Totality of Evidence-
Can we make better use of observational data?

Ian Hirsch-Lesley France

EFSPI Statistical Leaders Meeting

8th June 2011

Contents

- **Levels of evidence**
  - *RCT vs non-RCT information*
- **An interesting history**
  - *Smoking vs Lung Cancer*
    - *1939-1950*
    - *The 50 year study*
  - *1982 - The original argument against observational analyses/studies*
  - *2000 - The argument for observational analyses/studies*
- ***So why the interest now?***
- **Discussion: So what now?**

Contents

- **Levels of evidence**
  - *RCT vs non-RCT information*
- **An interesting history**
  - *Smoking vs Lung Cancer*
    - *1939-1950*
    - *The 50 year study*
  - *1982 - The original argument against observational analyses/studies*
  - *2000 - The argument for observational analyses/studies*
- ***So why the interest now?***
- **Discussion: So what now?**

The NEW ENGLAND JOURNAL of MEDICINE

Issue Index

A searchable index of tables of contents

FIND AN ISSUE
By Volume and Issue
Vol. No.
‹ Prev Issue Next Issue ›

TABLE OF CONTENTS FOR
June 22, 2000 Vol. 342 No. 25

SPECIAL ARTICLES

A Comparison of Observational Studies and Randomized, Controlled Trials 1878-1886
K. Benson, A.J. Hartz
Free Full Text

Randomized, Controlled Trials, Observational Studies, and the Hierarchy of Research Designs 1887-1892
J. Concato, N. Shah, R.I. Horwitz
Free Full Text

TABLE 1. GRADES OF EVIDENCE FOR THE PURPORTED QUALITY OF STUDY DESIGN.*

I	Evidence obtained from at least one properly randomized, controlled trial.
II-1	Evidence obtained from well-designed controlled trials without randomization.
II-2	Evidence obtained from well-designed cohort or case–control analytic studies, preferably from more than one center or research group.
II-3	Evidence obtained from multiple time series with or without the intervention. Dramatic results in uncontrolled experiments (such as the results of the introduction of penicillin treatment in the 1940s) could also be regarded as this type of evidence.
III	Opinions of respected authorities, based on clinical experience; descriptive studies and case reports; or reports of expert committees.

*The grades are those of the U.S. Preventive Services Task Force.[7]

Contents

- **Levels of evidence**
  - *RCT vs non-RCT information*
- **An interesting history**
  - *Smoking vs Lung Cancer*
    - *1939-1950*
    - *The 50 year study*
  - *1982 - The original argument against observational analyses/studies*
  - *2000 - The argument for observational analyses/studies*
- ***So why the interest now?***
- **Discussion: So what now?**

1939 to 1948

TOBACCO: A MEDICAL HISTORY

SIR RICHARD DOLL

TABLE IV Smoking and Lung Cancer Case-Control Studies Before 1950

	No. of Men		Nonsmokers, %		Heavy Smokers, %	
Author	Lung Cancer	Controls	Lung Cancer	Controls	Lung Cancer	Controls
Müller[30]	86	86	3.5	16.3	65	36
Schairer and Schöniger[32]	93	270	3.2	15.9	52	27
Wassink[33]	134	100	4.5	19.0	55	19

1 Doll R. Tobacco: a medical history. *J Urban Health* 1999;76:289-313.

German tobacco epidemiology by this time was the most advanced in the world. Franz H Muller in 1939 and Eberhard Schairer and Erich Schoniger in 1943 were the first to use case-control epidemiological methods to document the lung cancer hazard from cigarettes.

1939 to 1948

TOBACCO: A MEDICAL HISTORY

SIR RICHARD DOLL

TABLE IV Smoking and Lung Cancer Case-Control Studies Before 1950

	No. of Men		Nonsmokers, %		Heavy Smokers, %	
Author	Lung Cancer	Controls	Lung Cancer	Controls	Lung Cancer	Controls
Müller[30]	86	86	3.5	16.3	65	36
Schairer and Schöniger[32]	93	270	3.2	15.9	52	27
Wassink[33]	134	100	4.5	19.0	55	19

1 Doll R. Tobacco: a medical history. *J Urban Health* 1999;76:289-313.

German tobacco epidemiology by this time was the most advanced in the world. Franz H Muller in 1939 and Eberhard Schairer and Erich Schoniger in 1943 were the first to use case-control epidemiological methods to document the lung cancer hazard from cigarettes.

