



無料統計ソフトEZR (Easy R) の開発と臨床研究での活用の奨め

自治医科大学附属さいたま医療センター・血液科 教授 神田 善伸

1) はじめに

地域医療オープン・ラボ News Letter には 2008 年にも「造血幹細胞移植に関連する臨床研究」というタイトルで寄稿させていただきましたが、このたび統計ソフト EZR に関する寄稿のリクエストをいただき、せっかくの機会ですので「早い、うまい、安い」の三拍子揃った統計ソフト EZR を皆様にご紹介させていただきます(正確にいうと「簡単、多機能、無料」の三拍子)。

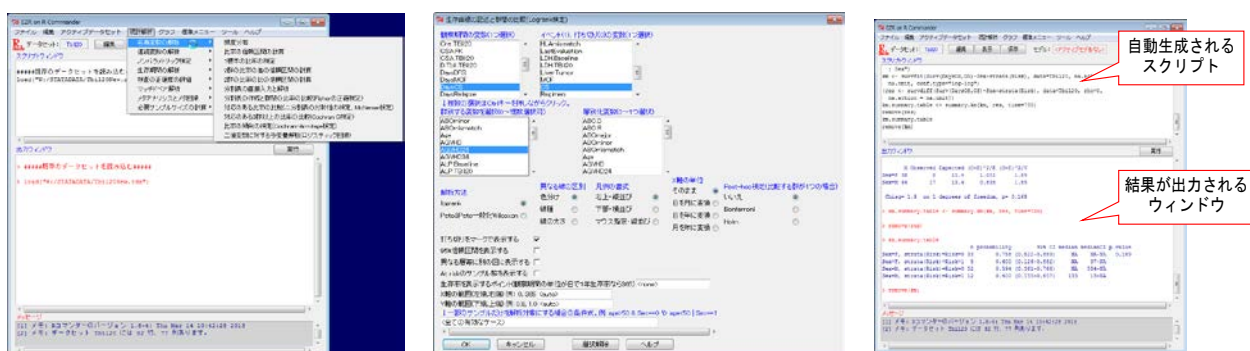


2) 統計ソフトEZRとは?

私は血液疾患を専門とする臨床医であり、統計学の専門家ではありません。私が医師になってからこれまでの 20 年余の間に行ってきたことは、日常診療で生じた臨床的クエスチョンを既存のエビデンスでは解決できない場合に、臨床決断のために必要なエビデンスを自ら構築することを目標として臨床研究を行い、その成果が世界の診療に役立つように国際誌に発表するという地道な繰り返しです。その中で、統計学は既存の論文の妥当性を解釈する上でも自ら研究を実施する上でも不可欠です。研究を実施する際に生物統計家の力を借りることができればよいのですが、生物統計家の少ない日本では全ての臨床研究に専門家が参加することはできません。そこで、臨床医にもある程度の統計の知識が必要ですし、統計ソフトを自ら活用しなければならない機会も必然的に生じます。

実際に使用する統計ソフトとしては、SAS、S-PLUS、SPSS、JMP、Stata などの統計解析ソフトが市販されていますが、いずれも個人で購入するには高価であり、解析の命令方法がわかりにくいなどの難点があります。一方、R は無料で用いることができるのでグループ内で共有するのに適していますし、既に FDA で用いられていることから信頼性も確かです。しかし、S 言語に基づくスクリプトの入力による解析は多くの人にとってなじみにくいという欠点がありました。R の追加機能パッケージである R コマンドーを R に組み込めば、かつての StatView (2002 年に販売終了) のようにマウス操作だけで簡単に解析を行うことができるようになります。しかし、R コマンドーに標準で組み込まれている統計解析の種類は限定されていました。そこで、R コマンドーのカスタマイズ機能を利用して、数多くの統計解析機能を組み込んだ統計ソフト EZR (Easy R) を自作しました(図 1、2)。少なくとも私がこれまでに発表してきた数々の臨床研究で行った統計解析手法についてはほぼ網羅しています。研究計画のために必要なサンプル数の計算も可能です。十数年前に Excel を使って苦労を重ねて行ったメタアナリシスも、この EZR を用いればマウス操作だけで瞬間的に完了してしまいます。

