

生存時間解析

survival analysis

ICR中級編（2007.02.01）

がん対策情報センター医学統計室

吉村 健一

例.

臨床試験

MAGIC Trial

The NEW ENGLAND
JOURNAL of MEDICINE

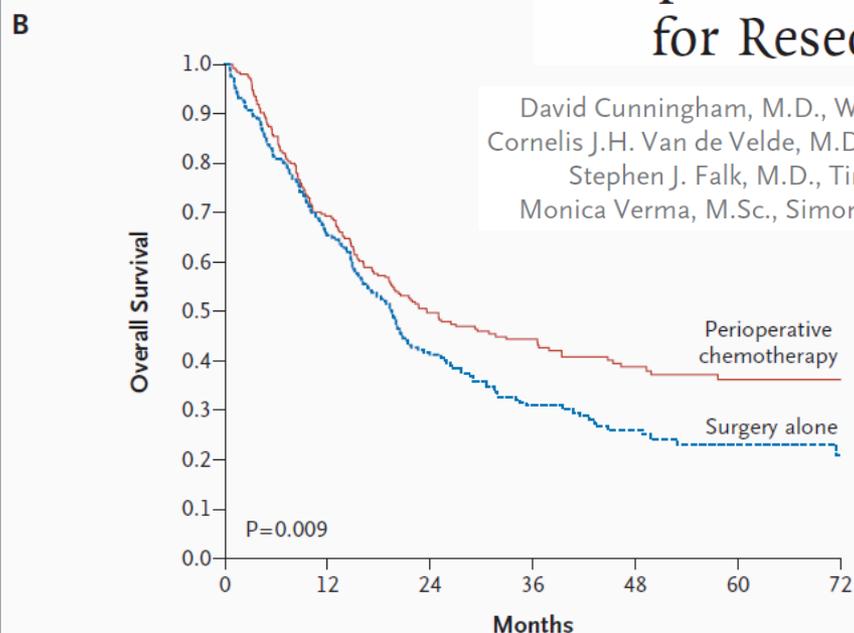
ESTABLISHED IN 1812

JULY 6, 2006

VOL. 355 NO. 1

Perioperative Chemotherapy versus Surgery Alone for Resectable Gastroesophageal Cancer

David Cunningham, M.D., William H. Allum, M.D., Sally P. Stenning, M.Sc., Jeremy N. Thompson, M.Chir., Cornelis J.H. Van de Velde, M.D., Ph.D., Marianne Nicolson, M.D., J. Howard Scarffe, M.D., Fiona J. Lofts, Ph.D., Stephen J. Falk, M.D., Timothy J. Iveson, M.D., David B. Smith, M.D., Ruth E. Langley, M.D., Ph.D., Monica Verma, M.Sc., Simon Weeden, M.Sc., and Yu Jo Chua, M.B., B.S., for the MAGIC Trial Participants*



No. at Risk	0	12	24	36	48	60	72
Perioperative chemotherapy	250	168	111	79	52	38	27
Surgery	253	155	80	50	31	18	9

Figure 1. Kaplan-Meier Estimates of Progression-free Survival (Panel A) and Overall Survival (Panel B).

ABSTRACT

Results

... As compared with the surgery group, the perioperative-chemotherapy group had a higher likelihood of overall survival (hazard ratio for death, 0.75; 95 percent confidence interval, 0.60 to 0.93; P = 0.009; five-year survival rate, 36 percent vs. 23percent) and ...

International Journal of Epidemiology 2005;**34**:467–474

Tooth loss is associated with increased risk of total death and death from upper gastrointestinal cancer, heart disease, and stroke in a Chinese population-based cohort

Christian C Abnet,^{1*} You-Lin Qiao,² Sanford M Dawsey,¹ Zhi-Wei Dong,² Philip R Taylor¹ and Steven D Mark³

Results Individuals with greater than the age-specific median number of teeth lost had **statistically significant 13% increased risk of total death [95% confidence interval (CI) 9–18%], 35% increased risk of upper GI cancer death (95% CI 14–59%),**

... These elevated risks were present in male smokers, male non-smokers, and females, nearly all never-smokers.

Table 3 The relative risk of tooth loss^a on total and cause-specific mortality in the General Population trial cohort, Linxian, PRC (1986–2001)

Mortality	RR ^b	95% CI	P ^c	Relative risk by sex and smoking Status ^d				Smoking interaction	
				Smoking	Cases	RR	95% CI	P	P
Death	1.13	1.09–1.18	<0.0001	Females	4346	1.07	1.01–1.14	0.024	
				Male never-smokers	1500	1.09	0.98–1.21	0.14	<0.0001
				Male ever-smokers	3516	1.24	1.16–1.32	4 × 10 ⁻¹⁰	
Upper GI cancer	1.35 ^e	1.14–1.59	0.0096 ^f	Females	1152	1.24	0.98–1.58	0.19	
				Male never smokers	394	1.59	1.03–2.45	0.041	0.0018
				Male ever-smokers	1079	1.39	1.06–1.83	0.0001	
Heart disease	1.28	1.17–1.40	<0.0001	Females	893	1.10	0.96–1.25	0.17	
				Male never-smokers	287	1.57	1.24–1.98	0.0002	0.037
				Male ever smokers	752	1.41	1.23–1.63	2 × 10 ⁻⁶	
Stroke	1.11	1.01–1.23	0.027		2866				0.54
Other cancer	1.07	0.89–1.28	0.44		514				Not tested

例.
コホート研究

臨床研究における生存時間解析

- 治療効果を調べる臨床試験ではしばしば、標準治療よりも新治療の方が生存時間を延長するかについて評価します
- 病気の原因を調べる疫学研究ではしばしば、ある曝露要因を持つものが持たないものに比べて疾病を罹患するリスク（ハザード）が何倍となるかについて評価します
 - 本日の内容は、生存時間解析について統計的に重要な考え方を中心に解説
 - cook-book、ソフトウェアの解説は行いません

生存時間と生存割合

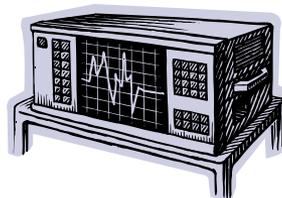
生存時間 survival time

- 何らかの**イベント（事象）**が発生するまで、しつこく（＝連続的に）追跡することで得られるイベント発生までの時間
 - より明示的に time-to-event data とも言う

腫瘍の増悪・再発



機器の故障



製品の劣化／寿命



死亡



time to death

イベント

退院



罹患／発症



感染



回復／社会への復帰

破産



生存時間解析 survival analysis

死亡

years to death

死亡までの年数

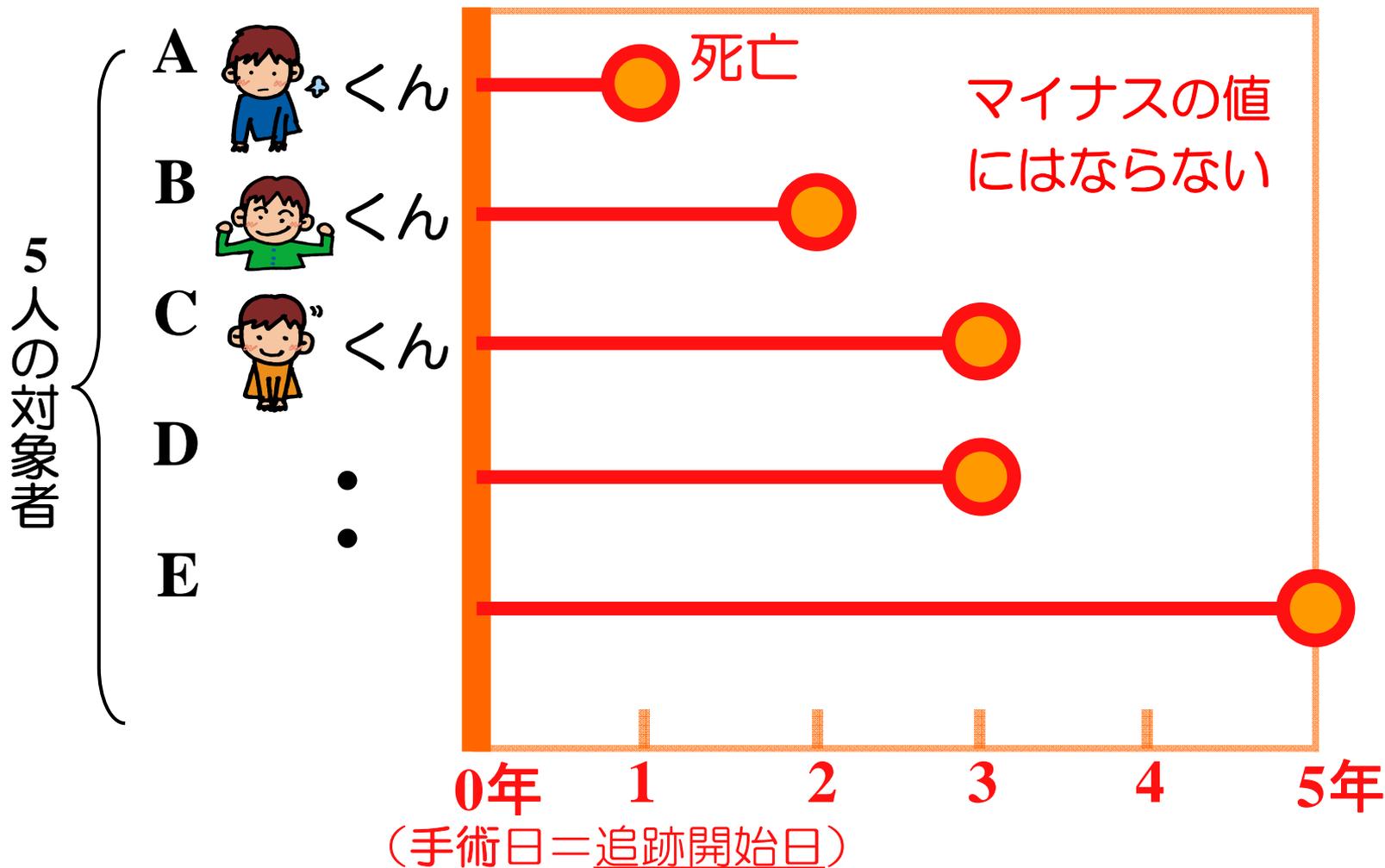
■ 生存時間に興味がある状況での解析

- 研究対象者が**死亡**するまでの年数
 - 白血病患者の寛解期間（＝寛解から**再発**までの月数）
 - 術後に**退院**するまでの日数
 - コホート研究の対象者が**がん罹患**した年齢
- など、時間の長短に興味がある状況

- Positiveな**イベント**（＝時間が短い方が好ましい）でも
- Negativeな**イベント**（＝ 長い方が ）でもよい

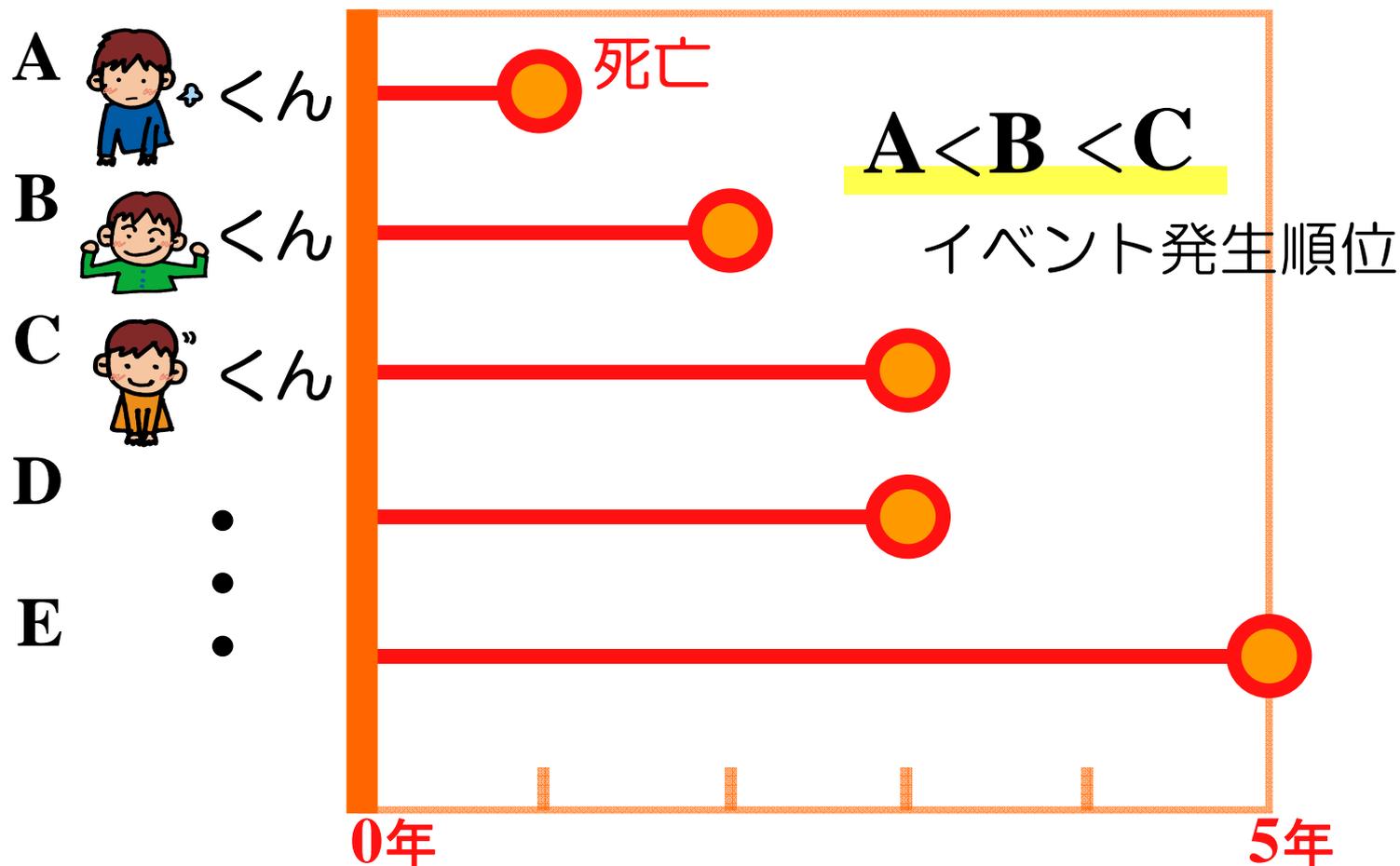
- time-to-death 以外の場合も慣習的に生存時間、生存時間解析
 - ここでも便宜上、死亡と生存時間に統一
 - 単位も年数（月数）に統一

術後の生存時間を測定しよう

追跡
時間 (年)

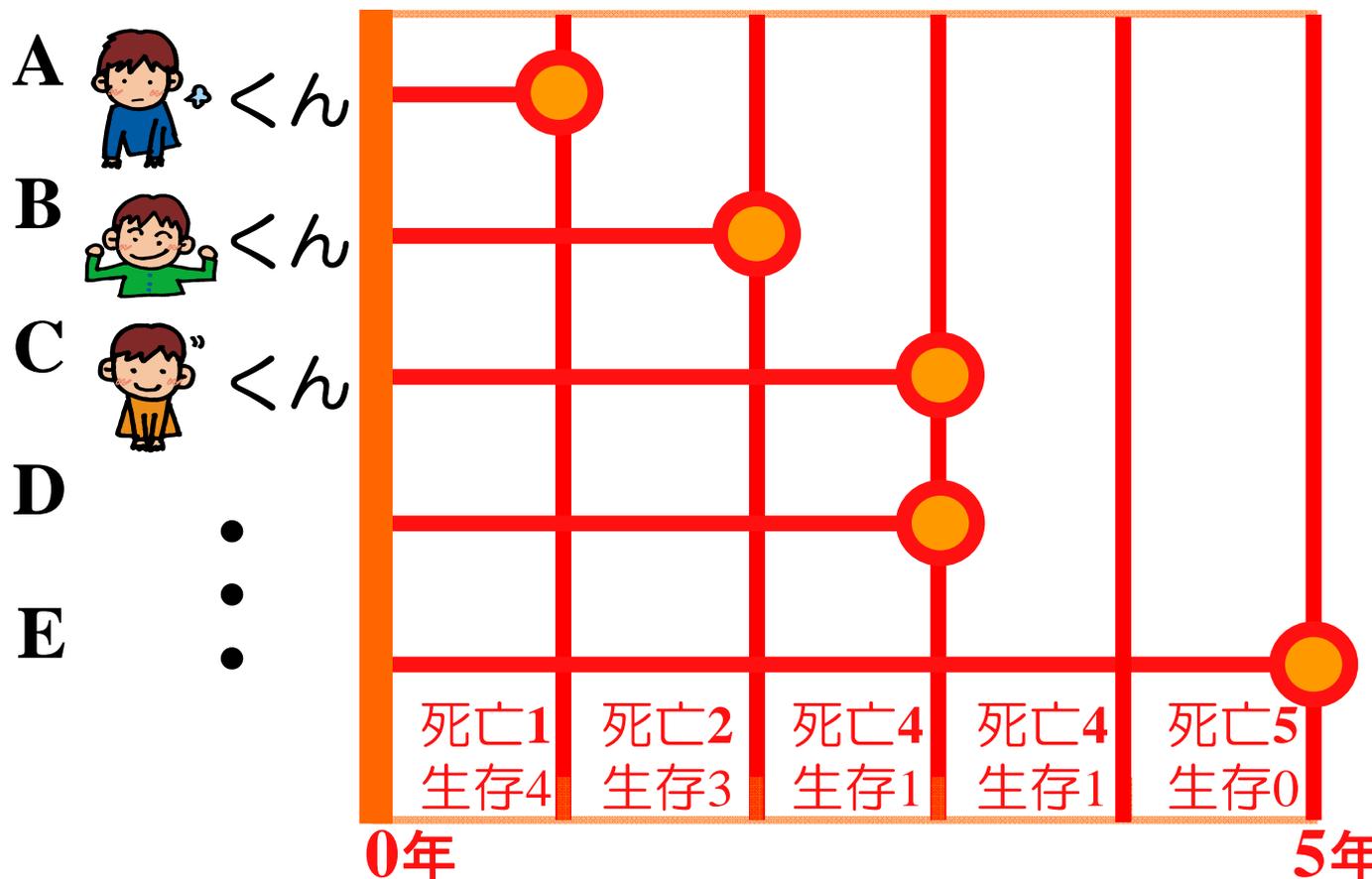
□ Aくんの生存時間は**1年**、Bくんは**2年**、Cくんは**3年**、、、

生存時間から分かること (1)

追跡
時間 (年)

□ AくんよりもBくんの方が長生き、Cくんの方がもっと長生き

生存時間から分かること (2)

追跡
時間 (年)

□ 1年生存割合は $\frac{4}{5}$ 、2年～は $\frac{3}{5}$ 、3年～は $\frac{1}{5}$ 、5年以降は 0

* 数学的に正しくは生存割合○、生存率×

生存曲線の特徴

1

真の生存曲線は滑らか

2

生存割合を求められる

曲線から割合が求まる

3

時間0では100%から始まる

研究開始時には全対象が生存

4

時間が経つにつれて
減少して、0%に近づく

経過時間が長いほど
生存者が少なくなり
最後には全対象が死亡

生存割合

100%

0

k 年
生存割合

 k

時間 (年)

