

USE OF REAL WORLD DATA: CHALLENGES IN THE ANALYSIS OF PATIENT REGISTRY DATA

Moderator: Chris Blanchette, PhD

Outline

- Introduction (Blanchette)
- The Severe Chronic Neutropenia International Registry (SCNIR): A Case Presentation (Kamble)
- General Overview of an Analysis Plan (Exuzides / Gemmen)
- Common Statistical Pitfalls in the Analysis of Registries (Ekwueme / Gibbons)
- Discussion



Patient Registry Special Interest Group

Team 3: Analysis

Classification, Strategy & Design Working Group

Team 3: Analysis Members

- Co-Chair: Shital Kamble MS, University of North Carolina - Charlotte
- Co-Chair: Christopher Blanchette PhD, Lovelace Respiratory Research Institute
- Eric Gemmen MA, Quintiles, Inc.
- Carl Gibbons BSc, Schering-Plough Ltd
- Alex Exuzides PhD, ICON Clinical Research
- Donatus Ekwueme PhD, US Centers for Disease Control and Prevention
- Joanna Lis PhD MSc, Sanofi-Aventis
- Anuprita Patkar PhD, ETHICON-A Johnson & Johnson Company
- Matt Reaney MSc, AHP Research
- Maznah Dahlui MD, MPH, University of Malaya

Tasks

- Support the development of a taxonomy of terms applicable to registries
- Support the development of a best practices document for working with registries

Statistical Analysis Plan

Describes the statistical techniques that will be used to evaluate the objectives of an observational study that uses registry data

Achievements

Identified 13 broad categories of terms

Including...

- Power/Sample Size Calculations
- Statistical Inference and Hypothesis Testing
- Timing of Analyses
- Available Variables for Analyses
- Main Analysis Techniques
- Treatment of Missing Observations
- Multiplicity Adjustments

Achievements

70 analysis terms defined

Including....

- Elixhauser Comorbid Disease Adjustment Method
- Ordinal Logistic/ Logit and Ordinal Probit Models
- Cox Proportional Hazards Regression Models
- Two-part Models
- Multilevel Models
- Propensity Score Methods
- Instrumental Variables
- Clinical Significance
- Statistical Significance
- Standard Errors
- Confidence intervals

The Severe Chronic Neutropenia International Registry (SCNIR): A Case Presentation

- Shital Kamble, MS

Outline- SCNIR

- Background
- Objectives
- Patient Enrollment
- Variables Available for Analyses
- Significance of SCNIR

SCNIR: Background

- SCNIR- organization dedicated to improve understanding and treatment for diseases causing severe neutropenia (absolute blood neutrophil level < $0.5 \times 10^9/L$) for months and/or years
- 1987: first clinical trial with the hematopoietic growth factor G-CSF (granulocyte - colony stimulating factor) was initiated
- March 1994: SCNIR established
- Clinical trial data from 1987-1994 were retrospectively transferred to the Registry and were added to the data of newly diagnosed SCN patients from 1994 onward
- 1994 – 2000: SCNIR was funded by Amgen Inc. for the collection of safety data on G-CSF (filgrastim) treatment annually reported to the FDA

SCNIR: Background

- Since 2000: data collection was expanded, but financial support of European part was dramatically decreased:
 - new neutropenia sub-diagnoses included
- Principal coordinating centers-
 - ✓ Seattle, WA, USA
 - ✓ Hannover, Germany

SCNIR's Administrative and Operational Structure

- 5 components:
 - ✓ an international Scientific Advisory Board of 10 physicians/hematologists;
 - ✓ a panel of European local liaison physicians who treat neutropenic patients;
 - ✓ the Registry data coordinating offices- responsible for data collection;
 - ✓ Amgen's International Clinical Safety Department; and
 - ✓ a Safety Monitoring Committee, which comprises 3 physicians and 2 scientists who address the risks and benefits of G-CSF therapy with regard to the hazard of MDS/AML arising during treatment.

Objectives of SCNIR

- document information on the natural course of SCN and its response to treatment with the blood-stimulating factor called granulocyte colony stimulating factor or G-CSF (Neupogen)
- determine incidence and outcome of various clinical events associated with SCN and its treatment including:
 - (a) osteoporosis, (b) vasculitis, (c) glomerulonephritis, (d) splenomegaly/hepatomegaly, (e) cytogenetic abnormalities, (f) myelodysplastic syndrome, and (g) leukemia
- evaluate the outcome for pregnancies in SCN patients, including patients receiving various therapies for neutropenia
- collect information and evaluate effectiveness of transplantation of blood forming cells and other treatments for SCN

SCNIR Patient Enrollment

- March 1994 through August 2004: 335 -1163 patients in 35 countries enrolled in the Registry
- Enrollment - primarily through information submitted on the standardized forms by the referring physician and reviewed by expert physicians of the Advisory Board for the SCNIR

Patients by Country*

Country	Patients	Country	Patients	Country	Patients
Germany	111	Switzerland	5	Australia	47
United Kingdom	60	Turkey	5	Canada	35
Sweden	26	Poland	4	Chile	12
Belgium	23	Ireland	4	Cuba	1
Spain	19	Czech Republic	3	Costa Rica	1
Norway	13	Serbia	2	New Zealand	1
Austria	12	Luxembourg	2	Japan	1
Netherlands	11	Denmark	1	Brazil	1
Italy	11	Portugal	1	Puerto Rico	4
Greece	10	Russia	1	United States	685
Israel	10	Honduras	1	TOTAL	1163
France	6	Morocco	1		