BRITISH MEDICAL JOURNAL

LONDON SATURDAY SEPTEMBER 30 1950

SMOKING AND CARCINOMA OF THE LUNG

PRELIMINARY REPORT

BY

RICHARD DOLL, M.D., M.R.C.P.

Member of the Statistical Research Unit of the Medical Research Council

AND

A. BRADFORD HILL, Ph.D., D.Sc.

Professor of Medical Statistics, London School of Hygiene and Tropical Medicine; Honorary Director of the Statistical Research Unit of the Medical Research Council

1950

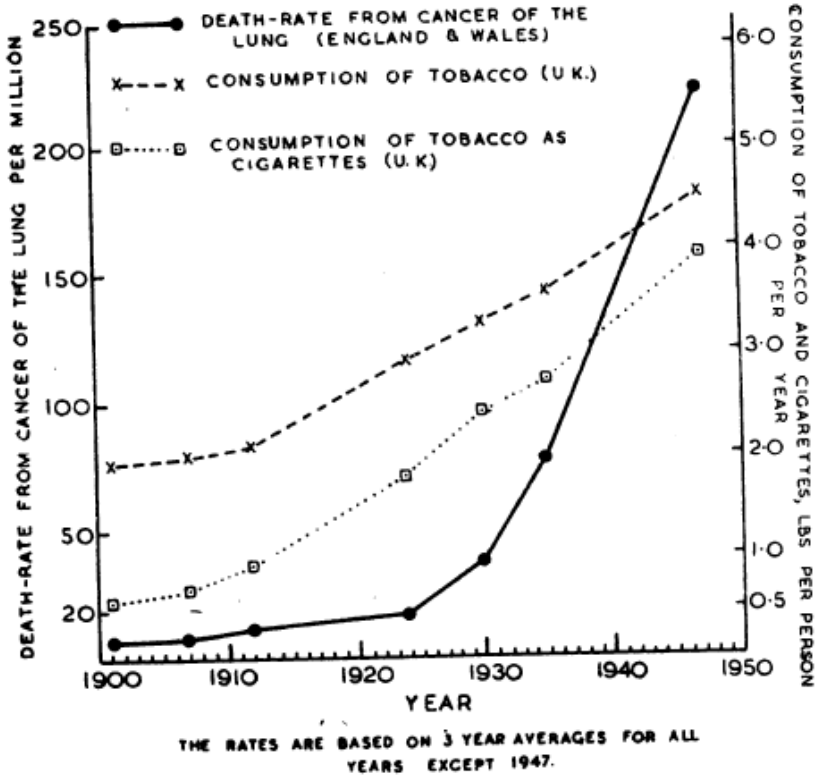

FIG. 2.—Death rate from cancer of the lung and rate of consumption of tobacco and cigarettes.

TABLE VII.—*Estimate of Total Amount of Tobacco Ever Consumed by Smokers; Lung-carcinoma Patients and Control Patients with Diseases Other Than Cancer*

Disease Group	No. Who have Smoked Altogether					Probability Test
	365 Cigs.–	50,000 Cigs.–	150,000 Cigs.–	250,000 Cigs.–	500,000 Cigs.+	
Males:						
Lung-carcinoma patients (647)	19 (2·9%)	145 (22·4%)	183 (28·3%)	225 (34·8%)	75 (11·6%)	$\chi^2=30·60$; n=4; P<0·001
Control patients with diseases other than cancer (622)..	36 (5·8%)	190 (30·5%)	182 (29·3%)	179 (28·9%)	35 (5·6%)	
Females:						
Lung-carcinoma patients (41)..	10 (24·4%)	19 (46·3%)	5 (12·2%)	7 (17·1%)	0 (0·0%)	$\chi^2=12·97$; n=2; 0·001< P < 0·01 (Women smoking 15 or more cigarettes a day grouped together)
Control patients with diseases other than cancer (28) ..	19 (67·9%)	5 (17·9%)	3 (10·7%)	1 (3·6%)	0 (0·0%)	

BRITISH MEDICAL JOURNAL

LONDON SATURDAY SEPTEMBER 30 1950

SMOKING AND CARCINOMA OF THE LUNG

PRELIMINARY REPORT

BY

RICHARD DOLL, M.D., M.R.C.P.

