Evidence-based medicine: data mining and pharmacoepidemiology research

B. B. Little[1,2,3], R. A. Weideman[3], K. C. Kelly[3] & B. Cryer[3]
[1]Department of Mathematics, Physics, and Engineering, Texas, USA
[2]Data Mining Research Institute, Tarleton State University, Stephenville, Texas, USA
[3]Veterans Affairs Medical Centre, Dallas, Texas, USA

Abstract

Evidence-based medicine is an important initiative of the U.S. National Institutes of Health. However, published data are limited. Medical knowledge can be advanced through pharmacoepidemiology and data mining research of very large databases, data warehouses. Before Vioxx™ (rofecoxib) and Celebrex™ (celecoxib) were publicly suspected of being associated with acute myocardial infarction and stroke, etodolac was evaluated for gastrointestinal safety using a data warehouse constructed from several data sources and data mining tools embedded with pharmacoepidemiologic methods. Results were published in a top tier medical journal, and established gastrointestinal safety of a generic COX-2 analgesic medication. This project demonstrates that the knowledge needed for evidence-based medicine can be augmented through pharmaco-epidemiology research using data mining and data warehousing.
Keywords: evidence-based medicine, pharmacoepidemiology, data mining, data warehousing.

1   Introduction

Evidence-based medicine is strongly advocated by the U.S. National Institutes of Health [1], and is a high funding priority. Between November 1992 and March 2006, 20,697 publications indexed by the National Library of Medicine list "evidence-based medicine" as a keyword (Figure 1). Evidence-based medicine is a structured analytical process in which published medical research is used to critically assess the efficacy and safety of different medical interventions. The

objective of evidence-based medicine is to improve physician decision-making and patient outcomes while containing costs. Assessments begin with clear empirical definitions of the clinical question (end-point) to be evaluated. Exhaustive and structured literature search identifies published evidence that directly bears on the clinical question. Predefined rules are applied to assess quality of published evidence, and the extent to which it bears directly on the clinical question. The objective nature of evidence-based medicine can provide a powerful decision-support tool for practising physicians, and ultimately improve patient outcomes [3].
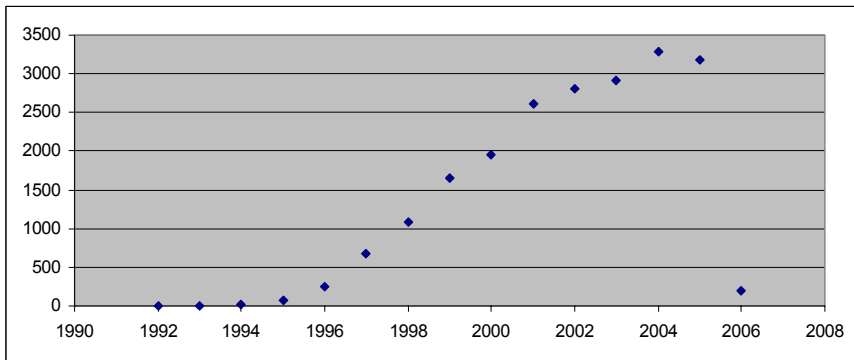

Figure 1:   Number of publications in U.S. National Library of Medicine with Keyword Phrase "Evidence-Based Medicine" (November 1992 to March 13, 2006).

Limitations of evidence-based medicine include: (1) shortage of coherent, consistent scientific evidence, (2) difficulties in applying evidence to the care of individuals patients, (3) barriers to the practise of high quality medicine, (4) need for clinicians to develop new skills (literature search and synthesis), (5) clinicians have limited time and resources, and (6) misperception that clinical skills and expertise are denigrated by evidence-driven practise [3]. Data mining in pharmacoepidemiologic research can (1) overcome evidence shortages, (2) help provide information applicable to individuals patients, (3) support best practises in medicine, (4) provide summary sources to minimize the need to learn new computer skills, or (5) invest time and resources, and (6) be an adjunct to clinical expertise, not a replacement.

Wide interest in evidence-based medicine exists. Unfortunately, empirical data to support evidence-based decisions in medicine are not widely available. For example, comprehensive evaluation of evidence-based maternal dose formulation in obstetrics found that only one drug was adequately studied to guide clinical practice [4]. In such situations, a void in medical evidence exists, and clinical expertise is the only guide for practise. The lack of information in certain areas is an opportunity for advancement of the use of technology to improve medical practise and benefit the human condition [3].

