



MEDIZINISCHE UNIVERSITÄT  
INNSBRUCK

# REGRESSION TREE CONSTRUCTION BY BOOTSTRAP: MODEL SEARCH FOR THE AUSTRIAN DRG-SYSTEM

---

ISCB Prague, 23-27 August 2009

**Thomas Grubinger, Conrad Kobel, Karl-Peter Pfeiffer**

Department for Medical Statistics, Informatics and Health  
Economics

Innsbruck Medical University

[thomas.grubinger@i-med.ac.at](mailto:thomas.grubinger@i-med.ac.at)



# Introduction

## Background

- DRG-systems are used to allocate resources to hospitals
- Regression trees [Breiman et al. 1984] can assist to model DRG-systems

## Problems

- Best models are often very complex
- Models are not always medically reasonable
- Manual performance degrading adaptations are necessary

## Proposed Solution

- Build diverse accurate trees to choose from
- Use of Bootstrap Bumping [Tibshirani & Knight 1997]



# The Austrian DRG-System

## Background

- In 1997 an Activity-based hospital financing system was introduced
- Resources are allocated to hospital based on their individual cases
- Basis is a 3-step classification procedure that result in homogenous patient groups (*LDF*)

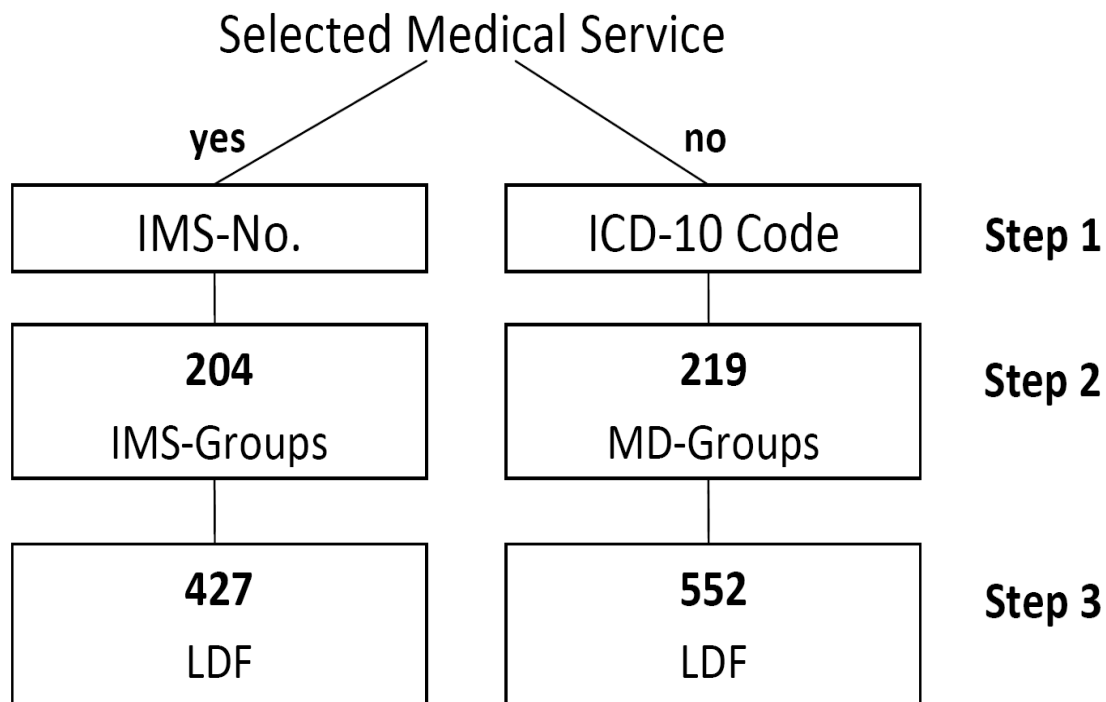
## Aim

- Consolidate rapidly increasing costs by reducing LOS
- Reduce costs by substituting inpatient care through ambulatory care
- Increase transparency of the hospitals' activities



# The Austrian DRG-System

## 3-Step Classification Procedure



**IMS** ... Individual Medical Services

**ICD** ... International Classification of Diseases

**MD** ... Main Diagnosis

**LDF** ... Leistungsorientierte Diagnosefallgruppen  
(Performance-oriented diagnoses groups)



# Regression Trees

## Recursive Partitioning

1. Examine every allowable split on each predictor variable.
2. Select and execute the split that minimizes the impurity measure in the nodes (e.g.: the MSE).
3. Recursion stops when no significant improvement can be achieved

