

野球のトラッキングデータに基づいた 肘内側側副靭帯損傷の要因解析

酒折 文武¹・圓城寺 啓人²・竹森 悠渡²・
西塚 真太郎²・保科 架風³

(受付 2017 年 1 月 31 日；改訂 3 月 22 日；採択 3 月 30 日)

要 旨

野球の投手における肘内側側副靭帯の損傷は近年増加しており、大きな問題となっている。予防の重要性にもかかわらず、そのリスク要因に関する科学的なコンセンサスが得られているとは言い難い。そこで本論文では、アマチュア野球の投手経験者とスポーツドクターの意見を参考にして、肘内側側副靭帯損傷のリスク要因の候補を再検討した。そして、先発投手とリリーフ投手とに層別してそれぞれロジスティック回帰モデルを立て、AIC を用いた変数選択により選択されたリスク要因について、調整オッズ比を算出した。その結果、先発投手については、球種数が少ないこと、リリース位置が体から横に離れていること、1 試合当たりの投球数が多いことがリスク要因であることがわかった。またリリーフ投手に関しては、球種数が少ないこと、リリース位置が体から横に離れていること、ファストボールの球速が速いこと、登板間隔が短いことがリスク要因であることがわかった。これらの結果は、他の研究成果の一部を肯定しているとともに、先発投手やリリーフ投手における 1 試合の投球数や登板間隔に関する重要な示唆を与えているといえる。

キーワード：オッズ比，ロジスティック回帰，スパースロジスティック回帰，Lasso。

1. はじめに

野球の投手における肘内側側副靭帯の損傷は近年増加しており、大きな問題となっている。それに伴い、肘内側側副靭帯再建手術、いわゆるトミー・ジョン手術がとりわけ米国のメジャーリーグ(MLB)では一般的となっている。我々日本人にも馴染みの深い、テキサス・レンジャーズに所属するダルビッシュ有投手も、2015 年にトミー・ジョン手術を受け 1 年以上の欠場を余儀なくされたことは記憶に新しい。

肘内側側副靭帯損傷の予防は非常に重要である一方で、そのリスク要因に関する科学的なコンセンサスが得られているとは言い難い(Whiteside et al., 2016)。例えば、アメリカスポーツ医学研究所は若年時からの蓄積によって故障が引き起こされると述べており、日本の全国高等学校野球選手権大会等での連戦連投、多投へのアメリカでの批判は多い。また、前述のダル

¹ 中央大学 理工学部：〒 112-8551 東京都文京区春日 1-13-27

² 中央大学大学院 理工学研究科：〒 112-8551 東京都文京区春日 1-13-27

³ 滋賀大学 データサイエンス教育研究センター：〒 522-8522 滋賀県彦根市馬場 1-1-1

ビッシュ有投手をはじめとする日本人選手や野球関係者の間では、1試合での投球数ではなく登板間隔が問題であり、MLBにおける先発投手の中4日での登板が故障の大きな原因であるとの意見がある。さらに、青少年に対する研究ではあるが、1試合の投球数やシーズンでの投球数が大きなリスク要因との指摘もある(Lyman et al., 2002)。その他、Olsen et al. (2006)では、青少年の投手に対する調査の結果をロジスティック回帰モデルにより分析し、怪我(多くは肘)のリスク要因として1年間に競技として投球する月数が多いこと、1試合あたりの投球数が多いこと、ファストボール(速球。球速の速い投球のこと)の球速が速いこと、疲労時に投球することが多いこと、の4つを挙げている。トミー・ジョン手術に限定しても、Keller et al. (2016)では、トミー・ジョン手術を受けた投手83名について、対照群83名との比較において、手術2年前の投球における各球種の球速には2群に有意差がなかったが、ファストボールの投球割合は手術群のほうが有意に高いと報告しており、ファストボールを多く投げる投手が肘内側側副靭帯損傷となる傾向があることを示唆している。

これらの研究に対して、Whiteside et al. (2016)では、近年の計測技術の向上により得られるようになった、投球に関する詳細な変数を用いてリスク要因を探っている。彼らは、手術群および対照群各104名の投手について、肘内側側副靭帯損傷のリスク要因となりうるデモグラフィック変数、セイバーメトリクス(野球における数理科学的なアプローチ)の指標FIPに加え、PITCHf/xという、球場に設置されたカメラの映像から、球速、球種、手からボールが離れたリリースポイントの座標、本塁を通過したときの座標、変化量、スピン量などを算出するシステムによる変数を用いてロジスティック回帰モデルによる分析を行っている。その結果、登板間隔が短い、投げる球種が少ない、リリースポイントの横位置が小さい(体に近い)、身長が低い、平均球速が速い、1試合あたりの投球数が多い、という6つをリスク要因として挙げている。しかしながら、モデルの立て方やリスク要因の候補の選定には一考の余地が残されている。例えば、投手の役割(先発投手かリリーフ投手か)によってこれらの要因の影響は大きく異なることが想定されるにもかかわらず、手術群と対照群のマッチングのみに使用されている。また、球速に関しては、球種の違いを考慮せずに全投球の平均球速のみしか使用していない。そして何よりも、リリースポイントが体に近いことは、Aguinaldo and Chambers (2009)などの他の研究と矛盾した結論である。

そこで本論文では、肘内側側副靭帯損傷のリスク要因についてさらなる検討を行った。まず、日本のアマチュア野球での投手経験者、スポーツドクター、トレーナーの意見を参考にし、改めてリスク要因の候補を再検討した。さらに、先発投手とリリーフ投手とに層別し、それぞれロジスティック回帰モデルを立て、AICを用いた変数選択により選択されたリスク要因の調整オッズ比を算出した。その結果、先発投手については、球種数が少ないこと、リリース位置が体から横に離れていること、1試合あたりの投球数が多いことがリスク要因であることがわかった。リリーフ投手に関しては、球種数が少ないこと、リリース位置が体から横に離れていること、ファストボールの球速が速いこと、登板間隔が短いことがリスク要因であることがわかった。これらの結果は、他の研究成果の一部を肯定しているとともに、先発投手やリリーフ投手における1試合の投球数や登板間隔に関する重要な示唆を与えているといえる。

本論文の構成は以下のとおりである。2節では分析に用いたデータとモデルについて説明する。3節では詳細な分析結果を述べる。それを受けて4節ではまとめと考察、そして今後の課題について述べる。

2. 研究手法

本論文では、2012年から2016年の間に肘内側側副靭帯損傷によるトミー・ジョン手術を行っ

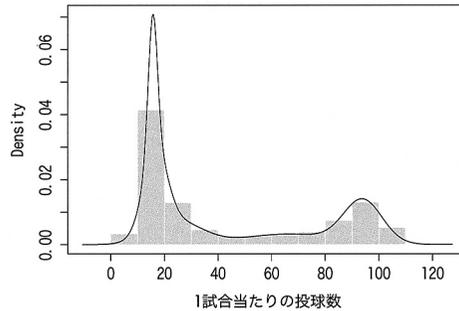


図 1. 全投手の 1 試合当たりの投球数の分布.