The objective of this paper is to demonstrate a general approach for the use of data warehousing, data mining, and pharmacoepidemiology technologies to advance evidence-based medicine. The aim of this demonstration project is to assess the gastrointestinal (GI) safety of etodolac, a generically available COX-2 selective analgesic. Patented COX-2 analgesics were removed from the US market because of an increased risk of acute myocardial infarction and stroke.

2 Materials and methods

The present investigation was approved by the Institutional Review Board for Use of Human Subjects, Veterans Affairs Medical Centre, Dallas, Texas, USA.
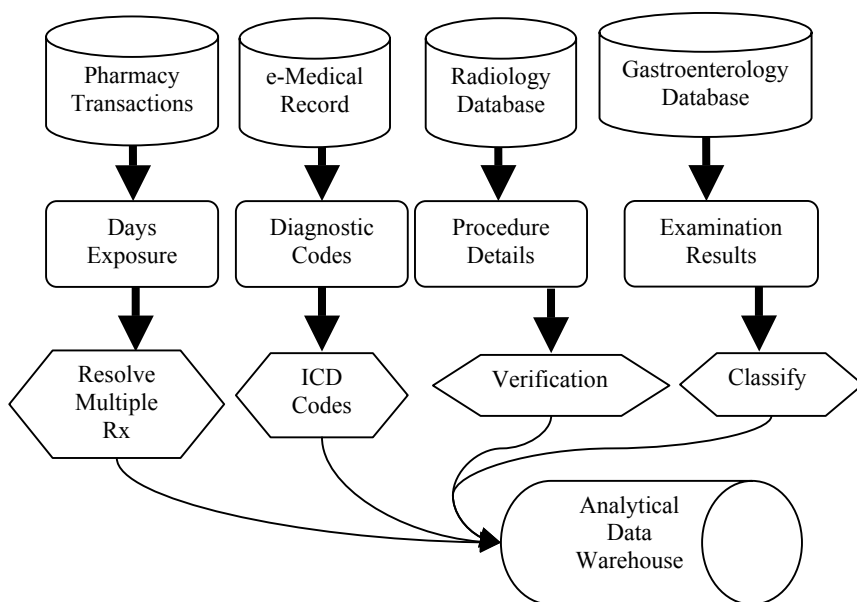

Figure 2:   Schema of data warehouse architecture.

2.1 Study subjects

The study subjects were U.S. Veterans who received health care and were prescribed naproxen or etodolac at the Veterans Affairs Medical Centre, Dallas, Texas, USA. Data were extracted from four separate databases that contain more than 294,000 unique patients.

2.2 Data warehouse architecture

The analytical data warehouse was constructed by standardizing and joining four very large transactional databases. SAS and SAS Warehouse were integrated used for joining and administering the databases used in this study (Figure 2).
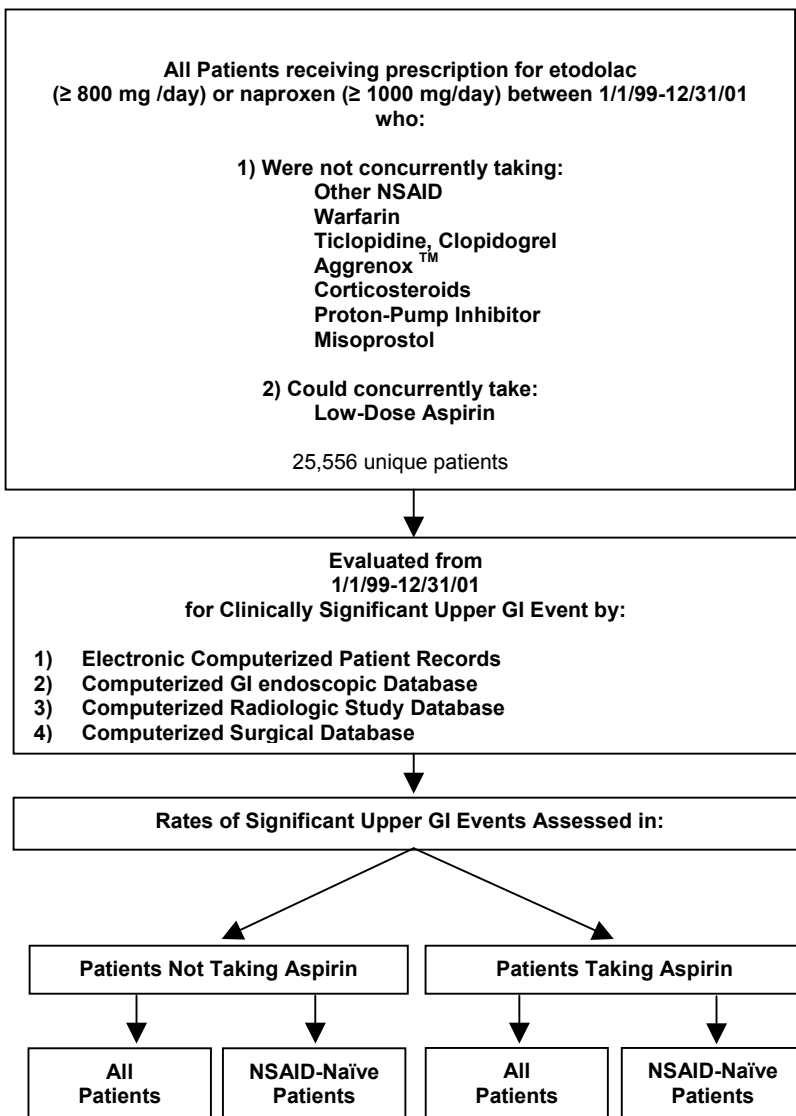
2.3 Inclusion criteria