生存時間解析の主な興味（医学領域）

目的

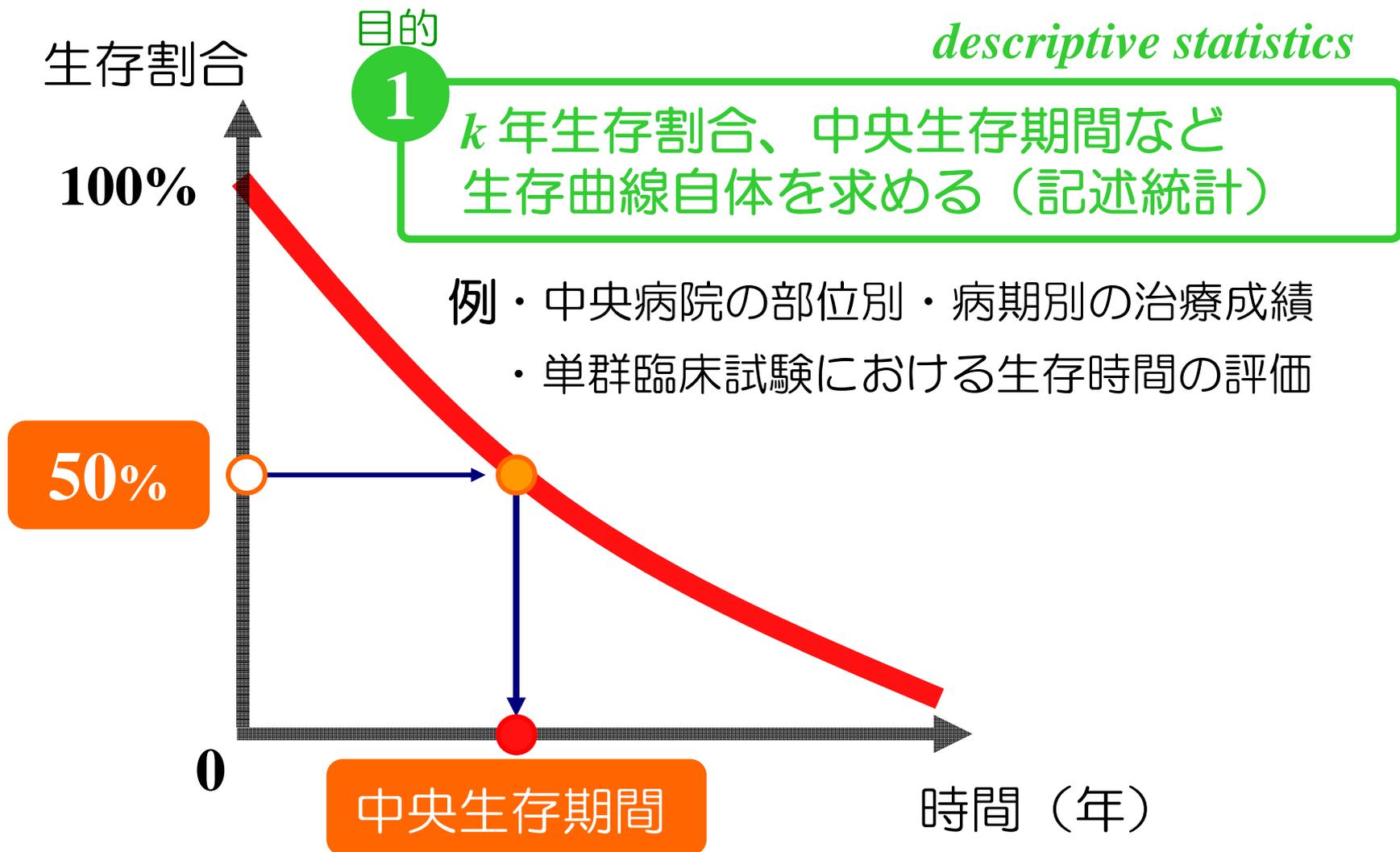
① 記述統計

② グループ間で生存曲線を比較

□ その一部として、

- 予後因子解析：予後の良い（悪い）サブグループの同定
- 予測因子解析：治療効果のあるサブグループの同定

生存時間解析の主な興味 (医学領域)



生存時間解析の主な興味 (医学領域)

目的

2

グループ間で生存曲線を比較する

例・臨床試験における治療群の比較

手術+補助化学療法群

手術単独群

曲線が上にあるほど
良い治療法に対応

補助化学療法をする
方が良さそう

生存割合

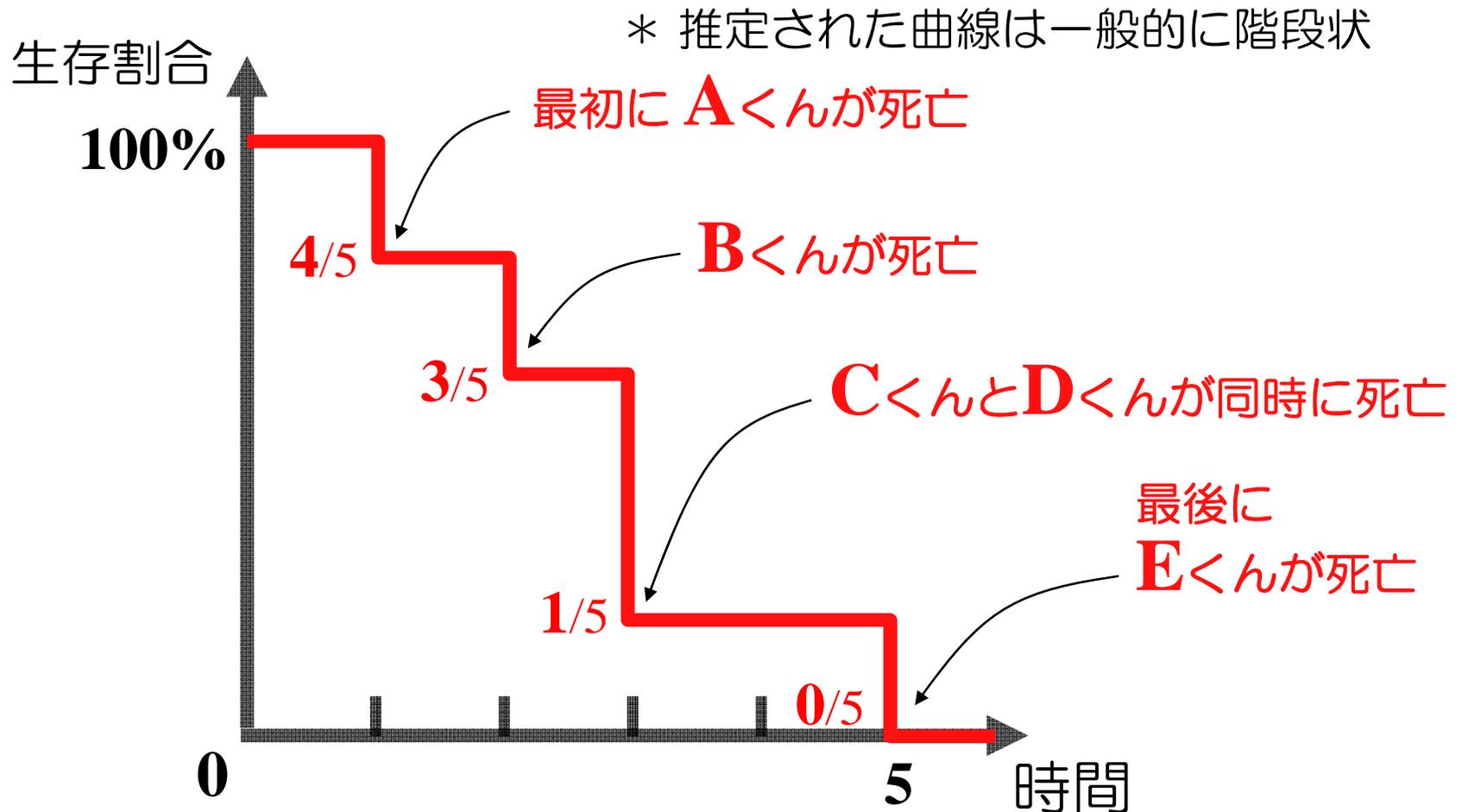
100%

0

時間 (年)

生存曲線の推定

- 経過時間（横軸）と生存割合（縦軸）の関係を表した曲線
 - 生存割合の減少が観察された死亡に対応して求まる

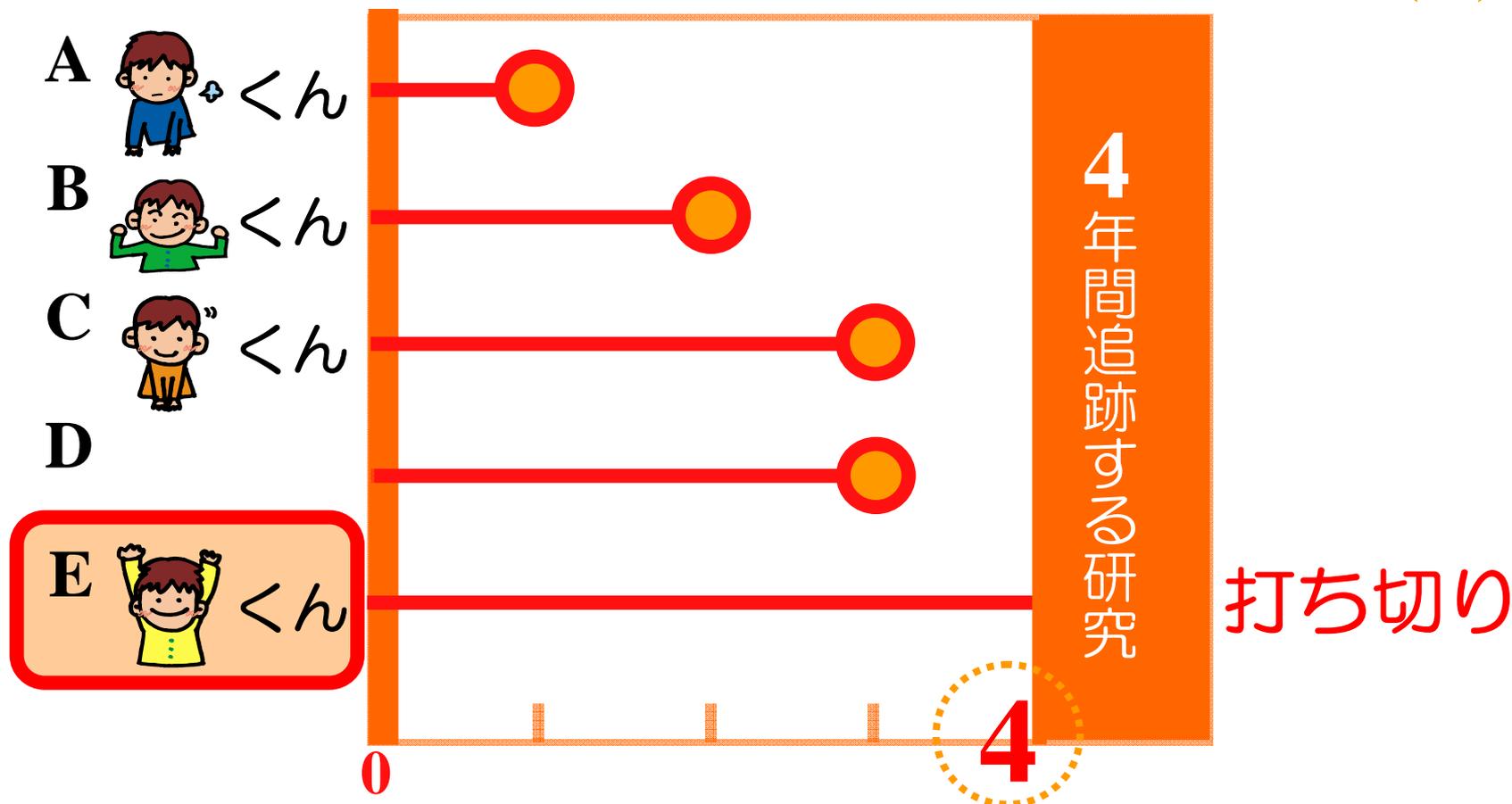




打ち切りと Kaplan-Meier法による生存曲線の推定

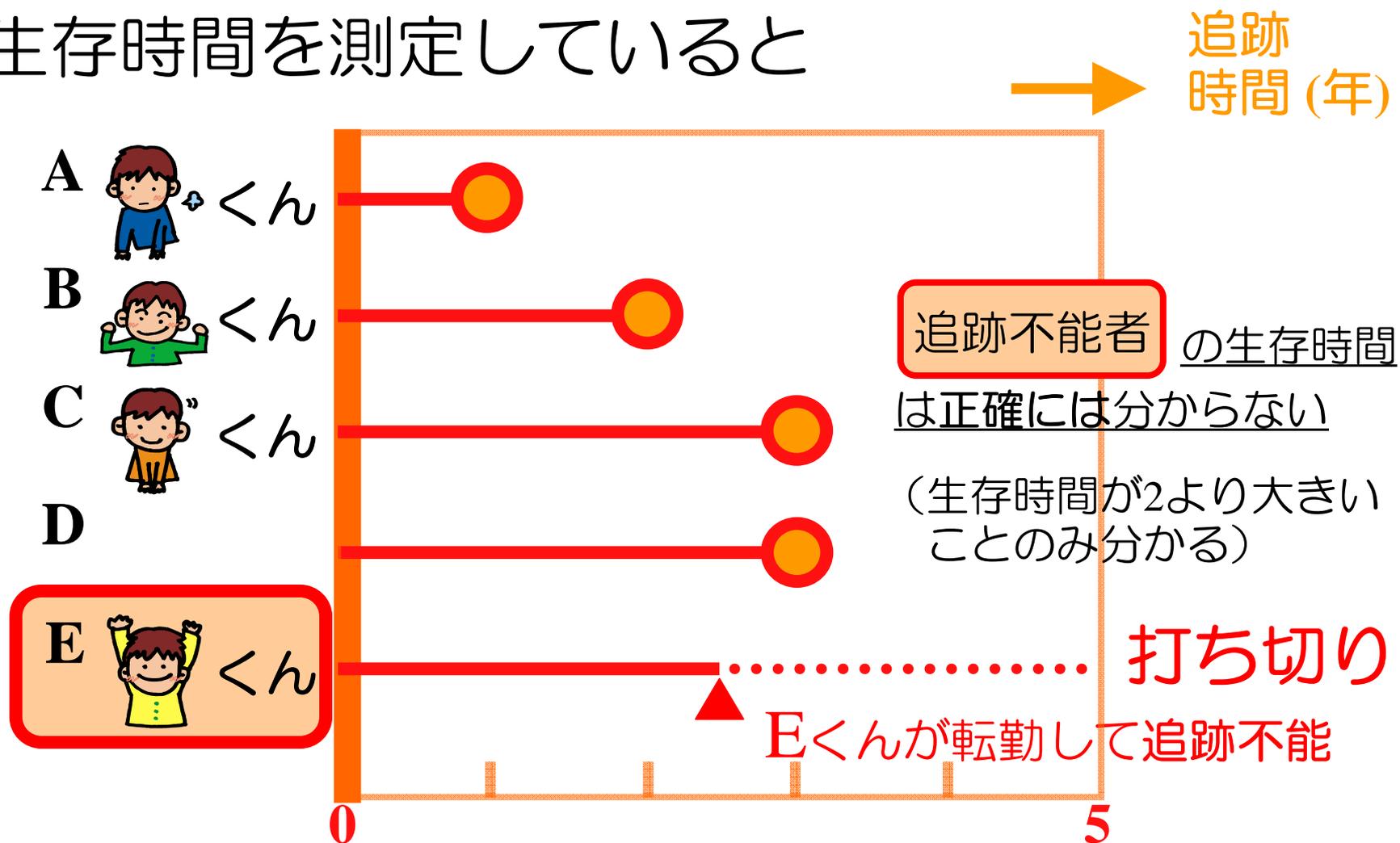
生存時間を測定していると

追跡時間 (年)



