions) on Risk Differences (RD):

(1) Choice of software matters. Why do p-values on same data differ so markedly for RD?

(2) When should we rely on RD (rather than OR or RR) as the metric for combining results?

(3) Will simulations of competing/alternative methods on the same datasets resolve these questions?

(4) When are permutation-test-based p-values (and confidence intervals) appropriate?

(5) Bayesian methods needs work on issues of nonconvergence (Warn 2002)

Practical problems with software

(1) Continuity corrections

Default sometimes not stated

Default is usually 0.5 (too large for rare outcomes)

Can be difficult to override and eliminate

Sweeting's alternative treatment arm method difficult to implement in standard software

Choice of continuity correction might be essential with rare outcomes

(2) Displaying results – rare outcomes need more significant digits for RD displays

Metan (Bradburn et al) – Current program allows for only f7.3 (123.567)

Wbstats(Thompson) -- Same problem with only 3 digits to right of decimal

MIX – Choice of output precision is limited

(3) How many digits needed?

Event rate of 5.7 per 1000 for MI for Avandia in 42 trials

Differences are on the order of 2 per 1000 persons

Observational studies as alternative to meta-analyses of RCTs

RCTs

Strengths	**Weaknesses**
Randomization	Loss to follow up
Control of unobserved confounders	Ascertainment bias (adverse outcomes)
Protocols for dose, duration, and follow up	ITT analyses might be suboptimal
Aggressive follow up	Insufficient sample sizes for comparing harm

Observational Studies

Strengths	Weaknesses
Large datasets representing populations	Confounding by indication
Rapidly available data (potentially)	Incomplete reporting of adverse outcomes
Follow up can be nearly 100% in closed system	Loss to follow up if data system < universal
Sufficient power to detect differences among subgroups or by subsets of composite outcomes	Complexity/inadequacy in controlling for confounding by observed covariates
	Inability to control for unobserved covariates
	Instrumental variable methods are problematic

Observational studies: outcome of myocardial infarction

Method	**Comparison (n events)**	**HR/ IRR**	**Confidence Interval**	**p-value**	**Source**
Cohort (Cox)	Avandia v Actos (375)	1.28	1.04 to 1.59		Gerrits (2007)
Cohort – propensity score matched (Cox)	Avandia vs others (323)	0.92	0.73 to 1.16		McAfee (2007)
Case-control	Avandia vs others (200)	1.76	1.27 to 2.44	<0.001	Lipscombe (2007)
Cohort	Avandia vs others (1078)	1.09	0.98 to 1.21	0.117	Wellpoint Rosen (2007)

Interim Analysis

Method	**Comparison (n events)**	**HR/IRR**	**Confidence Interval**	**p-value**	**Source**
Cohort (Cox)	Avandia (80)	1.16	0.75 to 1.81	0.50	Home (2007)

Practical challenges with choice of studies

(1) Combining apples and oranges (choice of studies to meta analyze) can lead to

(a) Misleading conclusions

(b) Violation of the principles of consistency of effect (Feinstein 2002)

(2) Heterogeneity becomes a tradeoff

Combining diverse studies improves power and precision

Stratifying analyses uncovers high-risk subclasses

Effect modification of safety outcomes should mirror same concerns as are applied to interaction of efficacy/effectiveness outcomes

But evidence of heterogeneity limited, and sample size requirements are huge

Conclusions:

(1) Avandia studies = “Scarce and fragile” data on adverse events (Mulrow 2007)

Nissen’s use of Peto might have been suboptimal, but

--standard 0.5 continuity corrections (Bracken 2007) and

--Bayesian meta-analysis” (Diamond 2007)

were probably no better and were perhaps worse.

(2) Statistical science and software lag behind clinical needs

Methods give different quantitative and qualitative results on same data

Continuity corrections options and implementations lag

More study on risk difference measures (absolute rather than relative effects)

(3) New methods should be tested via simulations across many scenarios

References:

Bax L, Yu LM, Ikeda N, Tsuruta H, Moons KGM: Development and validation of MIX: comprehensive free software for meta-analysis of causal research data. BMC Medical Research Methodology 2006, 6(50).

