1、　利用できるデータを選択し、
2、　その１つ１つのデータが研究に利用する際に留意すべきポイントがあるかをチェックして情報を付加し
3、　分析に必要な情報を容易にとりだせるよう用途別に分割し、

さまざまな統計処理を行うための数値を計算しやすいよう整理して共通分析用データセットとして作成し、本研究班においてより精度の高い、高度な分析を行うための環境を整備することを試みている。本年度は平成22年度から3年度(36ヶ月)分のデータを通年で１つのデータセットとして取り扱い、各種の分析処理を行うことが出来る基盤作成を目的とした。

B.方法

以下の処理を、研究班保有のDPCデータに対して行い、分析用データセットを作成することとした。

１、データの取り込み

参加医療機関から提供されたDPCデータ（FF1/3/4/D/E/F/外来ファイル）をDBに取り込む。その際本年度から実施した収集時のデータの暗号化に対応した処理を追加した。

２、エラーチェック

提出されたデータのエラーチェックを行い、エラーデータをデータセット内から除外する。また研究に使用する際留意する項目に対してチェックしフラグを付与する。

どのような条件についてエラー・留意とするかについては本年の研究として検討を行う。

３、DPCコード情報の一体的保有

平成20年度「DPC松田研究班版DPCコーダーの開発について」で作成したDPCコーディングツールを利用したDPCコードを分析用データセット内に取り込み、他のデータと一体的に保有する

