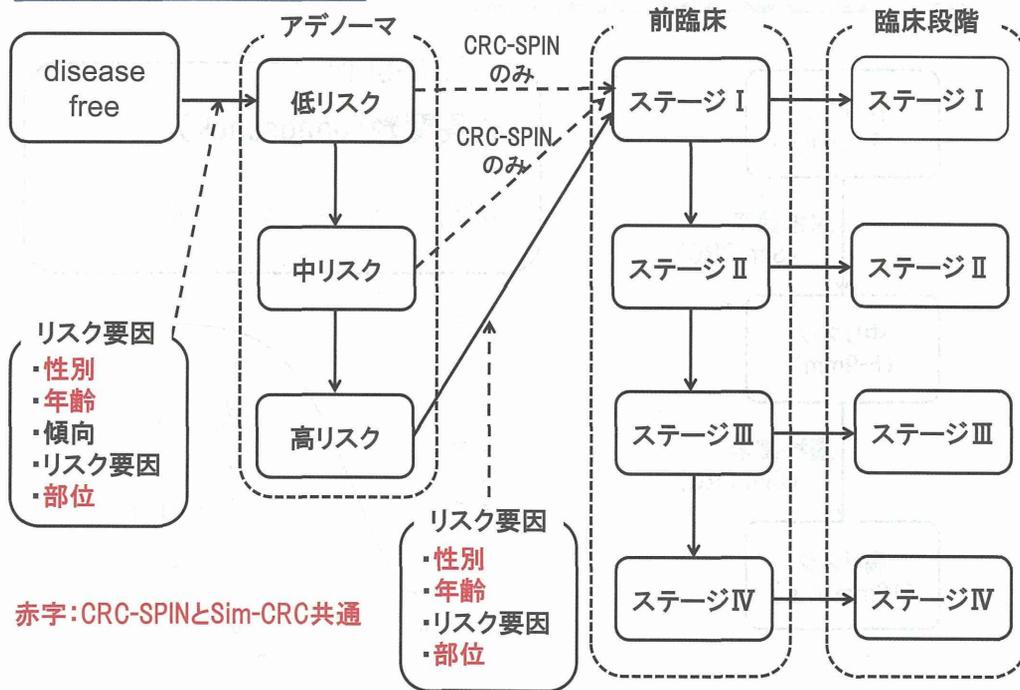


自然史モデル



9

遷移モデル (disease free ⇒ アデノーマ)

