

低線量放射線被曝によるがん発生に関する元データの誤差論に基づく考察

深田智*

(平成23年10月31日 受理)

Discussion of original data on cancer occurrences by exposure to low-level radiation based on error theory

Satoshi FUKADA

†E-mail of corresponding author: sfukada@nucl.kyushu-u.ac.jp

Hard controversy has been repeated on how large the number of solid cancers or leukemia as typical late effects is enhanced by exposure to low-level radiation or not. The most reliable data to answer it are published as Life Span Study (LSS) from Radiation Effects Research Foundation (RERF). However, the RERF report says that the Liner No-Threshold (LNT) model is applicable to the late radiation effects, because there is no reason apparently to deny it. In the present study, the error theory is introduced into the data of the RERF report, and quantitative discussion based on the error theory is made on how well the LNT model is applicable to the late effects in the low exposure range. So that, it is found that the LNT model is extrapolated also to the lower range of 0.1Gy-0.01Gy for occurrences of solid cancers and leukemia.

Key words: *Exposure to radiation, Cancer occurrences, Liner-no-threshold model, Error theory*

1. 緒 言

放射線被曝とがん発症との間に因果関係があるのは共通の認識である。因果関係の原因にはここで立ち入らず、定量性、言い砕けば、高い低いに関わらずいくらか放射線量を受けるとがん発症率はどれくらい増加するか、を問題にする。放射線の人体への影響で特に急性影響（白血球の減少、一時的あるいは永久不妊、白内障、脱毛等）について線量効果曲線が過去に求められ、放射線しきい値があることが広く認められており、しきい値も各障害毎に与えられている¹⁾。しかし被曝後年月が経過した後に現れる晩発影響に関しては過去多くの論争があり^{2,3)}、いまだに議論が収束していない。放射線被曝量とがん発症率（あるいは死亡率）との関係を定量的に明らかにし、放射線被曝の危険性、さらに原子力利用の効用と逆効果を見極め、他の多くの原因によるがん発症率あるいは死亡率⁴⁾と比較した上で放射線の安全利用を進めて行く必要がある。

ここで焦点を絞り、低線量放射線被曝によるがん発症確率が高くなるのか、ならないのかについて定量的に議論する。国際放射線防護委員会 (ICRP) がとる基本的考えは、線形しきい値なし (LNT) モデル⁵⁾であ

る。この考えに従い、放射線によるがん死亡リスクが、0.05/SvであることがICRPにより報告されている。その一方、特に0.1Gy以下の低線量域において、放射線影響にしきい値があるとするモデル²⁾、放射線ホルミシスを考えるモデル³⁾、あるいは逆に低線量で影響が増加する二相(Biphasic)モデル⁶⁾等がある。それぞれの説を読むとそれなりの言い分があり、強い弱いがあるにしても一見実験的あるいは経験的裏付けもある様に見える。しかし問題は、それぞれの説で示している事例が、ある集団には適用できても、別の集団には適用できないという事に気づく。例えば、Robert Ehrlichが述べている様に⁷⁾、長崎の男性被曝者の1970-1988年期間のがん以外の死亡率が0.5Gy-1Gyの間で有意に低いという論拠について、そのデータは確かに存在したとしても、なぜ広島では起こらず長崎だけなのか、なぜその時期の男性だけを選択するのかについての疑問がまったく晴れない。統計的に見て、無作為抽出した標本の集合が複数あるとき、各集合の平均が常に一定値を与える訳でなく、ある確率分布に従って現れるのは周知の事実で、特定の標本群だけで結果を判断する事はできない。これが科学である。

*1 先端エネルギー理工学専攻

放射線が発見されてから100年以上経過し、さまざまな放射線の利用が図られてきた一方、その負の面の合理的共通認識を得ることはきわめて重要なことと思われる。

後に詳しく説明するが、放射線障害を定量的に証明する上で難しいのは、まず放射線にはいろいろ種類があること、次いでがん発症は晩発障害であり、他にいろいろな原因で発症する慢性障害と症状的に同じで区別できないこと、また放射線による障害は特に低線量域では小さいので大規模な調査が必要であることにある。例えば、5%の有意水準で放射線障害を証明するには、0.1Gy (=100mGy)の被曝線量で必要な集団サイズが6390人、10mGyになると0.1%の過剰リスクを証明するには62万人必要となることが例に挙げられている⁹⁾。このような大規模な放射線障害を証明するのは、以下に示すように放射線影響研究所 (RERF) が長年にわたって実施している原爆被害者の寿命調査研究 (LSS)⁸⁾ しかないと思われる。この総理工報告では特に低線量被曝の影響を調べるため、報告されているデータを一吟味し、放射線障害についての定量的議論をおこない、放射線障害がどれだけ現れるのかを自分の目で調べたいと言うのが本研究の動機である。次章で放射線障害における低線量域のデータを調べるが、RERFの実施したLSSオリジナルデータを参考のため付録1に示す。

