

Quantitative Bias Analysis

Fiona Callaghan, MA MS PhD,
Amgen Inc., South San Francisco, CA

Disclosure: Opinions/statements in this course reflect personal opinions, not policy or opinions of Amgen, Inc. or FDA.

11 February 2016

What is a quantitative bias analysis?

What is it and why do it?

Simple selection bias

Selection bias due to differential initial participation

FDA Sentinel QBA methods

Misclassification bias

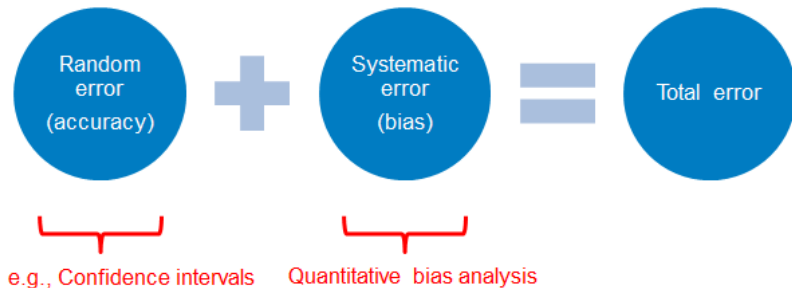
Section 1

What is a quantitative bias analysis?

What is quantitative bias analysis (QBA)?

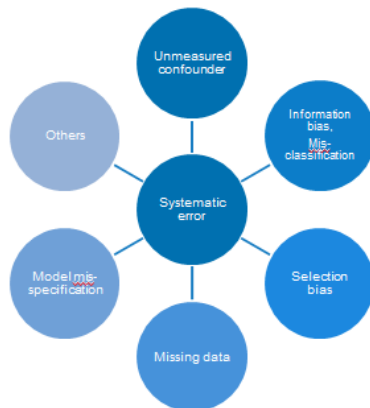
- Make some bias assumptions and calculate bias-adjusted estimates
- Why do QBA?
 - Bias is usually addressed in the 'limitations' section qualitatively
 - With a few assumptions, we can actually show that bias does or does not affect the final results quantitatively.

Types of error



- Systematic error=bias

Types of QBA problems



The QBA book: website

<https://sites.google.com/site/biasanalysis/>

bias.analysis

Navigation

- Applying Quantitative Bias Analysis to Epidemiologic Data
- Errata
- Collaboration
- Multiple Bias Model (Lash TL, Fink A)
- sasmacro SAS Macro (Fox MP, Lash TL, Greenland S)
- Links
- Sitemap

Use the links above to navigate to different pages on our site. Questions, comments, corrections, email bias.analysis@gmail.com

Applying Quantitative Bias Analysis to Epidemiologic Data

Use this page to download the accompanying spreadsheets and SAS code (see bottom of page) for the book:

Lash TL, Fox MP, Fink AK. [Applying Quantitative Bias Analysis to Epidemiologic Data](#). Springer. 2009.

You can find reviews of the textbook in the [American Journal of Epidemiology](#), [JASA](#), [JBSS](#) and [Biometrics](#).

This book collects and synthesizes methods for quantifying systematic errors that affect observational epidemiologic research.

This text provides the first-ever compilation of bias analysis methods for use with epidemiologic data. It guides the reader through the planning stages of bias analysis, including the design of validation studies and the collection of validity data from other sources. Three chapters present methods for corrections to address selection bias, uncontrolled confounding, and classification errors. Subsequent chapters extend these methods to multidimensional bias analysis, probabilistic bias analysis, and multiple bias analysis. The text concludes with a chapter on presentation and interpretation of bias analysis results.

Although techniques for bias analysis have been available for decades, these methods are considered difficult to implement. This text not only gathers the methods into one cohesive and organized presentation, it also explains the methods in a consistent fashion and provides customizable spreadsheets to implement the solutions. By downloading the spreadsheets (available at links provided in the text), readers can follow the examples in the text and then modify the spreadsheet to complete their own bias analyses. Readers without experience using quantitative bias analysis will be able to design, implement, and understand bias analyses that address the major threats to the validity of epidemiologic research. More experienced analysts will value the compilation of bias analysis methods and links to software tools that facilitate their projects.

Timothy L. Lash is a Professor of Epidemiology at Emory and Matthew P. Fox is an Associate Professor in the Center for Global Health & Development at the Boston University School of Public Health. Aliza K. Fink is a Project Manager at Macro International in Bethesda, Maryland. Together they have organized and presented many day-long workshops on the methods of quantitative bias analysis. In addition, they have collaborated on many papers that developed methods of quantitative bias analysis or used the methods in the data analysis.