・CRC-SPIN

遷移確率

← ポアソン回帰

- ・年齢
- ・性別
- ・部位

・Sim-CRC

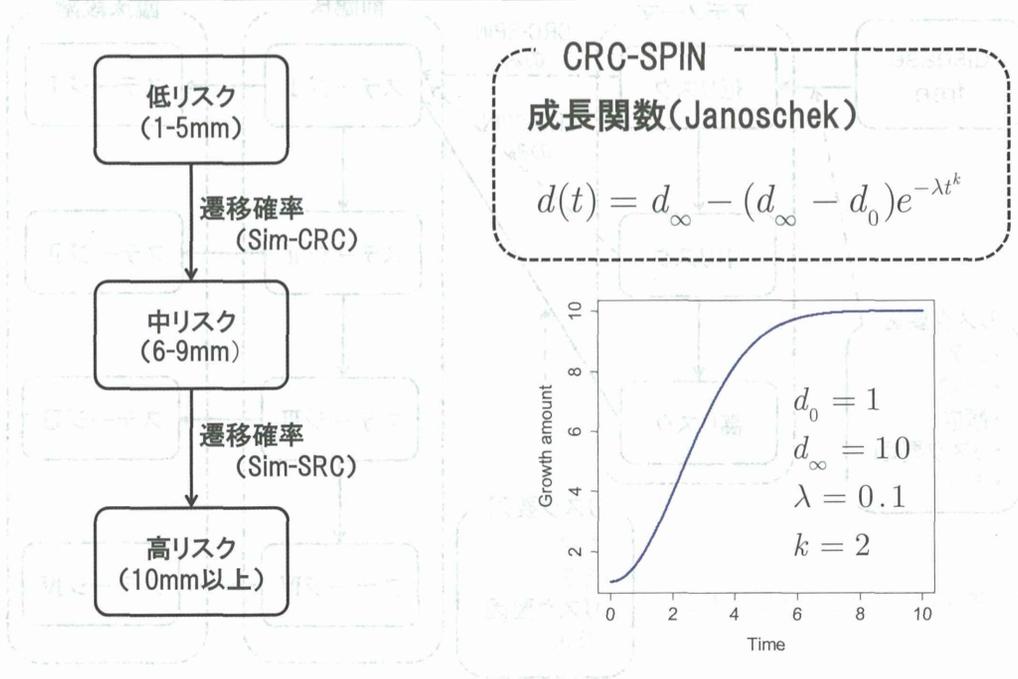
遷移確率

← ロジスティック回帰

- ・年齢
- ・性別
- ・リスク要因
- ・個人の傾向

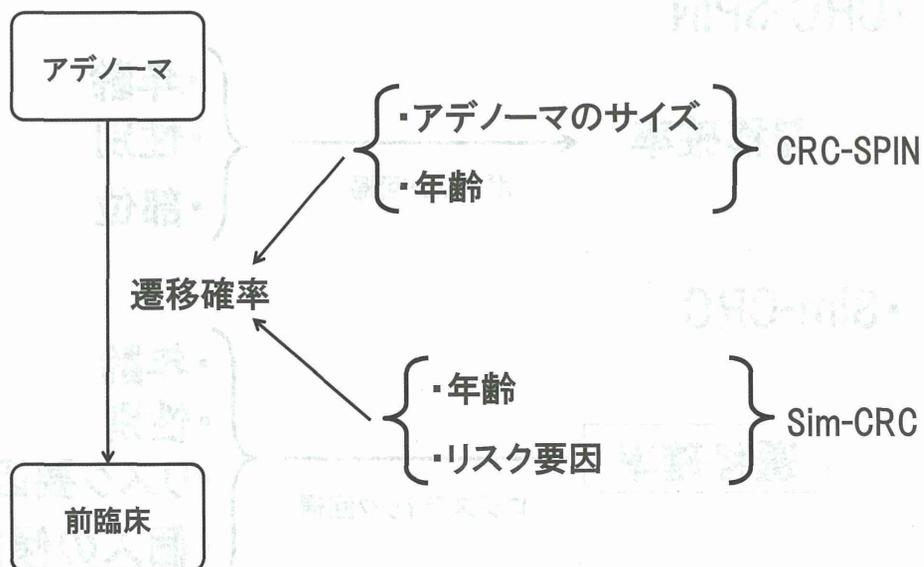
10

アデノーマ成長モデル



11

遷移モデル (アデノーマ→前臨床)



12

肝臓がんへの発展

大腸がん

腫瘍成長 → リスク

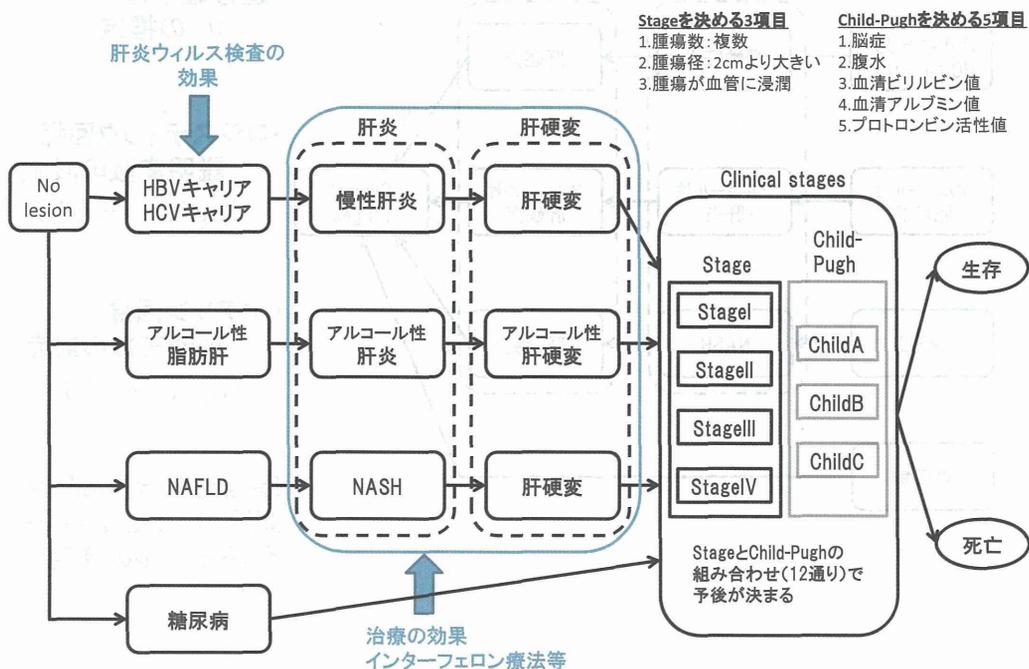
肝臓がん

{ HBV
HCV } キャリア

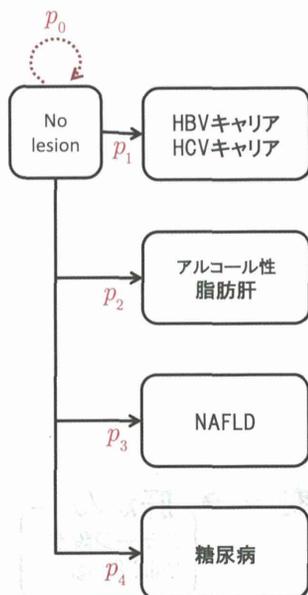


13

自然史モデル

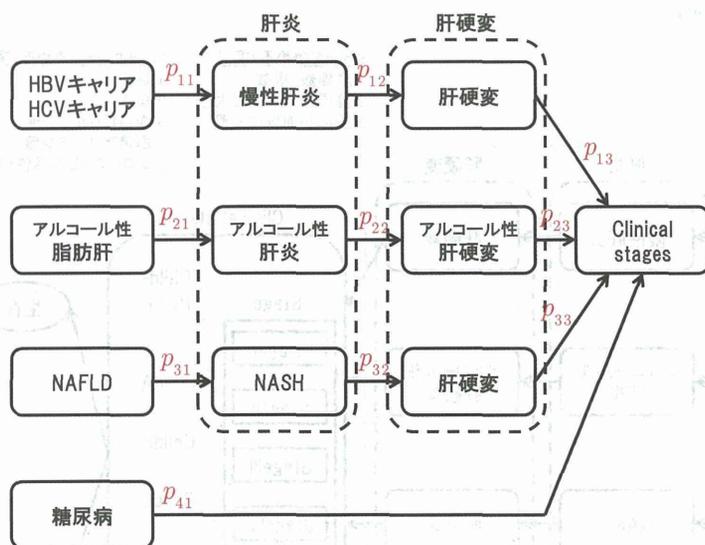


14



枝分かれのパス
 ⇒ 多項分布(多項ロジット回帰)
 (c.f. CRC-SPINの部位分け)

