

放射線疫学入門と広島・長崎原爆被爆者分析の課題

2019/6/1

「分野を横断した放射線疫学の研究会」 2019夏

濱岡 豊

hamaoka@fbc.keio.ac.jp

慶応大学商学部

広島、長崎被爆者データ

用いたデータは広島および長崎の放射線影響研究所（放影研）から入手したものである。放影研は、日本の厚生労働省（厚労省）ならびに米国のエネルギー省（DOE）により資金提供を（後者については、その一部を米国学士院に対するDOE研究助成金DE-HS0000031を通じて）受けている公益財団法人である。この報告書に示した結論は著者のものであり、必ずしも放影研またはその資金提供機関の判断を反映するものではない。

はじめに 放射線の健康影響を巡る混乱？

国立がん研究センターが放射線影響について緊急記者会見

100mSv未満の線量なら発がんリスクなし

2011/3/29

大滝隆行=日経メディカル

国立がん研究センター（理事長：嘉山孝正氏）は3月28日、緊急記者会見を開き、**福島第一原子力発電所の被災による現時点での放射性物質汚染の健康影響について、チェルノブイリ事故や広島・長崎の原爆生存者の追跡調査などのエビデンスから、「原子炉付近で作業を行っている人を除けばほとんど問題がない」とする見解を発表した。**

中略

また、同センター研究所長代理の中益齊氏は、「外部被曝による甲状腺癌リスクについて15歳以上の大人ではほとんどリスクが増加せず、15歳以下の小児においても100mSv以下であれば有意なリスク上昇は認められない」とした。

中略

一方、がん対策情報センターが「がん情報」統計部長の祖父江友孝氏は、日本の原爆被災者約10万人の追跡調査結果から、「100～200mSv以下の低線量域では、広島・長崎の原爆被爆者においても明らかな発がんリスクの増加は確認されていない」と強調した。



- はじめに
 - 放射線の健康影響を巡る混乱？
 - 本日の講演の目的
- 統計学&データ分析入門
- 原爆被曝者データ分析
- 原爆被曝者データ分析の課題
- 100mSv問題の原因？
- 参考文献

低線量放射線の健康影響について (原子力安全委員会H23/5/20)

- 「確率的影響」には、被ばくから一定の期間を経た後にある確率で、固形ガン、白血病等を発症することが含まれます。ガンのリスクの評価は、疫学的手法によるものが基礎となっています。広島や長崎で原子爆弾に起因する放射線を受けた方々の追跡調査の結果からは、100mSvを超える被ばく線量では被ばく量とその影響の発生率との間に比例性があると認められております。一方、100mSv以下の被ばく線量では、ガンリスクが見込まれるものの、統計的な不確かさが大きく疫学的手法によってガン等の確率的影響のリスクを直接明らかに示すことはできない、とされており、このように、100mSv以下の被ばく線量による確率的影響の存在は見込まれるものの不確かさがあります。
- そこでICRPは、100mSv以下の被ばく線量域を含め、線量とその影響の発生率に比例関係があるというモデルに基づいて放射線防護を行うことを推奨しております。また、このモデルに基づく全世代を通じたガンのリスク係数を示しております。それは100mSvあたり0.0055（100mSvの被ばくは生涯のガン死亡リスクを0.55%上乗せする。）に相当します。

■ 本講演の目的

- データ分析の考え方を簡単な例を通じて理解する。
- 放射線疫学の代表的な研究である広島・長崎の被爆者データ分析の方法と、その問題点を紹介する。
- 100mSvを巡る混乱の原因を考察する。

■ はじめに

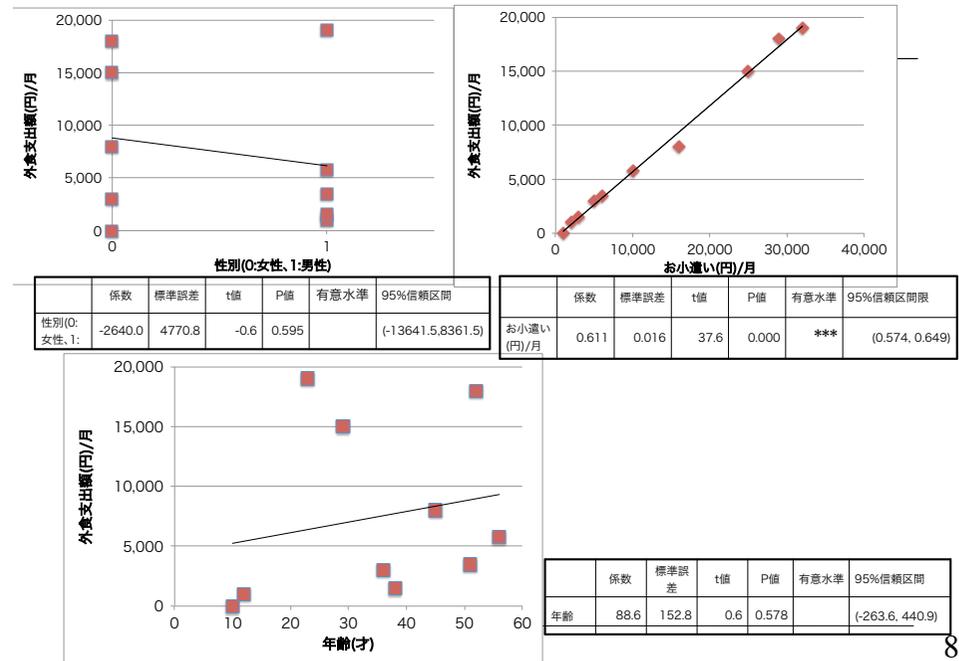
- 放射線の健康影響を巡る混乱?
- 本日の講演の目的
- 統計学&データ分析入門
- 原爆被曝者データ分析
- 原爆被曝者データ分析の課題
- まとめ
- 参考文献

分析のための準備運動1

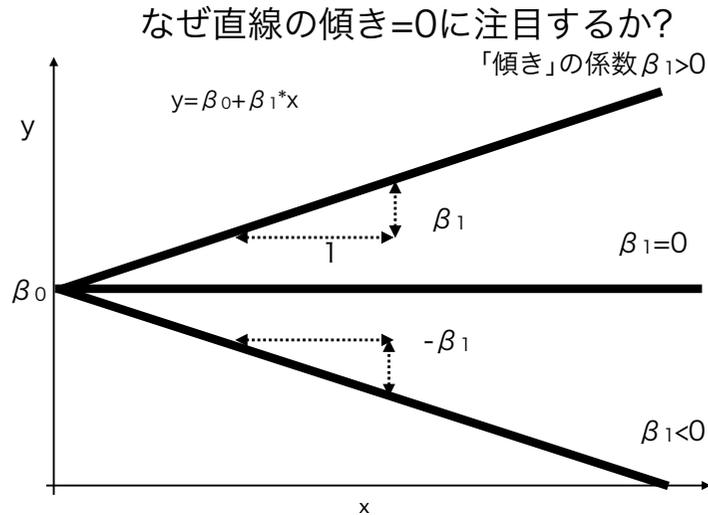
■ Q 下のようなデータがあります。外食支出額を決めるのはどの変数かを明らかにするには、どうすればよいのか?

番号	性別 (0:女性、1:男性)	年齢(才)	お小遣い (円/月)	外食支出額 (円/月)
1	0	10	1,000	0
2	0	29	25,000	15,000
3	0	36	5,000	3,000
4	0	45	16,000	8,000
5	0	52	29,000	18,000
6	1	12	2,000	1,000
7	1	23	32,000	19,000
8	1	38	3,000	1,500
9	1	51	6,000	3,500
10	1	56	10,000	5,800

データをグラフにまとめて、直線のあてはめ(回帰分析)



傾き=0に注目して仮説検定



9

統計的に「有意 significant」とは? とりあえず、(いい加減な)結果の読み方

	係数	標準誤差	t値	P値	有意水準	95%信頼区間
お小遣い (円)/月	0.611	0.016	37.6	0.000	***	(0.574, 0.649)

- 直線の傾きの「推定値」
 - 直線の傾きは 0.61
- t値
 - 「傾き=0」を検定するための指標。
 - これの絶対値が大きいほど、「傾き=0」でない可能性が大。
 - (サンプルサイズが大きいつき、t値の絶対値 > 1.96?)
- P値
 - 「傾き=0」が正しいとしたときにt値が上記の値となる確率。
 - この確率が小さい程、「傾き=0」は正しくなさそうと判断。
- 0.05=5%(研究によっては、10%)よりも小さければ、「傾き=0」ではないと考える。
 - P値 ≤ 0.05?
- 有意水準
 - P値に応じて下記のように表記。
 - P値が1%以下ならば「***」
 - 5%以下ならば「**」
 - (10%以下ならば「*」)
 - *がついているか?
 - 95%信頼区間 (100-5%)
 - この範囲の95%の確率で「(真の)傾き」の値が含まれている。
 - 区間に0が含まれているか?
 - 信頼区間の幅:誤差の小ささ

