

## Techniques de régression pour mesures répétées : exemple de l'analyse des chutes à l'hôpital

F. R. Herrmann

Dpt. of Rehabilitation et Geriatrics  
University Hospitals of Geneva, Switzerland



1

## Background

---

Some events (hospitalization, stroke, falls,..) occurs repeatedly but too often, only the first occurrence is described in the literature.

The repeated nature of falls provides an opportunity to describe a wide spectrum of statistical analysis techniques used for repeated risk modeling.

2

## Methods

---

Here we propose to improve probability estimates of conditions characterized by their repeated nature, like stroke or falls by using incidence data.

The selection of the appropriate model will depend on the research question, the study design and the type of the dependent variable.

3

## Epidemiological Measures

---

### Frequency

Outcomes

### Association

Strength of the relationship « Risk factor – Outcome »

### Impact

Factor contribution to an outcome frequency

4

## Frequency measures

### Prevalence (P)

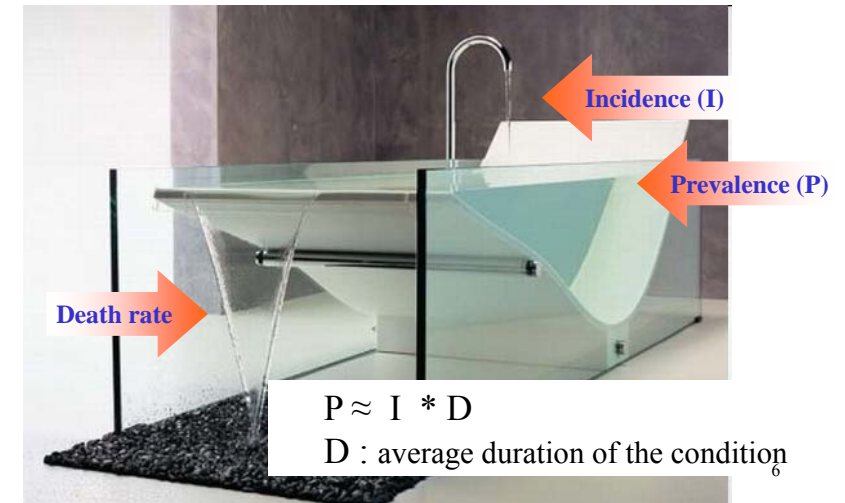
Number of individual with a condition during a time period or at a given time, in a defined population.

### Incidence (I)

Number of new cases with a condition

5

## Frequency measures



6

## Prevalence / cumulative incidence

Exposure	Outcome		Total
	I+	I-	
E+	A	B	A+B
E-	C	D	C+D
<b>Total</b>	<b>A+C</b>	<b>B+D</b>	<b>N</b>

Prevalence of exposure =  $A+B/N$   
 Prevalence of non exposure =  $C+D/N$   
 Prevalence /cumulative incidence of + outcome =  $A+C/N$

7

## Epidemiological Measures

### Frequency

Outcomes

### Association

Strength of risk factor - outcome relationship

### Impact

Factor contribution to an outcome frequency

8

## Risk

Exposure	Outcome Condition		Total
	I+	I-	
E+	A	B	A+B
E -	C	D	C+D
<b>Total</b>	A+C	B+D	<b>N</b>

Risk of I+ among the exposed =  $A / A+B$   
 Risk of I+ among the non-exposed =  $C / C+D$

9

## Relative risk (RR)

Exposure	Outcome Condition		Total
	I+	I-	
E+	A	B	A+B
E -	C	D	C+D
<b>Total</b>	A+C	B+D	<b>N</b>

Risk of I+ among the exposed =  $A / A+B$   
 Risk of I+ among the non-exposed =  $C / C+D$

$$RR = \frac{R(E+)}{R(E-)} = \frac{A/A+B}{C/C+D}$$

10

## Odds ratio (OR)

Exposure	Outcome Condition		Total
	I+	I-	
E+	A	B	A+B
E -	C	D	C+D
<b>Odds</b>	A / C	B / D	<b>N</b>

$$OR = \frac{\frac{A}{B}}{\frac{C}{D}} = \frac{A/B}{C/D} = \frac{A/C}{B/D} = \frac{AD}{CB}$$