た投手(手術群)と、損傷していない(トミー・ジョン手術を行っていない)投手(対照群)に対して、ケースコントロール研究によって肘内側副靭帯損傷の要因分析を行った。手術同年あるいは前年のように故障時期に近いデータを用いた場合、故障の兆候がすでに投球に現れている可能性がある。一方で、あまり過去のシーズンのデータを用いても、故障への影響が不明でありノイズを多く含んでしまい、さらに投手の投球スタイル自体が変化している可能性もある。そこで、手術群については、手術年の 2 年前のシーズンのデータを用いることとした。また、対照群は対応する手術群投手と同年のシーズンのデータを用いた。

2.1 手術群

Baseball Heat Maps (2016) の Disabled List Data を参照し、2012 年から 2016 年の間に肘内側副靭帯損傷によるトミー・ジョン手術を行った投手の中で、前々年の全投球数が 100 球以上の投手全 74 名を手術群とした。なお、複数回手術を行った投手は最初に手術を行った年を選んだ。

2.2 1 試合あたり投球数での分類

先発投手やロングリリーフ投手などの長いイニングを投げる投手と、中継ぎ投手や抑え投手のような短いイニングを投げる投手では、リスク要因が異なることが想定される。そこで、1 試合あたり投球数の大小によって 2 群にわけられるため、混合正規分布モデルによるクラスタリングを行った。

まず、2010 年～2014 年に投球した全投手における 1 試合当たりの投球数を確認したところ、図 1 のヒストグラムのように多峰型の分布であり、15 球と 95 球あたりに明確にピークがあることがわかった。そこで、これに混合正規分布を当てはめることとし、特異モデルである混合正規分布への適用については諸論があるところだが、BIC を用いてモデル選択を行った。各クラスターのサイズはすべて異なるとし、クラスター数(混合要素数)は 1 から 9 まで、各クラスターが等分散のモデル(E)と不等分散のモデル(V)を候補とした。各モデルの BIC の値は図 2 のようになり、5 クラスで不等分散のモデルが選択された。このモデルにおける各クラスターのパラメータの推定値は表 1 の通りである。

この結果を踏まえ、クラスター 1, 2, 3 が中継ぎ投手や抑え投手(以下リリーフ投手)、クラスター 4, 5 が先発投手やロングリリーフ投手(以下先発投手)と考え、クラスター 3 と 4 の分類の境界値、すなわち両クラスターへの帰属確率が等しくなる点を解析的に求め、この値(42.34)を超える選手を先発投手、それ以外をリリーフ投手と定めた。この基準をもとに上述の 74 名を先発とリリーフに振り分けたところ、ちょうど 37 名ずつであった。

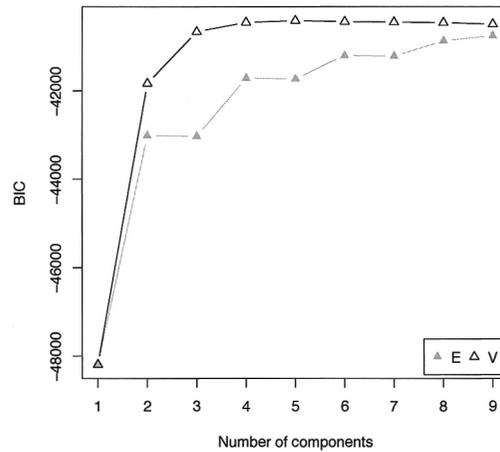


図 2. 混合正規分布モデルの BIC.

表 1. 混合正規分布モデリングの結果(混合要素数 5).

	クラスター 1	クラスター 2	クラスター 3	クラスター 4	クラスター 5
サイズ	0.179	0.276	0.167	0.140	0.238
平均	15.67	16.49	24.62	66.08	94.00
分散	2.933	19.428	91.547	299.907	51.538

2.3 対照群

先発およびリリーフそれぞれについて、選択した手術群の各投手に対して、同年に投球した同じ利き腕の投手 1 名ずつを対照群の投手として選定した。

選定にあたっては、過去に一度もトミー・ジョン手術を行っていない投手を対象とした。また、同姓同名の選手がおり、これらの選手は抽出対象から除外した。その上で、登板数や総投球数、投球スタイルなどの調べたい要因以外に故障に影響を与える可能性がある、体格や年齢・経験年数などをコントロールするために

- 手術前々年のシーズンにおける年齢
- 身長
- 体重
- メジャー経験年数
- 投球イニング

という 5 変数を Baseball Reference (2016) のウェブサイトから取得し、手術群の各投手と標準化ユークリッド距離が最も近い投手を選んだ。投球イニングを加えたのは、投球イニング自体の影響を除いた上で、1 試合の投球数と登板間隔のどちらが故障に影響を与えるのを見るためである。なお、対照群の投手で同じ投手が選ばれた場合には、年度が違った場合はそのまま使用し、年度まで同じ場合は次に距離の近い選手を選んだ。また、2.4 節での分析における欠測を避けるため、フォーシーム・ツーシーム・カットボールなどファストボールを 1 球も投げていない投手も除外した。以上の流れにより、先発とリリーフ合わせて各 74 名の手術群と対照群の投手をリストアップした。

表 2. 説明変数一覧.