図 1 EZR の操作画面



①メニューから解析方法を選択

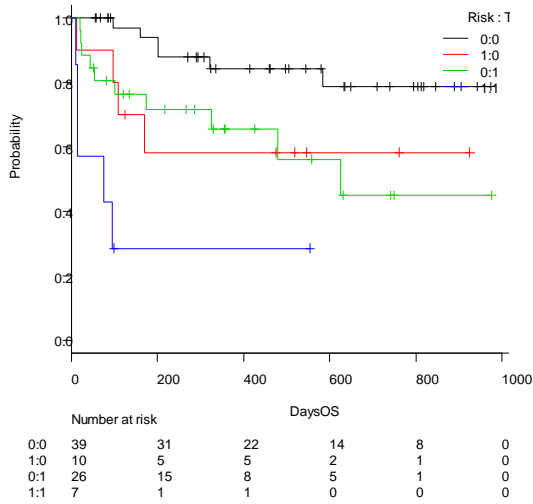
②マウスで変数や解析オプションを指定

③スクリプトと結果が表示される

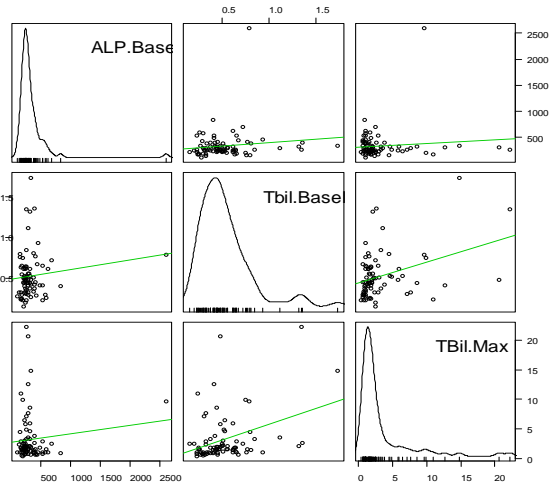
図2 EZRで作成される図のサンプル

(A 生存曲線、B 散布図行列、C メタアナリシス、D ROC曲線、E 反復測定分散分析、F 二元配置分散分析)