11

## Incidence rate ratio (IRR) Hazard ratio (HR)

Exposure	Issue		TI
	I+	PT	
E+	A	PT <sub>1</sub>	A/PT <sub>1</sub>
E -	C	PT <sub>2</sub>	C/PT <sub>2</sub>
<b>Total</b>	A+C	PT <sub>1</sub> +PT <sub>2</sub>	

$$IRR = \frac{\text{Incidence rate E+}}{\text{Incidence rate E-}} = \frac{TI_1}{TI_2} = \frac{A/PT_1}{C/PT_2}$$

PT = person-time

12

## RR, OR, IRR, HR

**Units:** none

**Range:** [ 0 ; +∞]

**Interpretation :**

RR, OR, IRR, HR < 1 : Exposure decreases the risk

RR, OR, IRR, HR = 1 : No risk – outcome association

RR, OR, IRR, HR > 1 : Exposure increases the risk

13

## Epidemiological Measures

### Frequency

Outcomes

### Association

Strength of the Factor - Outcome relationship

### Impact

Factor contribution to an outcome frequency

14

## Impact

**Risk differences = Attributable risk = %X-%Y**

**Number needed to treat (NNT)**

**Number needed-to-harm (NNH)**

$$\frac{1}{\text{attributable risk}}$$

15

## Attributable risk (AR)

Exposure	Outcome Condition		Total
	I+	I-	
E+	A	B	A+B
E-	C	D	C+D
<b>Total</b>	A+C	B+D	<b>N</b>

$$AR = R(E+) - R(E-) = \frac{A}{A+B} - \frac{C}{C+D}$$

16

## Results

**Results** are illustrated with a systematic data collection of falls occurring in a 298 beds, acute and rehabilitation geriatric teaching hospital.

Over a 10 y. period 7'795 falls among 13'949 patients.

**Petitpierre NJ, Trombetti A, Carroll I, Michel JP, Herrmann FR.** The FIM(R) instrument to identify patients at risk of falling in geriatric wards: a 10-year retrospective study. *Age Ageing* 2010.

17

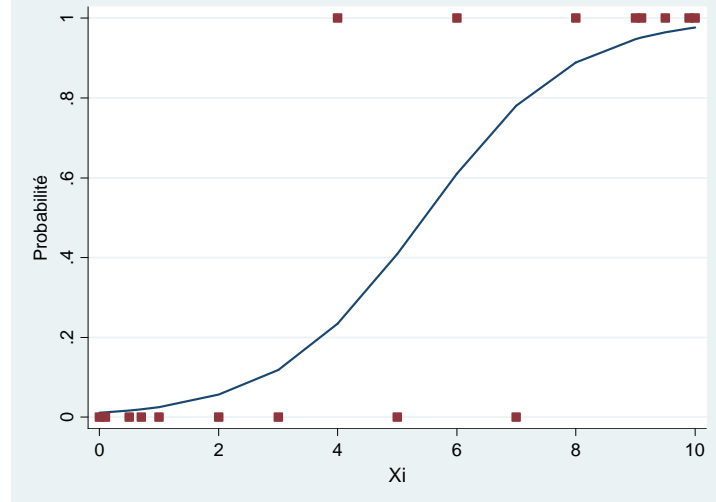
(Mouse Mickey. 1927 -...)