2. 広島・長崎における原爆放射線による被曝の晩発影響

2.1 線量-がん発生率・死亡率の関係

放射線影響の定量評価のため、RERFが広島と長崎における原爆被曝者のLSSを広くおこない、終戦後、長い年月が経過した後のがん発生率を要覧に公表している⁸⁾。付録1に彼らが報告している相対リスクデータを再掲しているが、図を見て分かる様に、特に0.1Gyの低線量域で、相対リスクが1にほぼ近く、この図では被曝が無いときとあるときでほとんど差が見られない。一応線形モデル等のフィッティングカーブも与えられているが、1.0Gy付近から特に0.1Gy以下までの低線量放射線被曝の影響の論拠として弱い様に見える。問題は、この領域の定量化であるが、序論でも述べた様に、この領域では特に多数のデータを必要とする。しかし、要覧では、男女の区別、年齢等により集団を分け、それぞれについてデータをプロットしており、そのため母集団標本数が少なくなり、誤差が広がっている。そこで、本研究では、男女の区別、年齢等の区分を一度取り外し、すべて被曝線量とがん発症率あるいは死亡率の結果として一元化して再プロットし、低線量被曝の影響を判断する事とした。例えば、このような操作をおこなったとき、二つの集団を

同一と見なして、データ解析する事になる。もともと二つの集団間に少し異なった統計データがあるときに、両者を混ぜて一つの集団としたときに現れる誤差の評価法について、付録2と付録3に示している。

Fig. 1は、2009年に発表された最新の要覧に挙げられたデータに基づいてプロットしたものである。被曝放射線量は、DS(02)としてまとめられた。丸印が報告された数値であり、実線や破線はここでの評価式である。先のDS(86)から見ると予測被曝線量は8%増加し、その結果放射線リスクはこれまでより少し下がっている。今回の線量評価値に基づいて、ICRPは福島第一原子力発電所事故対応のため、累積被曝線量の許容値の改訂を日本政府に提案した経緯がある。

この節では、LSS母集団(合計120,321人)のうちで、原爆被曝者の受けた放射線量と晩発性がん発症率との関係をRERF報告書の数値に基づいて再吟味する。Figs. 1, 2の横軸は、吸収線量(単位Gy (J/kg))である。Gyは単位から分かる様に、放射線被曝により体内1kgに吸収されたエネルギー量を表しており、体内に吸収されたγ線や中性子線が電離イオン化を起こし、細胞や遺伝子DNAに影響を及ぼす程度を定量的に表す数値である。この値を計算するには、γ線量はそのまま加え、核分裂発生中性子線については昔から取り入れられている生物学的効果の10倍を乗じた線量を加えることで評価する。かつてはこれを線量等量(単位Sv)と呼んで来た。現在は、線量等量を部位ごとにきめ細かく算定する様になっており、特定の組織の影響を調べるための組織加重係数を乗じた値をSv単位で表している。混乱を避けるため、要覧でも使用されている単位としてGyを踏襲した。基本的には、今回評価して

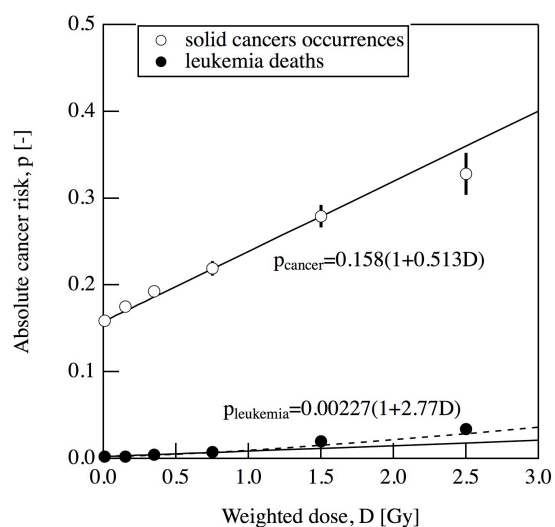


Fig. 1 Relation between weighted radiation dose and cancer occurrences or deaths in linear scale. The line is based on the LNT model

いるのは全身被曝であり、概要で使用されているGyをSvと読み替えてよいと思われる。

一つ注意すべき点は、 γ 線量は外部被曝のものであり、また中性子線についても考慮されているのは原爆から直接放射されたいわゆる即発中性子の寄与と体内で放射化された組織からの被曝の寄与であるという点である。核爆弾で発生した核分裂片からの遅発中性子の寄与や環境の残留放射能（特に β 崩壊の寄与）を呼吸等によって取り込む量がどの程度あるかをあらかじめ確認する必要がある。要覧でもその点は述べられており、爆心地近くでの誘導放射能の寄与は、20%程度（絶対値では広島で0.16Gy、長崎で0.06Gy-0.08Gy）と評価されている。また原爆破裂とその後の爆心地付近の大規模な火災により、強い上昇気流が発生し、それが気流に乗り黒い雨となって被曝の拡大を招いた。黒い雨の成分が煤塵すすとしても、核分裂片もその中に含まれ、一部は皮膚からあるいは呼吸とともに体内に吸収され、そこから発生する遅発中性子と β 線が内部被曝に寄与したと考えられる。その寄与は、広島市西部の乙斐・高須地区で10~30mGy、長崎市東部の西山地区で0.2Gy-0.4Gyと評価されている。さらに原爆破裂後、残留放射能がまだ高い市内に入ったいわゆる入市被曝者の多くも、LSS母集団の健康調査に含まれており、要覧を読む限り、横軸のGyと縦軸のがんリスクから被曝の晩発影響を定量的に調べることができると考えられる。

