```
        ``
      .0O1.^
      u$ON=1
      z00BAl
     l..=~.
    ;s<'''
    NRX~=-`
    z0c^<X^
    ~B0s~^`
     @@$H~'
    n$0=XN;.`
   iBBB0vU1=~'`
   `$@00cRr`vul
    FAHZuqr-'
    ZZUFA@Fl.`
   ;BRHv n$U^-
  `ARN1    ^@si
 'Onv~      O1.'
 cOqr       rs.`
 aUU`        ul`
`RO-          :.`
nn~`          -=.~l-`
=1^'..`       `..`
```

REVES - CUBA

2010

Which regression models to use with repeated incidence data

F. R. Herrmann

Dpt. of Rehabilitation et Geriatrics

University Hospitals of Geneva, Switzerland

Hôpitaux Universitaires de Genève

UNIVERSITÉ DE GENÈVE

FACULTÉ DE MÉDECINE

Background

The calculation of Disability-Free Life Expectancy (DFLE) by the Sullivan method makes use of information regarding the prevalence (proportion) of a condition.

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

Methods

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

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

Methods

The repeated nature of falls provides an opportunity to describe a wide spectrum of statistical analysis techniques used for repeated risk modeling. The selection of the appropriate model will depend on the research question, the study design and the type of the dependent variable, which can be either dichotomous (faller versus non faller), ordinal (non faller, one time and recurrent faller), continuous (number of falls over the study period) or time dependant (date and time of each fall) and will guide the choice of the corresponding regression model:


(c) 2006 Martin Waugh
www.liquidsculpture.com

Epidemiological Measures

Frequency

Outcomes

Association

Strength of the relationship « Risk factor – Outcome »

Impact

Factor contribution to an outcome frequency

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

Frequency measures

$P \approx I * D$

D : average duration of the condition

Prevalence / cumulative incidence

	Outcome		
Exposure	**I+**	**I-**	**Total**
E+	A	B	A+B
E -	C	D	C+D
Total	**A+C**	B+D	**N**

Prevalence of exposure = A+B / N

Prevalence of non exposure = C+D / N

Prevalence /cumulative incidence of + outome = A+C / N

Epidemiological Measures

Frequency

Outcomes

Association

Strength of risk factor - outcome relationship

Impact

Factor contribution to an outcome frequency

Risk

	Outcome Condition		
Exposure	**I+**	**I-**	**Total**
E+	A	B	A+B
E -	C	D	C+D
Total	A+C	B+D	**N**

Risk of I+ among the exposed = A / A+B

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

Relative risk (RR)

	Outcome Condition		
Exposure	**I+**	**I-**	**Total**
E+	A	B	A+B
E -	C	D	C+D
Total	A+C	B+D	**N**

Risk of I+ among the exposed = A / A+B

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

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

Odds ratio (OR)

	Outcome Condition		
Exposure	**I+**	**I-**	**Total**
E+	A	B	A+B
E -	C	D	C+D
Odds	A / C	B / D	**N**

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

Incidence rate ratio (IRR) Hazard ratio (HR)

	Issue		
Exposure	**I+**	**PT**	**TI**
E+	A	PT_1	A/PT_1
E -	C	PT_2	C/PT_2
Total	A+C	PT_1+PT_2	

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

PT = person-time

RR, OR, IRR, HR

Units: none

Range: [0 ; +∞]

Interpretation :

RR, OR, RTI < 1 : Exposure decreases the risk

RR, OR, RTI = 1 : No risk – outcome association

RR, OR, RTI > 1 : Exposure increases the risk

Epidemiological Measures

Frequency

Outcomes

Association

Strength of the Factor - Outcome relationship

Impact

Factor contribution to an outcome frequency

Impact

Risk differences = Attributable risk = %X-%Y

Number needed to treat (NNT)

Number needed-to-harm (NNH)

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

Attributable risk (AR)

	Outcome Condition		
Exposure	**I+**	**I-**	**Total**
E+	A	B	A+B
E -	C	D	C+D
Total	A+C	B+D	**N**

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


www.liquidsculpture.com
(c) 2006 Martin Waugh

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.

