# Database

**Open Access** 

# A suite of web applications to streamline the interdisciplinary collaboration in secondary data analyses

Ricardo Pietrobon<sup>\*1</sup>, Ulrich Guller<sup>2</sup>, Henrique Martins<sup>3</sup>, Andreia P Menezes<sup>4</sup>, Laurence D Higgins<sup>5</sup> and Danny O Jacobs<sup>6</sup>

Address: <sup>1</sup>Center for Excellence in Surgical Outcomes, Duke University Medical Center, Duke University, Durham, NC, USA, <sup>2</sup>Center for Excellence in Surgical Outcomes, University Hospital Basel, Department of General Surgery and Surgical Research, Basel, Switzerland, <sup>3</sup>Duke University Health System, Rio Claro/Sao Paulo, Brazil, <sup>4</sup>Reliable International Research, Center for Excellence in Surgical Outcomes, Campina Grande do Sul/ Parana – Brazil, <sup>5</sup>Department of Surgery, Duke University Medical Center, Durham, NC, USA and <sup>6</sup>Department of Internal Medicine, University of North Carolina at Chapel Hill, NC, USA

Email: Ricardo Pietrobon\* - rpietro@duke.edu; Ulrich Guller - uguller@yahoo.com; Henrique Martins - henrique@duke.edu; Andreia P Menezes - amenezes@duke.edu; Laurence D Higgins - l.higgins@duke.edu; Danny O Jacobs - Danny.Jacobs@duke.edu \* Corresponding author

Published: 14 December 2004

BMC Medical Research Methodology 2004, 4:29 doi:10.1186/1471-2288-4-29

This article is available from: http://www.biomedcentral.com/1471-2288/4/29

© 2004 Pietrobon et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<u>http://creativecommons.org/licenses/by/2.0</u>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received: 10 May 2004 Accepted: 14 December 2004

#### Abstract

**Background:** We describe a system of web applications designed to streamline the interdisciplinary collaboration in outcomes research.

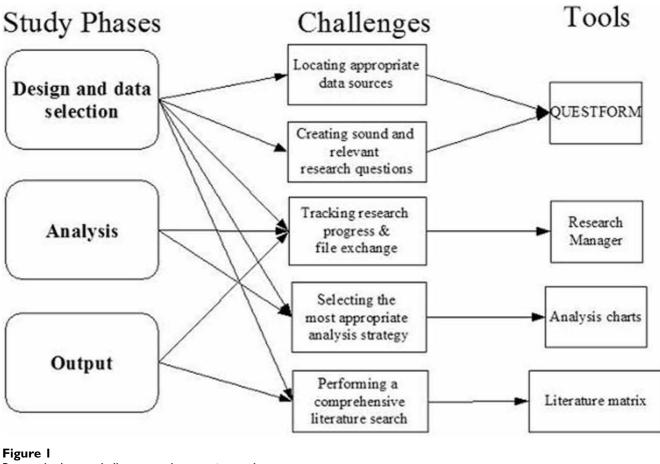
**Description:** The outcomes research process can be described as a set of three interrelated phases: design and selection of data sources, analysis, and output. Each of these phases has inherent challenges that can be addressed by a group of five web applications developed by our group. QuestForm allows for the formulation of relevant and well-structured outcomes research questions; Research Manager facilitates the project management and electronic file exchange among researchers; Analysis Charts facilitate the communication of complex statistical techniques to clinicians with varying previous levels of statistical knowledge; Literature Matrices improve the efficiency of literature reviews. An outcomes research question is used to illustrate the use of the system.

**Conclusions:** The system presents an alternative to streamline the interdisciplinary collaboration of clinicians, statisticians, programmers, and graduate students.

#### Background

In the last decade, the number of relevant data sources available for outcomes research has grown exponentially. In contrast, the number of individual researchers with clinical and statistical expertise required to explore these data sets increase at a much slower pace. As a result, an immense quantity of valuable clinical data are left untouched, never becoming clinical publications that could potentially improve health care.

The disproportion between data volume and number of qualified researchers can be explained by the growing complexity involved in outcomes research projects using secondary data analyses. Researchers have to formulate of



Research phases, challenges, and respective tools

a clinically relevant and methodologically sound research question, find appropriate data sources, perform statistical analyses, and generate a final manuscript that will be submitted for peer-review. Frequently, individual researchers have the training and time to perform a few of these steps, but the integration of all tasks calls for an interdisciplinary systems approach [1,2]. This interdisciplinary effort, however, is often challenged by communication problems among researchers with different backgrounds, particularly when physicians with an exclusive clinical education attempt to work in collaboration with quantitative researchers such as statisticians [3]. As a consequence, the output of such collaboration is either scarce or absent.