- Eくんの生存時間が4より大きいことは確かだが、正確な値は不明
 - 研究／試験期間中に全ての対象が死亡するとは限らない

生存時間を測定しているとき



- 転勤／転院、研究への同意の撤回などにより
追跡不能 loss to follow-up となる人がでてくる

打ち切り censoring

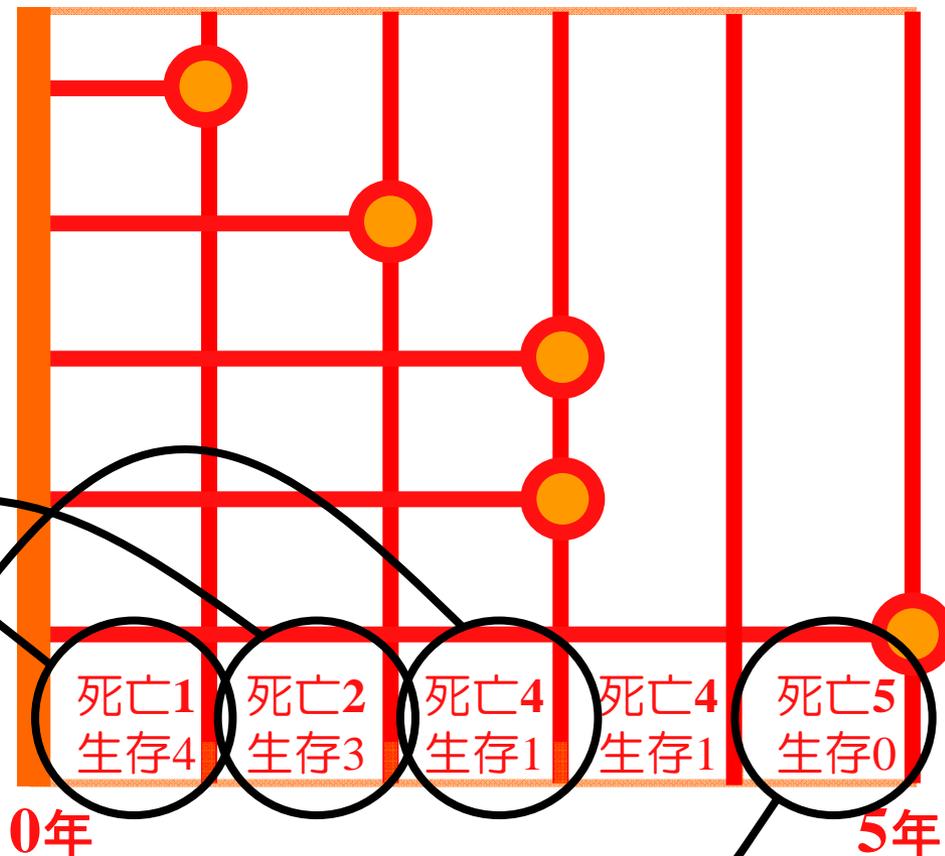
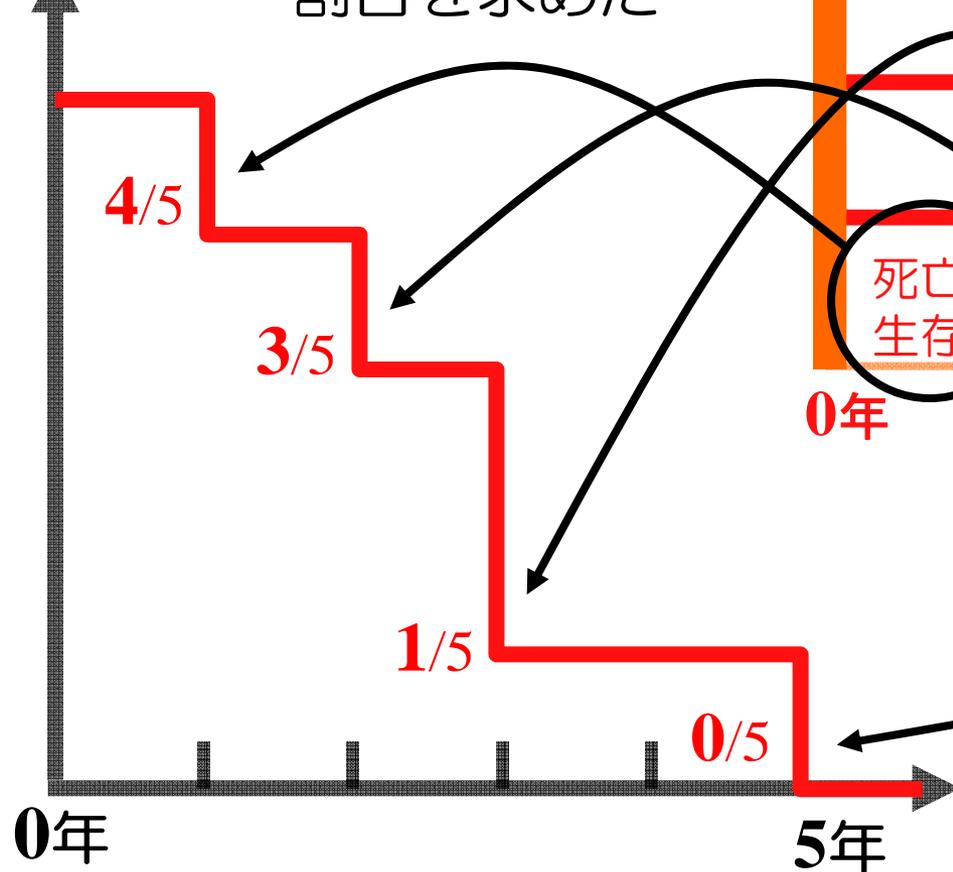
- ある時点まで生存していたことのみは分かるが、それ以降のいつ死亡したかは正確に分からない
 - (右側) 打ち切りという $2 < \text{生存時間} \leq \infty$
 - 左側打ち切り ($0 < \text{生存時間} < 2$) もあるがとりあえず忘れていい
- 生存時間を測定する研究では避けられない
 - 全対象が死亡するまで追跡しなければならないと非常に効率が悪い!
 - 生存時間解析を行う際には考慮すべき
 - 考慮できる特別な解析方法
 - Kaplan-Meier法、□ グランク検定、Cox回帰、、、、

目的

1 記述統計として生存曲線を求める方法

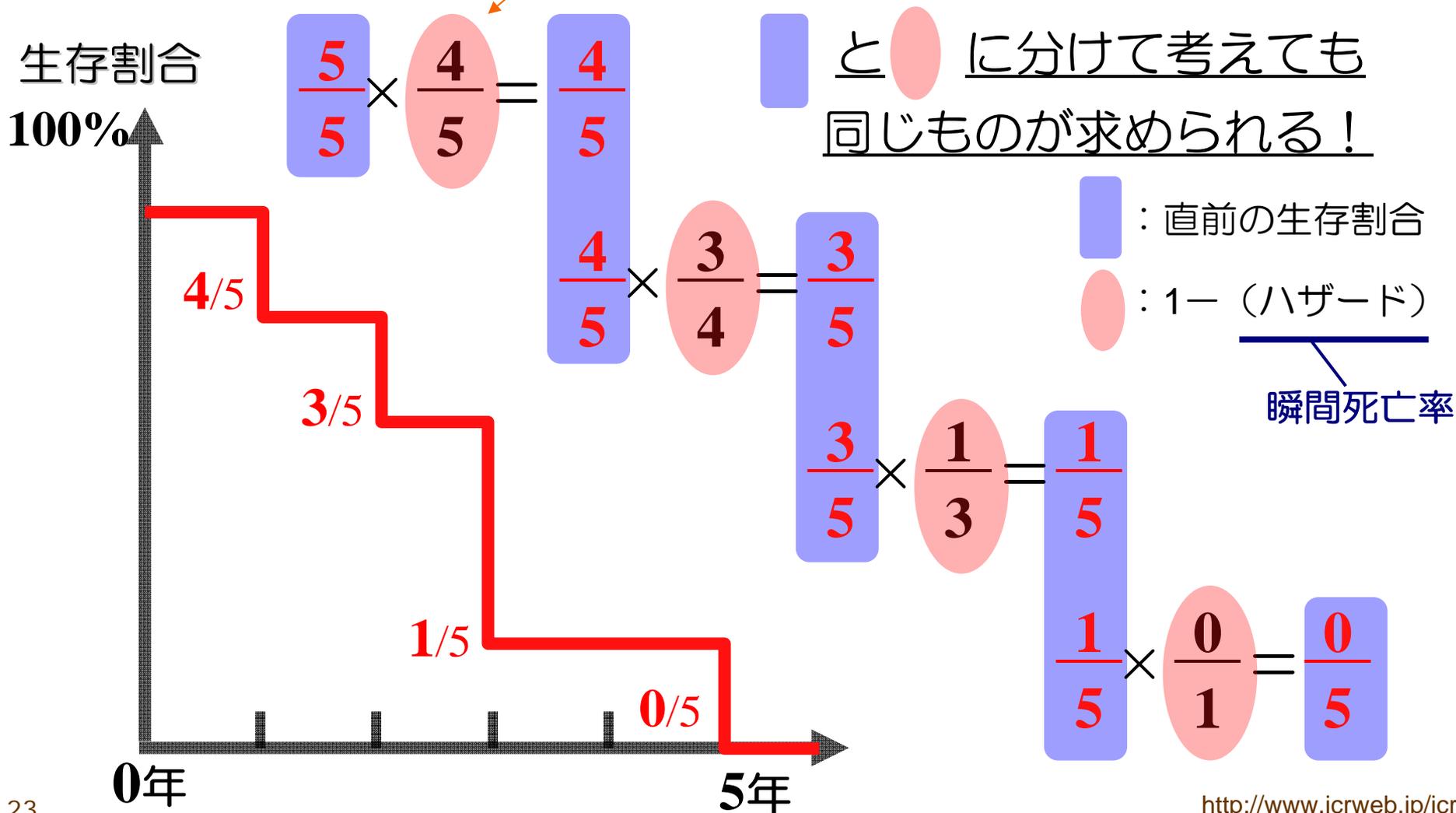
- 世界標準は、Kaplan-Meier法
 - 打ち切りを考慮して生存曲線、Kaplan-Meier曲線を描く（＝生存割合を求める）方法
 - ヒトの直感に極めて近い生存割合が推定される
＝ 数学的な性質も自然と良くなる！
 - A～Eに関して既に求めた生存割合も、Kaplan-Meier法によって求めたものと同じ

生存割合の求め方①

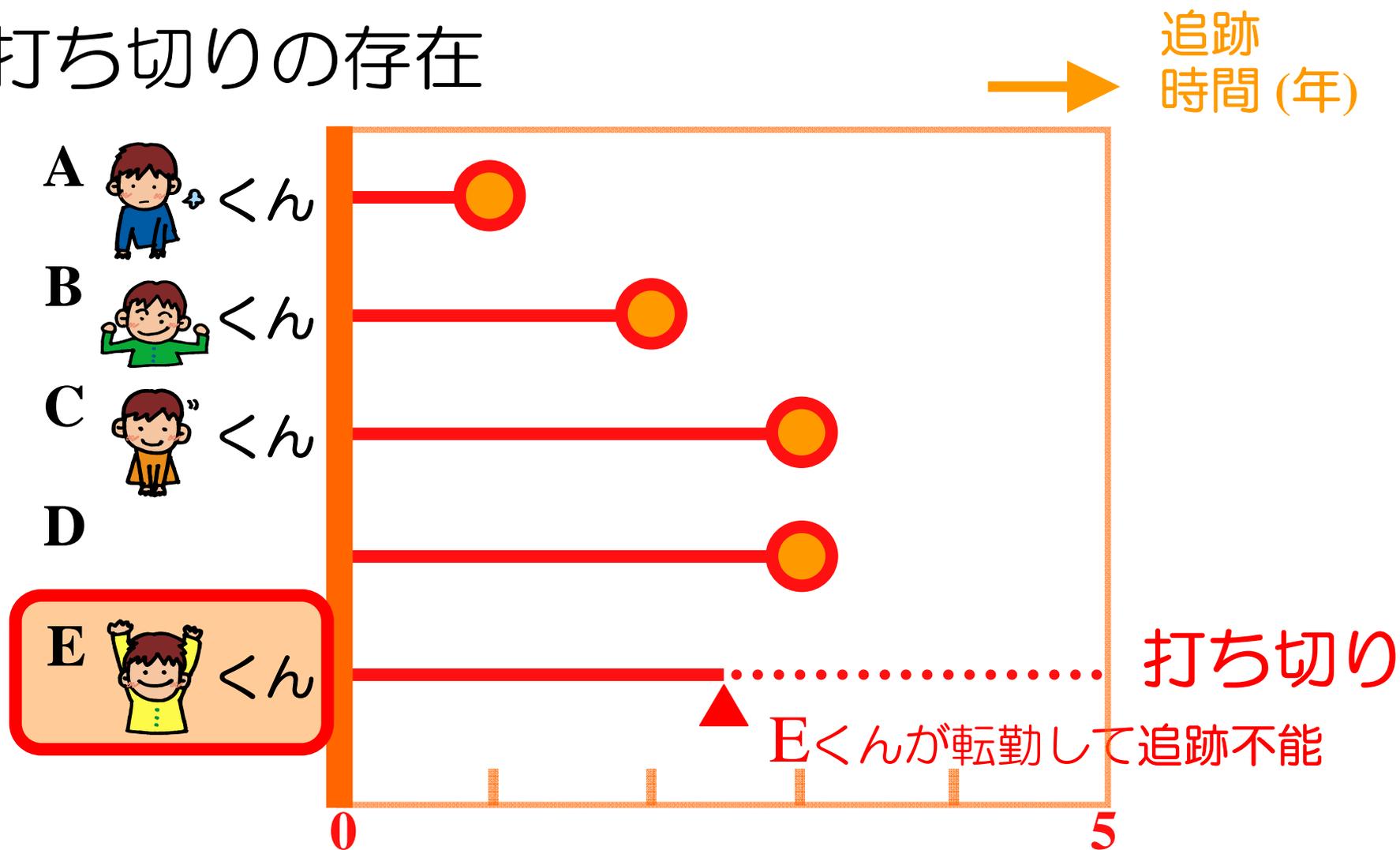
生存割合
100%先ほどは単純に
割合を求めた

生存割合の求め方② Kaplan-Meier法の考え方

今残っている人の中で更に生き残れる人の割合

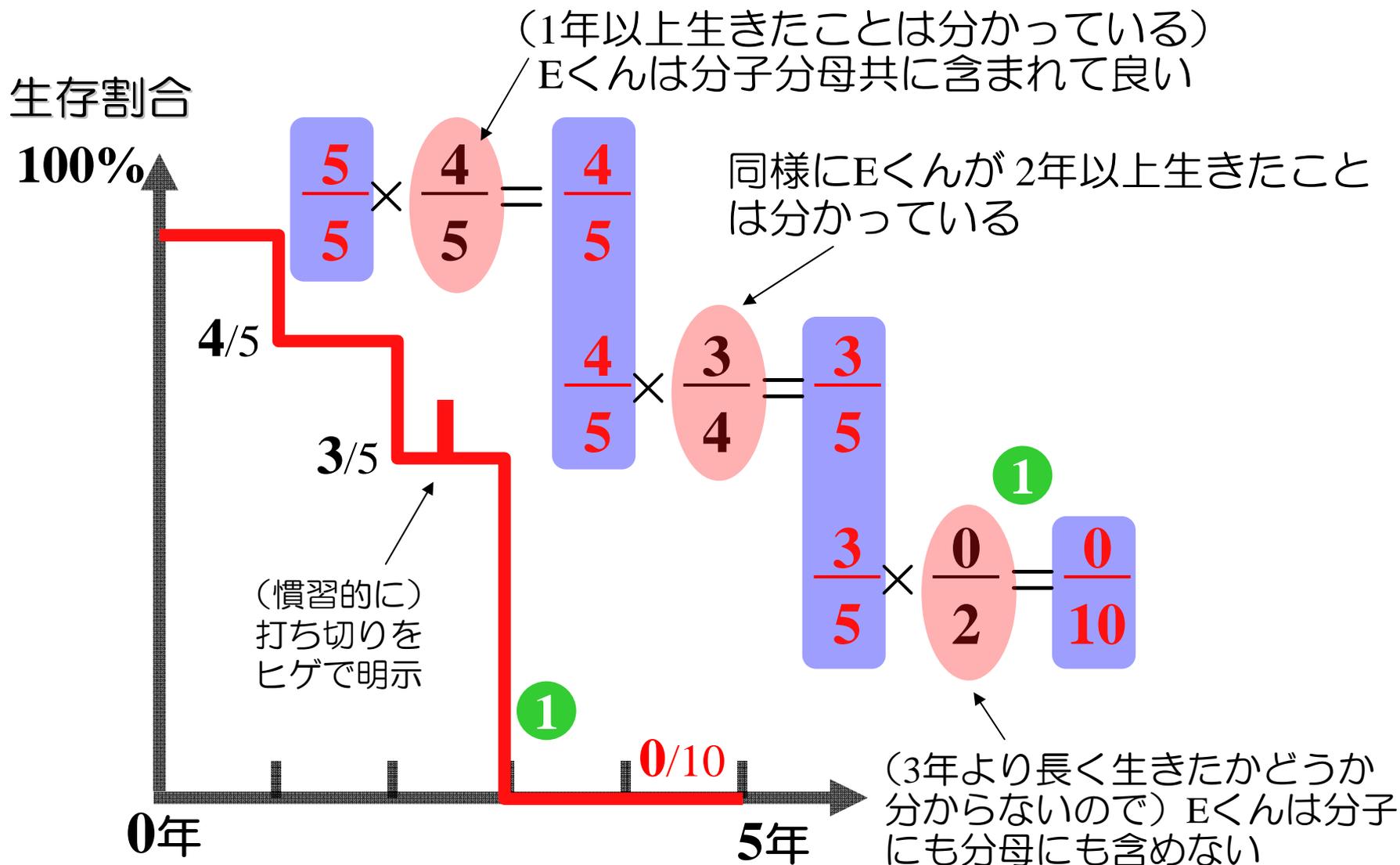


打ち切りの存在



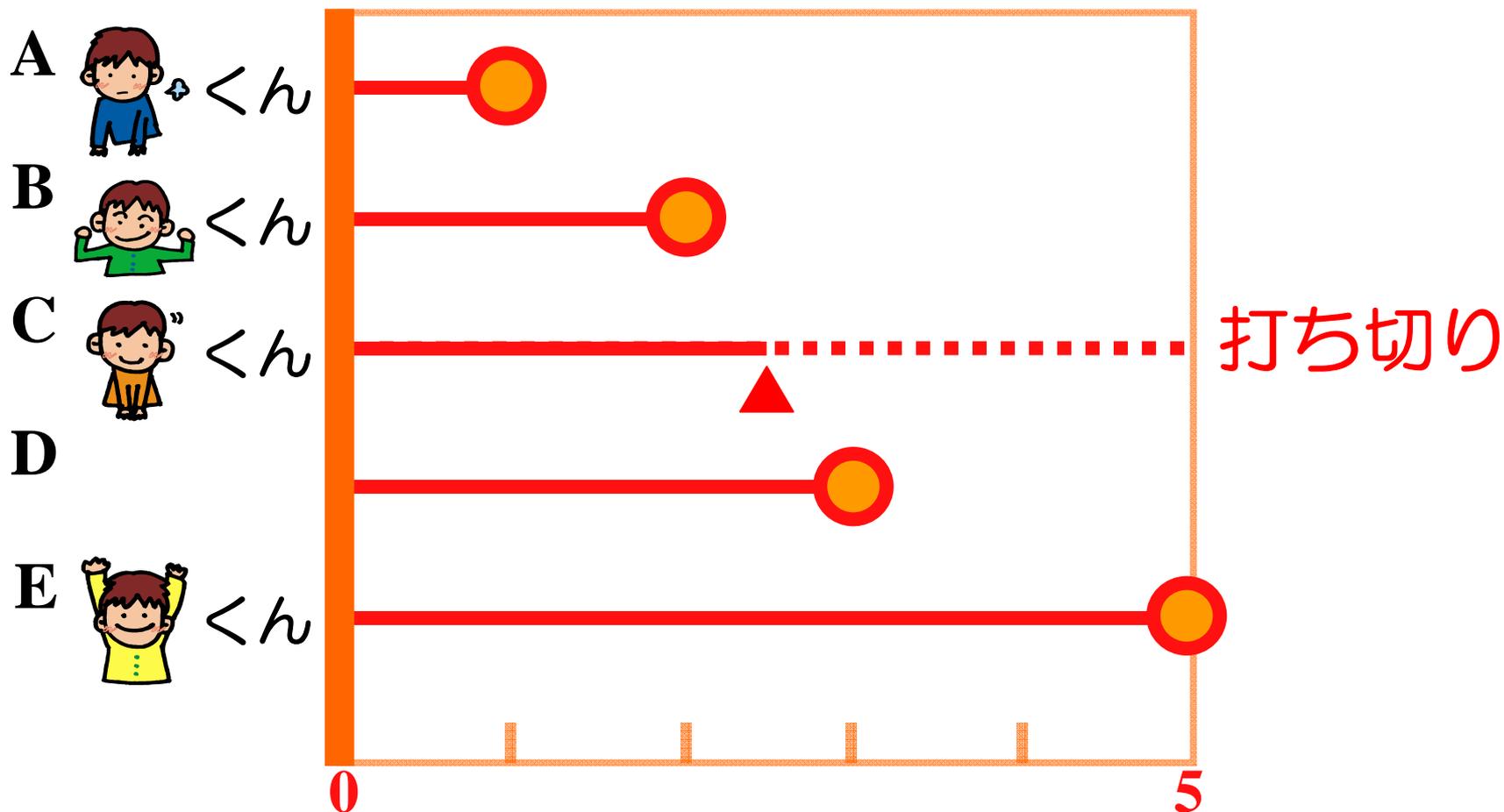
- Eくんは「2.5年の時点まで“生存”」したことは分かるが「それ以降のいずれの時点で“死亡”」したかは分からない

K-M法の打ち切り処理の考え方



打ち切りの存在 (Cくんの場合)

追跡時間 (年)