統計的に「有意 significant」

10

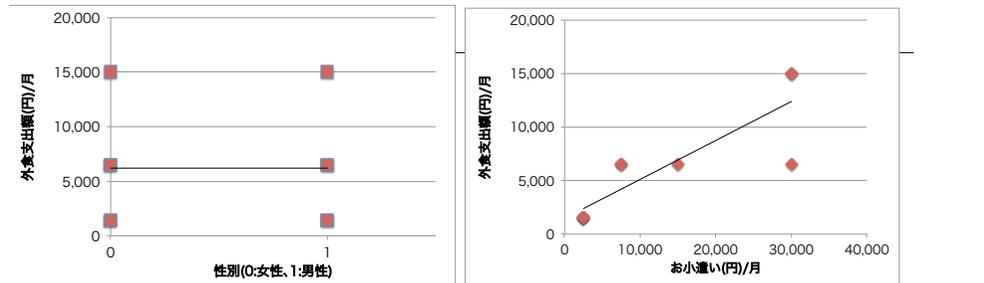
準備運動2

Q このデータについて、ある人は、次のように「カテゴリー化」して同様の分析をしました。このような方法のメリット、デメリットを挙げて下さい。

- 年齢
 - ~20才代
 - ~40才代
 - 41才~
- お小遣い
 - ~5000円
 - ~10000円
 - ~20000円
 - 20001円~
- 外食支出額(円)/月
 - ~3000円
 - ~10000円
 - 10001円~
- メリット
 - 匿名化可能
 - データ量を少なくできる
- デメリット
 - 情報の損失
 - カテゴリー化の任意性
 - カテゴリー数、各カテゴリーの区間
 - 例 年齢を3区分ではなく4,5区分しないのはなぜか?

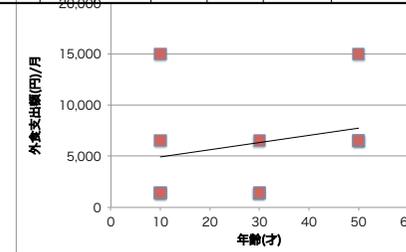
11

カテゴリー化したデータを用いて同様にプロット、推定



	係数	標準誤差	t値	P値	有意水準	95%信頼区間
カテゴリー化	0.000	489.984	0.0	1.000		(-8047, 92139)
オリジナル	-2640.0	4770.8	-0.6	0.595		(-13641.8361)

	係数	標準誤差	t値	P値	有意水準	95%信頼区間
カテゴリー化	0.364	0.075	4.8	0.001	***	(0.190, 0.537)
オリジナル	0.611	0.016	37.6	0.000	***	(0.574, 0.649)



	係数	標準誤差	t値	P値	有意水準	95%信頼区間
カテゴリー化	70.3	102.1	0.7	0.510		(-165.0, 305.6)
オリジナル	88.6	152.8	0.6	0.578		(-263.6, 440.9)

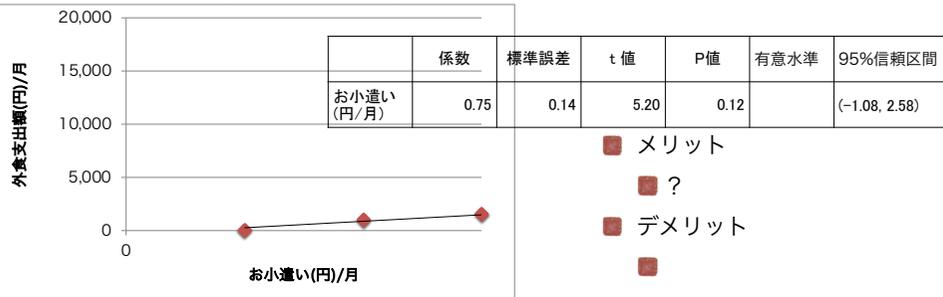
上段:カテゴリー化した推定結果

下段:オリジナルデータでの推定結果 12

準備運動3

Q このデータについて、お小遣いが少ない人(3000円以下)のみを用いて分析しました。このような方法のメリット、デメリットを挙げて下さい。

番号	性別 (0:女性、1:男性)	年齢(才)	お小遣い (円/月)	外食支出額 (円/月)
1	0	10	1,000	0
2	0	29	25,000	15,000
3	0	36	5,000	3,000
4	0	45	16,000	8,000
5	0	52	29,000	18,000
6	1	12	2,000	1,000
7	1	23	32,000	19,000
8	1	38	3,000	1,500
9	1	51	6,000	3,500
10	1	56	10,000	5,800



13

内容

- はじめに
 - 放射線の健康影響を巡る混乱?
 - 本日の講演の目的
- 統計学&データ分析入門
- 原爆被曝者データ分析
- 原爆被曝者データ分析の課題
- まとめ
- 参考文献

15

準備体操のまとめ

- 回帰分析の考え方
 - 二つの変数間の関係の有無を決定
- サンプルサイズが多い方が誤差は小さくなる。
 - サンプルサイズが大きい方がよい。
- カテゴリ化の問題
 - 情報の損失による検定力の低下。
 - カテゴリ設定の任意性
- 範囲を限定することの問題
 - サンプル数の減少および分散の減少により、検定力が低下。
- 仮説検定の考え方
 - 有意ではない の意味
 - 有意ではなくなる状況
 - 本当に関係が無い
 - 本当は関係があるが、分析が不適切
 - 個人データを集計してしまう
 - データ数が少ない
 - 有意ではない ≠ 影響(関係)がない

14

広島、長崎の被爆者データ

- 寿命調査(Life Span Survey: LSS)
 - Atomic Bomb Casualty Commission(ABCC; 現・放射線影響研究所)が収集
- 1950年時点での生存者のうち12万人(実際には1-3群(8.6万人)を分析することが多い)
 - 第1群(近距離被爆者)：爆心地から2,000 m以内で被爆した「基本群」被爆者全員
 - 第2群：爆心地から2,000-2,500 mで被爆した「基本群」全員
 - 第3群：(遠距離被爆者) 第1群と年齢・性が一致するように選ばれた、爆心地から2,500-10,000 mで被爆した人
 - 第4群(原爆時市内不在) 第1群と年齢・性が一致するように選ばれた、1950年の国勢調査時に広島・長崎に在住していたが原爆時には市内にいなかった人(2.7万人)。(NIC: Not in the City)

追跡調査の概要

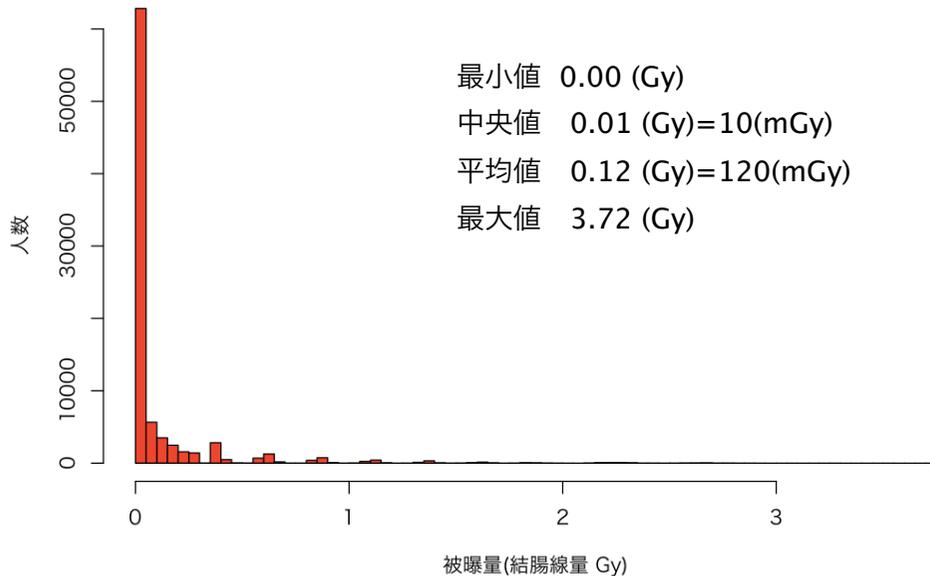
- 個人の登録
 - 被曝量の推計
 - 個人の被曝当時の所在地から被曝線量を推定
 - ガン登録データとの照合
 - ガン罹患(広島県、長崎県)
 - 死亡証明との照合
 - ガン、ガン以外の死因
 - その他病歴、死因
 - 約5-10年ごとに報告書、論文を発表
- 寿命調査 Life Span Survey (LSS)
 - 死亡についての分析
 - LSS13 1950-1997 Preston et al.(2003)
 - LSS14 1950-2003 Ozasa et al. (2012)
 - 罹患Incidenceについての分析
 - 1958-1998 Preston et al.(2007)
 - 1958-2009 Grant et al. (2017)