This article describes a suite of web applications developed to facilitate the process of converting outcome databases into clinical manuscripts, to streamline the interdisciplinary collaboration of researchers, and to connect all different steps of the outcomes research process. To illustrate its use, we will describe how a research project has been conducted using this system from its early phase of research question formulation to the completion of the final manuscript.

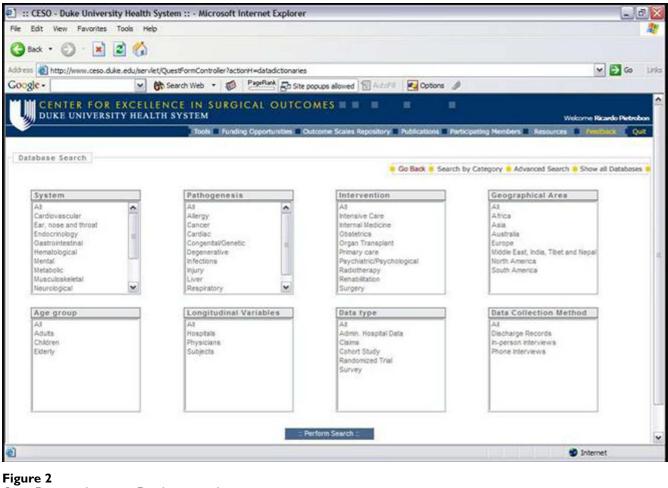
## **Construction and content**

The system of Web applications is composed by five different tools: QuestForm, Research Manager, Analysis Charts, and Literature Matrices. These tools were designed to assist researchers in each of the phases encountered in an outcomes research project involving secondary data analysis (Figure 1). All tools are freely available at a designated web site <u>http://www.ceso.duke.edu</u>. The following sections will describe each of the Web applications and their application in the answer of a real outcomes research question.

# QuestForm

#### General description

QuestForm, an acronym for "Question Formulation", is an application designed to assist researchers in the location of clinical databases and formulation of outcome



QuestForm application – Database search engine

research questions (Figure 2). Clinical databases contain raw data (observations from individual patients) from national administrative claim data, cohort studies, clinical trials, and registries (see <u>http://www.ceso.duke.edu</u> for an updated list). All databases have been de-identified and do not contain protected health information as specified by the Health Insurance Portability and Accountability Act (<u>http://www.hhs.gov/ocr/hipaa/</u>, accessed on Aug/04/ 2004). The application The application is built using Extensible HyperText Markup Language 1.0 (XHTML) [4], Java, and a relational database (MySQL 4.0)[5].

QuestForm starts by presenting researchers with three main strategies to find research databases: use of predetermined key words that describe the database as a unit (Figure 2), user-defined key words to describe variables present in the data dictionary of each database, and the presentation of a complete list of all databases. Once databases are located, researchers can read an overall summary about the database including details about number of subjects, sampling strategy, data ownership, and overall characteristics of the study population and associated procedures (Figure 3). Researchers then determine that the database most appropriate to answer the research question at hand, a JAVA screen is displayed for research question formulation (Figure 4). This screen presents all variables displayed in hierarchical categories. Variables are presented with the corresponding question and alternative responses. All variables can be inserted into a research question (Question Diagram) divided into the classical categories for an epidemiological question: Outcomes, Predictors, Confounders, Inclusion and Exclusion Criteria. Search engines are provided for ICD9-CM diagnosis and procedure codes (Figure 5), which can also be inserted into the Question Diagram. Finally, previously formulated Question Diagrams can be shared among researchers. This latter functionality allows researchers to both share Question Diagrams among members of the