All Patients receiving prescription for etodolac
(≥ 800 mg /day) or naproxen (≥ 1000 mg/day) between 1/1/99-12/31/01
who:

1) Were not concurrently taking:
    Other NSAID
    Warfarin
    Ticlopidine, Clopidogrel
    Aggrenox ^{TM}
    Corticosteroids
    Proton-Pump Inhibitor
    Misoprostol

2) Could concurrently take:
    Low-Dose Aspirin

25,556 unique patients

Evaluated from
1/1/99-12/31/01
for Clinically Significant Upper GI Event by:

1) **Electronic Computerized Patient Records**
2) **Computerized GI endoscopic Database**
3) **Computerized Radiologic Study Database**
4) **Computerized Surgical Database**

Rates of Significant Upper GI Events Assessed in:

Patients Not Taking Aspirin

Patients Taking Aspirin

| **All Patients** | **NSAID-Naïve Patients** | **All Patients** | **NSAID-Naïve Patients** |

Figure 3:   Inclusion criteria [5].

Table 1:   Diagnostic criteria for GI events [5].

Event	Confirmation
Gastric or duodenal perforation	Perforated lesion requiring surgery. Could involve a laparoscopic repair, but only if evidence of the perforation was unequivocal, such as free air in the abdomen visible on radiograph or direct visualization at surgery
Gastric outlet obstruction	Gastric outlet obstruction requiring diagnosis by investigator; diagnosis was required to be supported by endoscopy (e.g., ulcer with a tight edematous pyloric channel) or by radiographic results (e.g., dilated stomach, delayed barium emptying with clinical evidence of outlet obstruction and with an ulcer in the channel, severe outlet narrowing and edema)
Upper GI bleeding (At least one criterion present)	- Hematemesis with a lesion (ulcer or large erosion) on endoscopy or radiograph - Hematemesis or aspiration of bloody gastric fluid witnessed by health care provider; episode of melena witnessed by health care provider - Lesion (ulcer or large erosion) on endoscopy with evidence of active bleeding or stigmata of recent hemorrhage (visible vessel or clot attached to the base of an ulcer) - Melena with a lesion (ulcer or large erosion) on endoscopy or radiograph - Occult blood-positive stool with a lesion (ulcer or large erosion) on endoscopy or radiograph and with evidence of serious bleeding, including at least 1 of the following: - Decrease from baseline in hematocrit of $\geq 5\%$ points or in hemoglobin of > 15 g/L - Postural vital sign changes (increase in heart rate of ≥ 20/min and/or decrease in systolic blood pressure of ≥ 20 mm Hg and/or in diastolic blood pressure of $10\geq$mm Hg) - Transfusion of ≥ 2 units of blood - Blood in stomach on endoscopy or nasogastric aspiration
Symptomatic gastric or duodenal ulcer	1. Evidence of gastric or duodenal ulcer on endoscopy, at surgery, on contrast-enhanced radiography of the upper gastrointestinal tract, or at autopsy, or, 2. Abdominal pain or dyspepsia as patient indication followed by finding of gastric or duodenal ulcer

2.4 Analytical techniques

The endpoint of the analyses was whether or not study subjects had a bleeding ulcer or other GI event whilst taking one of the study medications. SAS (V.8.1) and SAS Enterprise Miner (V4.2) were used for the computations. PROC FREQ, PROC UNIVARIATE, and PROC LOGISTIC (Direct and Stepwise) were used in the analysis.