- ・多項分布 : $p_0 \sim p_4$ を推定
- ・多項ロジット : 説明変数の設定が必要

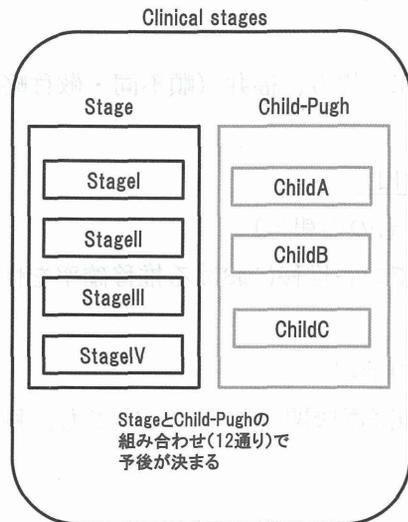


・遷移確率のみ
 p_{ij} の推定

・ロジスティック回帰
 説明変数の設定
 $\text{logit } p_{ij} = x'\beta$

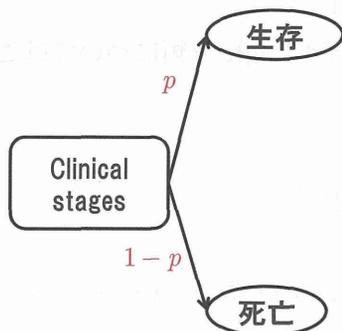
・ポアソン回帰
 説明変数の設定
 $\log \lambda_{ij} = x'\beta$

(感染症モデルSIRや
 SEIRといったコンパート
 メントモデルの活用
 も?)



- 腫瘍数
ロジスティック回帰
- 腫瘍径
成長曲線 (c.f. CRC-SPIN)
- 浸潤
ロジスティック回帰
- Child-Pugh
分布を, 予後の説明変数として活用.
あるいはChild決定過程をモデル化?

17



ロジスティック回帰
説明変数

- ステージやChild-Pughをどう扱うか?
- 元情報(腫瘍数など)をどう扱うか?

Cox回帰(生存時間解析)

× 生存関数の再現が困難?

18

平成 26 年度第 6 回 CISNET JAPAN WG 打ち合わせ

日 時：2014 年 11 月 28 日 13:00～16:00

場 所：広島大学霞キャンパス 基礎・社会医学研究棟 2 階セミナー室 2

参加者：【班員】坂本、加茂、雑賀、田中、

【協力者】秋田、伊森、

【オブザーバー】祖父江、相方、福井（順不同・敬称略）

1. 肝臓がん病態推移モデルとパラメータ（田中・秋田）

- 自然史モデルのパラメータ（現在そろっているものの報告）
 - ・ HCV 由来については感染後から肝臓罹患までの各推移における推移確率を性・年齢階級別に取得できている
 - ・ HBV 由来についても大久先生より提供可能の予定
 - ・ 肝炎ウイルス検査受検率および検査陽性後の医療機関受診率についても、科研費の報告書等から把握できている
- 肝病態の数理モデルの検討
 - ・ 全体図としては要因別の 2 パタンに分類
 - ① ウイルス性肝疾患（HBV と HCV 別に検討）
 - キャリア→慢性肝炎→肝硬変→肝臓・死亡
 - ② 非ウイルス性肝疾患（さらに 3 パタンに細分化）
 - ・ アルコール性脂肪肝→アルコール性肝炎→アルコール性肝硬変→肝臓・死亡
 - ・ NAFLD→NASH→肝硬変→肝臓・死亡
 - ・ 糖尿病（非アルコール・非脂肪性）→肝臓・死亡
 - ・ ステージ（「原発性肝臓取扱規約」における）と Child-Pugh 分類についてはこれまでの議論を継承する

2. 肝臓がん数理モデルとシミュレーション（加茂・坂本・伊森）

- モデル構築
 - ・ これまでの議論をもとに、もう一度 HBV と HCV 由来の分類についてモデルを構築し直す
 - ・ どの部分を固定なのか、シミュレーションの際のシナリオとするのかの議論が必要
- パラメータの推定
 - ・ これまで提供してもらったもの、および肝臓研究会からのものでまずスタートする
- シミュレーション
 - ・ どのようなシナリオを想定するか
 - 肝炎ウイルス検査の受検率を上げる
 - ウイルス検査陽性後の受療率を上げる
 - SVR の効果を受ける対象および効果の強度が変わる
 - ・ どのようなコホートを想定するか
 - クローズ・コホート：ある年のある年齢の集団の○年後のシナリオ別死亡数

オープン・コホート：ある年に対策を開始した場合の○年後のシナリオ別死亡数

3. 肝臓がん診断後の生存率パラメータについて (伊藤)

- 全国原発性肝癌追跡調査のデータベース（肝癌研究会）
会員からの個別研究の位置づけで申請し、データ利用可能
→伊藤、田中、片山（成人病センター）、相方を中心に申請の準備をする