2] :: CESO - Duke University Health S	rstem :: - Microsoft Internet Explorer	6	
File Edit View Favorites Tools Help			1
🔇 Back + 🔘 - 💌 😰 🐔			
Address Address Address Address	QuestFormController?actionH=description&idDataSet=2	💌 🛃 Go	Uria
Google - 🛛 🖌 🗑	Search Web 🔹 💋 Patriant 🗗 Site popups allowed 🔞 AutoFill 🛃 Options 🥒	and the state	
CENTER FOR EXCELLE	NCE IN SURGICAL OUTCOMES	Weicome Ricardo Pietrob	n
V	Tools 🖩 Funding Opportunities 🔳 Outcome Scales Repository 🖬 Publications 🖷 Participating Membérs 🖀 Reso	vrces 🖡 Resilback 🛙 Qu	
Database Search			
	Go Back Search by Category Advanced 5	earch 🧧 Show all Databases	
Database General Informat	ion		
Database Title:	Alameda - Alameda County Health and Ways of Living		
Keywords:	activities of daily living, chronic illnesses, community involvement, health behavior, health care services, indep arrangements, mental health, older adults, physical condition, social behavior, social life, social networks	endent living, living	
Contact information:	George A. Kaplan PO Box 1248 Anna Harbor, MI 48106-1248 (734) 647-5000 netmall@icpsr.umich.edu http://www.icpsr.umich.edu/cgi-bin/archive.prl?study=3083&path=ICPSR&allflag=open		
Data Dictionary:	Click here to view		
Database Description			
Description:	The Alameda County Health and Ways of Living Study is a collection which provides a 30-year follow-up with it Alameda County who were originally interviewed in 1965 for the first wave of the Health and Ways of Living Sti survey was to explore the influences of health practices and social relationships on the physical and mental h of the population. The first wave of the study, HEALTH AND WAYS OF LIVING STUDY, 1965 PANEL: (ALAMED/ (LCPSR 5688), collected information for 6.926 respondents (including anonzyimately 500 women and 65 years).	idy. The purpose of the ealth of a typical sample COUNTY, CALIFORNIA]	*
<b>e</b> ]		1 Internet	

Figure 3 QuestForm application – Overall description of the database

ongoing project as well as share previously formulated Question Diagrams with researchers from other teams. Once the question is fully formulated, researchers can save the question as in a graphical format known as Question Diagram.

#### Outcomes research application

Dr. Guller initiated the project searching for an existing database that would allow him to compare surgical outcomes between laparoscopic and open appendectomy procedures in the treatment of acute appendicitis. The outcomes were pre-specified as mortality and infection, although other existing outcomes would also be of interest. Although there are multiple single-institution studies attempting to answer this question [6,7], few studies have taken a population approach to test whether one procedure is superior to the other.

As a first step, Dr. Guller searched across previously formulated Question Diagrams to evaluate whether other studies could have used a similar research design. Since none was found in QuestForm, Dr. Guller searched across more than forty different databases for an existing database that would have the variables to answer his research question. After navigating through multiple data dictionaries, Dr. Guller found that the Nationwide Inpatient Sample (NIS) Release 6, 1997 [8] presented the variables and an adequate number of patients to answer his question.

Once the database was located, Dr. Guller selected the outcomes of interest (length of hospital stay, in-hospital complications, in-hospital mortality, rate of routine discharge), main predictors (laparoscopic versus open procedures), and confounders (age, gender, race, household income, comorbidity, hospital volume, location of the hospital, teaching status of hospital, and appendix

:: CESO - Duke U	Jniversity Health System :: - Microsoft Internet Explorer	- 0
ile Edit View Fa	varites Taols Help	
3 Back • 🕥 ·		
Schreiss ) http://www	v.ceso.duke.edu/serviet/QuestFormController	💌 🛃 Go 🛛 Link
loogle -	💌 🛞 Search Web 🔹 🦪 PageFlank 🖧 Site popups allowed 🔞 AutoFill 🛃 Options 🥒	
	FOR EXCELLENCE IN SURGICAL OUTCOMES = = = = = = = = = = = = = = = = = = =	Welcome Ricardo Pietrobon
<i>¥</i>	. Tools 🖬 Funding Opportuncies 🖬 Outcome Scales Repository 🖬 Publications 🖩 Participating Member	n 🖩 Resources 🐘 Feliliback 🛛 Quit
	Year     Anticipated Coordinations     Predictors     Outcomes     Confounders       Exclusion     Supported Folding     Supported Comments     Tables       B     V     IF     If	Database Vear_hear_stack Crouision_year_start Have_you_ever_beart Hear_attack Hear_attack Hear_attack
		Heart_have_you_ever Chest_pain_hosptalized Have_been_hosptalized Heart_attack_2nd Croulation_last_months
'ear 1st heart attack		Heart_have_you_ever Chest_cain_hospitalized Have_been_hospitalized Heart_attack_2nd
rear 1st heart attack	New Outstoo Dispran	Heart_have_you_ever     Chest_san_hosptaized     Have_been_hosptaized     Heart_stack_2nd     Crculation_last_months     G0 List
ear 1st heart attack	New Question Diagram Diagnostic Codes Procedure Codes Switch Database	Heart_have_you_ever Chest_pain_hospitalized Have_oeen_hospitalized Heart_attack_2nd Croulation_last_months Collist
rear 1st heart attack	Diagnostic Codes Procedure Codes Switch Database	Heart_have_you_ever Chest_pain_hospitalized Have_oeen_hospitalized Heart_attack_2nd Circulation_last_months Circulation_last_

**Figure 4** QuestForm application – Formulation of a Question Diagram

perforation), inserting each of them into the research question fields in QuestForm. Using built-in search engines for ICD9 codes, Dr. Guller created the definition for each of the above-mentioned variables and defined the inclusion and exclusion criteria. The final research question was then saved as a Question Diagram and immediately submitted to Dr. Pietrobon for feasibility evaluation. Dr. Pietrobon judged that the project was feasible and could be completed using the database indicated by Dr. Guller. At this point, the project was initiated and a detailed project management plan was established using Research Manager.