17

- Q 原爆被爆者の方の平均被曝量はどれくらい?(GyもしくはSv)

18

広島・長崎被曝者の被曝量の分布(86,611人。結腸線量)



19

LSS14報 (Ozasa et al. 2012)らの分析対象(NIC含まず)

線量 (Gy)	人数	都市		性別		被曝時 年齢	死亡数		
		広島	長崎	男性	女性		総数	固形ガン	白血病
-0.005	38,509	56.3%	43.7%	41.4%	58.6%	22.3	22,270	4,621	99
-0.1	29,961	75.9%	24.1%	41.2%	58.8%	22.1	17,292	3,653	78
-0.2	5,974	84.3%	15.7%	39.9%	60.1%	23.2	3,557	789	18
-0.5	6,356	79.7%	20.3%	39.0%	61.0%	23.4	3,898	870	27
-1.0	3,424	69.3%	30.7%	41.3%	58.7%	23.1	2,061	519	30
-2.0	1,763	65.3%	34.7%	46.1%	53.9%	22.2	1,127	353	39
2.0+	624	69.7%	30.3%	48.6%	51.4%	20.1	415	124	27
合計	86,611	67.5%	32.5%	41.2%	58.8%	22.4	50,620	10,929	318

注)死亡数・総数は固形ガン、白血病以外の死因も含む。

20

被曝量と健康との関係を明らかにする

■ Q このデータから、被曝量と(固形)ガン死の関係を把握するにはどうすればよいか?

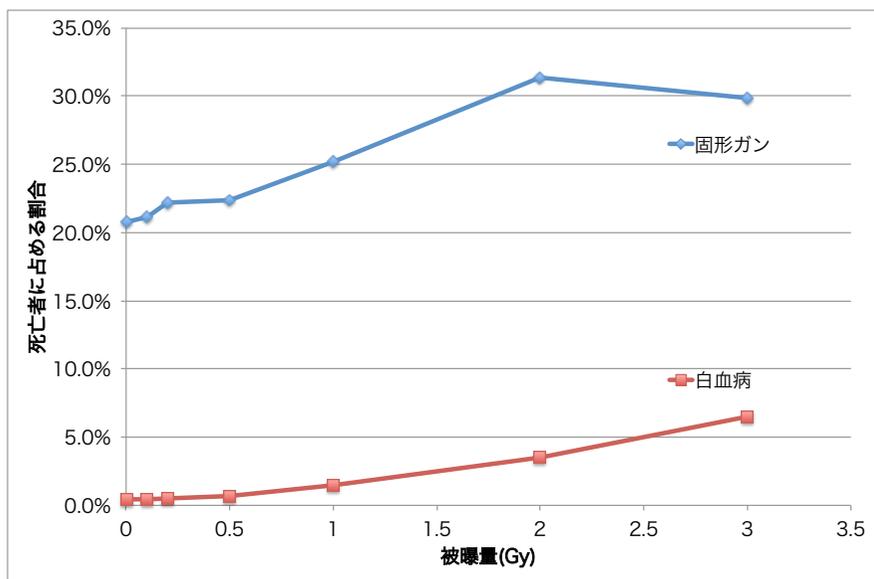
21

LSS14報 (Ozasa et al. 2012)らの分析対象(再掲)

線量 (Gy)	人数	都市		性別		被曝時 年齢	死亡数		
		広島	長崎	男性	女性		総数	固形ガン	白血病
-0.005	38,509	56.3%	43.7%	41.4%	58.6%	22.3	100.0%	20.7%	0.4%
-0.1	29,961	75.9%	24.1%	41.2%	58.8%	22.1	100.0%	21.1%	0.5%
-0.2	5,974	84.3%	15.7%	39.9%	60.1%	23.2	100.0%	22.2%	0.5%
-0.5	6,356	79.7%	20.3%	39.0%	61.0%	23.4	100.0%	22.3%	0.7%
-1.0	3,424	69.3%	30.7%	41.3%	58.7%	23.1	100.0%	25.2%	1.5%
-2.0	1,763	65.3%	34.7%	46.1%	53.9%	22.2	100.0%	31.3%	3.5%
2.0+	624	69.7%	30.3%	48.6%	51.4%	20.1	100.0%	29.9%	6.5%
合計	86,611	67.5%	32.5%	41.2%	58.8%	22.4	100.0%	21.6%	0.6%

22

プロットしてみると



23

実際の公開データ:連続量をカテゴリ化・集計して公開

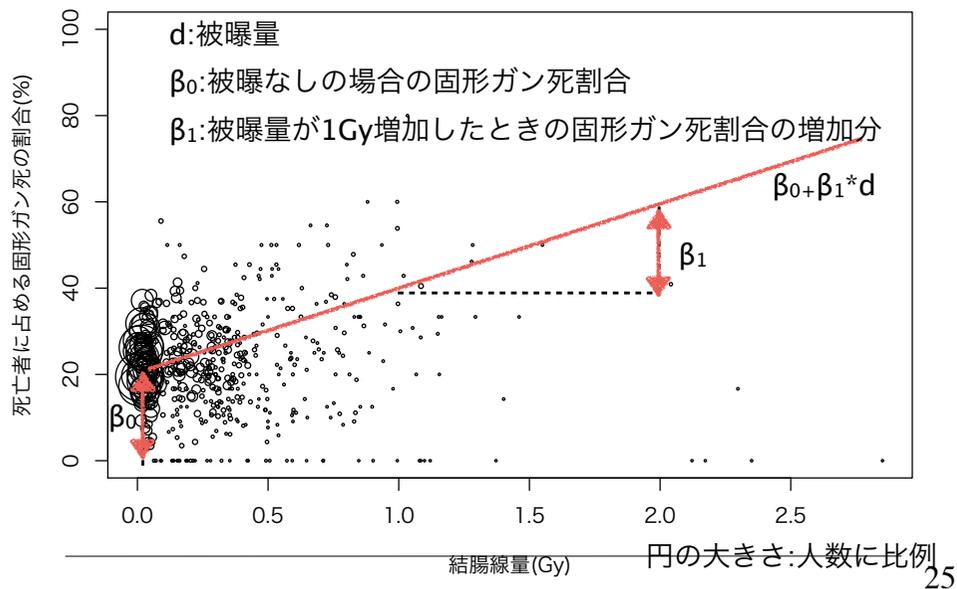
- 線量
- 22区分
- 都市
- 性別
- 被曝時年齢
- 到達年齢
- それぞれ5才刻み
- などを考慮する必要がある。

city	sex	gd3	ahs	agexca	agecat	ctime	doseca	subjec	pyr	agex	age	colon1	str	death	solid	leo
1	1	1	0	1	2	1	1	259	594.8	1.656	8.404	2.641	2	0	0	0
1	1	1	0	1	2	1	2	726	1701	1.712	8.432	10.54	11	0	0	0
1	1	1	0	1	2	1	3	277	571.2	1.862	8.505	29.09	21	1	0	0
1	1	1	0	1	2	1	4	151	335.2	1.786	8.466	49.08	51	1	0	0
1	1	1	0	1	2	1	5	109	252.3	1.726	8.44	70.32	71	0	0	0
1	1	1	0	1	2	1	6	79	177.8	1.712	8.432	90.46	91	0	0	0
1	1	1	0	1	2	1	7	98	209.6	1.807	8.474	112.1	111	1	0	0
1	1	1	0	1	2	1	8	61	135.1	1.87	8.511	137.8	141	0	0	0
1	1	1	0	1	2	1	9	64	142	1.7	8.415	163.2	161	1	0	0
1	1	1	0	1	2	1	10	63	131	1.94	8.546	187.4	181	0	0	0
1	1	1	0	1	2	1	11	63	161.8	1.506	8.329	222.3	221	0	0	0
1	1	1	0	1	2	1	12	60	118.4	2.026	8.56	271.9	271	2	0	0
1	1	1	0	1	2	1	13	118	258.2	1.777	8.459	393.4	393	1	0	0
1	1	1	0	1	2	1	14	59	144.9	1.659	8.405	592.8	61	1	0	0
1	1	1	0	1	2	1	15	4	8.594	2.052	8.602	811.3	811	0	0	0
1	1	1	0	1	2	1	16	3	3.782	3.063	9.108	1060	1060	0	0	0