以上の処理を行い、分析用データセットを作成する。

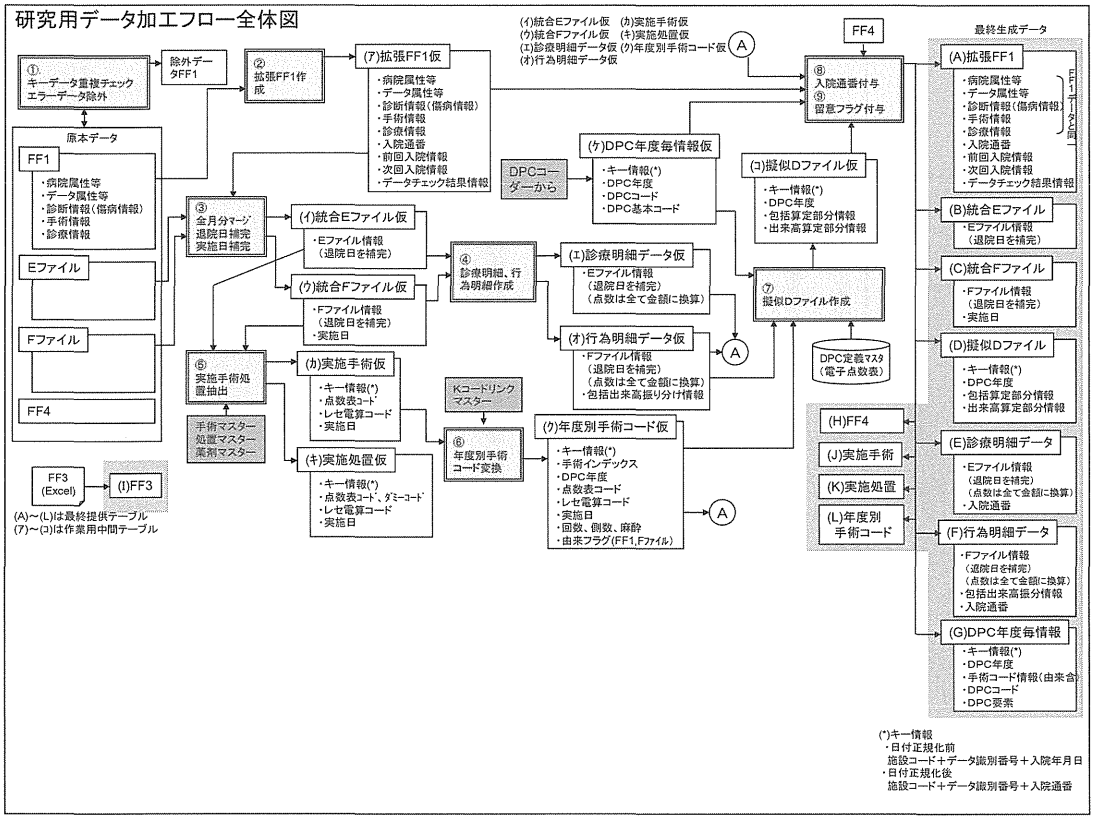
分析用データセットには次のファイルが含まれる

(A) 拡張FF1	様式１（FF1）情報に留意フラグ等の情報を付加したもの
(B) 統合Eファイル	Eファイルに退院日及び期間内入院回数を付加したもの
(C) 統合Fファイル	Fファイルに退院日・実施日等を付加したもの
(D) 疑似Dファイル	Dファイル作成ルールに基づいて擬似的にDファイルを生成
(E) 診療明細データ	統合Eファイルの点数部分を金額に置き換えたもの
(F) 行為明細データ	統合Fファイルの点数部分を金額に置き換えたもの
(G) DPC年度毎情報	運用時期別に生成されたDPCコード情報
(H) FF4	様式４
(J) 実施手術	Fファイルから手術関係のレコードのみを抽出したもの

(K) 実施処置　　Fファイルから処置・薬剤関係のレコードのみを抽出したもの

(L) 年度別手術コード 手術コードを診療報酬の運用年度別に変換したもの

尚、データ処理のフローは下に示す図のとおりである。



C.結果

本年度の研究期間においては平成25年度時点で伏見班保有のデータに対してB.方法で示した処理を行った。

以下特記すべき事項に関して記載する。

1、エラーチェックについて

本研究データセット作成については、以下のルールの下で、データエラーチェックおよび留意フラグ付与を行った。

表１　エラーチェック仕様

No.	ｴﾗｰ番号	チェック内容	チェック論理(NG 条件)	対象データ	種別	備考	フラグ内容等
1	ERR010	FF1 のキーが重複している	FF1 にて「施設コード＋データ識別番号＋入院年月日＋診療情報番号」が重複 （過去の FF1 との重複もチェックする。）	FF1	キー重複	・重複データは FF1 データを除外データ FF1 テーブルに、エラー情報と共に格納し、以降の処理には使用しない。 ・パターンとしては、「日帰り入退院＋同一日再入院」、「入院日一致、退院日不一致」、「入院日一致、退院日一致」があるが、ツールはそこまでは判定しない。 ・親様式1（診療情報番号＝0）がエラーの場合、子様式1（診療情報番号≧1）のレコードも一緒に除外する。 ・子様式1の重複は該当の子様式 1 のみ除外とし、親様式1および他の子様式1には影響を及ぼさない。	

2	ERR020	FF4 のキーが重複している	FF4 にて「施設コード＋データ識別番号＋入院年月日」が重複	FF4	キー重複	・重複データは FF1 データを除外データ FF1 テーブルに、エラー情報と共に格納し、以降の処理には使用しない。 ・親様式1と子様式1の両方を除外する。(過去に仮確定した子様式1も除外する。)	
3	ERR030	E ファイルのキーが重複している	E ファイルの「施設コード＋データ識別番号＋入院年月日＋データ区分＋順序番号＋該当月」が重複	E ファイル	キー重複	・重複データは FF1 データを除外データ FF1 テーブルに、エラー情報と共に格納し、以降の処理には使用しない。 ・親様式1と子様式1の両方を除外する。(過去に仮確定した子様式1も除外する。)	
4	ERR040	F ファイルのキーが重複している	F ファイルの「施設コード＋データ識別番号＋入院年月日＋データ区分＋順序番号＋行為明細番号＋該当月」が重複	F ファイル	キー重複	・重複データは FF1 データを除外データ FF1 テーブルに、エラー情報と共に格納し、以降の処理には使用しない。 ・親様式1と子様式1の両方を除外する。(過去に仮確定した子様式1も除外する。)	
5	ERR050	EF 統合ファイルのキーが重複している	EF 統合ファイルの「施設コード＋データ識別番号＋入院年月日＋データ区分＋順序番号＋行為明細番号＋該当月」が重複	EF 統合	キー重複	・重複データは FF1 データを除外データ FF1 テーブルに、エラー情報と共に格納し、以降の処理には使用しない。 ・親様式1と子様式1の両方を除外する。(過去に仮確定した子様式1も除外する。)	