#### Research manager

Research Manager is a Web application developed by our group designed to facilitate the project management of clinical research projects. Similar to QuestForm, Research Manager is licensed under the GNU Public License [10], which allows individuals to copy, modify, and freely distribute the software as long as the source code is provided.

Research Manager provides multiple features to facilitate project management of clinical research projects. All projects are displayed by category (e.g., cardiology, general surgery, etc) with a brief description. The internal content of all projects is password protected. All internal tasks within a project are assigned to individual researchers. Project administrators initially assign deadlines that can be modified by task leaders within three days. All participating members of the project receive weekly reports containing details about the activity and the latest electronic file within each task (Figure 6). These files can include research questions, data analysis files, synthesis of a literature review, and manuscript drafts. Project members can also customize the application to receive updates for every single file uploaded to Research Manager in real time if they decide to closely track the project. Expired tasks are marked in the weekly report sent to the entire team, thus providing an incentive for investigators to keep tasks within planned deadlines.

V-			CODE	×	G
NFECTIOUS AND PARASITIC DISEASES	DISEASES OF THE BLOOD AND BLOOD-FORMING ORGANS	(280) IRON DEFICIENCY ANEMIAS	(285.0) SIDEROBLASTIC ANEMIA (285.1) ACUTE	(285.21) ANEMIA IN END-STAGE RENAL DISEASE	
IEOPLASMS		[281] OTH DEFICIENCY	POSTHEMORRHAGIC ANEMIA	[285.22] ANEMIA IN	
NDOCRINE, NUTRITIONAL,		ANEMIAS	[285.2] ANEMIA IN CHRONIC	NEOPLASTIC DISEASE	
IETABOLIC, IMMUNITY IISORDES		[282] HEREDITARY HEMOLYTIC ANEMIAS	ILLNESS	[285:29] ANEMIA OF OTH CHRONIC ILLNESS	
ILOOD AND BLOOD FORMING		[283] ACQUIRED HEMOLYTIC ANEMIAS	(285.8) OTH SPEC ANEMIAS (285.9) ANEMIA, UNSPEC		
IENTAL DISORDERS		[284] APLASTIC ANEMIA			
ERVOUS SYSTEM AND SENSE		[285] OTH UNSPEC ANEMIAS			
RGANS		[288] COAGULATION DEFECTS			
IRCULATORY SYTEM		[287] PURPURA AND OTH HEMORRHAGIC CONDITIONS			
IGESTIVE SYSTEM		[288] DISEASES OF WHITE			
ENITOURINARY SYSTEM		BLOOD CELLS			
REGNANCY, CHILDBIRTH, ND THE PUERPERIUM		[289] OTH DISEASES OF BLOOD AND BLOOD-FORMING			
KIN AND SUBCUTANEOUS		ORGANS			
ND CONNECTIVE TISSUE					
ONGENITAL ANOMALIES					
CONDITIONS IN THE PERINATAL PERIOD					
YMPTOMS, SIGNS, AND ILL- DEFINED CONDITIONS					
JURY AND POISONING					
CODES					
CODES					

Figure 5 QuestForm application – Search engine for ICD9 codes

Research Manager helps identifying such problems and enables their early elimination, thus avoiding delays in the project completion. Finally, since weekly reports are generated to all participating members, Research Manager also provides peer-incentive for project members to complete their tasks in a timely manner.

