

Bayesian Adjustments for Misclassified Data

Lawrence Joseph

Bayesian Adjustments for Misclassified Data

Lawrence Joseph

Marcel Behr, Patrick Bélisle, Sasha Bernatsky, Nandini Dendukuri, Theresa Gyorkos, Martin Ladouceur, Elham Rahme, Kevin Schwartzman, Allison Scott

Bayesian Adjustments for Misclassified Data

Lawrence Joseph

Marcel Behr, Patrick Bélisle, Sasha Bernatsky, Nandini Dendukuri, Theresa Gyorkos, Martin Ladouceur, Elham Rahme, Kevin Schwartzman, Allison Scott

Department of Epidemiology
and Biostatistics

McGill University

Outline

Outline

- Statistical “trends” in medical journals

Outline

- Statistical “trends” in medical journals
- Misclassification in analysis of diagnostic testing data

Outline

- Statistical “trends” in medical journals
- Misclassification in analysis of diagnostic testing data
- User friendly software

Outline

- Statistical “trends” in medical journals
- Misclassification in analysis of diagnostic testing data
- User friendly software
- Misclassification in administrative database studies

Outline

- Statistical “trends” in medical journals
- Misclassification in analysis of diagnostic testing data
- User friendly software
- Misclassification in administrative database studies
- Extension to more complex situations (correlations, continuous data)

Outline

- Statistical “trends” in medical journals
- Misclassification in analysis of diagnostic testing data
- User friendly software
- Misclassification in administrative database studies
- Extension to more complex situations (correlations, continuous data)
- Conclusion

Trends in Medical Journals

Trends in Medical Journals

- “Recognition of a good idea”
 - Optional, but if you analyze data this way, reviewers generally will say it is a good idea

Trends in Medical Journals

- “Recognition of a good idea”
 - Optional, but if you analyze data this way, reviewers generally will say it is a good idea

- “Commonplace”
 - If you do not analyze data this way, reviewers generally will ask you to do it

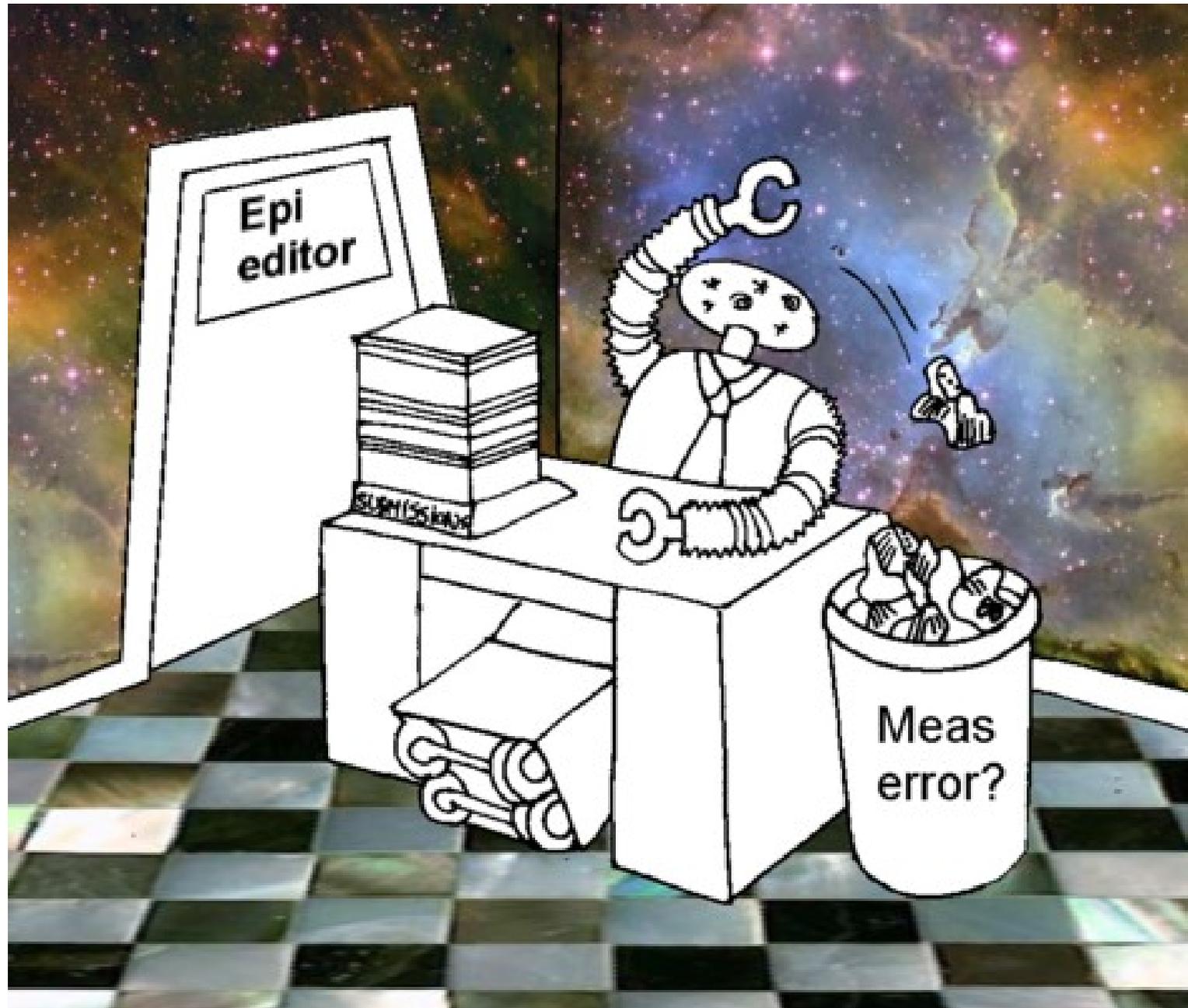
Trends in Medical Journals

- “Recognition of a good idea”
 - Optional, but if you analyze data this way, reviewers generally will say it is a good idea
- “Commonplace”
 - If you do not analyze data this way, reviewers generally will ask you to do it
- What is considered as a “Good Idea” or is “Commonplace” changes over time.