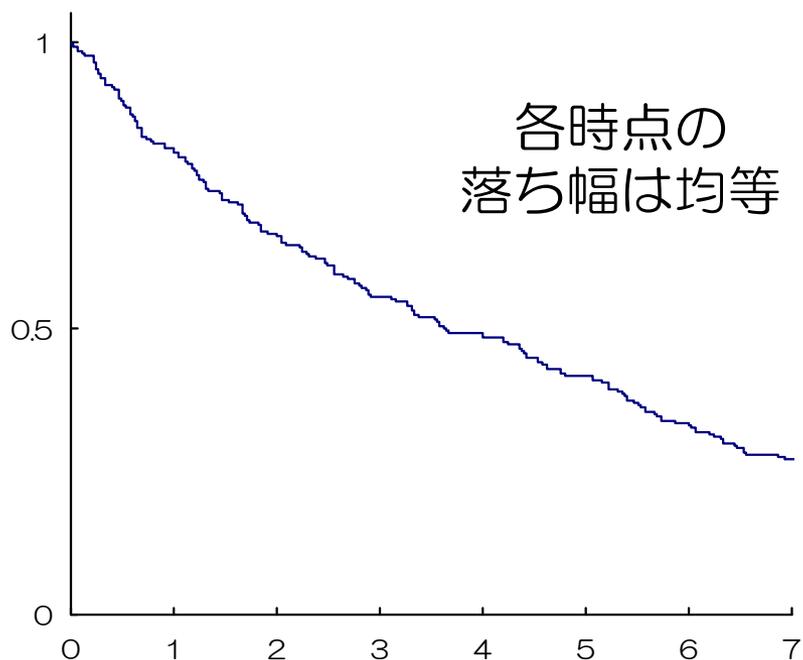


□ Cくんは「2.5年の時点まで“生存”」したことは分かるが

• • •

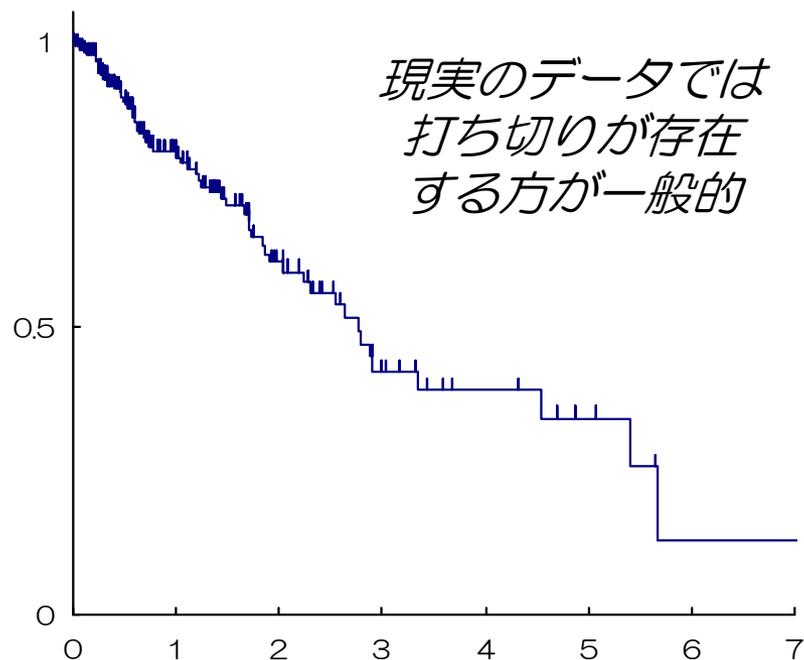
推定K-M曲線

同一の仮想データ



各時点の
落ち幅は均等

打ち切り無し
(理想的な状況)

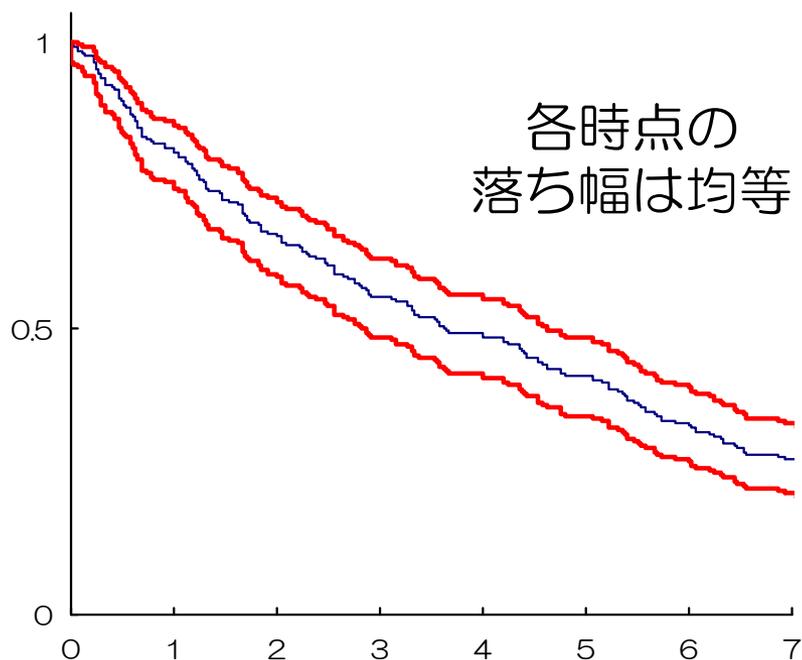


現実のデータでは
打ち切りが存在
する方が一般的

ランダムに
打ち切りを発生

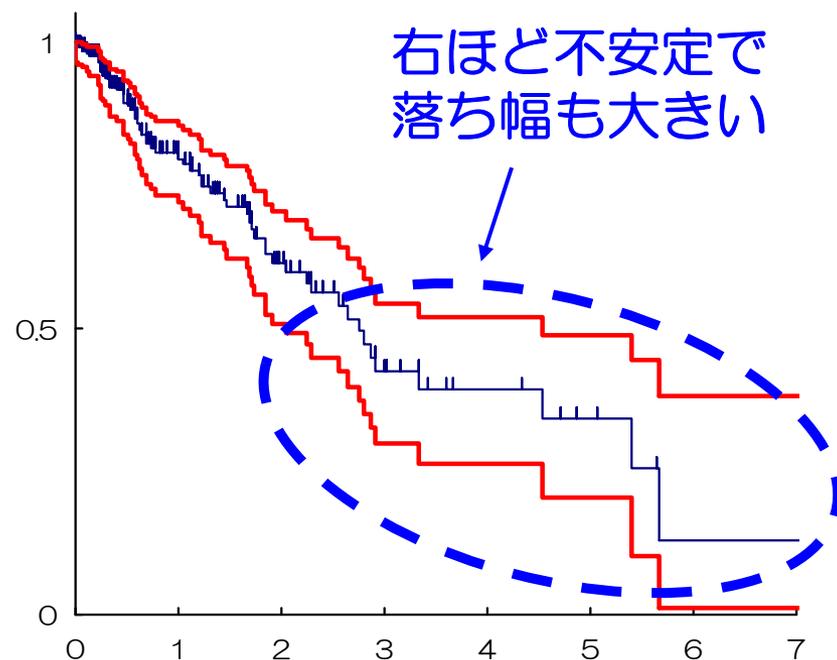
推定K-M曲線の精度を信頼区間で示すと

同一の仮想データ



各時点の
落ち幅は均等

打ち切り無し
(理想的な状況)

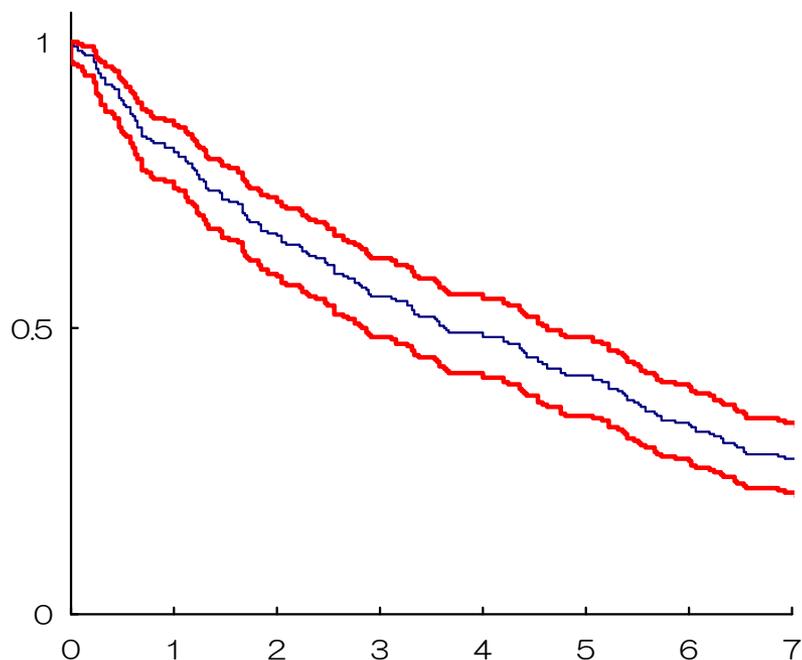


右ほど不安定で
落ち幅も大きい

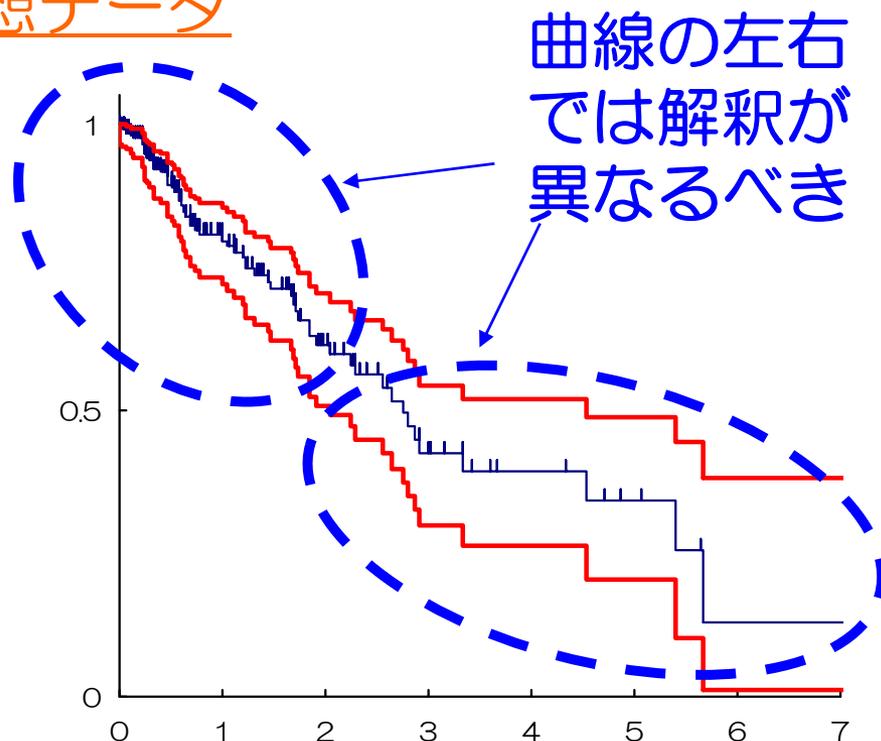
ランダムに
打ち切りを発生

推定K-M曲線の精度を信頼区間で示すと

同一の仮想データ



打ち切り無し
(理想的な状況)



ランダムに
打ち切りを発生

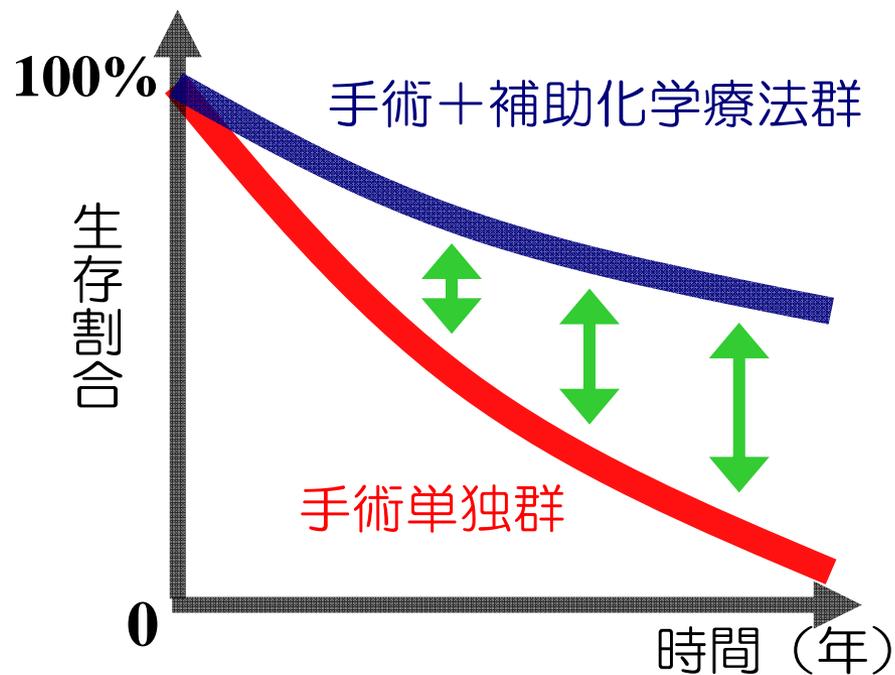
全体的に区間が広がるが、右ほどその影響を受けやすい



グループ間で
生存時間の比較をしたい！

目的

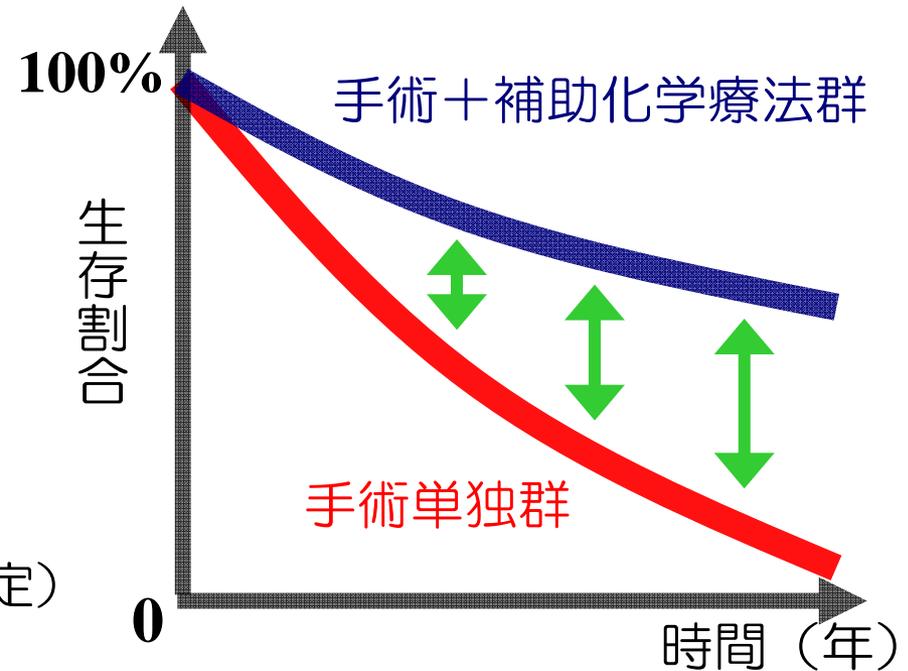
2 グループ間で生存曲線を比較したい!



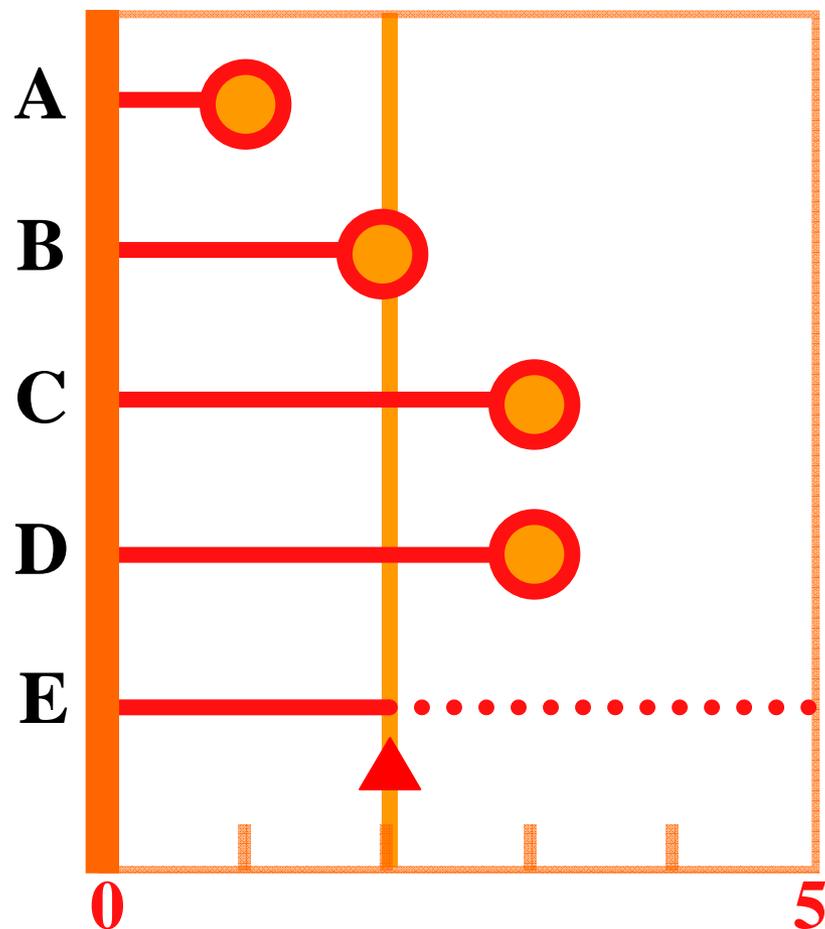
目的

2 グループ間で生存曲線を比較したい!

- 生存時間の**平均値**を求めて
群間差を計算したら
どうだろう?
 - 平均値の比較 (e.g. t検定)
 - **生存割合**を群間で比較
したらどうだろう?
 - 割合の比較 (e.g. Pearson χ^2 検定)
- 順に考えてみよう!