4. 今後の方向性について

- モデルの構築（12月中）
数理モデリングチーム（加茂、坂本、伊森）で実施し、何か結果を出す
- モデルに利用するパラメータの取得
肝癌研究会のデータの利用申請等（伊藤・田中・相方）
- 結果の解釈・モデルの見直し等（1月～2月）

5. その他

平成 26 年度がん政策研究成果発表会（2/19（木）：有楽町朝日スクエア）

6. 次回打ち合わせ

日時：2015 年 1 月 16 日（金） 13：00～16：00

場所：国立がん研究センター

平成 27 年度第 7 回 CISNET JAPAN WG 打ち合わせ

日 時：2015 年 1 月 16 日 13:00～16:00

場 所：国立がん研究センター 国際交流会館

参加者：【班員】坂本、加茂、雑賀、伊藤、田中

【協力者】、秋田、大久、伊森

【オブザーバー】祖父江、(順不同・敬称略)

1. 肝臓がん自然史 (加茂・雑賀)

- ・全体図としては要因別の 2 パタンに分類

① ウイルス性肝疾患 (HBV と HCV 別に検討)

キャリア→慢性肝炎→肝硬変→肝臓がん・死亡

② 非ウイルス性肝疾患 (さらに 3 パタンに細分化)

・アルコール性脂肪肝→アルコール性肝炎→アルコール性肝硬変→肝臓がん・死亡

・NAFLD→NASH→肝硬変→肝臓がん・死亡

・糖尿病 (非アルコール・非脂肪性) →肝臓がん・死亡

- ・ステージ (「原発性肝臓がん取扱規約」における) と Child-Pugh 分類の組み合わせが予後に影響する

2. 肝臓がん数理モデルとシミュレーション (秋田・大久・加茂・坂本・伊森)

- 肝臓がん細胞のサイズ (秋田・大久)

・腫瘍の倍加時間 (DT: Doubling Time) をもとに成長曲線を再現

患者別の診断時の腫瘍サイズと DT について記載のある論文*より、指数関数モデルを仮定して成長曲線を再現

*Nakajima T. et al. Hum Pathol, 2002, 33(1): 92-99

→現状は初期サイズを 1 と仮定しているが、シミュレーションの際には初期サイズをどのように設定するかなど検討が必要

→上記引用文献には発見時のサイズが記載されているので、このサイズと DT をモデル化することで発生時 (サイズ 0) からの成長曲線を推計できないか

- 肝臓がん腫瘍の発生数 (坂本)