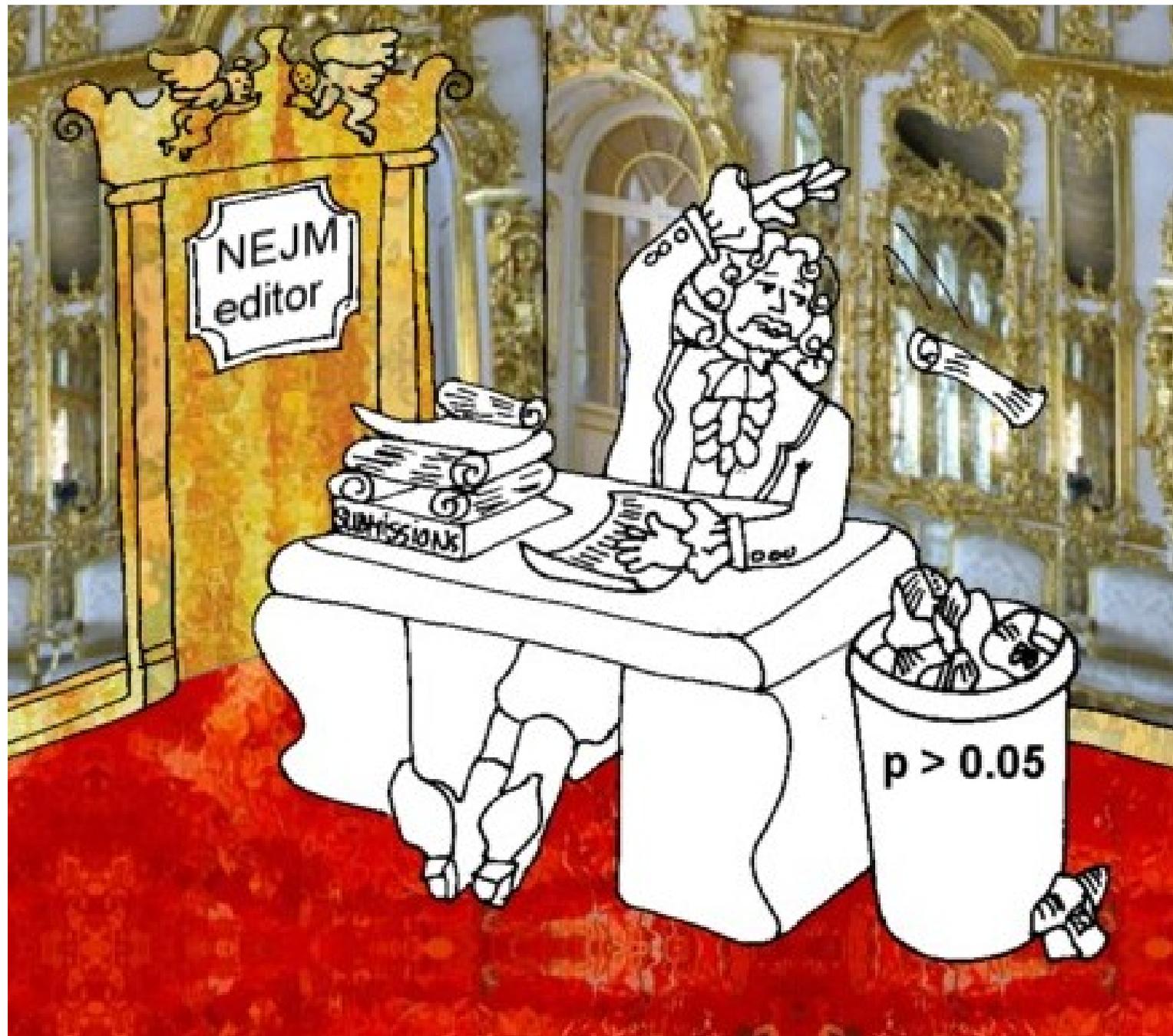




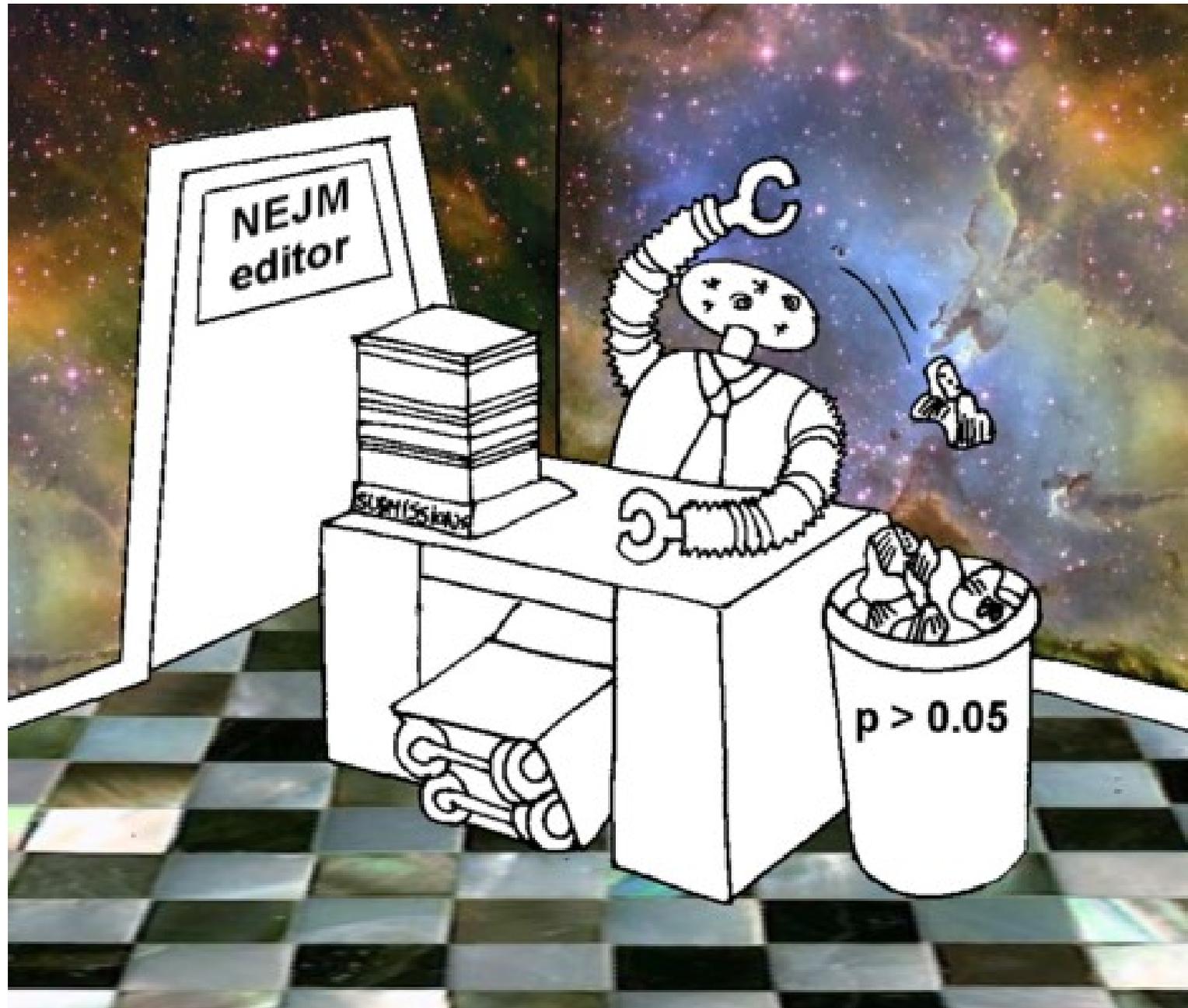












Current Trends in Medical Journals (2007)

Current Trends in Medical Journals (2007)

1. Bayesian analysis is “recognized as a good idea” but not required

Current Trends in Medical Journals (2007)

1. Bayesian analysis is “recognized as a good idea” but not required
2. Adjustment for measurement error is also “recognized as a good idea” but not required

Current Trends in Medical Journals (2007)

1. Bayesian analysis is “recognized as a good idea” but not required
2. Adjustment for measurement error is also “recognized as a good idea” but not required
3. Non-identifiability issues often means that $2 \Rightarrow 1$

Current Trends in Medical Journals (2007)

1. Bayesian analysis is “recognized as a good idea” but not required
2. Adjustment for measurement error is also “recognized as a good idea” but not required
3. Non-identifiability issues often means that $2 \Rightarrow 1$
4. Both ideas are increasing in popularity

Articles encouraging use of Bayesian methods in medicine

Articles encouraging use of Bayesian methods in medicine

- Malakoff D. **Bayes offers a new way to make sense of numbers.** Science 1999;286:1460-1464. [Science review about the rise of Bayesian analysis]

Articles encouraging use of Bayesian methods in medicine