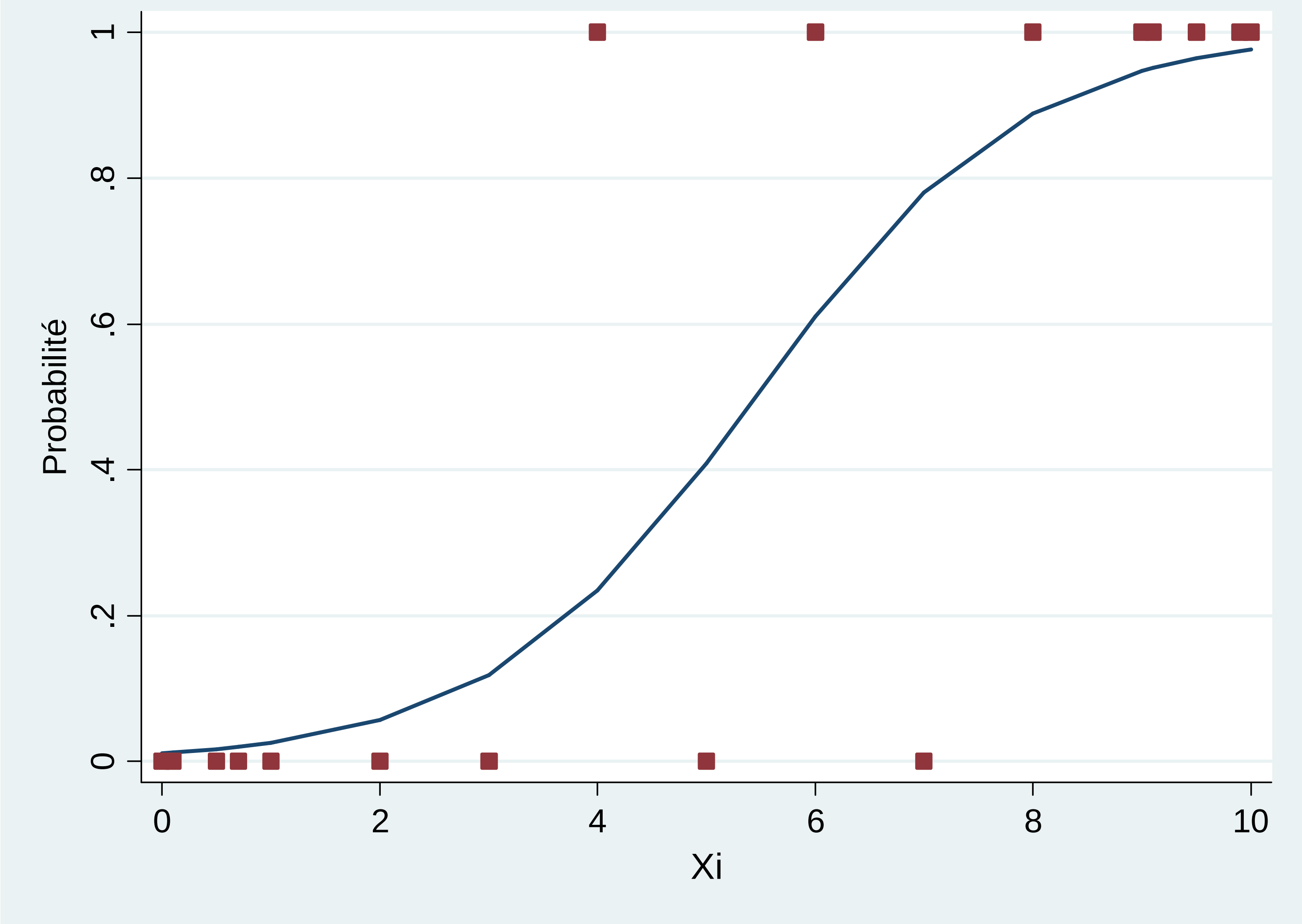
Petitpierre NJ, Trombetti A, Carroll I, Michel JP, Herrmann FR. The FIM(R) instrument to identify patients at risk of falling in geriatric wards: a 10-year retrospective study. *Age Ageing* 2010.

(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	Niumber of falls	Poisson Negative Binomiale
Time dependent binary	Date of each fall	Cox + Andersen-Gill

Logistic regression



Logistic regression

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

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

$$OR_i = e^{\beta_i}$$

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
```

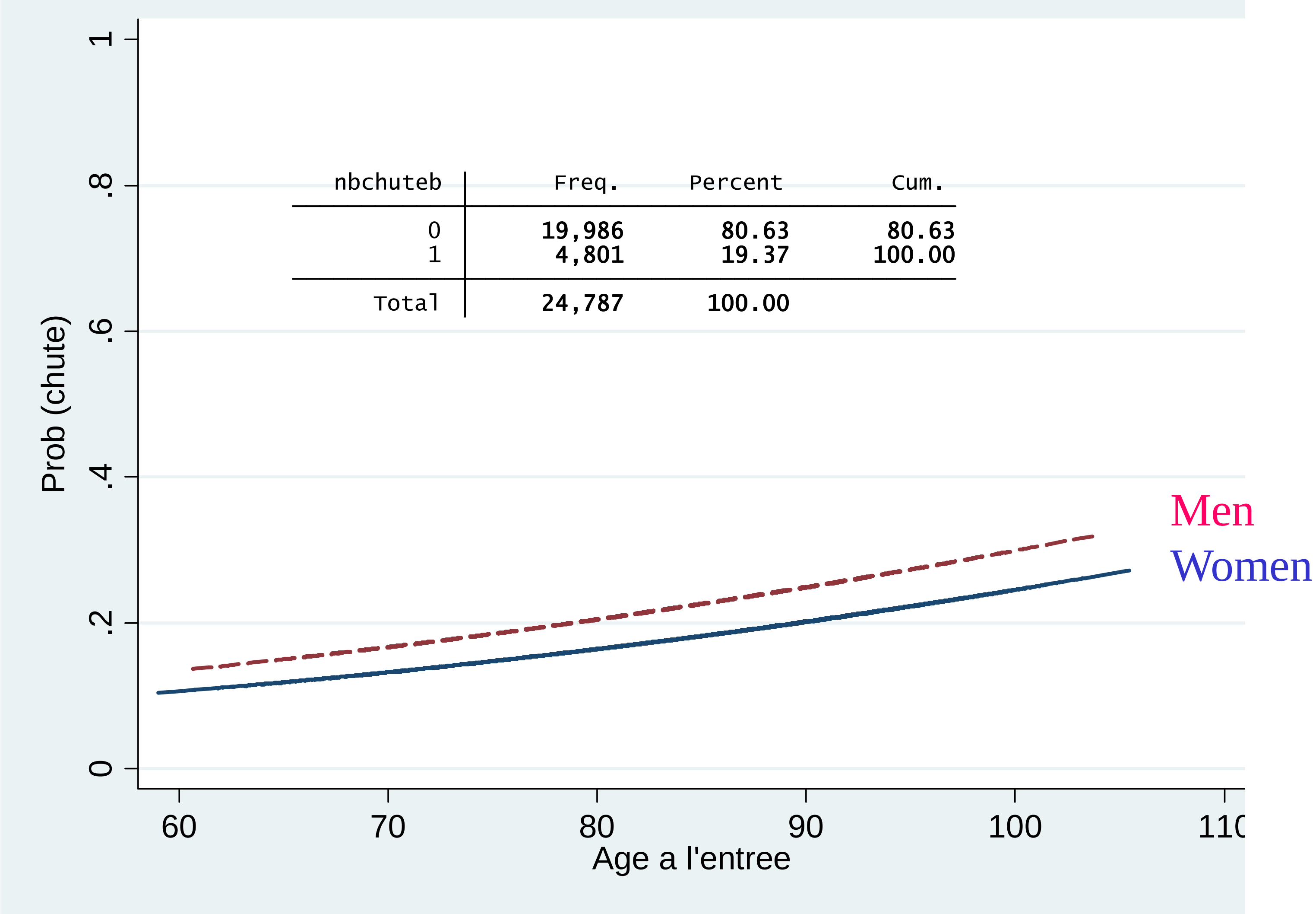
nbchuteb	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
sexe	1.312103	.0456875	7.80	0.000	1.225545	1.404776
ageentree	1.025537	.0024166	10.70	0.000	1.020811	1.030284

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)
```