放射研 ダウンロードできる研究データ
 寿命調査 第14報 ガンおよびガン以外の死亡率データ、1950-2003
<http://www.rerf.or.jp/library/dl/index.html>

24

「固形ガン死」についての実際のデータ
都市×性別×被曝時年齢×線量別集計と「線形モデル」



25

表 Ozasa et al.(2012)の推定結果
(固形ガン死:30才で被曝した方の70才時点での過剰リスク)

	推定値 (β_1 等)	標準 誤差	t値	P値	有意 水準	95%信頼区間
被曝量(結腸線量)Gy	0.42	0.05	8.40	<0.001	***	(0.32, 0.52)
被曝時年齢 (年齢-30)/10	-0.35	0.08	-4.25	<0.001	***	(-0.51,-0.19)
到達年齢 (年齢/70)	-0.86	0.42	-2.03	0.04	**	(-1.69,-0.03)
性別(男性-1、女性1)	0.34	0.09	3.92	<0.001	***	(0.17, 0.51)
サンプルサイズ	53782					
乖離度	18299.0					

注)***:1%水準で有意 **:5%水準で有意 *:10%水準で有意

- 1Gy被曝すると固形ガン死する方の割合が42%増加する。
- 被曝時年齢が高いとリスクは低下。
- 到達年齢が高いと被曝による(過剰相対)リスクは低下。
- 女性の方がリスクは高い

27

参考)各種の要因を考慮した定式化
過剰相対リスク(Excessive Relative Risk)モデル

固形ガンで亡くなった方の割合
=被曝していない人の固形ガン死亡率×(1+被曝者の特性× β_1 被曝量)

■ 意味

- 被曝によって、上式の下線部の分だけ死亡率が高まる(過剰相対リスク)
- 参考 前年売上 1.3倍
- 今年の売上=前年売上×(1+0.3)
- 1+0.3が相対リスク 0.3を過剰相対リスク と定義
- 固形ガンで亡くなった方の割合は被曝量に比例する。
- ただし、直線の傾き β_1 は性別、被曝時年齢、到達年齢によって異なる。

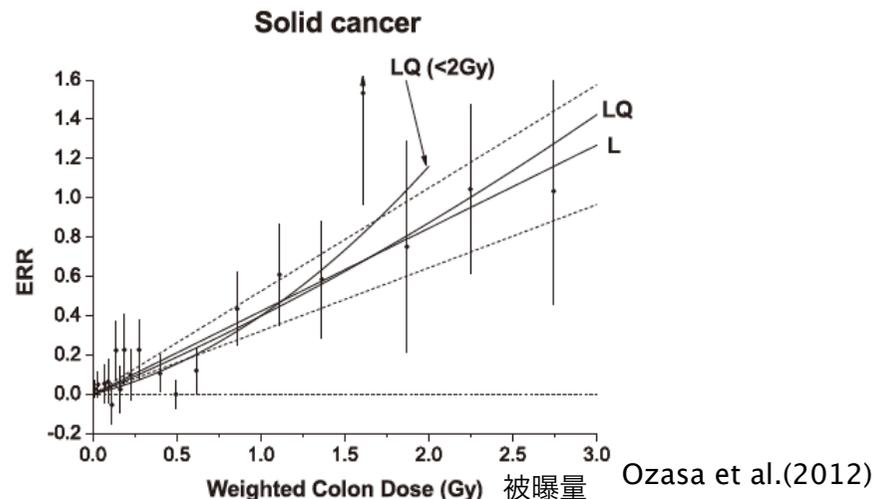
被曝者の特性

$(1 + \beta_2 \text{性別}) \times \exp\{\beta_3 \text{被曝時年齢} + \beta_4 \log(\text{到達年齢})\}$



26

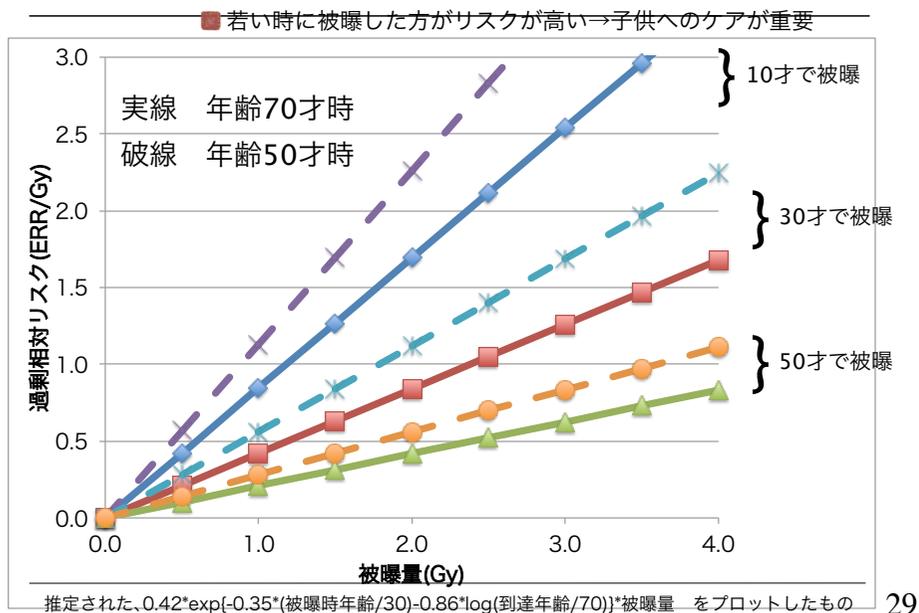
LSS14 (Ozasa et al. 2012)での推定結果(固形癌死)



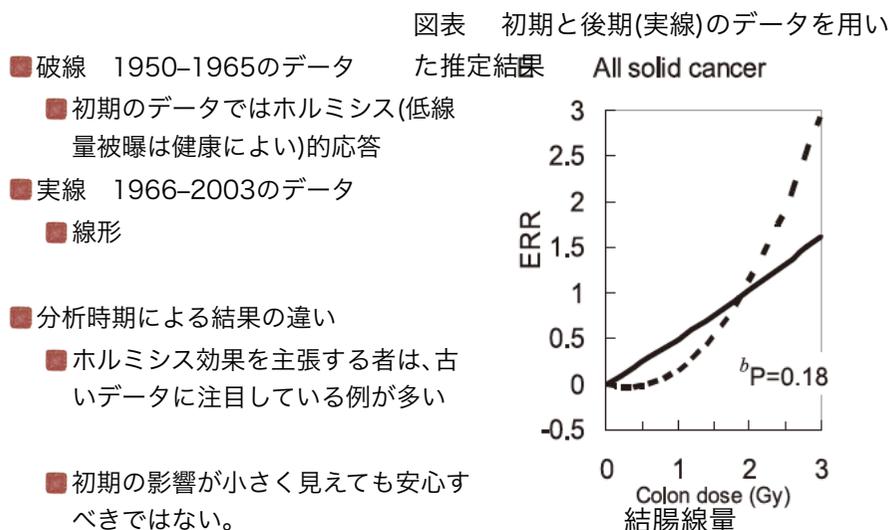
- 線形、1次+2次、線量ダミーモデルについてプロットしたもの。
- 30才で被曝、70才でのリスク

28

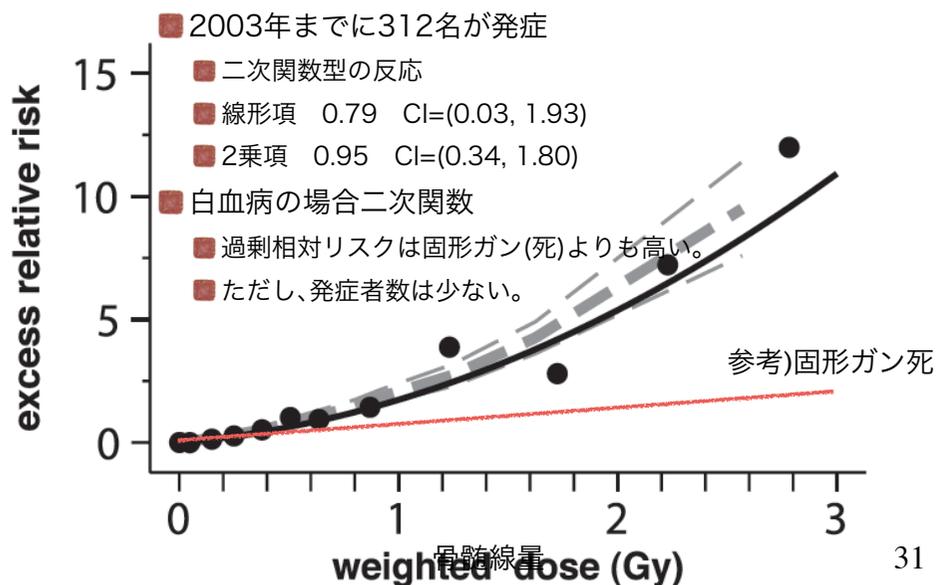
被曝時年齢、到達年齢別の固形ガン死の過剰相対リスク(男女平均)