6	ERR060	外来Eファイルのキーが重複している	外来Eファイルの「施設コード＋データ識別番号＋入院年月日＋データ区分＋順序番号＋該当月」が重複	外来E	キー重複	・重複データはFF1データを除外データFF1テーブルに、エラー情報と共に格納し、以降の処理には使用しない。 ・親様式1と子様式1の両方を除外する。（過去に仮確定した子様式1も除外する。）	
7	ERR070	外来Fファイルのキーが重複している	外来Fファイルの「施設コード＋データ識別番号＋入院年月日＋データ区分＋順序番号＋行為明細番号＋該当月」が重複	外来F	キー重複	・重複データはFF1データを除外データFF1テーブルに、エラー情報と共に格納し、以降の処理には使用しない。 ・親様式1と子様式1の両方を除外する。（過去に仮確定した子様式1も除外する。）	
8	ERR110	在院日数が1日未満である	入院日＞退院日	FF1	除外	・エラーデータはFF1データを除外データFF1テーブルに、エラー情報と共に格納し、以降の処理には使用しない。 ・親様式1（診療情報番号＝0）がエラーの場合、子様式1（診療情報番号≧1）のレコードも一緒に除外する。 ・子様式1のエラーの場合は、該当子様式1は除外するが、親様式1および他の子様式1は除外しない。 ・子様式1で退院日が'00000000'である場合は、許容する。親様式1はこのエラーとなる。	

9	ERR120	入院時年齢が0歳未満である	生年月日＞入院日	FF1	除外	・エラーデータはFF1データを除外データFF1テーブルに、エラー情報と共に格納し、以降の処理には使用しない。 ・親様式1（診療情報番号＝0）がエラーの場合、子様式1（診療情報番号≧1）のレコードも一緒に除外する。（過去に仮確定した子様式1も除外する。） ・子様式1のエラーの場合は、該当子様式1は除外するが、親様式1および他の子様式1は除外しない。	

10	ERR130	年月日が誤っている	実在しない年月日(13 月 1 日、7 月 32 日など)、および SQLserver の datetime 型で扱えない日付(1753 年 1 月 1 日以前)	ALL	除外	・エラーデータは FF1 データを除外データ FF1 テーブルに、エラー情報と共に格納し、以降の処理には使用しない。 ・対象は全ての日付情報とする ・あくまで実在しない年月日のチェックのみであり、手術日が入院日と退院日の間にあるか、などの相関チェックは行なわない。 ・'00000000'は許容する。 ・親様式1(診療情報番号＝0)がエラーの場合、子様式1(診療情報番号≧1)のレコードも一緒に除外する。(過去に仮確定した子様式1も除外する。) ・子様式1のエラーの場合は、該当子様式1は除外するが、親様式1および他の子様式1は除外しない。 ・様式1開始日、様式1終了日をチェック対象に追加。	
11	ERR140	(欠番)					
12	ERR150	(欠番)					
13	ERR160	統括診療情報番号が異常である(3日以内再	統括診療情報番号が0以上の数字でない	FF1	除外	・他のエラーデータ除外より最初に判定する。 ・エラーデータは FF1 データを除外データ FF1 テーブルに、エラー情報と共に格納し、以降の処理には使用しない。	

		入院集約レコード）					
14	ERR170	様式1対象期間が1日未満である	様式1開始日＞様式1終了日	FF1	除外	・エラーデータはFF1データを除外データFF1テーブルに、エラー情報と共に格納し、以降の処理には使用しない。 ・親様式1（診療情報番号＝0）がエラーの場合、子様式1（診療情報番号≧1）のレコードも一緒に除外する。 ・子様式1のエラーの場合は、該当子様式1は除外するが、親様式1および他の子様式1は除外しない。	
15	ERR180	親様式1において様式1開始日、様式1終了日が入院日、退院日に一致していない	統括診療情報番号＝0 かつ（様式1開始日が入院年月日と一致しない または 様式1終了日が退院年月日と一致しない）	FF1	除外	・エラーデータはFF1データを除外データFF1テーブルに、エラー情報と共に格納し、以降の処理には使用しない。 ・親様式1（診療情報番号＝0）がエラーの場合、子様式1（診療情報番号≧1）のレコードも一緒に除外する。（過去に仮確定した子様式1も除外する。） ・子様式1のエラーの場合は、該当子様式1は除外するが、親様式1および他の子様式1は除外しない。	