Figs.1, 2の縦軸を一意に絶対ガンリスク p と記述している。その内容は、全対象者44,635人（ ΣN_i ）を個人ごとの被曝線量に応じて6段階の被曝線量区分に分け、各放射線被曝線量を受けた母数(N_i)中に1958年から1998年の間に白血病以外のすべての固形がんを発症した人の数（ λ_i ）の確率リスク値 $p_{cancer,i}$ （ $=\lambda_i/N_i$ ）で表したものの(○)と、同様に全対象者49,204人を同じ被曝線量区分の母数に分け、1950年から2000年の間の白血病による死亡者数を、同じくその母数で割った確率リスク値 $p_{leukemia,i}$ （●）で表している。この母数表記から分かる様に、男女の区別、年齢の差をすべて同じ母集団で判断している。また縦軸の切片の値は、それぞれ37,307人あるいは60,792人の原爆被曝量が無視できる対象者母数（ N_0 ）における固形がん発症率 $p_{cancer,control}$ （実数で表すと1万人当たり1580人）と白血病死亡率 $p_{leukemia,control}$ （実数で1万人当たり25人）である。図から分かる様に、被曝量が少なくなるにつれて、被曝量が無視できる切片値に漸近するとともに、逆に1Gy以上の高線量被曝者は、受けなかった人より明らかに固形がん発症率あるいは白血病死亡率が高い。要覧には各部位別のがん発症率も調べられており、子宮がんや前立線がんの発症率は固形がん全体（平均）より1/3程度低く、逆に膀胱がんや乳がんは1.5倍から2

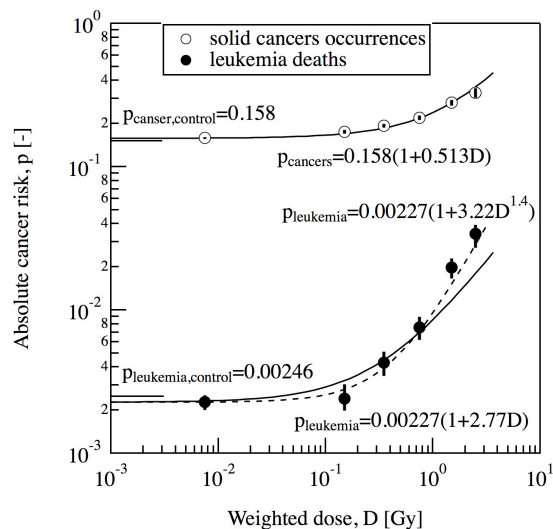


Fig. 2 Relation between weighted radiation dose and cancer occurrences or deaths in logarithmic scale. The line is based on the LNT model

倍程度高い。図の丸印値はすべての固形がん平均発症率である。1 Gy被曝すると、すべての固形がんを合わせて、被曝がないときより50%程度発症し易くなり、白血病死亡率は2.8倍高くなる。要覧では、固形がんの発症は47%増える（これは、過剰相対リスク, Excess Relative Risk, ERRである）としており、細かい数値の違いはともかくおおむね同じ内容と傾向である。

人間が地球上で生活していくとき、がんを発症する因子は、放射線被曝だけでなく他に遺伝的因子、生活習慣要因、ストレス因子、煙草や過度のアルコール等多くある。被曝の影響を線量-効果関係、特に低線量での影響を十分吟味する必要がある。そのため、Fig. 1の結果をFig. 2にある様に対数グラフに書き換え、特に低線量域における効果を見やすく描いている。まずがん発症数あるいは死亡数値の誤差を評価する。実験室で繰り返す事の出来る再現可能なデータではなく、各被曝線量条件で1回の実測数、1種類の標本数しか無いデータなので、標本平均や標本分散を統計操作に基づいて誤差予測できない。そこで数学的に減多に起こらない事象が理論的に二項分布、あるいはその極限のポアソン分布に従う事を利用して、一つ一つの実測統計データの誤差を評価する。各線量 D_i で例えば、固形がん発症率（ $p_{cancer,i}$ ）が決まっており、二項分布あるいはポアソン分布に従う発症数（ λ_i ）の最尤値が出現していると考えたら、各線量におけるがん発症実測数の標準偏差（ $\sqrt{\lambda_i}$ ）を求めることができる。それを固形がん発症率の誤差上限値と下限値（ $p_{cancer,i} \pm \sqrt{p_{cancer,i}/N_i}$ ）として誤差棒を用いてFigs. 1, 2に表している。白血病死亡率についても同様の操作を行っている。この誤差は、本来ランダム現象であるものが発現するときに含まれる理論的な誤差で

