

令和 4 年 6 月 9 日現在

機関番号：32612

研究種目：基盤研究(B)（一般）

研究期間：2018～2021

課題番号：18H03209

研究課題名（和文）異質性を考慮した因果効果の推定法の開発とその応用

研究課題名（英文）Development and application of heterogeneous causal effect estimation

研究代表者

星野 崇宏（HOSHINO, TAKAHIRO）

慶應義塾大学・経済学部（三田）・教授

研究者番号：20390586

交付決定額（研究期間全体）：（直接経費） 14,920,000円

研究成果の概要（和文）：本研究は近年の統計科学において最も重要かつ応用研究において利用されているルービン因果モデルの枠組みを拡張し、階層性のあるクラスターデータにおける因果効果の異質性や、同一個人でも何らかの要因により介入効果が異なる個人レベルの効果の異質性について統一的なモデル表現を行い、既存手法で生じるバイアスを回避する効率的な推定法を開発した。また、異質性の存在が複数RCT間での結果の不一致の問題につながることを明確にし、母集団情報を一部利用しながら異質性を考慮した選択バイアスの補正法を開発した。これらの方法論の開発と同時にマーケティング・医学・教育学での応用研究を行い提案した枠組みと方法論の妥当性を示した。

研究成果の学術的意義や社会的意義

無作為化実験が実行上あるいは倫理上難しい場合の研究における因果効果推定は医学における治療法の有効性の検証や政策効果の評価など社会的に非常に重要である。特に医学研究であれば個人差、補助金等企業の施策であれば企業による効果の違い（異質性）を正しく理解することは効率的な医療資源の供給や政策資源の配分において非常に意義がある。本研究では異質性を理解する際に生じるバイアスの除去を、場合によってはマクロレベルでしかデータが得られないといった現実場面にも利用可能な形で実施するなど、実用的ないくつかの方法を開発したという点で意義があると考えられる。

研究成果の概要（英文）：This study extends the framework of the Rubin causality model, one of the most important and applied models in statistical science in recent years, to develop a unified model for the heterogeneity of causal effects in hierarchical cluster data and individual-level effects where the intervention effect differs depending on some factors even for the same individual. We developed an efficient estimation method that avoids the bias introduced by existing methods. We also clarified that the presence of heterogeneity leads to the problem of inconsistent results across multiple RCTs, and developed a method to correct for selection bias that takes heterogeneity into account while partially utilizing population information. Simultaneously with the development of these methodologies, we conducted applied research in marketing, medicine, and education to demonstrate the validity of the proposed framework and methodologies.

研究分野：統計科学

キーワード：因果効果推定 階層モデル ベイズ推定 傾向スコア 選択バイアス

1. 研究開始当初の背景

近年の統計科学の理論的観点からも諸科学への応用の観点からも重要なトピックとして統計的因果推論がある。統計的因果推論にはいくつかの方向性があるが、特に利用されているのがルービン因果モデルに基づく統計的因果効果の推定法である。例えば統計的因果効果推定の代表的な方法である傾向スコアの元論文の引用件数は研究開始当初時点で2万回を超えており、医学や経済学、社会学など様々な分野で利用されている。これは反実仮想に基づく潜在的結果変数 (potential outcome) を想定した周辺平均や周辺分布の母数推定の問題に帰着する。因果効果や周辺分布の母数推定については、傾向スコアなどの手法が医学・経済学・社会学・経営学など多様な分野でこの10年ほどで頻繁に利用されるだけでなく、どのような仮定であればどのような推定法で一致性があり、より効率性があるかについての理論的研究は近年非常に多くなされており、本研究分野は近年の統計科学において非常に重要な分野である。