Member of the Statistical Research Unit of the Medical Research Council

AND

A. BRADFORD HILL, Ph.D., D.Sc.

Professor of Medical Statistics, London School of Hygiene and Tropical Medicine; Honorary Director of the Statistical Research Unit of the Medical Research Council

1950

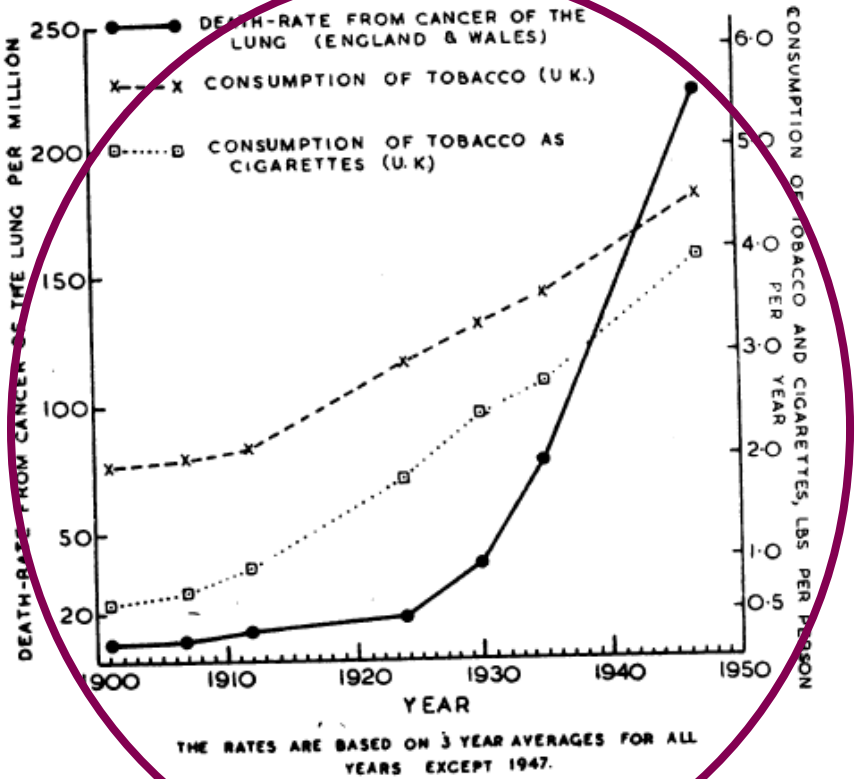

FIG. 2.—Death rate from cancer of the lung and rate of consumption of tobacco and cigarettes.

TABLE VII.—*Estimate of Total Amount of Tobacco Ever Consumed by Smokers; Lung-carcinoma Patients and Control Patients with Diseases Other Than Cancer*

Disease Group	No. Who have Smoked Altogether					Probability Test
	365 Cigs.–	50,000 Cigs.–	150,000 Cigs.–	250,000 Cigs.–	500,000 Cigs.+	
Males:						
Lung-carcinoma patients (647)	19 (2·9%)	145 (22·4%)	183 (28·3%)	225 (34·8%)	75 (11·6%)	$\chi^2=30·60$; n=4; P<0·001
Control patients with diseases other than cancer (622)..	36 (5·8%)	190 (30·5%)	182 (29·3%)	179 (28·9%)	35 (5·6%)	
Females:						
Lung-carcinoma patients (41)..	10 (24·4%)	19 (46·3%)	5 (12·2%)	7 (17·1%)	0 (0·0%)	$\chi^2=12·97$; n=2; 0·001< P < 0·01 (Women smoking 15 or more cigarettes a day grouped together)
Control patients with diseases other than cancer (28) ..	19 (67·9%)	5 (17·9%)	3 (10·7%)	1 (3·6%)	0 (0·0%)	

Cite this article as: BMJ, doi:10.1136/bmj.38142.554479.AE (published 22 June 2004)

Papers

Mortality in relation to smoking: 50 years' observations on male British doctors

Richard Doll, Richard Peto, Jillian Boreham, Isabelle Sutherland

1951 prospective study

This discovery stimulated much further research into the effects of smoking (not only on lung cancer but also on many other diseases), including a UK prospective study of smoking and death among British doctors that began in 1951 and has now continued for 50 years.[11–17] The decision that this study would be conducted among doctors was taken partly because it was thought that doctors might take the trouble to describe their own smoking habits accurately, but principally because their subsequent mortality would be relatively easy to follow, as they had to keep their names on the medical register if they were to continue to practise. Moreover, as most doctors would themselves have access to good medical care, the medical causes of any deaths among them should be reasonably accurately certified.

