Hungary-Serbia
IPA Cross-border Co-operation Programme

Teaching Mathematics and Statistics in Sciences HU-SRB/0901/221/088

Practical application of biostatistical methods in medical and biological research

Novi Sad, 2011.

Krisztina Boda PhD

Department of MedicalPhysics and Informatics, University of Szeged, Hungary

- **Why is a physician held in much higher esteem than a statistician?**
- **A physician makes an analysis of a complex illness whereas a statistician makes you ill with a complex analysis!**
- http://my.ilstu.edu/~gcramsey/StatOtherPro.html

Contents

- Introduction
  - Motivating examples
- Theory
  - Types of studies
  - Comparison of two probabilities
  - Multiplicity problems
  - Linear models
  - Generalized linear models, logistic regression, relative risk regression
- Practical application
  - Introductions
  - First version
  - Multivariate modelling
  - Correction of p-values

Introduction

- Investigation of risk factors of some illness is one of the most frequent problems in medical research.
- Such problems usually need hard statistics, multivariate methods (such as multiple regression, general linear or nonlinear models) .
- Motivating examples: investigation of risk factors of adverse respiratory events
  - use of laryngeal mask airway (LMA) – 60 variables about 831 children
  - respiratory complications in paediatric anaesthesia – 200 variables about 9297 children

Motivating example 1: Incidence of Adverse Respiratory Events in Children with Recent Upper Respiratory Tract Infections (URI)

- The laryngeal mask airway (LMA) is a technique to tracheal intubation for airway management of children with recent upper respiratory tract infections (URIs).
- The occurrence of adverse respiratory events was examined and the associated risk factors were identified to assess the safety of LMA in children.

- von Ungern-Sternberg BS., Boda K., Schwab C., Sims C., Johnson C., Habre W.: Laryngeal mask airway is associated with an increased incidence of adverse respiratory events in children with recent upper respiratory tract infections. Anesthesiology 107(5):714-9, 2007. IF: 4.596

Data about 831 children

- Independent (exploratory) variables (risk factors??)
  - Demography
    - Gender, age, weight, etc.
  - Medical history
    - Asthma, cough, allergy, smoking, etc.
  - Symptoms of URI
    - Fever, moist cough, runny nose, etc.
  - Medication and maintenance of anaesthesia
    - Surgery, airway management, etc.
- Dependent (outcome) variables
  - Respiratory adverse events: laryngospasm, bronchospasm, airway obstruction, cough, oxigen desaturation, overall (any of them)
  - Intraoperative / in the recovery room

Variables in the data file

*lma-gva-cleaned-corrected30.sav [DataSet2] - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

	Name	Type	Width	Decimals	Label	Values	Missing
5	cold	Numeric	8	0		{0, no}...	
6	runnynos	Numeric	8	0	runny nose	{1, last 2 week	
7	clearnos	Numeric	8	0	clear runny nose	{0, no}...	
8	greenose	Numeric	8	0	nasty, green runny nose	{0, no}...	
9	fever	Numeric	8	0	fever	{0, no}...	
10	drycough	Numeric	8	0	dry cough	{0, no}...	
11	moistcof	Numeric	8	0	moist cough	{0, no}...	
12	asthma	Numeric	8	0	history of asthma	{0, no}...	
13	cough	Numeric	8	0	Nocturnal chronic cough	{0, no}...	
14	allergy	Numeric	8	0	history of allergy	{0, no}...	
15	smoking	Numeric	8	0	passive smoking	{0, no}...	
16	procedur	Numeric	8	0	surgical procedure	{1, lower abdo	
17	registra	Numeric	8	0	doctor	{1, registrar}...	
18	inhalat	Numeric	8	0	inhalation induction	{0, no}...	
19	intraven	Numeric	8	0	intravenous induction	{0, no}...	
20	midazola	Numeric	8	0	premedication	{0, no}...	
21	narcotic	Numeric	8	0		{0, no}...	
22	regional	Numeric	8	0		{0, no}...	
23	caudal	Numeric	8	0		{0, no}...	
24	lmasize	Numeric	8	1		None	
25	reinforc	Numeric	8	0		{0, no}...	
26	lignocai	Numeric	8	0	jelly put on the cuff	{0, no}...	
27	attempts	Numeric	8	0	number of attempts to insert	None	
28	laryngos	Numeric	8	0		{0, no}...	
29	bronchos	Numeric	8	0		{0, no}...	
30	obstruc	Numeric	8	0		{0, no}...	
31	desat	Numeric	8	0		{0, no}...	
32	lmacough	Numeric	8	0	respiratory complications	{0, no}...	
33	removal	Numeric	8	0		{0, deep}...	
34	rrlaryng	Numeric	8	0	laryngospasm in recovery	{0, no}...	
35	rrbronch	Numeric	8	0	bronchospasm in recovery	{0, no}...	
36	rrobstr	Numeric	8	0	airway obstruction in recover	{0, no}...	
37	rrdesat	Numeric	8	0	desaturation in recovery	{0, no}...	
38	rrcough	Numeric	8	0	cough in recovery	{0, no}...	
39	oxygen	Numeric	8	0	oxygen in recovery	{0, no}...	
40	timeoxyg	Numeric	8	2	total time oxygen administrati	None	

Data View | Variable View

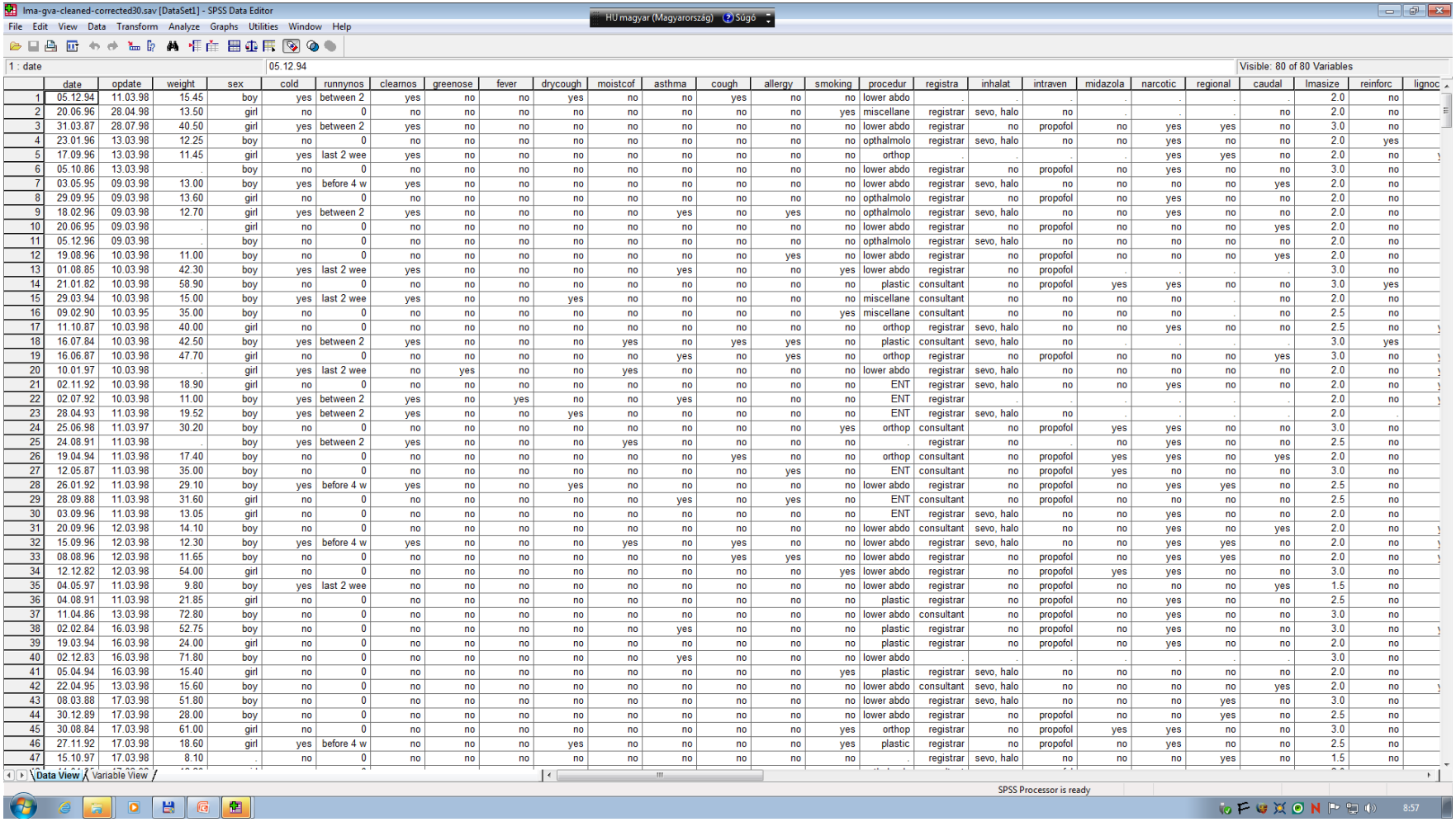
*lma-gva-cleaned-corrected30.sav [DataSet2] - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

	Name	Type	Width	Decimals	Label	Values	Missing
45	pradvers	Numeric	8	2	Adverse event during the proc.	{.00, No}...	None
46	rradvers	Numeric	8	2	Adverse event in r.room	{.00, No}...	None
47	adverse	Numeric	8	2	Adverse event	{.00, No}...	None
48	preins	Numeric	8	2	pre insertion	None	None
49	onins	Numeric	8	2	on insertion	None	None
50	postins	Numeric	8	2	post insertion	None	None
51	onrem	Numeric	8	2	on removal	None	None
52	postrem	Numeric	8	2	post removal	None	None
53	intrven0	Numeric	8	2	intravenous induction recoded	{.00, no}...	None
54	intrvenp	Numeric	8	2	intravenous induction recoded	{.00, propofol}...	None
55	intprno	Numeric	8	2	intravenous induction recoded	{.00, propofol}...	None
56	intprthi	Numeric	8	2	intravenous induction recoded	{.00, propofol}...	None
57	registr1	Numeric	8	2		{.00, 1 'registrar' 1 'consultant'}...	None
58	nmlaryn1	Numeric	8	2		None	None
59	nmlaryn	Numeric	8	2		None	None
60	laryng	Numeric	8	2	Laryngospasm during the proc. or in the r.room	{.00, No}...	None
61	prother	Numeric	8	2	Adverse events except laryng during the proc	{.00, No}...	None
62	bronch	Numeric	8	2	Bronchospasm during the proc. or in the r.room	{.00, No}...	None
63	obstr	Numeric	8	2	Airway obstr during the proc. or in the r.room	{.00, No}...	None
64	des	Numeric	8	2	Desaturation obstr during the proc. or in the r.room	{.00, No}...	None
65	coug	Numeric	8	2	Cough during the proc. or in the r.room	{.00, No}...	None
66	coldsymt	Numeric	8	2		None	None
67	group	Numeric	8	2		{.00, No cold}...	None
68	uri	Numeric	8	2	Children with recent URI	{.00, no}...	None
69	filter_$	Numeric	1	0	uri = 1 (FILTER)	{0, Not Selected}...	None
70	perlaryng	Numeric	8	2	Laryngospasm during the proc.	{.00, No}...	None
71	perbronch	Numeric	8	2	Bronchospasm during the proc.	{.00, No}...	None
72	perobstr	Numeric	8	2	Airway obstr during the proc.	{.00, No}...	None
73	perdesat	Numeric	8	2	Desaturation obstr during the proc.	{.00, No}...	None
74	percough	Numeric	8	2	Cough during the proc.	{.00, No}...	None
75	ENT	Numeric	8	2		{.00, all other surgical proc.}...	None
76	respevent	Numeric	8	2	number of resp. events during the proc. or in the rr	{.00, 0}...	None
77	induction	Numeric	8	2		{.00, sevo/halo}...	None
78	attem1	Numeric	8	2	Number of attempts	{.00, 1}...	None
79	allcomp	Numeric	8	2	All complications during the proc. or in the r.room	{.00, No}...	None
80	PRE_1	Numeric	11	5	Predicted probability	None	None

Data View | Variable View

The data file (part)