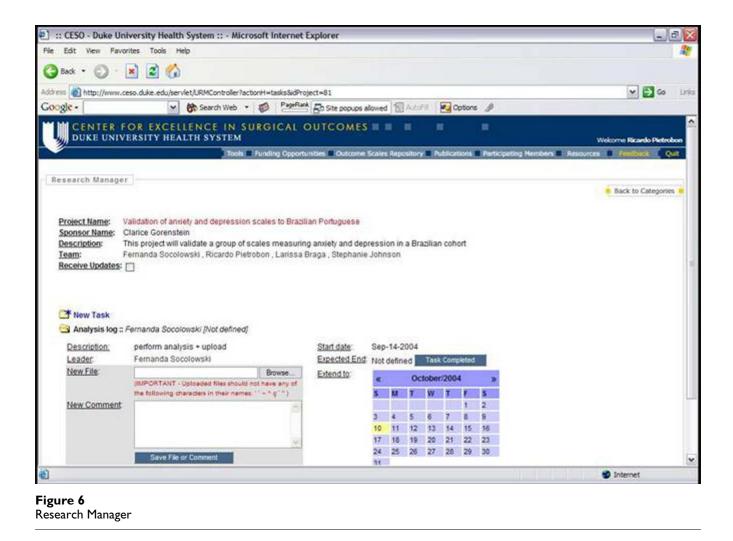
#### Outcomes research application

Once the Question Diagram was evaluated by the clinical epidemiologist, this project was transferred to Research Manager. Contact information for each of the researchers involved and deadlines for completion of the main phases of the research process were set, including data extraction and cleaning, data analysis, literature review, and manuscript writing. Weekly reports were generated to update investigators on all tasks of the project, including different versions of the statistical analysis, modifications in the research question, and manuscript sections. With an established project plan, the statistician in charge selected the best methods for analysis. Since the database is a random sample of the United States and requires special survey analysis methods, it was necessary that all involved researchers understood the statistical approach by using Analysis Charts.

#### Analysis charts

Analysis Chart is a tool designed to enhance the understanding of statistical methods to a format that is understandable by clinical researchers with different previous levels of statistical knowledge. As such, it is important in the design as well as the analysis phases of a project (Figure 7). The application was built using Extensible HyperText Markup Language (XHTML 1.0) [4] in combination with Cascading Style Sheets [11].

Analysis charts are composed by cascading links that display information about quantitative methods in progres-



sive levels of complexity called "layers of information". Each layer explains the statistical method with an increasing level of complexity. In this manner, clinicians interested in simply understanding the method to evaluate whether it can be applied to a research project can simply read the first three layers. In contrast, researchers interested in a direct application of the method to available research data can follow all layers and their respective references. Most commonly, complex techniques are presented using five layers of information. Layer 1 summarizes the general goal of the method. Layer 2 presents previous clinical applications so that the researcher can visualize situations in which the method may be realistically applied. Layer 3 describes the data requirements for the application of the statistical technique. Layer 4 describes the basic statistical underpinnings of the method, initially breaking down equations and then reassembling them. Finally, layer 5 presents a list of available software packages for the implementation of the technique as well as cases studies where all previous layers are applied to real data sets. Each layer ends with a section containing selected references that explain the topic in more detail.

#### Outcomes research application

While deciding on the most appropriate analysis strategy for the Question Diagram, Drs. Pietrobon and Guller consulted the Analysis Chart searching for the most appropriate statistical methods of analysis. Given the nature of the research question and that the NIS database has a sample design, Drs. Guller and Pietrobon opted for an approach involving multiple and logistic regression models while adjusting for sampling weights, strata, and clusters. With a defined analysis protocol, the research question was then transferred to a statistical programmer trained in the translation of Question Diagrams into statistical code. This process was closely evaluated by Drs. Pietrobon and Guller, who scrutinized the statistical code and results from a clinical and statistical perspectives. Once the

:: CESO - Duke Univ	rersity Health System :: - Microsoft Internet Explorer	- <u>0</u>
ile Edit View Favorit	ies Tools Help	
3 Back • 🕥 • 🗗	1 🖻 😚	
ddress a http://www.ces	o.duke.edu/servlet/CesoController?action?f=layers8menufium=18id=1	💌 🛃 Go 🛛 Lin
loogle -	💌 🚯 Search Web 🔹 🦪 PageRank 🖧 Site popups allowed 🗑 AutoFill 💽 Options 🥒	State of the second
	DR EXCELLENCE IN SURGICAL OUTCOMES = = = = = = = = = = = = = = = = = = =	Weicome Ricardo Pietrobon
<i>y</i>	🖬 Tools 📕 Funding Opportunities 📕 Outcome Scales Repository 🖉 Publications 🖩 Participating Member	s 🛚 Assources 🖷 Teatback 🕴 Que
Item Response The Layer 1	ory	Layers = Layer 1 = Layer 2
underlying trait measu social impairment, or responses, and mode	the study of the association between responses given to items (questions) of a scale and the degree of an red by this same scale. Examples of traits include any construct that cannot be measured directly, such as disability, depression. There are two main classes of IRT models: logistic models for items with dichotomous (yes/no) is for items with categorical response options (e.g., no, a little, a lot, all the time).	» Layer 3 » Layer 4 » Layer 5
properties can #7). This limita contrast, IRT p 2. Accounting for the possibility	e scale is population independent (Embretson, 2000 #6). In scales validated under the CTT paradigm, their vary substantially across patient populations with different degrees of severity of the underlying trait (Pietrobon, 2002 tion generates uncertainty when a scale is to be used in a population where it has not been previously tested. In roperties are invariant across populations. measurement error. CTT takes the scores obtained from scales to be taken at face value, i.e., without accounting for of measurement error. Since measurement error is inherent to any human measurement, this assumption is clearly rast. IRT provides standard errors for final scores of individual patients.	
This has sever	em. While CTT focuses on the scale as a unit, IRT focuses on the item and its association with the underlying trait, al practical applications. First, if two scales measure the same trait, IRT allows for their scores to be comparable mponent items are completely different. These same items can then be used to created custom scales or create adaptive tests.	
1		Dinternet