x_1	登板間隔の中央値 (日)
x_2	登板数
x_3	1 試合当たりの平均投球数
x_4	球種数
x_5	ファストボールの投球割合
x_6	カーブの投球割合
x_7	チェンジアップの投球割合
x_8	ファストボールの平均球速 (マイル毎時)
x_9	ファストボールの最大球速と平均球速の差 (マイル毎時)
x_{10}	ファストボールの横方向の変化量の平均 (インチ)
x_{11}	ファストボールの縦方向の変化量の平均 (インチ)
x_{12}	ファストボールの回転数の平均
x_{13}	リリース位置の横座標の平均 (ライン)
x_{14}	リリース位置の横座標の標準偏差 (ライン)
x_{15}	リリース位置の縦座標の平均 (ライン)
x_{16}	リリース位置の縦座標の標準偏差 (ライン)

2.4 変数

リスク要因の検討には、まず、先行研究である Whiteside et al. (2016) で用いている変数を参考にした。その上で、日本のアマチュア野球の経験者と、スポーツドクター、トレーナーへのインタビューを行い、以下のような意見を得た：

- 登板数や 1 試合当たりの投球数が多い投手ほど、疲労が蓄積されるため怪我に繋がるのではないか。
- 疲労によって体が開き肘が下がること、逆に肘の位置が高すぎるにより負担がかかるため、リリース位置も考慮すべきである。
- 中 4 日の登板間隔が短いというダルビッシュ有投手らの意見は尤もであり、短い登板間隔で疲労回復ができていないのかは疑問が残る。
- 球種によってボールの握りや肘の捻りが様々なので、球種数が多いと怪我しやすい可能性がある。
- 力を入れて投球すれば、きれいな投げ方でも肘に負担がかかり怪我に繋がると考えられるため、ファストボールの球速や、最大球速との球速差などを考慮すべきである。
- カーブ、チェンジアップなどの変化球も肘に負担がかかると考えられるので含めるべきである。

これらを踏まえて、2.1 節で抽出した手術群と対照群各 74 名、計 148 の投手に対し、肘内側副韌帯損傷の要因と考えられる説明変数として表 2 にまとめた 16 変数を用いた。いずれも PITCHf/x データから算出した。本研究では R の pitchRx パッケージ (Sievert, 2014, 2015) を用いて PITCHf/x データを取得した。以下、これらの 16 変数について説明する。

まず、日本のプロ野球よりもトミー・ジョン手術が頻繁に行われている MLB では、中 4 日での登板が主流であり、登板間隔が短いことが怪我の原因であるという意見が聞かれる。そこで、登板間隔に関する変数として、登板間隔(日)の中央値 x_1 と、登板数 x_2 を用いた。 x_1 で中央値を用いた理由は、故障者リスト(DL)入りやマイナー落ちなどの影響を除くためである。次に、投球数や登板回数の多い投手ほど怪我につながりやすいと考えられるため、1 試合当たりの平均投球数 x_3 を用いた。そして、球種が少ないほど故障しやすいという Whiteside et

al. (2016) の指摘を受けて、投球した球種数(最大で 10 種類) x_4 を用いた。さらに、各球種の影響を見るために、フォーシーム・ツーシーム・カットボールなどを合わせたファストボール、カーブ、チェンジアップの 3 球種の投球割合 x_5, x_6, x_7 を用いた。スライダの投球割合も検討したが、ファストボールの投球割合 x_5 との相関が -0.9 程度と非常に強かったため変数には加えなかった。

また, Keller et al. (2016) で指摘されるファストボールの影響をさらに細かく見るため, ファストボールの平均球速(マイル毎時) x_8 , 最大球速と平均球速の差(マイル毎時) x_9 , 横方向および縦方向の変化量の平均(インチ) x_{10}, x_{11} , 回転数の平均 x_{12} を用いた。投球フォームによる影響を考慮するため, 投球時にボールをリリースした位置の横座標, 縦座標の平均(ライン = $1/12$ マイル) x_{13}, x_{15} をそれぞれ用いた。加えて, 疲労による体の開きやそれに伴う肘下がりなどによる平均的なリリース位置からのずれの具合を測るために, リリース位置の横方向, 縦方向それぞれの標準偏差(ライン) x_{14}, x_{16} を用いた。カーブやチェンジアップについても同様の変数群を加えたいところであるが, そもそもこれらの球種を投げない投手も少なくないため, 今回は変数に加えなかった。

2.5 分析方法

上記のようにして選んだりリスク要因の候補全 16 変数を説明変数とし, トミー・ジョン手術の有無(手術群 or 対照群)を目的変数として, 先発とリリーフとに分けてそれぞれロジスティック回帰モデル

$$(2.1) \quad \text{logit}(p_i) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_{16} x_{i16} = \beta^T \mathbf{x}_i$$

により各説明変数による故障のリスクに関する分析を行った。最尤法により推定した係数 $\hat{\beta}_j$ および正規近似による β_j の信頼区間 $[l_j, u_j]$ から, 各説明変数の調整オッズ比 $e^{\hat{\beta}_j}$, および母オッズ比の 95% 信頼区間 $[e^{l_j}, e^{u_j}]$ を求めた。

説明変数の数が 16 と多いため, AIC を基準とするステップワイズ法(変数増減法)によって変数選択を行い, 選ばれた変数に対する調整オッズ比と 95% 信頼区間を求めた。

これらの 2 つのロジスティック回帰モデルでは判別を目的とはしていないが, 参考のために確率 0.5 を境界としたときの判別結果と正判別率を求めた。また, モデル M の適合度を評価するために, スルモデル M_0

$$(2.2) \quad \text{logit}(p_i) = \beta_0$$

との尤度比検定の結果と, Nagelkerke の疑似決定係数(Nagelkerke, 1991)

$$(2.3) \quad R^2 = \frac{1 - \left\{ \frac{L_{M_0}(\hat{\beta}_0)}{L_M(\hat{\beta})} \right\}^{2/n}}{1 - L_{M_0}(\hat{\beta}_0)^{2/n}}$$

の値を求めた。ここで, $L_M(\beta)$ はモデル M の下での尤度, ハットはそのモデルの下での最尤推定量, n はサンプルサイズである。後者は重回帰における決定係数の考え方を一般化し, 上限が 1 となるように調整したものである。

AIC を用いた変数選択では, 説明変数の数が多いため総当たり法ではなくステップワイズ法を用いたが, 別のアプローチとして考えられるのは L_1 正則化法である。 L_1 正則化法を用いることにより, 実質的に L_0 での正則化と考えられる総当たり法よりも高速に分析を行うことができる。ここでは, モデル(2.1)に Lasso (Friedman et al., 2010; Hastie et al., 2009)による L_1 正則化を加えたスパースロジスティック回帰モデルによる分析も合わせて行った。Lasso による

表 3. 手術群と対照群における 5 変数に関する平均の差の検定. 上：先発投手, 下：リリーフ投手.