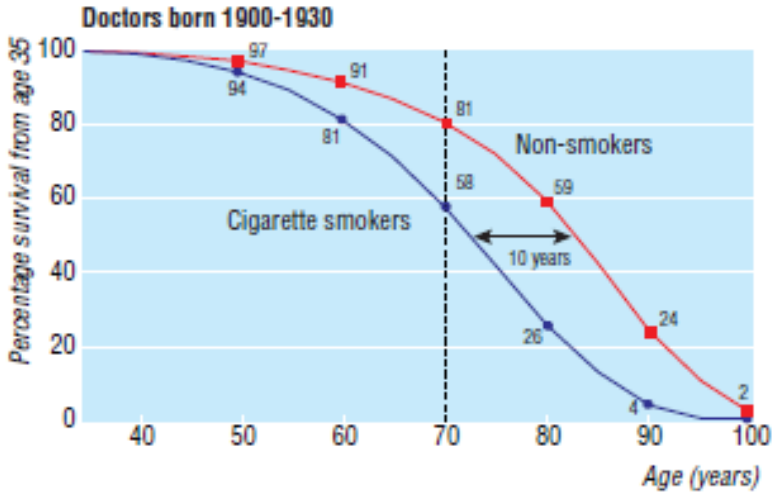

Fig 3 Survival from age 35 for continuing cigarette smokers and lifelong non-smokers among UK male doctors born 1900-1930, with percentages alive at each decade of age

1951 to today

Cite this article as: BMJ, doi:10.1136/bmj.38142.554479.AE (published 22 June 2004)

Papers

Mortality in relation to smoking: 50 years' observations on male British doctors

Richard Doll, Richard Peto, Jillian Boreham, Isabelle Sutherland

1951 prospective study

This discovery stimulated much further research into the effects of smoking (not only on lung cancer but also on many other diseases), including a UK prospective study of smoking and death among British doctors that began in 1951 and has now continued for 50 years.[11–17] The decision that this study would be conducted among doctors was taken partly because it was thought that doctors might take the trouble to describe their own smoking habits accurately, but principally because their subsequent mortality would be relatively easy to follow, as they had to keep their names on the medical register if they were to continue to practise. Moreover, as most doctors would themselves have access to good medical care, the medical causes of any deaths among them should be reasonably accurately certified.

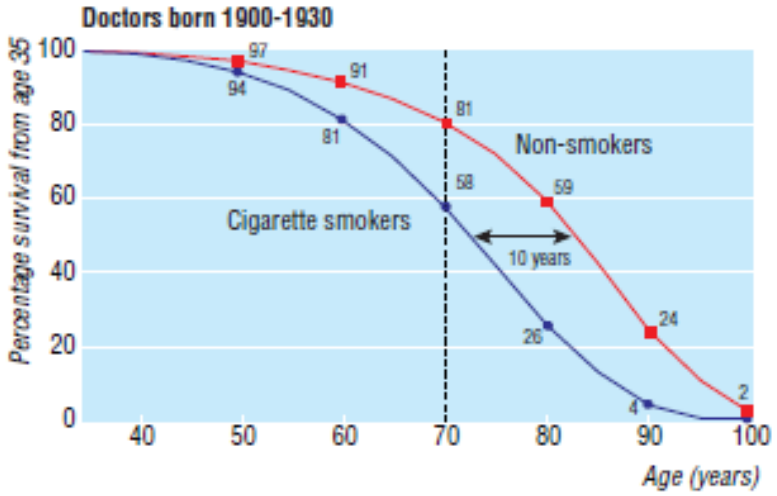

Fig 3 Survival from age 35 for continuing cigarette smokers and lifelong non-smokers among UK male doctors born 1900-1930, with percentages alive at each decade of age

1951 to today

HENRY SACKS, Ph.D., M.D.
THOMAS C. CHALMERS, M.D.
HARRY SMITH, Jr., Ph.D.
New York, New York