16	ERR190	様式1開始日が入院と退院の範囲外である	様式1開始日＜入院年月日 または 様式1開始日 ＞ 退院年月日	FF1	除外	・エラーデータはFF1データを除外データFF1テーブルに、エラー情報と共に格納し、以降の処理には使用しない。 ・親様式1（診療情報番号＝0）がエラーの場合、子様式1（診療情報番号≧1）のレコードも一緒に除外する。 ・子様式1のエラーの場合は、該当子様式1は除外するが、親様式1および他の子様式1は除外しない。	
17	ERR200	様式1終了日が入院と退院の範囲外である	様式1終了日＜入院年月日 または 様式1終了日 ＞ 退院年月日	FF1	除外	・エラーデータはFF1データを除外データFF1テーブルに、エラー情報と共に格納し、以降の処理には使用しない。 ・親様式1（診療情報番号＝0）がエラーの場合、子様式1（診療情報番号≧1）のレコードも一緒に除外する。 ・子様式1のエラーの場合は、該当子様式1は除外するが、親様式1および他の子様式1は除外しない。	
18	ERR510	EファイルとFファイルが不整合である（Fファイルデー	Eファイルの各レコードに対して、同じ月のFファイルに同じ「施設コード＋データ識別番号＋入院年月日＋データ区分＋順序番号」を持つレコードが存在しな	Eファイル、Fファイル	フラグ	・親様式1（診療情報番号＝0）のみチェック対象。 ・子様式1（診療情報番号≧1）はチェック対象外。フラグは0を設定する。	0:エラー無し 1:Fファイルデータ欠落有り

		タが非存在である)	い。				
19	ERR520	EファイルとFファイルが不整合である(Eファイルデータ非存在である)	Fファイルの各レコードに対して、同じ月のEファイルに同じ「施設コード＋データ識別番号＋入院年月日＋データ区分＋順序番号」を持つレコードが存在しない。	Eファイル、Fファイル	フラグ	・親様式1(診療情報番号＝0)のみチェック対象。 ・子様式1(診療情報番号≧1)はチェック対象外。フラグは0を設定する。	0:エラー無し 1:Eファイルデータ欠落有り
20	ERR530	入院期間外のEFファイルレコードが存在する	Eファイルの実施日＜FF1の入院日 または FF1の退院日＜Eファイルの実施日	FF1,Eファイル、統合Eファイル	フラグ	・親様式1(診療情報番号＝0)のみチェック対象。 ・子様式1(診療情報番号≧1)はチェック対象外。フラグは0を設定する。	0:エラー無し 1:入院期間外Eファイルデータ有り

21	ERR540	E ファイル、F ファイルの退院日が FF1 と異なる	「施設コード＋データ識別番号＋入院年月日」が FF1 と E ファイル、F ファイルで同一であるが、退院日が異なる。	FF1,E ファイル、F ファイル EF 統合ファイル	フラグ	・親様式 1（診療情報番号＝0）のみチェック対象。 ・子様式 1（診療情報番号≧1）はチェック対象外。フラグは 0 を設定する。	0:エラー無し 1:E ファイルの退院日が FF1 と異なる 2:F ファイルの退院日が FF1 と異なる 3:E ファイルと F ファイルの両方の退院日が FF1 と異なる 4:EF 統合ファイルの退院日が FF1 と異なる
22	ERR541	D ファイルの退院日が FF1 と異なる	「施設コード＋データ識別番号＋入院年月日」が FF1 と D ファイルで同一であるが、退院日が異なる。	FF1,D ファイル	フラグ	・親様式 1（診療情報番号＝0）のみチェック対象。 ・子様式 1（診療情報番号≧1）はチェック対象外。フラグは 0 を設定する。	0:エラー無し 1:D ファイルの退院日が FF1 と異なる