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

Logistic regression



nbchuteb	Freq.	Percent	Cum.
0	19,986	80.63	80.63
1	4,801	19.37	100.00
Total	24,787	100.00	

Ordered logistic regression

$$P(\text{issue}_j = i) = P\left(k_{i-1} < \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \ldots + \beta_k x_{kj} + u_i \le k_i\right)$$

u_i	Follows a logistic distribution
k	Number of outcome
K_i	Cutpoint

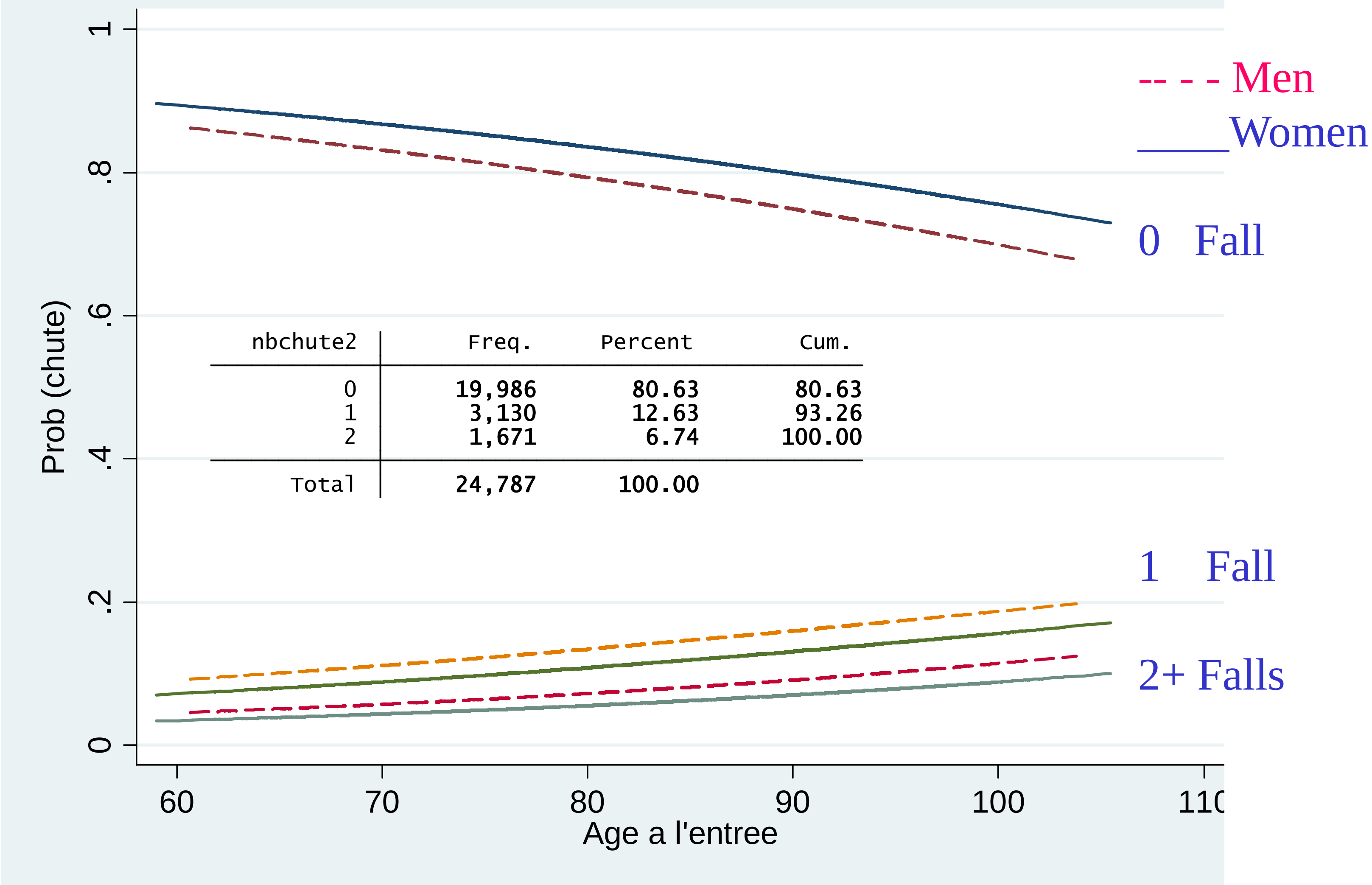
Ordered logistic regression

xi:ologit nbchute2 sex ageentree , or cluster(nopatient)

```
Ordered logistic regression                       Number of obs   =      24787
                                                  Wald chi2( 2)    =     141.19
                                                  Prob > chi2     =     0.0000
Log pseudolikelihood =  -15202.904                 Pseudo R2       =     0.0054

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

nbchute2	Odds Ratio	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
sexe	1.330441	.0504039	7.54	0.000	1.23523	1.432991
ageentree	1.025463	.0025662	10.05	0.000	1.020446	1.030505