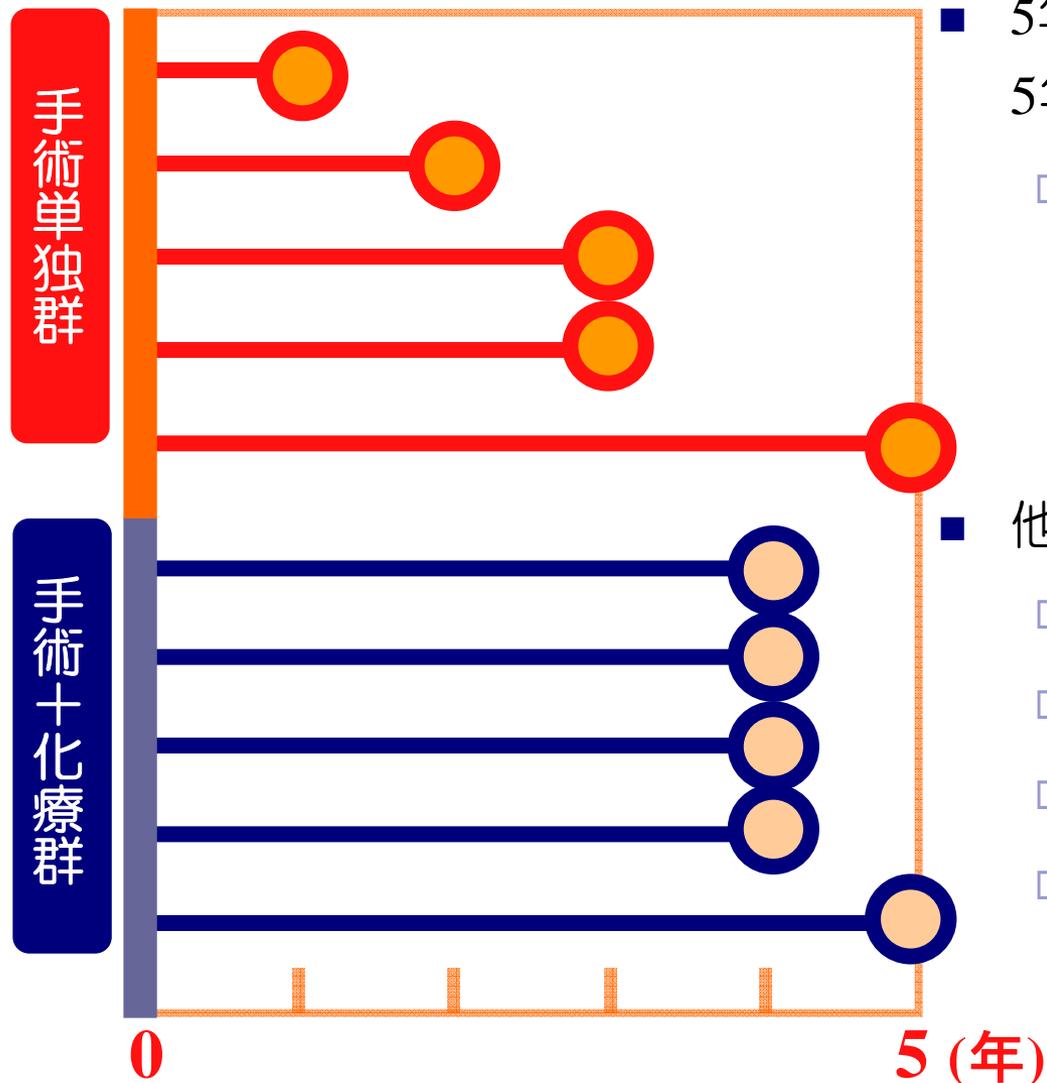


平均値を求めてみよう！



- A ~ Dの4人のみなら
 - 平均は $(1+2+3+3) / 4 = 2.25$
- Eも含めると、
 - 平均は $(1+2+3+3+?) / 5 = ?$
計算できない、計算不能
- 生存時間で平均値は使わない
 - 代わりに中央値
 - 生存期間中央値
median survival time (MST)

生存割合を比べてみよう



- 5年の研究終了時点で比較すると、5年生存割合の差 $0-0=0$
 - でも、みるからに手術群の生存時間は長い
 - 時間の長い短いが考慮できていない
- 他の年で比べることもできる
 - 1年では、差が $20-0=20\%$
 - 2年では、差が $40-0=40\%$
 - 3年では、差が $80-0=80\%$
 - 4年では、差が $80-80=0\%$
 - 恣意的に選ぶのは必ずしも適切でない

生存時間解析では平均や割合が適切でない

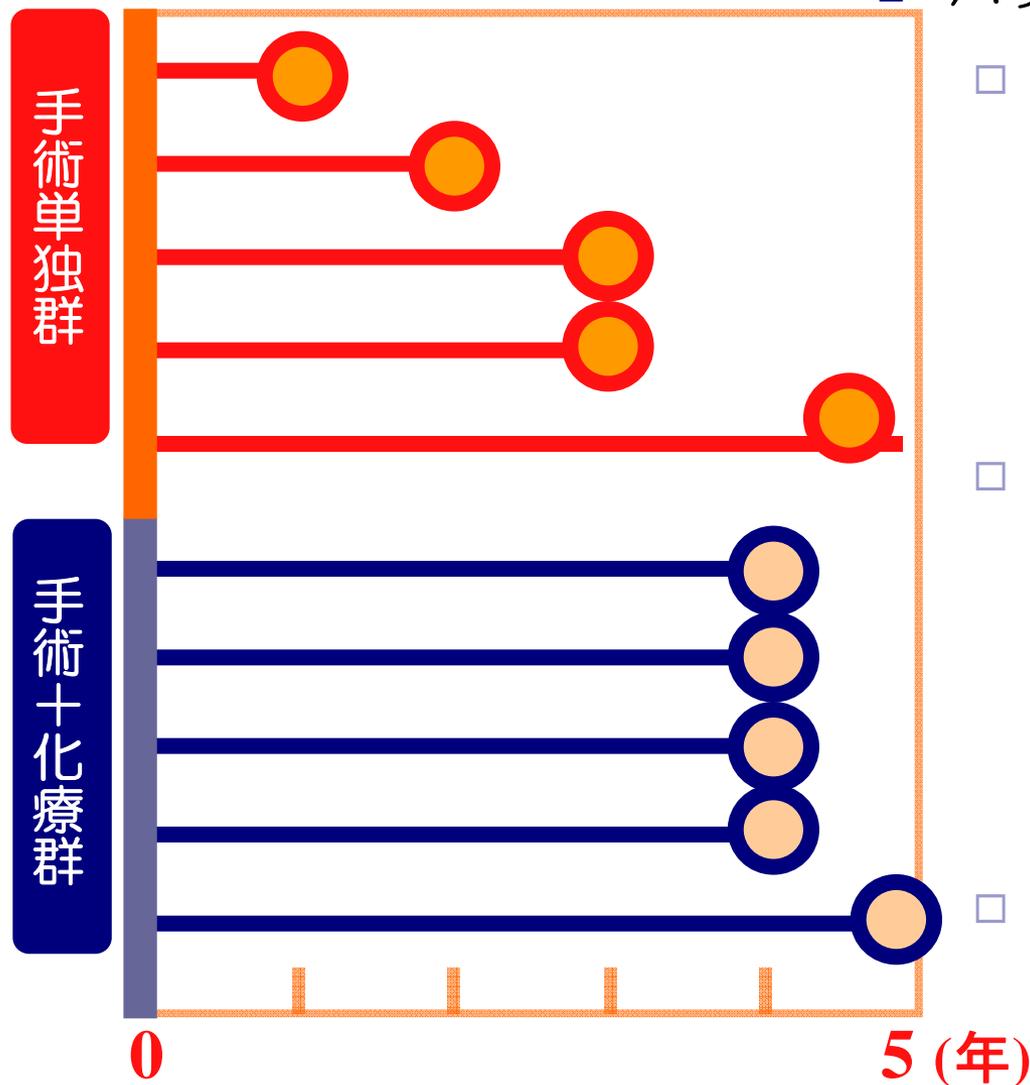
- 平均は打ち切りがあると求められない
- 生存割合では適切に比較できない
 - 時間の長さも考慮して、群間比較したい
 - 中央値でも同じく恣意的
- 生存時間解析ではハザード（=率）という概念を導入
 - 打ち切りも時間の長さも考慮できる指標

- 単純には

$$\text{ハザード} = \frac{\text{観察死亡数}}{\text{追跡人年 (人×年)}}$$

で近似

ハザード（瞬間死亡率、死亡の速度）



- ハザード=1年あたりの死亡率

□ 手術単独群

$$\frac{5}{1+2+3+3+5} = \frac{5}{14}$$

= 0.36/年
1年100人あたり36人が死亡

□ 手術+化療群

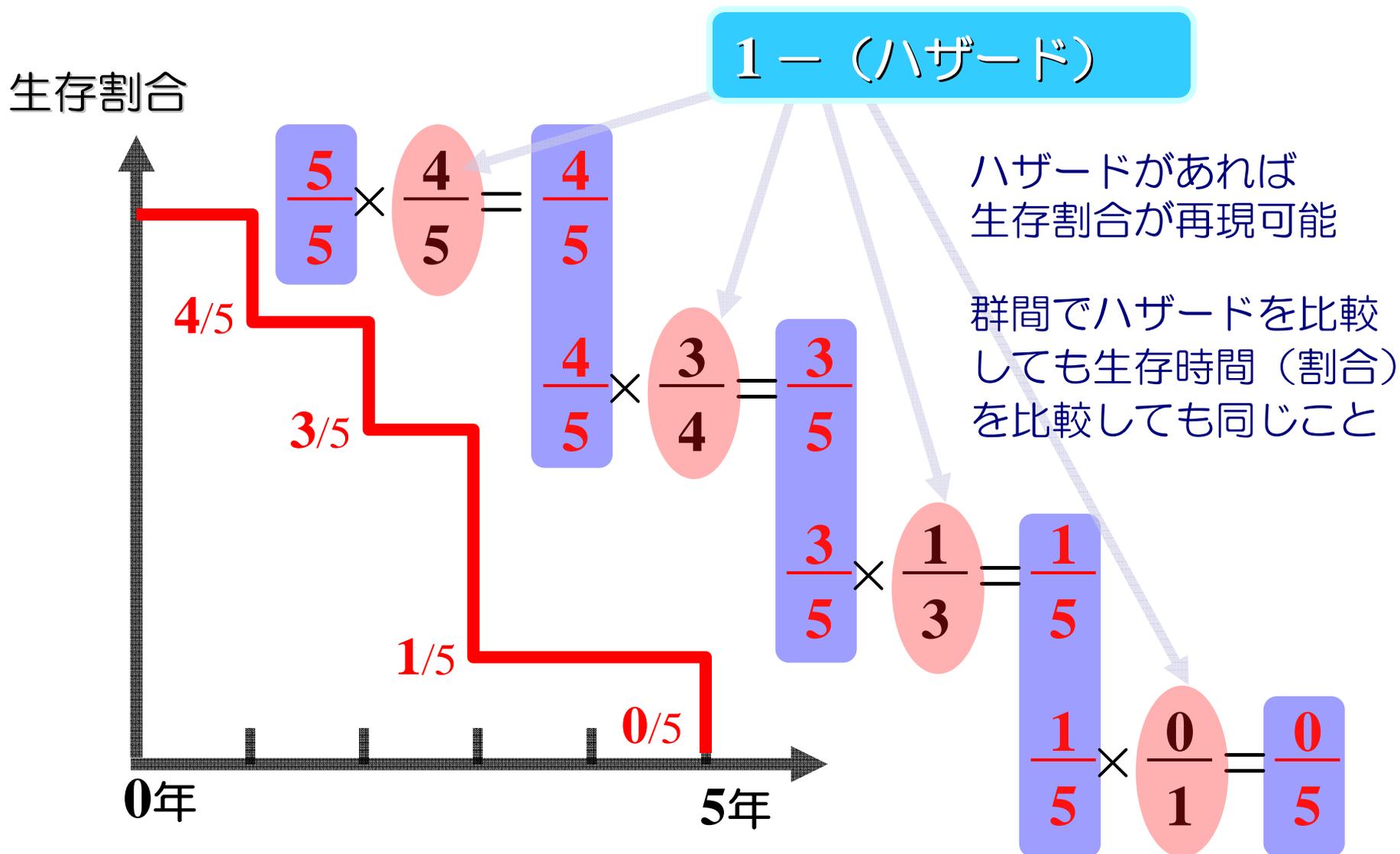
$$\frac{5}{4+4+4+4+5} = \frac{5}{21}$$

= 0.24/年
1年100人あたり約24人が死亡

- ハザードの比をとると 1.5

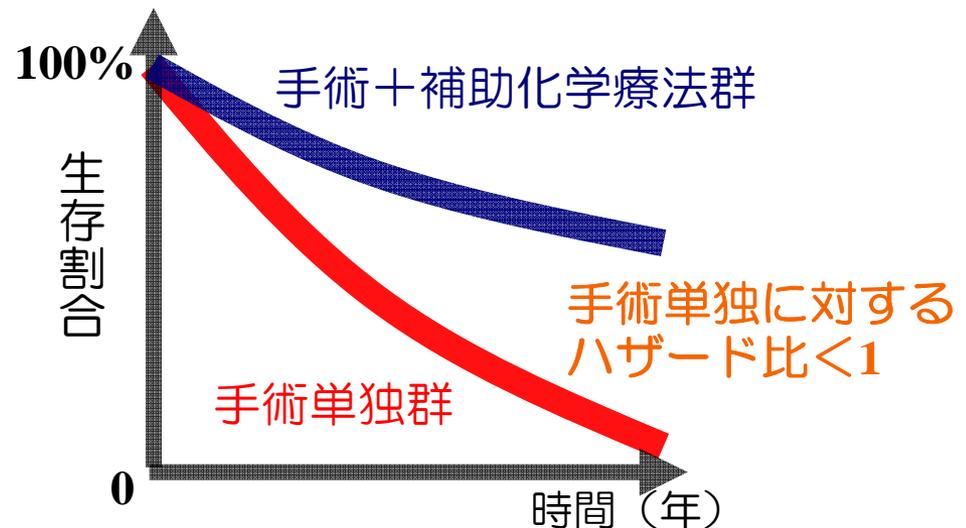
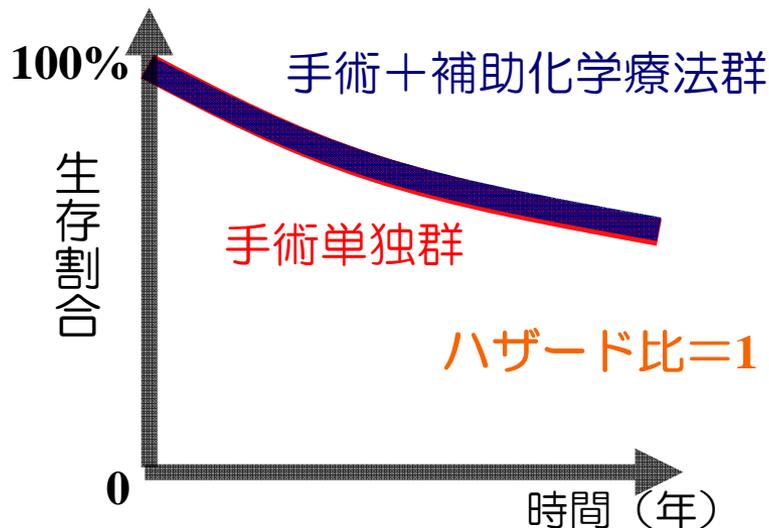
- 化療しないとハザード1.5倍

ハザードと生存時間は一対一の対応



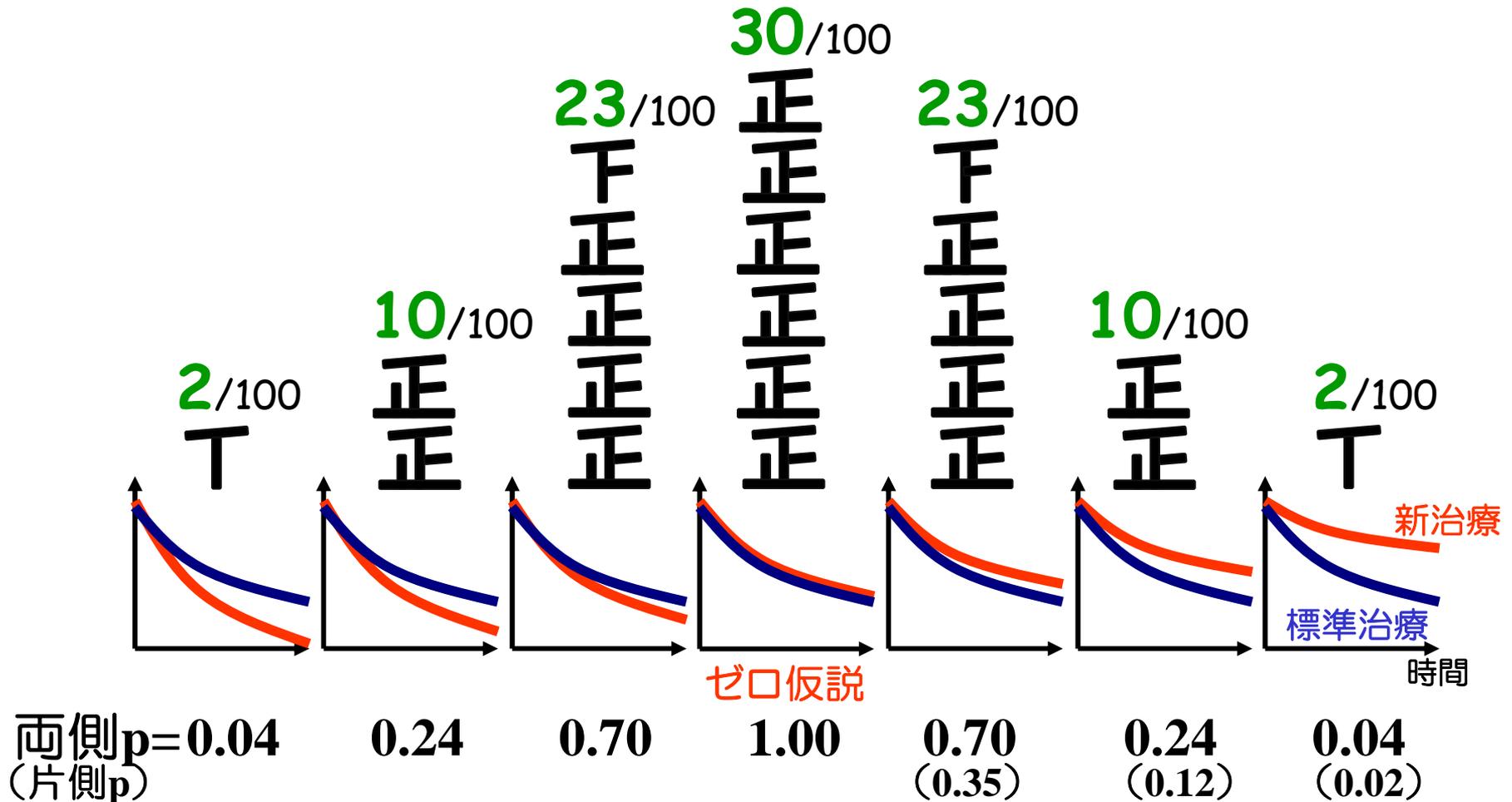
ハザードにより群間で生存曲線を比較しよう！

- 帰無仮説：群間でハザードが等しい
- 対立仮説：群間でハザードが異なる
 - ハザードを比較することにより生存曲線を比較
 - 生存曲線が等しい場合、ハザード比=1となる
- 仮説検定により帰無仮説を否定できれば対立仮説を採用
 - 方法：ログランク検定、Cox回帰など（後述）