23	ERR550	入院基本料または特定入院料を算定しない日がある	F ファイルでデータ区分 90(入院基本料)または 92(特定入院料)の点数がない日がある。	E ファイル	フラグ	・親様式1(診療情報番号＝0)のみチェック対象。 ・子様式1(診療情報番号≧1)はチェック対象外。フラグは 0 を設定する。	0:エラー無し 1:入院基本料または特定入院料を算定した前に、入院基本料または特定入院料を算定しない日がある(ただし、2 のケースを除く) 2:入院基本料または特定入院料を算定した後に、退院前に入院基本料または特定入院料を算定しない日がある
24	ERR610	医科レセプトのみでない	FF4 で「1.医科レセプトのみ」以外	FF4	フラグ	・FF4 にレコードがない場合もフラグ設定する。 ・親様式1(診療情報番号＝0)のみチェック対象。 ・子様式1(診療情報番号≧1)はチェック対象外。フラグは 0 を設定する。	0:エラー無し(医科レセプトのみ) 2:歯科レセプトあり 3:保険請求なし 4:保険と他制度の併用 5:その他 9:FF4 に対応レコード無し

25	ERR710	24 時間以内の死亡である	FF1 で「24 時間以内死亡の有無」が「24 時間以内死亡の有り」または「救急患者として搬送され、入院前に処置室、手術室等で死亡有り」	FF1	フラグ		0:エラー無し(24 時間以内死亡無し) 1:24 時間以内死亡有り 2:救急患者として搬送され、入院前に処置室、手術室等で死亡有り
26	ERR720	生後 7 日以内の死亡である	退院時転帰が死亡、かつ退院日が生年月日から 7 日以内(生年月日当日は 1 日とカウント)	FF1	フラグ		0:エラー無し(生後7日以内の死亡無し) 1:生後7日以内の死亡有り
27	ERR730	治験対象である	FF1 で「治験実施の有無」が「有り	FF1	フラグ		0:エラー無し(治験実施無し) 1 治験実施有り
28	ERR740	移植手術を実施している	F ファイルに移植手術に該当する K コードに該当するレセ電算コードを持つ。移植手術の K コードは厚労省告示に基づく。(H15 第 75 号、H18 第 138 号、H20 第 93 号、H22 第 93 号)	F ファイル	フラグ	・親様式 1(診療情報番号＝0)のみチェック対象。 ・子様式 1(診療情報番号≧1)はチェック対象外。フラグは 0 を設定する。	0:エラー無し(移植手術無し) 1:移植手術有り

29	ERR750	厚生労働大臣が定める者である	Fファイルに厚労省告示に該当する医科点数表コードを持つ。(H16 第 107 号、H18 第 139 号、H22 第 94 号、H22 第 94 号,第 96 号,第 197 号)	Fファイル	フラグ	・親様式1(診療情報番号＝0)のみチェック対象。 ・子様式1(診療情報番号≧1)はチェック対象外。フラグは 0 を設定する。	0:エラー無し(厚生労働大臣が定める者でない) 1:H16 年度の厚生労働大臣が定める者である 2:H18 年度の厚生労働大臣が定める者である 3:H20 年度の厚生労働大臣が定める者である 4:H22 年度の厚生労働大臣が定める者である
30	ERR760	一般病棟外への移動がある	FF1 で「一般病棟外への移動あり」がある(～H16)。「精神病棟への入院あり」または「その他の病棟への入院有り」がある(H17～)。	FF1	フラグ		0:エラー無し(一般病棟外への入院無し) 1:一般病棟外への入院有り
31	ERR770	年齢が 120 歳以上である	入院時年齢が 120 歳以上	FF1	フラグ		0:エラー無し(入院時年齢 120 歳未満) 1:入院時年齢 120 歳以上

