Probabilistic Asthma Case Finding A pilot study using the CHICA System

Vibha Anand, MS[1]
Stephen M Downs MD, MS[1,2]

[1] Children's Health Services Research, IU School of Medicine
[2] Regenstrief Institute Inc.
Indiana, USA

Introduction

- In pediatric primary care, eligible preventive services may exceed time available in a visit
- Decision support systems must prioritize what reminders are most important
- Bayesian belief networks provide a strategy for assessing patient risk and selecting reminders

Challenge

- Bayesian belief networks are acyclic directed graphs
- Nodes are random variables and arcs represent probabilistic dependencies
- “Learning” the probabilistic relationships can be challenging
- Expert judgment may be necessary where data are not available

Study Objective

- Compare the performance of two Bayesian belief networks for identifying children who will develop asthma
  - Network A developed by a clinical expert
  - Network B mined from data by a learning algorithm
- Probabilities for both networks were derived from data

Methods – Model Development

- Domain expert model:
  - Developed by a pediatrician trained in informatics and decision sciences
  - Using Netica Software (http://www.norsys.com)
- Mined model:
  - Directed Acyclic Graph (DAG) derived from clinical dataset
  - Using WinMine Toolkit software (http://research.microsoft.com/~dmax/WinMine)
- Both models trained on clinical data using the Netica software

Data Source

- 16,000 cases were extracted from the Regenstrief medical record system (RMRS)
- Cases randomly split
  - 11,000 training cases
  - 5000 test cases
- An additional 2000 cases collected prospectively to evaluate the model
  - These data were incomplete (missing values)

Variables in the Dataset

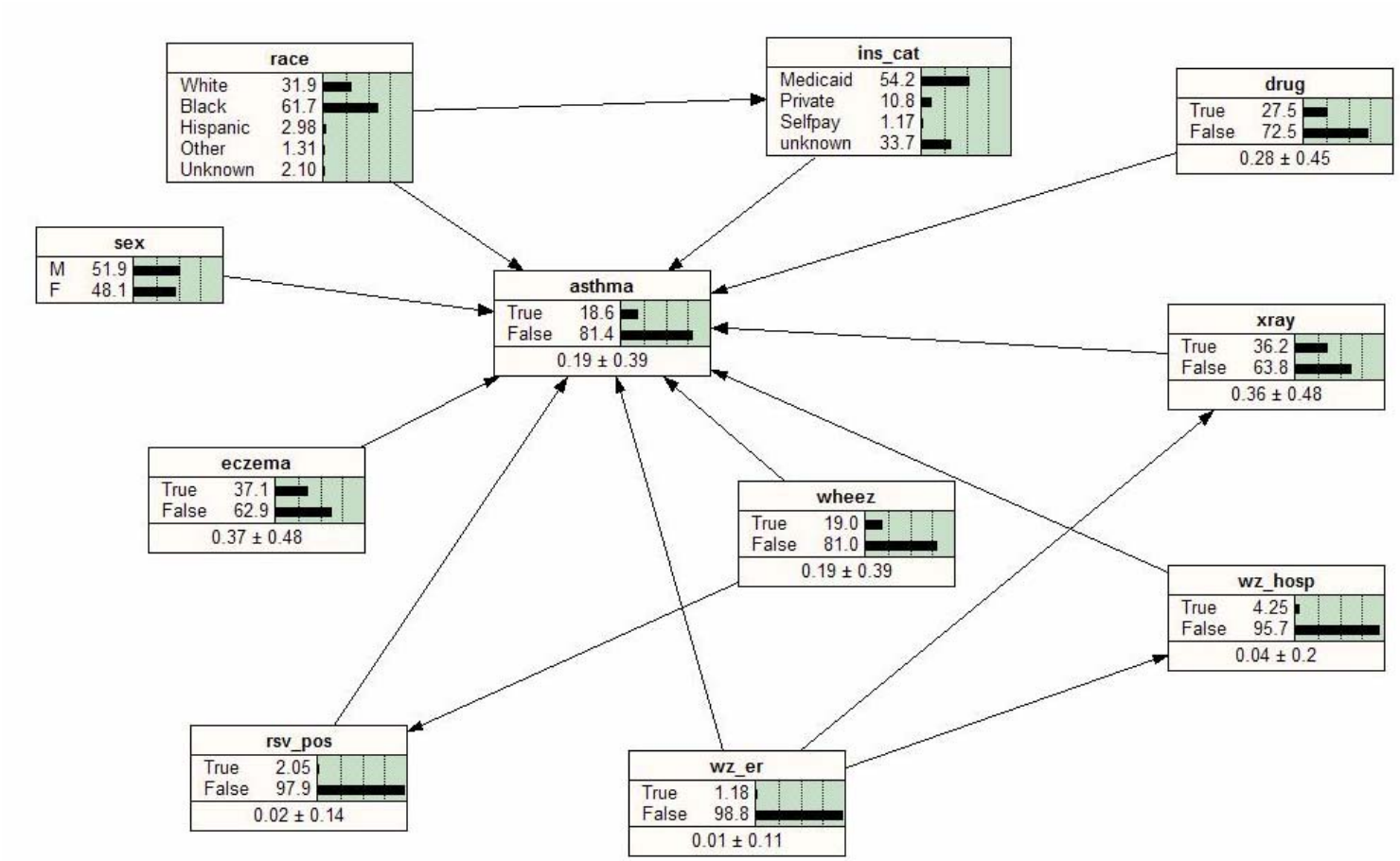
- Predictors measured before age 2 years
- Outcome asthma after 5 years