	date	opdate	weight	sex	cold	runnynos	clearnos	greenose	fever	drycough	moistcof	asthma	cough	allergy	smoking	procedur	registra	inhalat	intraven	midazola	narcotic	regional	caudal	lmasize	reinforc	lignoc
1	05.12.94	11.03.98	15.45	boy	yes	between 2	yes	no	no	yes	no	no	yes	no	no	lower abdo	.	.	.	.	.	.	.	2.0	no	
2	20.06.96	28.04.98	13.50	girl	no	0	no	no	no	no	no	no	no	no	yes	miscellane	registrar	sevo, halo	no	.	.	.	no	2.0	no	
3	31.03.87	28.07.98	40.50	girl	yes	between 2	yes	no	no	no	no	no	no	no	no	lower abdo	registrar	no	propofol	no	yes	yes	no	3.0	no	
4	23.01.96	13.03.98	12.25	boy	no	0	no	no	no	no	no	no	no	no	no	opthalmolo	registrar	sevo, halo	no	no	yes	no	no	2.0	yes	
5	17.09.96	13.03.98	11.45	girl	yes	last 2 wee	yes	no	no	no	no	no	no	no	no	orthop	.	.	.	.	yes	yes	no	2.0	no	
6	05.10.86	13.03.98	.	boy	no	0	no	no	no	no	no	no	no	no	no	lower abdo	registrar	no	propofol	no	yes	no	no	3.0	no	
7	03.05.95	09.03.98	13.00	boy	yes	before 4 w	yes	no	no	no	no	no	no	no	no	lower abdo	registrar	sevo, halo	no	no	no	no	yes	2.0	no	
8	29.09.95	09.03.98	13.60	girl	no	0	no	no	no	no	no	no	no	no	no	opthalmolo	registrar	no	propofol	no	yes	no	no	2.0	no	
9	18.02.96	09.03.98	12.70	girl	yes	between 2	yes	no	no	no	no	yes	no	yes	no	opthalmolo	registrar	sevo, halo	no	no	yes	no	no	2.0	no	
10	20.06.95	09.03.98	.	girl	no	0	no	no	no	no	no	no	no	no	no	lower abdo	registrar	no	propofol	no	no	no	yes	2.0	no	
11	05.12.96	09.03.98	.	boy	no	0	no	no	no	no	no	no	no	no	no	opthalmolo	registrar	sevo, halo	no	no	no	no	no	2.0	no	
12	19.08.96	10.03.98	11.00	boy	no	0	no	no	no	no	no	no	no	yes	no	lower abdo	registrar	no	propofol	no	no	no	yes	2.0	no	
13	01.08.85	10.03.98	42.30	boy	yes	last 2 wee	yes	no	no	no	no	yes	no	no	yes	lower abdo	registrar	no	propofol	.	.	.	.	3.0	no	
14	21.01.82	10.03.98	58.90	boy	no	0	no	no	no	no	no	no	no	no	no	plastic	consultant	no	propofol	yes	yes	no	no	3.0	yes	
15	29.03.94	10.03.98	15.00	boy	yes	last 2 wee	yes	no	no	yes	no	no	no	no	no	miscellane	consultant	no	no	no	no	.	no	2.0	no	
16	09.02.90	10.03.95	35.00	boy	no	0	no	no	no	no	no	no	no	no	yes	miscellane	consultant	no	no	no	no	.	no	2.5	no	
17	11.10.87	10.03.98	40.00	girl	no	0	no	no	no	no	no	no	no	no	no	orthop	registrar	sevo, halo	no	no	yes	no	no	2.5	no	
18	16.07.84	10.03.98	42.50	boy	yes	between 2	yes	no	no	no	yes	no	yes	yes	no	plastic	consultant	sevo, halo	no	.	.	.	.	3.0	yes	
19	16.06.87	10.03.98	47.70	girl	no	0	no	no	no	no	no	yes	no	yes	no	orthop	registrar	no	propofol	no	no	no	yes	3.0	no	
20	10.01.97	10.03.98	.	girl	yes	last 2 wee	no	yes	no	no	yes	no	no	no	no	lower abdo	registrar	sevo, halo	no	no	no	no	no	2.0	no	
21	02.11.92	10.03.98	18.90	girl	no	0	no	no	no	no	no	no	no	no	no	ENT	registrar	sevo, halo	no	no	yes	no	no	2.0	no	
22	02.07.92	10.03.98	11.00	boy	yes	between 2	yes	no	yes	no	no	yes	no	no	no	ENT	registrar	.	.	.	.	.	.	2.0	no	
23	28.04.93	11.03.98	19.52	boy	yes	between 2	yes	no	no	yes	no	no	no	no	no	ENT	registrar	sevo, halo	no	.	.	.	.	2.0	.	
24	25.06.98	11.03.97	30.20	boy	no	0	no	no	no	no	no	no	no	no	yes	orthop	consultant	no	propofol	yes	yes	no	no	3.0	no	
25	24.08.91	11.03.98	.	boy	yes	between 2	yes	no	no	no	yes	no	no	no	no	.	registrar	no	.	no	yes	no	no	2.5	no	
26	19.04.94	11.03.98	17.40	boy	no	0	no	no	no	no	no	no	yes	no	no	orthop	consultant	no	propofol	yes	yes	no	yes	2.0	no	
27	12.05.87	11.03.98	35.00	boy	no	0	no	no	no	no	no	no	no	yes	no	ENT	consultant	no	propofol	yes	no	no	no	3.0	no	
28	26.01.92	11.03.98	29.10	boy	yes	before 4 w	yes	no	no	yes	no	no	no	no	no	lower abdo	registrar	no	propofol	no	yes	yes	no	2.5	no	
29	28.09.88	11.03.98	31.60	girl	no	0	no	no	no	no	no	yes	no	yes	no	ENT	consultant	no	propofol	no	no	no	no	2.5	no	
30	03.09.96	11.03.98	13.05	girl	no	0	no	no	no	no	no	no	no	no	no	ENT	registrar	sevo, halo	no	no	yes	no	no	2.0	no	
31	20.09.96	12.03.98	14.10	boy	no	0	no	no	no	no	no	no	no	no	no	lower abdo	consultant	sevo, halo	no	no	yes	no	yes	2.0	no	
32	15.09.96	12.03.98	12.30	boy	yes	before 4 w	yes	no	no	no	yes	no	yes	no	no	lower abdo	registrar	sevo, halo	no	no	yes	yes	no	2.0	no	
33	08.08.96	12.03.98	11.65	boy	no	0	no	no	no	no	no	no	yes	yes	no	lower abdo	registrar	no	propofol	no	yes	yes	no	2.0	no	
34	12.12.82	12.03.98	54.00	girl	no	0	no	no	no	no	no	no	no	no	yes	lower abdo	registrar	no	propofol	yes	yes	no	no	3.0	no	
35	04.05.97	11.03.98	9.80	boy	yes	last 2 wee	no	no	no	no	no	no	no	no	no	lower abdo	registrar	no	propofol	no	no	no	yes	1.5	no	
36	04.08.91	11.03.98	21.85	girl	no	0	no	no	no	no	no	no	no	no	no	plastic	registrar	no	propofol	no	yes	no	no	2.5	no	
37	11.04.86	13.03.98	72.80	boy	no	0	no	no	no	no	no	no	no	no	no	lower abdo	consultant	no	propofol	no	yes	no	no	3.0	no	
38	02.02.84	16.03.98	52.75	boy	no	0	no	no	no	no	no	yes	no	no	no	plastic	registrar	no	propofol	no	yes	no	no	3.0	no	
39	19.03.94	16.03.98	24.00	girl	no	0	no	no	no	no	no	no	no	no	no	plastic	registrar	no	propofol	no	yes	no	no	2.0	no	
40	02.12.83	16.03.98	71.80	boy	no	0	no	no	no	no	no	yes	no	no	no	lower abdo	.	.	.	.	.	.	.	3.0	no	
41	05.04.94	16.03.98	15.40	girl	no	0	no	no	no	no	no	no	no	no	yes	plastic	registrar	sevo, halo	no	no	no	no	no	2.0	no	
42	22.04.95	13.03.98	15.60	boy	no	0	no	no	no	no	no	no	no	no	no	lower abdo	consultant	sevo, halo	no	no	no	no	yes	2.0	no	
43	08.03.88	17.03.98	51.80	boy	no	0	no	no	no	no	no	no	no	no	no	lower abdo	registrar	sevo, halo	no	no	no	yes	no	3.0	no	
44	30.12.89	17.03.98	28.00	boy	no	0	no	no	no	no	no	no	no	no	no	lower abdo	registrar	no	propofol	no	no	yes	no	2.5	no	
45	30.08.84	17.03.98	61.00	girl	no	0	no	no	no	no	no	no	no	no	yes	orthop	registrar	no	propofol	yes	yes	no	no	3.0	no	
46	27.11.92	17.03.98	18.60	girl	yes	before 4 w	no	no	no	yes	no	no	no	no	yes	plastic	registrar	no	propofol	no	yes	no	no	2.5	no	
47	15.10.97	17.03.98	8.10	.	no	0	no	no	no	no	no	no	no	no	no	.	registrar	sevo, halo	no	no	no	yes	no	1.5	no	

Some univariate results

Table 1. Demographic and Medical History of Children with and without a Recent Upper Respiratory Tract Infection

Variable	Without URI (n = 608)	URI (n = 223)	*P* Value*
Age, mean (SE), yr	6.9 (0.18)	5.1 (0.27)	<0.0001*
Weight, mean (SE), kg	26.4 (0.71)	21.8 (1.01)	<0.0001*
Male sex, %	63.9	32.3	0.325
History of asthma, %	16.1	22.0	0.052
Nocturnal chronic cough, %	11.2	15.7	0.096
History of allergy, %	17.3	16.6	0.917
Passive smoking, %	17.4	21.5	0.191
Symptoms of URI, %			
Fever	3.9	18.4	<0.0001
Dry cough	9.0	32.3	<0.0001
Moist cough	6.4	29.1	<0.0001

* *P* value from Mann–Whitney U test; other *P* values from Fisher exact test.
URI = upper respiratory tract infection.

Table 2. Incidence of Respiratory Complications in the Two Groups of Children

	No URI, % (n = 608)	URI, % (n = 223)	OR	95% CI	*P* Value
Overall complications in the perioperative period					
Laryngospasm	3.1	7.6	2.558	1.305–5.016	0.007†
Bronchospasm		0.9	—	—	0.072
Airway obstruction	7.1	6.3	0.880	0.472–1.642	0.759
Oxygen desaturation	11.4	19.3	1.863	1.228–2.825	0.004†
Cough	7.5	17.9	2.730	1.728–4.313	<0.0001*
Overall‡	19.1	31.8	1.981	1.401–2.803	<0.0001*

Question

- Which are the real risk factors of the respiratory adverse events?

Motivating example 2: Investigation of risk factors of respiratory complications in paediatric anaesthesia

- Perioperative respiratory adverse events in children are one of the major causes of morbidity and mortality during paediatric anaesthesia. We aimed to identify associations between family history, anaesthesia management, and occurrence of perioperative respiratory adverse events.

- von Ungern-Sternberg BS., Boda K., Chambers NA., Rebmann C ., Johnson C., Sly PD, Habre W.:: Risk assessment for respiratory complications in paediatric anaesthesia: a prospective cohort study, The Lancet, 376 (9743): 773-783, 2010.

Data

- We prospectively included all children who had general anaesthesia for surgical or medical interventions, elective or urgent procedures at Princess Margaret Hospital for Children, Perth, Australia, from Feb 1, 2007, to Jan 31,2008.
- On the day of surgery, anaesthetists in charge of paediatric patients completed an adapted version of the International Study Group for Asthma and Allergies in Childhood questionnaire. RESPIRATORY COMPLICATIONS without boxes.doc
- We collected data on family medical history of asthma, atopy, allergy, upper respiratory tract infection, and passive smoking.
- Anaesthesia management and all perioperative respiratory adverse events were recorded.
- **9297** questionnaires were available for analysis.
- Number of variables: more than **300**.

Statistical methods and problems

- Check the data base – are data consequently coded, etc.
- Univariate methods
- Correction of univariate p-values to avoid the inflation of the Type I error
- Examining relationship (correlation) between variables
- Multiple regression modelling
  - Possible problems to find a reasonable model:
    - Number of independent variables – not too much, not too small
    - Avoid multicollinearity
    - Good fit
    - Checking interactions
    - Comparison of models
- …

Univariate methods

Description of contingency tables (Agresti)

- Notation
  - *X* categorical variable with *I* categories
  - *Y* categorical variable with *J* categories
- Variables can be cross tabulated. The table of frequencies is called contingency table or cross-classification table with *I* rows and *J* columns, *I*x*J* table.
- Generally, *X* is considered to be independent variable and *Y* is a dependent variable(outcome)

TABLE 2.1 Cross-Classification of Aspirin Use and Myocardial Infarction

	Myocardial Infarction		
	Fatal Attack	Nonfatal Attack	No Attack
Placebo	18	171	10,845
Aspirin	5	99	10,933

Source: Preliminary report: Findings from the aspirin component of the ongoing Physicians' Health Study. *New Engl. J. Med.* **318**: 262–264 (1988).

Probability distributions

- π_{ij}: the probability that (X, Y) occurs in the cell in row i and column j. The probability distribution $\{\pi_{ij}\}$ ***is the joint distribution of X and Y***
- The ***marginal distributions*** are the row and column totals that result from summing the joint probabilities.
- $\pi_{j|i}$: Given that a subject is classified in row i *of X,* $\pi_{j|i}$ is the probability of classification in column *j of Y, j=1, . . . , J.*
- The probabilities $\{\pi_{1|i}, \pi_{2|i}, \ldots, \pi_{J|i}\}$ form the *conditional distribution* of *Y at category i of X.*
- A principal aim of many studies is to compare conditional distributions of *Y at various levels of explanatory variables.*

TABLE 2.3 Notation for Joint, Conditional, and Marginal Probabilities

	Column		
Row	1	2	Total
1	π_{11} $(\pi_{1\|1})$	π_{12} $(\pi_{2\|1})$	π_{1+} (1.0)
2	π_{21} $(\pi_{1\|2})$	π_{22} $(\pi_{2\|2})$	π_{2+} (1.0)
Total	π_{+1}	π_{+2}	1.0

Types of studies

- **Case-controll (retrospective**). The smoking behaviour of 709 patients with lung cancer was examined For each of the 709 patients admitted, researchers studied the smoking behaviour of a noncancer patient at the same hospital of the same gender and within the same 5-year grouping on age .
- **Prospective.** Groups of smokers and non-smokers are observed during years (30 years) and the outcome (cancer) is observed at the end of the study.
- **Clinical trials**– randomisation of the patients
- **Cohort studies** – subjects make their own choice about whether to smoke, and the study observes in future time who develops lung cancer.
- **Cross-sectional studies** – samples subjects and classifies them simultaneously on both variables.

TABLE 2.5 Cross-Classification of Smoking by Lung Cancer

	Lung Cancer	
Smoker	Cases	Controls
Yes	688	650
No	21	59
Total	709	709

Source: Based on data reported in Table IV, R. Doll and A. B. Hill, *British Med. J.*, Sept. 30, 1950, pp. 739–748.

- *Prospective studies* usually condition on the totals for categories of *X* and regard each row of *J* counts as an independent multinomial sample on *Y*.
- *Retrospective studies* usually treat the totals for *Y* as fixed and regard each column of *I* counts as a multinomial sample on *X*.
- In *cross-sectional studies,* the total sample size is fixed but not the row or column totals, and the *IJ* cell counts are a multinomial sample.

Comparison of two proportions

- Notation in case 2x2-es: instead of $\pi_{2|i}$ =1- $\pi_{1|i}$, simply π_1-π_2
- Difference (absolute risk difference) π_1-π_2
  - It falls between -1 and 1
  - The response Y is statistically independent of the row classification when the difference is 0
- Ratio (relative risk, risk ratio, RR) π_1/π_2
  - It can be any nonnegative number
  - A relative risk of 1.0 corresponds to independence
  - Comparing probabilities close to 0 or 1, the differences might be negligible while their ratio is more informative
- Odss ratio, OR, here Ω
  - For a probability of π success, the odds are defined to be Ω= π/(1- π)
  - Odds are nonnegative. Ω>1, when a success is more likely than a failure.
  - Getting probability from the odds: π= Ω/(Ω+1)
  - Odds ratio

$$\theta = \frac{\Omega_1}{\Omega_2} = \frac{\pi_1/(1-\pi_1)}{\pi_2/(1-\pi_2)}$$

  - Odds ratio when the cell probabilities π_{ij} are given $\Omega_i = \pi_{i1}/\pi_{i2}, i=1,2$

$$\theta = \frac{\pi_{11}/\pi_{12}}{\pi_{21}/\pi_{22}} = \frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}}.$$

Odds ratio (OR) and relative risk (RR)

- when each probability is small, the odds ratio provides a rough indication of the relative risk when it is not directly estimable

$$\text{odds ratio} = \text{relative risk}\left(\frac{1-\pi_2}{1-\pi_1}\right).$$

Odds ratio and logistic regression

- Logistic regression models give the estimation of odds ratio (adjusted or unadjusted).
- It has no distributional assumption, the algorithm is generally convergent.
- The use of logistic regression is popular in medical literature.

Case-control studies and OR

Illness / *Risk factorr*	*groups* Case	Control	Σ
present	*a:* 40	*b:* 20	60
absent	*c:* 60	*d:* 80	140
Σ	100	100	200

- Odds Ratio: $$OR = \frac{\text{odds of illness for smokers}}{\text{odds of illness for nonsmokers}} = \frac{a/b}{c/d} = \frac{a/c}{b/d} = \frac{ad}{bc} = \frac{40 \cdot 80}{20 \cdot 60} = 2{,}67$$
- OR=1: independency, OR<<1: strong negative association, OR>>1: strong positive association.
- Interpretation. The illness is 2,67-times more likely to occur among smokers than among non-smokers.
  Comment. Although the retrospective sample is not representative for the ill/healthy ratio (a/b and c/d), we get correct estimation, because the ratios a/c and b/d are correct.
- In case of several risk factors, the common effect of several risk factors can be analysed using logistic regression, and adjusted odds ratios can be calculated.