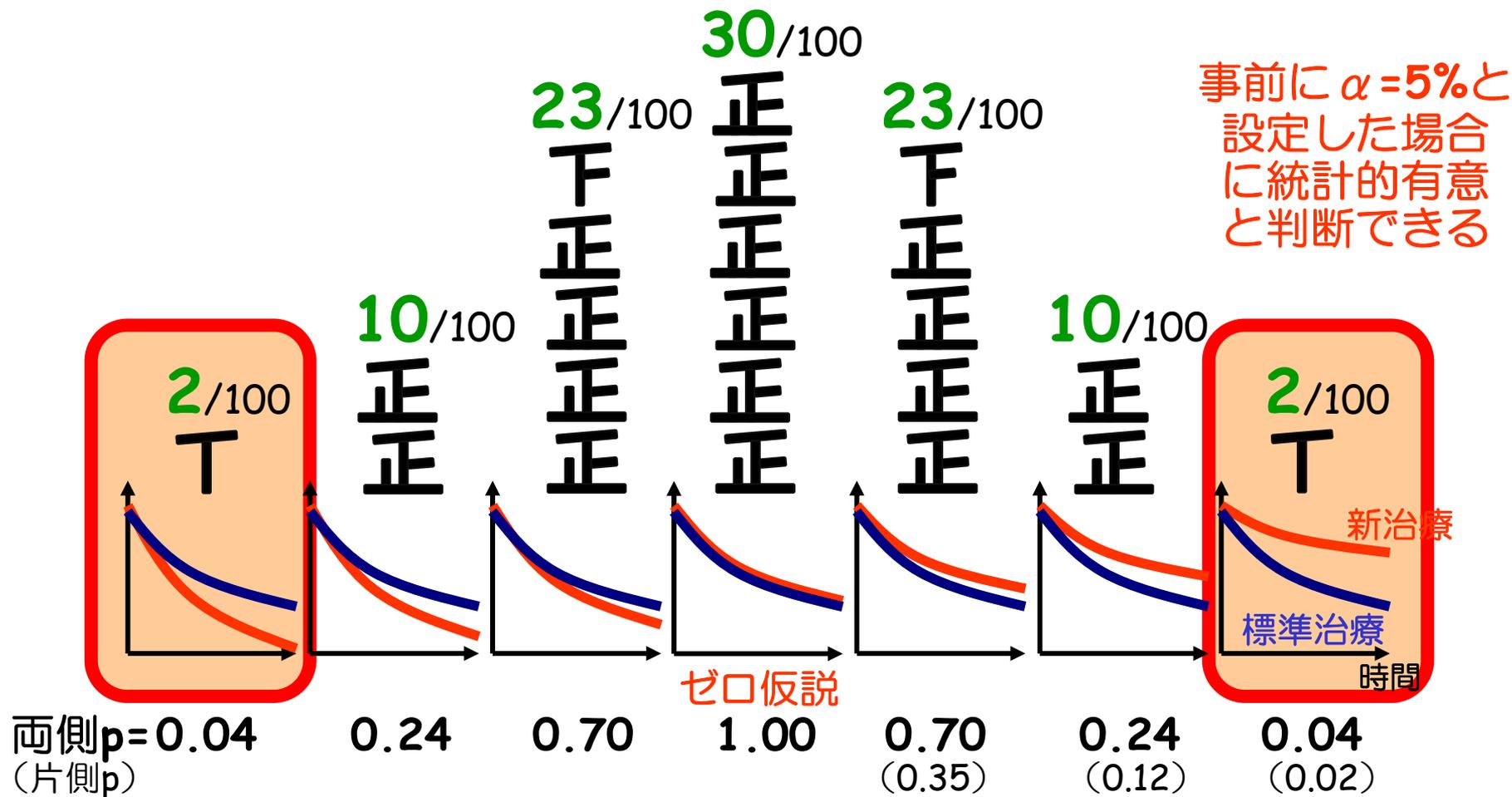
参考. 生存時間解析における仮説検定

- ゼロ仮説（帰無仮説）が正しい場合に得られる結果のバラツキを求め、観察されたデータの不自然さをp値で評価



参考. 生存時間解析における仮説検定

- ゼロ仮説（帰無仮説）が正しい場合に得られる結果のバラツキを求め、観察されたデータの不自然さをp値で評価





生存曲線を比較する際の注意点

- 検査に依存する生存時間
- 打ち切りによるバイアス
- 生存曲線全体を比較するのが標準
- Cox回帰は万能でない（時間があれば）

生存曲線を比較する際の注意点

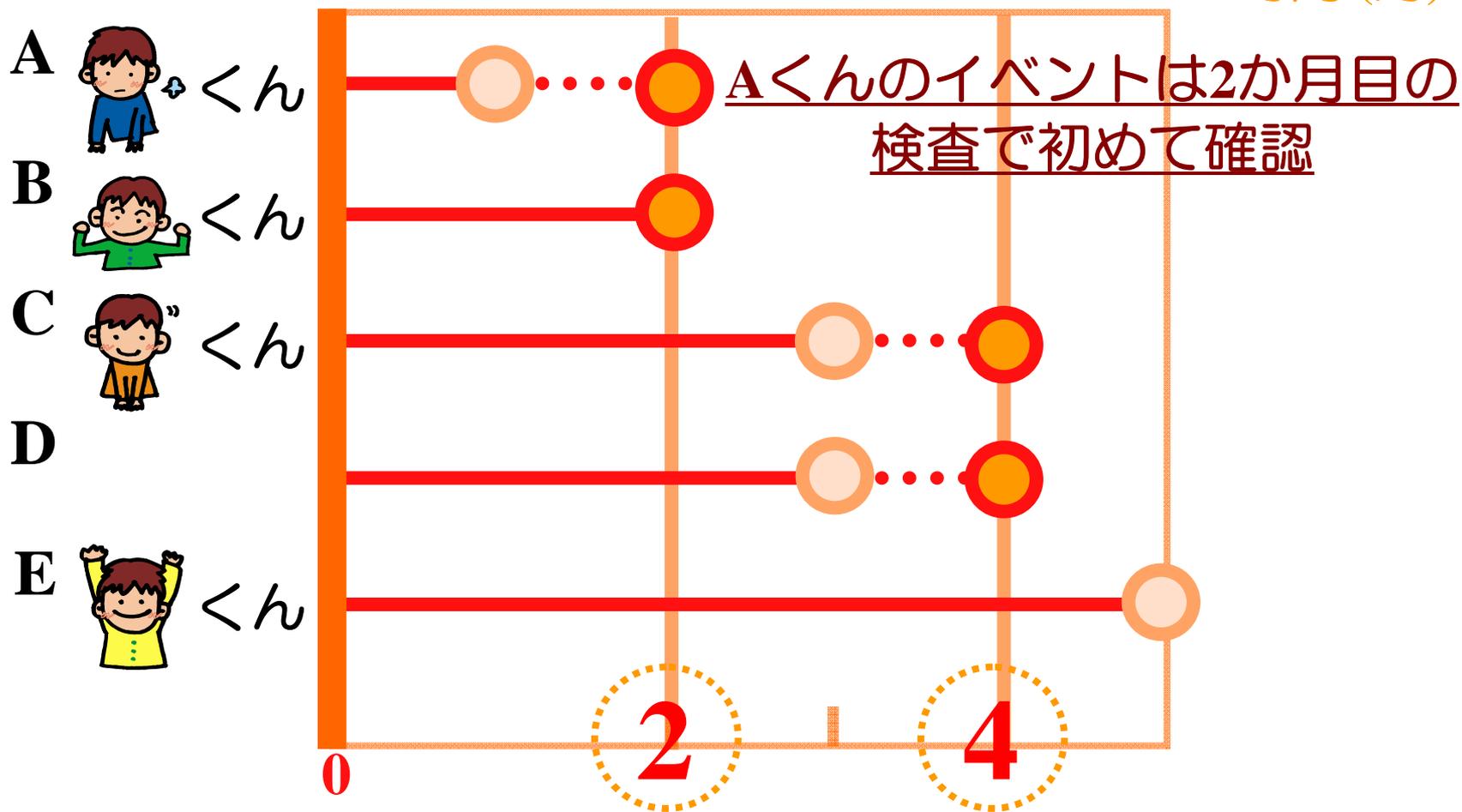
- 検査に依存する生存時間
- 打ち切りによるバイアス
- 生存曲線全体を比較するのが標準
- Cox回帰は万能でない（時間があれば）

検査に依存する生存時間

- 生存時間の誤差はイベントの種類に依存
 - 誤差が小さいイベント：死亡のみ
 - 死亡日は（日単位でも）正確に測定可能
 - 誤差が大きいイベント：再発など死亡以外
 - イベントの確認が検査に基づく場合、正確な測定は不能
 - 骨転移した正確な日付は測定不能
 - 臨床研究で、代わりに測定できるのは「骨転移を検査で初めて確認できた日」
≠ 骨転移した正確な日付

2か月ごとに再発イベントを調べると

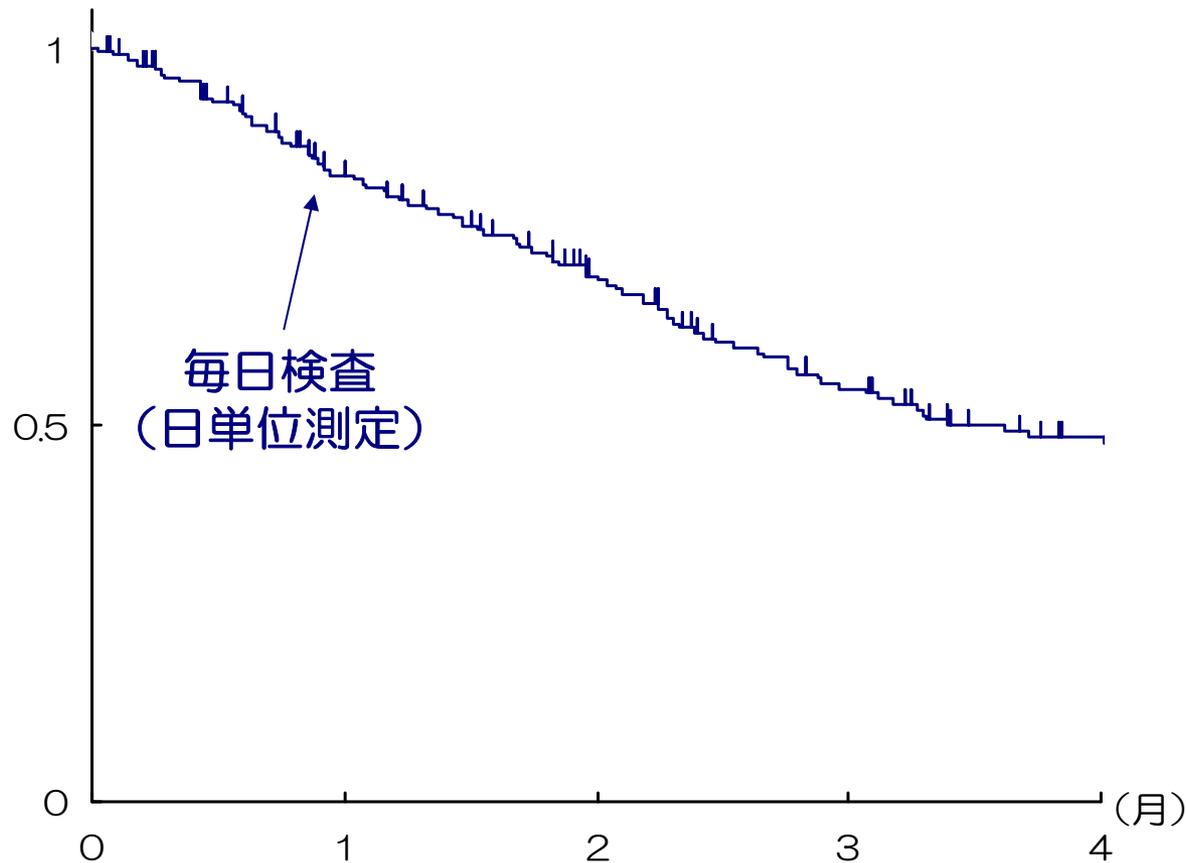
追跡
時間(月)



- Aくんは0~2年のどこかでイベントが起こっていたことのみがデータから分かる（統計学的には区間打ち切りの問題という）

群間比較への影響

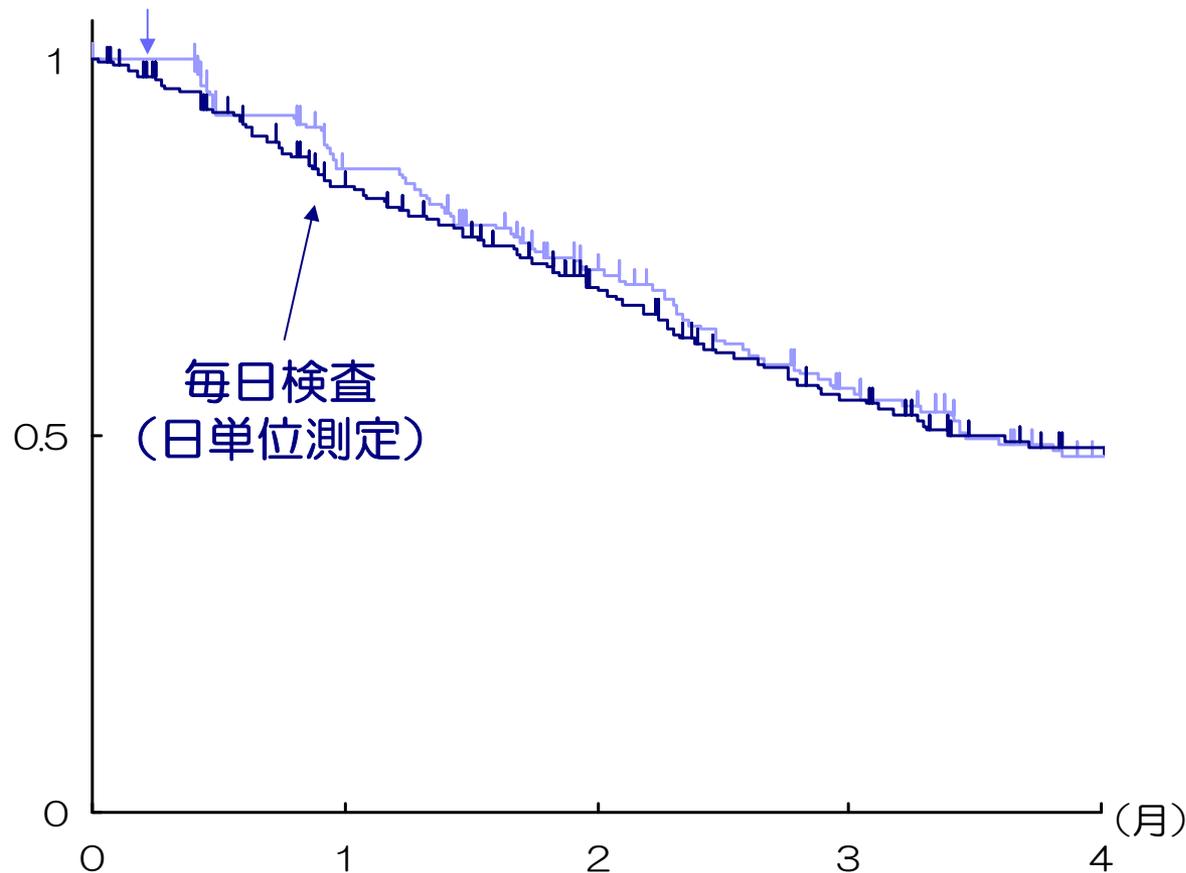
各対象者の検査間隔を空けてみると (シミュレーション)



群間比較への影響

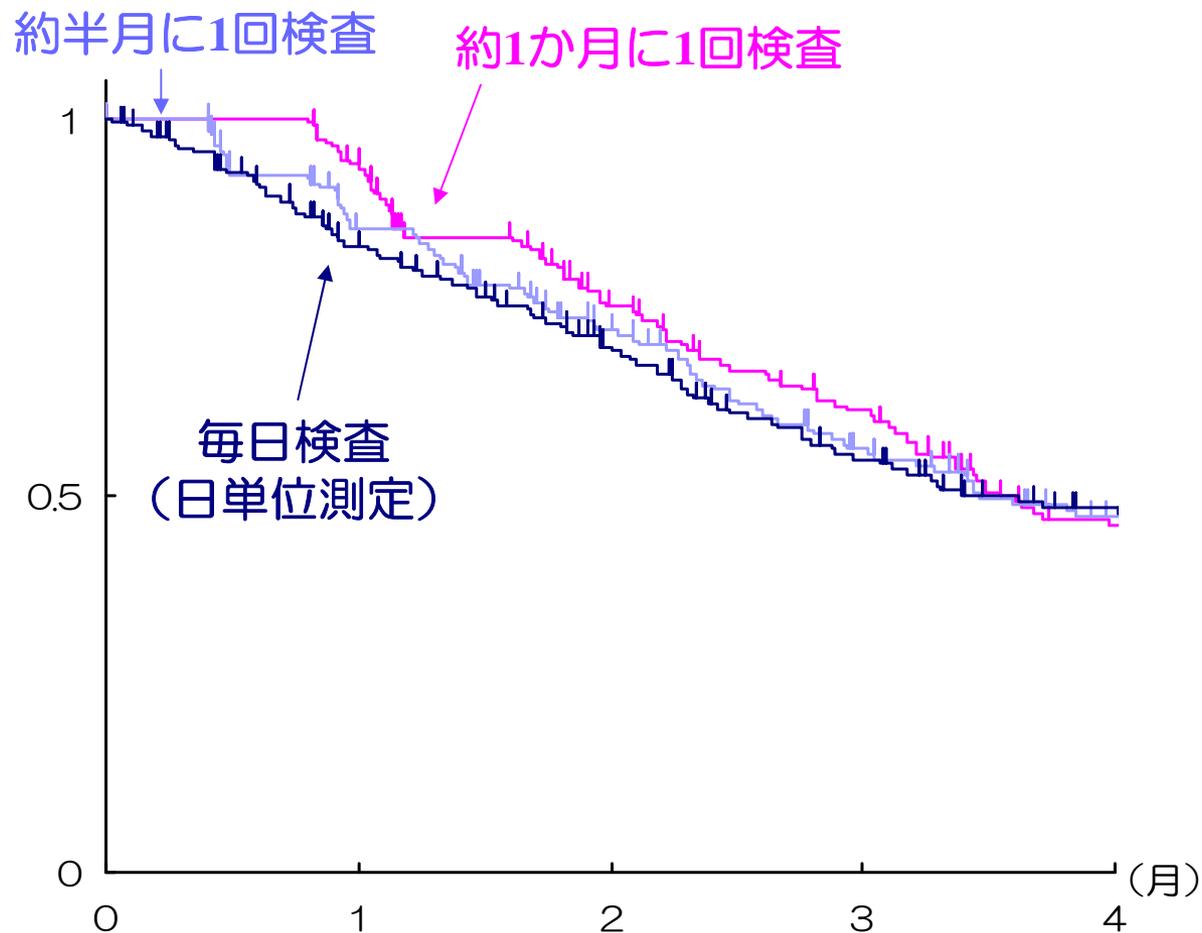
各対象者の検査間隔を空けてみると (シミュレーション)