Analysis Chart

results were deemed to be accurate, Dr. Pietrobon started the literature review using a Literature Matrix.

#### Literature matrices

Literature matrices consist of a comprehensive but not necessarily exhaustive review of the literature focused on a narrow clinical topic (Figure 8). Each article is analyzed using the following criteria: study objectives, data sources, outcome variables, primary predictor variables, confounders, statistical analysis, results, established knowledge, and shortcomings. Each literature matrix is saved as an XHTML file that can be visualized in web browsers as well as imported in any commercial or open source spreadsheet applications such as Microsoft Excel<sup>®</sup> or Open Office Calc [12].

The advantage of the Literature Matrix as implemented in the Web suite is its availability over the web to the whole outcomes research community. This allows Literature Matrices to be constantly updated, with authors receiving their due credit in a list of contributors. Literature Matrices also enable researchers to obtain a complete summary of the literature without going through the cumbersome process of copying and reading a manuscript for the first time. In contrast, researchers' time can be spent more efficiently in reviewing what has already been compiled and attempting to expand the Literature Matrix with other relevant bibliographic references.

#### Outcomes research application

In order to evaluate the literature, a graduate student performed a thorough literature search. All relevant articles were copied, read, and the data extracted according to the established categories. Once the Literature Matrix had been completed, Drs. Guller compared the current project results with results published in the literature. The structured information in Literature Matrices also allowed Dr. Guller to compare the strengths and weaknesses of the current project in relation to previous publications. Once

9	5le Edit	t View Insert Format I	ools Data Window	Help				Type a ques	ition for help 🔹 🕳 🗗
1	8.4		an. 119.	N - 1 × 9. Σ	- 01 31 00 4	1 100%			
Irial		and the second se	and the second se		(1)课课(田)	and the second se	S		
-	F1	<ul> <li>&amp; Outcomes</li> </ul>				-			
	A	8	C	D	E	F	G	Н	1
1	No.	Reference	Objective	Data_Sources	Inclusion/Exclu sion	Outcomes	Main_Effects	Statistical_Analys is	Results
2		Factors affecting length of stay and need for rehabilitation after hip and knee arthroplasty. 1 Journal of Arthroplasty.	factors that predict the length of stay on surgical service after hip and knee	admitted to the Albany Medical Center for elective total hip/total knee	*Fracture, neoplasms excluded	"Length of stay "Need for rehabilitation	Marital status, BMI, COPD, diabetes, organic heart disease, RA, prior joint	Pearson's correlation coefficient	with increased length of stay (p=.0011) and need for discharge to rehabilitation
		Patient characteristics that affect the outcome of total hip arthroplasty: 2 a review. ?CJS, Vol. 41, K. Tang et al. total	orthopedic literature to determine the effect of patient	identified via MEDLINE from 1966 onward with keywords 109 (PC4 on so	not evaluate the effect of patient characteristics on the outcome -2 patients 12	function as defined as limitation in activity weat snee	weight, Marital status and social support, socioeconomic	Meta-analysis	?
		Knee Arthroplasty in Diabetic Patients. The Journal of Arthroplasty, 3 Vol. 16 No. 1, 2001. Kay Wastereman et al.	TO Determine the	patients in Singapore General Hospital between 1993	knees) died of unrelated causes. *2 patients were	Score 'Mean Function Score 'Incidence of UTI postoperative	Surgical unit	Pearson's correlation coefficient, Questionnaires	time (>2 hours) in diabetic patients increase risk factor for deep joint
		Patient Comorbidity: Relationship to Outcomes of Total Knee 4 Arthroplasty. Clinical	effects of medical and muscoloskeletal comorbidity on	treated consecutively with total knee arthroplasty were		outcomes 'length of stay 'total and specific hospital	*cost of implants and supplies *need for surgical	correlation coefficient, t test, analysis of variance (ANOVA).	more risk factors has significantly
		factors on Perioperative morbidity in Knee Arthroplasty. Clinical Orthopaedics and	specific preoperative and postoperative parameters	total knee endoprotheseses implanted under standard		*postoperative nonsurgical	*Concomitant Diseases *Lutz and Klose Risk Score *American	test, Mann- Whitney U test, Student's t test, and chi square	with increased number of nonsurgical perioperative
		5 Related Research, notiziman er al. Effect of Baseline Functional Status and Pain on Outcomes of Total Hip 5 Arthroplasty. The	regarding their to examine one relationship between baseline pain and functional status	conditions in 203 weacare beneficiaries receiving total hip replacement (ICD 85.51)	"Hip tracture "Bilateral replacement with an interval of one month or less	complications nevel or accurvy *presence and severity of pain with walking *need for	Society of age genuer "comorbidities "functional status and pain "pain with walking	test, log regression cm square tests, t tests, ordinary least squares, logistic regression analysis, STATA	complications Patients who rad pain with walking, who had severe pail with walking, who participated in
			To assess the	(100 00 01)	month of less		waa waalang	analysis. STATA	
•	* *\S	heet1 / Sheet2 / Sheet3 /	ananinasina anin			[<]	2		