## Regression models and falls

Dependent Var.	Statistical Unit	Regression
Binary	Non faller Faller	Logistic General linear model
Polytomous	Non faller One time faller Recurrent faller	Ordered logistic regression
Discrete	Number of falls	Poisson Negative Binomiale
Time dependent binary	Date of each fall	Cox + Andersen-Gill

18

## Logistic regression



19

## Logistic regression

$$y = \text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon_i$$

$$p = \frac{e^y}{1 + e^y} = \frac{e^{\text{logit}(p)}}{1 + e^{\text{logit}(p)}} = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon_i}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon_i}}$$

$$\text{OR}_i = e^{\beta_i}$$

20

# Logistic regression

xi:logistic nbchuteb sex ageentree

Logistic regression	Number of obs =	24787
	LR chi2(2) =	159.94
	Prob > chi2 =	0.0000
Log likelihood = -12103.581	Pseudo R2 =	0.0066

	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
nbchuteb					
sex	1.312103	.0456875	7.80	0.000	1.225545 1.404776
ageentree	1.025537	.0024166	10.70	0.000	1.020811 1.030284

xi:logistic nbchuteb sex ageentree, cluster(nopatient)

Logistic regression	Number of obs =	24787
	Wald chi2(2) =	135.83
	Prob > chi2 =	0.0000
Log pseudolikelihood = -12103.581	Pseudo R2 =	0.0066

(Std. Err. adjusted for 13949 clusters in nopatient)

	Odds Ratio	Robust Std. Err.	z	P> z	[95% Conf. Interval]
nbchuteb					
sex	1.312103	.0496708	7.18	0.000	1.218274 1.413159
ageentree	1.025537	.0025927	9.97	0.000	1.020468 1.030631

21

# Logistic regression: verification of log linear assumption via quartiles (not recommended)

xtile ageentreeq4 = ageentree , nq(4)  
xi:logistic nbchuteb ib1.ageentreeq4

ageentree	Freq.	Percent	Cum.
1	6,199	25.01	25.01
2	6,195	24.99	50.00
3	6,201	25.02	75.02
4	6,192	24.98	100.00
Total	24,787	100.00	

xi:logistic nbchuteb ib1.ageentreeq

Logistic regression	Number of obs =	24787
	LR chi2(3) =	89.91
	Prob > chi2 =	0.0000
Log likelihood = -12138.593	Pseudo R2 =	0.0037

	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
nbchuteb					
ageentreeq					
2	1.257631	.0598122	4.82	0.000	1.145699 1.380499
3	1.377974	.0647011	6.83	0.000	1.256822 1.510804
4	1.522846	.070626	9.07	0.000	1.390526 1.667756

22

# Logistic regression: Verification of log linear assumption via equally spaced interval

logistic nbchuteb i.age10

age10	Freq.	Percent	Cum.
60	790	3.19	3.19
70	5,717	23.06	26.25
80	12,675	51.14	77.39
90	5,508	22.22	99.61
100	97	0.39	100.00
Total	24,787	100.00	

Logistic regression	Number of obs =	24787
	LR chi2(4) =	99.04
	Prob > chi2 =	0.0000
Log likelihood = -12134.029	Pseudo R2 =	0.0041

	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
nbchuteb					
age10					
70	1.426757	.1642071	3.09	0.002	1.138634 1.787787
80	1.820831	.2032454	5.37	0.000	1.463041 2.266119
90	2.081443	.2374854	6.42	0.000	1.664353 2.603057
100	2.821794	.7098811	4.12	0.000	1.723407 4.620222

# Logistic regression: Verification of log linear assumption via equally spaced interval

logistic nbchuteb i.age5r if age5r > 55

Logistic regression	Number of obs =	24787
	LR chi2(8) =	105.37
	Prob > chi2 =	0.0000
Log likelihood = -12130.865	Pseudo R2 =	0.0043

Freq.	nbchuteb	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
124	age5r					
666	65	1.198864	.3788063	0.57	0.566	.6453793 2.227025
1,880	70	1.570002	.4709604	1.50	0.133	.8720906 2.826434
3,837	75	1.712455	.5074279	1.82	0.069	.9580639 3.060863
5,893	80	2.023439	.5969175	2.39	0.017	1.13497 3.607413
6,782	85	2.215025	.6527065	2.70	0.007	1.243235 3.946427
4,482	90	2.458422	.7260282	3.05	0.002	1.378088 4.385667
1,026	95	2.303632	.6977952	2.75	0.006	1.272257 4.171106
97	100	3.293318	1.220098	3.22	0.001	1.593248 6.807444