- Malakoff D. **Bayes offers a new way to make sense of numbers.** Science 1999;286:1460-1464. [Science review about the rise of Bayesian analysis]
- Dunson D. **Practical advantages of Bayesian analysis of epidemiologic data.** American Journal of Epidemiology 2001;153:1222-1226.

Articles encouraging use of Bayesian methods in medicine

- Malakoff D. **Bayes offers a new way to make sense of numbers.** Science 1999;286:1460-1464. [Science review about the rise of Bayesian analysis]
- Dunson D. **Practical advantages of Bayesian analysis of epidemiologic data.** American Journal of Epidemiology 2001;153:1222-1226.
- Goodman S. **Of P-values and Bayes: A modest proposal.** Epidemiology. 2001;12:295-7. [Encourages use of Bayes Factors rather than p -values]

Articles encouraging use of Bayesian methods in medicine

- Malakoff D. **Bayes offers a new way to make sense of numbers.** Science 1999;286:1460-1464. [Science review about the rise of Bayesian analysis]
- Dunson D. **Practical advantages of Bayesian analysis of epidemiologic data.** American Journal of Epidemiology 2001;153:1222-1226.
- Goodman S. **Of P-values and Bayes: A modest proposal.** Epidemiology. 2001;12:295-7. [Encourages use of Bayes Factors rather than p -values]
- Greenland S. **Multiple-bias modelling for analysis of observational data.** JRSSA 2005;168:267-306. [Encourages use of Bayesian methods for bias adjustments and measurement error/misclassification]