Use the links at the bottom of the page to download the spreadsheets and SAS code from each chapter.

You can download a [sample chapter](#) here as well

Copies of the book can be ordered from the [publisher's website](#) or from online publishers like [Amazon](#).

The QBA book: website

- Chapter 1-3 background
- Chapter 4- onwards has specific problems.
- The website has Excel spreadsheets to help with simple bias analysis
- SAS macros to help with more complex bias analysis
- There is also an R package 'episensr' that has software for all the techniques in the book

Section 2

Simple selection bias

Selection bias

- Differential initial participation bias is a kind of selection bias that occurs when both the exposure and the outcome affect the probability of initial participation in a study.
- To calculate the adjusted risk estimate you need:
 1. Observed data (counts from a 2x2 table representing the number of subjects exposed/not exposed and cases/controls: a, b, c, and d)
 2. Outcome (OR or RR)
 3. Selection proportions for each of the four exposure/outcome combinations: S_a, S_b, S_c, S_d .
- For example, S_a is the proportion of all potential participants who were exposed and had the outcome, who actually participated in the study.

Selection bias: example

- Association between mobile phone use and uveal melanoma (Stang et al, 2009)
- Cases were identified during September 2002 to September 2004 from a tertiary care hospital in Germany
- 94% of cases and 55% of controls agreed to participate.
- Some non-participants agreed to answer a brief survey.

Selection bias: example

	Exposed Mobile phone +	Unexposed Mobile phone -
Uveal melanoma + /Cases	a=136	c=107
Uveal melanoma - /Controls	b=297	d=165
Total	433	272

$$\begin{aligned}
 \text{Observed OR} &= \frac{a \times d}{b \times c} \\
 &= \frac{136 \times 165}{297 \times 107} \\
 &= 0.71
 \end{aligned}$$

Selection bias: example

- Evidence of selection bias
- Based on the survey of non-participants, and non-participation rates, researchers estimated participation rates in each of the four exposure/outcome categories.

	Exposed Mobile phone +	Unexposed Mobile phone -
Uveal melanoma + /Cases	$S_a=0.94$	$S_b=0.85$
Uveal melanoma - /Controls	$S_c=0.64$	$S_d=0.25$

Selection bias: example

$$\begin{aligned}\text{Adjusted OR} &= \frac{a/S_a \times d/S_d}{b/S_b \times c/S_c} \\ &= \frac{136/0.94 \times 165/0.25}{297/0.85 \times 107/0.64} \\ &= 1.63\end{aligned}$$

Section 3

FDA Sentinel QBA methods

Sentinel QBA

- Working group outlined a test case of QBA methods for mini-sentinel (Sept 2015)



MINI-SENTINEL METHODS

ANALYTICAL METHODS TO ASSESS ROBUSTNESS OF DRUG SAFETY MONITORING RESULTS

Prepared by: Joshua J. Gagne, PharmD, ScD,¹ Shirley V. Wang, PhD,¹ Matthew Fox, PhD,² Timothy Lash, PhD,³ Wesley Eddings, PhD,¹ Sascha Dublin, MD, PhD,⁴ Tobias Gerhard, PhD,⁵ Bruce Psaty, MD, PhD,⁶ Judy Maro, PhD,⁷ Catherine Rogers, MPH,⁷ Rita Ouellet-Hellstrom, PhD, MPH,⁸ Simone Pinheiro, ScD, MSc,⁸ Mark Levenson, PhD,⁹ Richard Forshee,¹⁰ Azadeh Shoaibi, PhD, MS, MHS,¹¹ Sebastian Schneeweiss, MD, ScD¹

http://mini-sentinel.org/work_products/Statistical_Methods/

[Mini-Sentinel_Methods_Analytical-Methods-to-Assess-Robustness-of-Drug-Safety-Monitoring-Results.pdf](#)

Sentinel QBA: Example

- What is the risk of angioedema for new lisinopril users compared to new beta-blocker users?
- No prior use of lisinopril or ACE inhibitors or beta-blockers in prior 183 days for either group
- Single code ICD-9 995.1 recorded any position, IP or OP or ER
- Censoring: discontinuation of medication, death, disenrollment
- Database: large commercial insurer 2008-2013

	Lisinopril	Beta-blockers
Angioedema	664	276
No angioedema	230856	231244
Total	231520	231520