・基礎統計データはすべて第 19 回全国原発性肝臓がん追跡調査報告 (2006～2007) (日本肝臓病学会、平成 26 年 5 月発行) より引用

・ベースモデル：生涯に発生する腫瘍数は Poisson 分布する

・シミュレーション結果の概要

生涯腫瘍発生数の分布：平均 2.83、分散 2.86

診断時平均年齢での腫瘍発生数の分布：平均 1.76、分散 1.77

診断時年齢での腫瘍発生数の分布：平均 1.75、分散 1.89

・課題

過剰分散をどのように取り込むか

腫瘍発生割合 (1 年あたり) は年齢を通して一定としているが、年齢とともに増加すると考

えるべきではないか（上記サイズのモデル等との整合性をどうするか）

腫瘍数の分布のデータには腫瘍数ゼロがないが、これをどう考えるか（罹患率で考慮するか）

腫瘍が発生しているのに診断されていない場合をどのように考慮するか

- 肝癌の脈管侵襲（加茂）

基礎統計の情報がなさすぎてモデルの検討に至っていない

数理モデル検討チームからアイディアなどが欲しい

- 最終的なシミュレーションイメージ（前回からの検討課題）

- ・どのようなシナリオを想定するか

肝炎ウイルス検査の受検率を上げる

ウイルス検査陽性後の受療率を上げる

SVR の効果を受ける対象および効果の強度が変わる

- ・どのようなコホートを想定するか

クローズ・コホート：ある年のある年齢の集団の○年後のシナリオ別死亡数

オープン・コホート：ある年に対策を開始した場合の○年後のシナリオ別死亡数

3. その他データからの検討事項（伊藤）

- 全国原発性肝癌追跡調査のデータベース（肝癌研究会）

会員からの個別研究の位置づけで申請したが、利用許可下りず

- 地域がん登録、院内がん登録のデータ

初回診断等に限って、情報収集できることをまとめる

4. 今後の方向性について

- モデルの構築

数理モデリングチーム（加茂、坂本、伊森）で実施し、何か結果を出す

- 結果の解釈・モデルの見直し等

- 今年度の報告書作成（3/4 締め切り）

5. その他

- ・各分担研究者の経理処理などを期限内に依頼

- ・倫理報告書を年度内に提出依頼、年度明け提出

6. 次回打ち合わせ

日時：2015年2月17日（火） 13:00～16:00

場所：大阪大学大学院医学系研究科

肝癌罹患・診断のシミュレーション

腫瘍発生数・時点（担当：坂本）

発生割合の算出について

第 19 回全国原発性肝癌追跡調査報告（2006～2007）（日本肝癌研究会，平成 26 年 5 月発行）より

● 基礎統計データ

新規登録症例（2006/1/1～2007/12/31） 20,850 例（男性 14,715 例，女性 6,135 例）

➤ 肝細胞癌 19,754 例（肝細胞癌 男性 14,014 例，女性 5,740 例）

➤ 臨床診断：臨床的に原発性肝癌と診断された年齢（p.36, Table 13）

◇ 肝細胞癌 18,659 例（男性 13,218 例，女性 5,441 例），5 歳区分

◇ 平均年齢 69.2 歳（男性 67.4 歳，女性 70.9 歳）

➤ 画像診断：腫瘍数の分布（p.46, Table 44）

◇ 肝細胞癌 18,467 例（男女区分なし，腫瘍数不明の 548 例を除く）

◇ 1 個～5 個，6 個以上の 6 カテゴリー

◇ 腫瘍数の条件付平均 2.08，条件付分散 2.86（6 個以上は 6 として計算）

仮定（いろいろ無理があるかもしれませんが）

● 腫瘍発生年齢の範囲

➤ $t_0 = 20$ ：年齢の下限（起点：腫瘍発生数を 0 とする）

➤ $t_u = 100$ ：年齢の上限（寿命ではない）

● 生涯において発生する腫瘍数 N の分布は Poisson 分布（平均・分散とも θ ）に従う。

● 臨床診断において，腫瘍はすべてもれなく発見されるとする。

腫瘍が 1 個でも発見された場合，原発性肝癌（肝細胞癌）であると診断されるとする。

腫瘍が発見された人における腫瘍数の（条件付）分布

$$P(N = n | N \geq 1) = \frac{e^{-\theta} \theta^n}{1 - e^{-\theta}} \frac{1}{n!} \quad (n=1,2,..), \quad E(N | N \geq 1) = \frac{\theta}{1 - e^{-\theta}} \approx \frac{1}{1 - \theta/2}$$

（条件付分散も計算可能であるが少し複雑）

パラメータ θ の推定値

● モーメント法（条件付平均から近似計算）： $\hat{\theta} = 1.04$

● 最尤法（腫瘍数の条件付分布から）： $\hat{\theta} = 1.74$

腫瘍発生数・時点の分布

- 腫瘍数 0 をどのように扱うかが問題（データがない）
 - 1 個目の腫瘍発生機構だけを分けて考えるのはやや不自然では？
 - 年齢の起点を固定する方がモデリングとしては簡単
 - ◇ より精密には、前段階（肝硬変、肝炎など）によって、発生割合のパラメータを変える。
 - ◇ 腫瘍数 0 のデータがないと、Zero-inflated Poisson 分布のあてはめは困難。

仮定

- 腫瘍発生時点：Poisson 過程
- 腫瘍発生割合（1 年あたり） λ_0 ：年齢範囲を通じて一定とする（秋田先生の資料を参照）
 - 時点 t までの累積発生率： $\Lambda(t) = \lambda_0(t - t_0)$
 - 平均診断時年齢（ $t_m = 69.2$ ）における累積発生率： $\Lambda(t_m) = \lambda_0(t_m - t_0)$
 - ◇ 前ページの推定値に基づいて、 $\hat{\theta} = \lambda_0(t_m - t_0)$ より算出した λ_0 の値を用いる。
- 原発性肝癌（肝細胞癌）と診断された年齢と腫瘍の発生は「独立」である。

シミュレーションの方法

- 生涯腫瘍発生数 N ：平均 $\hat{\theta}$ の Poisson 分布に従って生成
- 腫瘍発生時点の生成
 - 与えられた N のもとでの腫瘍発生時点 T_1, \dots, T_N の条件付分布：区間 (t_0, t_u) の一様分布
- 結果の集計
 - 所与の年齢における腫瘍数の分布
 - 診断年齢時点の腫瘍数の分布
 - ◇ 診断時年齢：実際の度数分布（5 歳区分）に基づいて生成

結果の要約

- 生涯腫瘍発生数の分布：平均 2.83, 分散 2.86
- 平均診断時年齢での腫瘍発生の分布：平均 1.76, 分散 1.77
- 診断時年齢での腫瘍数の分布：平均 1.75, 分散 1.89
 - 腫瘍数 1 以上に限ると、条件付平均 2.16, 条件付分散 1.44

課題

- 実際の腫瘍数の分布は散らばりが大きい：過剰分散 (over-dispersion) の問題
 - 腫瘍発生割合は人によって異なると考えるべき（前段階による違い、確率的変動）
- 腫瘍発生割合はむしろ一定でない（年齢とともに増減）と考えるのが自然
 - 非同次 (non-homogeneous) Poisson 過程
- 実際に肝癌と診断される年齢の度数分布は腫瘍数 0 の人を含んでいないのに、シミュレーションによる診断年齢時点の腫瘍数は 0 になることもある
 - 腫瘍数 1 以上の人限定すると、診断年齢の分布はやや大きい方に偏る
- 腫瘍が発生しているのに肝癌と診断されない場合を考慮していない