あり、それ以外の恣意的な外部要因で含まれる誤差は除かれる。問題は男女の差、あるいは年齢による偏差がもしあるとしたら、この理論的な誤差をどれだけ増すかを知る必要がある。付録3に、二項分布あるいはポアソン分布に従う2種類の分布が加えられて一つの分布として現れるときのリスクの計算式を示した。結果は、付録3の(A-19)式になる。もし、男女間に10%の差があるときには、全体の誤差は、10%程度増加する。このことも含めて付録図 Fig. A2に確率密度関数と標準偏差を評価している。LSSのデータは、極めて多数の母数 N_i から評価されており、ポアソン分布の可能性が高いと考えられる。評価した誤差棒から判断して低線量あるいは高線量被曝に関わらず広島と長崎の原爆被害者の各実測数 (λ_i) は、十分な精度があり、いずれの線量でも実測値の誤差は0.1Gyの線領域までの外挿できる事を示している。従って、放射線リスクを考える上で、広島と長崎の被害者のデータの集積は重要な示唆を与えると考えられる。

2.2 過剰ガンリスクと余剰ガンリスク

LSSデータを一元化し、絶対リスクを用いる事によって放射線の晩発影響の傾向が読み取れる事が分かった。そこで問題の原爆放射線吸収線量と、被曝による固形がん発症率 $p_{cancer,i}$ 、白血病死亡率 $p_{leukemia,i}$ の各値から被曝していないときの夫々の値 $p_{cancer,control}$ と $p_{leukemia,control}$ との差を求めた結果をFig. 3に示す。この縦軸の値は過剰絶対リスク (EAR, Excess Absolute Risk), R_k と呼ばれる。誤差の伝搬評価には次式を使っている。

$$R_k = p_{k,i} - p_{k,control} \pm \sqrt{p_{k,i}/N_{k,i} + p_{k,control}/N_0} \quad (1)$$

k=solid cancers or leukemia

固形がん発症率では直線関係が0.1Gyあるいはそれ以下の10mGy程度まで成立していると見てよい。一方、白血病死亡率では傾き1の直線近似では差が大きいものの0.1Gy程度まで差を認識できる。この傾きを線量Dの1.4乗で増加する式で整理すると、図の破線と白血病死亡率●との差はだいぶ小さくなる。この場合、直線近似は低線量域で過小評価になっておらず、ひとまずこれまでの先例に従ってしきい値なしの直線近似モデル (LNT) で話を進めるのが安全側に立つ姿勢である。なお詳しく見るとFig. 2で固形がん発症リスクは被曝量0に近づくにつれて、control値にうまく漸近するが、白血病ガン死亡リスクは、被曝量が0.1Gyのときは、原子爆弾による被曝量が0と見なせるときに比べて若干リスクが下がっている。これはホルミシスが原因ではなく、誤差から当然そのような場合もあり得る。相

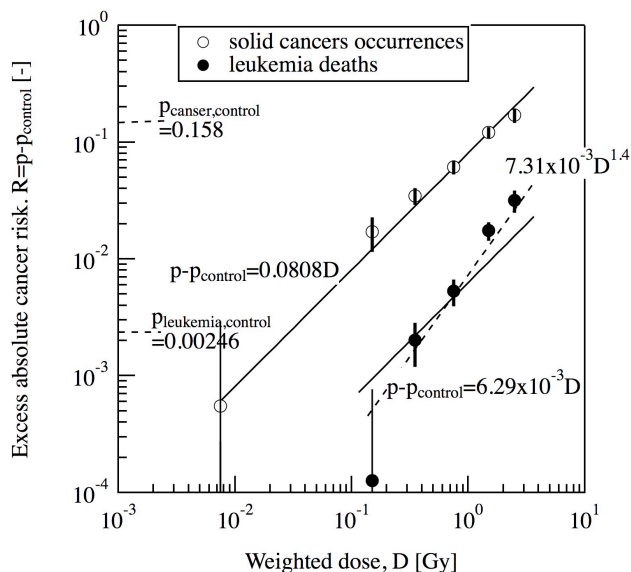


Fig. 3 Relation between weighted radiation dose and excess absolute risks of cancers in logarithmic scale. The solid line is based on the LNT model. The broken line is an empirical one.

対誤差で負の値が出るのを避けるため、有効数字を下げ、低線量で $p_{leukemia,control}=0.0025$ 値を選んで計算した。ここまでの評価値をまとめると、絶対リスクは次式に書くことができる。

$$p_{cancer} = 0.081D + 0.158 \quad D > 0.01 \text{Gy} \quad (2)$$

$$p_{leukemia} = 6.3 \times 10^{-3} D + 0.0025 \quad D > 0.1 \text{Gy} \quad (3)$$

すべての固形がんの発生リスクは 8.1×10^{-2} (1/Gy)であり、実数で表すと1万人当り被曝無しの1580人から100 mSv全身被曝すると1661人に増加し、白血病死亡リスクは 6.3×10^{-3} (1/Gy)であり、1万人あたり被曝無しの25人から、100mSv被曝で31人に増加すると見積もることができる。低線量被曝の効果については昔から議論も多く、直線近似するLNTモデルの他、放射線ホルミシスを考えるモデル、晩発性がん発症にも急性と同じくしきい値を取り入れるモデル、二相モデル等がある。要覧にも述べられている通り、固形がんについては、LNTを否定する何の根拠もないことは今回の評価でも確かめられた。白血病死亡率については、直線関係を仮定するには誤差が大きい様に思われるが、固形がんとは違って被曝後2年ほどで発生し始め、6-8年の比較的早期に現れ、現在では過剰発生がほとんどない等、固形がんとは少し異なった発症形態をとる事も、低線量で傾向が異なって見える一因かもしれない。