23



## General linear model

```
glm chute sexe ageentree, family(bin) link(log) eform vce(cluster nopatient)
```

```
Generalized Linear models           No. of obs   =   24787
Optimization      : ML              Residual df  =   24784
Deviance          = 24208.33594      Scale parameter =   1
Pearson           = 24773.71739      (1/df) Deviance = .9767728
                                           (1/df) Pearson = .9995851

Variance function: V(u) = u*(1-u)    [Bernoulli]
Link function     : g(u) = ln(u)     [Log]

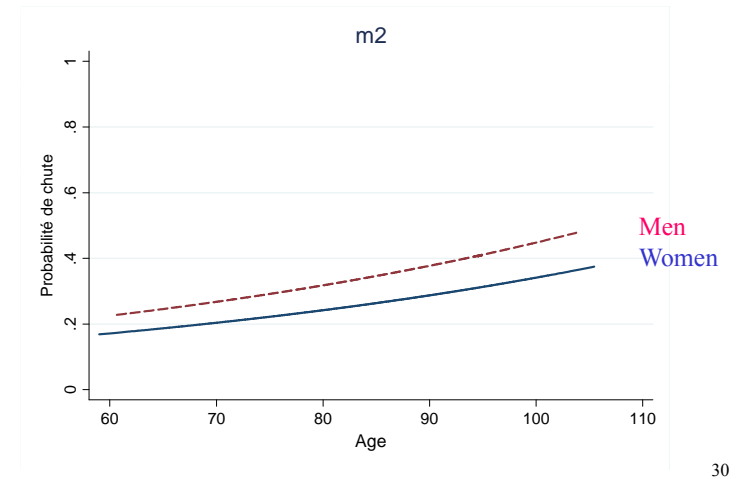
Log pseudolikelihood = -12104.16797   AIC          = .9768966
                                           BIC          = -226558
```

(Std. Err. adjusted for 13949 clusters in nopatient)

nbchuteb	Risk Ratio	Robust Std. Err.	z	P> z	[95% Conf. Interval]
sexe	1.240248	.0368743	7.24	0.000	1.170042 1.314668
ageentree	1.020314	.0020524	10.00	0.000	1.0163 1.024345

29

## General linear model



## Poisson regression

### Model a discrete, positive variable

- Rare event ( $N < 100$ )
- ie: number of falls
- $E(Y) = \text{Var}(Y) = \lambda$
- $\lambda$  parameter allows to modify the shape of the distribution

31

## Poisson regression

$$\Pr[Y_i = y_i] = \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!}, y_i = 0, 1, 2, \dots$$

$$\log \lambda_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik}$$

# of expected event

$$E[y_i | x_i] = \lambda_i = e^{x_i' \beta}$$

32





# Binomial negative regression

- Extension of the Poisson model to correct for over dispersion
- Include a noise parameter

$$\log \lambda_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \sigma \varepsilon_i$$

# Binomial negative regression

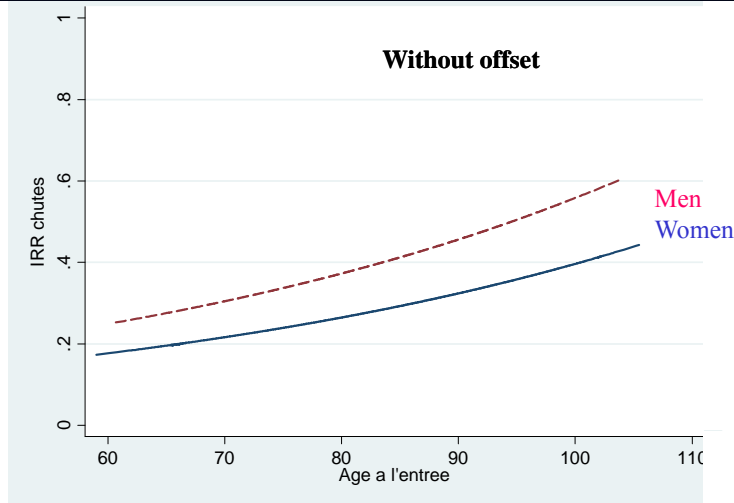
xi:nbreg nbchute sexe ageentree , irr cluster(nopatient)