平成 27 年度第 8 回 CISNET JAPAN WG 打ち合わせ

日 時：2015 年 2 月 17 日 13:00～16:00

場 所：大阪大学環境医学講座会議室

参加者：【班員】田中（電話）、坂本、加茂、雑賀、伊藤

【協力者】秋田（電話）、大久（電話）、片山（電話）、伊森、松田

【オブザーバー】祖父江、（順不同・敬称略）

1. 事務連絡（加茂）

- ・ 2 月 19 日・がん政策の発表会において用いるスライドを回覧&修正
- ・ 報告書および経理書類の〆切について

2. 肝臓がん数理モデルとシミュレーション（坂本・秋田）

- ・ 腫瘍個数について（坂本）

非同次のポアソン過程を用いてモデルを洗練

肝がん研究会のデータをほぼ再現可能になった。

多数（6 個以上）に関するフィッティングが上手く機能していない、

または情報が少ないため仮説等の設定が必要

- ・ 腫瘍成長について（秋田）

倍加時間のモデリングに関して、特性がありそう

初期から 1500 日を境にして、2 群に分割した解析を実施。ただし、このままだとシミュレーションに適用できないため、このままでシミュレーションに適用する方法の検討もしくはモデルの中に 2 群に分けて検討できる要素を検討する必要がある。

他の説明変数がありそう

倍加時間の解析に関して、後日雑賀先生から新たな提案があった

（次回サテライトの議題）

3. 今後の展開

- ・ 胃がんについて

ピロリ菌と胃がんに関するモデリングを来年度より手がける

北大の浅香先生にコンタクトを取り、加藤先生を紹介

がん政策の発表会において加藤先生とコンタクトを取る予定 ← 取りました。

- ・ 今後の活動方針について

これまでの、全員で 1 つのテーマに当たるスタイルを来年度より改めたい

臓器毎に「シナリオ担当（まとめ役）」、「疫学（データ収集）担当」、「数理担当」を決め、ある程度独立で活動を進める方向で

4. 次回予定

5 月 20 日（水）10 時～13 時 大阪大学

V. 数理部門ミーティングの記録

日 時：2014 年 12 月 24 日 9:30~12:00

場 所：岡山大学環境科学講座会議室

参加者：坂本、加茂、秋田、伊藤（順不同・敬称略）

今回の主テーマ：肝臓がんにおける臨床段階のモデリング

1. 腫瘍個数のモデリング（坂本）

- ・腫瘍発生レート λ を推定
- ・数値実験の結果、 $\lambda = \frac{1}{80} \sim \frac{1}{40}$ が妥当か？
- ・今後の課題としては $\lambda(t) = \lambda_0 \exp(\beta(t - t_0))$ の形での推定。

2. 腫瘍サイズのモデリング（秋田）

- ・倍加時間に関する論文あり ($N = 34$)。これを基にモデリングできないか？
- ・指数成長モデル（倍加時間をモデル化） $S(t) = s_0 e^{ct}$ ：パラメータは s_0 と c
- ・ $S(t + \Delta t) = s_0 e^{c(t+\Delta t)} = 2S(t) = 2s_0 e^{ct} \quad \Leftrightarrow \quad c\Delta t = \log 2$
 $\Leftrightarrow \quad \log c = \log \Delta t + \log \log 2$
 Δt （倍加時間）に分布を仮定して c を推定する。
- ・今後の課題としては、年齢と性別が影響を与えているかを上記論文から調べる。

3. 脈管浸襲の確率（加茂）

- ・浸襲の確率は、サイズと個数が説明変数となっており、タブレートされたデータからはパラメータ推定ができなかった。
- ・モデル： $p(t) = \left(1 + e^{-\alpha t}\right)^{-1} \prod_{k=1}^n \left(1 + e^{-\beta S(t)}\right)^{-1}$ からパラメータ α と β を推定する。
- ・今後の課題としては、個数とサイズの情報が上記より得られるので、これを用いて実際の推定を行う。

4. 次回までの課題

- ・個数のモデリング：パラメータの概数を推定する（坂本）。
- ・サイズのモデリング：論文から、サイズに影響を与える要因を探索する（秋田）。
- ・浸襲の確率：個数とサイズのデータを用いてパラメータを推定する（加茂）。
- ・データの探索：年齢や性別の上記データに関する情報を探索する（伊藤）。

5. 今後の方向性について

- ・ 班会議が約2ヶ月間隔であるため、その間に必要に応じて数理部門のサテライトミーティングを開催する。
- ・ モデルやプログラミングなどの詳細な検討を行う場とする。
- ・ 基本的には関西地方で開催し、メンバーは、坂本、加茂、伊藤、秋田、伊森とする。

6. 次回サテライトミーティング予定

日時：2015年2月5日（木） 13:00～17:00

場所：岡山大学環境科学講座会議室（秋田先生は電話参加の予定）

脈管侵襲の有無について・・・加茂

$p(t)$: 時刻 t において, 脈管侵襲を起こす確率

$$p(t) = \left(\frac{1}{1 + e^{-\alpha t}} \right) \prod_{k=1}^n \left(\frac{1}{1 + e^{-\beta S_k(t)}} \right) \dots (*)$$

ここで $S_k(t)$: 腫瘍の大きさで $S_k(t) = S_0 e^{c(t-X_1-\dots-X_{i-1})}$ である.

決定する必要があるパラメータは α と β .

α : 時間の経過が侵襲に与える影響 (強度)

β : サイズの大きさが侵襲に与える影響 (強度)

パラメータ推定には, 個々に関してサイズと個数と年齢の情報が必要であるが, それが無い場合現時点ではラフな推定しか出来ない.

サイズの平均は Table 43 より 4.008135 .

個数の平均は Table 44 より 2.193959 .

年齢の平均は Table 178 より 67.76619 .

侵襲のパーセンテージは 0.08347025 .