Variable	Values
Race	White, Black, Hispanic, Other, Unknown
Sex	Male, Female
Eczema	True, False
Wheeze	ICD9 or clinic billing diagnosis before age 2 (True, False)
Asthma	ICD9 (493.*) or any clinic billing diagnosis after age 5 or at least 3 drugs from a specified list within 12 months after age 5 (True, False)
X-ray	Chest x-ray before age 2 (True, False)
Drug	Drugs from a specified list before (True, False)
Wz_hosp	Inpatient admission with hospital ICD9 as wheezing (True, False)
Wz_er	Any ER visit with billing ICD9 as wheezing (True, False)
Ins_cat	Insurance category - first available insurance in the same year of the first wheezing diagnosis (True, False)

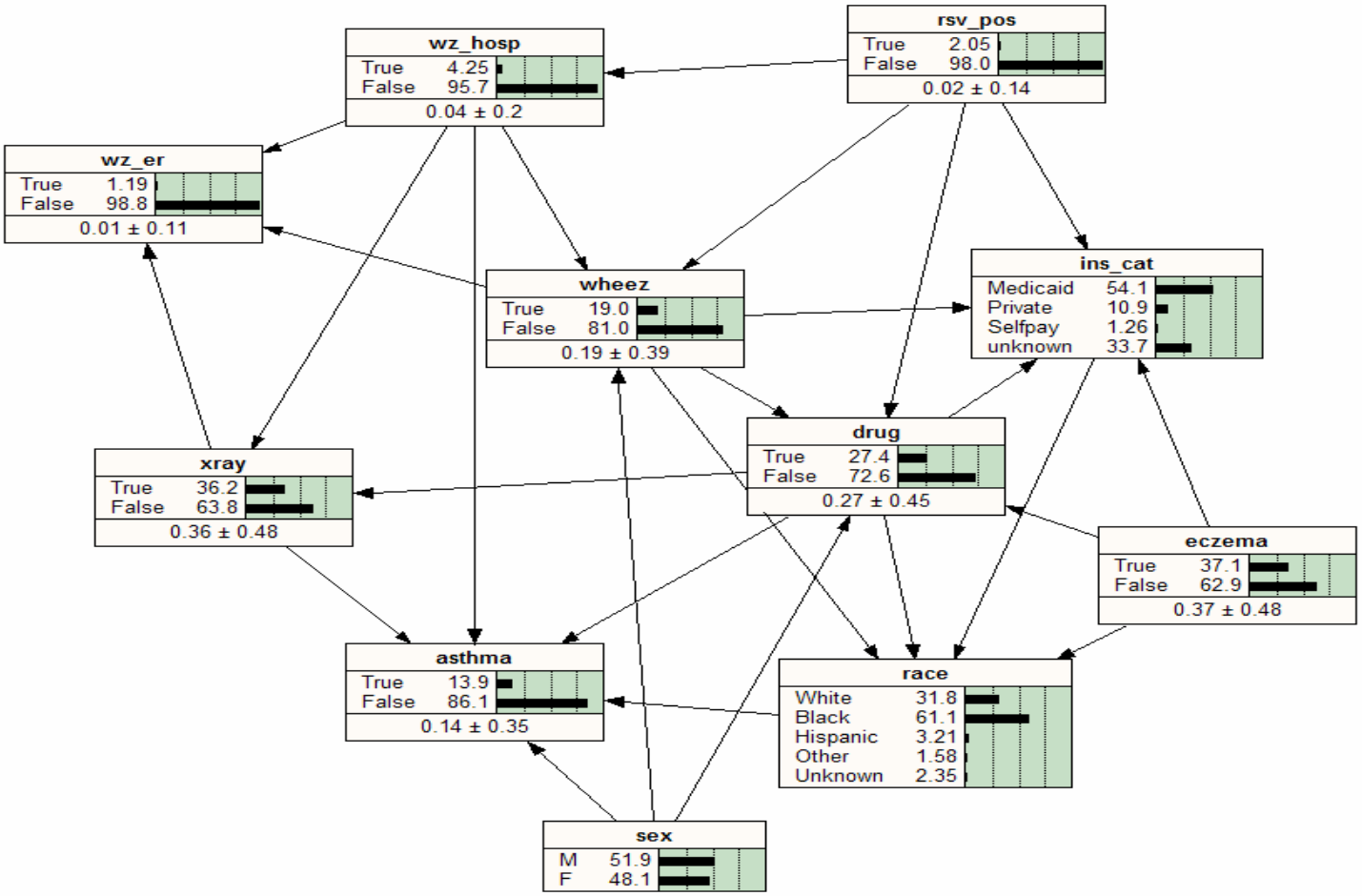
Evaluation

- Bayesian belief networks were trained on 11,000 training cases
- Ability of each network to discriminate children who develop asthma evaluated
  - 5000 test cases
  - 2000 prospective cases
- Power to discriminate was compared using receiver operator characteristic (ROC) curves

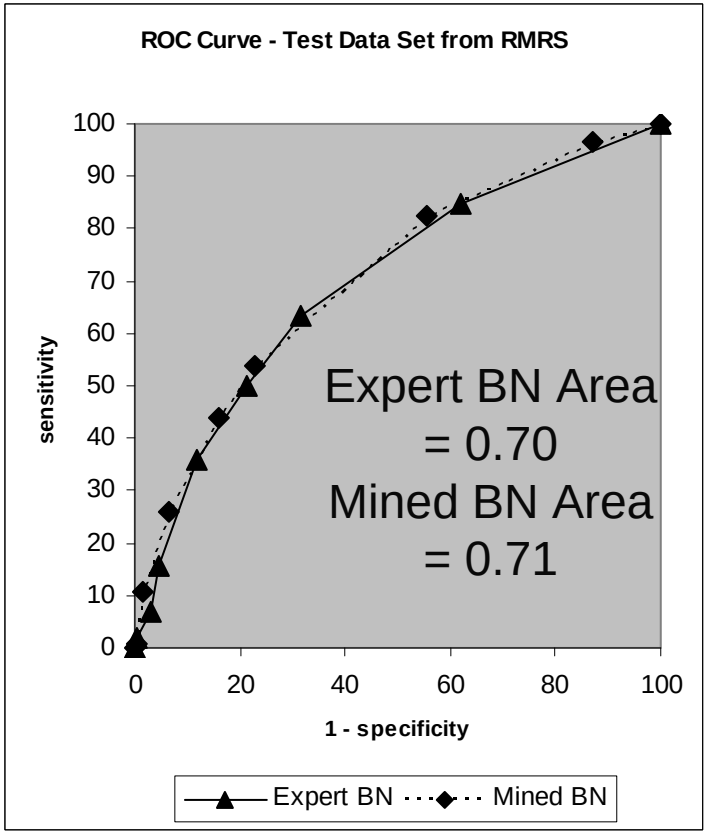
Belief Network Model (Domain Expert)