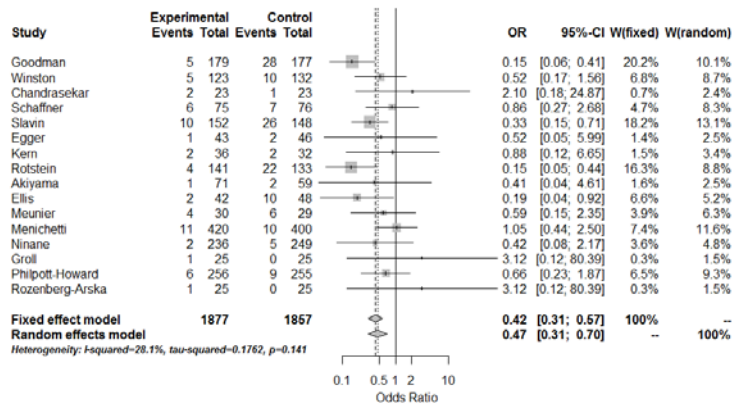
A



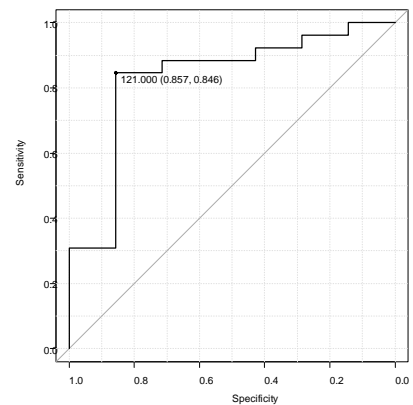
B



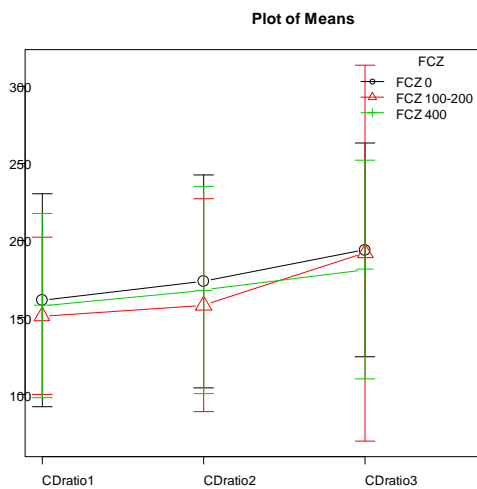
C



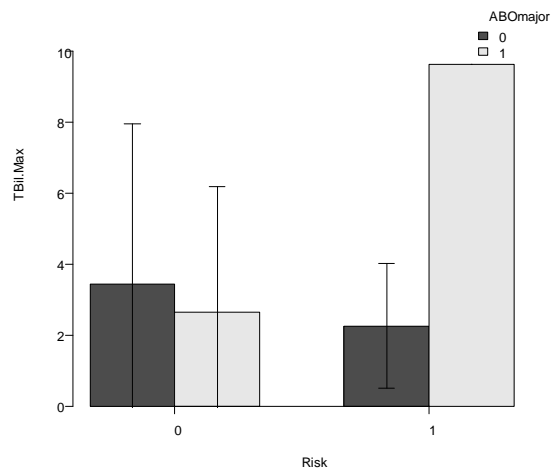
D



E



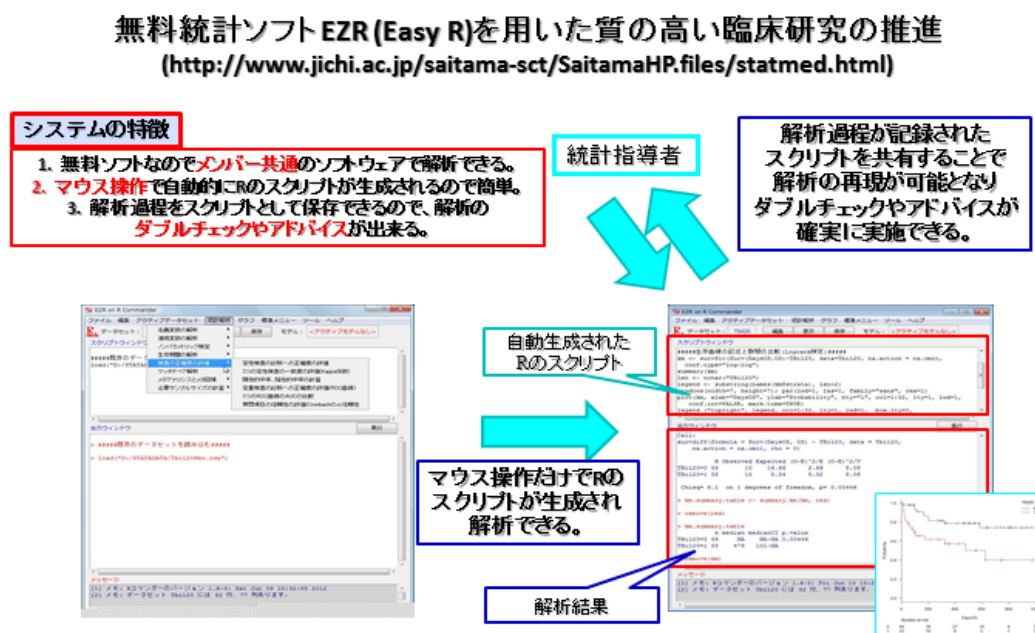
F



3) EZRを使用する利点

EZRを使用する利点として、前述のように、無料で使用できるので個人でも気軽に入手できること、グループで共通のソフトウェアとして用いることができること、マウス操作だけで高度な統計解析も実施できることなどがあります。また、学会発表、論文発表を意識して作成したソフトですので、解析結果を発表形式に自動整形して出力するという機能も備えられています。EZRのもう一つの重要な利点は、マウス操作で解析をしていると自動的にRの統計スクリプトが生成され、そのスクリプトを保存することができるということです。このスクリプトを読むことでRのスクリプトを学習することができますし、統計解析の過程を記録したり共有したりすることによって統計解析過程に問題がないか、指導者がチェックすることができるようになります(図3)。この点は特に全国各地で活躍されている自治医科大学の卒業生の皆様にとってもお役に立つのではないかと思います。統計解析に不安がある場合には、データと解析スクリプトのファイルを臨床研究支援センターにお送りいただければ(あるいは全く解析方法がわからないという段階でも構いません)、アドバイスができると思いますので、お気軽にご相談ください。連絡先は臨床研究支援センターのホームページ(<http://www.jichi.ac.jp/laboratory/clinical/>)に記載されていますので、学内者、学外者を問わず、ご利用ください。