Prospective study and RR

Illness / *Risk factorr*	*groups* Case	Control	Σ
present	*a:* 40	*b:* 20	60
absent	*c:* 60	*d:* 80	140
Σ	100	100	200

Relative Risk (*RR*):

$$RR = \frac{\text{risk of illness for smokers}}{\text{risk of illness for smokers}} = \frac{a/(a+b)}{c/(c+d)} = \frac{80/1000}{40/1000} = \frac{0{,}08}{0{,}04} = 2$$

Interpretation. The probability (risk) of illness is twice illness among smokers than among non-smokers.

When the incidence of illness is small in both groups ($a<<b$, $c<<d$), then $RR \approx OR$, i.e., the relative risk can be well approximated by the odds ratio

$$RR = \frac{a/(a+b)}{c/(c+d)} = \frac{80/1000}{40/1000} = 2 \approx OR = \frac{a/b}{c/d} = \frac{80/920}{40/960} = 2{,}087$$

Case-control studies and the odds ratio

- In case-control studies we cannot estimate some conditional probabilities
- Here, the marginal distribution of lung cancer is fixed by the sampling design (i.e. 709 cases and 709 controls), and the outcome measured is whether the subject ever was a smoker.
- We can calculate the conditional distribution of smoking behaviour, given lung cancer status: for cases with lung cancer, this is 688/709, and for controls it is 650/709.
- In the reverse direction (which would be more relevant) we cannot estimate the probability of disease, given smoking behaviour.
- When we know the proportion of the population having lung cancer, we can use Bayes' theorem to compute sample conditional distributions in the direction of main interest

TABLE 2.5 Cross-Classification of Smoking by Lung Cancer

	Lung Cancer	
Smoker	Cases	Controls
Yes	688	650
No	21	59
Total	709	709

Source: Based on data reported in Table IV, R. Doll and A. B. Hill, *British Med. J.*, Sept. 30, 1950, pp. 739–748.

2.21 For a diagnostic test of a certain disease, π_1 denotes the probability that the diagnosis is positive given that a subject has the disease, and π_2 denotes the probability that the diagnosis is positive given that a subject does not have it. Let ρ denote the probability that a subject does have the disease.

a. Given that the diagnosis is positive, show that the probability that a subject does have the disease is

$$\pi_1 \rho / [\pi_1 \rho + \pi_2 (1 - \rho)].$$

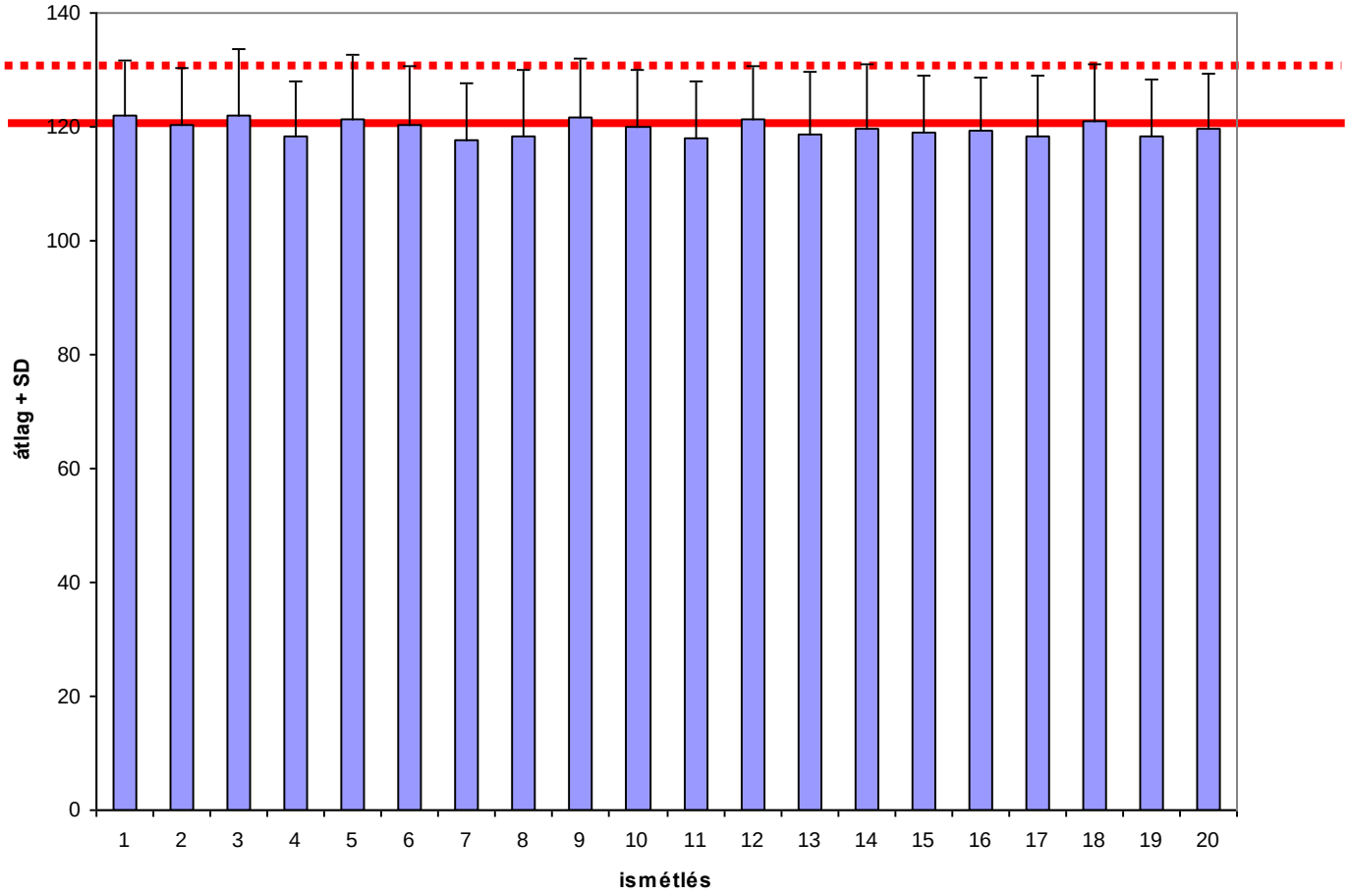
Let "pos" denote positive diagnosis, "dis" denote subject has disease.

$$P(dis|pos) = \frac{P(pos|dis)P(dis)}{P(pos|dis)P(dis) + P(pos|no\ dis)P(no\ dis)}$$

Comparison of several samples using univariate methods

The repeated use of t-tests is not appropriate

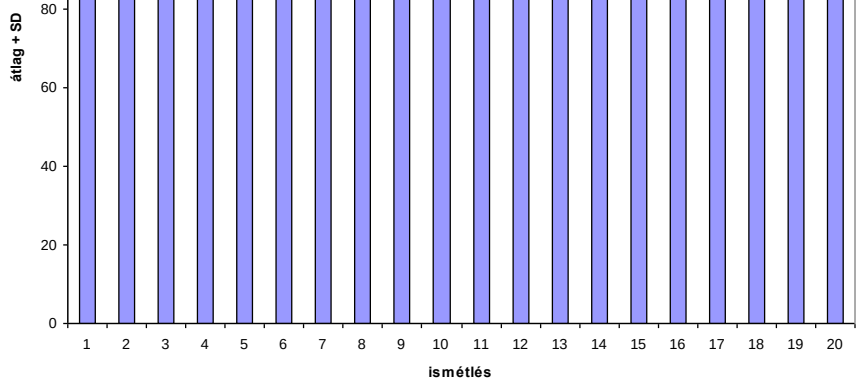
Mean and SD of samples drawn from a normal population N(120, 10^2), (i.e. μ=120 and σ=10)



Pair-wise comparison of samples drawn from the same distribution using *t*-tests

Variable	T-test for Dependent Samples: p-levels (veletlen) Marked differences are significant at p < .05000										
	s10	s11	s12	s13	s14	s15	s16	s17	s18	s19	s20
s1	0.30407	0.07484	0.78173	0.15872	0.22271	0.15123	0.21106	0.02826	0.65675	0.04878	0.22301
s2	0.94385	0.32693	0.44510	0.45003	0.79924	0.46849	0.73289	0.35108	0.58983	0.31241	0.84292
s3	0.36469	0.10013	0.83458	0.15161	0.30077	0.15297	0.20104	0.13663	0.71210	0.09278	0.34899
s4	0.33509	0.91259	0.06954	0.81184	0.49090	0.64673	0.52137	0.99453	0.17286	0.97725	0.33843
s5	0.49261	0.13965	0.99830	0.23623	0.42063	0.18648	0.36294	0.14388	0.86579	0.14724	0.39985
s6	0.90480	0.28520	0.59216	0.42988	0.77452	0.49416	0.67473	0.39279	0.70786	0.33013	0.79602
s7	0.15756	0.87779	0.05375	0.63178	0.36101	0.52659	0.53	0.90	0.09261	0.81870	0.26351
s8	0.46222	0.85891	0.15671	0.87889	0.62412	0.78943	0.6	0.32	0.13600	0.92358	0.56453
s9	0.41991	0.04018	0.87536	0.16744	0.35766	0.73997	0.5	0.09	0.75776	0.06879	0.37176

***p*-values (detail)**



The increase of type I error

- It can be shown that when t tests are used to test for differences between multiple groups, the chance of mistakenly declaring significance (Type I Error) is increasing. For example, in the case of 5 groups, if no overall differences exist between any of the groups, using two-sample t tests pair wise, we would have about 30% chance of declaring at least one difference significant, instead of 5% chance.
- *In general, the t test can be used to test the hypothesis that **two** group means are not different. To test the hypothesis that **three ore more** group means are not different, analysis of variance should be used.*

- Each statistical test produces a ‘p’ value
- If the significance level is set at 0.05 (false positive rate) and we do multiple significance testing on the data from a single clinical trial,
- then the overall false positive rate for the trial will increase with each significance test.

Multiple hypotheses

- (H_{01} and H_{02} and... H_{0n}) null hypotheses, the appropriate significance levels $\alpha_1, \alpha_2, \ldots, \alpha_n$
- How to choose α_i-s that the level of hypothesis (H_{01} and H_{02} and... H_{0n}) be greater than a given α ? $\alpha \in (0,1)$

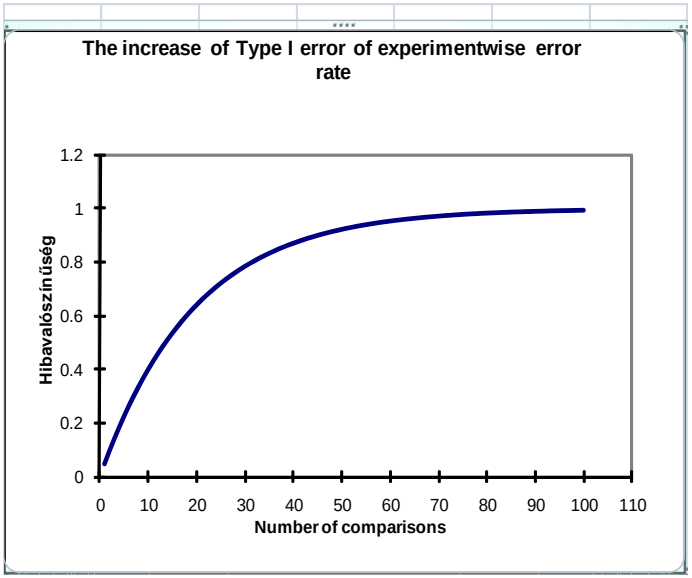
Increase of type I error

Gigen n null hypotheses, H_{oi}, i=1,2,...,n with significance level α

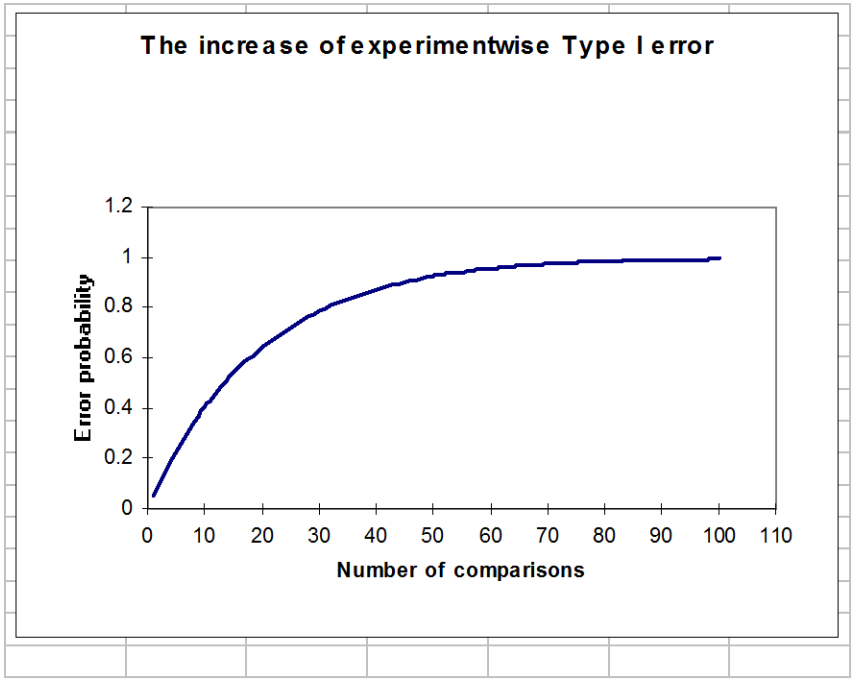
When the hypotheses are independent, the probability that at least one null hypothesis is falsely rejected, is: $1-(1-\alpha)^n$

When the hypotheses are not independent, the probability that at least one null hypothesis is falsely rejected $\leq n\alpha$.

$$1-\left(1-\alpha\right)^n \leq 1-\left(1-n\alpha\right)=n\alpha$$



- False positive rate for each test = 0.05
- Probability of incorrectly rejecting ≥ 1 hypothesis out of *N* testing
- $= 1 - (1\text{-}0.05)^{N}$