To compare the use of randomized controls (RCTs) and historical controls (HCTs) for clinical trials, we searched the literature for therapies studied by both methods. We found six therapies for which 50 RCTs and 56 HCTs were reported. Forty-four of 56 HCTs (79 percent) found the therapy better than the control regimen, but only 10 of 50 RCTs (20 percent) agreed. For each therapy, the treated patients in RCTs and HCTs had similar outcomes. The difference between RCTs and HCTs of the same therapy was largely due to differences in outcome for the control groups, with the HCT control

1982

The data presented suggest that such biases in patient selection may irretrievably bias the outcome of the HCT. It has been claimed that retrospective adjustment for prognostic factors can be used to produce an estimate of the effect of the treatment alone, but the studies we reviewed with such adjustments (either by the original authors or by us) showed nearly the same treatment effect as unadjusted studies. These adjustments were relatively crude and do not take into account possible interactions between prognostic factors. Recently, more sophisticated step-wise multiple regression procedures have been advocated [116], but there is as yet little evidence to suggest such procedures can better recognize ineffective therapies.

The accuracy of RCTs, on the other hand, could be improved by greater attention to sample size in planning studies. A recent review of 71 "negative" RCTs [117] found that a potential 25 percent improvement could have been missed in 57, and a potential 50 percent improvement in 34. At the planning stage of a trial, consideration of the size of the benefit sought and the number of patients needed to demonstrate it can keep the possibility of this sort of type II error at acceptable levels, but with increases in the cost and duration of the study.

A possible solution is to reconsider the nearly automatic use of a p value of less than 0.05 as the critical point at which a difference is felt to be statistically significant. Perhaps well-designed and well-blinded RCTs with little chance for bias should be considered positive when α is less than 0.10 or 0.20. This would increase the proportion of positive trials and save time and money. On the other hand, our data suggest that the opportunity for bias is so large in HCTs that when α is less than 0.01 or even 0.001, the therapy still may not be effective. The decision about what significance level to accept should also take into account other factors, including the prevalence of the disease, the medical and economic costs of the disease and of the therapy and the best pretrial estimate of the likelihood that the new therapy represents an advance.

HENRY SACKS, Ph.D., M.D.
THOMAS C. CHALMERS, M.D.
HARRY SMITH, Jr., Ph.D.
New York, New York

To compare the use of randomized controls (RCTs) and historical controls (HCTs) for clinical trials, we searched the literature for therapies studied by both methods. We found six therapies for which 50 RCTs and 56 HCTs were reported. Forty-four of 56 HCTs (79 percent) found the therapy better than the control regimen, but only 10 of 50 RCTs (20 percent) agreed. For each therapy, the treated patients in RCTs and HCTs had similar outcomes. The difference between RCTs and HCTs of the same therapy was largely due to differences in outcome for the control groups, with the HCT control

1982

The data presented suggest that such biases in patient selection may irretrievably bias the outcome of the HCT. It has been claimed that retrospective adjustment for prognostic factors can be used to produce an estimate of the effect of the treatment alone, but the studies we reviewed with such adjustments (either by the original authors or by us) showed nearly the same treatment effect as unadjusted studies. These adjustments were relatively crude and do not take into account possible interactions between prognostic factors. Recently, more sophisticated step-wise multiple regression procedures have been advocated [116], but there is as yet little evidence to suggest such procedures can better recognize ineffective therapies.

The accuracy of RCTs, on the other hand, could be improved by greater attention to sample size in planning studies. A recent review of 71 "negative" RCTs [117] found that a potential 25 percent improvement could have been missed in 57, and a potential 50 percent improvement in 34. At the planning stage of a trial, consideration of the size of the benefit sought and the number of patients needed to demonstrate it can keep the possibility of this sort of type II error at acceptable levels, but with increases in the cost and duration of the study.

A possible solution is to reconsider the nearly automatic use of a p value of less than 0.05 as the critical point at which a difference is felt to be statistically significant. Perhaps well-designed and well-blinded RCTs with little chance for bias should be considered positive when α is less than 0.10 or 0.20. This would increase the proportion of positive trials and save time and money. On the other hand, our data suggest that the opportunity for bias is so large in HCTs that when α is less than 0.01 or even 0.001 the therapy still may not be effective. The decision about what significance level to accept should also take into account other factors, including the prevalence of the disease, the medical and economic costs of the disease and of the therapy and the best pretrial estimate of the likelihood that the new therapy represents an advance.

2000

The New England Journal of Medicine

Special Articles

A COMPARISON OF OBSERVATIONAL STUDIES AND RANDOMIZED, CONTROLLED TRIALS

KJELL BENSON, B.A., AND ARTHUR J. HARTZ, M.D., PH.D.