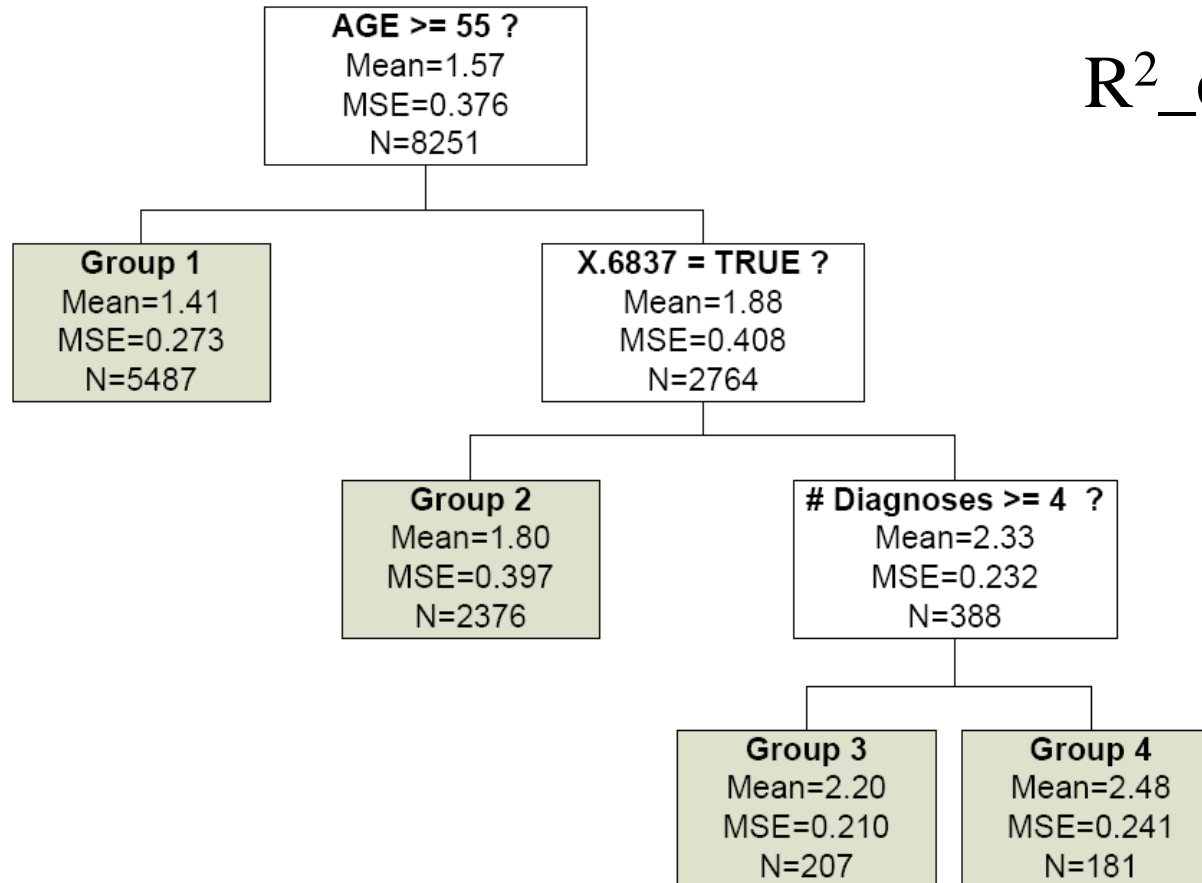
## Pruning after Recursive Partitioning

- Constructed trees are often too big
- Trees are usually pruned to smaller size



# Regression Trees

Example:



$R^2_{\text{example}} \approx 0.31$



# Bootstrap Bumping

## Basic Idea

- Diverse models are build on different bootstrap samples
- Solutions are single trees

## Algorithm

1. A set of bootstrap samples  $z^* = (z^{*1}, z^{*2}, \dots, z^{*B})$  is drawn from the training set  $z$
2. Models are fit to each bootstrap sample giving prediction  $\hat{f}^{*b}(x)$  each bootstrap  $b = (1, 2, \dots, B)$
3. The best trees are selected based on their performance on the original training-set  $z$
4. The best trees are compared to the standard CART algorithm on the test-set