Fig.3で求めた過度放射線リスク値を次式で定義する過剰相対リスク (ERR, Excess Relative Risk), E_k として計算し直した結果をFig. 4にプロットする。

$$E_k = \frac{P_k - P_{k,control}}{P_{k,control}} \quad k=\text{solid cancers or leukemia (4)}$$

Fig. 4において、より低線量まで外挿を認め、10mSv被曝すると、固形がんでは0.5%ガン発症率リスクが増加し、白血病では3%（直線近似）あるいはD^{1.4}の近似では、0.5%増加と評価できる。

がん発症は確定的影響ではなく、確率的影響である。人間には放射線で引き起こされたガン遺伝子DNAを修復する能力があり、その因子をできるだけ定量的に評価するためには、集積線量は同じであっても、原爆被害者の様に最初に一時期に被曝する場合と、少量の線量を長く受ける場合とで影響が異なると考えられる。定性的には損傷した遺伝子の修復能力がどのように働くかで違いを証明する必要があるが、これを定量的に表す十分なデータが無く、いまのところ影響は同じと考えられている。同様に放射線の遺伝的影響も長く調べられているが、放射線の遺伝的影響が増加したという報告例はない。

最近の医学データとして、広島県に住む56万の男性と59万人の女性、岡山県に住む38万人の男性と42万人の女性を母集団として比較したLSSに基づく固形がん死亡率と被曝量との関係を求めた大規模な調査結果が報告された⁹⁾。その結果も0.1Gy以下の線量でも有意な差が現れているのが確認されている。また、医療放射線によるがん発症率増加の可能性が予測されており、特に日本で高いことが見積もられている¹⁰⁾ことも最後に追記する。

3. 結 言

以上、RERFによる原爆被曝者のLSSから求められた被曝量と固形がん発症率あるいは白血病死亡率調査を誤差論に基づいて再吟味したところ、以下の結論を得た。

1. もとの RERF 要覧では被曝吸収線量と固形がん発症率あるいは白血病死亡率を男女別あるいは被曝年齢別に分け、調査報告している。第一義的に重要な事は、人が放射線によりどのような率でがん発症あるいは死亡するかを見極めることである。例えば、男女間で若干の差が見られても、男女のがん発症率で合計して見て、データ精度を高めプロットすると、より精度高く評価できる事が分かった。
2. 母集団が二項分布あるいはポアソン分布に従うとして、誤差を評価したところ、固形がん発症率と原爆被曝線量での直線関係は、0.1Gy 以下まで十分な精度があり、LNT 仮説が適用できる事が認められた。白血病でも、0.1Gy 程度まで死亡率と原

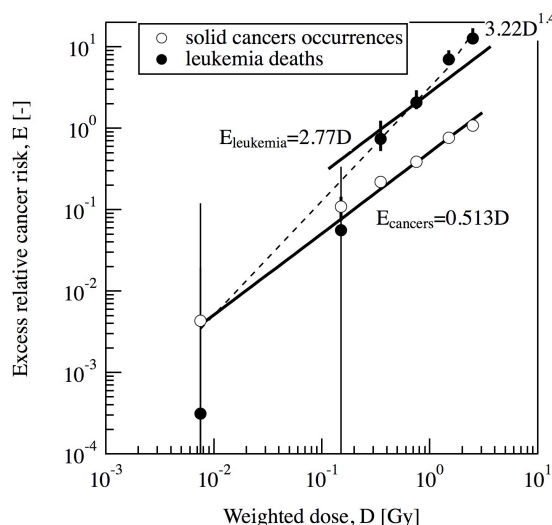


Fig. 4 Relation between weighted radiation dose and excess relative risks of cancers in logarithmic scale. The line is based on the LNT model

爆被曝量との間に線形の相関が見られる事が分かった。

3. 相対リスクは、被曝量0で1に収束し、定性的なガンリスクを理解しやすいが、余剰リスクは定量的なリスクに適しており、被曝量とガンリスクの関係をより明らかに見ることができる。

おわりに

2011年3月11日東北地方太平洋沖でのマグニチュード9.0の東日本大震災と引き続いて起こった後の津波で未曾有の被害が生じた。福島第一原子力発電所で起こった電源喪失事故で、炉の冷却機能が完全に失われ、放射能の放出と水素爆発の結果原子炉建家は極めて甚大な被害が生じ、現在でもなお、多くの被害者が避難されており、その影響の深さに思わず声も無い状況である。

原子炉から放出された放射能と現在でも観測される放射線についてこの間、新聞やテレビでも多くの報道がなされ、多くの見識者の論評がある。しかしながら一部の論評を聞くと、誤った情報に基づく、あるいはそれほど根拠の無い憶測に基づいた論評をしている場合がある。地震で家を失い、津波で家が流され、肉親兄弟、子供が多く失われている状況で、福岡に在住しながら、科学と教育に身を置く人間として、一つ思い立ったことは、いままで私が原子力の分野に関わって来て、遠く昔大学時代に習った放射線の晩発影響について思い返し、インターネット等で得られる情報をもとに、できるだけ正確な情報を発信できないかと思い、放射線、特に低線量放射線、あるいは環境放射線についての情報と、他のリスクとの関係等を一つの文章にまとめた、これが本文です。