以上の情報をそのまま式 (*) に代入すると

$$0.08347025 = \left(\frac{1}{1 + e^{-67.76619\alpha}} \right) \left(\frac{1}{1 + e^{-4.008135\beta}} \right)^{2.193959}$$

となる. 未知パラメータが2つなので決定できない.

仕方ないので, 乱数を発生させてパラメータを推定する.

$$\text{logit } p \sim N(\text{logit } p, \sigma_1^2)$$

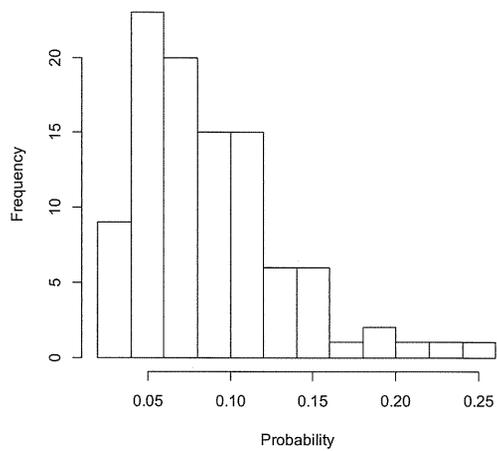
$$\text{log } t \sim N(\mu, \sigma_2^2)$$

$$\text{log } S \sim N(\mu, \sigma_3^2)$$

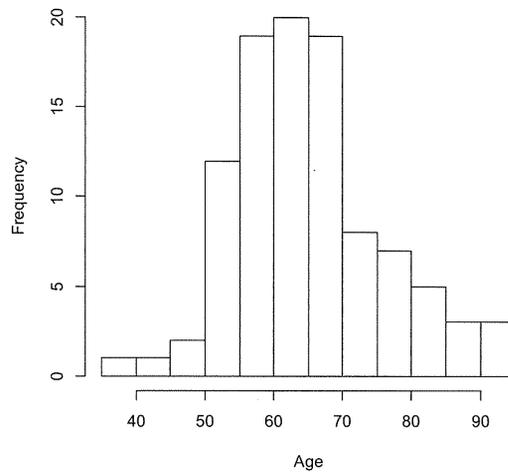
$$n \sim \text{Pois}(\lambda)$$

と仮定して 100 個の観測値を発生させ, 最小二乗推定により $\hat{\alpha}$ と $\hat{\beta}$ を得る.

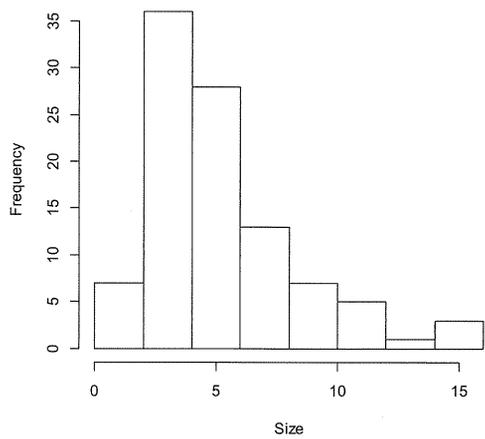
侵襲のパーセンテージ



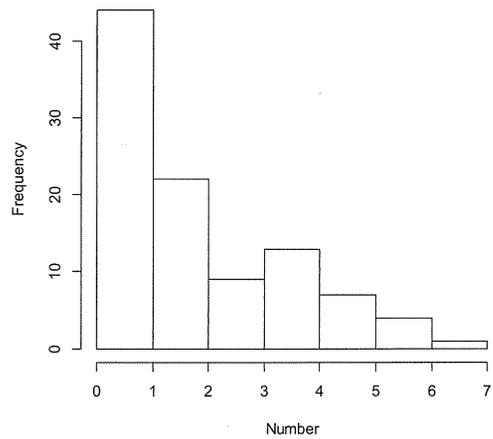
年齢



サイズ



個数



腫瘍発生数および時点の分布

- 疑問点：1 個目の腫瘍発生機構を分けて考えるのは不自然では？
 - 年齢の起点を固定した方が簡単
 - 前段階（肝硬変，肝炎など）によって，Rate function のパラメータを変えては？
- 簡単のため，年齢の下限・上限を固定
 - $t_0 = 20$ ：年齢の下限（起点：腫瘍発生数を 0 とする）
 - $t_u = 100$ ：年齢の上限
- Homogeneous Poisson process（発生リスク一定）：秋田先生のスライド
- Non-homogeneous Poisson Process（発生リスクが年齢とともに増加することを想定）
 - Rate function: $\lambda(t) = \lambda_0 \exp\{\beta(t - t_0)\}$
 - ◇ $\beta = 0$ のとき $\lambda(t) = \lambda_0$ (homogeneous)
 - Cumulative rate function: $\Lambda(t) = \int_{t_0}^t \lambda(s) ds$
 - ◇ $\beta = 0$ のとき $\Lambda(t) = \lambda_0(t - t_0)$
 - ◇ $\beta \neq 0$ のとき $\Lambda(t) = (\lambda_0/\beta)[\exp\{\beta(t - t_0)\} - 1]$
 - 生涯腫瘍発生数 N ：平均 $\Lambda(t_u)$ の Poisson 分布
 - 与えられた N のもとでの腫瘍発生時点 T_1, \dots, T_N の条件付分布：
累積分布関数 $F(t) = \Lambda(t)/\Lambda(t_u)$ の分布
 - ◇ $(0,1)$ 上の一様乱数 U_1, \dots, U_N
 - ◇ $T_j = F^{-1}(U_j)$ とする
 - パラメータ λ_0, β の設定
 - ◇ $\lambda(t_0), \lambda(t_m)$ を与える (t_m : 所与の年齢)