3   Results

3.1 Baseline patient characteristics

Patients in the study groups were similar, except etodolac patients were several years older than naproxen patients.

Table 2:   Baseline patient characteristics [5].

Characteristics	Etodolac (n=9,607)	Naproxen (n=8,382)
Patient-yrs exposure	2600	2996
Age, mean ± SD years	57.6±14.0	54.6±13.8
≥ 50 years	72.2%	63.7%
Men, (%)	91.0%	88.0%
NSAID-naïve (%)	59.5%	58.7%
Rheumatoid arthritis (%)	2.0%	2.1%
Congestive Heart Failure (%)	6.4%	6.2%
Diabetes Mellitus (%)	14.8%	14.7%
Concurrent Low-Dose Aspirin (%)	18.2%	16.1%
History of Ulcer (%)	4.7%	3.7%
Mean time in DVAMC System prior to study, (mean ± SD years)	4.4 ± 3.7	4.4 ± 3.5

Time to GI event (bleeding ulcer or other GI event, see Table 3) analysis revealed a wide separation between etodolac and naproxen within weeks of analgesic therapy (Figure 4).

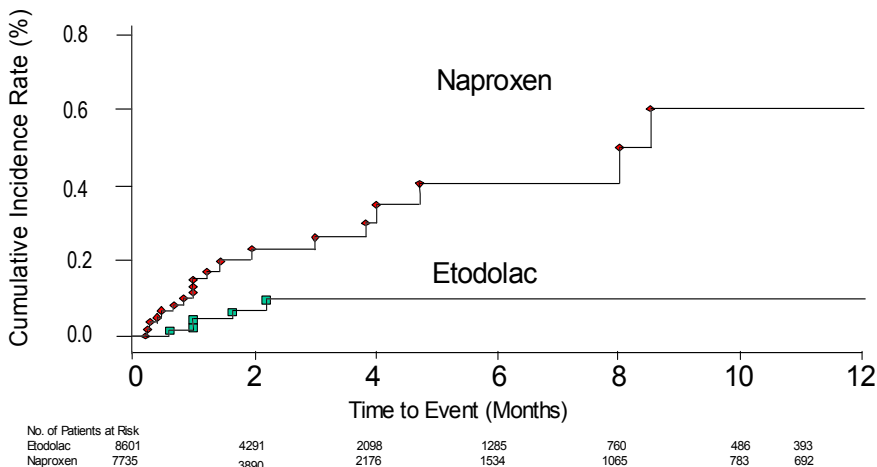| No. of Patients at Risk						
Etodolac	8601	4291	2098	1285	760	486	393
Naproxen	7735	3890	2176	1534	1065	783	692

Figure 4:   Clinically significant upper GI events in patients not taking aspirin [5].