Quote from Greenland 2005

“Conventional analytic results do not reflect any source of uncertainty other than random error, and as a result readers must rely on informal judgments regarding the effect of possible biases. When standard errors are small these judgments often **fail to capture sources of uncertainty** and their interactions adequately. Multiple-bias models provide alternatives Typically, the **bias parameters** in the model are **not identified by the analysis data** and so the results depend completely on priors for those parameters. **A Bayesian analysis is then natural . . .**”

Another quote from Greenland 2005

“Conventional analyses can be characterized as:

Another quote from Greenland 2005

“Conventional analyses can be characterized as:

(a) **Employ frequentist statistical methods based on assumptions which may be grossly violated and are not testable with the data under analysis:**

Another quote from Greenland 2005

“Conventional analyses can be characterized as:

(a) **Employ frequentist statistical methods based on assumptions which may be grossly violated** and are not testable with the data under analysis:

(i) the study exposure is randomized within levels of controlled covariates

Another quote from Greenland 2005

“Conventional analyses can be characterized as:

(a) **Employ frequentist statistical methods based on assumptions which may be grossly violated** and are not testable with the data under analysis:

(i) the study exposure is randomized within levels of controlled covariates

(ii) selection, participation and missing data are random

Another quote from Greenland 2005

“Conventional analyses can be characterized as:

(a) **Employ frequentist statistical methods based on assumptions which may be grossly violated** and are not testable with the data under analysis:

(i) the study exposure is randomized within levels of controlled covariates

(ii) selection, participation and missing data are random

(iii) **there is no measurement error**

Another quote from Greenland 2005

“Conventional analyses can be characterized as:

(a) **Employ frequentist statistical methods based on assumptions which may be grossly violated** and are not testable with the data under analysis:

(i) the study exposure is randomized within levels of controlled covariates

(ii) selection, participation and missing data are random

(iii) **there is no measurement error**

(b) **Address possible violations of assumptions with speculative discussions.** If they like the results, researchers argue that the biases are inconsequential. If they dislike the results they focus on possible biases.”

Example: Diagnostic testing for Strongyloides infection

Stool Examination

		+	-	
Serology	+	38	87	125
	-	2	35	37
		40	122	162

Example: Diagnostic testing for Strongyloides infection

Stool Examination

		+	-	
Serology	+	38	87	125
	-	2	35	37
		40	122	162

- Prevalence $\approx 25\%$? Prevalence $\approx 75\%$?

Example: Diagnostic testing for Strongyloides infection

Stool Examination

		+	-	
Serology	+	38	87	125
	-	2	35	37
		40	122	162

- Prevalence $\approx 25\%$? Prevalence $\approx 75\%$?
- Sensitivity? Specificity?

Problem: Non-identifiability

Problem: Non-identifiability

- Table has 3 degrees of freedom

Problem: Non-identifiability

- Table has 3 degrees of freedom
- There are five unknown parameters (prev + sens and spec from each test)

Problem: Non-identifiability

- Table has 3 degrees of freedom
- There are five unknown parameters (prev + sens and spec from each test)
- Thus we have non-identifiability: There is an infinite number of solutions (estimates of the five parameters) that fit the data equally well

Frequentist Solution

Frequentist Solution

- With 3 df, can only estimate 3 parameters at a time

Frequentist Solution

- With 3 df, can only estimate 3 parameters at a time
- Solution: Pick any two parameters, and fix their values, use data to estimate other three

Frequentist Solution

- With 3 df, can only estimate 3 parameters at a time
- Solution: Pick any two parameters, and fix their values, use data to estimate other three
- Problems:

Frequentist Solution

- With 3 df, can only estimate 3 parameters at a time
- Solution: Pick any two parameters, and fix their values, use data to estimate other three
- Problems:
 - Which 2 to pick as “known”?

Frequentist Solution

- With 3 df, can only estimate 3 parameters at a time
- Solution: Pick any two parameters, and fix their values, use data to estimate other three
- Problems:
 - Which 2 to pick as “known”?
 - What if “known” values are inaccurate?

Frequentist Solution

- With 3 df, can only estimate 3 parameters at a time
- Solution: Pick any two parameters, and fix their values, use data to estimate other three
- Problems:
 - Which 2 to pick as “known”?
 - What if “known” values are inaccurate?
 - Even if exactly correct values selected for “known” parameters, confidence intervals are too narrow, as uncertainty in constrained values ignored.

Bayesian Solution

Bayesian Solution

- Treat all five parameters as equal

Bayesian Solution

- Treat all five parameters as equal
- Place a prior distribution over each parameter

Bayesian Solution

- Treat all five parameters as equal
- Place a prior distribution over each parameter
- At least two priors must be “informative”, to get around identifiability problem

Bayesian Solution

- Treat all five parameters as equal
- Place a prior distribution over each parameter
- At least two priors must be “informative”, to get around identifiability problem
- Have priors, can write down likelihood, get posteriors for all parameters via Bayes Theorem

Bayesian Solution

- Treat all five parameters as equal
- Place a prior distribution over each parameter
- At least two priors must be “informative”, to get around identifiability problem
- Have priors, can write down likelihood, get posteriors for all parameters via Bayes Theorem
- Contains the frequentist solution as a special case, with point priors on two parameters and uniform priors on other three parameters . . . very unrealistic solution when looked at in this way

Bayesian Solution - Need For User Friendly Software

Bayesian Solution - Need For User Friendly Software

- Thousands of articles discussing new statistical methods are published every year