Calibration

- シミュレーション：所与の年齢に対する腫瘍数の分布
 - パラメータを変えたときの平均値をシミュレート
- 実際のデータ：検査時点（年齢は異なる）での腫瘍数の分布？

日 時：2015 年 2 月 5 日 14：00～16：30

場 所：岡山大学環境科学講座会議室

参加者：坂本、加茂、伊森、秋田（スカイプ参加）（順不同・敬称略）

【肝臓がんにおける臨床段階のモデリング】

1. 腫瘍個数のモデリング（坂本）

- ・ポアソン分布により個数および時点の分布を推定するモデル開発。
- ・パラメータに個人差を許すモデルを構築。
- ・数値実験により、肝がん研究会の結果をほぼ再現可。

2. 腫瘍サイズのモデリング（秋田）

- ・倍加時間に関する論文（ $N = 34$ ）のデータに関する詳細解析。
- ・倍加時間ではなく、実際の期間を使った解析を行った。
- ・倍加時間 1500 日を境に、2 群に分け、それぞれの成長レートを推定。
- ・成長レートに影響してそうな説明変数を用いて、変化係数のようなモデリングは出来ないか？

3. 脈管浸襲の確率（加茂）

- ・データがバイナリでしか存在せず、バイナリ乱数でシミュレーションするしかない。
- ・詳細データについては、文献調査を続ける。

4. 次回までの課題

- ・ステージングを決定する変数のモデルはほぼ固まったので、これらを融合するプログラミング
- ・臨床段階までの推移のモデル化に手を付ける。

5. 次回サテライトミーティング予定

日時：2015 年 3 月 17 日（火） 14：00～18：00

場所：札幌医科大学基礎棟 102 会議室

肝癌罹患・診断のシミュレーション 腫瘍発生数・時点 (担当: 坂本)

発生割合の算出について

第 19 回全国原発性肝癌追跡調査報告 (2006~2007) (日本肝癌研究会, 平成 26 年 5 月発行) より

● 基礎統計データ

新規登録症例 (2006/1/1~2007/12/31) 20,850 例 (男性 14,715 例, 女性 6,135 例)

- 肝細胞癌 19,754 例 (肝細胞癌 男性 14,014 例, 女性 5,740 例)
- 臨床診断: 臨床的に原発性肝癌と診断された年齢 (p.36, Table 13)
 - ◇ 肝細胞癌 18,659 例 (男性 13,218 例, 女性 5,441 例), 5 歳区分
 - ◇ 平均年齢 69.2 歳 (男性 67.4 歳, 女性 70.9 歳)
- 画像診断: 腫瘍数の分布 (p.46, Table 44)
 - ◇ 肝細胞癌 18,467 例 (男女区分なし, 腫瘍数不明の 548 例を除く)
 - ◇ 1 個~5 個, 6 個以上の 6 カテゴリー
 - ◇ 腫瘍数の条件付平均 2.08, 条件付分散 2.86 (6 個以上は 6 として計算)

仮定 (いろいろ無理があるかもしれませんが)

- 腫瘍発生年齢の範囲
 - $t_0 = 20$: 年齢の下限 (起点: 腫瘍発生数を 0 とする)
 - $t_u = 100$: 年齢の上限 (寿命ではない)
- 生涯において発生する腫瘍数 N の分布
 - Poisson 分布 (平均・分散とも θ),
 - θ も確率的に変動 (個人差を考慮), 平均 μ , 分散 v^2 の対数正規分布
- 臨床診断において, 腫瘍はすべてもれなく発見されるとする.
腫瘍が 1 個でも発見された場合, 原発性肝癌 (肝細胞癌) であると診断されるとする.

腫瘍が発見された人における腫瘍数の (条件付) 分布

$$P(N = n | N \geq 1, \theta) = \frac{e^{-\theta} \theta^n}{1 - e^{-\theta}} \frac{1}{n!} \quad (n=1,2,..)$$

$$E(N | N \geq 1, \theta) = \frac{\theta}{1 - e^{-\theta}}, \quad V(N | N \geq 1, \theta) = \frac{\theta^2}{1 - e^{-\theta}} + \frac{\theta}{1 - e^{-\theta}} - \frac{\theta^2}{(1 - e^{-\theta})^2}$$

$$E(N | N \geq 1) \approx \frac{\mu}{1 - e^{-\mu}}, \quad V(N | N \geq 1) \approx \frac{\mu^2}{1 - e^{-\mu}} + \frac{\mu}{1 - e^{-\mu}} - \frac{\mu^2}{(1 - e^{-\mu})^2} + \frac{1 - e^{-\mu} + \mu e^{-\mu}}{(1 - e^{-\mu})^2} v^2 \quad (\text{デルタ法})$$

θ を定数パラメータとしたときの推定値

- モーメント法 (条件付平均から近似計算): $\hat{\theta} = 1.04$
- 最尤法 (腫瘍数の条件付分布から): $\hat{\theta} = 1.74$

θ を確率変数としたときの μ, v^2 の推定値

- 最尤法 (かなり粗い近似): $\hat{\mu} = 1.74$
- モーメント法 (条件付分散から近似計算): $\hat{v}^2 = 0.922$