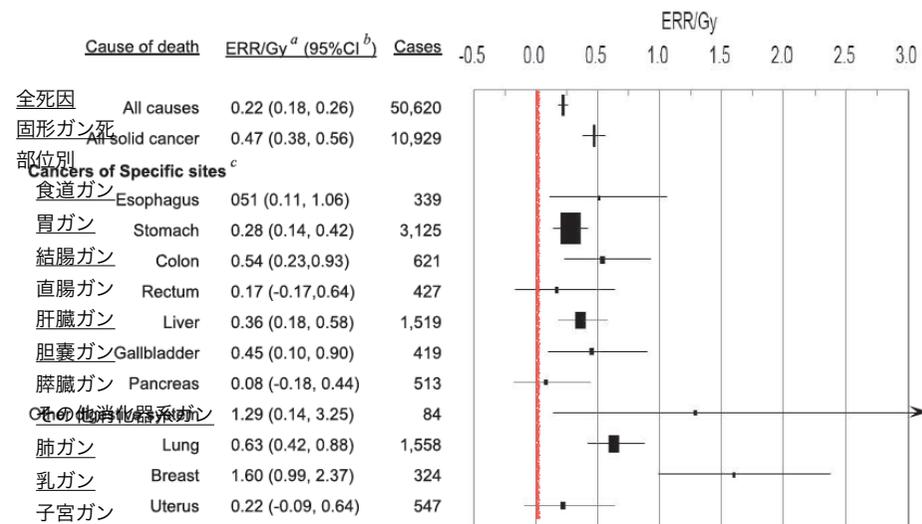
分析対象期間による結果の差異



その他のガン:白血病の発症(Hsu et al. 2013)



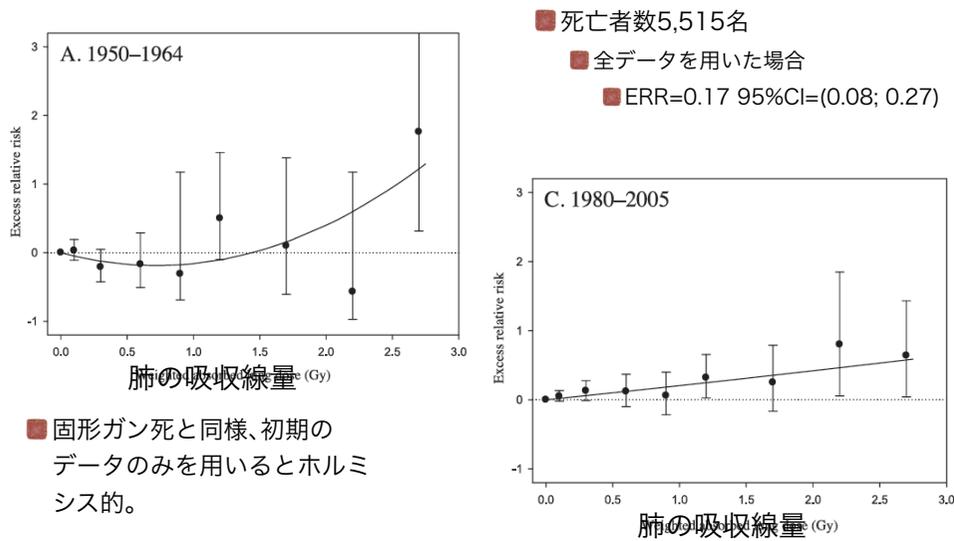
図表 様々な固形ガン死について推定した結果(95%信頼区間を图示)



注)固形ガン死のERRが0.42ではないのは、被曝者の特性による影響の変化は考慮していないため。

ガン以外の影響の分析

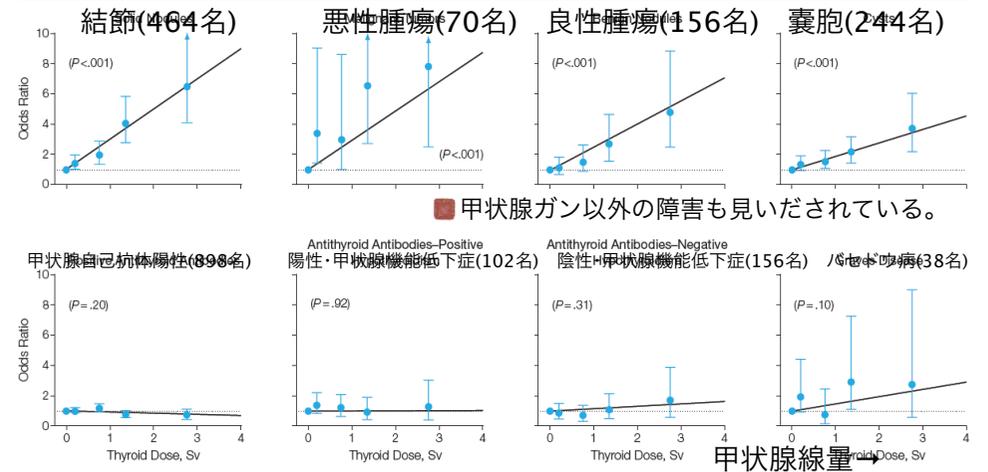
図表 呼吸器系疾病による死亡:分析対象時期を変更(Pham et al. 2013)



33

甲状腺障害 (Imaizumi et al. 2006)

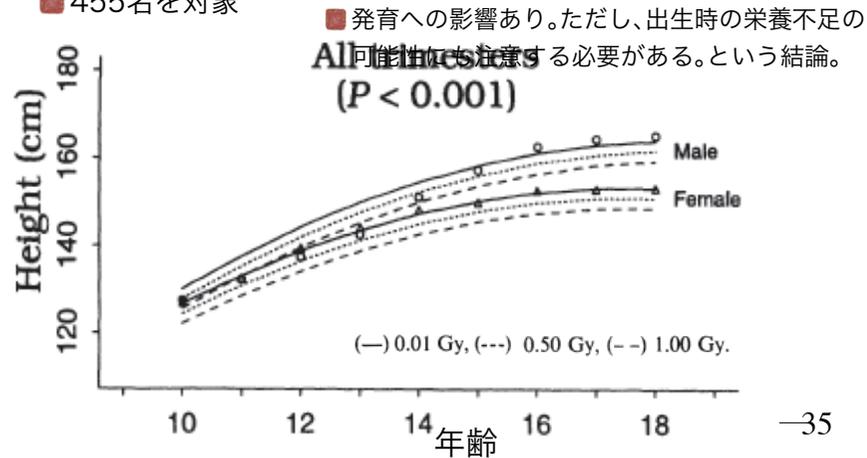
■ 被爆者・健康調査対象者のうち該当年に検査を受け、被曝量の判明している3,185名を対象(平均年齢70才)。



胎児の被曝:成長への影響

■ 年齢と身長 (Otake et al. 1993)

- 性別、被曝量別
- 455名を対象

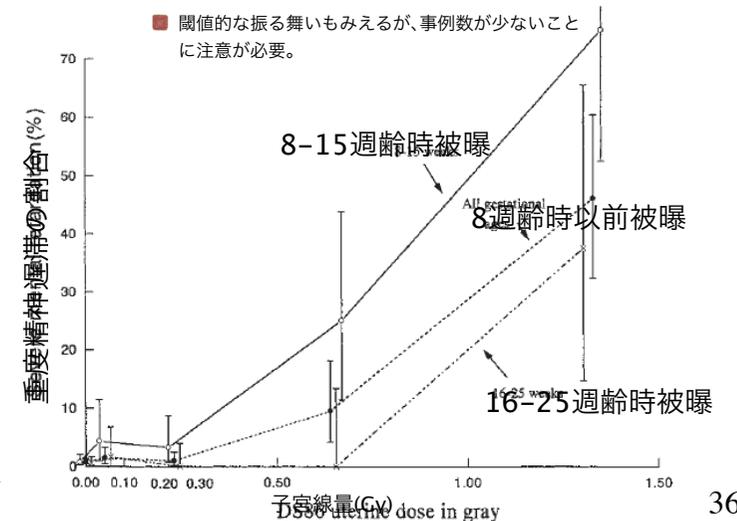


35

胎児の被曝:重度精神遅滞 (Otake 1998)