# Evaluation

## Evaluation over Different Tree Sizes

- Predictive accuracy is estimated using 10-fold CV
- Use of tree sizes ranging from 2-16 internal nodes
- Final tree size can be chosen manually based on statistical accuracy as well as on economic and medical consideration

## Evaluation Criteria

- Compare the best bootstrapped based tree with the standard CART tree
- Assess the number of bootstrap-based trees that are more accurate than the standard CART tree
- Compare the accuracy of different tree complexities



# Evaluated Data-Sets

DATA-SET	DESCRIPTION	POPULATION SIZE	#VARIABLES
HDG0106	Parkinson's disease	6155	114
HDG0202	Malignant neoplasms	3933	55
HDG0304	Eye-diagnoses	9067	41
HDG0502	Acute Affections of the respiratory tract and middle atelectasis	8251	100
MEL0101	Interventions on the skull	875	60
MEL0203	Small interventions in connective tissue and soft tissue	17268	58
MEL0401	Interventions on the outer and middle ear, designed to treat a liquorrhoe	4102	44
MEL0501	Interventions on the esophagus, stomach and diaphragm	3432	86

Table 1: Description of the evaluated data-sets



# Results: Improvement of the Best Bootstrapped Tree

TREE SIZE	HDG0106	HDG0202	HDG0304	HDG0502	MEL0101	MEL0203	MEL0401	MEL0501	AVG
2	0.00	1.12	2.55	0.71	1.20	3.74	3.34	-1.52	<b>1.39</b>
3	0.00	2.78	3.33	1.65	5.96	1.88	3.92	-1.97	<b>2.19</b>
4	-0.36	5.57	3.52	1.23	5.77	3.30	4.28	-1.05	<b>2.78</b>
5	0.42	3.18	3.85	2.30	7.43	0.26	3.81	-0.84	<b>2.55</b>
6	-0.24	4.38	5.47	1.13	9.65	12.03	2.33	4.41	<b>4.90</b>
8	-0.11	6.05	1.75	1.15	1.06	12.91	2.67	3.63	<b>3.64</b>
10	-0.06	3.99	3.16	0.69	-2.93	5.09	1.94	2.83	<b>1.84</b>
12	-0.42	4.14	3.24	1.75	2.89	1.61	1.24	4.95	<b>2.43</b>
14	-1.87	3.35	1.82	1.20	-0.36	0.00	2.15	2.17	<b>1.06</b>
16	-0.76	2.11	2.52	1.27	1.38	1.18	1.89	0.65	<b>1.28</b>

Table 2: Relative average improvement of the best bootstrapped tree compared to the standard CART tree.



# Results: Diversity of Candidate Models

TREE SIZE	HDG0106	HDG0202	HDG0304	HDG0502	MEL0101	MEL0203	MEL0401	MEL0501	MEDIAN
2	0	0	0	3	1	0	0	4	<b>0</b>
3	0	5	0	0	9	0	38	3	<b>1.5</b>
4	0	17	6	1	25	4	43	0	<b>5</b>
5	2	20	26	20	38	7	9	1	<b>14.5</b>
6	1	18	34	37	52	11	32	0	<b>25</b>
8	15	18	36	18	5	6	10	13	<b>14</b>
10	2	14	76	5	2	68	19	32	<b>16.5</b>
12	14	8	13	14	2	162	1	14	<b>13.5</b>
14	8	1	2	42	70	100	3	1	<b>5.5</b>
16	5	15	5	107	0	157	7	1	<b>6</b>

Table 3: Number of bootstrapped based trees with improved accuracy compared to the standard CART tree.