Bayesian Solution - Need For User Friendly Software

- Thousands of articles discussing new statistical methods are published every year
- Only a very small percentage of these find use in real applications

Bayesian Solution - Need For User Friendly Software

- Thousands of articles discussing new statistical methods are published every year
- Only a very small percentage of these find use in real applications
- An even smaller proportion of new methods find use by researchers other than the developers

Bayesian Solution - Need For User Friendly Software

- Thousands of articles discussing new statistical methods are published every year
- Only a very small percentage of these find use in real applications
- An even smaller proportion of new methods find use by researchers other than the developers
- Reasons:

Bayesian Solution - Need For User Friendly Software

- Thousands of articles discussing new statistical methods are published every year
- Only a very small percentage of these find use in real applications
- An even smaller proportion of new methods find use by researchers other than the developers
- Reasons:
 - Too difficult for most non-statisticians to understand

Bayesian Solution - Need For User Friendly Software

- Thousands of articles discussing new statistical methods are published every year
- Only a very small percentage of these find use in real applications
- An even smaller proportion of new methods find use by researchers other than the developers
- Reasons:
 - Too difficult for most non-statisticians to understand
 - Lack of recognition for applied work (not here!)

Bayesian Solution - Need For User Friendly Software

- Thousands of articles discussing new statistical methods are published every year
- Only a very small percentage of these find use in real applications
- An even smaller proportion of new methods find use by researchers other than the developers
- Reasons:
 - Too difficult for most non-statisticians to understand
 - Lack of recognition for applied work (not here!)
 - Even if understandable, lack of time to program

Bayesian Solution - Need For User Friendly Software

- Thousands of articles discussing new statistical methods are published every year
- Only a very small percentage of these find use in real applications
- An even smaller proportion of new methods find use by researchers other than the developers
- Reasons:
 - Too difficult for most non-statisticians to understand
 - Lack of recognition for applied work (not here!)
 - Even if understandable, lack of time to program
- Useful to provide user friendly software

Example: BayesDiagnosticTests Software

Example: BayesDiagnosticTests Software

- Windows based exe file

Example: BayesDiagnosticTests Software

- Windows based exe file
- Input data + priors through user friendly “fill in the blanks” windows

Example: BayesDiagnosticTests Software

- Windows based exe file
- Input data + priors through user friendly “fill in the blanks” windows
- Runs a Gibbs sampler using WinBUGS by itself

Example: BayesDiagnosticTests Software

- Windows based exe file
- Input data + priors through user friendly “fill in the blanks” windows
- Runs a Gibbs sampler using WinBUGS by itself
- When ready (typically a few minutes) pops up posterior distributions + graphs

Example: BayesDiagnosticTests Software

- Windows based exe file
- Input data + priors through user friendly “fill in the blanks” windows
- Runs a Gibbs sampler using WinBUGS by itself
- When ready (typically a few minutes) pops up posterior distributions + graphs
- Extensive manual, free help via email

Example: BayesDiagnosticTests Software

- Windows based exe file
- Input data + priors through user friendly “fill in the blanks” windows
- Runs a Gibbs sampler using WinBUGS by itself
- When ready (typically a few minutes) pops up posterior distributions + graphs
- Extensive manual, free help via email
- Let’s look at the results from the Strongyloides example