参考文献

- 1) 大泉幸雄:“放射線による急性放射線障害”, 医学書院, ISBN 978-4-260-00900-3 (1994).
- 2) 近藤宗平:“人は放射線になぜ弱い”, 講談社第3版, ISBN4062572389 (1999).
- 3) T. D. Lucky, 松平寛通訳:“放射線ホルミシス”, ソフトサイエンス社 (1990).
- 4) B. L. Cohen, 近藤駿介訳:“私はなぜ原子力を選択するか”, ERC 出版, ISBN4-900622-05-2.
- 5) ICRP-Publication 99, Annals of ICRP 35 (2005).
- 6) ECRP, “2010 Recommendations of European Committee on Radiation Risk”, (2010).
<http://www.euradcom.org/2011/ecrr2010.pdf>
- 7) R. Ehrlich, 垂水雄二訳:“トンデモ科学の見破り方—もしかしたら本当かもしれない9つの奇説”, 草思社, ISBN-13:978-4794212825 (2001).
- 8) 財団法人放射線影響研究所: 要覧(2009).
- 9) T. Watanabe, M. Miyao, R. Honda, Y. Yamada, Environ. Health Prev. Med., 13 (2008) 264-270.
- 10) A. B. Gonzales, S. Darby, Lancet, 363 (2004) 345-351.
- 11) D. A. Pierce, D. L. Preston, Radiation Research, 154 (2000) 178-186.

付 録

A1. RERF オリジナルデータ

付録図 Fig. A1は最近のRERF要覧に挙げられたデータ^{8,11)}をそのまま記載したものである。横軸は、本文と同じ換算吸収線量であり、Gy単位で表している。縦軸は、相対リスク（本書ではEで表す）であり、分母は被曝の無い母数集団での固形がん発生確率、分子はある吸収線量における固形がん発生率の比を取っている。データは、男女の区別、被曝時での年齢によりデータを区分しているが、このプロット上では、同じ記号で表されている。このように同じ記号で表すと、例えば男女の差をこのプロットでは判断しない事になる。従って二種類の加重平均を求めてプロットする方が一つ一つの標本でプロットする方が誤差が小さくなるというのが本説の根拠である。

濃い実線がLNTによるモデルで、薄い実線がデータの間をなめらかなにぞった式である。さらに破線は、誤差の幅を記載している。0.1Gyまでは線量が増えるに従って、●印が増加している様に見えるけれども、相対リスクが1.1以下の場合、10%以下の放射線障害リスクが現れない状況でこのようなプロットでは結果をうまく表現できない様になる。また破線の最良近似式で表した結果は、デー

ターを滑らかにつないでいる様に見えるが、現実的にこのようなモデルで整理できる理論的裏付けは記載されていない。

A2. 二項分布あるいはポアソン分布に従うときの平均値、標準偏差、および誤差表記

付録A2に記載する内容は、通常確率統計学の教科書に記載されているものであるが、考察の材料を与えるため、あえて記載する。

ある外的条件の下で、1個1個の固体に事象Aが生じる確率を p_1 とし、一定とする。母数 N_1 個の独立した個体があるとき、その N_1 個の個体中に事象Aが現れる個体数 x は、最小値0から最大値 N_1 まで広がる可能性がある。このとき個体数 x が現れる確率を表す確率密度関数 $f(x)$ は、次の二項分布関数である。

$$f(x) = {}_{N_1} C_x p_1^x (1-p_1)^{N_1-x} \quad (\text{A-1})$$

確率 p_1 が1に比べて十分に小さくかつ N が大きくなると、現れる個体数の平均値を λ_1 と書くと、(A-1)式の $f(x)$ は次のポアソン分布に簡略化できることが理論的に求められる。

$$f(x) = \frac{\lambda_1^x}{x!} e^{-\lambda_1} \quad (\text{A-2})$$

二項分布でもポアソン分布でも、 $f(x)$ をすべての x について足し合わせると、1になる。

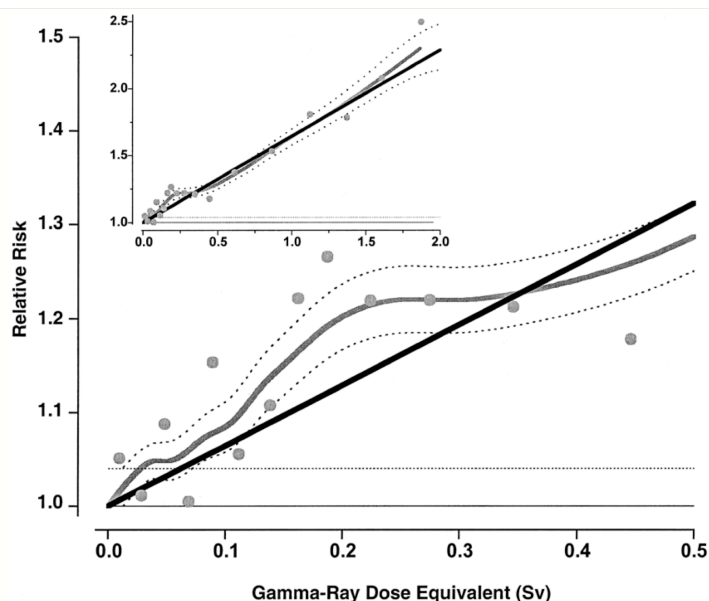


Fig. A1 Relation between weighted radiation dose and relative risks of solid cancers in linear scale. The lines are based on the LNT model and a curvilinear model¹¹⁾