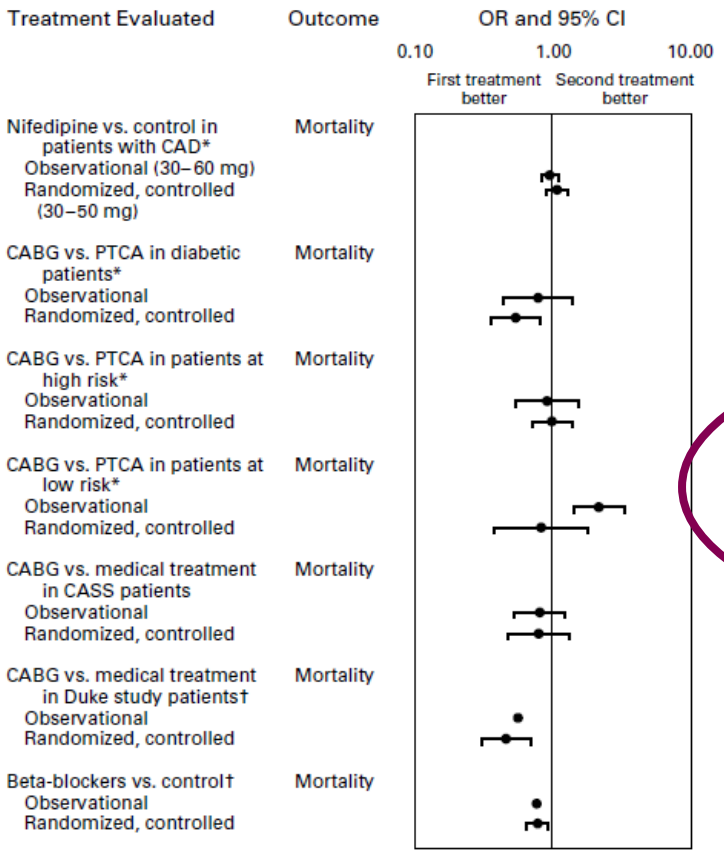

Figure 1. Results of Observational Studies and Randomized, Controlled Trials of Cardiologic Treatments.

The figure is based on data from eight articles.[13-20] Some articles contain data from more than one study. OR denotes odds ratio, CI confidence interval, CAD coronary artery disease, CABG coronary-artery bypass graft surgery, PTCA percutaneous transluminal coronary angioplasty, CASS Coronary Artery Surgery Study, and Duke the Duke University Cardiovascular Disease Databank. Asterisks indicate studies that reported relative risks rather than odds ratios. Daggers indicate studies that reported neither a confidence interval nor a P value for the odds ratio.

Conclusions We found little evidence that estimates of treatment effects in observational studies reported after 1984 are either consistently larger than or qualitatively different from those obtained in randomized, controlled trials. (N Engl J Med 2000;342:1878-86.)

RANDOMIZED, CONTROLLED TRIALS, OBSERVATIONAL STUDIES, AND THE HIERARCHY OF RESEARCH DESIGNS

JOHN CONCATO, M.D., M.P.H., NIRAV SHAH, M.D., M.P.H., AND RALPH I. HORWITZ, M.D.