# Results: Example HDG0502

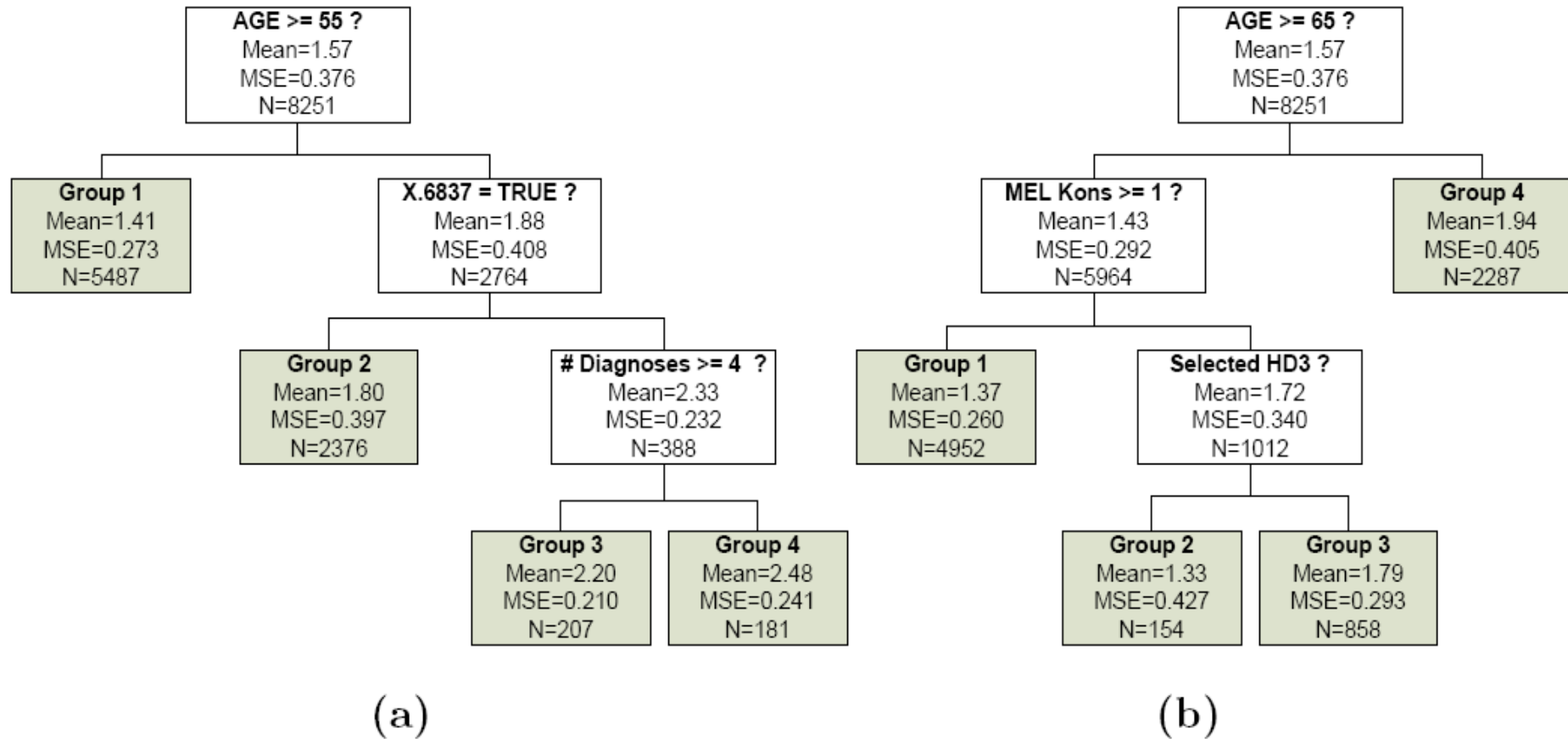


Fig. 1: Two different trees constructed by bumping from the HDG.0502 data.