/cut1	3.642564	.2156241			3.219948	4.065179
/cut2	4.848833	.2164997			4.424501	5.273164

Ordered logistic regression



nbchute2	Freq.	Percent	Cum.
0	19,986	80.63	80.63
1	3,130	12.63	93.26
2	1,671	6.74	100.00
Total	24,787	100.00	

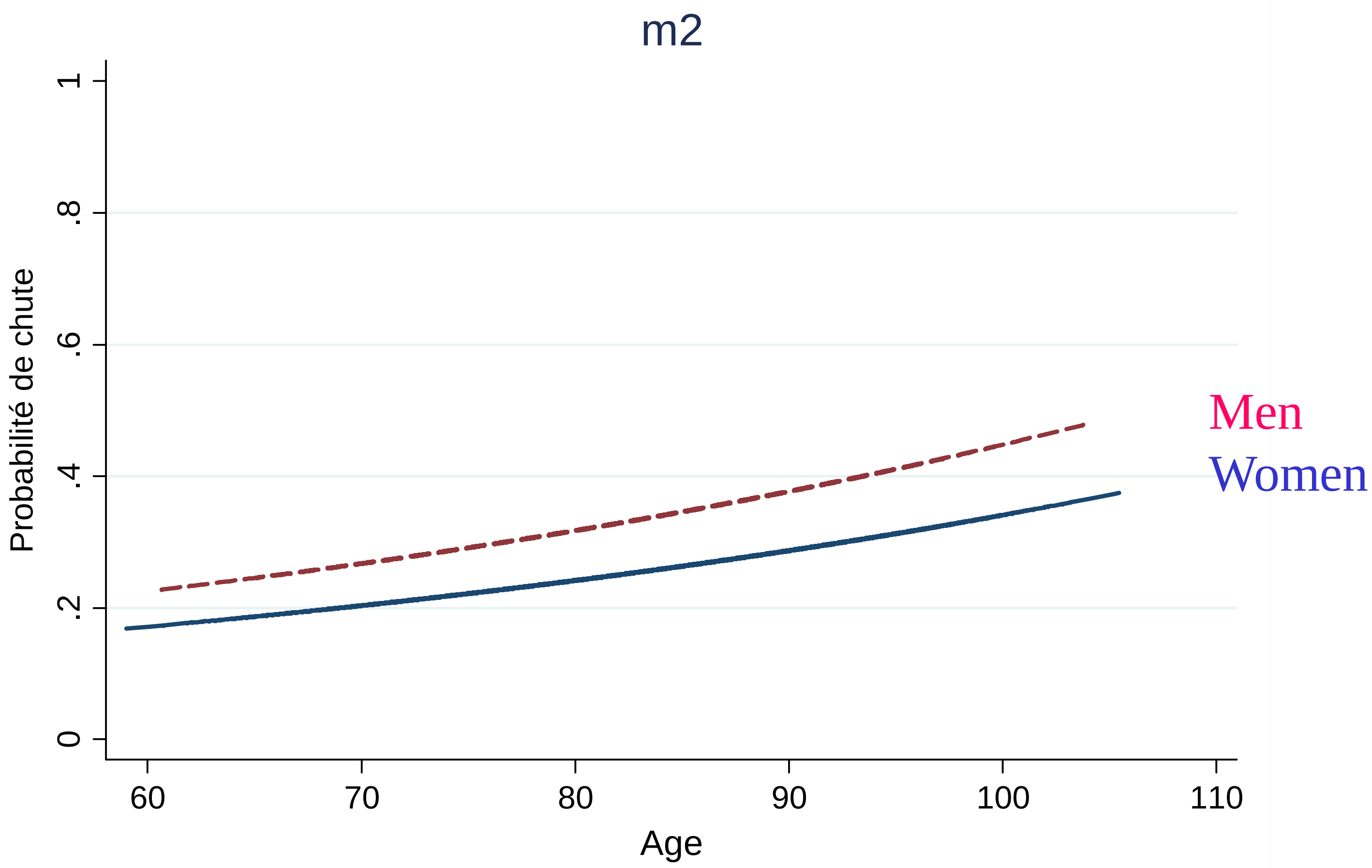
General linear model

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

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

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

                                                   AIC             =   .9768966
Log pseudolikelihood = -12104.16797                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

General linear model



Poisson regression

Model a discrete, positive variable

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

Poisson regression

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

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

# of expected event

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

Poisson regression

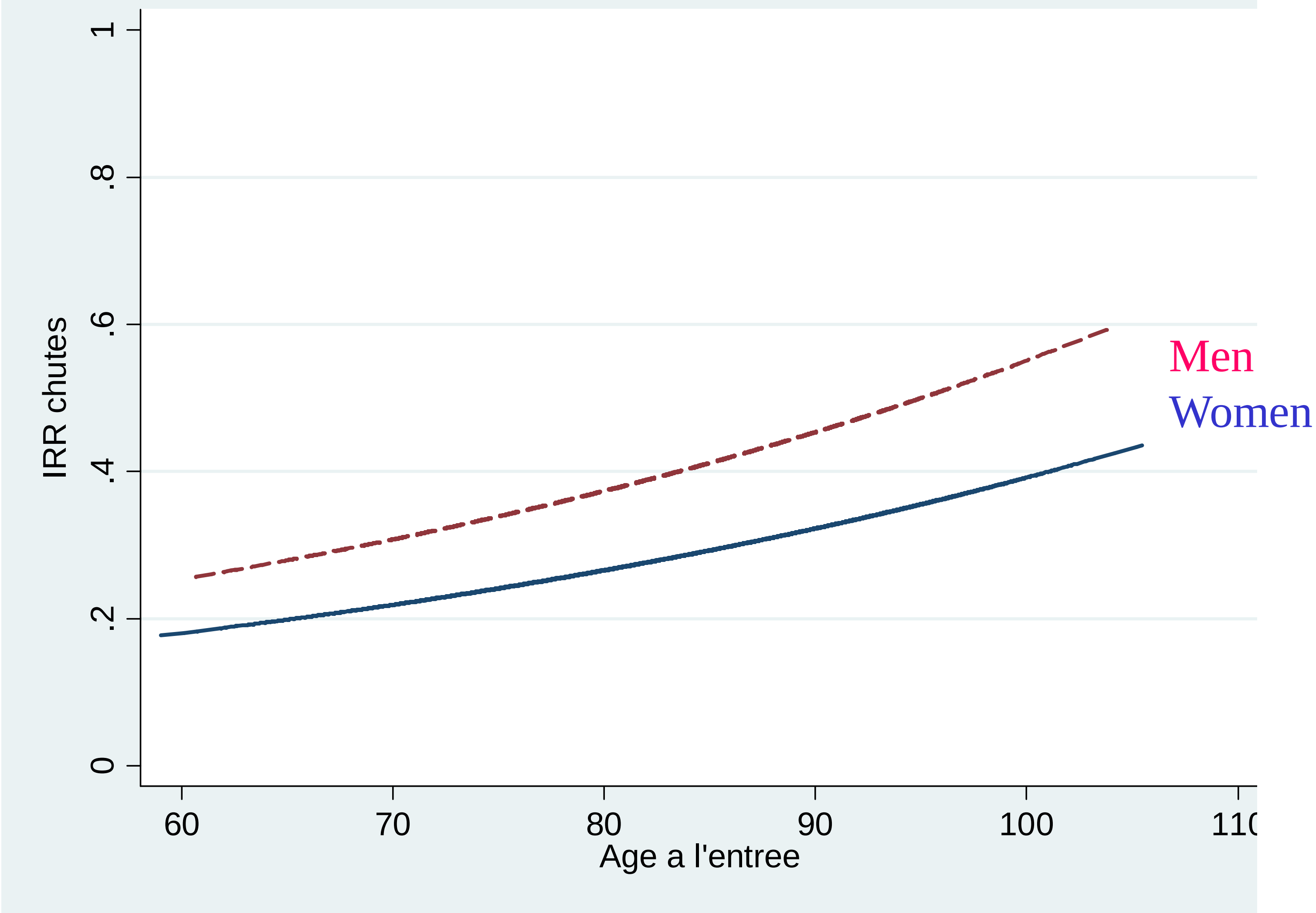
xi:poisson nbchute sexe ageentree , irr cluster(nopatient)

Pseudo R2 = 0.0077