Observed data

Enter the number of patients with each combination of tests' results

		Test 2	
		positive	negative
Test 1	positive	38	87
	negative	2	35

Tests labels

Test 1: Serology
Test 2: Stool

Gibbs sampler specifications >>

Prior information on prevalence

Prevalence _____

Prior beta parameters α β

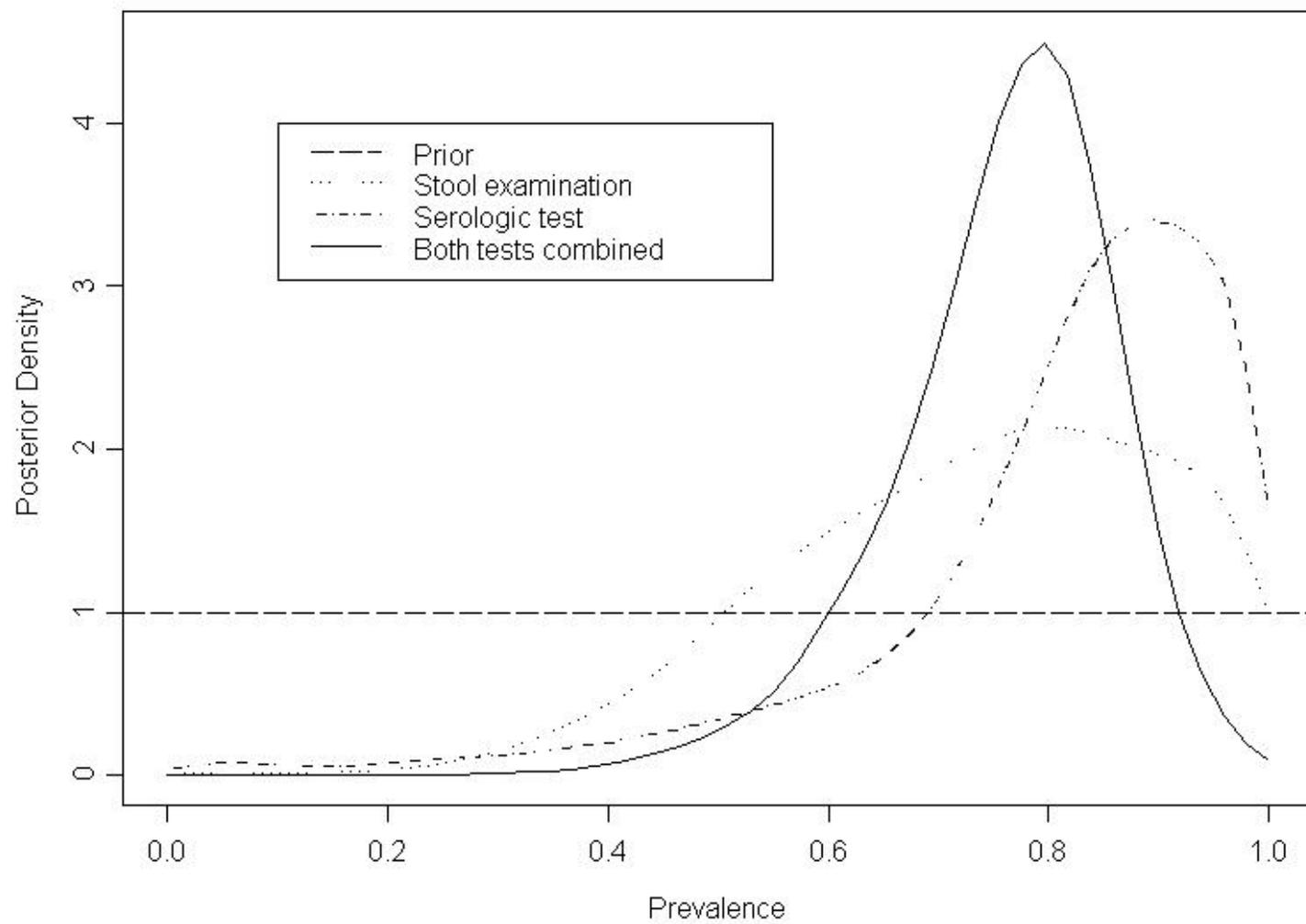
Disease name (optional)

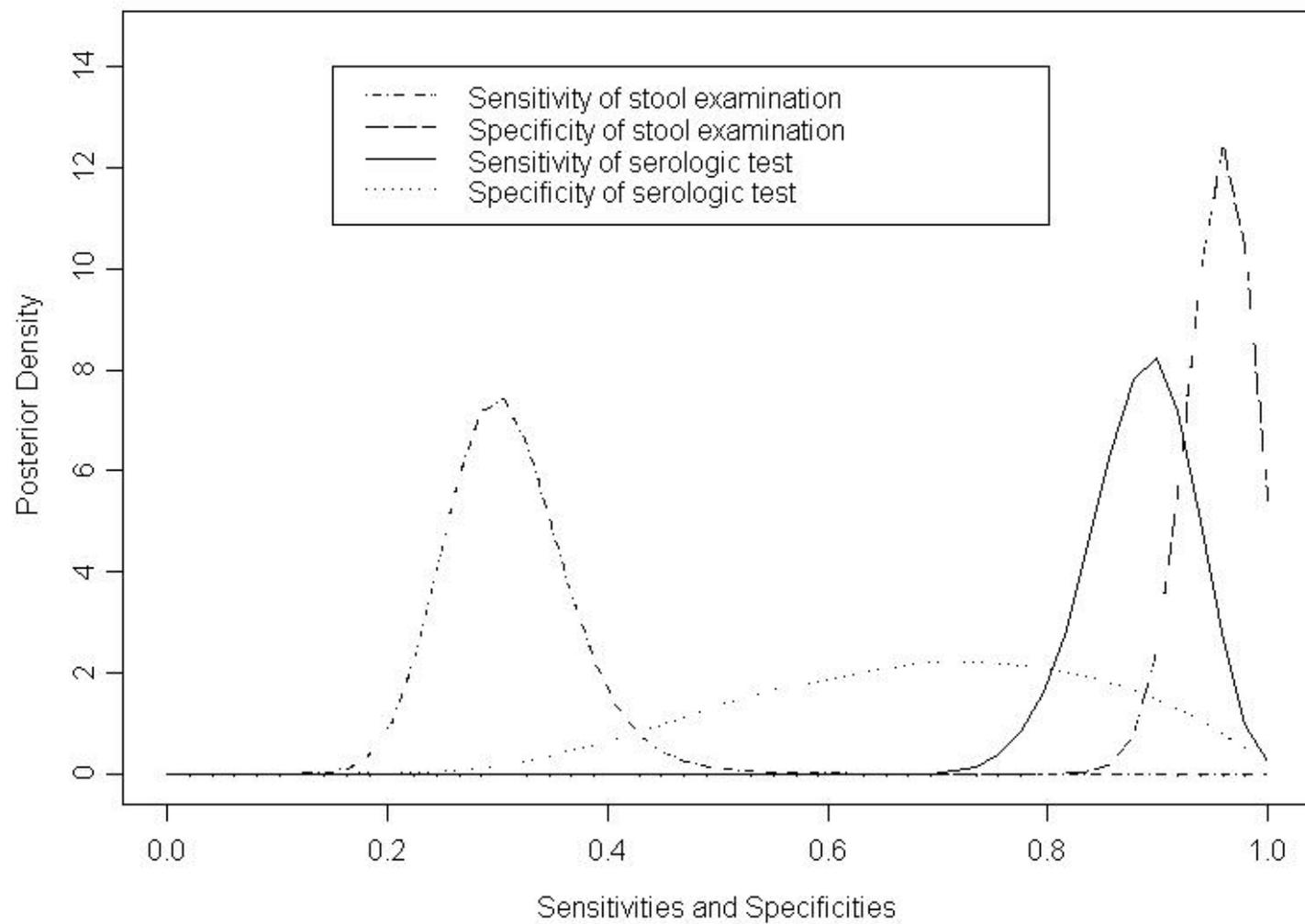
(μ, σ)
 \leftrightarrow
 (α, β)