Negative binomial regression      Number of obs =      **24787**  
 Dispersion = **mean**      Wald chi2(2) =      **111.29**  
 Log pseudolikelihood = **-17481.849**      Prob > chi2 =      **0.0000**

(Std. Err. adjusted for **13949** clusters in nopatient)

nbchute	IRR	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
sexe	<b>1.408448</b>	<b>.0579207</b>	<b>8.33</b>	<b>0.000</b>	<b>1.29938</b>	<b>1.526671</b>
ageentree	<b>1.020393</b>	<b>.0027368</b>	<b>7.53</b>	<b>0.000</b>	<b>1.015043</b>	<b>1.025771</b>
/lnal pha	<b>1.273565</b>	<b>.0352127</b>			<b>1.204549</b>	<b>1.34258</b>
al pha	<b>3.573569</b>	<b>.125835</b>			<b>3.335255</b>	<b>3.828911</b>

# Binomial negative regression



# Binomial negative regression

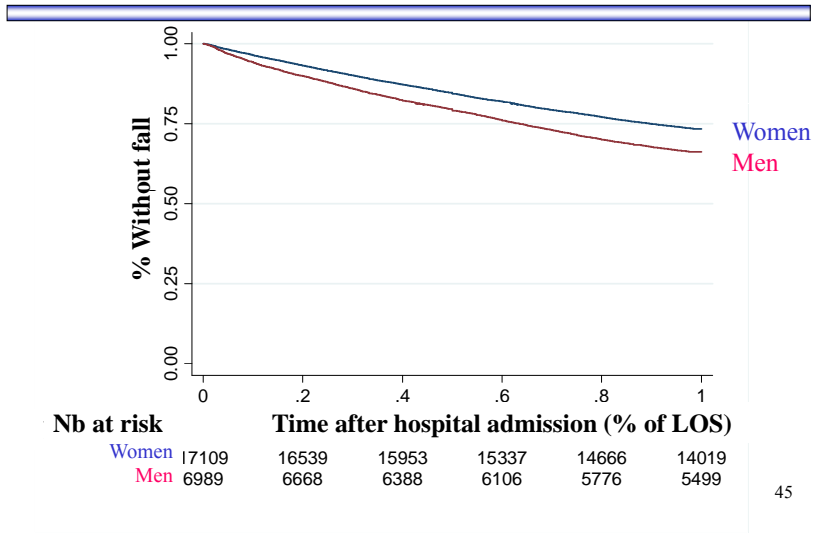
$$P(y = r) = \frac{(\lambda_i t_i)^r e^{-\lambda_i t_i}}{r!}$$

$$\log \lambda_i = \log(t_i) + \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \sigma \varepsilon_i$$

Adjusted for the time of exposure (los)