	手術群	対照群	t 値	P 値
年齢	25.62 ± 3.29	26.00 ± 2.76	-0.536	0.59
身長	190.50 ± 5.22	190.50 ± 4.90	0.000	1.00
体重	98.93 ± 8.15	98.20 ± 7.89	0.394	0.69
経験年数	4.19 ± 3.47	4.24 ± 3.16	-0.070	0.94
イニング	117.07 ± 65.95	118.28 ± 70.77	-0.076	0.94

	手術群	対照群	t 値	P 値
年齢	27.22 ± 3.61	27.00 ± 3.06	0.278	0.78
身長	189.54 ± 5.78	189.95 ± 5.59	-0.312	0.76
体重	99.36 ± 9.80	98.38 ± 8.24	0.466	0.64
経験年数	3.97 ± 2.69	4.14 ± 2.51	-0.268	0.79
イニング	46.56 ± 20.61	46.37 ± 22.17	0.038	0.97

推定量は

$$(2.4) \quad \hat{\beta}_{\text{lasso}} = \arg \min_{\beta} \left\{ -\frac{1}{n} \log L(\beta) + \lambda \sum_{j=1}^{16} |\beta_j| \right\}$$

により与えられ, 正則化パラメータ λ の値によっては縮小推定とともにいくつかの係数が正確に 0 と推定され, 変数選択の役割も果たす. ここでは, λ の選択には 10-fold クロスバリデーションを用いた.

3. 分析結果

分析結果は以下の通りである.

3.1 対照群選定の妥当性の確認

対照群の投手を選定する際に用いた 5 変数が手術群と対照群において差がないことを確認するため, 変数ごとに等分散を仮定した平均の差の検定を行った. P 値は表 3 の通りとなり, 各変数に有意な差がないことがわかった. なお, これら 5 変数は手術群と対照群とで調整されているため, 以降の分析には用いていない.

3.2 各リスク要因の分布

リスク要因の候補となる各説明変数の要約統計量は表 4 の通りである. 先発投手とリリーフ投手では当然ながら, 登板間隔 x_1 , 登板数 x_2 , 1 試合当たり投球数 x_3 が大きく異なることがわかる. さらに, 球種数 x_4 , 変化球の投球割合 x_6, x_7 , ファストボールの球速 x_8 , 球速差 x_9 , 横変化量 x_{10} , 回転数 x_{12} , リリース位置 x_{13}, x_{15} などほとんどの変数においても先発投手とリリーフ投手に差がみられることがわかる. また, 手術群と対照群を比較すると, 先発投手ではリリース位置横 x_{13} 以外には大きな差が認められないのに対し, リリーフ投手では登板間隔 x_1 , 1 試合当たり投球数 x_3 , ファストボール球速 x_8 , リリース位置横 x_{13} などに差がみられる.

なお, 説明変数間の相関はほとんど無く, 相関係数の絶対値が 0.5 を超えるものは先発投手では登板数 x_2 と球種数 x_4 (0.53), ファストボールの縦変化量 x_{11} と回転数 x_{12} (0.59), リリー

表 4. 各リスク要因の要約統計量(平均 ± 標準偏差).

変数	先発投手 ($n = 74$)		リリーフ投手 ($n = 74$)	
	手術群 ($n = 37$)	対照群 ($n = 37$)	手術群 ($n = 37$)	対照群 ($n = 37$)
登板間隔 x_1	5.49 ± 0.49	5.62 ± 2.68	2.57 ± 0.85	3.41 ± 1.36
登板数 x_2	21.57 ± 10.48	22.14 ± 11.73	51.32 ± 23.17	46.35 ± 23.23
1 試合当たり投球数 x_3	90.23 ± 12.31	86.63 ± 17.53	16.87 ± 5.39	19.19 ± 7.27
球種数 x_4	6.49 ± 1.15	6.73 ± 1.48	5.49 ± 1.15	5.97 ± 1.09
ファストボール割合 x_5	0.53 ± 0.21	0.49 ± 0.22	0.54 ± 0.24	0.48 ± 0.24
カーブ割合 x_6	0.08 ± 0.08	0.08 ± 0.08	0.04 ± 0.09	0.06 ± 0.10
チェンジアップ割合 x_7	0.11 ± 0.08	0.12 ± 0.08	0.06 ± 0.06	0.09 ± 0.08
ファストボール球速 x_8	91.15 ± 2.25	91.14 ± 2.13	93.32 ± 2.45	92.02 ± 2.91
ファストボール球速差 x_9	3.88 ± 1.05	3.87 ± 1.12	3.50 ± 1.38	3.27 ± 1.08
ファストボール横変化量 x_{10}	-1.91 ± 5.45	-2.08 ± 5.15	-1.88 ± 5.63	-0.84 ± 6.14
ファストボール縦変化量 x_{11}	7.93 ± 1.64	7.8 ± 1.77	7.75 ± 2.72	7.92 ± 2.03
ファストボール回転数 x_{12}	2028.47 ± 351.28	2013.13 ± 260.14	2126.23 ± 276.44	2092.68 ± 297.8
リリース位置横 x_{13}	22.77 ± 5.55	19.64 ± 6.82	24.15 ± 8.49	20.68 ± 8.67
リリース位置縦 x_{14}	71.6 ± 3.91	72.62 ± 4.37	70.3 ± 4.08	70.46 ± 5.41
リリース位置横ばらつき x_{15}	3.14 ± 1.05	3.15 ± 0.90	3.67 ± 1.37	3.78 ± 1.44
リリース位置縦ばらつき x_{16}	2.52 ± 0.70	2.42 ± 0.49	2.65 ± 0.95	2.52 ± 0.84

ス位置の横のばらつき x_{15} と縦のばらつき x_{16} (0.55), リリーフ投手では, 登板間隔 x_1 と登板数 x_2 (-0.62), 登板間隔 x_1 と 1 試合当たり投球数 x_3 (0.60), 登板数 x_2 と 1 試合当たり投球数 x_3 (-0.52), ファストボールの球速 x_8 と横変化量 x_{10} (-0.58)のみであった.