*As of August 1, 2004 (Dale et al., 2005); European enrollment= 342 patients

Enrollment by Diagnostic Category*

Diagnosis	# of Patients
Congenital, Total	526
Severe congenital neutropenia	422
Glycogen storage disease	42
Shwachman-Diamond	37
Barth	10
Myelokathexis	8
Immunodeficiency	7
Cyclic	205
Idiopathic	349
Autoimmune	68
Other	15
TOTAL	1163

*As of August 1, 2004 (Dale et al., 2005)

Variables Available for Data Analyses

- All data collected by the SCNIR are analyzed periodically (usually annually)
- Patient Demographics (age, race/ethnicity, sex, city, state, country, etc.)
- Clinical Characteristics
 - Type and frequency of infectious episodes
 - Type and frequency of non-infectious clinical events
 - Analysis of physical assessments
- Treatment
 - Applied therapies
 - Side-effects of treatment
 - Treatment safety
 - Non-Responsiveness
- Secondary Events
 - MDS/leukemia
 - Bone marrow transplant
- Death

SCNIR- Significance

- SCNIR serves as a comprehensive resource for educating physicians, patients and their families by providing the most up to date information on the natural history of SCN and its treatment options
- SCNIR serves as a model for expanding our understanding on the causes, consequences, and treatment of rare diseases

The Patient Registry SAP:

General Overview and Analysis Considerations

- Alex Exuzides, PhD
- Eric Gemmen, MA

Analysis Core Questions

- Are the objectives/hypotheses predefined or *post hoc*?
- What is the patient population?
- What data are collected?
- What statistical methods are used?
- How are missing data treated?

Purpose of the Registry

- Not all registries have clear *a priori* hypotheses
- Some registries evolve from an effort to discover associations
- Other registries pursue only certain preliminary findings
- Important to disclose how the registry is developed

Patient Population

- Target population:
 - The population to which the study findings are meant to apply
- Assessable/Evaluable population:
 - Subset of the target population available for the study
- Indented population:
 - Subset of the accessible population sampled according to the registry design
- Actual population:
 - Patients who actually participate in the registry

Data Collection

- Patient characteristics that do not change
 - Demographics
- Baseline patient characteristics
 - Clinical measurements
 - Patient-reported outcomes
- Follow-up clinical and patient-reported data
 - Not necessarily at equal intervals
 - Not necessarily at pre-defined times
- Health care resource utilization
 - Hospitalizations
 - ER visits
 - Physician visits
 - Laboratory tests

Data Analysis

- Common descriptive statistics
 - Mean, Median
 - Proportions, Rates
 - Incidence, Prevalence
- Common statistical tests and models
 - t-test
 - Wilcoxon Rank-Sum test
 - Chi-square test
 - Generalized Linear Models (GLM)
 - Generalized Estimating Equations (GEE)
 - Cox-Proportional Hazards

Handling of Missing Data

- MCAR
- MAR
- MNAR

Multiplicity Adjustment

- Multiplicity is the potential increase in Type I Error that occurs when statistical tests are used repeatedly (Tukey 1977)
 - If n independent comparisons are performed, the familywise error (FWE) rate is given by:
$$1-(1-\alpha_{\text{per comparison}})^n$$
- Multiple endpoints
- Multiple cohorts/treatment arms
- Multiple looks at the data (interim analyses)
- Multiple subgroup analyses

Multiplicity Adjustment

- Regulatory guidelines on use of PRO to support label claims recommend adjusting for multiple comparisons in PRO (FDA 2006; EMEA 2005)
- Bonferroni's method (Tukey et al. 1985)
- Hierarchical testing (Westfall et al. 2001)
- Hochberg's method (Hochberg 1988)

Statistical Issues: Common Pitfalls

- Carl Gibbons, BSc
- Donatus Ekwueme, PhD

Statistical Issues: Common Pitfalls

- Statistical significance?
- Matching
- Subscriptions and data access
- Data management challenges
- Software issues

Statistical significance?

- What if your registry is very large?
- Jumbo registries have capacity to report miniscule p-values even on unremarkable effects.
- Is disproving H_0 truly your priority?
- Consider focusing your discussion around clinical significance
- Since any effect you find in a huge registry is probably also present in the population (assuming no confounding), the 'so what factor' gains prominence.

Matching

- Why do matching?
- Selecting Control- common techniques
- Temptation of over-matching
 - Practical concerns
 - Statistical concerns

Subscriptions and data access

- Subscriptions to many datasets are annual, some are paid!
- Receiving updates is rarely simple
 - Even in 'professional' registries, minute changes to data layout and format are common
 - "Full" updates are rarely well documented, so differences in results obtained from old/new data can be difficult to explain accurately
 - Substantial time must be allocated to importing and re-checking the database post update
- Timing of updates is rarely perfect (usually around deadlines!)

Data management challenges

- Large registries (>100GB) gaining popularity
 - Storage capacity increases every year
 - Even desktop PCs capable of analyzing such data
- Programming challenges prioritize reducing the data to a manageable size
- Big server setups can handle such volumes but are costly and tricky to set up & update
- Desktop PCs can manage if you slice the data into "bite-size" chunks of 1GB or less.
- Consider developing small "practice datasets" on which to test your code before committing to a full analysis run.

Software issues

- Tab-delimited text file with an accompanying, clear dictionary is the fail-safe format for data exchange
- Smaller registries may deliver their data via "prosumer" platforms such as Access, Excel, MS SQL Server.
- Such platforms are usually employed by providers to improve "referential integrity", speed up data entry, or make the job of exchanging data with clients easier.
- If run on ordinary desktop PCs, these programs start to falter as data volume increases.
- Version 'upgrades' of these platforms may be associated with additional cost to the client if they don't have the latest version of Access, etc.