この $f(x)$ を使うと、事象Aの現れる個体数 x の平均値 λ_1 は一般的に次式から求められる。

$$\lambda_1 = \sum_{x=0}^{\infty \text{ or } N_1} x f(x) \quad (\text{A-3})$$

二項分布の足し算を最大値 N_1 までとり、ポアソン分布では無限大まで計算する。代入すると、二項分布とポアソン分布とも次式の様に求められる。

$$\lambda_1 = N_1 p_1 \quad (\text{A-4})$$

$x=\lambda_1$ のとき $f(x)$ が最大値となる。

この $f(x)$ で記述される x の分布の分散は

$$\sigma_1^2 = \sum_{x=0}^{N_1 \text{ or } \infty} (x - \lambda_1)^2 f(x) \quad (\text{A-5})$$

で定義され、代入して計算すると二項分布について

$$\sigma_1^2 = N_1 p_1 (1 - p_1) \quad (\text{A-6})$$

となる。ポアソン分布では、 $\sigma_1^2 = \lambda_1$ である。

母数 N_1 の母集団において事象Aの現れる最も確からしい個体数は、誤差も含めて表すと、

$$\lambda_1 \pm \sigma_1 \quad (\text{A-7})$$

となる。いま逆に現れた観察数 λ_1 からAが起こる平均出現確率 $p_{1,m}$ を推定する場合、最尤値を誤差とともに表記すると

$$p_{1,m} = \frac{\lambda_1 \pm \sigma_1}{N_1} \quad (\text{A-8})$$

となる。

さらに外的条件が変わり、事象Aの起こる確率が p_1 から p_2 に変わったとする。異なった母数 N_1 と N_2 において、それぞれ確率 p_1 と p_2 で事象Aが現れることが予め決まっているとき、各母数中で事象Aの起こる確率の比 R （相対確率）を誤差も含めて書くとポアソン分布では次式で表せる。

上の(A-9)式では次の誤差の伝搬式を用いて解析している。例えば x_1, x_2 の二つの独立変数に誤差を含むときの二次元の分散は次式で計算される。

$$R = \frac{\left(\frac{\lambda_2}{N_2}\right)}{\left(\frac{\lambda_1}{N_1}\right)} \left(1 \pm \sqrt{\frac{1}{\lambda_1} + \frac{1}{\lambda_2}}\right) \quad (\text{A-9})$$

$$\sigma^2 = \left(\frac{\partial f}{\partial x_1}\right)^2 \sigma_{x_1}^2 + \left(\frac{\partial f}{\partial x_2}\right)^2 \sigma_{x_2}^2 \quad (\text{A-10})$$

同様の手法を過剰確率 E に適用すると、次式を得る。

$$E = \left(\frac{\lambda_2}{N_2} - \frac{\lambda_1}{N_1}\right) \pm \sqrt{\frac{\lambda_1}{N_1^2} + \frac{\lambda_2}{N_2^2}} \quad (\text{A-11})$$

A3. 二つの二項分布あるいはポアソン分布を含むときの平均値と標準偏差、および誤差の表記

二つの集団（例えば男女等）において事象Aの現れる個体数に差がわずかにある場合を一つの集団と見なしたときの誤差を考える。二つの集団で異なった出現確率 p_1, p_2 、平均出現数 λ_1, λ_2 、標準偏差 σ_1, σ_2 にある分布集団を同一の母集団と見なし、Aの出現確率の平均値 p_m とその標準偏差 σ_m がどのように表されるかを求める。いま二つの集団の母数をそれぞれ N_1, N_2 とし、合計した集団の全母数における割合を $\alpha_1 (=N_1/(N_1+N_2))$ 、 $\alpha_2 (=N_2/(N_1+N_2))$ で表すと、 $\alpha_1 + \alpha_2 = 1$ の関係にある。両方の集団とも二項分布（以下の式で、式番号に”a”を加えたもの）あるいはその極限であるポアソン分布（以下の式で、式番号に”b”を加えたもの）で表され、合計しても独立に存在すると仮定すると、二つの集団を加えた母集団における確率密度関数 $f(x)$ は次式で与えられる。

$$f(x) = \alpha_{1N_1} C_x p_1^x (1 - p_1)^{N_1 - x} + \alpha_{2N_2} C_x p_2^x (1 - p_2)^{N_2 - x} \quad (\text{A-12a})$$

$$f(x) = \alpha_1 \frac{\lambda_1^x}{x!} e^{-\lambda_1} + \alpha_2 \frac{\lambda_2^x}{x!} e^{-\lambda_2} \quad (\text{A-12b})$$