32	ERR780	手術が輸血のみである	FF1 の手術1～5に輸血だけしか存在しない （輸血管理料は含まないこと）	FF1	フラグ		0:エラー無し（手術なし、または、輸血以外の手術あり） 1:手術が輸血のみ
33	ERR790	短期滞在手術基本料を算定している	E ファイルに短期滞在手術基本料に該当するレセ電コードがある。 '190076710','190076810','190125310','190130410','190130510'	E ファイル	フラグ		0:エラー無し（短期滞在手術基本料なし） 1:短期滞在手術基本料あり

D.考察

本研究の結果から、提出されているデータについて、一定数のエラーが含まれていることがわかった。特にキー情報の重複や必要なデータが欠損している症例などは、分析に影響度が大きいので、それを確実に除去できるようになったことは成果である。

また、分析の方向性によって、使用の可否が決まる入院患者レコードがあるということが本研究の過程で判明した。たとえば、入院の途中で保険適応になる患者について、それを分析に含めるか否か、その場合の在院日数はどの範囲を指すべきなのか、などといった点は、研究の目的に応じて、それを研究者が容易に判断できる環境にあることは結果の妥当性を維持するためにも重要である。今回の研究で、留意が必要な入院レコードに対してそのフラグメント化ができたことは大きな成果であったといえる。今後の開発においてもさらなる留意コードが必要かについてその使い勝手とともに検討していきたい。

現時点ではいまだにいくつかのテーブルについては一般的な研究者の持つ環境ではハンドルが難しいサイズのレコード数を持っている状況にある。今後、このデータセットを使って、研究者が共通して使える集計データを作成し、より容易に研究ができるデータ環境を構築していく必要がある。この点が来年度の課題である。

E.結論

本年度、平成24年度分までの分析用データセットの作成が完了し、運用することもできた。

次年度以降、より容易に研究ができるデータ環境を構築していく研究を行っていきたい。

平成 25 年厚生労働科学研究費補助金(政策科学総合研究事業)

ナイーブベイズ分類による副傷病を用いた在院日数の推定

神経系疾患領域（MDC01）の DPC 分類に関する検討

藤野 善久	産業医科大学医学部	准教授
村松 圭司	産業医科大学医学部	専修医
久保 達彦	産業医科大学医学部	講師
村上 玄樹	産業医科大学病院	講師
松田 晋哉	産業医科大学医学部	教授

研究目的： 現在の DPC 分類と併用可能な、副傷病を用いた医療資源必要度分類のための方法として、ナイーブベイズ分類による副傷病を用いた在院日数の推定について検討する。

分析方法： 平成 24 年度の DPC 研究班データに含まれる MDC01 に該当する 452,992 件のデータを対象とした。副傷病は ICD10 の 3 桁までの情報を用いた。在院日数を 50,75,90 パーセンタイルで 4 分割したものを実際の在院日数分類とした。ナイーブベイズ分類の応用である、Complement Naive Bayes(CNB)を用いて、在院日数分類を推定した。

結果： ナイーブベイズ分類による在院日数分類の推定精度は、実際の在院日数の分類との重み付け一致率は約 80%～90%であった。また、推定された在院日数分類は、在位日数の平均を反映していた。

考察： ナイーブベイズ分類を用いた副傷病による医療資源必要度分類は、副傷病を選択したりグルーピングしたりする処理が必要なく、全ての MDC 群において適用可能な方法である。特に、DPC データでは、一人の患者が複数の副傷病を持ち、かつ、データ上に出現する副傷病 ICD コードが 600～800 あるため、次元の呪いによる問題を抱えている。今後は、これらの手法と比較しながら、実応用に関する長所・短所を検討する必要がある。

A. 研究目的

同一 DPC において、重症度の違いを反映させ、必要な医療資源の分類を精緻化する試みとして、Comorbidity Complication

Procedure (CCP) Matrix の必要性が提唱されている。

発想としては、現在の DPC コードの分岐とは異なる情報を用いて、重症度を反映することである。そこで、重症度を評価するために、副傷病による情報の活用が考えられる。