■ 1,554名中、重度精神遅滞30名(簡単な計算ができない等)。被曝時期別にプロット。

- 細胞分裂が盛んな時期の被曝は影響が大きい
- 閾値的な振る舞いもみえるが、事例数が少ないことに注意が必要。

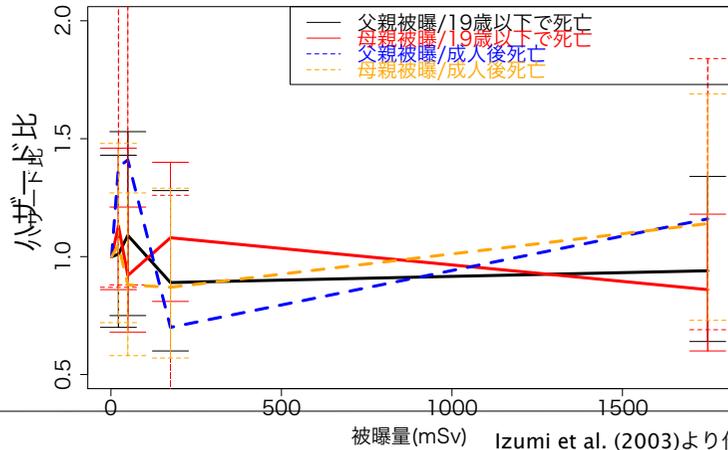


36

被爆者の子供

ガン以外の死亡(Izumi et al. 2003)

- 被爆者の子供41,010名への調査
- ガン以外で死亡した1,125名について父親、母親の被曝量との関連を分析。
 - ハザード比=各被曝量グループでの死亡率/被曝量がもっとも低い0-4mSvの方の死亡率



Izumi et al. (2003)より作成。37

LSS分析の問題点

- LSS13報 (Preston et al., 2013)、14報(Ozasa et al. 2012)の批判的にレビューによって課題が見いだされた(濱岡2015)。それらのうち、3点(赤字)について紹介する。

	問題点		LSS 13 1950-1997 (Preston et al.)	LSS 14 1950-2003 (Ozasa et al.)
データ処理	個人レベルデータの集計	統計的検定力の低下	✓	✓
モデルの定式化	LQモデルにおける多重共線性	不安定な推定結果	✓	✓
	探索的な閾値推定モデル。	閾値の統計的有意性を検定できない	✓	✓
モデルの推定	低線量域にサンプルを限定した推定。		✓	
	線量2Gy以下に限定したL,Q,LQモデルの推定	統計的検定力の低下		✓
	広島、長崎をまとめて分析	両都市の差異を無視	✓	✓
モデル選択	モデルの診断に必要な情報が示されていない。	モデル診断が不可能に。	✓	✓
	不完全なモデル選択	混乱した結果。最適なモデルが選択されない。	✓	✓

内容

- はじめに
 - 放射線の健康影響を巡る混乱?
 - 本日の講演の目的
- 統計学&データ分析入門
- 原爆被曝者データ分析
- 原爆被曝者データ分析の課題
- まとめ
- 参考文献

問題点1) 個人レベルデータを集計することによる情報損失、検定力低下

- LSSでは、個人レベルのデータを線量、性別、都市、年齢などによって層別集計。
 - 線量については: 22 区間
 - (0.005, 0.02, 0.04, 0.06, 0.08, 0.1, 0.125, 0.15, 0.175, 0.2, 0.25, 0.3, 0.5, 0.75, 1.0, 1.25, 1.5, 1.75, 2.0, 2.5, 3Gy +)
- 二つの問題
 - 区間設定の任意性 (区間の数と範囲)
 - LSS 14データの場合、86,611名中、38,509名が最小の線量区分(~0.005Gy、つまり~5mGy)に分類されている。
 - 個人データの集計による情報損失と検定力低下

データの集計による情報損失の例

	Data	Variance
Raw data	1,2,3,4,5,6,7,8,9,10	Var(x)=9.17
Categorized data	1~5 x 5 samples 6~10 x 5 samples	Var(x)=6.94

集計による影響を確認する

- LSS 14では線量を22区分
 - (-0.005, -0.02, -0.04, -0.06, -0.08, -0.1, -0.125, -0.15, -0.175, -0.2, -0.25, -0.3, -0.5, -0.75, -1.0, -1.25, -1.5, -1.75, -2.0, -2.5, -3.0, 3.0Gy~)
 - サンプルサイズ: 53782
- これを12,6区分にさらに集計。
 - 12区分 (-0.02, -0.06, -0.1, -0.15, -0.2, -0.3, -0.75, -1.25, -1.75, -2.5, 2.5Gy~)
 - サンプルサイズ: 33973
 - 6区分 (-0.06, -0.15, -0.3, -1.25, -2.5, 2.5Gy~)
 - サンプルサイズ: 22257
- これらに対して二つのモデルを推定。
 - 線形モデル $\beta_1 d$
 - (統計的に)閾値推定モデル

0	$(d < \tau)$
$\beta_2 (d - \tau)$	$(d \geq \tau)$
- 推定
 - RのOptim で尤度を最大化

41

問題点2) 低線量に限定した推定
(LSS13報 Preston et al.2003)

- 線形モデルについて低線量域にサンプルを限定して推定。125 mSvまで用いると有意となる
 - Q このようにサンプルを限定することは妥当か?

線量範囲 (Sv)	ERR/Sv	t値	P値	
0-0.05	0.93	1.09	0.150	
0-0.1	0.64	1.16	0.300	
0-0.125	0.74	1.95	0.025	**
0-0.15	0.56	1.75	0.045	**
0-0.2	0.76	2.62	0.003	***
0-0.5	0.44	3.67	<0.001	***
0-1	0.47	4.70	<0.001	***
0-2	0.54	7.71	<0.001	***
0-4	0.47	9.40	<0.001	***

注)P値は片側検定。出所) LSS13報Preston et al.(2003)のTable 4に加筆

推定結果

a) 線形モデル

	22区分			11区分			6区分		
	推定値	標準誤差	t値	推定値	標準誤差	t値	推定値	標準誤差	t値
線量 : ERR (/Gy)	0.413	0.051	8.07 **	0.408	0.052	7.84 **	0.391	0.053	7.34 **
性別	0.340	0.088	3.88 **	0.331	0.089	3.72 **	0.340	0.092	3.70 **
到達年齢	-0.334	0.084	-4.00 **	-0.347	0.086	-4.04 **	-0.364	0.092	-3.97 **
到達年齢	-0.949	0.382	-2.49 **	-0.878	0.390	-2.25 **	-0.823	0.407	-2.02 **
N	53782			33973			22257		
AIC	33285			26520			21115		
BIC	33760			26973			21548		

b) 閾値推定モデル

	22区分			11区分			6区分		
	推定値	標準誤差	t値	推定値	標準誤差	t値	推定値	標準誤差	t値
線量 : ERR (/Gy)	0.417	0.071	5.86 **	0.408	0.074	5.55 **	0.385	0.073	5.25 **
閾値	-0.023	0.264	-0.09	0.003	0.304	0.01	0.037	0.356	0.10
性別	0.345	0.105	3.29 **	0.330	0.108	3.07 **	0.332	0.114	2.91 **
到達年齢	-0.338	0.096	-3.53 **	-0.346	0.100	-3.46 **	-0.358	0.107	-3.34 **
到達年齢	-0.985	0.562	-1.75 *	-0.874	0.577	-1.52	-0.774	0.619	-1.25
N	53782			33973			22257		
AIC	33287			26522			21117		
BIC	33782			26994			21568		