3.3 各リスク要因のオッズ比

表 5 は, リスク要因ごとに 1 変量のロジスティック回帰分析を行い, 各変数のオッズ比を算出したものである. 係数の分布の正規近似により求めた 95% 信頼区間を併記した.

登板間隔 x_1 , 1 試合当たり投球数 x_3 , カーブの投球割合 x_6 については, 先発投手とリリーフ投手でオッズ比が大きく異なることがわかる. しかし, 多くの変数においては差がそれほど見られず, さらにオッズ比の信頼区間が 1 を含んでいることもわかる.

なお, ファストボール等の投球割合 x_5, x_6, x_7 は 0 から 1 のみを取りうるため, オッズ比の値自身は解釈しにくい. 逆に, 2000 程度の大きな値を取るファストボール回転数 x_{12} も同様である.

3.4 ロジスティック回帰の結果と調整オッズ比

全変数を用いたロジスティック回帰分析の結果と, 各変数の調整オッズ比および信頼区間を示したものが表 6 である.

5% 有意となった変数は, 先発投手においてはリリース位置横 x_{13} のみ, リリーフ投手については登板間隔 x_1 のみであった. 説明変数間の相関関係の影響から, 球種数 x_4 , リリース位置縦ばらつき x_{16} の調整オッズ比の値が元のオッズ比とは大きく異なっている. また, 各球種の投球割合 x_5, x_6, x_7 のオッズ比についても大きく異なっているが, これらの値は前述の通り非常に小さな値をとるため解釈しづらい.

リリーフ投手においても, 各球種の投球割合 x_5, x_6, x_7 を除いては, 登板間隔 x_1 , 1 試合当たり投球数 x_3 など他変数との相関が比較的強い変数ではオッズ比と調整オッズ比は多少異なるが, ファストボール球速差 x_9 やリリース位置横ばらつき x_{15} などのその他の一部の変数に

表 5. 各リスク要因のオッズ比.

変数	先発投手 ($n = 74$)		リリーフ投手 ($n = 74$)	
	オッズ比	95%信頼区間	オッズ比	95%信頼区間
登板間隔 x_1	0.9625	[0.6981, 1.2511]	0.4519	[0.2391, 0.7532]
登板数 x_2	0.9953	[0.9543, 1.0377]	1.0095	[0.9896, 1.0304]
1 試合当たり投球数 x_3	1.0164	[0.9858, 1.0503]	0.9411	[0.8620, 1.0139]
球種数 x_4	0.8672	[0.6008, 1.2293]	0.6739	[0.4317, 1.0193]
ファストボール割合 x_5	2.2649	[0.2558, 22.0998]	2.8578	[0.4318, 20.9748]
カーブ割合 x_6	1.9658	[0.0045, 899.2441]	0.1232	[0.0005, 16.7727]
チェンジアップ割合 x_7	0.1602	[0.0004, 53.2924]	0.0017	[0.0000, 0.9381]
ファストボール球速 x_8	1.0009	[0.8087, 1.2390]	1.2007	[1.0093, 1.4536]
ファストボール球速差 x_9	1.0105	[0.6559, 1.5588]	1.1726	[0.8059, 1.7657]
ファストボール横変化量 x_{10}	1.0064	[0.9217, 1.0996]	0.9696	[0.8947, 1.0486]
ファストボール縦変化量 x_{11}	1.0446	[0.7962, 1.3750]	0.9706	[0.7907, 1.1811]
ファストボール回転数 x_{12}	1.0002	[0.9987, 1.0017]	1.0004	[0.9988, 1.0021]
リリース位置横 x_{13}	1.0872	[1.0075, 1.1838]	1.0495	[0.9942, 1.1130]
リリース位置縦 x_{14}	0.9406	[0.8353, 1.0516]	0.9928	[0.8993, 1.0947]
リリース位置横ばらつき x_{15}	0.9869	[0.6063, 1.6016]	0.9453	[0.6664, 1.3206]
リリース位置縦ばらつき x_{16}	1.3689	[0.6309, 3.2856]	1.1808	[0.6997, 2.1017]

おいても差が見られる。

また、AIC を基準としたステップワイズ法(変数増減法)によって変数選択を行った結果が表 7 であり、確率 0.5 を境界としたときの判別結果や正判別率と、尤度比検定の結果、疑似決定係数の値が表 8 にまとめられている。

先発投手に関しては、1 試合当たり投球数 x_3 、球種数 x_4 、リリース位置横 x_{13} という 3 変数が選択された。まず、球種数が少ないほど故障しやすく、球種が 1 つ少ないと故障のオッズが $1/0.7509 = 1.33$ 倍(95% 信頼区間は $1/1.1013 = 0.91$ 以上 $1/0.4952 = 2.02$ 以下)となることがわかる。これは、Whiteside et al. (2016) とほぼ同様の結果である。また、リリース位置が体から横に離れるほど故障しやすく、横に 1 インチ(= 12 ライン)離れると故障のオッズが $1.0999^{12} = 3.14$ 倍(95% 信頼区間は $1.0160^{12} = 1.21$ 以上 $1.2039^{12} = 9.27$ 以下)となることがわかる。これは、Whiteside et al. (2016) と正反対の結果である。しかしながら、Whiteside et al. (2016) の結果は横手投げの投手のほうが上手投げの投手よりも肘への負荷が有意に大きいという指摘(Aguinaldo and Chambers, 2009)と矛盾するため、今回の結果は妥当であるといえる。そして、1 試合当たりの投球数が多いほど故障しやすく、1 試合当たり投球数が 1 球多いと故障のオッズが 1.03 倍(95% 信頼区間は 0.99 以上 1.06 以下)となることがわかる。これも Whiteside et al. (2016) とほぼ同様の結果である。一方、登板間隔は選択されなかった。このことは、主に日本人投手の言う、1 試合での投球数ではなく登板間隔こそがリスクであるという意見に反する結果であり、MLB での投手起用の方針を支持するものである。とはいえ、尤度比検定の結果は有意ではあるが、正判別率や疑似決定係数の値からは、リスク要因の選定に改善の余地が残されている。また、あくまでも MLB の投手における結果であり、これが日本人投手にも同じことが言えるかはさらなる議論の余地がある。