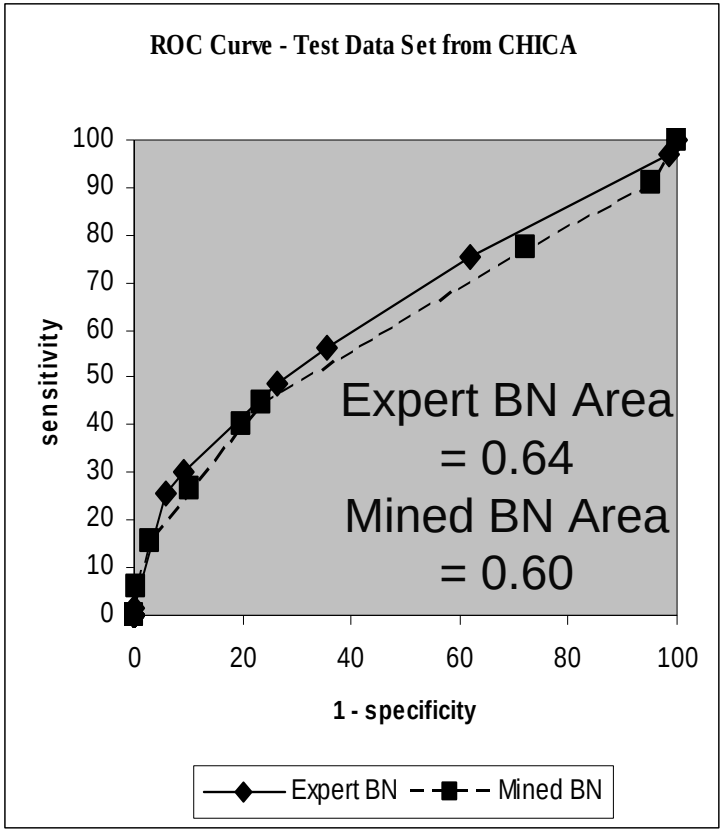
Belief Network Model (Mined from Data)



Results



ROC curve derived from 5000 test cases



ROC curve derived from 2000 cases collected prospectively

- Data Mined model had similar discriminative power to Domain Expert model.
- Slightly degraded performance when tested with prospectively collected data - data were incomplete.

Acknowledgements

- Drs. Downs, Biondich, Carroll
- Primary Care Clinic, Wishard Hospital, Indiana University
- Grant Support - NLM (1 K22 LM009160-01), AHRQ -04-0015, and MCHB (U22MC06969) for development of CHICA system

References

- Downs, S.M. and H. Uner, *Expected value prioritization of prompts and reminders.* Proc AMIA Symp, 2002: p. 215-9.
- Biondich, P.G., et al., *Using adaptive turnaround documents to electronically acquire structured data in clinical settings.* AMIA Annu Symp Proc, 2003: p. 86-90.
- Biondich, P.G., et al., *A modern optical character recognition system in a real world clinical setting: some accuracy and feasibility observations.* Proc AMIA Symp, 2002: p. 56-60.
- Anand, V., et al., *Child Health Improvement through Computer Automation: the CHICA system.* Medinfo, 2004. 11(Pt 1): p. 187-91.
- Downs, S.M., et al., *Human and System Errors, Using Adaptive Turnaround Documents to Capture Data in a Busy Practice. .* Proc AMIA Symp, 2005.
- Biondich, P.G., et al., *Automating the recognition and Prioritization of Needed Preventive Services: Early Results from the CHICA System.* AMIA Annu Symp Proc., 2005: p. 51-5.
- Jenders, R.A., et al., *Medical decision support: experience with implementing the Arden Syntax at the Columbia-Presbyterian Medical Center.* Proc Annu Symp Comput Appl Med Care, 1995: p. 169-73.
- McDonald, C.J. and et al, *The Regenstrief Medical Record System: 20 years of experience in hospitals, clinics, and neighborhood health centers. MD Computing.* 1992. 9(4): p. 206-17.
- Downs, S.M., J. Arbanas, and L. Cohen, *Computer supported preventive services for children. in Nineteenth Annual Symposium on Computer Applications in Medical Care. New Orleans, LA*. 1995, Hanley & Belfus, Inc.
- Pearl, J., *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. 1988.
- *Netica. Application for Belief Networks and Influence Diagrams. Norsys Software Corp.* 1997.
- Chickering, D.M., *The WinMine Toolkit - MSR-TR-2002-103.* 2002.
- Hanley, J.A. and B.J. McNeil, *The meaning and use of the area under a receiver operating characteristic (ROC) curve.* Radiology, 1982. 143(1): p. 29-36.

Contact Details

Vibha Anand: vanand@iupui.edu

Stephen M Downs: stmdowns@iupui.edu

1-317-278 0552