```
Poisson regression                                Number of obs   =      24787
                                                  Wald chi2( 2)    =     115.55
Log pseudolikelihood =  -20374.312                 Prob > chi2     =     0.0000

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

nbchute	IRR	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
sexe	1.40594	.0571526	8.38	0.000	1.29827	1.522541
ageentree	1.019584	.0026049	7.59	0.000	1.014492	1.024703

Poisson regression



Observed and predicted probabilies

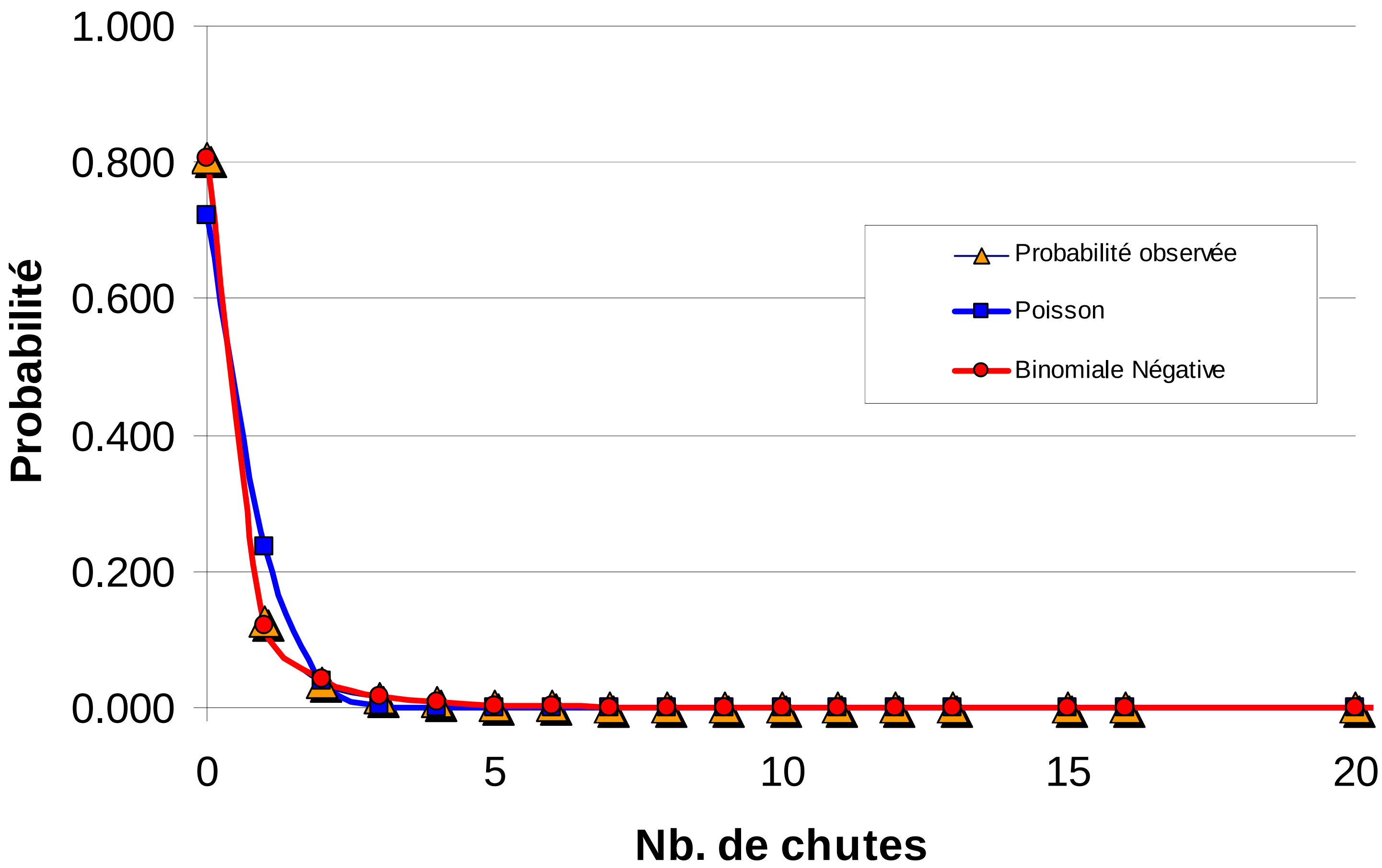
Falls #	Freq	Observed Prob.	Poisson	Negative binomial
0	19986	0.806	0.722	0.807
1	3130	0.126	0.235	0.119
2	930	0.038	0.038	0.041
3	360	0.015	0.004	0.017
4	193	0.008	0.000	0.008
5	97	0.004	0.000	0.004
6	35	0.001	0.000	0.002
7	20	0.001	0.000	0.001
8	11	0.000	0.000	0.000
9	4	0.000	0.000	0.000
10	5	0.000	0.000	0.000
11	5	0.000	0.000	0.000
12	3	0.000	0.000	0.000
13	2	0.000	0.000	0.000
15	3	0.000	0.000	0.000
16	1	0.000	0.000	0.000
20	1	0.000	0.000	0.000
21	1	0.000	0.000	0.000
	24787			