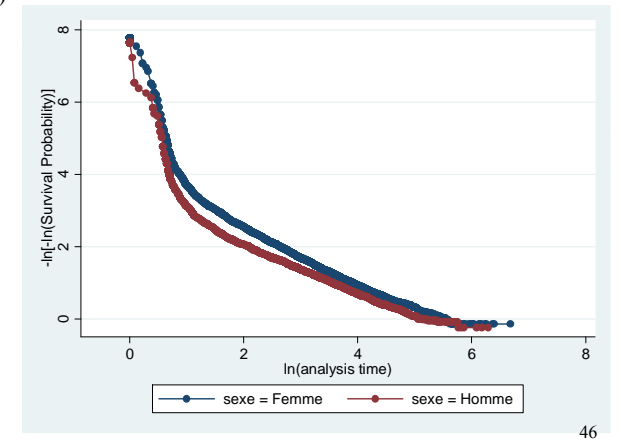


# Cox regression



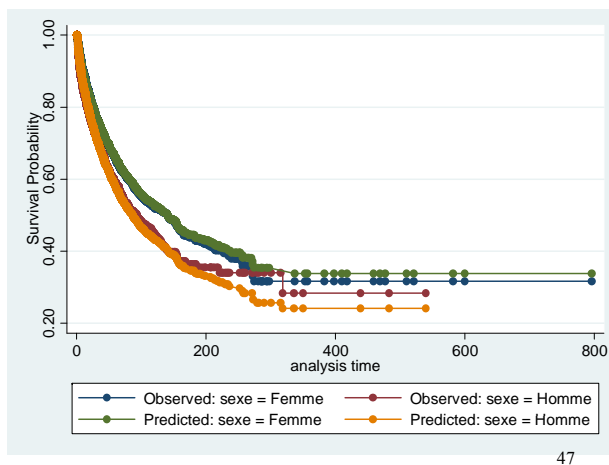
# Cox regression: Verification 1 of proportional hazard assumption

stphplot, by(sexe)



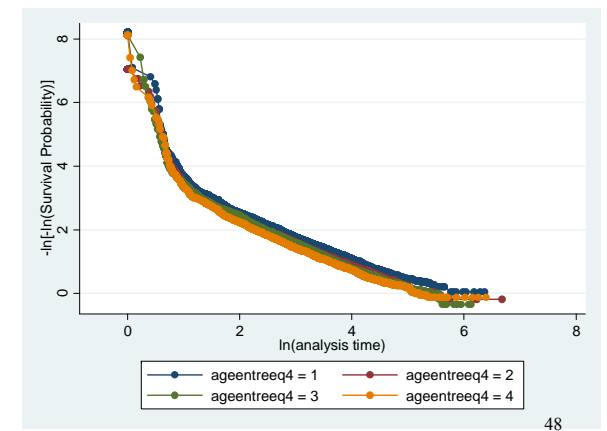
# Cox regression: Verification 2 of proportional hazard assumption

stcoxkm, by(sexe)



# Cox regression: Verification 3 of proportional hazard assumption

stphplot, by(ageentreeq4)



# Cox regression: Verification 4 of proportional hazard assumption

estat phtest, det

## Test of proportional -hazards assumption

Time: Time

	rho	chi 2	df	Prob>chi 2
sexe	-0.04118	6.73	1	0.0095
ageentree	-0.00580	0.13	1	0.7208
global test		6.73	2	0.0346

note: robust variance-covariance matrix used.

# Cox regression (modified according to Andersen-Gill)

stset tbf3, fail(nbchuteb==1) exit(time.) id(nopatent) enter(time 0)

stcox sexe ageentree, efron robust nolog

Cox regression -- Efron method for ties

No. of subjects = 13925 Number of obs = 26634  
 No. of failures = 7780  
 Time at risk = 635394.0909

Log pseudolikelihood = -67316.882 Wald chi2(2) = 127.47  
 Prob > chi2 = 0.0000

(Std. Err. adjusted for 13925 clusters in nopatient)

_t	Haz. Ratio	Robust Std. Err.	z	P> z	[95% Conf. Interval]
sexe	1.51168	.0605413	10.32	0.000	1.397559 1.63512
ageentree	1.016567	.0027541	6.06	0.000	1.011183 1.021979

Andersen PK and Gill RD. Cox's Regression Model for Counting Processes: A Large Sample Study- *Ann. Stat.* 1982; 4 (10): 1100-20.