しかしながら、副傷病を用いた重症度の評価にはいくつかの課題がある。まず、ICD10 においては、3 桁分類でも約 2000 項目がある。このように多数の項目から、どの疾病が重症度（医療資源の消費）に影響しているかを検討するような場合、いわゆる次元の呪い、や p>>n 問題と呼ばれる統計学上の課題があげられる。また、このように特定の副傷病を選択する手法は、DPC ごとや、MDC ごとに実施しなければならず、実務上、膨大な作業を必要とする。また、疾病を選択した場合、大多数の選択されなかった副傷病について配慮されないという、臨床現場の実情との乖離も問題となる。

本研究では、副傷病による重症度（医療資源の消費）を評価する方法の一つとして、ナイーブベイズ分類による方法を提案し、検証を行う。

B. 分析方法

（１）対象データ

平成 24 年度の DPC 研究班データ（以下、研究班データ）を用いた。今回は、検討のため MDC01 に該当する 452,992 件のデータを対象とした（図 1　神経系疾患（DPC01）の DPC6 桁コード別頻度）。

（２）分析方法

在院日数(length of stay, 以下 LOS)について、50、75, 90 パーセンタイルで 4 分割を行った（図 2　在院日数による分類）。この LOS を 4 つのクラスに分類した変数を Los4 とする。(Los4: 0,1,2,3)。すなわち、この Los4 を、副傷病を用いて推測することが今回の目的である。

本研究では、LOS クラス(los4)を副傷病によって分類するためにナイーブベイズ分類を用いた。実際には、ナイーブベイズ分類の応用である、Complement Naive Bayes(CNB)を用いた。これは、補集合を用いて推測することで、通常のナイーブベイズ分類よりも、分類するクラスが３つ以上ある場合に有効とされている。

CNB の計算過程において、以下の処理を加えた。

① zero cell 問題：副傷病が１回も出現しない LOS 分類が存在する場合、P(副傷病|分類)=0 となり、事後確率が常にゼロとなってしまう。これを避けるために、全セルに１を加える Laplace smoothing を実施した。

② underflow の問題：複数の副傷病による尤度を掛け合わせると、小数点以下限りなくゼロに近い数字となり、計算用のソフトなどで正しく対応できない場合が生じる。Map(maximum a posterior)を得るためには、大小関係が判別できれば良いので、対数変換を行う。Underflow の問題とは異なるが、行列計算を実行する際にも計算が簡易になるメリットもある。

（３）検証方法

（３．１）推定の安定性（K-fold cross-validation）

学習に利用したデータを用いて推計した場合は、推定が過適合することがある。そのためデータを無作為に分割し、学習用データと検証用データに分けて検証を実施する。今回は、データを無作為に 20 分割し、20 個のデータセットを作成した。このうち 19 個のデータセットを用いて学習させ、残り１個のデータで推測の精度を検証する。この作業を、データセットを入れ替えながら 20 回施行する。このような検証を k(20)-fold cross-validation 法と呼ぶ。

実際には、452,992 件のデータのうちランダムに 450,000 件を抽出し、それをさらに 20 分割して用いた。

（３．２）分類の正答率の検証

検証用データを１つ用いて、推定された LOS 分類と、実際の LOS 分類の正答率を検証した。全体での検証、疾患別による検証、副傷病の数別による検証を行った。

（３．３）LOS の平均の検証

検証用データを１つ用いて、推定された LOS 分類別の LOS 平均を検証した。全体での検証、疾患別による検証、副傷病の保有数別による検証、頻度の多い副傷病別による検証を行った。

C．結果

１）推定の安定性（K-fold cross-validation）の検証結果

20 個に分割したデータを用いた cross-validation の結果を示す（図 3 cross-validation 結果）。粗一致率の平均は 49%（0.49）であり、SE は 0.003 であった。平均に対して、SE が極めて低いことからも、推定の安定性は robust であると言える。

２）分類の正答率の検証結果

20 分割したデータ・セットのうち、19 個をトレーニング用、残りの 1 個を検証用データとして検証した結果を示す。

Los4 が、実際の LOS の分類クラスであり、map という変数が、ベイズ推定によって副傷病から推測された LOS クラスである。全体の粗一致率は 49.5%であった（図 4　粗正答率の結果）。さらに、実際のクラスに推定値が近い場合から遠い場合まで、漸次重み付けを実施して計算した一致率は 82.0%と良好な結果を示した（図 5　重み付けの正答率）。