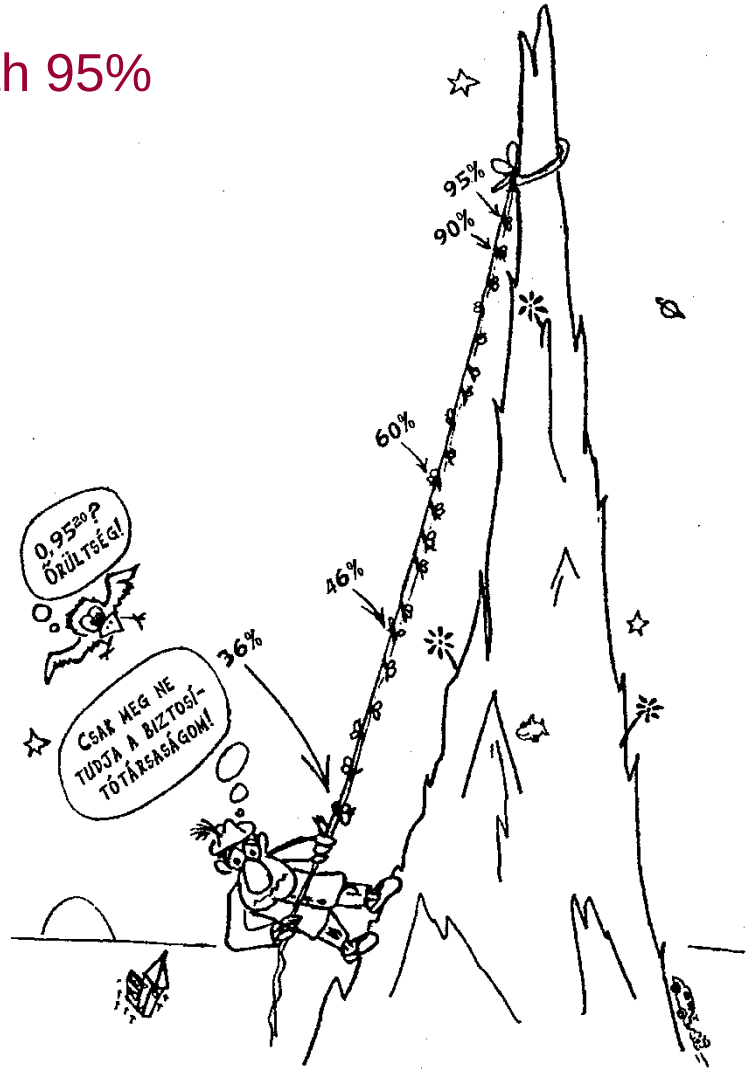
Correction of the unique p-values by the method of Bonferroni-Holm (step-down Bonferroni)

- Calculate the p-values and arrange them in increasing order $p_1 \leq p_2 \leq ... \leq p_n$
- H_{0i} is tested at level. $\frac{\alpha}{n+1-i}$
- If any of them is significant, then we reject the hypothesis (H_{01} and H_{02} and... H_{0n}) .
- Example. n=5

p_1	$\alpha/5=0.01$	if $p_1 \geq 0.01$, stop (there is no significant difference)
p_2	$\alpha/4=0.0125$	if $p_2 \geq 0.0125$, stop
p_3	$\alpha/3=0.0166$	…
p_4	$\alpha/2=0.025$	….
p_5	$\alpha/1=0.05$	

Knotted ropes: each knot is safe with 95% probability

- The probability that two knots are „safe” =0.95*0.95 =0.9025~90%
- The probability that 20 knots are „safe” $=0.95^{20}=0.358$~36%
- The probability of a crash in case of 20 knots is ~64%



10. ábra. Nemtörődöm doktor, amint a nemzetközi szakirodalom által javasolt számos, egyenként meglehetősen biztonságos csomóval összekötözött mászókötélen függ. Ez az utolsó felvétel Nemtörődöm doktorról. Egy naiv elképzelésnek esett áldozatul, azt hitte, hogy a tudomány megbízhatósági kritériumait a hegymászásra is alkalmazni lehet

Correction of p-values using PROC MULTTEST is SAS software

The SAS System

The Multtest Procedure

p-Values

Test	Raw	Stepdown Bonferroni	Hochberg	False Discovery Rate
1	0.9999	1.0000	0.9999	0.9999
2	0.2318	0.9272	0.9272	0.5795
3	0.3771	1.0000	0.9999	0.6285
4	0.8231	1.0000	0.9999	0.9999
5	0.0141	0.0705	0.0705	0.0705

Linear models

The General Linear Model(GLM)

The general form of the linear model is

$$\boldsymbol{y} = \boldsymbol{X}\beta + \varepsilon,$$

where

$\boldsymbol{y}$ is an n x1 response vector,

$\boldsymbol{X}$ is an n x p matrix of constants (“design” matrix), columns are mainly values of 0 or 1 and values of independent variables,

β is a p x 1 vector of parameters, and

ε is an n x 1 random vector whose elements are independent and all have normal distribution N(0, σ^2).

For example, a linear regression equation containing three independent variables can be written as $Y = \beta_0 + \beta_1 X_1 + \beta_2 X2 + \beta_3 X_3, + \boldsymbol{\varepsilon}$, or

$$\boldsymbol{y}=\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \boldsymbol{X}=\begin{bmatrix} 1 & x_{11} & x_{12} & x_{13} \\ 1 & x_{21} & x_{22} & x_{23} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & x_{n3} \end{bmatrix}, \boldsymbol{\beta}=\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} \boldsymbol{\varepsilon}=\begin{bmatrix} \varepsilon_0 \\ \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Limitations

Normal distribution – what happens when normality does not hold?

Constant variance – What happens when variance is not constant?

Dependent variable – what happens when dependent variable is categorical or binary?

The generalized linear model

A generalized linear model has three components:

1. **Random component**. Response variables $Y_1, \ldots, Y_N$ which are assumed to share the same distribution from the exponential family;
2. A set of parameters $\boldsymbol{\beta}$ and explanatory variables

$$\boldsymbol{X} = \begin{bmatrix} \mathbf{x}_1^T \\ \vdots \\ \mathbf{x}_n^T \end{bmatrix} = \begin{bmatrix} x_{11} \cdots x_{1p} \\ \vdots \quad \vdots \\ x_{n1} \cdots x_{np} \end{bmatrix}$$

3. A monotone, differentiable function $\boldsymbol{g}$ – called **link function** such that

$$g(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta}$$

where $\mu_i = E(Y_i)$.

The exponential family of distributions

- The density function :

$$f(y) = \exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right\}$$

- Θ: canonical parameter
- Φ: dispersion (or scale) parameter

Generalized linear models

Random component	Link	Linear component	Model
Normal	**Identity**	**Continuous**	**Regression**
Normal	Identity	Categorical	Analysis of variance
Normal	Identity	Mixed	Analysis of covariance
Binomial	**Logit**	**Mixed**	**Logistic regression**
Poisson	Log	Mixed	Loglinear analysis
Polinomial	Gen.logit	Mixed	Polin.regr.
Binary	**Log**	**Mixed**	**Rel.risk.regr.**

The model of binary logistic regression

Given *p* independent variables: $\boldsymbol{x}'=(x_1, x_2, \ldots, x_p)$ and a dependent variable *Y* with values 0 and 1. Let's denote $P(Y=1|\boldsymbol{x})=\pi(\boldsymbol{x})$: the probability of success given $\boldsymbol{x}$.

The model is

$$g(\boldsymbol{x}) = \ln\left[\frac{\pi(\boldsymbol{x})}{1-\pi(\boldsymbol{x})}\right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p$$

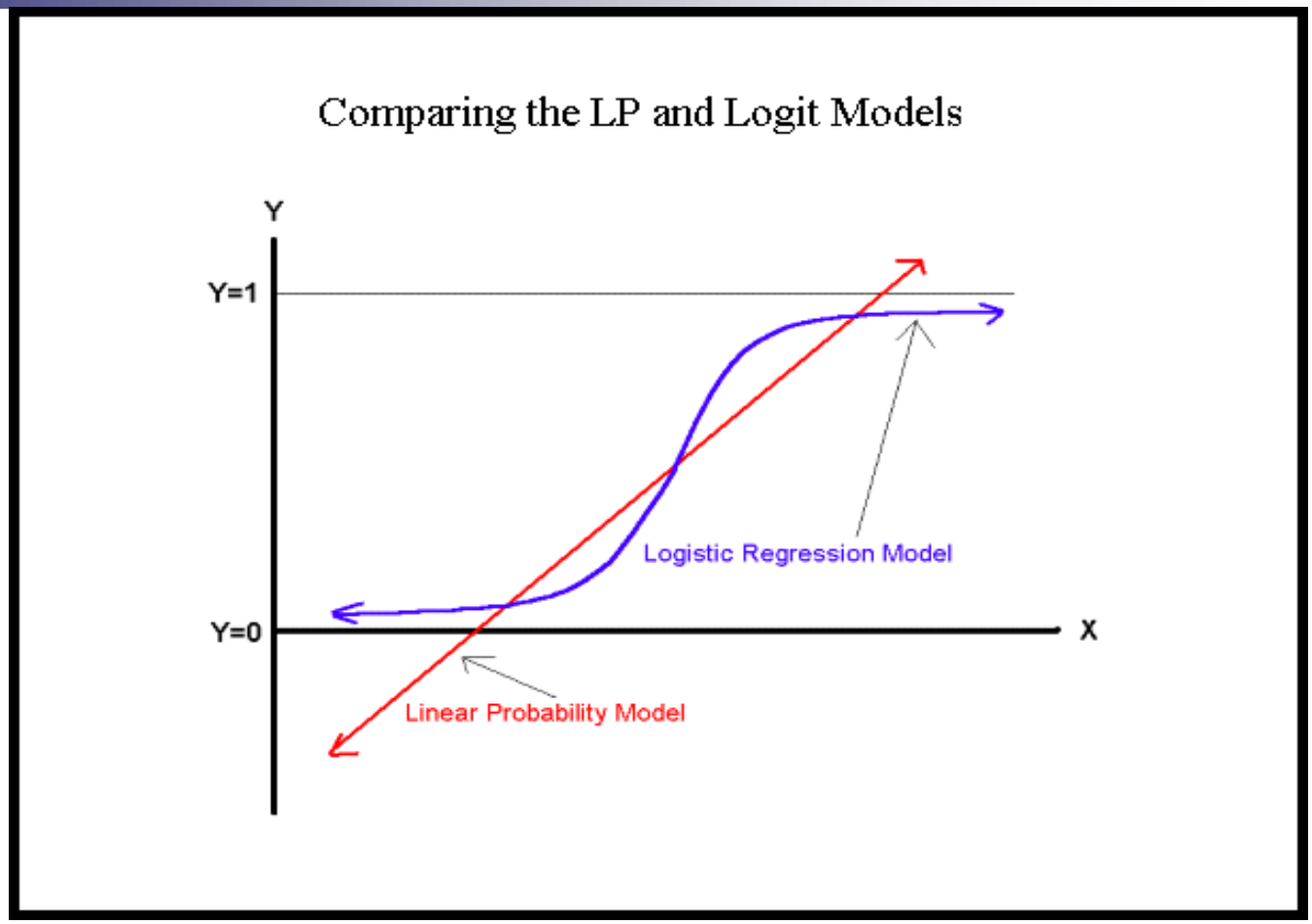
or

$$\pi(x) = \frac{e^{g(\boldsymbol{x})}}{1+e^{g(\boldsymbol{x})}} = \frac{1}{1+e^{-g(\boldsymbol{x})}}$$

g(x): logit transformation. G(x)=ln(OR). Properties:

- It is a linear function of the parameters
- $-\infty < g(x) < +\infty$

- if $\beta_0+\beta_1 X = 0$, then $\pi(x) = .50$
- if $\beta_0+\beta_1 X$ is big, then $\pi(x)$ is close to 1
- if $\beta_0+\beta_1 X$ is small, then $\pi(x)$ is close to 0



An Introduction to Logistic Regression

John Whitehead
Department of Economics East Carolina University
http://personal.ecu.edu/whiteheadj/data/logit/

Multiple logistic regression

- The independent variables can be categorical or continuous variables

$$g(\boldsymbol{x}) = \ln\left[\frac{\pi(x)}{1-\pi(x)}\right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p$$

- Categorical variable encoding:
  - binary: 0-1
  - In case of *k* possible values, we form *k*-1 „dummy" variables.
    - Reference category encoding:
    - The variable has 3 possible values: white, black, other. The dummy variables are:

	D1	D2
White	0	0
Black	1	0
Other	0	1

Interpretation of β_1 in case of dichotomous independent variable

While x changes from 0 to 1, the change in logit is β_1

The estimate of OR is exp(β_1),

$$g(x) = \ln\left[\frac{\pi(x)}{1-\pi(x)}\right] = \beta_0 + \beta_1 x$$

$$g(1) - g(0) = (\beta_0 + \beta_1 \cdot 1) - (\beta_0 + \beta_1 \cdot 0) = \beta_1$$

$$g(1) - g(0) = \ln\frac{\pi(1)}{1-\pi(1)} - \ln\frac{\pi(0)}{1-\pi(0)} = \ln\frac{\frac{\pi(1)}{1-\pi(1)}}{\frac{\pi(0)}{1-\pi(0)}} = \ln(OR)$$

$$e^{\beta_1} = OR$$

In case of several independent variables, exp(β_i)-s are „adjusted” ORs

Fitting logistic regression models

- maximum likelihood method: maximum of the log likelihood -> solution of the likelihood equations by iterations.
- Testing for the significance of the coefficients
  - Wald test
  - Likelihood ratio test
  - Score test

Testing for significance of the coefficients I.

Wald test in case of one independent variable

H0: $ß_1=0$.

Test statistic: compare the maximum likelihood estimate of the slope parameter, $\hat{\beta}_1$, to an estimate of its standard error. The resulting ratio under the null hypothesis will follow a standard normal distribution.

$$W = \frac{\hat{\beta}_1}{\hat{S}_E(\hat{\beta}_1)}$$

Problem: the Wald test behaves in an aberrant manner, often failing to reject the null hypothesis when the coefficient was significant. (Hauck and Donner (1977, J. Am.Stat) – they recommended that likelihood ratio test be used).

Example

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	age	-.063	.020	10.246	1	.001	.939	.903	.976
	Constant	-.853	.141	36.709	1	.000	.426		

a. Variable(s) entered on step 1: age.

$$W = \frac{-0.06324}{0.019756} = 3.201 \quad W^2 = 10.24 \sim \chi^2$$ distribution with 1 degrees of freedom

Interpretation of $ß_1$: it is an estimated log odds ratio. While x changes from 0 to 1, the change in logit is β_1. But the meaningful change must be defined for a continuous variable.

Testing for significance of the coefficients II.
Likelihood ratio test in case of one independent variable

- Does the model that includes the variable in question tell us more about the outcome variable than the model that does not include that variable?
- In linear regression we use an ANOVA table, where we partition the total sum of squares into SS due to regression and residual SS.
- Here we use D=Deviance -2lnL:

- Good fit: likelihood =1 $\rightarrow$ -2lnL=0
- Bad fit: likelihood =0 $\rightarrow$ -2lnL$\approx\infty$.

The better the fit, the smallest is -2lnL.

Comparison of the change of D:
D(with the variable) -D(without the variable) is distributed by χ^2 with 1 degrees of freedom

Example.
Without the variable age: -2lnL= 871.675
With the variable age: -2lnL= 864.706
Difference: 6.969 $\chi^2_{0.05,1}$ =3.841, p < 0.05

We need the variable „age”

Testing possible interactions using likelihood ratio test

Example.