Mean = .3255739
Variance = .8024941

Poisson probability
lambda = .3255739

Negative binomial
With mean = .3255739 &
over dispersion = 3.690428

Observed and predicted probabilies



Binomial negative regression

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

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

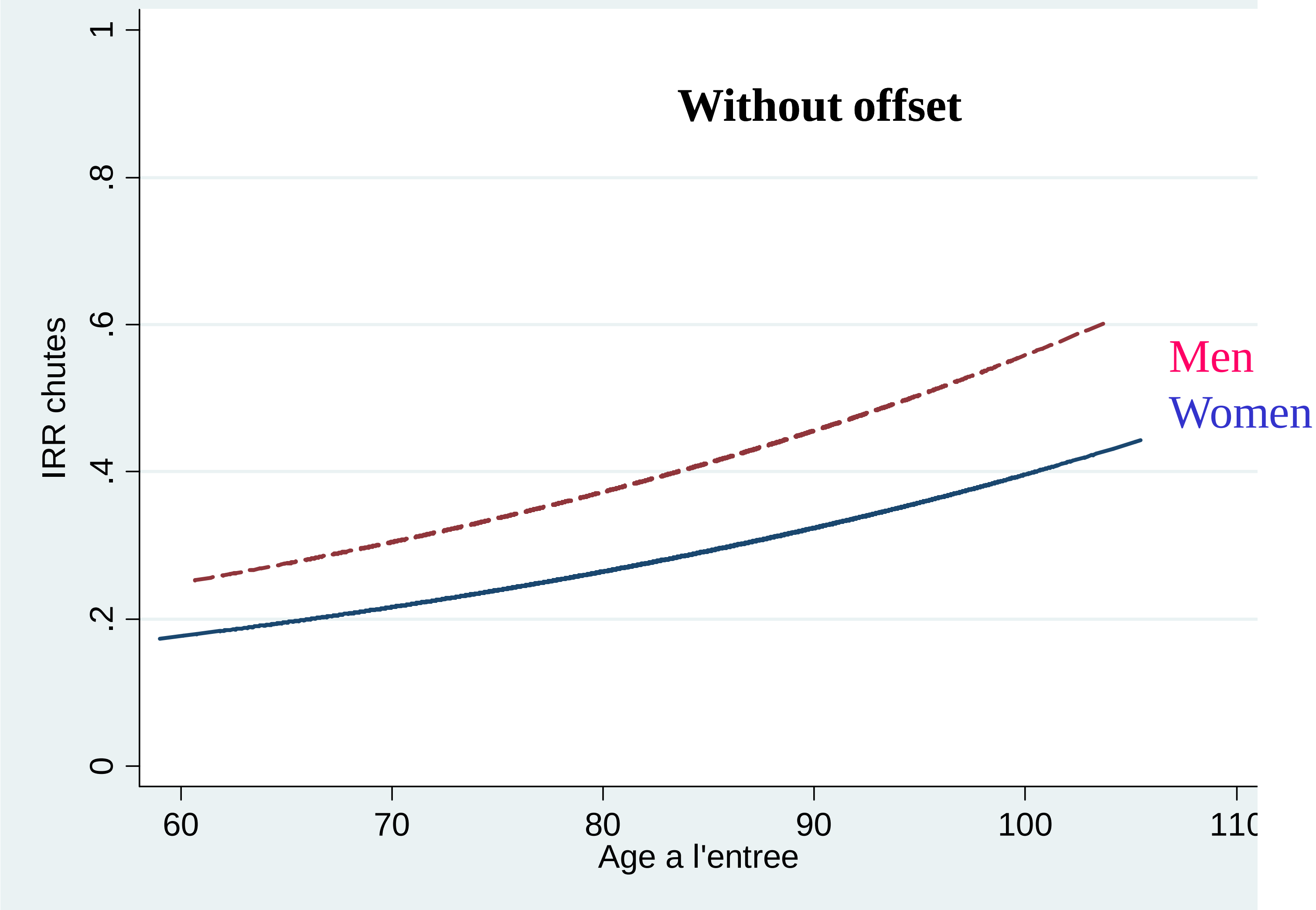
Binomial negative regression

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	1.408448	.0579207	8.33	0.000	1.29938	1.526671
ageentree	1.020393	.0027368	7.53	0.000	1.015043	1.025771
/lnalpha	1.273565	.0352127			1.204549	1.34258
alpha	3.573569	.125835			3.335255	3.828911

Binomial negative regression



Binomial negative regression

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

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

Adjusted for the time of exposure (los)