図3 EZRによる臨床研究の推進



4) EZRのダウンロード

無料統計ソフトEZRはさいたま医療センター血液科のホームページ(<http://www.jichi.ac.jp/saitama-sct/>)から無料でダウンロードできます。Windows、Mac OSX、Linuxに対応しており、すでに数多くの英文論文でEZRが使用されています。解析機能一覧を文末の表に示します。簡単なマニュアル、論文に記載する文例、実際の使用論文リストは「EZRの使い方、変更履歴」のページに記載されていますので参考にしてください。EZRの開発を報告した参考文献1の論文(Open accessですので全文閲覧可能です。<http://www.nature.com/bmt/journal/v48/n3/pdf/bmt2012244a.pdf>)も役立つのではないかと思います。より詳細な使用方法を知りたい方は、統計、臨床研究の基礎的な解説に加えてEZRの使用方法を解説した書籍(参考文献2)が中外医学社から発行されていますのでご覧になってください。残念ながら本書は無料ではありませんが、購入すると著者の印税収入に貢献できるという利点があります。さいたま医療センターの血液科ではEZRを利用して後期研修医を含む全スタッフが臨床研究に取り組んでいます。EZRがみなさまの研究に役立つことを祈念いたします。

参考文献

1. Kanda Y. Investigation of the freely-available easy-to-use software “EZ R” (Easy R) for medical statistics. Bone Marrow Transplant. 2013;48,452-458
2. 神田善伸 EZR でやさしく学ぶ統計学 ～EBM の実践から臨床研究まで～ 中外医学社