With variables sex and age: -2lnL= 864.706

With sex, age and sex*age: -2lnL= 864.608

Difference: 0.098 $p > 0.05$

The model without interaction is as good as the model with the interaction -> we keep the simpler model

Testing goodness of fit

- Pearson chi-square (Model-chi-square, deviance-D):
  This statistic tests the overall significance of the model. It is distributed as $\chi 2$, the degrees of freedom is the number of independent variables
- Pseudo R^2:
  It is similar to the R^2 in the linear regression. It lies between 0 and 1.
- Hosmer-Lemeshow test
  If the result is not significant, the fit is good (???)
- Classification tables. Based on the predicted probabilities, classification of cases is possible. The „cut" point is generally 0.5.

Classification Table[a]

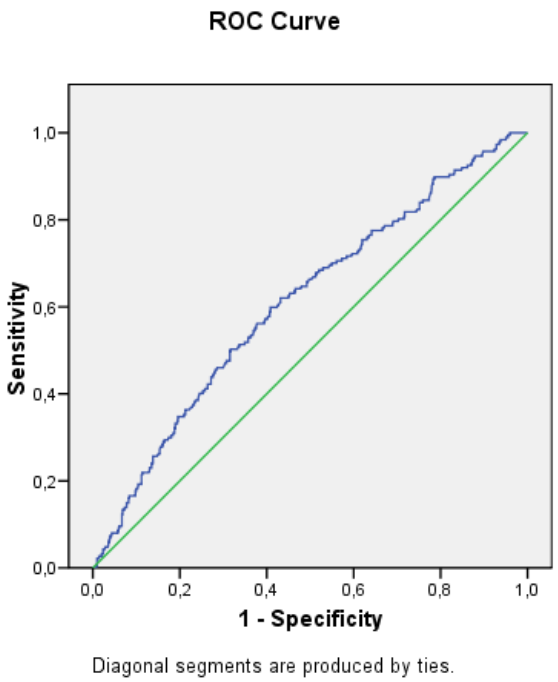
Observed			Predicted		
			All complications during the proc. or in the r. room		Percentage Correct
			No	Yes	
Step 1	All complications during the proc. or in the r.room	No	509	135	79.0
		Yes	122	65	34.8
	Overall Percentage				69.1

a. The cut value is .250

specificity

sensitivity

ROC curves



Diagonal segments are produced by ties.

A plot of Sensitivity vs. 1−Specificity.
In case of complete separation, the curve becomes an upper triangle.
In case of complete equality, the cure becomes a line (green).
Area under the curve can be calculated. The difference from 0.5 can be tested

Area Under the Curve

Test Result Variable(s): Predicted probability

Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.610	.023	.000	.564	.656

The test result variable(s): Predicted probability has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

Steps of model-building

- Choosing candidate variables
  - Univariate statistics (t-test, $\chi 2$ test)
  - „candidate” variables: test result is $p<0.25$
  - Based on medical findings, some nonsignificant variables can be involved
- Testing the „importance” of variables
  - Wald test
  - likelihood ratio
  - stepwise regression
  - best subset
- Check the assumption of linearity in the logit
- Testing interactions
- Goodness of fit
- interpretation

Possible problems

- Irrelevant variables in the model might cause poor model-fit
- Omitting important variables might cause bias in the estimation of coefficients
- Multicollinearity:

- When the independent variables are correlated, there are problems in estimating regression coefficients.
- The greater the multicollinearity, the greater the standard errors. Slight changes in model structure result in considerable changes in the magnitude or sign of parameter estimates.

Relative risk regression (log binomial regression)

$$g(x) = \ln[\pi(x)] = \beta_0 + \beta_1 x$$

$$g(1) - g(0) = (\beta_0 + \beta_1 \cdot 1) - (\beta_0 + \beta_1 \cdot 0) = \beta_1$$

$$g(1) - g(0) = \ln \pi(1) - \ln \pi(0) = \ln \frac{\pi(1)}{\pi(0)} = \ln(RR)$$

$$e^{\beta_1} = RR$$

Problem:
The estimated probability must be between 0 and 1, i.e., $\beta_0 + \beta_1 x \leq 0$. When the method does not converge, then we get a wrong estimation of the RR-s. In case of logistic regression there is no such problem

Overdispersion

- In practice, count observations often exhibit variability exceeding that predicted by the binomial or Poisson. This phenomenon is called *overdispersion*. For example, the sample variance is greater then the sample mean. The reason of this phenomenon is generally the heterogeneity of data.
- Overdispersion does not occur in normal regression models (the mean and the variance are independent parameters), but in case of Poisson and binomial distribution the variance and the mean are not independent.

Evaluation of logistic regression model for data of Example 1.

Univariate analysis: χ^2 test or Mann-Whitney U-test.

Children with recent URI * All complications during the proc. or in the r.room Crosstabulation

			All complications during the proc. or in the r. room		
			No	Yes	Total
Children with recent URI	no	Count	492	116	608
		% within Children with recent URI	80.9%	19.1%	100.0%
	URI	Count	152	71	223
		% within Children with recent URI	68.2%	31.8%	100.0%
Total		Count	644	187	831
		% within Children with recent URI	77.5%	22.5%	100.0%

Risk Estimate

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for Children with recent URI (no / URI)	1.981	1.401	2.803
For cohort All			

Table 2. Incidence of Respiratory Complications in the Two Groups of Children

	No URI, % (n = 608)	URI, % (n = 223)	OR	95% CI	*P* Value
Overall complications in the perioperative period					
Laryngospasm	3.1	7.6	2.558	1.305–5.016	0.007†
Bronchospasm		0.9	—	—	0.072
Airway obstruction	7.1	6.3	0.880	0.472–1.642	0.759
Oxygen desaturation	11.4	19.3	1.863	1.228–2.825	0.004†
Cough	7.5	17.9	2.730	1.728–4.313	<0.0001*
Overall‡	19.1	31.8	1.981	1.401–2.803	<0.0001*

Logistic regression with one independent variable (URI)

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	871.675[a]	.017	.026

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B) Lower	Upper
Step 1[a]	uri	.684	.177	14.926	1	.000	1.981	1.401	2.803
	Constant	-1.445	.103	195.969	1	.000	.236		

a. Variable(s) entered on step 1: uri.

Table 2. Incidence of Respiratory Complications in the Two Groups of Children

	No URI, % (n = 608)	URI, % (n = 223)	OR	95% CI	*P* Value
Overall complications in the perioperative period					
Laryngospasm	3.1	7.6	2.558	1.305–5.016	0.007†
Bronchospasm		0.9	—	—	0.072
Airway obstruction	7.1	6.3	0.880	0.472–1.642	0.759
Oxygen desaturation	11.4	19.3	1.863	1.228–2.825	0.004†
Cough	7.5	17.9	2.730	1.728–4.313	<0.0001*
Overall‡	19.1	31.8	1.981	1.401–2.803	<0.0001*

Logistic regression with two independent variables (URI and age)

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	864.706[a]	.026	.039

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I. for EXP(B) Lower	Upper
Step 1[a]	uri	.598	.180	10.996	1	.001	1.818	1.277	2.588
	age	-.052	.020	6.735	1	.009	.949	.912	.987
	Constant	-1.102	.163	45.694	1	.000	.332		

a. Variable(s) entered on step 1: uri, age.

Adjusted OR

Without the variable age: -2lnL= 871.675
With the variable age: -2lnL= 864.706
Difference: 6.969 $\chi^2_{0.05,1}$ =3.841, $p < 0.05$

We need the variable „age”

Logistic regression with interaction

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	864.608[a]	.026	.039

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	uri	.525	.294	3.195	1	.074	1.690	.951	3.006
	age	-.056	.024	5.568	1	.018	.945	.902	.991
	age by uri	.014	.044	.099	1	.754	1.014	.929	1.106
	Constant	-1.077	.180	35.634	1	.000	.341		

a. Variable(s) entered on step 1: uri, age, age * uri .

With variables sex and age: -2lnL= 864.706
With sex, age and sex*age: -2lnL= 864.608
Difference: 0.098 p > 0.05

The model without interaction is as good as the model with the interaction -> we keep the simpler model

Logistic regression with several independent variables

Table 3. Odds Ratios and 95% Confidence Intervals for the Risk Factors Associated with the Occurrence of Perioperative Respiratory Adverse Events

	Univariate		Multivariate First Model		Multivariate Final Model	
	OR	CI	OR	CI	OR	CI
URI	**2.0***	**1.4–2.8**	**1.777†**	**1.107–2.854**	**1.828†**	**1.3–2.6**
Age	**0.9***	**0.8–0.9**	**0.953†**	**0.915–0.992**	**0.95†**	**0.91–0.98**
ENT surgery	1.3	0.8–2.2				
Asthma	1.19	0.8–1.8				
Nocturnal cough	1.20	0.7–1.9				
Allergy	0.95	0.6–1.5				
Passive smoking	0.89	0.6–1.4				
Clear runny secretions	**1.53**	**1.1–2.1**	1.052	0.657–1.682		
Green runny secretions	**1.79**	**1.0–3.2**	1.465	0.748–2.869		
Fever	1.59	0.9–2.8				
Dry cough	1.26	0.8–2.0				
Moist cough	1.46	0.9–2.3	1.052	0.626–1.768		
LMA size	**0.62**	**0.5–0.9**				
Reinforced LMA	1.18	0.7–2.0				
Lignocaine	0.69	0.4–1.1	0.702	0.443–1.113		
Number of attempts	**3.06***	**1.9–5.0**				
Consultant	0.95	0.7–1.4				
Inhalational induction	1.21	0.8–1.6				
Propofol induction	0.89	0.6–1.3				
Thiopentone induction	1.02	0.5–1.9				
Midazolam	1.35	0.8–2.2				
Opioid	**1.47***	**1.0–2.2***				
Removal of LMA						
Deep *vs.* awake	0.70	0.3–1.5				

* $P < 0.05$ after the correction by step-down Bonferroni method. † $P < 0.05$, Wald test.

CI = confidence interval; ENT = ear, nose, and throat; LMA = laryngeal mask airway; OR = odds ratio; URI = upper respiratory tract infection.

Correction of univariate p-values

Table 2. Incidence of Respiratory Complications in the Two Groups of Children

	No URI, % (n = 608)	URI, % (n = 223)	OR	95% CI	*P* Value
Overall complications in the perioperative period					
Laryngospasm	3.1	7.6	2.558	1.305–5.016	0.007†
Bronchospasm		0.9	—	—	0.072
Airway obstruction	7.1	6.3	0.880	0.472–1.642	0.759
Oxygen desaturation	11.4	19.3	1.863	1.228–2.825	0.004†
Cough	7.5	17.9	2.730	1.728–4.313	<0.0001*
Overall‡	19.1	31.8	1.981	1.401–2.803	<0.0001*
Intraoperative complications					
Laryngospasm	3.5	6.9	2.044	1.005–4.157	0.069
Bronchospasm		1.0	—	—	0.073
Airway obstruction	4.7	4.4	0.926	0.426–2.012	1.000
Oxygen desaturation	2.6	5.0	1.972	0.861–4.513	0.107
Cough	4.6	8.9	2.008	1.071–3.766	0.034
Overall‡	9.5	15.2	1.713	1.063–2.760	0.035
Complications in the recovery room					
Laryngospasm	0.3	1.9	5.561	1.011–30.589	0.047
Bronchospasm			—	—	—
Airway obstruction	3.5	3.4	0.966	0.402–2.319	1.000
Oxygen desaturation	10.3	18.3	1.944	1.248–3.027	0.005†
Cough	4.6	14.0	3.409	1.955–5.942	<0.0001*
Overall‡	14.7	25.4	1.978	1.342–2.916	0.001*

* $P < 0.05$ after the correction by step-down Bonferroni method. † $P < 0.08$ after the correction by step-down Bonferroni method. ‡ Overall = percentage of individuals having at least one specific complication.

CI = confidence interval; OR = odds ratio; URI = upper respiratory tract infection.

Evaluation of logistic regression and relative regression models for data of Example 2.

Investigation of risk factors of respiratory complications in paediatric anaesthesia

- Background: Incidence of Adverse Respiratory Events in Children with Recent Upper Respiratory Tract Infections (URI) –Example 1. (Anesthesiology 2007; 107:714–9).
- Data: Outcome variables - complications (5 types):
  - Bronchospasm
  - Laryngospasm
  - Cough
  - Desaturation
  - <95%Airway obstruction
  - Overall
- Any of them might occur
  - at induction
  - during maintenance
  - On recovery - the three together are called intraoperative compl.
  - PACU (recovery room) – a 4 together are called perioperative complications

Risk factors

- Characteristics of the patient
  - Cold
    - Currently, <2 weeks, <4 weeks, none
    - Runny nose (several categories), cough (dry/moist), fewer
  - wheezing
  - Rhinitis
  - Eczema
  - The same factors in the family
    - mother/father/brother/>1 relatives
- Characteristic of anaesthesia
  - Maintained by registrar or consultant
  - Induction of anaesthesia
  - Maintenance of anaesthesia
  - Airway management (face mask/LMA/ETT) – further details
  - Timing
- Events at the recovery room (PACU)
- Original questionnaire RESPIRATORY COMPLICATIONS without boxes.doc

First steps

- Correcting mistakes in data base (! !)
- Univariate tests (all complications, all cases, too much) - χ^2 tests and odds ratios
- For example, odds of a female for bronchospasm: 81:3661=0.022125
- odds of a male 82:5472=0.01498
- A male has 0.01498/0.022125=0.6765 times less odds

Crosstab

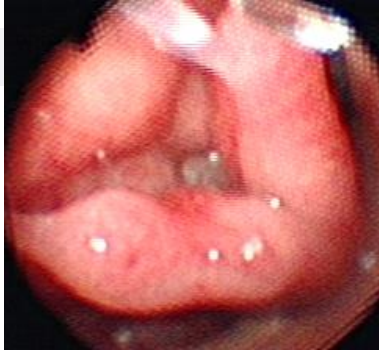
			Bronch			Overall p	p related to the first category	OR (unadjusted)	95%CI, lower	95%Ci upper
			no	Yes	Total					
sex	female	Count	3 661	81	3 742	0.015		0.677	0.497	0.923
		% within Bronch	40.1%	49.7%	40.3%					
	male	Count	5 472	82	5 554					
		% within Bronch	59.9%	50.3%	59.7%					
Total		Count	9 133	163	9 296					
		% within Bronch	100.0%	100.0%	100.0%					

Crosstab

			Bronch							
			no	Yes	Total					
When were the last symptoms	NONE	Count	6 067	74	6 141	0.000	0.000	2.737	1.854	4.042
		% within Bronch	66.5%	45.4%	66.1%		0.000	3.236	2.134	4.909
	Currently	Count	1 198	40	1 238		0.373	1.281	0.743	2.208
		% within Bronch	13.1%	24.5%	13.3%					
	<2 weeks	Count	836	33	869					
		% within Bronch	9.2%	20.2%	9.4%					
	<4 weeks	Count	1 024	16	1 040					
		% within Bronch	11.2%	9.8%	11.2%					