リリーフ投手に関しては、登板間隔 x_1 、球種数 x_4 、ファストボール球速 x_8 、リリース位置横 x_{13} という 4 変数が選択された。まず、先発投手と同様、球種数が少なく、リリース位置が体から横に離れるほど故障しやすいことがわかる。そのオッズ比は、球種数は $1/0.6697 = 1.49$ (95% 信頼区間は $1/1.0776 = 0.93$ 以上 $1/0.4034 = 2.48$ 以下)、リリース位置横(インチ)は

表 6. 先発投手およびリリーフ投手におけるロジスティック回帰分析の結果(全変数). 上: 先発投手, 下: リリーフ投手.

	推定値	標準誤差	z 値	P 値	調整オッズ比	95% 信頼区間
切片	-1.3358	19.9426	-0.0670	0.9466		
登板間隔 x_1	0.1757	0.1508	1.165	0.2440	1.1921	[0.8496, 1.6416]
登板数 x_2	-0.0051	0.0363	-0.1415	0.8875	0.9949	[0.9244, 1.0682]
1 試合当たり投球数 x_3	0.0436	0.0236	1.8473	0.0647	1.0446	[0.9988, 1.0971]
球種数 x_4	-0.5173	0.2845	-1.8182	0.0690	0.5961	[0.3276, 1.0159]
ファストボール割合 x_5	2.9403	1.6738	1.7566	0.0790	18.922	[0.8132, 639.926]
カーブ割合 x_6	-0.2055	3.7229	-0.0552	0.9560	0.8142	[0.0005, 1342.5372]
チェンジアップ割合 x_7	-3.7282	4.5957	-0.8113	0.4172	0.0240	[0.0000, 176.7731]
ファストボール球速 x_8	-0.0325	0.2040	-0.1592	0.8735	0.9680	[0.6426, 1.4448]
ファストボール球速差 x_9	-0.1262	0.3016	-0.4184	0.6756	0.8814	[0.4799, 1.5977]
ファストボール横変化量 x_{10}	0.0419	0.0682	0.6145	0.5389	1.0428	[0.9117, 1.1953]
ファストボール縦変化量 x_{11}	0.1739	0.2377	0.7316	0.4644	1.1899	[0.7476, 1.9285]
ファストボール回転数 x_{12}	-0.0003	0.0015	-0.1739	0.8619	0.9997	[0.9968, 1.0027]
リリース位置横 x_{13}	0.1181	0.0536	2.205	0.0275	1.1253	[1.0185, 1.2598]
リリース位置縦 x_{14}	-0.0435	0.0788	-0.5517	0.5811	0.9575	[0.8161, 1.1168]
リリース位置横ばらつき x_{15}	0.2240	0.3522	0.6360	0.5248	1.2511	[0.6186, 2.5490]
リリース位置縦ばらつき x_{16}	0.6418	0.6174	1.0396	0.2985	1.8999	[0.5974, 7.1152]

	推定値	標準誤差	z 値	P 値	調整オッズ比	95% 信頼区間
切片	-16.0479	14.4209	-1.1128	0.2658		
登板間隔 x_1	-1.2147	0.5081	-2.3904	0.0168	0.2968	[0.0942, 0.7129]
登板数 x_2	-0.0124	0.0192	-0.6464	0.5180	0.9877	[0.9499, 1.0254]
1 試合当たり投球数 x_3	0.0358	0.0703	0.5084	0.6112	1.0364	[0.8978, 1.1918]
球種数 x_4	-0.3389	0.2872	-1.1800	0.2380	0.7126	[0.3962, 1.2422]
ファストボール割合 x_5	0.5391	1.5732	0.3427	0.7318	1.7145	[0.0791, 42.3171]
カーブ割合 x_6	-1.9066	3.7203	-0.5125	0.6083	0.1486	[0.0001, 259.6404]
チェンジアップ割合 x_7	-3.0066	4.4531	-0.6752	0.4996	0.0495	[0.0000, 3060.9778]
ファストボール球速 x_8	0.2065	0.1558	1.3256	0.1850	1.2293	[0.9118, 1.6961]
ファストボール球速差 x_9	-0.0711	0.3130	-0.2272	0.8202	0.9313	[0.5046, 1.7648]
ファストボール横変化量 x_{10}	0.0381	0.0599	0.6363	0.5246	1.0389	[0.9250, 1.1743]
ファストボール縦変化量 x_{11}	-0.0437	0.1755	-0.2489	0.8035	0.9573	[0.6644, 1.3503]
ファストボール回転数 x_{12}	0.0005	0.0011	0.4724	0.6367	1.0005	[0.9984, 1.0027]
リリース位置横 x_{13}	0.0430	0.0404	1.0649	0.2869	1.0440	[0.9654, 1.1336]
リリース位置縦 x_{14}	0.0166	0.0687	0.2417	0.8090	1.0167	[0.8879, 1.1670]
リリース位置横ばらつき x_{15}	-0.2086	0.2526	-0.8256	0.4090	0.8117	[0.4868, 1.3383]
リリース位置縦ばらつき x_{16}	0.2091	0.4650	0.4497	0.6529	1.2326	[0.5024, 3.2197]

$1.0528^{12} = 1.85$ (95% 信頼区間は $0.9596^{12} = 0.88$ 以上 $1.1252^{12} = 4.12$ 以下)である. 次に, ファストボールの球速が速いほど, そして登板間隔が短いほど故障しやすいことがわかる. オッズ比はそれぞれ, 1.23 (95% 信頼区間は 1.00 以上 1.08 以下), $1/0.4211 = 2.37$ (95% 信頼区間は $1/0.7453 = 1.34$ 以上 $1/0.2047 = 4.89$ 以下)である. これは, Whiteside et al. (2016)における, (ファストボールに限らない)平均球速が速いほど, そして登板間隔が短いほど故障しやすいという結果に対応している. 判別結果や尤度比検定, 疑似決定係数の結果から, リリーフ投手に関する分析結果は妥当であるといえる.