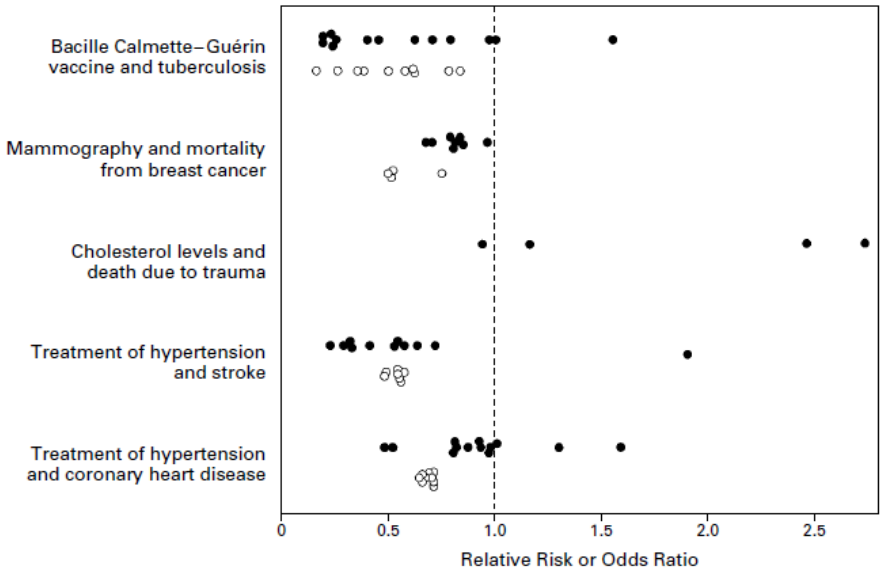

Figure 1. Range of Point Estimates According to Type of Research Design.

The studies evaluated bacille Calmette–Guérin vaccine and active tuberculosis (13 randomized, controlled trials and 10 case–control studies), screening mammography and mortality from breast cancer (8 randomized, controlled trials and 4 case–control studies), cholesterol levels and death due to trauma among men (4 of 6 randomized, controlled trials [2 trials did not provide point estimates]; the results of the 14 cohort studies were not reported individually), treatment of hypertension and stroke among only the men in the studies (11 randomized, controlled trials and 7 cohort studies), and treatment of hypertension an disease among only the men in the studies (13 of 14 randomized, controlled trials [1 trial did not provide poir 9 cohort studies). Solid circles indicate randomized, controlled trials, and open circles observational studies.

2000

Conclusions The results of well-designed observational studies (with either a cohort or a case–control design) do not systematically overestimate the magnitude of the effects of treatment as compared with those in randomized, controlled trials on the same topic. (N Engl J Med 2000;342:1887-92.)

Contents

- **Levels of evidence**
  - *RCT vs non-RCT information*
- **An interesting history**
  - *Smoking vs Lung Cancer*
    - *1939-1950*
    - *The 50 year study*
  - *1982 - The original argument against observational analyses/studies*
  - *2000 - The argument for observational analyses/studies*
- ***So why the interest now?***
- **Discussion: So what now?**

So why the interest now?

- **Statisticians are becoming more involved in strategic elements of research** e.g.

  - *Quantifying high interest Disease Areas*
  - *Valuation of global comparator therapies*
  - *Overall prevalence and burden of illness*
  - *Key market/cost/recruitment drivers*

- **Real world information is becoming more important for payers** e.g.

  - *Value based pricing (e.g. in UK)*
  - *Evidence based pricing*

So why the interest now?

- **Easier access to non-RCT data**

  - *Claims data*
  - *Electronic medical records*
  - *Prescribing information*
  - *Lab data*
  - *Disease registries*
  - *Better technology to collect "Real world" effectiveness data e.g. via mobile phones etc*

- **Payers and regulators use non-RCT data to form decisions**
  - Therefore these databases are being analyzed

So why the interest now?

- **<u>As clinical statisticians we are being asked to input into observational analyses</u>**

  - Interpretation dependent on good design
  - Realisation that we have access to larger databases not available via RCTs
  - Can be used to supplement RCTs

Payers and regulators use non-RCT data to make “informed” decisions

From IQWiG press release

26.06.2009 Insulin analogue glargine possibly increases cancer risk

Data evaluated of approximately 130,000 diabetes patients insured in the German Local Health Care Fund (AOK) - IQWiG and WIdO publish joint analysis

The risk of cancer possibly increases if patients with diabetes use the long-acting insulin analogue glargine instead of human insulin. The Institute for Quality and Efficiency in Health Care (IQWiG), in collaboration with the "Wissenschaftliches Institut der AOK” (WIdO), the research institute of the German Local Health Care Fund, analysed the data of almost 130,000 patients with diabetes in Germany who had been treated with either human insulin or the insulin analogues lispro (trade name: Humalog), aspart (Novorapid) or glargine (Lantus) between January 2001 and June 2005.

Payers and regulators use non-RCT data to make "informed" decisions

From the RPM report

Friday, November 20 2009 Plavix Label Change: Good For Effient Now, Bad For Brands in the Long Run?

By Michael McCaughan

FDA is strengthening a warning against coadministration of PPIs with Plavix. In the logic of blockbuster markets, the bad news for Bristol is good news for Lilly and its competitor Effient. But this isn't an old model story…

Because there are third parties involved: payors and pharmacy benefit managers. The interaction between PPIs and Plavix was first publicized by Aetna and by Medco, both of whom used claims data to suggest an association between PPI use and diminished outcomes for patients treated with Plavix.