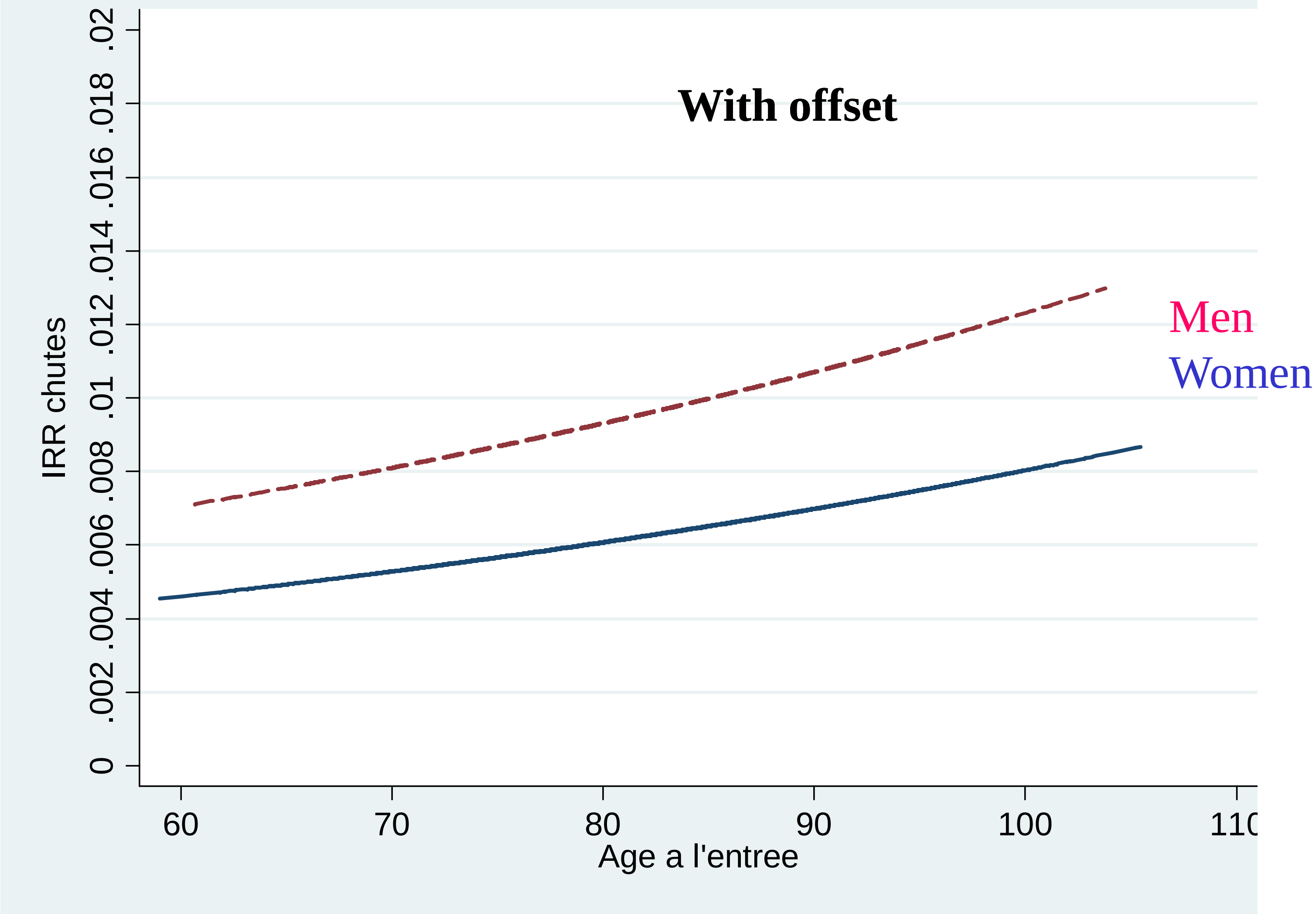
Binomial negative regression

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

```
Negative binomial regression                      Number of obs   =      24787
Dispersion           = mean                        Wald chi2(2)    =     139.61
Log pseudolikelihood = -15868.716                  Prob > chi2     =     0.0000

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

nbchute	IRR	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
sexe	1.532328	.0585667	11.17	0.000	1.421734	1.651526
ageentree	1.014042	.0026012	5.44	0.000	1.008957	1.019153
logdursj	(offset)					
/lnalpha	.5209404	.0452081			.4323341	.6095467
alpha	1.68361	.0761129			1.54085	1.839597

Binomial negative regression



Cox regression

Hazard function

$$h(t) = \lim_{dt->0} \frac{\text{prob}((t \le T < t + dt)/(T \ge t)}{dt}$$

$$h(t) = \frac{f(t)}{S(t)} = -\frac{S'(t)}{S(t)} = -\frac{d}{dt}\ln[S(t)]$$

$$S(t) = \exp\left[-\int_0^t h(u)\,du\right] = \frac{1}{e^{\int_0^t h(u)\,du}}$$

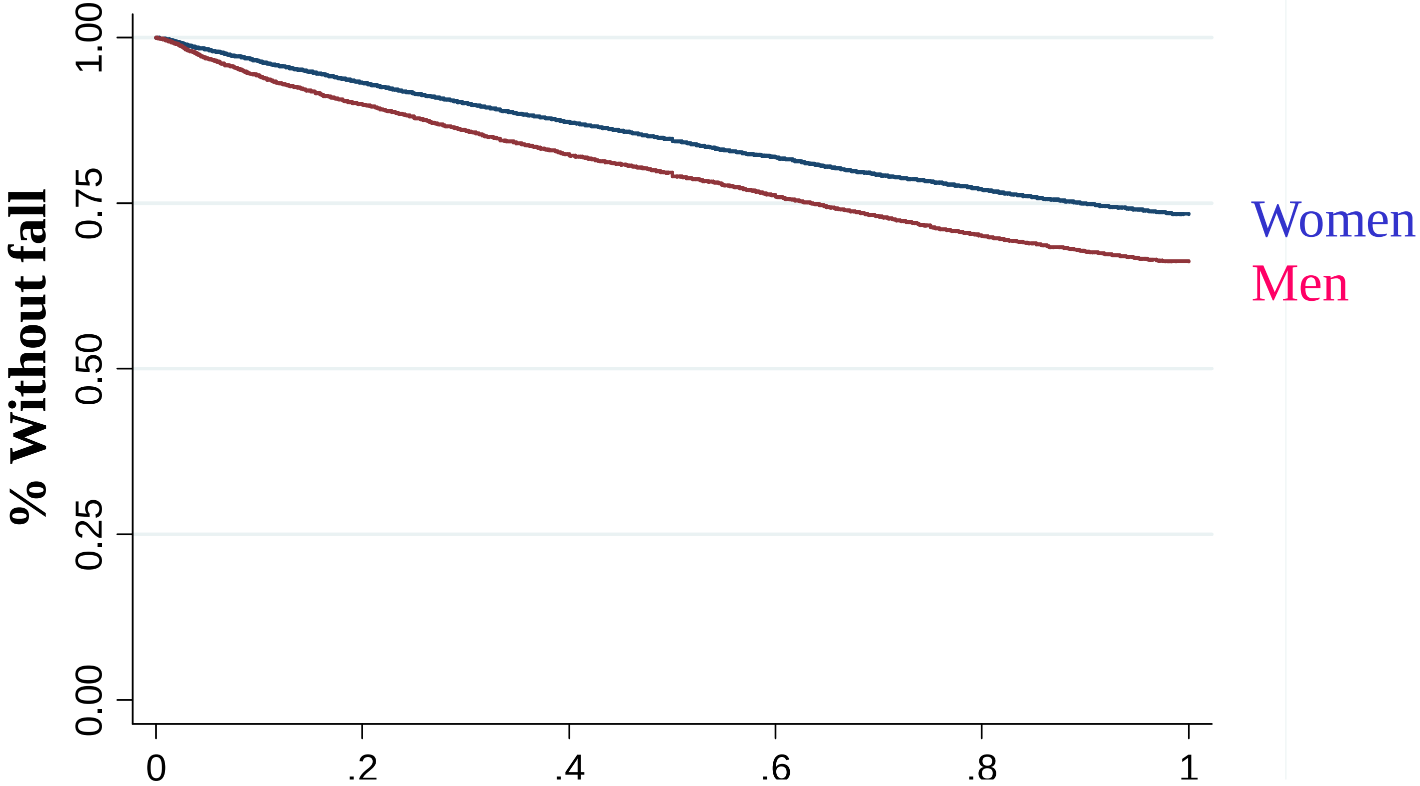
Cox regression

stset timep, id(seqadmin) failure(chuteb==1) origin(time 0) exit(time 1)
stcox sexe ageentree, vce(cluster nopatient)

```
Cox regression -- Breslow method for ties

No. of subjects      =        13920                     Number of obs   =      20119
No. of failures      =         3936
Time at risk         =  518931.4166
                                                        Wald chi2(2)    =     145.77
Log pseudolikelihood =   -34812.176                     Prob > chi2     =     0.0000

