



TITLE:

Establishment of a predictive model for
GVHD-free, relapse-free survival after
allogeneic HSCT using ensemble learning(
Abstract_要旨)

AUTHOR(S):

Iwasaki, Makoto

CITATION:

Iwasaki, Makoto. Establishment of a predictive model for GVHD-free, relapse-free survival after allogeneic HSCT using ensemble learning. 京都大学, 2023, 博士(医学)

ISSUE DATE:

2023-01-23

URL:

<https://doi.org/10.14989/doctor.k24312>

RIGHT:

京都大学	博士 (医 学)	氏 名	岩崎 惇
論文題目	Establishment of a predictive model for GVHD-free, relapse-free survival after allogeneic HSCT using ensemble learning (アンサンブル学習を用いた造血幹細胞移植後予後予測モデルの開発)		
(論文内容の要旨)			
<p>Graft-versus-host disease (GVHD) -free, relapse-free survival (GRFS)は、同種造血幹細胞移植 (HSCT) 後の重篤な合併症を伴わない生存を測る複合エンドポイントとして有用であるが、生存、再発、GVHD といった個々のエンドポイントに関しても正確に予測する事は、移植後の予後を理解する上で重要である。造血幹細胞移植後の予後は、疾患、患者とドナーの Human Leukocyte Antigen の一致度、ドナーソースの種類 (骨髄・末梢血・臍帯血) 等、非常に多くの移植前因子に影響を受けることが知られており、機械学習手法の応用が近年進められている。しかしながら、生存時間解析で広く使用されている統計学的手法の Cox 比例ハザードモデルと比較して、予後予測能を高める手法の検討は十分に進んでいるとは言い難い。本研究では、古典的統計学解析手法と複数の機械学習モデルを、スタッキングと呼ばれるアンサンブル学習手法を用いる事で統合して新たなメタモデルを作成し、古典的統計解析手法や既存の機械学習手法と C-index 及び Calibration を用いて予後予測能を比較した。アンサンブル学習の基礎となるモデルには、Cox 比例ハザードモデル、Random Survival Forest、Dynamic DeepHit、ADABOOST、XGBoost、Extra Tree Classifier、Bagging Classifier、及び Gradient Boosting Classifier を用いた。全生存、再発、非再発死亡、GVHD 発症といった様々なエンドポイントにおいて従来の統計学的解析や機械学習モデルと比較して高い C-index を示した (ensemble model: 0.670; Cox-PH: 0.668; Random Survival Forest: 0.660; Dynamic DeepHit: 0.646)。アンサンブルモデルは、全生存 (C-index: 0.763)、再発 (C-index: 0.793)、非再発死亡 (C-index: 0.777)、グレード II-IV 急性 GVHD (C-index: 0.656)、慢性 GVHD (C-index: 0.583)に関しても、既存の手法より高い C-index を示した。同モデルはまた、GRFS (0.023)、OS (0.210)、relapse (0.044)、グレード II-IV 急性 GVHD (0.017)、及び慢性 GVHD (0.258)において、最も低い Integrated Calibration Index を示した。本研究における結果から、アンサンブル学習を用いる事で、既存の統計学的手法や機械学習手法よりも優れた予後予測能を示すモデルが作成できる可能性を示しており、移植方法の選択に役立つ事が期待できると考えている。</p>			

(論文審査の結果の要旨)

同種造血幹細胞移植の治療を選択する上で、患者及びドナーに関わる数多くの移植前因子から、生存、再発、移植後合併症といった様々な予後予測する必要がある、近年、機械学習を用いた生存時間解析手法が検討されている。しかし、既存のモデルと比較して、予後予測能を高める手法に関しては十分に検討が進んでいない。

本研究では、統計学的解析手法と複数の機械学習モデルを、スタッキングを用いたアンサンブル学習を行い統合し、新たなメタモデルを作成した。重篤な合併症を伴わない生存を測る複合エンドポイントである、Graft-versus-host disease (GVHD) -free, relapse-free survival (GRFS)を主要評価項目とし、C-indexに基づく識別能、Integrated Calibration Index(ICI)に基づく較正能を既存のモデルと比較した。アンサンブル学習モデルは、GRFS (C-index/ICI: 0.670/0.023)。全生存 (0.763/0.210)、再発 (0.793/0.044)、急性GVHD (0.656/0.017)、慢性GVHD (0.583/0.258)に関して、従来の統計学的解析や機械学習モデルと比較して高いC-index、低いICIを示し、優れた識別能・較正能が確認できた。SHapley Additive exPlanationsを用いて、各変数の予後への影響を検討したところ、GRFSではドナーソース選択、全生存では疾患リスクの影響が特に強い事が示唆された。

以上の研究は、生存時間解析における精度の高い予後予測モデルの作成に貢献し、造血幹細胞移植の治療選択に寄与するところが多い。

したがって、本論文は博士 (医学) の学位論文として価値あるものと認める。

なお、本学位授与申請者は、令和 4 年 10 月 28 日実施の論文内容とそれに関連した試問を受け、合格と認められたものである。

要旨公開可能日： 年 月 日以降