付録A2の $f(x)$ と同様に、すべての x について上の二つの $f(x)$ をそれぞれ足し合わせると、1になる。

(A-3)式の $f(x)$ に(A-12)式を代入して、整理すると平均出現確率 p_m は二項分布とポアソン分布とも同じ次式に表せる。

$$p_m = \alpha_1 p_1 + \alpha_2 p_2 \quad (\text{A-13})$$

平均出現数 $\lambda_m (=p_m(N_1+N_2))$ を p_1, p_2 で表すと、次式となる。

$$\lambda_m = N_1 p_1 + N_2 p_2 \quad (\text{A-14a})$$

$$\lambda_m = \lambda_1 + \lambda_2 \quad (\text{A-14b})$$

同様に平均出現確率の分散 σ_m^2 について、付録A2の(A-5)式の $f(x)$ に(A-12)式を、 λ_m に(A-14)式を代入して整理すると、二項分布とポアソン分布それぞれ最終的に次式を得る。

$$\sigma_m^2 = \frac{p_1(1-p_1) + p_2(1-p_2)}{N_1 + N_2} + \frac{N_1 N_2 (p_1 - p_2)^2}{(N_1 + N_2)^2}$$

$$\sigma_m^2 = \frac{p_1 + p_2}{N_1 + N_2} + \frac{N_1 N_2 (p_1 - p_2)^2}{(N_1 + N_2)^2} \quad (\text{A-15a})$$

$$(\text{A-15b})$$

(A-15)式の右辺第一項が通常二項分布あるいはポアソン分布本来の分散項で、右辺第二項が二つの分布の平均が異なる事により現れた分散項である。 N_1 と N_2 を大きく取る限り、あるいは p_1 と p_2 の差が小さいものを合わせると、誤差が小さくなる事が(A-15)式から推測できる。以上までの式をまとめて記載すると、次式となる。

$$p_m = \frac{N_1 p_1 + N_2 p_2}{N_1 + N_2} \pm \sqrt{\frac{p_1(1-p_1) + p_2(1-p_2)}{N_1 + N_2} + \alpha_1 \alpha_2 (p_1 - p_2)^2} \quad (\text{A-16a})$$

$$p_m = \frac{N_1 p_1 + N_2 p_2}{N_1 + N_2} \pm \sqrt{\frac{p_1 + p_2}{N_1 + N_2} + \alpha_1 \alpha_2 (p_1 - p_2)^2} \quad (\text{A-16b})$$

観測データから λ_m が求まったとき、逆に p_m をその誤差とともに求める手段を考える。これまでの議論に基づき、観測数 λ_m から出現確率 p_m を誤差 σ_p とともに記載すると次式となる。

$$p_m = \frac{\lambda_m}{N_1 + N_2} \pm \sigma_p \quad (\text{A-17})$$

さらに、ここで二つの集団の出現確率の差を定義する。

$$\Delta p = p_1 - p_2 \quad (\text{A-18})$$

これを使って、(A-15)式の誤差部分を書き換えると、ポアソン分布について、次式となる。

$$\sigma_p = \pm \sqrt{\frac{2\lambda_m}{(N_1 + N_2)^2} + \frac{\alpha_2 - \alpha_1}{N_1 + N_2} \Delta p + \alpha_1 \alpha_2 \Delta p^2} \quad (\text{A-19})$$

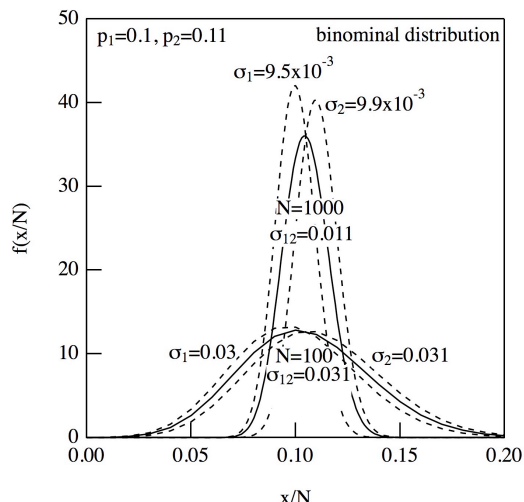


Fig. A2 Comparison of probability density function between $N=100$ and 1000

最初母数 N_1 だけの統計を取る場合に比べて、母数 $N_1 + N_2$ と大きくして統計を取ると、第一項は小さくなる。逆に第二項以降が加わるが、もし $\alpha_1 = \alpha_2 = 0.5$ の様に同程度の母数を含む集団を加えると、第一項は母数が増える分、誤差は小さくなり、この第二項は消え、第三項として $\Delta p^2/4$ が加わる事になる。従って、元の二つの確率差に比べて集団を加える事により σ_p は約半分に減少する。

Fig. A2は $p_1=0.1$ と $p_2=0.11$ と10%だけ違った出現確率を持つ確率密度関数と(図の破線)、二つの分布を足し合わせて一つの分布と考えたときの確率密度関数(図の実線)を表している。横軸は、出現数を母集団の個数で割った出現確率で表している。母集団の個数が $N=100$ 個のときに比べて、 $N=1000$ になると、(A-15)式で明らかな様に誤差は小さくなり、確率密度関数はシャープになる。二つの確率を加え合わせる事により、少し広がりが増すが、母集団の個数を増やす効果の方が大きい事が、この計算例からも定性的に認められる。