The rates of upper GI events were significantly lower among patients treated with etodolac compared to naproxen (Figure 5A). Concomitant aspirin use caused the difference between etodolac and naproxen to disappear (Figure 5B).
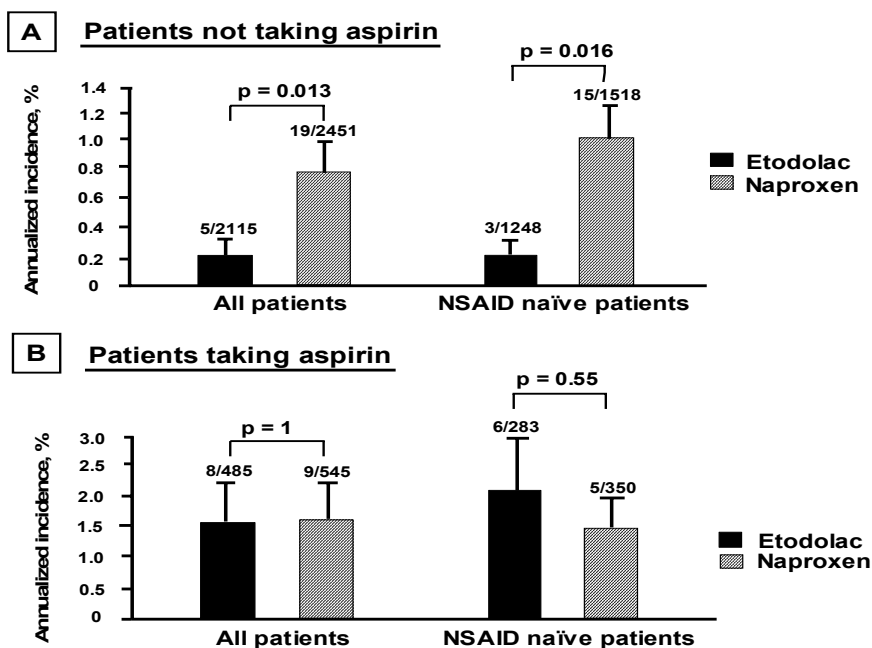

Figure 5:   Rates of clinically significant upper GI events (number of events / patients years) [5].

The rate of GI events per patient-year (pt-yrs) was higher among naproxen- than etodolac-exposed patients, and the most frequent site was upper GI.

Table 3:   Clinically significant upper GI events [5].

Clinically significant Upper GI Event	Etodolac (2600 pt-yrs)	Naproxen (2996 pt-yrs)
Symptomatic Gastroduodenal Ulcer	6	9
Upper GI Bleeding*	7	17
Perforation	0	2
Total	13	28

*The cause or source of upper GI bleeding was gastric ulcer in 7 patients, duodenal ulcers in 13, and other GI sources in 4 patients.

Aspirin is shown to alter the safety profile of etodolac, making it just as likely to be associated with a GI event as naproxen (Table 4).

Table 4:   Annualized incidence rates and adjusted odds ratio of CSUGI events for Etodolac versus Naproxen [5].

	Annualized Incidence Etodolac/Naproxen	Odds Ratio	0.95 CI	P
Not Taking Aspirin				
All Patients	0.24 / 0.78	0.24	0.09 to 0.63	0.01
NSAID-Naïve Patients	0.24 / 0.99	0.24	0.18 to 0.61	0.01
Taking Aspirin				
All Patients	1.65 / 1.65	0.75	0.28 to 1.99	NS
NSAID-Naïve Patients	2.12 / 1.43	1.24	0.35 to 4.42	NS

4   Discussion

Collectively, in the absence of aspirin, these data indicate that a generic COX-2 agent is not associated with GI events when not used concomitantly with aspirin [5], and may be a safe substitute for rofecoxib and celecoxib, pending publication of our ongoing analysis of the association between the generic drug and acute myocardial infarction. Using data warehouse and data mining technology with pharmacoepidemiology research, disparate data sources can remedy the shortage of coherent, consistent scientific evidence (see [3]). Massive amounts of data remain not analyzed because they are not in a common format and reside in stovepipe databases. Evidence-based medicine can be advanced through the use of data warehouse architecture, data mining analysis, and pharmaco-epidemiology.

References

[1]   Evidence-based Medicine Working Group.: Evidence-based medicine: a new approach to teaching the practice of medicine. *Journal of the American Medical Association* 208:2420-5, 1992.
[2]   Ellrodt, G., Cook, D.J., Lee, J., Cho, M., Hunt, D., Weingarten, S.: Evidence-based disease management. *Journal of the American Medical Association* 278:1687-1692, 1997.
[3]   Straus, S.E., McAlister, F.A.: Evidence-based medicine: Commentary of criticisms. *Canadian Medical Association Journal* 163:837-841, 2000.
[4]   Little, B.B.: Pharmacokinetics during pregnancy: Evidence-based maternal dose formulation. *Obstetrics and Gynecology* 93: 858-868, 1999.
[5]   Weideman, R.A., K.C. Kelly, S. Kazi, A. Cung, K.W. Roberts, H.J. Smith, G.A. Sarosi, B.B. Little, B. Cryer: Risks of clinically significant upper gastrointestinal toxicity with etodolac and naproxen: A historical cohort analysis. *Gastroenterology* 127:1322-1328, 2004.