次に、DPC6 桁コードを用いて、脳血管、脳梗塞、てんかん、その他の４群に分類した（図 6　疾患群の分類）。それぞれの疾患群別に、正答率の一致性について検証した。

脳血管では 79%（図 7　DPC=脳血管群における正答率）、脳梗塞では 81%（図 8　DPC＝脳梗塞群における正答率）、てんかんでは 93%（図 9　DPC=てんかん群における正答率）、その他では 83%（図 10　DPC=その他の群における正答率）の重み付けスコアによる一致性を得た。

さらに、副傷病の数別に、正答率について検証を行った。当然だが、副傷病がない場合は、LOS 分類の推定に影響はなく、推定 LOS 分類は事前確率が最も高いグループに分類されている（図 11　副傷病が 0 個の

場合の正答率)。また、副傷病が１つの場合も、事後分布に与える影響はほとんどなく、副傷病によって LOS 分類の推定に影響があったのは、3604 件中の 13 件のみであった（図 12　副傷病が 1 個の場合の正答率)。

副傷病が２個、３個、４個の場合の重み付け正答率は、それぞれ、83%、81%、78%であった。(図 13　副傷病が 2 個の場合の正答率、図 14　副傷病が 3 個の場合の正答率、図 15　副傷病が 4 個の場合の正答率)

３）LOS 平均の検証結果

推定された LOS クラスが、実際の LOS をどの程度反映しているかについて、LOS の平均による検証結果を示す。

全体では、推定された LOS クラスは、平均で 26、37、50、62 日と漸増していた（図 16　推定された分類別の LOS 推定平均)。この傾向は、傾向性の検定においても有意であった。

推定された LOS のクラスと LOS の平均の検証について、さらに疾患群別の結果を示す。脳血管（図 17　DPC＝脳血管群における推定された分類別の LOS 平均)、脳梗塞（図 18　DPC＝脳梗塞群における推定された分類別の LOS 平均)、その他の群（図 19 DPC＝その他の群における推定された分類別の LOS 平均）のいずれも、推定された LOS クラスが高くなるほど、LOS 平均値も増えていた。これらは傾向性も有意であった。

てんかん群では、推定による LOS クラスが最も高い群における LOS の平均が低かった（図 20　DPC＝てんかん群における推定された分類別の LOS 平均)。これは、てんかん群のｎ数および、推定された LOS クラスが最も高い群のｎが少ないためと推測される。

推定された LOS のクラスと LOS の平均の検証について、さらに副傷病別の結果を示す。今回の検証データでは、副傷病は最大で４つまで記録されていた。付けられていた副傷病の数別に、推定された LOS 推定クラスと LOS の平均について結果を示す。

副傷病の数が１個の場合、事後確率はほとんど影響されず、推定された分類が事前分布と置換されたのは 13 件のみであり、LOS 平均の精度も高くなかった（図 21　副傷病が１個の場合:推定された分類別の LOS 平均)。

一方、副傷病の数が 2 個から４個の場合は、推定された分類が高くなるほど、LOS 平均も概ね大きくなっていた（図 22 副傷病が２個の場合：推定された分類別の LOS 平均、図 23 副傷病が３個の場合：推定された分類別の LOS 平均、図 24 副傷病が４個の場合：推定された分類別の LOS 平均)。

頻度が 50 以上の副傷病を頻度順に示す（図 25　頻度順副傷病（度数５０以上))。頻度が多い副傷病を上位から１０個選び、副傷病別に、推定された LOS 分類と LOS 平均を示す。

概ね、推定された分類が大きいクラスほど、LOS 平均が大きくなっていた。ただし、脳血管疾患の続発・後遺症(図 31　I69　脳血管疾患の続発・後遺症における推定 LOS 分類と LOS 平均)や脳梗塞(図 33　I63　脳梗塞における推定 LOS 分類と LOS 平均）のように、副傷病による分類があまり影響していない場合は、LOS 平均を必ずしも反映していなかった。

D．考察