表 EZRの解析機能一覧（その他、主成分分析、因子分析などのRコマンドーの標準解析機能も実施可能）

- ① 名義変数の解析
 - 頻度分布
 - 比率の信頼区間の計算
 - 2群の比率の差・比率の比の信頼区間の計算
 - 分割表の直接入力と解析
 - 分割表の作成と群間の比率の比較(カイ二乗検定、Fisherの正確検定)
 - 多重比較検定(Bonferroni法、Holm法)
 - 対応のある比率の比較(McNemar検定)
 - 比率の傾向の検定(Cochran-Armitage検定)
 - 二値変数に対する多変量解析(ロジスティック回帰)
- ② 連続変数の解析
 - 連続変数の要約
 - 平均値の信頼区間の計算
 - 外れ値の検定と除外(Smirnov-Grubbs検定)
 - 正規性の検定(Shapiro-Wilk検定)
 - 2群の等分散性の検定(F検定)
 - 3群以上の等分散性の検定(Bartlett検定)
 - 1標本のt検定
 - 2群間の平均値の比較(t検定)
 - 対応のある2群間の平均値の比較(paired t検定)
 - 3群以上の間の平均値の比較(一元配置分散分析 one-way ANOVA)
 - ANOVAの多重比較検定
 - 対応のある2群以上の間の平均値の比較(反復測定分散分析)
 - 対応のない複数の因子での平均値の比較(multi-way ANOVA)
 - 共分散分析(ANCOVA)
 - 相関係数の検定(Pearsonの積率相関係数)
 - 線形回帰(単回帰、重回帰)
- ③ 連続変数の解析(ノンパラメトリック)
 - 2群間の比較(Mann-Whitney U検定)
 - 対応のある2群間の比較(Wilcoxon符号付順位和検定)
 - 3群以上の間の比較(Kruskal-Wallis検定)
 - Kruskal-Wallis検定の多重比較検定
 - 対応のある3群以上の間の比較(Friedman検定)
 - 連続変数の傾向の検定(Jonckheere-Terpstra検定)
 - 相関係数の検定(Spearmanの順位相関係数)
- ④ 生存期間の解析
 - 生存曲線の記述と群間の比較(Logrank検定、一般化Wilcoxon検定)
 - 多重比較検定(Bonferroni法、Holm法)
 - 生存期間の傾向の検定(Logrank trend検定)
 - 生存期間に対する多変量解析(Cox比例ハザード回帰)
 - 時間依存性変数を含む生存期間に対する多変量解析
 - 競合するイベントの累積発生率の記述と群間の比較(Gray検定)
 - 競合するイベントの累積発生率に対する多変量解析(Fine-Gray法)
- ⑤ 検査の有用性の評価
 - 定性検査の診断への有用性の評価
 - 2つの定性検査の一致度の評価(Kappa係数)
 - 陽性適中率、陰性適中率の計算
 - 定量検査の診断への有用性の評価(ROC曲線)
 - 2つのROC曲線のAUCの比較
 - 質問項目の信頼性の評価(Cronbachの α 信頼性係数)
- ⑥ メタアナリシスとメタ回帰
 - 比率の比較のメタアナリシスとメタ回帰
 - 平均値の比較のメタアナリシスとメタ回帰
 - ハザード比のメタアナリシスとメタ回帰
- ⑦ マッチドペア解析
 - マッチさせたコントロールの抽出
 - マッチさせたサンプルの比率の比較(Mantel-Haenszel検定)
 - マッチさせたサンプルの比率の多変量解析
 - マッチさせたサンプルの生存率の多変量解析
- ⑧ 必要サンプルサイズの計算
 - 閾値奏効率、期待奏効率からのサンプルサイズ
 - 1群の比率の信頼区間をある幅におさめるためのサンプルサイズ
 - 1群の比率を既知の比率と比較するためのサンプルサイズ、検出力
 - 2群の比率の比較のためのサンプルサイズ、検出力
 - 2群の比率の比較(非劣性)のためのサンプルサイズ
 - 1群の平均値の信頼区間の幅に基づくサンプルサイズ
 - 2群の平均値の比較のためのサンプルサイズ、検出力
 - 対応のある2群の平均値の比較のためのサンプルサイズ、検出力
 - 2群の生存曲線の比較のためのサンプルサイズ、検出力
- ⑨ グラフ
 - 棒グラフ
 - 円グラフ
 - 幹葉表示
 - ヒストグラム
 - 折れ線グラフ
 - 反復測定データの折れ線グラフ
 - 箱ひげ図
 - ドットチャート
 - 整列チャート(waterfall plot)
 - 散布図
 - 散布図行列
 - 他の因子で調整した生存曲線
 - 競合するイベントの累積発生率を積み重ねて表示