Unifactorial results

Example laryngospasm



Laryngospasm – odds ratios

Overall incidence	3.5 %
Registrar vs. consultants	2.5 (1.9-3.4), p < 0.001
No cold	
Current cold	3.2 (2.4-4.2), p = 0.005
Cold < 2 wks	4.3 (3.3-5.7), p < 0.001
Cold 2 - 4 wks	0.4 (0.2-0.8), p < 0.001

Impact of different symptoms

	Currently	2 weeks	2-4 weeks
Current cold	**3.2**	**4.3**	**0.4**
Clear nose	**2.0** (1.5-2.8), p<0.001	**2.1** (1.5-3.0), p<0.001	1.1 (0.6-2.0), p=0.67
Green nose	**5.0** (3.2-7.9), p<0.001	**8.2** (5.5-12.3), p<0.001	**0.1** (0.0-0.6), p=0.02
Dry cough	**2.3** (1.5-3.3), p<0.001	**2.2** (1.4-3.6), p<0.001	0.5 (0.2-1.3), p=0.15
Moist cough	**4.3** (3.1-6.0), p<0.001	**7.9** (5.7-10.9), p<0.001	**0.1** (0.0-0.6), p=0.01
Fever	**2.5** (1.1-5.4), p=0.024	**6.3** (3.8-10.5), p<0.001	0.6 (0.2-1.5), p=0.26

Medical history – related odds ratios

Ever eczema	1.9 (1.5-2.4), p < 0.001
Eczema < 12 months	2.0 (1.5-2.6), p < 0.001
Rhinitis < 12 months	1.0 (0.7-1.4), p = 0.930
Ever asthma	1.5 (1.1-1.9), p = 0.006
Wheezing episodes during last 12 months vs. none	
1-3	1.6 (1.2-2.3), p = 0.005
4-12	3.1 (2.1-4.7), p < 0.001
> 12	3.4 (2.0-6.1), p < 0.001
Wheezing during exercise	3.5 (2.7-4.6), p < 0.001
Dry cough at night	4.2 (3.3-5.3), p < 0.001

Multivariate analysis

- Given one binary outcome variable and a lot of independent variables (5-times)
- Model: INSTEAD OF a logistic regression relative risk regression (instead of a logit link log link – we get the estimation of the RR, not the OR)

Example. y=bronchospasm (1=yes, 0=no) x=sex (0 female, 0 male). Logistic regression

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	8.036	1	.005
	Block	8.036	1	.005
	Model	8.036	1	.005

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	1869.586[a]	.001	.005

a. Estimation terminated at iteration number 7 because parameter estimates changed by less than .001

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	Sex	-.414	.146	8.084	1	.004	.661	.497	.879
	Constant	-3.627	.103	1242.756	1	.000	.027		

a. Variable(s) entered on step 1: Sex.

Classification Table[a]

			Predicted		
			Bronchospasm periop		Percentage Correct
	Observed		.00 no	1.00 yes	
Step 1	Bronchospasm periop	.00 no	9104	0	100.0
		1.00 yes	193	0	.0
	Overall Percentage				97.9

a. The cut value is .500

Example. y=bronchospasm (1=yes, 0=no)
x=sex (0 female, 0 male) + age. Logistic regression

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	8.036	1	.005
	Block	8.036	1	.005
	Model	8.036	1	.005

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	1869.586[a]	.001	.005

a. Estimation terminated at iteration number 7 because parameter estimates changed by less than .001.

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	9.046	2	.011
	Block	9.046	2	.011
	Model	9.046	2	.011

LR:9.046-8.036=1.01

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	1868.575[a]	.001	.005

a. Estimation terminated at iteration number 7 because parameter estimates changed by less than .001.

LR:1869.586-1868.575=1.01

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	Sex	-.415	.146	8.112	1	.004	.661	.497	.879
	age	-.015	.015	.996	1	.318	.985	.955	1.015
	Constant	-3.533	.138	657.532	1	.000	.029		

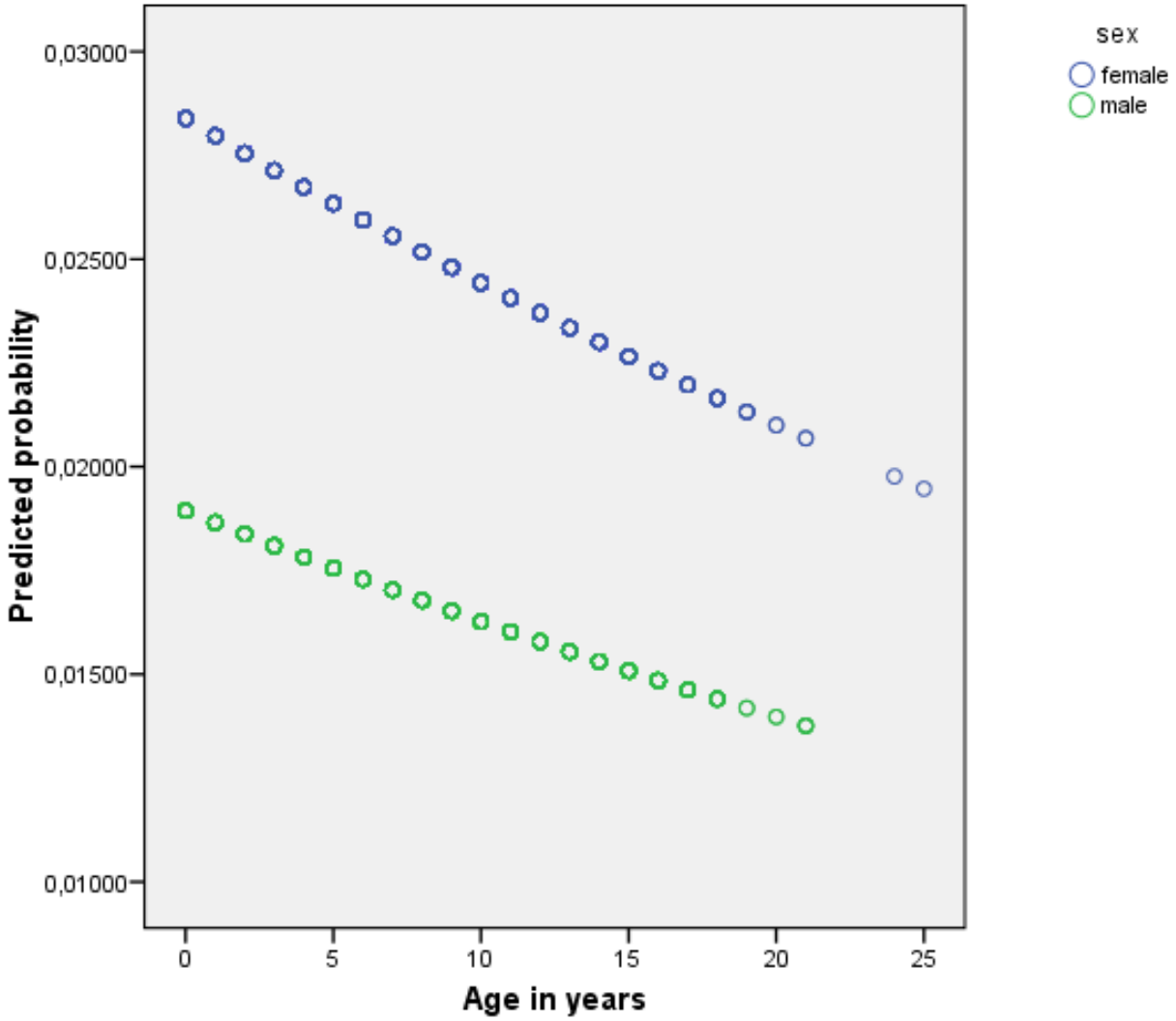
a. Variable(s) entered on step 1: Sex, age.

Classification Table[a]

			Predicted		
			Bronchospasm periop		Percentage Correct
	Observed		.00 no	1.00 yes	
Step 1	Bronchospasm periop	.00 no	9104	0	100.0
		1.00 yes	193	0	.0
	Overall Percentage				97.9

a. The cut value is .500

Estimated probabilities



Example. y=bronchospasm (1=yes, 0=no) x=sex (0 female, 0 male) +age. Rel.risk. regression

Omnibus Test[a]

Likelihood Ratio Chi-Square	df	Sig.
9.021	2	.011

Dependent Variable: Bronchospasm periop
Model: (Intercept), Sex, age

a. Compares the fitted model against

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test			Exp(B)	95% Wald Confidence Interval for Exp(B)	
			Lower	Upper	Wald Chi-Square	df	Sig.		Lower	Upper
(Intercept)	-3.563	.1342	-3.826	-3.300	704.838	1	.000	.028	.022	.037
[Sex=1]	-.405	.1424	-.684	-.126	8.088	1	.004	.667	.505	.882
[Sex=0]	0[a]	.	.	.	.	.	.	1	.	.
age	-.015	.0152	-.045	.015	.970	1	.325	.985	.956	1.015
(Scale)	1[b]									

Dependent Variable: Bronchospasm periop
Model: (Intercept), Sex, age

a. Set to zero because this parameter is redundant.

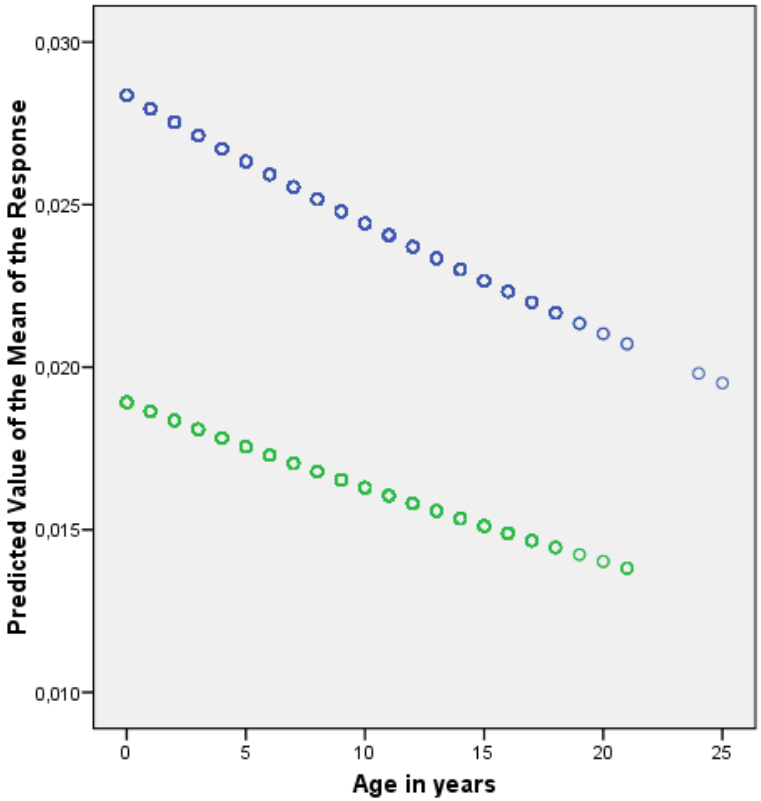
b. Fixed at the displayed value.

Variables in the Equation

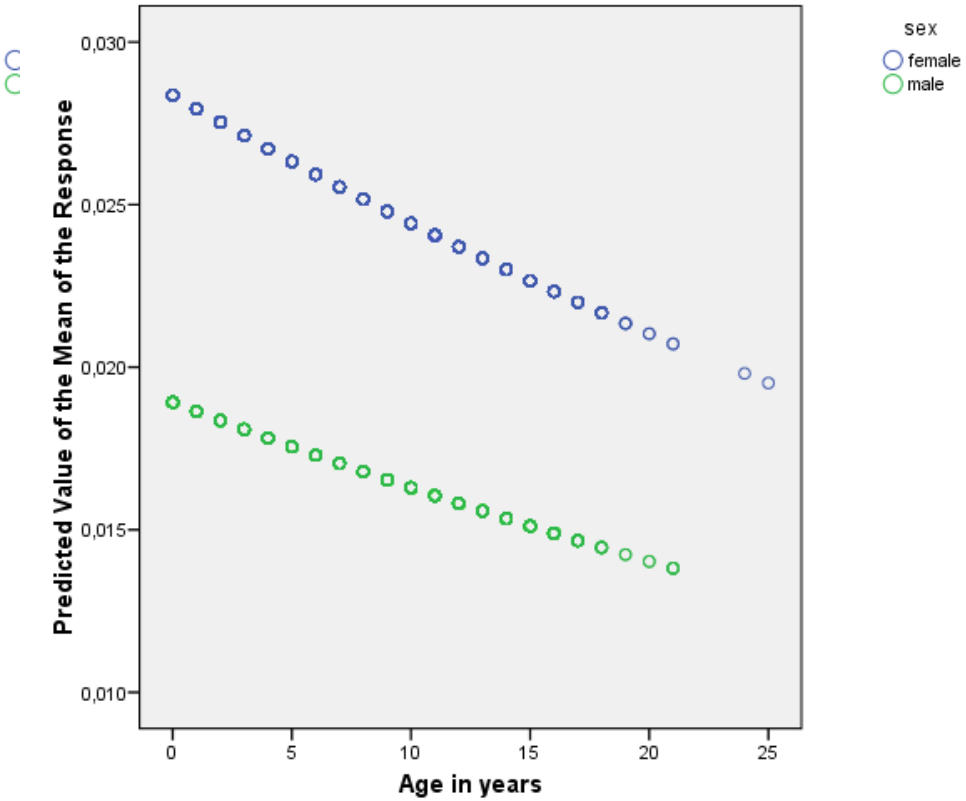
		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	Sex	-.415	.146	8.112	1	.004	.661	.497	.879
	age	-.015	.015	.996	1	.318	.985	.955	1.015
	Constant	-3.533	.138	657.532	1	.000	.029		

a. Variable(s) entered on step 1: Sex, age.

Log. regr.



Rel.riks. reg



The phenomenon of multicollinearity (example from another study)

Univariate logistic regressions

Variable	Code	Coeff	St.Err.	Wald	df	p
No. of oocytes	OOCYT	0.052	0.019	7.742	1	0.005
No. of mature oocytes	MII	0.066	0.022	8.687	1	0.003

Multivariate model (variables together)

Variable	Code	Coeff	St.Err.	Wald	df	p
No. of oocytes	OOCYT	0.011	0.045	0.063	1	0.802
No. of mature oocytes	MII	0.053	0.054	0.991	1	0.320

Simplifications

- We collapsed the last three complications, so we performed only 3 multivariate modelling
- We performed multivariate analysis only for the „overall” complication
- The problem of multicollinearity – we had a lot of variable expressing the same thing. The physician could not decide which is more important.