**Figure 8** Literature Matrix

this phase was completed, Dr. Guller proceeded to the final writing of the manuscript using Output Templates.

#### Outcomes research application

At the end of the project, Dr. Guller combined the Question Diagram, Analysis Chart, and Literature Matrices to write the final manuscript. While Analysis Charts provided the information concerning the statistical techniques used in this study, Literature Matrices provided the basis for comparison of the study results against previous publications. Although not included in the final manuscript, Dr. Guller also had access to multiple analysis files through Research Manager to orient him in each of the steps taken during the research question formulation, data analysis, and literature review.

#### Use of web application by clinical researchers

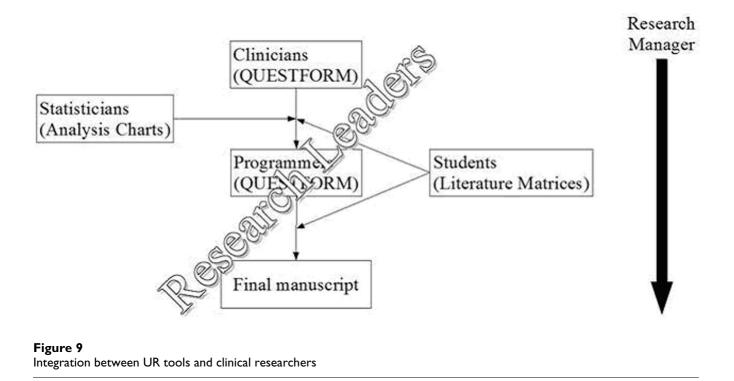
The use of web applications by individual clinical researchers can be summarized in Figure 9.

# Utility

Since the concept of streamlining the interdisciplinary collaboration preceded the existence of the current suite of web applications, different versions of the central idea have been gradually applied since the second half of 2002. Although our usability, qualitative, and economic studies to evaluate this application are still ongoing, we have noticed a significant improvement in the number and quality of our publications as evidenced by the increasing acceptance rates of our manuscripts for publication in peer-reviewed journals. The Web applications are currently used in research projects involving Duke University and other universities in the United States and abroad, where they have been shown to facilitate inter-institutional collaborations.

#### Discussion

Improving the efficiency of an interdisciplinary approach to secondary data analyses has multiple potential benefits.



These include an increase in the overall clinical significance of the final publication, decrease in the number of failed projects, decrease in the time for completion of individual projects, improvement in the education of future outcomes researchers, decrease in the cost-benefit ratio for individual outcomes research projects, and, perhaps most importantly, an automation of repetitive tasks. This last factor is crucial since eliminating repetitive tasks will allow researchers to concentrate on the design of innovative projects [2].

Surprisingly, few systems and applications have been described to solve the problem of complex interdisciplinary collaboration between clinicians and statisticians. Isolated approaches have usually focused on specific portions of the outcomes research process, without attempting to integrate them into a cohesive system. For example, Marshall [13] has proposed the use of a secure Internet web site for collaborative medical research and data collection. While this system seems to achieve its proposed objectives, it does not improve the process of guiding teams in the translation of data into useful clinical information. Other systems have approached the process in a more comprehensive manner. The Research Toolbox [14], for example, is a software application that combines databases for literature searches in addition to providing templates for the scientific output. The system is applicable to any type of research, but lacks the ability to connect researchers over the World Wide Web. It also does not address the formulation of research questions from existing databases, selection of statistical techniques, exchange of manuscripts, or project management. Finally, the Webbased Medical Information Retrieval System (WebMIRS) project, funded by the National Library of Medicine [15], allows researchers not only to evaluate the database content but also to perform the data extraction of specific subsets of the data set. In spite of its high performance as a research tool, WebMIRS is currently restricted to one single publicly available database (National Health and Examination Survey – NHANES), and does not contribute to other phases of the outcomes research process.