Bracken MB, Rosiglitizone and cardiovascular risk (letter). N Engl J Med. 2007. 357:937-938.

Bradburn MJ, Deeks JJ, Berlin JA, Localio AR. Much ado about nothing: a comparison of the performance of meta-analytic methods with rare events. Statist Med. 2007;26:53-77.

Couzin J. Heart attack risk overshadows Science. 2007:316:1550-1551.

Diamond GA, Kaul S. Rosiglitizone and cardiovascular risk (letter). N Engl J Med. 2007;357:938-939.

Drazen JM. Rosiglitazone – continued uncertainty about safety. (editorial). N Engl J Med. 2007;357:63-64

Feinstein AR. Principles of Medical Statistics. Boca Raton, FL: Chapman &Hall/CRC; 2002: Section 25.4.2.2.

Friedrich JO, Adhikari NKJ, Beyene J. Inclusion of zero total event trials in meta-anlaysis maintains analytic consistency and incorporates all available data. BMC Medical Research Methodology. 2007:7:5.

Garthwaite PH. Confidence intervals from randomization tests. Biometrics. 1996;52:1387-93.

Gerrits CM, Bhattacharya M, Manthena S, Baran R, Perez A, Kupfer S. A comparison of pioglitazone and rosiglitazone for hospitalization for acute myocardial infarction in type 2 diabetes. Pharmacoepidemiology and Drug Safety. (2007)

Home PD, Pocock SJ, Beck-Nielsen H, Gomis R, Hanefeld M, Jones NP, Komajda M, McMurray JJV. Rosiglitzone evaluated for cardiovascular outcomes – An interim analysis. N Engl J Med. 2007;357:28-38.

Lipscombe LL, Gomes T, Levesque LE, Hux JE, Juurlink DN, Alter DA. Thiazolidinediones and cardiovascular outcomes in older patients with diabetes. JAMA. 2007;298:2634-2643.

McAfee AT, Koro C, Landon J, Ziyadeh N, Walker AM. Coronary heart disease outcomes in patients receiving antidiabetic agents. Pharmacoepidemiology and Drug Safety. 207;16:711-25.

Mulrow CD, Cornell JE, Localio AR. Rosiglitizone: a thunderstorm from scarce and fragile data. Ann Intern Med. 2007;147:585-587.

Nathan DM. Rosiglitione and cardiotoxicity – weighing the evidence. N Engl J Med. 2007;357:64-66.

Nissen SE, Wolski K. Effect of Rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. N Engl J Med. 2007;356:2457-2471.

Psaty BM, Furberg CD. The record on rosiglitazone and the risk of myocardial infarction. (editorial). N Engl J Med. 2007;67-69.

Rosen CJ. The rosiglitazone story – Lessons from an FDA advisory committee meeting. N Engl J Med. 2007:1-3.

Shuster JJ, Jones LS, Salmon DA. Fixed vs random effects meta-analysis in rare event studies: The Rosiglitazone link with myocardial infarction and cardiac death. Stat. Med. 2007;26:4375-4385.

Sweeting MJ, Sutton AJ, Lambert PC. What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data. Stat Med. 2004;23:1351-1375.

Thompson JR, Palmer TM, Moreno, S. Bayesian analysis in Stata using WinBUGS. *The Stata Journal* 2006;6(4):530-549. (http://www2.le.ac.uk/departments/health-sciences/extranet/BGE/genetic-epidemiology/gedownload/information/)

Tian L, Cia T, Piankov N, Cremieux P-Y, Wei LJ. Effectively combining independent 2x2 tables for valid inferences in meta analysis with all available data but no artificial continuity corrections for studies with zero events and its application to the analysis of Rosigliatazone's cardiovascular disease related event data. Boston: Harvard University Working Paper Series (No. 69); 2007. http://www.bepress.com/harvardbiostat/paper69

Warn DE, Thompson SG, Spiegelhalter DJ. Bayesian random effects meta-analysis of trials with binary outcomes: methods for absolute risk difference and relative risk scales. Statist Med. 2002;21:1601-1623.