# Results: Accuracy of Models with Different Complexities

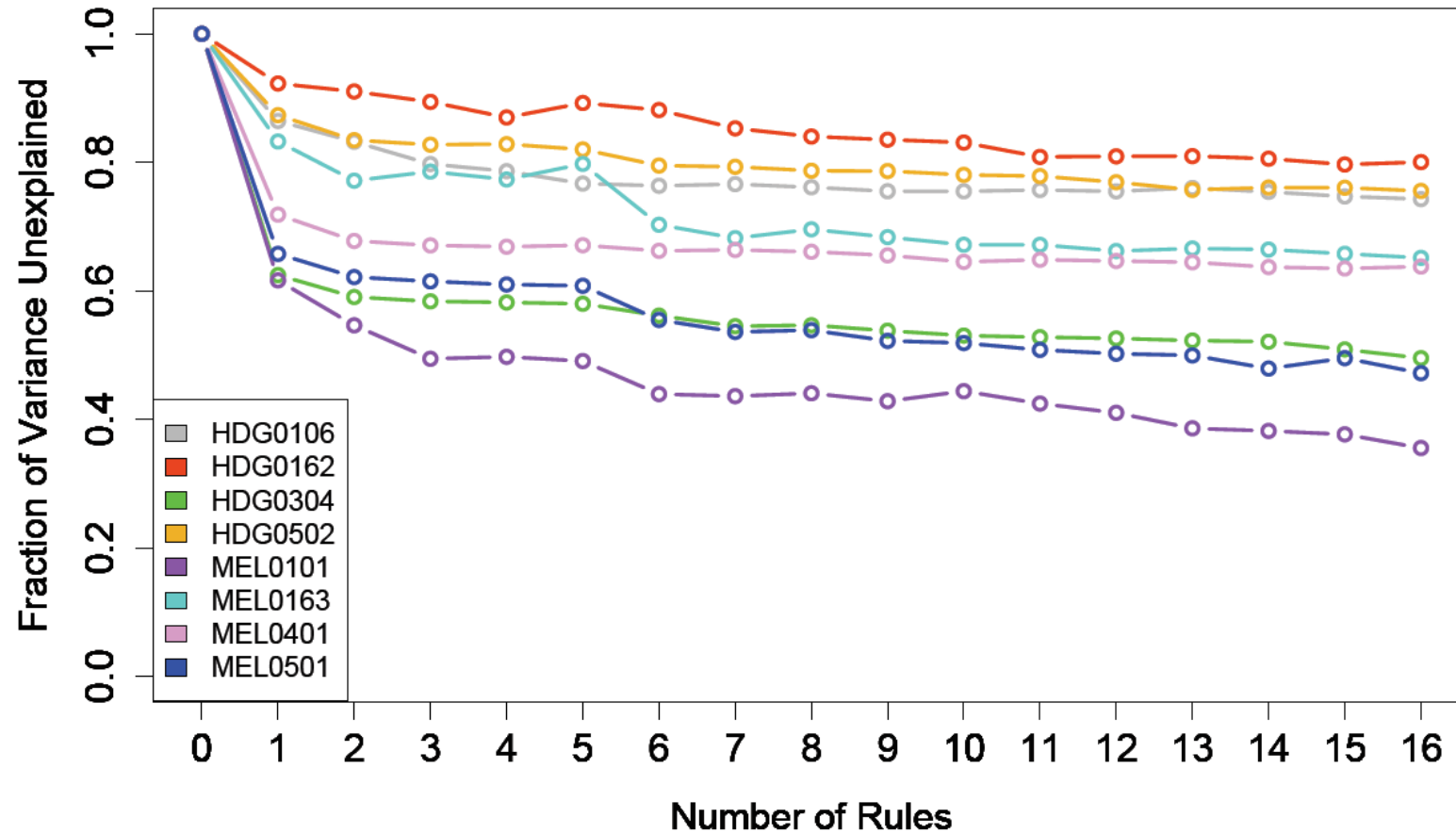


Fig. 2: Reduction of the MSE obtained by the best bootstrapped tree for different tree sizes



# Conclusions

- Bootstrap bumping can be used to find diverse and accurate trees
- Often smaller models are found that offer the same accuracy as more complex trees built by the standard CART algorithm
- Medical domain experts can choose among different models
- Manual performance degenerating adoptions of models are minimized
- In the future alternatives to bootstrap bumping will be evaluated



# References

1. Tibshirani R, Knight K. Model Search by Bootstrap "Bumping". *Journal of Computational and Graphical Statistics* 1999; 8(4): 671–86.
2. Breiman L, Friedman JH, Olshen RA, Stone CJ. *Classification and Regression Trees*. Belmont: Wadsworth. Crc Pr Inc; 1984.
3. Bundesministerium für Gesundheit, Familie und Jugend. Leistungsorientierte Krankenanstaltenfinanzierung. LKF Systembeschreibung, 2009. URL:[http://www.bmg.gv.at/cms/site/attachments/6/4/5/CH0719/CMS1159516854629/systembeschreibung\\_2009.pdf](http://www.bmg.gv.at/cms/site/attachments/6/4/5/CH0719/CMS1159516854629/systembeschreibung_2009.pdf) (accessed June 2, 2009).