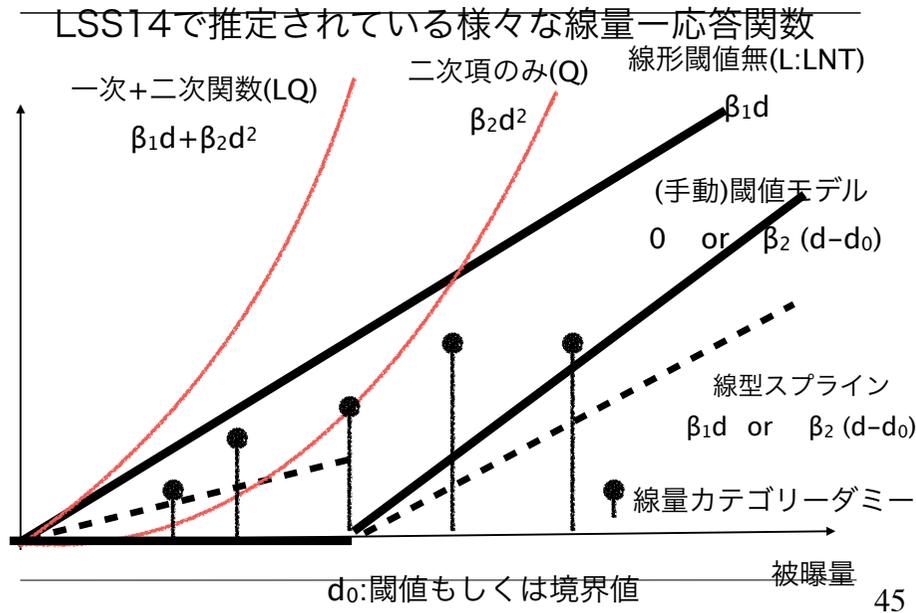
42

UNSCEAR (United Nations Scientific Committee on the Effects of Atomic Radiation)も同様の不適切な分析

- UNSCEAR2006レポート
 - 線量推定における誤差を考慮したモデルで同様の分析。固形癌死については200mSv、固形癌発症については250mSvまでのサンプルを用いると有意になる。
- UNSCEAR2011サマリーレポート
 - "Statistically significant elevations in risk are observed at doses of 100 to 200 mGy and above. UNSCEAR(2011, parag. 25)"

44

問題点3)不完全なモデル選択



モデル選択をしていないため混乱した記述

- 通常は複数のモデルを推定し、それらの適合度を比較して最良のモデルを決定する。LSST4では、それが行われていない。LSST4 (Ozasa et al.2012) の要約
- 線形モデルの固形癌のERR(30歳で被曝、70歳に到達時。男女平均。)は 0.42 /Gy [95% 信頼区間: 0.32, 0.53]であった。線形モデルを支持?
- (線形スプライン関数によると) 閾値0.2Gyを示唆?
- ERRが有意になるのは、0 から0.20 Gyまでを用いた場合であった。フォーマルな線量、閾値分析によると閾値ゼロが示された。つまり、閾値ゼロが最良の推定値であった。
- (赤字、下線は濱岡による補記) 線形モデルを支持?

LSST4 (Ozasa et al. 2012)での推定(線形スプライン・モデル)

- 低線量サンプルに限定する不適切な分析 (LSST3)にかわって、全サンプルを使って「線形スプライン・モデル」を推定
 - $\beta_1 d$ ($d < d_0$)
 - $\beta_2 (d - d_0)$ ($d \geq d_0$)
- 境界値 d_0 の上下で直線の傾きが異なる ($d_0=20, 30, \dots$ mGyとして推定)。
- 境界値 d_0 よりも低線量側の傾きの推定値と95%信頼区間をプロット。
- 有意となる(低線量側の係数=0という仮説が棄却される)のは、0 ~0.20 Gyまでのサンプルを用いた場合。
- この結果に注目して閾値モデルを主張する者も存在。
- ただし、線形モデルとの比較はされていない。

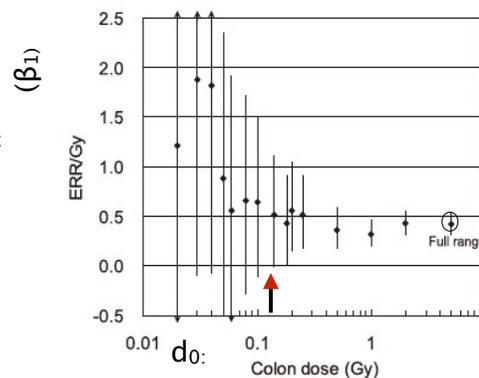


FIG. 5. Excess relative risk per Gy (ERR/Gy) for all solid cancer for selected dose ranges. The figure shows the ERR/Gy and 95% CI for a dose range from zero to a given dose based on the linear model for the full data that allowed for different ERRs below and above the given dose and taking radiation effect modifiers as common to the two dose ranges. The increased ERR/Gy in the low-dose levels less than 0.1 Gy corresponds to the estimates of ERR higher than the expected linear line in Fig. 4.

LSST4 (Ozasa et al. 2012)での推定結果 (固形ガン死):L、Q、LQモデルの比較

TABLE 6
Parameter Estimates of the Dose-Response Models for Excess Relative Risk (ERR) for all Solid Cancer in the Full Dose Range and for the Range of 0-2 Gy

Dose range model ^a	Full			<2 Gy		
	L ^b	LQ	Q	L	LQ	Q
β_1 : linear	0.42	0.36	-	0.44	0.22	-
β_2 : quadratic	-	0.038	0.22	-	0.18	0.33
Effect modification						
σ : sex (female = 1; male = -1)	0.34	0.35	0.40	0.28	0.29	0.29
τ : age at exposure (year)	-0.035	-0.034	-0.035	-0.033	-0.034	-0.035
ν : attained age (log(age/70))	-0.86	-0.86	-0.90	-0.84	-0.89	-0.97
Deviance	18301.2	18300.4	18324.9	17557.3	17551.6	17557.2
df	53147	53146	53147	49577	49576	49577
Test (vs. LQ model)	$P = 0.36$	-	$P < 0.001$	$P = 0.02$	-	$P = 0.02$

Note. Bolded columns are the selected models.

^a The ERR model was defined as $\lambda_0(c,s,b,a) [1 + \rho(d) \cdot \exp(\tau e + \nu \ln(a)) \cdot (1 + \sigma s)]$, where d is colon dose, s is sex, b is birth year, e is age at exposure, and a is attained age. $\rho(d)$ was $\beta_1 d$ for the linear model, $\beta_1 d + \beta_2 d^2$ for the linear-quadratic model, and γd^2 for the quadratic model. τ , ν and σ are coefficients for effect modification.

^b L: linear, LQ: linear-quadratic, Q: quadratic.

LSS14で推定された線量-応答関数とモデル適合度評価法

LSS14 (Ozasa et al. 2012)

- Model 1 L:線形 $\beta_1 d$
- Model 2 1次+2次(LQ) $\beta_1 d + \beta_2 d^2$
- Model 3 2次(Q) $\beta_2 d^2$
- Model 1-3 は全サンプルだけでなく、2Gy未満に限定して推定。
- Model 4 (Manual search) Threshold ($d_0=10,20,30\text{Mg},,$)
 0 ($d < d_0$)
 $\beta_2 (d-d')$ ($d \geq d_0$)
- Model 5 Dose category dummy
 ■ 15 categories
- Model 6 (Manual search) Linear spline (L1L2)($d_0=10,20,30\text{Mg},,$)
 $\beta_1 d$ ($d < d_0$)
 $\beta_2 (d-d')$ ($d \geq d_0$)
- Model 7 2 Gyで屈曲
 L1, L1Q1, or Q1 ($d < 2\text{Gy}$)
 L2, L2Q2, or Q2 ($d \geq 2\text{Gy}$)
- Model 8 (統計的に) 閾値推定
 0 ($d < \tau$)
 $\beta_2 (d-\tau)$ ($d \geq \tau$)