# Summary of regression models

Regression Model	Parameter	Short	Sex			Age		
			Value	95 % CI	Value	95 % CI		
Logistic	Odds ratio	OR	1.32	1.23 1.42	1.03	1.02 1.03		
General linear model	Risk ratio	RR	1.25	1.18 1.32	1.02	1.02 1.03		
Ordered logistic regression	Odds ratio	OR	1.34	1.24 1.44	1.03	1.02 1.03		
Poisson	Incidence rate ratio	IRR	1.40	1.30 1.51	1.02	1.02 1.03		
Negative binomial	Incidence rate ratio	IRR	1.40	1.30 1.51	1.02	1.02 1.03		
Negative binomial + offset	Incidence rate ratio	IRR	1.53	1.42 1.65	1.01	1.01 1.02		
Cox modified according to Andersen-Gill	Hazard ratio	HR	1.51	1.40 1.64	1.02	1.01 1.02		

Herrmann FR, Petitpierre NJ. Techniques de régression pour l'analyse des facteurs de risque de chute. *Annales de Gérontologie* 2009;2(4):225-29.

# Discussion

The results produced by the different models are quite equivalent (risk of falls 1.2 to 1.5 times higher in men, and increases significantly by 1.2 to 2.6 % with each year of age) but addresses different research question:

## Discussion

---

**Logistic model** predict who will fall or not

**Poisson model** address the number of falls

It assumes that the probability of observing a fall must be the same in each individual, over time and not depending on the previous falls suffered by the individual, (occurrence independence).

Navarro A. *et al.*, *Prev Med* **48**, 298-302 (2009).

53

## Discussion

---

**Binomial models** address the number of falls

“It assumes that the count of falls follows a Poisson whose  $\lambda$  parameter, in turn, follows a Gamma distribution (i.e. the parameter varies, rather than being fixed).”

**Cox model** predicts the speed at which falls occur

The term “hazard” is used to describe the concept of risk of occurrence in an infinitesimal interval after time  $t$ , conditional on the subject having survived to time  $t$ .

Navarro A. *et al.*, *Prev Med* **48**, 298-302 (2009).

54

## Discussion

---

**Cox model** predicts the speed at which falls occur

“The term “hazard” is used to describe the concept of risk of occurrence in an infinitesimal interval after time  $t$ , conditional on the subject having survived to time  $t$ .”

- Hazard ratio (HR) for two subjects at risk, with fixed covariate vectors, remains constant over time.
- Probability of observing a fall do not need to be constant along the follow-up.

Navarro A. *et al.*, *Prev Med* **48**, 298-302 (2009).

55

## Discussion

---

**Cox model**

The event can occur only once

- a) analyze only the first occurrence
- b) treat each observation as independent, not taking into account that there may be more than one per individual.

Navarro A, Ancizu I, *Prev Med* **48**, 298-302 (2009)

56

## Discussion

---

### Andersen–Gill model

“Marginal models incorporate recurrence which involves estimating the coefficients ignoring the dependence between observations and subsequently correcting the naive variance through robust estimators. But the hazard of suffering a fall is independent of previous falls that the same individual may have experienced, (occurrence independence).”

Navarro A, Ancizu I, *Prev Med* **48**, 298-302 (2009)  
Andersen PK, Gill RD, 1982. *Ann. Stat.* **10**, 1100–1120.

57

## Discussion

---

### Prentice–Williams–Peterson (PWP)

“An individual cannot be at risk of suffering fall  $s$  until they have suffered fall  $s-1$ . For this reason this model is commonly referred to as a “conditional model”. Hence this is a proportional hazards model with time-dependent strata, where the dependence between times of falls is handled through stratifying by the number of previous occurrences and presenting each strata its own baseline hazard.”

Navarro A, Ancizu I, *Prev Med* **48**, 298-302 (2009)  
Prentice, RL, Williams, BJ, Peterson, AV. *Biometrika* **68**, 373–379(1981)

58

## Medline Bibliometrics (10.1.2011)

---

N	%	Key words
554	10.9	Logistic
8	0.2	General linear model
0	0.0	Ordered logistic
45	0.9	Poisson
9	0.2	Binomial negative
81	1.6	Cox
2	0.0	Andersen–Gill
1	0.0	Prentice–Williams–Peterson
<b>5095</b>	<b>100.0</b>	<b>Falls risk factors</b>

59

## Conclusions

---

For commodity reasons or lack of the appropriate software many studies with repeated outcomes reports only the occurrence of a first event, but to limit information loss, model dealing with repeated measure design are recommended so that all observed events are considered in risk modeling, thus avoiding data loss.