                            (Std. Err. adjusted for 13920 clusters in nopatient)
```

_t	Haz. Ratio	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
sexe	1.358143	.0466334	8.92	0.000	1.269751	1.452688
ageentree	1.021364	.002374	9.09	0.000	1.016722	1.026028

Cox regression



Nb at risk **Time after hospital admission (% of LOS)**

	0	.2	.4	.6	.8	1
Women	17109	16539	15953	15337	14666	14019
Men	6989	6668	6388	6106	5776	5499

Cox regression
(modified according to Andersen–Gill)

stset tbf3, fail(nbchuteb==1) exit(time .) id(nopatient) 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
                                                   Wald chi2(2)     =    127.47
Log pseudolikelihood =   -67316.882                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 Value	Sex 95 % CI		Age Value	Age 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.


www.liquidsculpture.com
(c) 2006 Martin Waugh

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

Cox model predicts the speed at which falls occur

Poisson and binomial models address the number of falls

Logistic model predict who will fall or not

Medline Bibliometrics (2.5.2010)

N	%	Key words
515	10.9	Logistic
48	1.0	General linear model
0	0.0	Ordered logistic
41	0.9	Poisson
8	0.2	Binomial negative
88	1.9	Cox
4734	**100.0**	**Falls risk factors**

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.

Conclusions

The predicted value obtained after risk modeling of repeating events can the be used instead of prevalence data in the Sullivan method.