Factor analysis

- We performed factor analysis based on almost every independent variables.
- We have got reasonable factors.
- Instead of producing new artificial variables by factor analysis, we collapsed original variables belonging to the factors using the „or" logical operator. In multivariate models, age, gender, hayfever, airway management (TT, LMA or face mask) and the new collapsed variables (airway sensitivity, eczema, family history and anaesthesia) were examined.
  - Airwsusc1: wheezing>3 times or asthmaexercise or dry night cough or cold<2 weeks
  - Familyw: rhinitis or eczema or asthma or smoke int he family (>2 persons)
  - Anaest: Registrar or change anaesth or induction anaest.
- We decided to use the combined variables variables to examine the following complications: (1) Laryngospasm periop, (2) Brochospasm periop, (3) all others periop.
- Details: collapse.doc

Rotated Component Matrix[a]

	Component				
	1	2	3	4	5
BHR at exercise	.824				
dry night cough	.784		.153		
Wheezing >3 attacks	.722				
eczema last 12 months		.922			
ever eczema	.170	.897			
Rhinitis >2 persons in the family			.714		
Eczema >2 persons in the family			.664		
Asthma >2 persons in the family	.123		.660		
indanaest2				.735	-.139
Cold <2weeks			.108	.562	
ENT	.125			.334	
Airway management who?					.712
changeofanaesthetist				.351	.544
Smoke Mum and Dad			.135	-.120	.522

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 5 iterations.

Simplifications

- Simplifications of variables – where possible (worse scenario based on univariate statistics)
  - Asthma in the family, >2 persons
  - Hayfever in the family, >2 persons
  - Eczema in the family, >2 persons
  - Smoking in the family, Mother and Father
- Upper respiratory tract infection (URI) <2 weeks:
  - calles also positive respiratory history or airway susceptibility

Table 2. Incidence of respiratory adverse events in the 2 groups of children.

	Healthy n=7041		Positive respiratory history[1] n=2256		RR	95%CI		p-value	Absolute risk reduction	95% CI	
Overall complications in the perioperative period[2]											
Bronchospasm	52	0.7%	141	6.3%	8.463	6.179	11.590	<0.0001*	5.51%	4.49%	6.53%
Laryngospasm	151	2.1%	200	8.9%	4.134	3.365	5.079	<0.0001*	6.72%	5.50%	7.94%
Cough	319	4.5%	368	16.3%	3.600	3.123	4.151	<0.0001*	11.78%	10.18%	13.38%
Desaturation <95%	455	6.5%	464	20.6%	3.183	2.822	3.590	<0.0001*	14.11%	12.34%	15.87%
Airway obstruction	178	2.5%	154	6.8%	2.700	2.188	3.333	<0.0001*	4.30%	3.19%	5.40%
Overall [3]	693	9.8%	699	31.0%	3.148	2.866	3.457	<0.0001*	21.14%	19.11%	23.17%
Intraoperative complications											
Bronchospasm	30	0.4%	133	5.9%	13.835	9.336	20.501	<0.0001*	5.47%	4.49%	6.45%
Laryngospasm	142	2.0%	180	8.0%	3.956	3.191	4.904	<0.0001*	5.96%	4.80%	7.13%
Cough	267	3.8%	286	12.7%	3.343	2.849	3.922	<0.0001*	8.88%	7.44%	10.33%
Desaturation <95%	373	5.3%	389	17.2%	3.254	2.847	3.721	<0.0001*	11.94%	10.30%	13.59%
Airway obstruction	130	1.8%	136	6.0%	3.265	2.579	4.132	<0.0001*	4.18%	3.15%	5.21%
Overall [3]	584	8.3%	595	26.4%	3.179	2.866	3.527	<0.0001*	18.08%	16.15%	20.01%
Complications in the recovery room											
Bronchospasm	19	0.3%	68	3.0%	11.168	6.731	18.531	<0.0001*	2.74%	2.03%	3.46%
Laryngospasm	77	1.1%	91	4.0%	3.688	2.733	4.977	<0.0001*	2.94%	2.09%	3.79%
Cough	156	2.2%	162	7.2%	3.241	2.614	4.017	<0.0001*	4.96%	3.85%	6.08%
Desaturation <95%	163	2.3%	168	7.4%	3.216	2.607	3.969	<0.0001*	5.13%	3.99%	6.27%
Airway obstruction	39	0.6%	39	1.7%	3.121	2.007	4.852	<0.0001*	1.17%	0.61%	1.74%
Stridor	28	0.4%	30	1.4 %	3.344	2.002	5.584	<0.0001*	0.93%	0.44%	1.43%
Overall [3]	290	4.1%	307	13.6%	3.303	2.834	3.851	<0.0001*	9.49%	8.00%	10.98%

[1] Positive respiratory history: URI<2 weeks or wheezing at exercise or > 3 times wheezing during last 12 months or nocturnal dry cough

[2] Intraoperative complications + PACU

[3] Bronchospasm or Laryngospasm or Cough or Desaturation <95% or Airway obstruction

* p<0.0001 after the correction by step-down Bonferroni method

Table 3 a Relative risk and 95% confidence interval (CI) for the risk factors associated with the occurrence for perioperative bronchospasm.

Variable	Univariate				Multivariate			
	p	RR	95%CI		p	RR	95%CI	
Age	0.325	0.985	0.956	1.015	-	-	-	-
Gender	0.004	0.667	0.505	0.882				
Hayfever	0.000	2.915	2.153	3.947				
Upper respiratory tract infection (URI) <2 weeks	0.000	2.146	1.498	3.075				
Wheezing at exercise	0.000	7.730	5.870	10.178				
Wheezing >3 times in the last 12 months	0.000	7.168	5.307	9.680				
Nocturnal dry cough	0.000	10.510	7.932	13.927				
Airway sensitivity	**0.000**	**8.463**	**6.179**	**11.590**	**0.000**	**5.653**	**4.089**	**7.816**
Eczema in the last 12 months	0.000	3.158	2.359	4.227				
Ever eczema	0.000	4.575	3.444	6.077				
Eczema	**0.000**	**4.533**	**3.416**	**6.016**	**0.000**	**2.601**	**1.950**	**3.470**
Asthma in the family, >2 persons	0.000	4.415	3.082	6.325				
Hayfever in the family, >2 persons	0.000	3.753	2.426	5.808				
Eczema in the family, >2 persons	0.028	2.190	1.089	4.401				
Smoking in the family, Mother and Father	0.000	2.603	1.894	3.576				
Family history	**0.000**	**2.932**	**2.212**	**3.887**	**0.000**	**1.863**	**1.413**	**2.458**
Airway managed by registrar vs. pediatric anesthesia consultant	0.000	3.847	2.473	5.984				
Inhalational induction of anesthesia	0.000	2.381	1.791	3.167				
Change of anesthesiologist during airway management	0.000	4.094	2.646	6.335				
Anesthesia	**0.000**	**3.872**	**2.163**	**6.929**	**0.000**	**3.078**	**1.727**	**5.484**
ENT surgery	0.043	1.458	1.012	2.101	-	-	-	-
Face mask vs. laryngeal mask (LMA)	0.118	1.933	0.846	4.418	0.304	1.538	0.677	3.493
Face mask vs. tracheal tube (TT)	0.000	5.105	2.252	11.574	0.002	3.523	1.564	7.937

Table 3 b Relative risk and 95% confidence interval (CI) for the risk factors associated with the occurrence for perioperative laryngospasm[1]

Variable	Univariate				Multivariate			
	p	RR	95%CI		P	RR	95%CI	
age	0.000	0.894	0.871	0.918	0.000	0.903	0.879	0.926
Gender	0.038	0.805	0.655	0.988				
Hayfever	0.820	1.036	0.762	1.409				
Upper respiratory tract infection (URI) <2 weeks	0.000	3.341	2.657	4.202				
Wheezing at exercise	0.000	3.279	2.605	4.128				
Wheezing >3 times in the last 12 months	0.000	2.644	1.955	3.577				
Nocturnal dry cough	0.000	3.973	3.224	4.897				
Airway sensitivity	**0.000**	**4.134**	**3.365**	**5.079**	**0.000**	**3.261**	**2.654**	**4.008**
Eczema in the last 12 months	0.000	1.912	1.507	2.426				
Ever eczema	0.000	1.848	1.493	2.288				
Eczema	**0.000**	**1.917**	**1.553**	**2.365**	**-**	**-**	**-**	**-**
Asthma in the family, >2 persons	0.000	3.767	2.877	4.932				
Rhinitis in the family, >2 persons	0.000	3.108	2.222	4.347				
Eczema in the family, >2 persons	0.000	3.127	2.093	4.671				
Smoking in the family, Mother and Father	0.000	3.005	2.403	3.758				
Family history	**0.000**	**3.391**	**2.765**	**4.158**	**0.000**	**2.571**	**2.101**	**3.146**
Airway managed by registrar vs. pediatric anesthesia consultant	0.000	2.353	1.791	3.091				
Inhalational induction of anesthesia	0.000	3.202	2.574	3.984				
Change of anesthesiologist during airway management	0.000	4.479	3.310	6.061				
Anesthesia	**0.000**	**4.248**	**2.713**	**6.652**	**0.000**	**3.098**	**1.985**	**4.836**
ENT surgery	0.000	1.853	1.446	2.374	0.042	1.293	1.01	1.656
Face mask vs. laryngeal mask (LMA)	0.000	6.716	2.501	18.036	0.001	5.227	1.954	13.985
Face mask vs. tracheal tube (TT)	0.000	11.629	4.326	31.260	0.000	7.572	2.825	20.295

Table 3 c Relative risk and 95% confidence interval (CI) for the risk factors associated with the occurrence of perioperative cough, desaturation and airway obstruction.

Variable	Univariate				Multivariate			
	p	RR	95%CI		p	RR	95%CI	
Age	0.000	0.941	0.930	0.952	0.000	0.954	0.943	0.964
Gender	0.744	1.018	0.917	1.129				
Hayfever	0.000	1.382	1.207	1.581				
Upper respiratory tract infection (URI) <2 weeks	0.000	1.973	1.734	2.244				
Wheezing at exercise	0.000	3.043	2.732	3.390				
Wheezing >3 times in the last 12 months	0.000	2.572	2.236	2.958				
Nocturnal dry cough	0.000	3.443	3.118	3.803				
Airway sensitivity	**0.000**	**3.048**	**2.761**	**3.366**	**0.000**	**2.371**	**2.142**	**2.624**
Eczema in the last 12 months	0.000	1.887	1.681	2.118				
Ever eczema	0.000	1.770	1.592	1.967				
Eczema	**0.000**	**1.824**	**1.644**	**2.023**	**0.000**	**1.254**	**1.138**	**1.382**
Asthma in the family, >2 persons	0.000	2.551	2.206	2.951				
Rhinitis in the family, >2 persons	0.000	2.298	1.919	2.751				
Eczema in the family, >2 persons	0.000	3.023	2.499	3.658				
Smoking in the family, Mother and Father	0.000	1.950	1.728	2.200				
Family history	**0.000**	**2.086**	**1.879**	**2.315**	**0.000**	**1.545**	**1.403**	**1.701**
Airway managed by registrar vs. pediatric anesthesia consultant	0.000	1.932	1.698	2.199				
Inhalational induction of anesthesia	0.000	1.971	1.779	2.183				
Change of anesthesiologist during airway managment	0.000	4.483	3.978	5.053				
Anesthesia	**0.000**	**2.168**	**1.827**	**2.572**	**0.000**	**1.797**	**1.521**	**2.124**
ENT surgery	0.000	1.884	1.673	2.121	0.080	1.098	0.989	1.219
Face mask vs. laryngeal mask (LMA)	0.009	1.440	1.096	1.892	0.169	1.207	0.923	1.580
Face mask vs. tracheal tube (TT)	0.000	3.757	2.873	4.913	0.000	2.703	2.073	3.525

	Univariate						Multivariate (n=9256)	
	Yes		No		RR (95% CI)	p value	RR (95% CI)	p value
	Total	Value	Total	Value				
Age	4·95 (4·67)		6·41 (4·81)		0·94 (0·93–0·95)	<0·0001	0·95 (0·94–0·96)	<0·0001*
Male	5554	767 (14%)	3743	508 (14%)	1·02 (0·92–1·13)	0·74	..	..
Hayfever	1163	209 (18%)	8088	1052 (13%)	1·38 (1·21–1·58)	<0·0001	..	..
Positive respiratory history								
Upper respiratory tract infection <2 weeks	869	215 (25%)	8420	1056 (13%)	1·97 (1·73–2·24)	<0·0001	..	..
Wheezing at exercise	872	306 (35%)	8386	967 (12%)	3·04 (2·73–3·39)	<0·0001	..	..
Wheezing >3 times in past 12 months	478	156 (33%)	8819	1119 (13%)	2·57 (2·24–2·96)	<0·0001	..	..
Nocturnal dry cough	1161	421 (36%)	8100	853 (11%)	3·44 (3·12–3·80)	<0·0001	..	..
Any of the above	2256	630 (28%)	7041	645 (9%)	3·05 (2·76–3·37)	<0·0001	2·37 (2·14–2·62)	<0·0001*
Eczema								
In the past 12 months	1307	300 (23%)	7942	966 (12%)	1·89 (1·68–2·12)	<0·0001	..	..
Ever (excluding past 12 months)	2181	442 (20%)	7038	806 (11%)	1·77 (1·59–1·97)	<0·0001	..	..
Any of the above	2235	465 (21%)	7021	801 (11%)	1·82 (1·64–2·02)	<0·0001	1·25 (1·14–1·38)	<0·0001†
Family history								
Asthma in ≥2 family members	571	160 (28%)	8040	883 (11%)	2·55 (2·21–2·95)	<0·0001	..	..
Rhinitis in ≥2 family members	349	96 (28%)	8336	998 (12%)	2·30 (1·92–2·75)	<0·0001	..	..
Eczema in ≥2 family members	210	75 (36%)	8507	1005 (12%)	3·02 (2·50–3·66)	<0·0001	..	..
Both parents smokers	1075	259 (24%)	8222	1016 (12%)	1·95 (1·73–2·20)	<0·0001	..	..
Any of the above	1808	427 (24%)	7489	848 (11%)	2·09 (1·88–2·32)	<0·0001	1·55 (1·40–1·70)	<0·0001*
Anaesthesia								
Airway managed by registrar	6219	1015 (16%)	3078	260 (8%)	1·93 (1·70–2·20)	<0·0001	..	..
Inhalational induction of anaesthesia	3597	707 (20%)	5686	567 (10%)	1·97 (1·78–2·18)	<0·0001	..	..
Change of anaesthesiologist during airway management	269	150 (56%)	9021	1122 (12%)	4·48 (3·98–5·05)	<0·0001	..	..
Any of the above	7398	1140 (15%)	1899	135 (7%)	2·17 (1·83–2·57)	<0·0001	1·80 (1·52–2·12)	<0·0001*
Type of surgery								
Otolaryngology	1189	276 (23%)	8108	999 (12%)	1·88 (1·67–2·12)	<0·0001	1·10 (0·99–1·22)	0·080
Airway management device used								
Laryngeal mask vs face mask	5586	520 (9%)	820	53 (6%)	1·44 (1·10–1·89)	0·009	1·21 (0·92–1·58)	0·17
Tracheal tube vs face mask	2891	702 (24%)	820	53 (6%)	3·76 (2·87–4·91)	<0·0001	2·70 (2·07–3·53)	<0·0001*

Data are mean (SD) or number (%). RR=relative risk. *p<0·0001 after correction by the step-down Bonferroni method. †p=0·0003 after correction.