約半月に1回検査



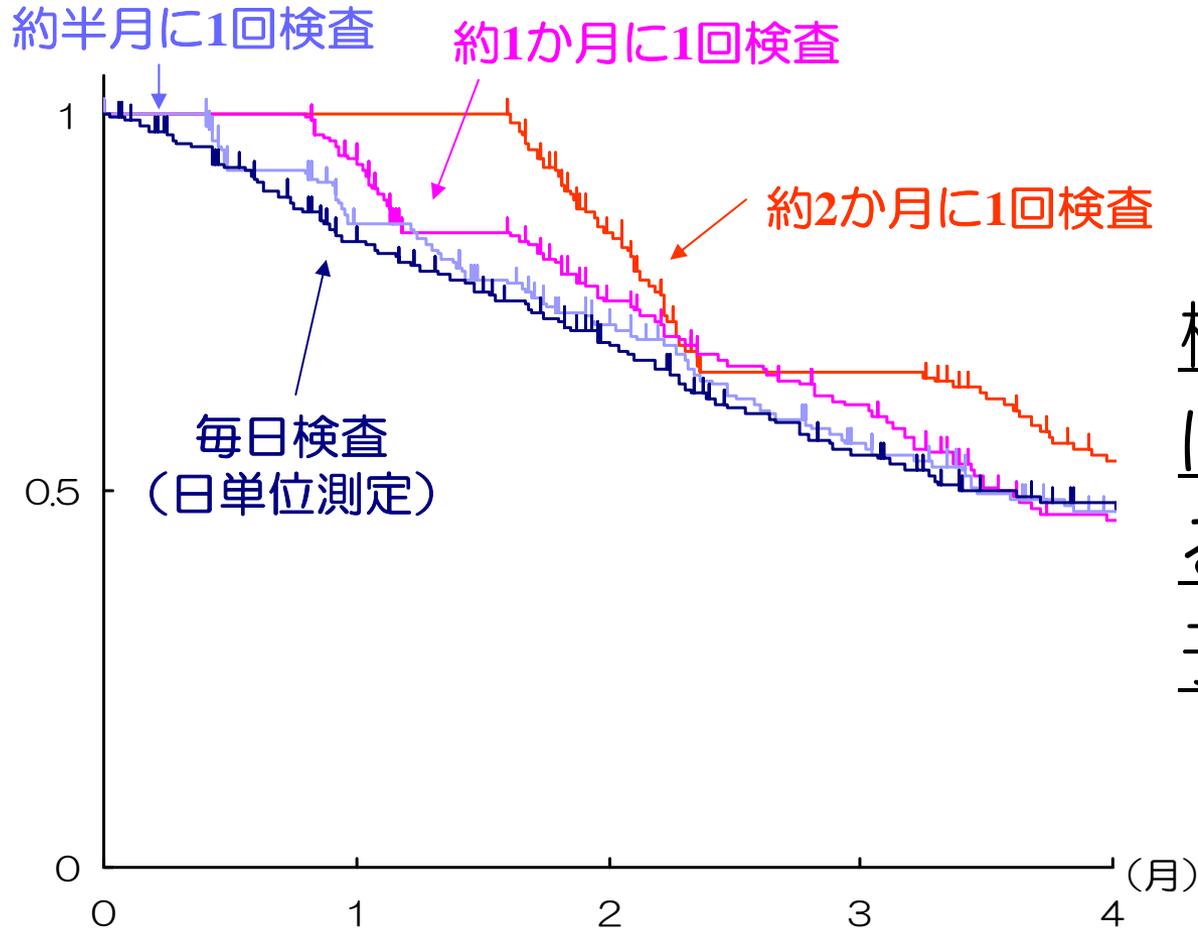
群間比較への影響

各対象者の検査間隔を空けてみると (シミュレーション)



群間比較への影響

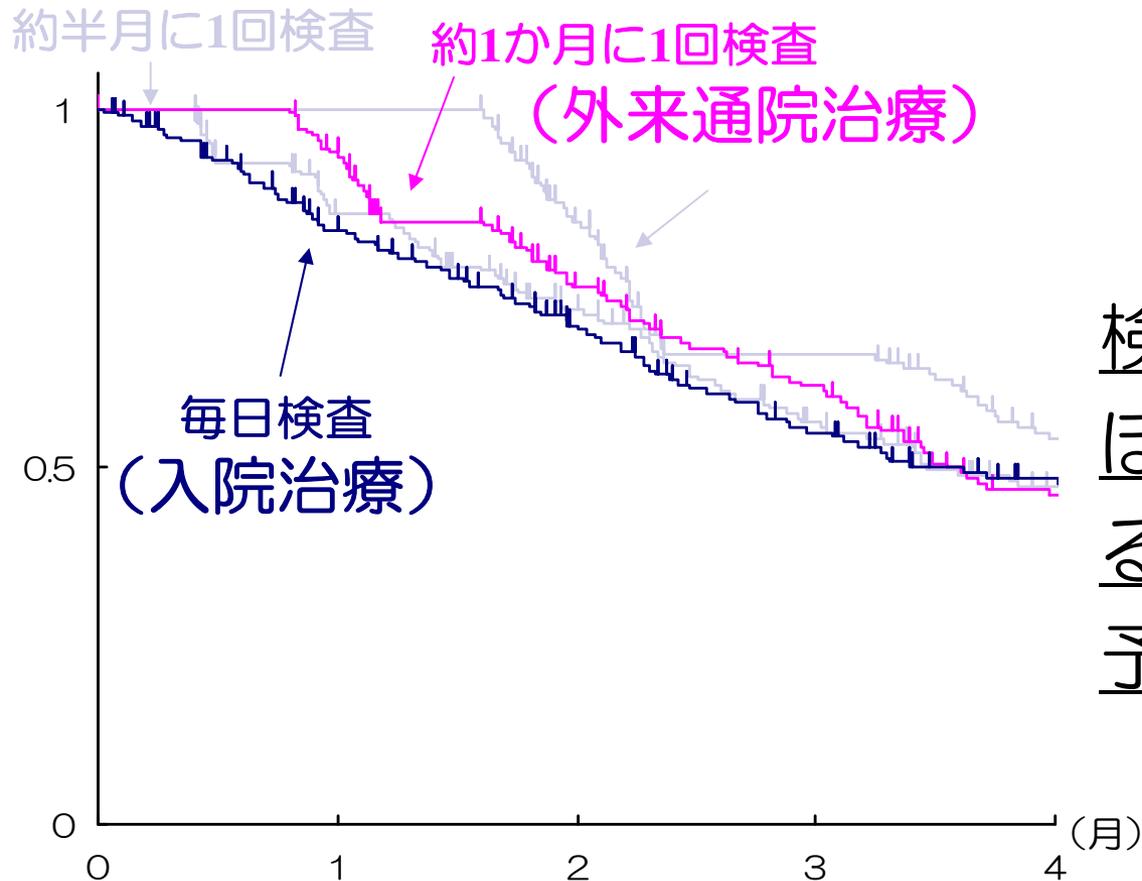
各対象者の検査間隔を空けてみると (シミュレーション)



検査間隔を空ける
ほど÷検査を怠け
るほど見かけ上の
予後は良い方向に

群間比較への影響

各対象者の検査間隔を空けてみると (シミュレーション)



検査間隔を空ける
ほど ÷ 検査を怠け
るほど見かけ上の
予後は良い方向に

群間で検査間隔が等しくない場合の群間比較には要注意!

生存曲線を比較する際の注意点

- 検査に依存する生存時間
- 打ち切りによるバイアス
- 生存曲線全体を比較するのが標準
- Cox回帰は万能でない（時間があれば）

打ち切りによるバイアス

- 興味のあるイベントと打ち切りが関連するとバイアスが生じる
 - **情報のある打ち切り**：状態の悪い患者さんが転院して脱落
 - こちらは安易に打ち切りにすべきでない
 - **情報の無い打ち切り**：病状とは関係なく転院して脱落

- 不適切な打ち切り処理によるバイアス
 - その研究でイベントが生じなかった対象に関して、イベントが本当に無かったことを確認した最終日（最終無イベント確認日）が正確に測定できていないとバイアスが生じる

打ち切りによるバイアス (1)

- 興味のあるイベントと打ち切りが関連するとバイアスが生じる
 - 再発までの時間 time-to-progression に対する解析で「再発せずに死亡した時点」で打ち切り

バイアスを減らすためには・・・

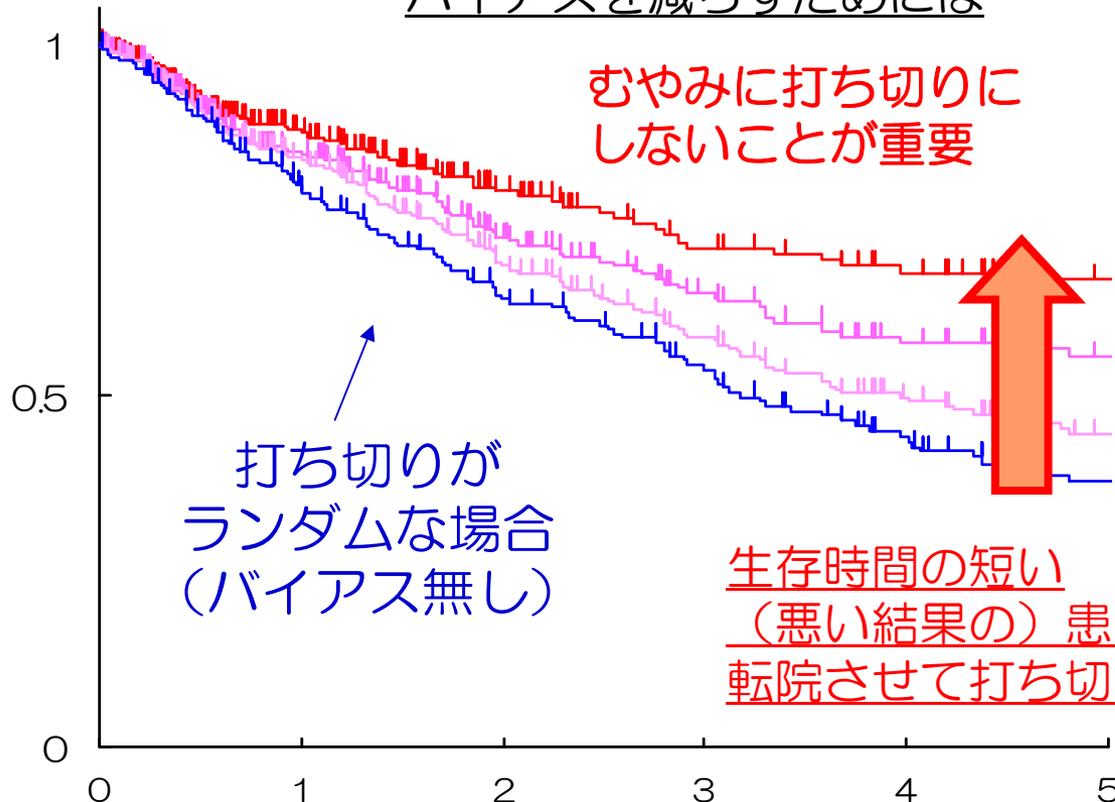
1. 関係する全てのイベントが生じるまでの時間を評価
e.g. 無増悪生存期間
(検証的臨床試験に向く)

むやみに打ち切りにしないことが重要

2. cumulative incidenceを評価
(血液腫瘍では標準法の1つ)

見かけ上の生存曲線は
良い方向に偏る

生存時間の短い
(悪い結果の) 患者さんを意図的に
転院させて打ち切りにすると・・・
(シミュレーション)



打ち切りが
ランダムな場合
(バイアス無し)

打ち切りによるバイアス (2)

JCOG9205試験

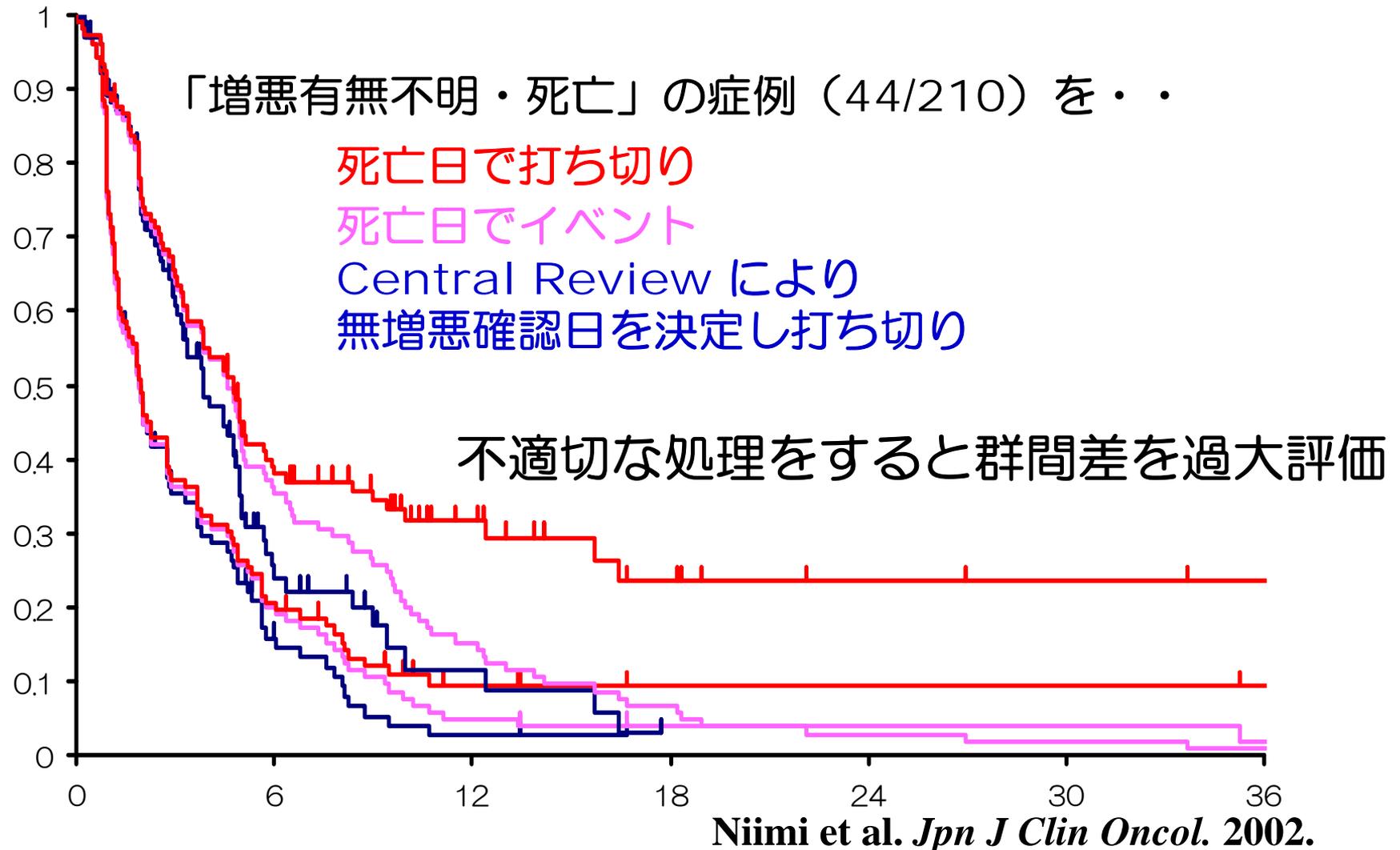
例. 無増悪生存期間における打ち切りの影響

Niimi et al. *Jpn J Clin Oncol.* 2002.

打ち切りによるバイアス (2)

JCOG9205試験

例. 無増悪生存期間における打ち切りの影響





生存曲線を比較する際の注意点

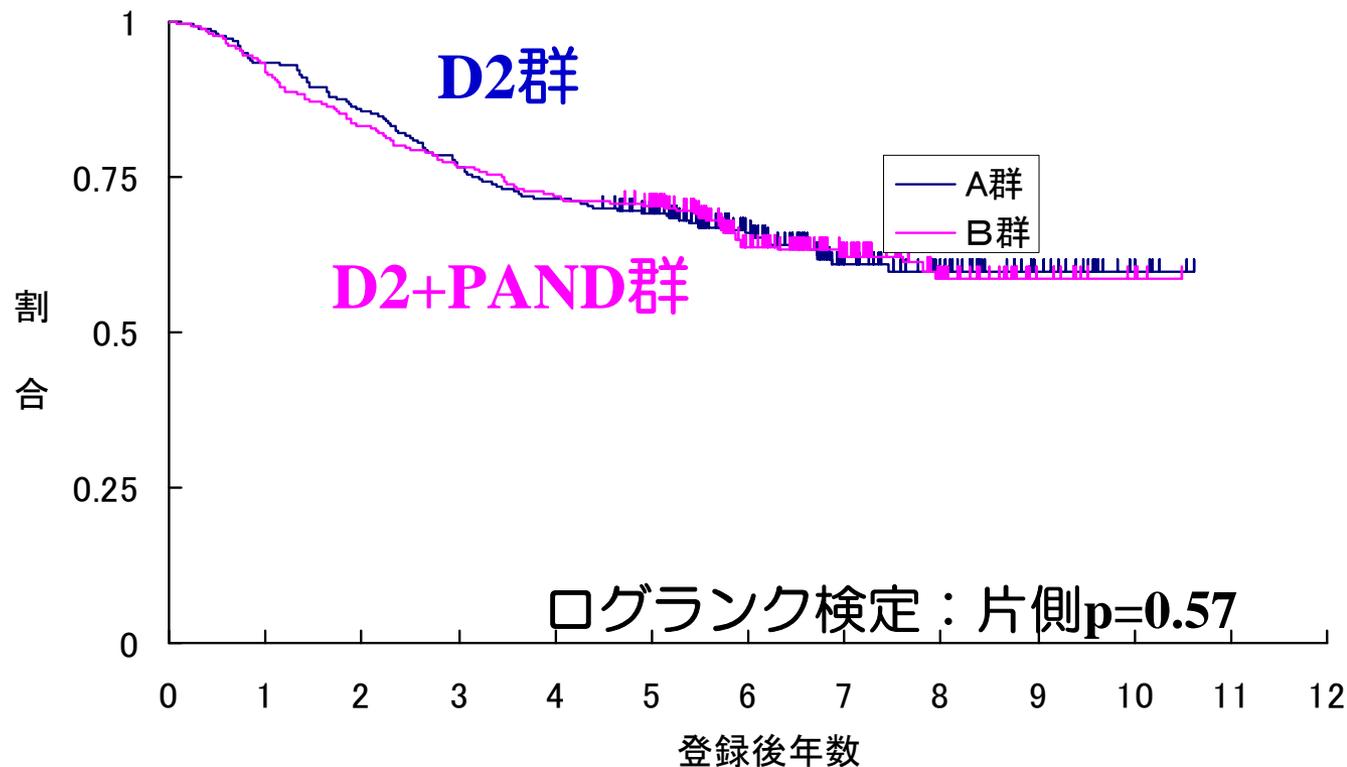
- 検査に依存する生存時間
- 打ち切りによるバイアス
- **生存曲線全体を比較するのが標準**
- Cox回帰は万能でない（時間があれば）