Next >>

Results: Strongyloides Example

		Stool Examination		Serology	
	Prev	Sens	Spec	Sens	Spec
Prior Information	0.50 0.03 – 0.98	0.24 0.07 – 0.47	0.95 0.89 – 0.99	0.81 0.63 – 0.92	0.72 0.31 – 0.96
Stool Examination Alone	0.74 0.41 – 0.98	0.30 0.21 – 0.47	0.95 0.88 – 0.99		
Serology Alone	0.80 0.23 – 0.99			0.83 0.73 – 0.92	0.58 0.22 – 0.94
Both Tests Combined	0.76 0.52 – 0.91	0.31 0.22 – 0.44	0.96 0.91 – 0.99	0.89 0.80 – 0.95	0.67 0.36 – 0.95





Application to Administrative Database Research

Application to Administrative Database Research

- Primary data collection can be expensive

Application to Administrative Database Research

- Primary data collection can be expensive
- Researchers are increasingly using information collected in administrative databases (e.g. RAMQ)

Application to Administrative Database Research

- Primary data collection can be expensive
- Researchers are increasingly using information collected in administrative databases (e.g. RAMQ)
- Such databases typically contain substantial proportions of misclassification errors (e.g. diagnoses).

Example: Prevalence of OA in 65+ from RAMQ data

Three imperfect clues about OA are available:

Example: Prevalence of OA in 65+ from RAMQ data

Three imperfect clues about OA are available:

- ICD-9 diagnostic code for OA

Example: Prevalence of OA in 65+ from RAMQ data

Three imperfect clues about OA are available:

- ICD-9 diagnostic code for OA
- At least one prescription for acetaminophen or an NSAID, but not methotrexate or plaquenil

Example: Prevalence of OA in 65+ from RAMQ data

Three imperfect clues about OA are available:

- ICD-9 diagnostic code for OA
- At least one prescription for acetaminophen or an NSAID, but not methotrexate or plaquenil
- Received injection common in OA, an arthroplasty or a tibial osteotomy

Methods

Methods

- Similar to case with two tests, but can use data from one, two, or all three tests (7 combinations in all)

Methods

- Similar to case with two tests, but can use data from one, two, or all three tests (7 combinations in all)
- Priors not needed if all three tests are used

Methods

- Similar to case with two tests, but can use data from one, two, or all three tests (7 combinations in all)
- Priors not needed if all three tests are used
- Idea is to compare results from a variety of models, to check robustness of prevalence estimates

OA data from RAMQ data base

Test 1 Phys. Diagnosis	Test 2 Medication	Test 3 Medical Acts	Number of individuals observed
+	+	+	11,816
+	+	-	57,222
+	-	+	3,320
+	-	-	25,651
-	+	+	9,610
-	+	-	260,923
-	-	+	5,002
-	-	-	595,415

OA data from RAMQ data base

Test 1 Phys. Diagnosis	Test 2 Medication	Test 3 Medical Acts	Number of individuals observed
+	+	+	11,816
+	+	-	57,222
+	-	+	3,320
+	-	-	25,651
-	+	+	9,610
-	+	-	260,923
-	-	+	5,002
-	-	-	595,415

- Test 1 +ve = $11,816 + 57,222 + 3,320 + 25,651 = 98,009$
- N = 968,959

Results with no Misclassification Adjustment

Results with no Misclassification Adjustment

- Naive estimate using physician diagnosis error is 10.1%,
95% Credible Interval (CrI) 10.1-10.2

Results with no Misclassification Adjustment

- Naive estimate using physician diagnosis error is 10.1%, 95% Credible Interval (CrI) 10.1-10.2
- Very narrow interval, but accounts only for uncertainty due to random variation, not for extra variability due to misclassification

Results with no Misclassification Adjustment

- Naive estimate using physician diagnosis error is 10.1%, 95% Credible Interval (CrI) 10.1-10.2
- Very narrow interval, but accounts only for uncertainty due to random variation, not for extra variability due to misclassification
- How much confidence can we place in this seemingly very accurate estimate?

Results with Misclassification Adjustment

	Prev	Sens 1	Sens 2	Sens 3	Spec 1	Spec 2	Spec 2
Prior distribution	50.0 0.0-100	75.0 70.0-80.0	75.0 70.0-80.0	25.0 20.0-30.0	95.0 90.0-100	60.0 55.0-65.0	95.0 90.0-100
One Test							
Physician diagnosis alone	11.5 4.5-14.2	72.8 63.2-82.4			95.4 92.8-99.9		
Prescribed medication alone	9.5 3.3-22.0		72.3 62.5-81.9			71.1 66.3-75.6	
Medical procedure alone	10.6 5.2-18.6			22.1 14.7-33.8			99.2 97.9-99.9
Two Tests							
Combination of physician diagnosis and prescribed medication	11.8 8.6-14.8	75.1 68.4-81.4	76.1 73.9-78.4		98.9 98.1-99.1	70.5 70.1-71.3	
Combination of physician diagnosis and medical acts	9.8 6.4-13.7	74.1 63.2-83.2		23.5 15.9-31.1	96.7 94.3-99.6		99.2 98.6-99.3
Combination of prescribed medication and medical acts	10.0 6.8-16.4		77.6 72.2-83.3	24.0 18.2-38.2		70.6 68.2-72.5	99.6 99.3-99.9
Three Tests							
3 tests using non-informative priors	14.8 14.5-15.1	58.2 57.0-59.0	78.3 77.6-79.0	18.2 17.8-18.5	98.1 98.0-98.3	72.4 72.3-72.6	99.5 99.5-99.5