尤度比検定

最大尤度

最大尤度



本研究で追加

- Model 7 2 Gyで屈曲

モデルの推定結果と適合度の比較

モデル	推定値と有意水準								適合度					
	閾値,境界値		L1		Q1		LもしくはL2		QもしくはQ2		LL	AIC	BIC	
	推定値	有意水準	推定値	有意水準	推定値	有意水準	推定値	有意水準	推定値	有意水準				
1 線形	L						0.413	***			-16618.5	33285.0	33759.8	
2 線形+二次	LQ						0.342	***	0.049		-16617.9	33285.8	33780.5	
3 二次	Q						0.238	***			-16629.3	33306.7	33781.5	
4 (手動)閾値	0+L2	1	f	0			0.399	***			-16636.1	33320.2	33795.1	
	0+L2	5	f	0			0.398	***			-16636.1	33320.2	33795.1	
	0+L2	10	f	0			0.398	***			-16636.1	33320.3	33795.1	
	0+L2	20	f	0			0.401	***			-16636.1	33320.1	33794.9	
	0+L2	50	f	0			0.403	***			-16635.4	33318.8	33793.6	
	0+L2	100	f	0			0.411	***			-16632.5	33313.0	33787.8	
	0+L2	125	f				0.398	***			-16631.9	33311.9	33786.7	
	0+L2	200	f				0.422	***			-16622.9	33293.8	33768.7	
5 線量カテゴリダミー変数											推定不能			
6 線型スプライン	L1+L2	1	f	19.341			0.416	***			-16618.4	33286.8	33781.5	
	L1+L2	5	f								推定不能			
	L1+L2	10	f	-2.252			0.411	***			-16618.4	33286.7	33781.4	
	L1+L2	20	f	0.979			0.416	***			-16618.4	33286.9	33781.5	
	L1+L2	50	f	0.905	*		0.418	***			-16618.2	33286.4	33781.1	
	L1+L2	100	f	0.667	*		0.417	***			-16618.3	33286.6	33781.2	
	L1+L2	125	f	0.154			0.412	***			-16618.1	33286.3	33780.9	
7 2Gyで屈折	L1+L2	200	f	0.548	***		0.417	***			-16618.3	33286.6	33781.2	
	L1+L2	2000	f	0.421	***		0.395	***			-16618.4	33286.8	33781.5	
	L1Q1+L2Q2	2000	f	0.194	**	0.207	***	0.622		-0.100	-16615.1	33284.2	33818.4	
	L1Q1+L2	2000	f	0.658	*	0.666	*	1.409	*		-16627.1	33306.3	33820.7	
8 閾値推定	Q1+Q2	2000	f			0.361	***			0.150	***	-16618.0	33286.1	33780.7
		-23.160					0.417	***			-16618.3	33286.9	33781.6	

注)***:1%水準で有意 **:5%水準で有意 *:10%水準で有意 f:固定パラメータ

再分析

- データ
 - LSS 14 の固形癌死(全サンプル)
- 推定
 - 統計ソフトRのoptimで推定。
- モデル選択
 - 情報量基準
 - 尤度比検定は、包含されたモデル間でしか比較できないが、情報量基準は、包含されていないモデルでも比較可能
 - 赤池の情報量基準
 - AIC=乖離度+ 2k
 - ベイジアン情報量基準
 - BIC =乖離度+ k*log(N)
 - ここで、
 - k: 推定されるパラメータの数
 - N: サンプルサイズ:LSS14では約50,000なのでlog(N)-10
 - これらの値が小さい方が少ないパラメーターでデータとの乖離が小さい、よいモデルとされる。
 - BICの方がより厳しい基準

内容

- はじめに
 - 放射線の健康影響を巡る混乱?
 - 本日の講演の目的
- 統計学&データ分析入門
- 原爆被曝者データ分析
- 原爆被曝者データ分析の課題
- 100mSv問題の原因?
- 参考文献

「100mSv以下では有意ではない」の問題

■ 分析上の問題

- 原爆被爆者データに関しては、まともに分析すれば、閾値無し線形モデルが妥当であるはずだが、それがなされていない。

■ 解釈、政策上の問題

- 有意ではない=影響がないと解釈している？
- 本当に100mSv以下で影響があるかないか不確実ならば、影響があることを想定した対応を行うべきだが、それはされていない。

53

「100mSv以下では有意ではない」を巡る混乱の原因

■ 分析対象

- ばらつきが大きいデータ

■ 分析方法とその結果

- 分析時期による結果の違い
- 部位による分析結果の違い
- 交絡因子測定の有無
- 研究による結果の違い(特に長期低線量被曝)

■ 統計学の知識、分析レベル

- 不適切な分析
 - 個別データの集計による情報損失
 - 複数の線量-応答関数の中でどれが最良かが分析されていない。
 - 低線量域に限定するとサンプルサイズが減少して検定力が低下

■ 分析手法の問題

■ 研究アプローチ

- 異なるアプローチ(疫学、動物実験、分子レベル)で異なる知見

■ 制度など

- 個別研究→知見とりまとめ→施策反映の時間的ラグ
- 人によって注目する点異なる
- 新しい個別研究 or 古い情報がそれなりに統一された情報など

■ 専門家の言動による混乱や不信の発生

- 単純な誤り?
- 知識のつまみ食い・偏った紹介

■ メディアの報道

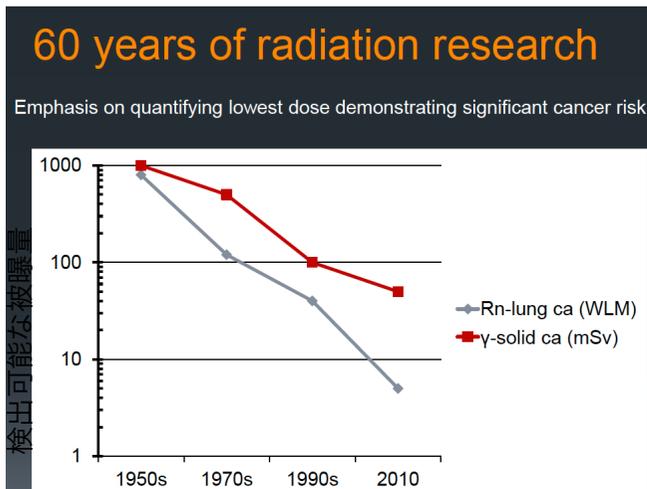
- 統計的な知識不足に基づく誤解?

■ 国などの偏った施策

- 体系的な勧告からのつまみ食い

54

図 統計的にガンリスクが有意となる線量の推移



■ 対象者の増加、対象期間の長期化、分析手法の進化などによって、リスクを検出できる線量は低下。

■ これから見ても低線量被曝の影響があることを前提に行動すべきである。

55

おわりに

■ 本講演の目的

- データ分析の基本的を理解する。
 - 回帰分析
 - 有意とは
- 放射線疫学の代表的な研究である広島・長崎の被爆者データ分析の方法を理解する。
- その問題点を理解する。
 - 3つの問題点
 - データの集計
 - 線量・サンプルを限定
 - モデル選択せず
- 100mSvを巡る混乱の原因を考察する。

■ どうすればよいか?

- 他分野の(統計やデータ分析)専門家による批判的レビュー、再分析のためのデータ提供
- 市民も批判的に文献を読む(放射線疫学に関しては、英語+統計の基礎程度わかればok)
- 「データ分析の基本的を理解する。」を紹介したのもそのため。
- 再分析を可能とする
 - (匿名化個票)データ公開

56

主要参照文献

- Cardis, Elisabeth, Howe, Geoffrey, Ron, Elaine, Bebeshko, Vladimir, Bogdanova, Tetyana, Bouville, Andre, . . . Zvonova, Irina. (2006). Cancer consequences of the Chernobyl accident: 20 years on. *Journal of Radiological Protection*, 26(2), 127-140. doi: 10.1088/0952-4746/26/2/001
- Grant, E. J., A. Brenner, H. Sugiyama, R. Sakata, A. Sadakane, M. Utada, E. K. Cahoon, C. M. Milder, M. Soda, H. M. Cullings, D. L. Preston, K. Mabuchi, and K. Ozasa (2017), "Solid Cancer Incidence among the Life Span Study of Atomic Bomb Survivors: 1958-2009," *Radiat Res.*
- Ozasa, Kotaro, Yukiko Shimizu, Akihiko Suyama, Fumiyoshi Kasagi, Midori Soda, Eric J. Grant, Ritsu Sakata, Hiromi Sugiyama, and Kazunori Kodama (2012), "Studies of the Mortality of Atomic Bomb Survivors, Report 14, 1950-2003: An Overview of Cancer and Noncancer Diseases," *Radiation Research*, 177 (3), 229-43.
- Preston, DL, Y Shimizu, DA Pierce, A Suyama, and K Mabuchi (2003), "Studies of Mortality of Atomic Bomb Survivors. Report 13. Solid Cancer and Noncancer Disease Mortality: 1950-1997," *Radiation Research*, 160 (4), 381-407.
- Schubauer-Berigan, M. K., & Daniels, R. D. 2013. Risk of cancer from exposure to 100 mSv within 5 years: What can we learn from epidemiology? Paper presented at the MELODI Workshop, Brussels: Belgium.
- 濱岡 豊(2012) LSS(Life Span Study) 13報 被爆者データのRによる分析 <http://nonuke2011.blogspot.jp/2012/05/lsslife-span-study-13-preston-dl-y.html>
- 濱岡豊 (2015a), "福島甲状腺データの分析," 岩波書店, 科学 (7月号)
- 濱岡豊 (2015b), "広島・長崎被曝者データの再分析," 岩波書店, 科学 (9月号).
- 濱岡豊 (2015c), "長期低線量被曝からの知見と課題," 岩波書店, 科学 (10月号).