Sentinel QBA: Example

- Initial analyses (not corrected for bias): propensity score matching
- Propensity score adjusted for: Age, prior allergic reaction, diabetes, ischemic heart disease, prescription NSAIDs
- Unadjusted results: $RR=2.4$ (95%CI 2.1-2.8) and 2.3 (95%CI 1.9-2.8) from a Cox PH model

Sentinel QBA: Example, “front-end” sensitivity analyses

Table 2. Results from three “front end” sensitivity analyses varying parameters from primary analysis

Scenario	Description	HR (95% CI)
Reference	Primary analysis	2.32 (1.94 to 2.79)
1	Additional confounding adjustment with empirical variables identified hdPS	2.49 (2.05 to 3.02)
2	Extend new users washout and covariate assessment windows to 365 days	2.62 (2.13 to 3.21)
3	Change in risk window to first 30 days following initial exposure	2.98 (2.25 to 3.95)
4	1 and 2 above simultaneously	2.52 (2.06 to 3.14)
5	1 and 3 above simultaneously	3.50 (2.59 to 4.75)

hdPS, high-dimensional propensity score; HR, hazard ratio; CI, confidence interval

- QBA adjusted values range from 2.5 to 3.5 (unadjusted was 2.3)
- Note: hdPS was a method to identify “likely” confounders

Sentinel QBA: Example, “back-end” sensitivity analyses

Table 3. Results of quantitative bias analysis correcting for outcome misclassification (Reference Risk Ratio, 2.41)

Scenario	Assumed specificity	Assumed sensitivity	Corrected risk ratio
1	0.999777494	0.2	2.73
2	0.999777494	0.4	2.73
3	0.999777494	0.6	2.73
4	0.999777494	0.8	2.73
5	0.999777494	1.0	2.73
6	0.999891798	0.2	2.55
7	0.999891798	0.4	2.55
8	0.999891798	0.6	2.55
9	0.999891798	0.8	2.55
10	0.999891798	1.0	2.55
11	1.0	0.8	2.41
12	0.99	1.0	0.81

- Doesn't matter what the sensitivity is, only specificity is important
- RR generally underestimated without adjustment

Section 4

Misclassification bias

Misclassification bias

- An example of more sophisticated techniques.
- Misclassification bias occurs when some exposed subjects are labeled non-exposed and vice versa, and the misclassification is related to the outcome.
- SIMEX (simulation and extrapolation) method adjusts the odds ratio (OR) by resampling according to the estimated sensitivity and/or specificity

Cook JR, Stefanski LA (1994). Simulation-extrapolation estimation in parametric measurement error models.

Journal of American Statistical Association, 89, 1314-28

Misclassification bias: example

- In heart failure patients, what is the risk of 30-day hospitalization for patients with stroke compared to patients without stroke at baseline?
- One algorithm to identify stroke in claims data has a validated sensitivity = 82% and specificity = 93%

Tirschwell, D. L. and W. T. Longstreth (2002). "Validating Administrative Data in Stroke Research." *Stroke* 33(10): 2465-70.

Misclassification bias: example

- Install library
- Read in data
- Look at first 5 rows:

```
> library(simex)
> hf = read.table(file="Y:/hfdata.2011.2012.csv",sep=",", header=T)
> hf[1:5,]
  hf.ENROLID hf.comorb.STRK hf.Hosp30day.allcause
1 875931501      No stroke                1
2 2060826701      No stroke                0
3 27457582701      No stroke                0
4 28419528901      No stroke                0
5 28096163501      Stroke                 0
```

Misclassification bias: example

- Build logistic regression model (“naive” /not adjusted for bias)
- $OR = \exp(0.385) = 1.5$

```
> nave = glm(Hosp30day.allcause~ comorb.STRK,family= binomial,data =hf, x=TRUE, y=TRUE)
> summary(naive)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-2.530062	0.005359	-472.14	<2e-16 ***
comorb.STRKStroke	0.385192	0.018181	21.19	<2e-16 ***

Misclassification bias: example

- Enter sensitivity (82%) and specificity (93%)

```
> mc = matrix(c(0.93,0.07,0.18,0.82),nrow=2)
```

```
> mc
```

	No stroke	Stroke
No stroke	0.93	0.18
Stroke	0.07	0.82

Misclassification bias: example

- Use SIMEX to correct for bias
- Corrected $OR = \exp(0.71) = 2.0$

```
> mod.strk = mcsimex(naive, mc.matrix = mc, SIMEXvariable = comorb.STRK, B=100)
> mod.strk
```

Jackknife variance:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-2.536858	0.005499	-461.36	<2e-16 ***
comorb.STRKStroke	0.712671	0.027697	25.73	<2e-16 ***