60

# Effect of Music-Based Multitask Training on Gait, Balance, and Fall Risk in Elderly People

A Randomized Controlled Trial

Andrea Trombetti, MD; Melany Hars, PhD; François R. Herrmann, MD, MPH; Reto W. Kressig, MD; Serge Ferrari, MD; René Rizzoli, MD

## Methods

- 12-month RCT
- 134 community-dwelling individuals > 65y at increased risk of falling.
- Intervention group (n=66) vs delayed intervention control group (n=68).

**Intervention** : 6-month multitask exercise program performed to the rhythm of piano music. 1-hour weekly class.

61

## Outcomes

- Change in gait variability under dual-task condition from baseline to 6 months was the primary end point.
- Secondary outcomes = changes in:
  - Balance
  - functional performances
  - fall risk

Trombetti A., Hars M, Herrmann FR *et al.*, Effect of Music-Based Multitask Training on Gait, Balance, and Fall Risk in Elderly People: A Randomized Controlled Trial. *Arch Intern Med* (2010).

62

## Results

### At 6 months, in the intervention group)

- Reduction in stride length variability (−1.4%;  $P < .002$ ) under dual-task condition
- Improvement in balance and functional tests
- Fewer falls (IRR, 0.46; 95% CI 0.27-0.79)
- Lower risk of falling (RR, 0.61; 95% CI 0.39-0.96)

### In the delayed intervention control

- Similar changes during the second 6-month period with intervention

Trombetti A., Hars M, Herrmann FR *et al.*, Effect of Music-Based Multitask Training on Gait, Balance, and Fall Risk in Elderly People: A Randomized Controlled Trial. *Arch Intern Med* (2010).

63

Table 4. Falls at the 6-Month Follow-up

Outcomes	Early Intervention (n=66)	Delayed Intervention (n=68)	Unadjusted	Adjusted <sup>a</sup>	Method
Falls, rate <sup>b</sup>	24 (0.7)	54 (1.6)			
IRR (95% CI)			0.46 (0.27-0.79) <sup>d</sup>	0.49 (0.27-0.91) <sup>c</sup>	Negative binomial regression model
Participants with ≥1 fall, No. (%)	19 (28.8)	32 (47.1)			
RR (95% CI)			0.61 (0.39-0.96) <sup>c</sup>	0.69 (0.44-1.07)	Log-binomial regression model
Participants with multiple (≥2) falls, No. (%)	3 (4.6)	16 (23.5)			
RR (95% CI)			0.19 (0.06-0.63) <sup>d</sup>	0.21 (0.06-0.67) <sup>d</sup>	Log-binomial regression model
Survival analysis					
HR (95% CI)			0.53 (0.30-0.94) <sup>c</sup>	0.55 (0.31-0.99) <sup>c</sup>	Cox proportional hazards model
HR (95% CI)			0.46 (0.27-0.78) <sup>d</sup>	0.46 (0.27-0.79) <sup>d</sup>	Andersen-Gill model

Abbreviations: CI, confidence interval; HR, hazard ratio; IRR, incidence rate ratio; RR, relative risk.  
<sup>a</sup>Adjusted for age, history of falls over the previous 12 months, simplified Tinetti test performance, and total number of frailty criteria (according to Fried *et al*)<sup>1</sup> met.

<sup>b</sup>Fall rates per person per year.

<sup>c</sup> $P < .05$ .

<sup>d</sup> $P < .01$ .

Trombetti A., Hars M, Herrmann FR *et al.*, Effect of Music-Based Multitask Training on Gait, Balance, and Fall Risk in Elderly People: A Randomized Controlled Trial. *Arch Intern Med* (2010).

64



## Conclusion

In community-dwelling older people at increased risk of falling, a 6-month music-based multitask exercise program

- improved gait under dual-task condition,
- improved balance,
- and reduced both the rate of falls and the risk of falling.