It is not just that payors capitalized on a safety issue: they really drove the regulatory response and the application of a newly discovered pharmacogenomic marker. In Medco's case at least, Chief Medical Officer Robert Epstein told us, the whole idea was to find a way to test the emerging theory that CYP2C19 genotyping may predict Plavix response. Since Medco didn't have genotyping data on patients in its database, it looked at concomitant use of omeprazole instead, since the PPI is a known inhibitor of the 2C19 pathway.

FDA's first public health alert followed the Aetna and Medco claims studies; the latest one came after Bristol and Sanofi conducted a drug interaction study confirming the observational results. That's certainly not a regulatory model sponsors are eager to consider--especially since we would be willing to bet that the observational research that triggered the warning cost Medco much less than the clinical trial the sponsors were forced to conduct to confirm it.

Medco, at least, isn't done. As we reported here, the company is now taking the next step, conducting a large scale observational study to test the hypothesis that the superior efficacy demonstrated by Lilly in its head-to-head study of Effient vs. Plavix can be explained by the inclusion of poor metabolizers of Plavix in the comparator group.

Payers and regulators use their data to make "informed" decisions
From the RPM report

Friday, November 20 2009 Plavix Label Change: Good For Effient Now, Bad For Brands in the Long Run?

By Michael McCaughan

FDA is strengthening a warning against coadministration of PPIs with Plavix. In the logic of blockbuster markets, the bad news for Bristol is good news for Lilly and its competitor Effient. But this isn't an old model story…

Because there are third parties involved: payors and pharmacy benefit managers. The interaction between PPIs and Plavix was first publicized by Aetna and by Medco, both of whom used claims data to suggest an association between PPI use and diminished outcomes for patients treated with Plavix.

It is not just that payors capitalized on a safety issue: they really drove the regulatory response and the application of a newly discovered pharmacogenomic marker. In Medco's case at least, Chief Medical Officer Robert Epstein told us, the whole idea was to find a way to test the emerging theory that CYP2C19 genotyping may predict Plavix response. Since Medco didn't have genotyping data on patients in its database, it looked at concomitant use of omeprazole instead, since the PPI is a known inhibitor of the 2C19 pathway.

FDA's first public health alert followed the Aetna and Medco claims studies; the latest one came after Bristol and Sanofi conducted a drug interaction study confirming the observational results. That's certainly not a regulatory model sponsors are eager to consider--especially since we would be willing to bet that the observational research that triggered the warning cost Medco much less than the clinical trial the sponsors were forced to conduct to confirm it.

Medco, at least, isn't done. As we reported here, the company is now taking the next step, conducting a large scale observational study to test the hypothesis that the superior efficacy demonstrated by Lilly in its head-to-head study of Effient vs. Plavix can be explained by the inclusion of poor metabolizers of Plavix in the comparator group.

Contents

- **Levels of evidence**
  - *RCT vs non-RCT information*
- **An interesting history**
  - *Smoking vs Lung Cancer*
    - *1939-1950*
    - *The 50 year study*
  - *1982 - The original argument against observational analyses/studies*
  - *2000 - The argument for observational analyses/studies*
- ***So why the interest now?***
- **Discussion: So what now?**

Totality of Evidence- *an we make better use of observational data?*

Discussion

1) Use of non RCT data

To what extent can epidemiological studies, observational databases and other data sources be used:

-to help design RCTs
-to quantify safety signals not readily seen in RCTS (e.g. those with low prevalence or long term)
-to support reimbursement with payers (such as comparative effectiveness and real usage)
-and would we ever see a situation where they could replace RCTs?

2) Current thinking

-Are we investigating or evaluating opportunities in this area? Can you describe examples where observational data studies have been used to help inform internal or external decision making?

3) Current statistical capability

-Given these databases are being analysed already, is it better that we get good statisticians involved rather than simply point out the inherent weaknesses of such analyses?
-As a profession do we understand enough about observational analyses to either promote them or not?
-Do we need to make sure we have different types of skills in addition to those used for more traditional forms of evidence?

4)The Future

-Do we need more generalists with applied knowledge or more specialists with expert skills?
-Should we be developing partnerships with other professionals for example statisticians at Healthcare Providers, epidemiologists etc. ?