表 7. 先発投手およびリリーフ投手におけるロジスティック回帰分析の結果(変数選択後).
上：先発投手, 下：リリーフ投手.

	推定値	標準誤差	z 値	P 値	調整オッズ比	95% 信頼区間
切片	-2.3455	1.887	-1.2430	0.2139		
1 試合当たり投球数 x_3	0.0250	0.0174	1.4382	0.1504	1.0253	[0.9919, 1.0633]
球種数 x_4	-0.2864	0.2011	-1.4246	0.1543	0.7509	[0.4952, 1.1013]
リリース位置横 x_{13}	0.0952	0.0428	2.2258	0.0260	1.0999	[1.0160, 1.2039]

	推定値	標準誤差	z 値	P 値	調整オッズ比	95% 信頼区間
切片	-15.0922	10.1447	-1.4877	0.1368		
登板間隔 x_1	-0.8650	0.3276	-2.6401	0.0083	0.4211	[0.2047, 0.7453]
球種数 x_4	-0.4010	0.2480	-1.6168	0.1059	0.6697	[0.4034, 1.0776]
ファストボール球速 x_8	0.2019	0.1080	1.8690	0.0616	1.2237	[0.9982, 1.5333]
リリース位置横 x_{13}	0.0514	0.0323	1.5904	0.1117	1.0528	[0.9896, 1.1252]

表 8. 変数選択後のモデルにおける先発投手およびリリーフ投手の判別結果. 左：先発投手, 右：リリーフ投手.

手術群と判別	対照群と判別	合計	手術群と判別	対照群と判別	合計
手術群	24	13	28	9	37
対照群	16	21	11	26	37
合計	40	34	39	35	74

正判別率: 0.61	正判別率: 0.73
$\chi^2 = 8.025, P$ 値: 0.0455	$\chi^2 = 20.422, P$ 値: 0.0004
Nagelkerke の疑似 R^2 : 0.137	Nagelkerke の疑似 R^2 : 0.322

3.5 スパースロジスティック回帰

Lasso を用いた正則化ロジスティック回帰モデルによるソリューションパスと, 10-fold クロスバリデーションにより選ばれた正則化パラメータの値 $\hat{\lambda}$ を示したのが図 3 である. 図の横軸は正則化パラメータ λ の対数, 縦軸は偏回帰係数を表す. 各曲線が一つの変数に対応しており, 正則化パラメータの値に対応する係数の推定値を表している. また, 図中の縦の点線が, クロスバリデーションによって選ばれた正則化パラメータの値 $\hat{\lambda}$ を表しており, そのときの各曲線の値が, この分析での最終的な推定値である.

クロスバリデーションで選択された λ の値 $\hat{\lambda}$ に基づくと, 先発投手についてはすべての係数が 0 と推定されてしまったことが図からわかる. リリーフ投手に関する係数の推定値 $\hat{\beta}_{\text{lasso}}$ およびそれを元に計算した“調整オッズ比” $e^{\hat{\beta}_{\text{lasso}}}$ は表 9 のようになった.

先発投手については, すべての変数が選択されず, 3.4 節のモデルでの説明力の弱さが強調される結果となった. またリリーフ投手については, 3.4 節の変数選択後のモデルで選ばれた変数は全て選ばれており, lasso の縮小推定の効果によりオッズ比はいずれも 1 に近づいていることがわかる. その近づき具合は概ね, 元のオッズ比の平方根程度である. すなわち, 例えばファストボールの球速 x_8 が 1 マイル増えたときのオッズ比を 0.5 マイル増えたときのオッズ比程度に過小評価していることになる. もちろん, この過小評価の度合いは λ に依存するため, 常にこのような結果が得られるわけではない. また, 上記変数に加えてチェンジアップの投球割合 x_7 が影響を及ぼしているとの結果であった.

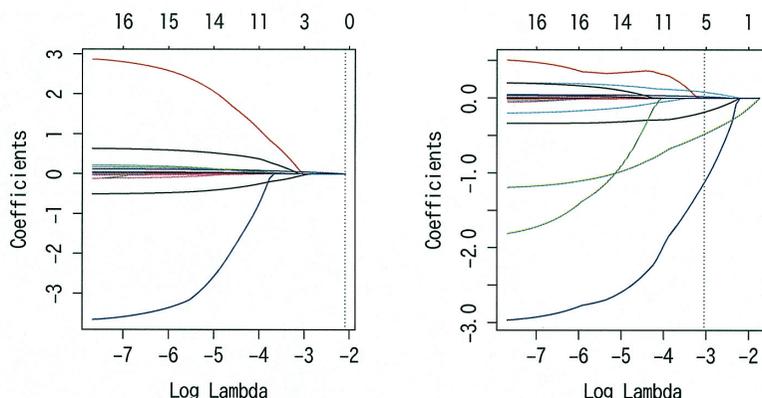


図 3. Lasso によるソリューションパス. 左: 先発投手, 右: リリーフ投手. 各折れ線が変数に対応しており, クロスバリデーションで選ばれた正則化パラメータの値(縦点線)を用いると, 先発投手では全ての変数の係数が 0 と推定され, リリーフ投手では上から順にファストボール球速(水色), リリース位置横(青色), 球種数(黒色), 登板間隔(緑色), チェンジアップ割合(青色)以外は 0 と推定されている.

表 9. Lasso による推定値(リリーフ投手).

	推定値	調整オッズ比
登板間隔 x_1	-0.4830	0.6169
球種数 x_4	-0.1866	0.8297
チェンジアップ割合 x_7	-1.1275	0.3238
ファストボール球速 x_8	0.0801	1.0834
リリース位置横 x_{13}	0.0185	1.0187

以上を踏まえると, リリーフ投手についてはステップワイズ法での結果と(オッズ比は過小評価しているものの)ほぼ類似した結果であると言えよう. 別の見方をすれば, ステップワイズ法では局所解に落ちている可能性もあるが, いずれにしても結果が妥当であることが確認できたとも言える.