生存曲線全体を比較するのが標準

- 順位に基づく検定（ログランク検定、一般化Wilcoxon検定など）
 - イベントが観察されるごとにハザード（≡生存割合）を群間比較

JCOG9501試験（胃癌外科グループ）

Sasako *et al.* ASCO 2006

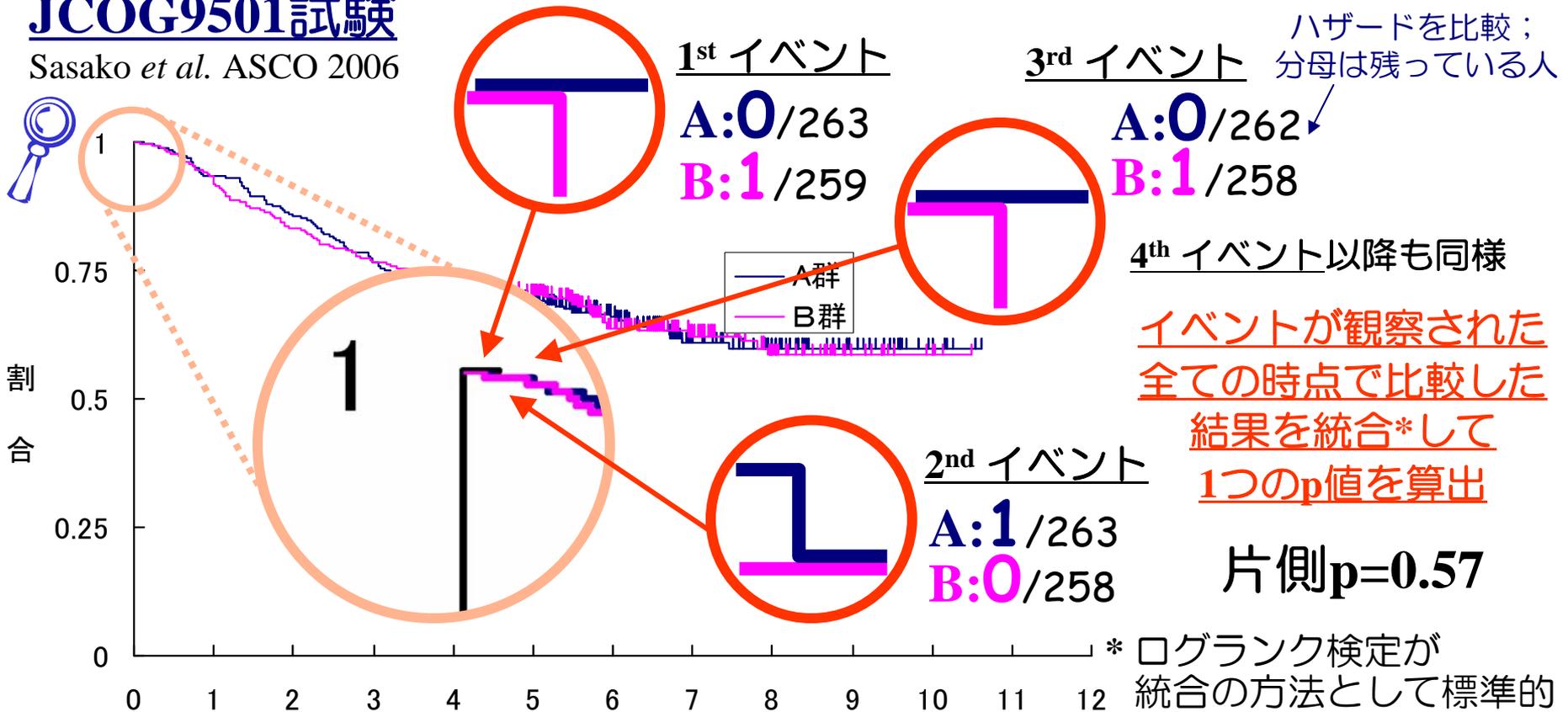


生存曲線全体を比較するのが標準

- 順位に基づく検定 (ログランク検定、一般化Wilcoxon検定など)
 - イベントが観察されることにハザード (≡生存割合) を群間比較

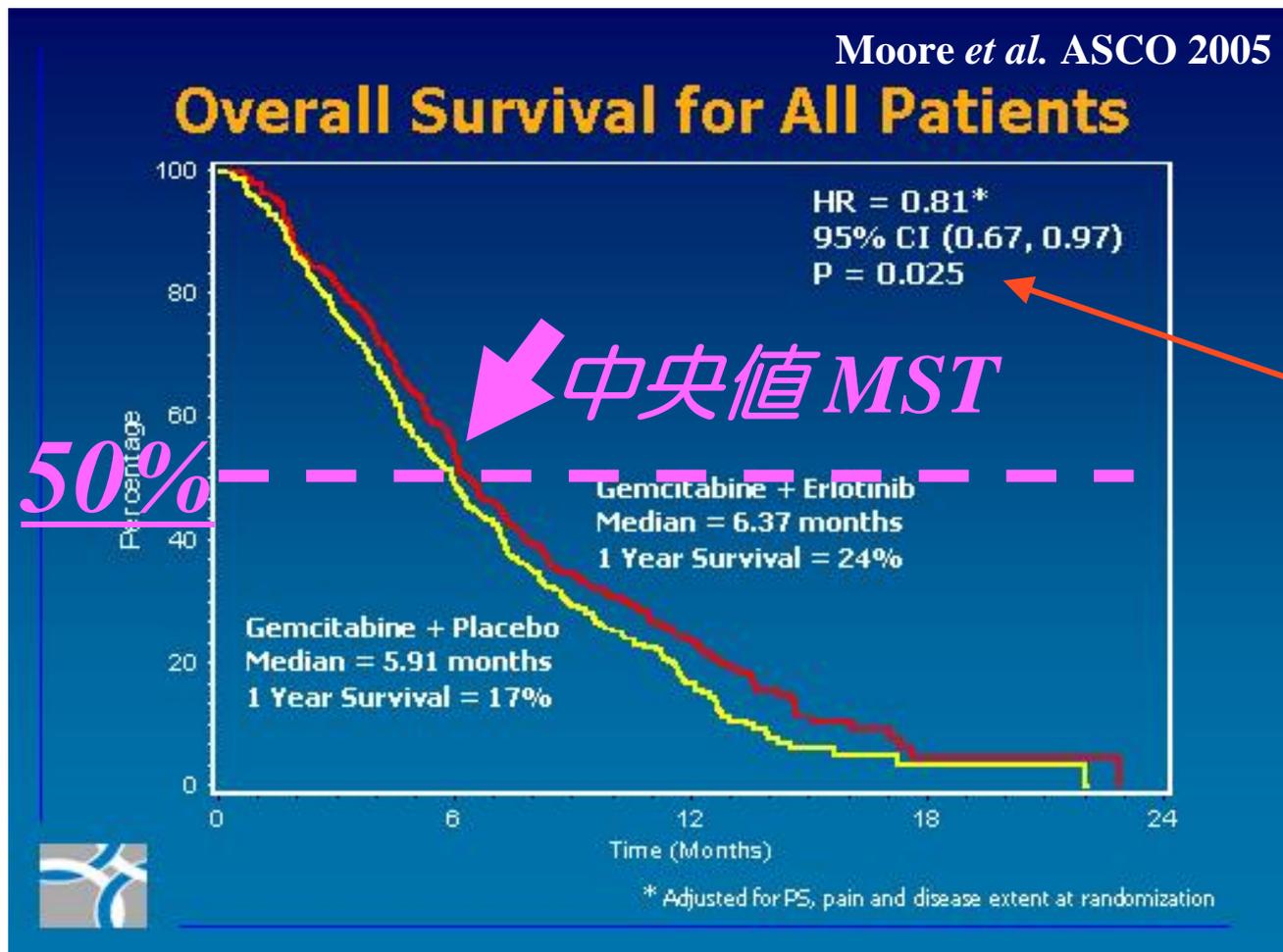
JCOG9501試験

Sasako et al. ASCO 2006



5年生存割合や中央値(MST)を比較しているのではない!

例. NCIC-CTGによる進行膵癌の比較試験



臨床家によく
尋ねられる典型例

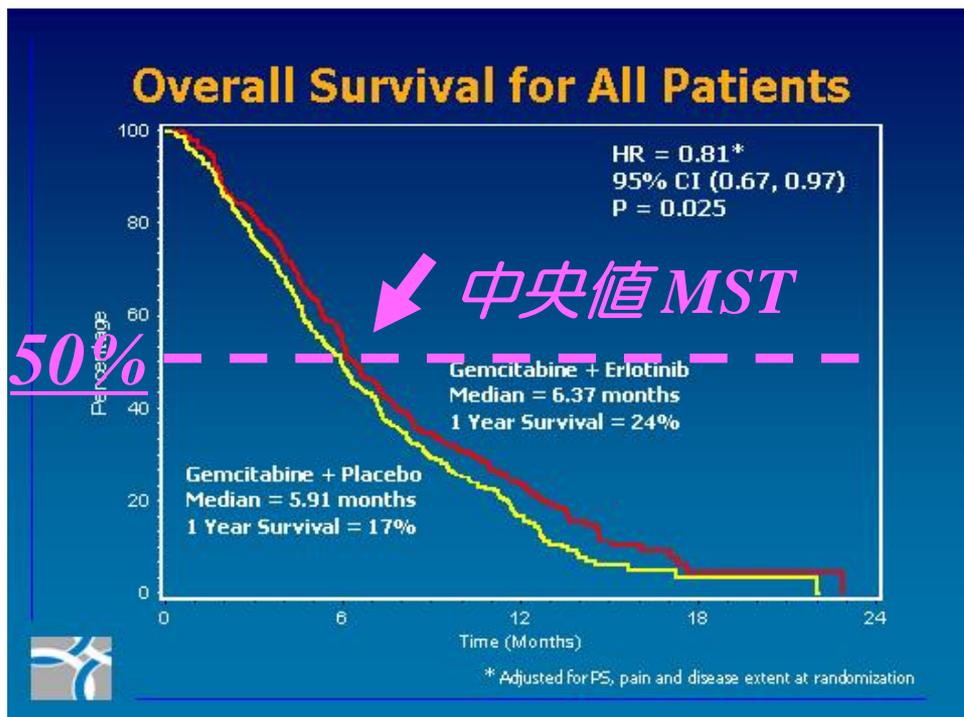
ログランク検定
による両側 $p=2.5\%$

事前に定めた
 $\alpha=5.0\%$ よりも小

統計的有意!

- ところが 6.4か月 vs. 5.9か月、2週間の差しかない！
 - 統計家がいろいろイカサマをしたのでは・・・！？

ログランク検定は生存曲線全体を比較 (MSTではない)



直感的にも・・・
よくみるとMST周辺だけ
差が小さくなっている
(偶然縮まった可能性も十分ある)

- 有意なp値が示すのは「2つの曲線は全体的にみると異なる」
- MSTも誤差を含む
 - 信頼区間付きで評価すべき
- 曲線全体 (ハザード比) から参考として概算すると
 - 単剤群中央値6か月の場合
併用群中央値7.4か月
 - 95%信頼区間を付けると
[6.2か月, 9.0か月]
- 勿論、臨床的にどうかが一番重要

生存曲線を比較する際の注意点

- 検査に依存する生存時間
- 打ち切りによるバイアス
- 生存曲線全体を比較するのが標準
- **Cox回帰は万能でない（時間があれば）**

Cox回帰 *Cox regression*

- *Cox proportional hazards model*: 英国の大統計家Sir Coxによる
- 説明変数1つ (e.g. 群) のみであればログランク検定と同等の結果が得られる
- ログランク検定ではハザード (瞬間死亡率) をイベント観察時点ごとに比較したが、これに対してモデル化

□ **ハザード** = (標準治療のハザード) ←

× (新治療の上乗せ効果)

× (全身状態による差)

× (病期による差)

× . . .

この部分を特定
しなくて良いのが
Cox回帰の利点

余談だが、
全く同じアイデアを
後から遅れて提案した
米国の経済学者が
ノーベル経済学賞
(2000年のMcFadden)

- **その他の変数を調整**した下で興味のある変数の効果 (ここでは群間のハザード比) を評価可能
 - モデルによる調整について詳しくは次回

臨床医がよく多変量解析と表現しますが

■ 統計用語の（もはや慣習的な）誤用

- 「多変量 multi-variate」：結果変数が沢山あること
 - 例. 多変量Cox回帰：臨床医と統計家の間で異なるものを想像しうる
- 「多変数 multi-variable」：説明変数が沢山あること
 - 例. 複数の説明変数をCox回帰に含めることで調整して解析する

■ 統計用語を厳密に用いれば、、、

臨床医がよく云うところの「多変量解析」は「単変量解析」

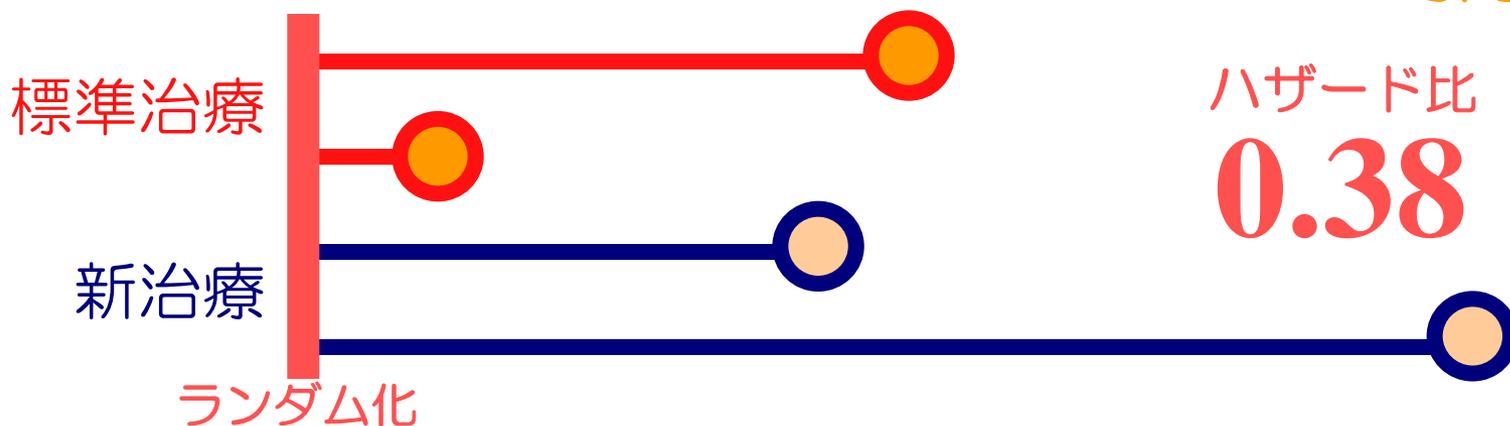


Piantadosi “Clinical Trials, 2nd” Wiley. 2005.

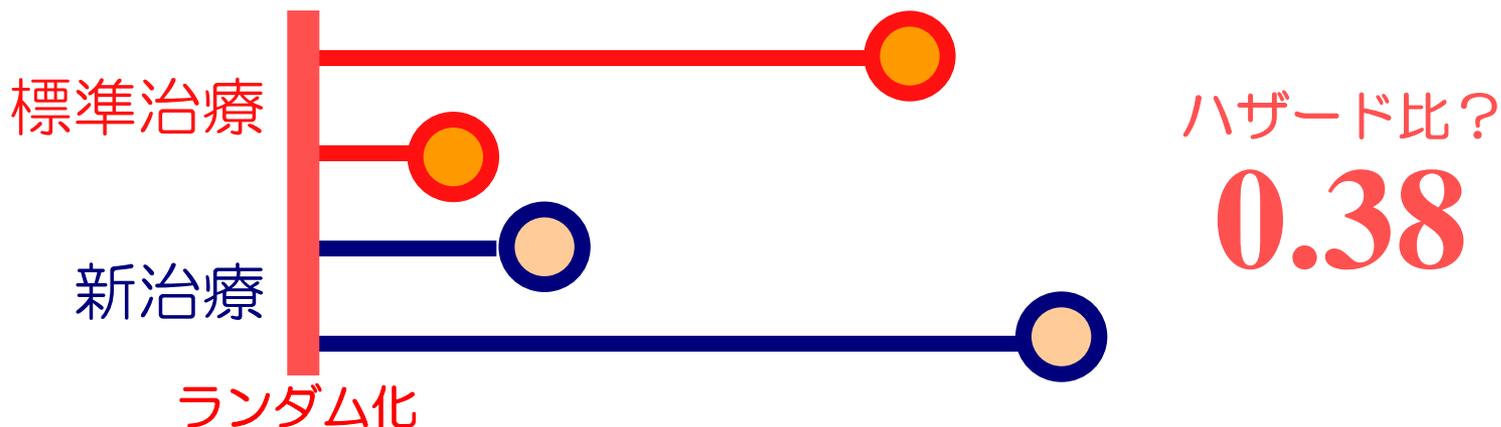
群間で比較しているのは順位

→ 追跡時間

試験
①



試験
②



◇ 両試験とも等しい (ログランク検定のp値も等しくなる)

Cox回帰 (*adjusted analysis*) は万能か

- どちらの結果を信用するか?
 - 遺伝子Xと糖尿病罹患の間には有意な関連がみられた
 - しかしながら、先行研究で報告されていた30のリスク要因全てをCox回帰を用いて調整してみたら
遺伝子Xと糖尿病罹患の間には有意な関連がみられなかった
- case by case、明確なcook-bookはない
 - 観察研究 vs. ランダム化研究：ランダム化ならばまず前者が良い
 - (Cox回帰に限らず) モデルによる結果が正しいためには
「モデルが正しい」という強い前提を要する
 - 「モデル」は所詮「モデル」：他にも適切なモデルがあるのかも
 - 次回に詳しく

Take Home Messages (まとめ)

- 生存時間は何らかのイベント（事象）が発生するまでの時間
- Kaplan-Meier法を用いれば、打ち切りが存在する下でも生存曲線を適切に推定できる
- 生存曲線の群間比較は曲線全体（ハザード）を比較することによって行うのが標準、ログランク検定はその1つの標準法
- 検査間隔や打ち切りの種類・処理によってバイアスが生じうることに十分に留意して研究計画、解析を行うべき



参考.用語

- 生存時間 (survival time, survival data) の他にも・・・
 - failure time, failure time data 故障時間
 - time-to-event, time-to-event data } ほぼ同義で使用可

- 生存時間解析 (survival analysis) の他にも
 - time-to-event (data) analysis
 - lifetime (data) analysis
 - failure-time (data) analysis } ほぼ同義で使用可

参考. がん領域で用いられる主な「生存時間」

	対象	起算日	イベント
生存期間 Overall Survival	全例 (登録例/適格例)	登録日 (ランダム割付日)	全死亡 (死因を問わない)
無増悪生存期間 Progression-free Survival Time to Progression	全例	登録日	増悪/全死亡 (または原病死)
治療成功期間 TTF Time to Treatment Failure Failure-free survival Event-free survival	全例	登録日	治療中止 (または毒性中止) /増悪/全死亡
無病生存期間 Disease-free Survival CR期間 Complete Response Duration	術後： 全例 (治癒切除例) 進行例： CR例	術後： 登録日 進行例： CR判定日 /登録日	再発/全死亡
奏効期間 Response Duration	奏効例 (CR/PR)	PR/CR判定日 /登録日	増悪/全死亡

* 奏効期間の群間比較は一般に適切でないことに注意