Although this newly developed system of applications provides a significant improvement in the way secondary data analyses are conducted, it still has limitations. First, because of the lack of a formal evaluation of the effectiveness of this system, we are unable to quantify its real time and cost saving benefits. Second, the system is currently restricted to secondary analyses and does not allow for the planning of prospective data collection. Although one of the main advantages of formulating a research question based on existing data sets is the bounded nature of the process, future applications should attempt to create rules and algorithms that may guide prospective data collection.

## Conclusion

In summary, we have experienced that this system has significant advantages over the traditional manner of conducting outcomes research based on secondary data analyses. This tool may have important applications, not only resulting in an improvement in the overall efficiency of the outcomes research process, but also affecting the way new outcomes researchers are trained and introduced to a research environment.

## Availability and requirements

The Web application is available at <u>http://</u> www.ceso.duke.edu

#### **Abbreviations**

XHTML Extensible HyperText Markup Language 1.0

JAVA: by Sun Microsystems

ICD9-CM: International Classification of Diseases, Ninth Revision - Clinical Modification

NIS: Nationwide Inpatient Sample

GNU: GNU's Not linux General Public License

WebMIRS: Web-based Medical Information Retrieval System

NHANES: National Health and Examination Survey

**UR: Uniform Resource** 

#### **Competing interests**

The author(s) declare that they have no competing interests

#### **Authors' contributions**

Ricardo Pietrobon - design, manuscript drafting

Ulrich Guller - design, manuscript revision for important intellectual content

Henrique Martins - design, manuscript revision for important intellectual content, software programming

Andreia P Menezes - design, manuscript revision for important intellectual content

Laurence D. Higgins - design, manuscript revision for important intellectual content

Danny O. Jacobs - design, manuscript revision for important intellectual content

## Acknowledgements

We thank Juliana and Guilherme Serzedello for assistance with manuscript formatting.

#### References

- Ι. Moses L, Louis TA: Statistical consulting in clinical research: the two-way street. Stat Med 1984, 3(1):1-5.
- Kong SX, Wertheimer Al: Outcomes research: collaboration 2. among academic researchers, managed care organizations, and pharmaceutical manufacturers. Am J Manag Care 1998, 4(1):28-34.
- Cheung YB, Tan SB, Khoo KS: The need for collaboration 3. between clinicians and statisticians: some experience and examples. Ann Acad Med Singapore 2001, 30(5):552-5
- 4. W3C, XHTML 1.0 The Extensible HyperText Markup Language (Second Edition) [http://www.w3.org/TR/xhtml1/]
- 5. MySQL [http://www.mysql.com/]
- Long KH, Bannon MP, Zietlow SP, Helgeson ER, Harmsen WS, Smith 6 CD, Ilstrup DM, Baerga-Varela Y, Sarr MG: A prospective randomized comparison of laparoscopic appendectomy with open appendectomy: Clinical and economic analyses. Surgery 2001, 129(4):390-400
- Maxwell JG, Robinson CL, Maxwell TG, Maxwell BG, Smith CR, Brinker CC: Deriving the indications for laparoscopic appen-7. dectomy from a comparison of the outcomes of laparoscopic and open appendectomy. Am J Surg 2001, 182(6):687-692.
- The Healthcare Cost and Utilization Project Nationwide 8. Inpatient Sample. Release 6. 1997.
- 9. The Ultimate Team Organisation Software (TUTOS) [http:// /www.tutos.org/]
- 10. GNU General Public Licenses [http://www.gnu.org/licenses/ 
   licenses.html]

   II.
   W3C, Cascading Style Sheets [http://www.w3.org/Style/CSS/]
- OpenOffice.org 1.02 [http://www.openoffice.org/] 12.
- 13. Marshall WW, Haley RW: Use of a secure Internet Web site for collaborative medical research. Jama 2000, 284(14):1843-9.
- 14. Ltd., M.P., Research Toolbox System [http://www.research toolbox.com/]. MCI (Pty) Ltd. Web site no longer available on October 10, 2004.
- 15. Web-based Medical Information Retrieval System [http:// archive.nlm.nih.gov/proj/webmirs/index.php]

## **Pre-publication history**

The pre-publication history for this paper can be accessed here:

http://www.biomedcentral.com/1471-2288/4/29/prepub