4. おわりに

本研究では, 先発投手やロングリリーフ投手など長いイニングを投げる投手(先発投手)と, 中継ぎ投手や抑え投手など短いイニングを投げる投手(リリーフ投手)とに分けて, それぞれ肘内側側副韌帯損傷のリスク要因について検証した.

その結果, 先発投手については, 球種数が少ないこと, リリース位置が体から横に離れていること, 1 試合当たりの投球数が多いことがリスク要因であることがわかった. また, これらのリスクを定量的に評価することができた. 分析結果は, MLB での投手起用の方針を支持し, 主に日本での論調となっている「1 試合での投球数は多くてもよい, 登板間隔が短いのが問題である」という意見に反する結論となった. とはいえ, 先発投手に関してはリスク要因の選定に改善の余地が残されており, また, これが日本人投手にも同じことが言えるかはさらなる議論の余地がある.

またリリーフ投手に関しては, 球種数が少ないこと, リリース位置が体から横に離れている

こと、ファストボールの球速が速いこと、登板間隔が短いことがリスク要因であることがわかった。すなわち、先発とは異なり、リリーフ投手では1試合の投球数よりも登板間隔が重要であるとの結果が得られた。

また、本研究では変数選択法としてAICを用いたステップワイズ法と、Lassoによるスパースロジスティック回帰モデルとを用いた。リリーフ投手については非常に類似した結果が得られた一方で、先発投手についてはスパースロジスティック回帰モデルではよい推定結果が得られなかった。これは、ある程度の説明力がある説明変数がある場合にはステップワイズ法でもスパースロジスティック回帰モデルでも同様の結果が得られるが、説明力が十分に高くない場合にはスパースロジスティック回帰モデルではその影響をうまく捉えることができないことを示唆している。本研究の主たる目的ではないが、両方法の結果の違いを副産物的に理解することができた。

謝 辞

適切かつ有益な助言を頂いたことに対し、2人の匿名査読者と編集委員の先生に感謝する。本研究の着想に至るには、日本統計学会スポーツ統計分科会主催のスポーツデータ解析コンペティション、および情報・システム研究機構の新領域融合研究プロジェクト『社会コミュニケーション』データ中心科学リサーチコモンズ事業『人間・社会データ』の支援が不可欠であった。また、著者の1名は青山学院大学経営学部グローバルビジネス研究所の支援を受けた。本研究の一部は科学科研究費(基盤研究(C)(一般)No.15K00054)の助成を受けている。

参 考 文 献

- Aguinaldo, A. L. and Chambers, H. (2009). Correlation of throwing mechanics with elbow valgus load in adult baseball pitchers, *The American Journal of Sports Medicine*, **37**, 2043–2048.
- Baseball Heat Maps (2016). <http://www.baseballheatmaps.com/disabled-list-data/>, 2016年12月19日アクセス。
- Baseball Reference (2016). <http://www.baseball-reference.com>, 2016年12月19日アクセス。
- Friedman, J., Hastie, T. and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent, *Journal of Statistical Software*, **33**, 1–22.
- Hastie, T., Tibshirani, R. and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed., Springer, New York.
- Keller, R. A., Marshall, N. E., Guest, J. M., Okoroha, K. R., Jung, E. K. and Moutzouros, V. (2016). Major League Baseball pitch velocity and pitch type associated with risk of ulnar collateral ligament injury, *Journal of Shoulder and Elbow Surgery*, **25**, 671–675.
- Lyman, S., Fleisig, G. S., Andrews, J. R. and Osinski, E. D. (2002). Effect of pitch type, pitch count, and pitching mechanics on risk of elbow and shoulder pain in youth baseball pitchers, *The American Journal of Sports Medicine*, **30**, 463–468.
- Nagelkerke, N. J. D. (1991). A note on a general definition of the coefficient of determination, *Biometrika*, **78**, 691–692.
- Olsen, S. J., Fleisig, G. S., Dun, S., Loftice, J. and Andrews, J. R. (2006). Risk factors for shoulder and elbow injuries in adolescent baseball pitchers, *The American Journal of Sports Medicine*, **34**, 905–912.
- Sievert, C. (2014). Taming PITCHf/x Data with pitchRx and XML2R, *The R Journal*, **6**, <http://journal.r-project.org/archive/2014-1/sievert.pdf>.

- Sievert, C. (2015). pitchRx: Tools for Harnessing ‘MLBAM’ ‘Gameday’ Data and Visualizing ‘pitchfx’, R package version 1.8.2.
- Whiteside, D., Martini, D. N., Lepley, A. S., Zernicke, R. F. and Goulet, G. C. (2016). Predictors of ulnar collateral ligament reconstruction in major league baseball pitchers, *The American Journal of Sports Medicine*, **44**, 2202–2209.

A Statistical Analysis of Medial Collateral Ligament Injury Using Baseball Tracking Data in MLB

Fumitake Sakaori¹, Hiroto Enjoji², Yuto Takemori²,
Shintaro Nishizuka² and Ibuki Hoshina³

¹Faculty of Science and Engineering, Chuo University

²Graduate School of Science and Engineering, Chuo University

³The Center for Data Science Education and Research, Shiga University

The incidence of ulnar collateral ligament (UCL) reconstruction surgeries among baseball pitchers has increased in recent decades. Despite the importance of preventing UCL injuries, there is as yet no scientific consensus regarding the risk factors for such injuries.

In this paper, we reconsidered candidate risk factors for UCL injuries, referring to the opinions of an amateur pitcher and a sports doctor, and then obtained adjusted odds ratios for selected risk factors via a logistic regression model and stepwise variable selection using AIC. The results revealed the following risk factors: for starting pitchers, smaller repertoire of pitch types, horizontal release location farther from the body, and a greater mean pitch count per game; and for relief pitchers, smaller repertoire of pitch types, horizontal release location farther from the body, greater mean pitch speed of fast balls, and fewer days between consecutive games. These results support previous studies of the risk factors of UCL injuries, and provide important suggestions regarding pitch count per game and mound interval for both starter and relief pitchers.