Comparison of Estimates

Comparison of Estimates

- Usual estimate: 10.1% (10.1-10.2), width = 0.1

Comparison of Estimates

- Usual estimate: 10.1% (10.1-10.2), width = 0.1
- Adj estimate (3 tests): 14.8% (14.5-15.1), width = 0.6

Comparison of Estimates

- Usual estimate: 10.1% (10.1-10.2), width = 0.1
- Adj estimate (3 tests): 14.8% (14.5-15.1), width = 0.6
- Bias \approx 50%, CrI width grows by a factor of 6

Comparison of Estimates

- Usual estimate: 10.1% (10.1-10.2), width = 0.1
- Adj estimate (3 tests): 14.8% (14.5-15.1), width = 0.6
- Bias \approx 50%, CrI width grows by a factor of 6
- Other plausible estimates range from 3.3% to 22%

Comparison of Estimates

- Usual estimate: 10.1% (10.1-10.2), width = 0.1
- Adj estimate (3 tests): 14.8% (14.5-15.1), width = 0.6
- Bias \approx 50%, CrI width grows by a factor of 6
- Other plausible estimates range from 3.3% to 22%
- Must admit true answer not really known

Comparison of Estimates

- Usual estimate: 10.1% (10.1-10.2), width = 0.1
- Adj estimate (3 tests): 14.8% (14.5-15.1), width = 0.6
- Bias \approx 50%, CrI width grows by a factor of 6
- Other plausible estimates range from 3.3% to 22%
- Must admit true answer not really known
- All estimates depend on unverifiable assumptions

Comparison of Estimates

- Usual estimate: 10.1% (10.1-10.2), width = 0.1
- Adj estimate (3 tests): 14.8% (14.5-15.1), width = 0.6
- Bias \approx 50%, CrI width grows by a factor of 6
- Other plausible estimates range from 3.3% to 22%
- Must admit true answer not really known
- All estimates depend on unverifiable assumptions
- Problems carry over to regressions based on risk factors for OA, etc.

Extensions

Extensions

- Different numbers of tests

Extensions

- Different numbers of tests
- Correlations among dichotomous tests

Extensions

- Different numbers of tests
- Correlations among dichotomous tests
- Continuous diagnostic test results (Para + Non-Para)

Extensions

- Different numbers of tests
- Correlations among dichotomous tests
- Continuous diagnostic test results (Para + Non-Para)
- Combinations of continuous and dichotomous tests

Extensions

- Different numbers of tests
- Correlations among dichotomous tests
- Continuous diagnostic test results (Para + Non-Para)
- Combinations of continuous and dichotomous tests
- Correlations among continuous tests

Extensions

- Different numbers of tests
- Correlations among dichotomous tests
- Continuous diagnostic test results (Para + Non-Para)
- Combinations of continuous and dichotomous tests
- Correlations among continuous tests
- Hierarchical models for diagnostic test data

Conclusions

- Important to consider measurement/misclassification error in Epi (coming trend?)

Conclusions

- Important to consider measurement/misclassification error in Epi (coming trend?)
- For real impact:
 - not enough to develop methods
 - need user friendly software

Conclusions

- Important to consider measurement/misclassification error in Epi (coming trend?)
- For real impact:
 - not enough to develop methods
 - need user friendly software
- Beware of unadjusted results from Administrative Database research

References

- Joseph L, Gyorkos T, Coupal L. Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. *American Journal of Epidemiology* 1995;141(3):263-272. (**Dichotomous tests**)
- Rahme E, Joseph L, and Gyorkos T. Bayesian sample size determination for estimating binomial parameters from data subject to misclassification. *Applied Statistics* 2000;49(1):119-228. (**Design of studies with misclassification**)
- Dendukuri N, Joseph L. Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests. *Biometrics* 2001;57(1):208-217. (**Dependent tests**)
- Bernatsky S, Joseph L, Bélisle P, Boivin J, Rajan R, Moore A, Clarke A. A Bayesian hierarchical model for estimating the properties of cancer ascertainment methods in cohort studies. *Statistics in Medicine* 2005;24:2365-2379. (**Hierarchical model**)
- Carabin H, Marshall C, Joseph L, Riley S, Olveda R, McGarvey S. Estimating and modelling the dynamics of the intensity of infection with *Schistosoma japonicum* in villagers of Leyte, Philippines. Part I: A Bayesian cumulative logit model. *American Journal of Tropical Medicine and Hygiene* 2005;72(6):745-753. (**Logistic regression with misclassification**)
- Ladouceur M, Rahme E, Pineau C, Joseph L. Robustness of prevalence estimates derived from misclassified data from administrative databases. *Biometrics* 2007;63:272-279. (**Application to Administrative Data**)
- Scott A, Joseph L, Bélisle P, Behr M, Schwartzman K. Bayesian estimation of tuberculosis clustering rates from DNA sequence data. *Statistics in Medicine* 2007 (to appear). (**Continuous tests**)