Table 7: **Risk factors associated with perioperative cough, desaturation, or airway obstruction**

SPSS command

```
GENLIN
  Bronchp (REFERENCE=FIRST)
  BY Airwsusc1 Ecz Familyw anaest airwman1 airwman2
  (ORDER=DESCENDING)
  /MODEL
  Airwsusc1 Familyw Ecz anaest airwman1 airwman2
  INTERCEPT=YES
  DISTRIBUTION=BINOMIAL
  LINK=LOG
 /CRITERIA METHOD=Fisher(1) SCALE=1 COVB=MODEL
  MAXITERATIONS=100 MAXSTEPHALVING=5
  PCONVERGE=1E-006(ABSOLUTE)
  SINGULAR=1E-012
  ANALYSISTYPE=3 CILEVEL=95 LIKELIHOOD=FULL
 /MISSING CLASSMISSING=EXCLUDE
 /PRINT CPS DESCRIPTIVES MODELINFO FIT SUMMARY SOLUTION(EXPONENTIATED)
  HISTORY(1).
```

Iteration History

Iteration	Update Type	Number of Step-halvings	Log Likelihood[a]	Parameter (Intercept)	[Airwsusc1=1.00]	[Familyw=1.00]	[Ecz=1.00]	[Anaest=1.00]	[airwman1=1.00]	[airwman2=1.00]	(Scale)
0	Initial	0	-1622.970	-2.603609	.586187	.241612	.323985	.207150	.027234	.265989	1
1	Scoring	0	-979.984	-3.725293	.760321	.313236	.435436	.299653	.062280	.406510	1
2	Newton	0	-810.400	-4.939239	1.081661	.441552	.632300	.495755	.135855	.664498	1
3	Newton	0	-771.490	-6.101388	1.459448	.569899	.844339	.792959	.257608	.987343	1
4	Newton	0	-766.149	-6.833161	1.682051	.617562	.941553	1.038349	.380784	1.198677	1
5	Newton	0	-765.968	-7.026670	1.730405	.622326	.955653	1.118692	.427035	1.255880	1
6	Newton	0	-765.968	-7.037537	1.732235	.622393	.956000	1.124206	.430231	1.259285	1
7	Newton	0	-765.968	-7.037571	1.732237	.622393	.956001	1.124226	.430241	1.259296	1
8	Newton[b]	0	-765.968	-7.037571	1.732237	.622393	.956001	1.124226	.430241	1.259296	1

Redundant parameters are not displayed. Their values are always zero in all iterations.
Dependent Variable: Bronchospasm periop
Model: (Intercept), Airwsusc1, Familyw, Ecz, Anaest, airwman1, airwman2

a. The full log likelihood function is displayed.

b. All convergence criteria are satisfied.

Goodness of Fit[b]

	Value	df	Value/df
Deviance	51.168	41	1.248
Scaled Deviance	51.168	41	
Pearson Chi-Square	56.254	41	1.372
Scaled Pearson Chi-Square	56.254	41	
Log Likelihood[a]	-765.968		
Akaike's Information Criterion (AIC)	1545.936		
Finite Sample Corrected AIC (AICC)	1548.736		
Bayesian Information Criterion (BIC)	1559.035		
Consistent AIC (CAIC)	1566.035		

Dependent Variable: Bronchospasm periop
Model: (Intercept), Airwsusc1, Familyw, Ecz, Anaest, airwman1, airwman2

a. The full log likelihood function is displayed and used in computing information criteria.

b. Information criteria are in small-is-better form.

Omnibus Test[a]

Likelihood Ratio Chi-Square	df	Sig.
343.961	6	.000

Dependent Variable: Bronchospasm periop
Model: (Intercept), Airwsusc1, Familyw, Ecz, Anaest, airwman1, airwman2

a. Compares the fitted model against the intercept-only model.

Tests of Model Effects

Source	Type III Wald Chi-Square	df	Sig.
(Intercept)	660.178	1	.000
Airwsusc1	109.823	1	.000
Familyw	19.420	1	.000
Ecz	42.263	1	.000
Anaest	14.548	1	.000
airwman1	1.056	1	.304
airwman2	9.233	1	.002

Dependent Variable: Bronchospasm periop
Model: (Intercept), Airwsusc1, Familyw, Ecz, Anaest, airwman1, airwman2

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test			Exp(B)	95% Wald Confidence Interval for Exp(B)	
			Lower	Upper	Wald Chi-Square	df	Sig.		Lower	Upper
(Intercept)	-7.038	.4850	-7.988	-6.087	210.553	1	.000	.001	.000	.002
[Airwsusc1=1.00]	1.732	.1653	1.408	2.056	109.823	1	.000	5.653	4.089	7.816
[Airwsusc1=.00]	0[a]	.	.	.	.	.	.	1	.	.
[Familyw=1.00]	.622	.1412	.346	.899	19.420	1	.000	1.863	1.413	2.458
[Familyw=.00]	0[a]	.	.	.	.	.	.	1	.	.
[Ecz=1.00]	.956	.1471	.668	1.244	42.263	1	.000	2.601	1.950	3.470
[Ecz=.00]	0[a]	.	.	.	.	.	.	1	.	.
[Anaest=1.00]	1.124	.2947	.547	1.702	14.548	1	.000	3.078	1.727	5.484
[Anaest=.00]	0[a]	.	.	.	.	.	.	1	.	.
[airwman1=1.00]	.430	.4186	-.390	1.251	1.056	1	.304	1.538	.677	3.493
[airwman1=.00]	0[a]	.	.	.	.	.	.	1	.	.
[airwman2=1.00]	1.259	.4144	.447	2.072	9.233	1	.002	3.523	1.564	7.937
[airwman2=.00]	0[a]	.	.	.	.	.	.	1	.	.
(Scale)	1[b]									

Dependent Variable: Bronchospasm periop
Model: (Intercept), Airwsusc1, Familyw, Ecz, Anaest, airwman1, airwman2

a. Set to zero because this parameter is redundant.

b. Fixed at the displayed value.

Likelihood ration test for the variable age

Omnibus Test[a]

Likelihood Ratio Chi-Square	df	Sig.
344.110	7	.000

Dependent Variable: Bronchospasm periop
Model: (Intercept), Airwsusc1, Familyw, Ecz, Anaest, airwman1, airwman2, age

a. Compares the fitted model against the intercept-only model.

Omnibus Test[a]

Likelihood Ratio Chi-Square	df	Sig.
343.961	6	.000

Dependent Variable: Bronchospasm periop
Model: (Intercept), Airwsusc1, Familyw, Ecz, Anaest, airwman1, airwman2

a. Compares the fitted model against the intercept-only model.

Chi-square (with age)	=344.11	df=7
Chi-square (without age)	=343.961	df=6
Difference:	0.149	df=1

Not significant at 0.05 level

So adding variable age does not increase significantly the model chi-square, i.e., does not decrease significantly the deviance D=-2logL.

Part of the review of New England Journal of Medicine

9. Which "…statistically significant variables were not included into the set of candidate variables"? What was the rationale for this exclusion?

10. With so many variables evaluated, was there a power analysis to justify the number of subjects, number of RAEs, and the number of variables in question? Type I errors should be discussed.

11.
Was there some statistical addressing the multiple comparisons, such as a Bonferonni (or equivalent) correction?

The authors could explore using propensity scores to which may assist in giving some idea of adjusted absolute risk reduction.

Next: Lancet

- There were no main problems concerning statistics
- But based on question of the reviewers, we had to put new univariate statistics into the text of the manuscript.
- What can we do against the increase of Type I error?

Other problems during the analysis

- I misunderstood the meaning of some variables (recovery room – at recovery)
- The problem of decimal digits
- The problem of frequencies

Correction of p-values: Step-down Bonferroni method

- I corrected every p-values occuring in the tables or text, and they remain significant at $p<0.05$ level (sample size: 10000, $p=10^{-27}$!!!)
- Based on new requests, the number of p-values changed during the process
- Repeated 4 times
- Question: publish original or corrected p-values?
- Result: corrected p-value were published – it contradicts to the level of confidence intervals which were not corrected

Table 5. Risk factors for perioperative bronchospasm, laryngospasmon the timing of symptoms and all respiratory adverse events (bronchospasm, laryngospasm, desaturation, severe coughing, airway obstruction, stridor) as compared to no symptom.

Data are presented as relative risk (RR) and 95% confidence interval.

	Bronchospasm			Laryngospasm			All complications		
	Currently	**<2 weeks**	**2-4 weeks**	**Currently**	**<2 weeks**	**2-4 weeks**	**Currently**	**<2 weeks**	**2-4 weeks**
Clear runny nose	2.0 (1.3-3.0) p=0.001*	1.1 (0.6-2.0) p=0.738	1.1 (0.5-2.2) p=0.900	2.0 (1.5-2.7) p<0.0001***	2.0 (1.5-2.9) p<0.0001***	1.1 (0.7-1.9) p=0.672	1.5 (1.3-1.8) p<0.0001***	1.4 (1.1-1.7) p=0.001*	1.0 (0.7-1.3) p=0.740
Green runny nose	1.9 (0.9-4.3) p=0.107	2.4 (1.1-4.9) p=0.023	0.8 (0.3-1.8) p=0.514	4.4 (3.0-6.5) p<0.0001***	6.6 (4.8-9.1) p<0.0001***	0.1 (0.01-0.6) p=0.015	3.1 (2.6-3.8) p<0.0001***	3.4 (2.8-4.1) p<0.0001***	0.2 (0.1-0.4) p<0.0001***
Dry cough	1.7 (0.96-2.9) p=0.071	2.1 (1.2-3.8) p=0.015	0.6 (0.2-1.8) p=0.327	2.2 (1.5-3.1) p<0.0001**	2.1 (1.4-3.3) p=0.001*	0.5 (0.2-1.3) p=0.155	1.7 (1.4-2.1) p<0.0001***	1.9 (1.5-2.3) p<0.0001***	0.3 (0.2-0.6) p<0.0001***
Moist cough	3.3 (2.1-5.0) p<0.0001***	4.0 (2.6-6.3) p<0.0001***	0.3 (0.1-1.1) p=0.069	3.9 (2.9-5.2) p<0.0001***	6.5 (5.0-8.5) p<0.0001***	0.1 (0.01-0.6) p=0.012	3.1 (2.6-3.5) p<0.0001***	3.4 (2.9-4.0) p<0.0001***	0.5 (0.3-0.7) p<0.0001**
Fever	4.2 (2.0-8.7) p<0.0001**	2.0 (0.8-5.3) p=0.164	0.8 (0.3-2.4) p=0.645	2.3 (1.1-4.8) p=0.020	5.3 (3.5-8.0) p<0.0001***	0.6 (0.2-1.5) p=0.259	2.9 (2.2-3.8) p<0.0001***	2.9 (2.3-3.8) p<0.0001***	0.5 (0.3-0.9) p=0.017

* : p<0.05 after the correction by step-down Bonferroni method
** : p<0.01 after the correction by step-down Bonferroni method
***: p<0.001 after the correction by step-down Bonferroni method

Consequences

- We published the paper in the Lancet.
  - Title: Risk assessment for respiratory complications in paediatric anaesthesia: a prospective cohort study
    Author(s): von Ungern-Sternberg BS, Boda K, Chambers NA, et al.
    Source: **LANCET** Volume: **376** Issue: **9743** Pages: **773-783** Published: **SEP 4 2010**
- The big sample size is important
- Appropriate data set is important
- Good cooperation with the physician is necessary
- Statistician should know a little bit biology
- Was this statistics good enough? Can we continue the research? Propensity score analysis?

Reactions

- We have already references
- It is really interesting meanly from medical point of view
- The week-end Australian , West Australian

Triggers found on asthma risk

ADAM CRESSWELL
HEALTH EDITOR

CHILDREN with a dry night-time cough, past or present eczema or who wheeze during exercise have more than eight times the risk of suffering an asthma attack while under anaesthetic — a situation that can be fatal if corrective drugs fail.

The same factors put affected children at four times the risk that their vocal cords will lock up while under sedation, a situation that requires the prompt use of muscle-relaxing drugs to allow air back into the lungs.

The findings — the result of Australian research involving nearly 10,000 children undergoing operations at a Perth hospital — suggest doctors can more accurately predict which children are at highest risk for bad reactions by asking whether they share these risk factors.

The research also indicates that a child who has recently had a cold or another airway infection is at increased risk for an anaesthetic reaction for only two weeks afterwards — the first time doctors have had a clear idea of how long the post-infectious danger period lasts.

Serious adverse events are extremely rare among children under anaesthetic. The research, published yesterday in the British medical journal *The Lancet*, showed that 1392, or 15 per cent, of the 9297 children for whom full data was available had respiratory adverse events during or soon after their operations. However, pediatric anaesthetist Britta von Ungern-Sternberg, who ran the study at Perth's Princess Margaret Hospital for Children, said most of these events were minor and transient, and only a tiny sub-group experienced lasting harm. "It's pinpointed the risk factors — how we can predict the children who are going to have complications," she said.

Andrew Davidson, staff anaesthetist at the Royal Children's Hospital in Melbourne, said it was "a significant study because it has really crystallised a lot of the previous 'maybes' about which children are likely to get complications".

New surgery check

CATHY O'LEARY

Doctors at Princess Margaret Hospital have developed a checklist for children having surgery which uses details such as a history of asthma and exposure to passive smoking to reduce the risk of complications during operations.

Their study of more than 9200 children, published in the medical journal The Lancet, used a revised medical questionnaire about the child's health and family history of disease as well as a physical examination to identify children most at risk of adverse events.

Professor Britta von Ungern-Sternberg, who chairs paediatric anaesthesia at PMH, said the checklist could identify those at higher risk and help determine their pre-operative care and the anaesthetic used.

References

1. A. Agresti: Categorical Data Analysis 2nd. Edition. Wiley, 2002
2. A.J. Dobson: An introduction to generalized linear models. Chapman &Hall, 1990.
3. D.W. Hosmer and S.Lemeshow: Applied Logistic Regression. Wiley, 2000.
4. T. Lumley, R. Kronmal, S. Ma: Relative Risk Regression in Medical Research: Models, Contrasts, Estimators,and Algorithms. UW Biostatistics Working Paper Series University of Washington, Year 2006 Paper 293 http://www.bepress.com/uwbiostat/paper293
5. SAS Institute, Inc: The MIXED procedure in SAS/STAT Software: Changes and Enhancements through Release 6.11. Copyright © 1996 by SAS Institute Inc., Cary, NC 27513.
6. SPSS Advanced Models 9.0. Copyright © 1996 by SPSS Inc P.

- If had only one day left to live, I would live it in my statistics class:
  it would seem so much longer.
- Mathematical Jokes: Statistics