統計的因果効果推定が利用されるのは、多くの場合ヒトを対象とするようなランダム化比較実験が倫理的な理由や政策介入の評価など実施できない場合、あるいはランダム化を行ったとしても不遵守 (noncompliance) が生じるような研究分野である。ヒトを対象とする以上、治療法であれ政策介入、マーケティング施策であれ効果の個人差という意味での異質性が存在するのは当然である。治療法は患者の状態や遺伝子に、教育法は能力や学習歴に、施策は個人の嗜好やインセンティブの強さによってその効果が異なる。また治療や施策が施設や地域、集団レベルで異なるという意味の異質性も存在する。これらは統計学的にはクラスターデータでの因果効果として理解できるが、この場合にはこれまでの因果効果推定の研究でも考慮されてきた「どの条件に割り当てられるか」の内生性 (様々な変数に影響を受ける効果) だけではなく「どのクラスターに所属するか」の内生性も考慮する必要が生じる。従って施設・集団・個体ごとの固定効果や変量効果を導入する、クラスター別に因果効果を推定するだけでは大きなバイアスが生じる。このような場合の因果効果推定の方法論的研究としてまだあまり進んでいないが、応用研究を見据えた際には (具体的には以下にも記載する) 「異質性の存在する場合での因果効果」の重要性の理解、さらにそこでの推定法の開発は非常に重要である。

申請者は因果効果研究の理論的研究と応用研究を実施しているが、社会科学分野において非常に重要とされる個人差や異質性という観点から既存手法の問題点について認識していた。その上で近年 RCT の一般化可能性についての議論 (例えば Cole ら *Am.J.of.Epidem.*,2010; Van Poucke ら *J.Med.Intern.*2016; Jeffrey ら *New.Eng.J.Med.*,2017) や社会心理学研究の再現性 (Science,2015,2016) など、RCT やそのメタ分析であっても研究間の一貫性がない、あるいは実務 (医学では治療、経済経営では政策や販促の実施場面など) での再現性がないなどの問題点は参加者やクラスターの異質性の問題として理解できると考えた。また実際にクラスターへの割り当てが無作為でない場合にはシミュレーションからクラスター間で因果効果の推定に一貫性がないこと、既存手法にバイアスがあることが分かり、記載した研究内容が統計学的にも応用研究においても非常に重要であることを認識していた。

2. 研究の目的

本研究は近年の統計科学において最も重要かつ応用研究において利用されているルービン因果モデルの枠組みを拡張し、店舗と消費者、病院と患者、クラスと生徒のような階層性のあるクラスターデータにおける因果効果の異質性や、同一個人でも何らかの要因により介入効果が異なる個人レベルの効果の異質性について統一的なモデル表現を行い、既存手法で生じるバイアスを回避する効率的な推定法を開発することを目的とする。さらに提案された枠組みを適応的デザインに応用し、異質性を考慮しながら最適な介入法を個別に提示するデザインの開発を行う。また、異質性の存在が複数 RCT 間での結果の不一致や既存のメタ分析での解析の問題につながることを明確にし、母集団情報を一部利用しながら異質性を考慮した選択バイアスの補正法を開発する。これらの方法論の開発と同時にマーケティング・医学・教育学での応用研究を行い提案した枠組みと方法論の妥当性を示す。

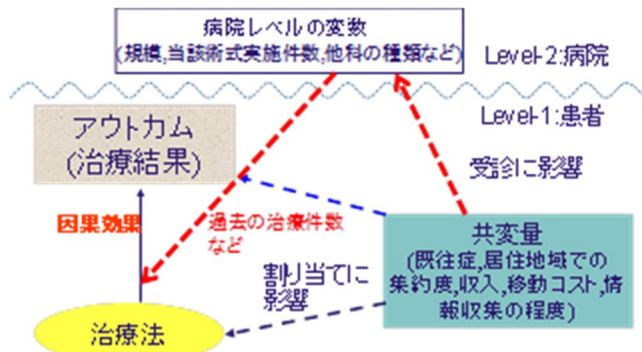
3. 研究の方法

本研究での異質性を考慮した因果効果推定について具体的な研究内容とその独自性をより具体的に述べると以下の4点である。

(1) クラスターあるいは個別ごとの介入の最適化
教育では学校や学級によって施設やクラスサイズ、教員の指導歴などの要因が異なることによ

り教授法や学習法の効果が異なる。医療であれば当該治療の実施回数(経験効果)、病院の救急医療体制や合併症に対する複数診療科の存在、コメディカルの人件など施設要因によって治療法の効果が異なる(右図)。マーケティングでも店舗やサービスによる施策の異質性が存在する。

マーケティングでは顧客ごとの one to one マーケティング、医療での個別化医療、教育での教授法の個別最適化などクラスターや施設による因果効果の異質性の解析からどのようなタイプのクラスター・施設であればどの介入法が最適かを理解することは重要であるが、既存手法では集団全体での最適介入しかわからない。また通常の解析法や既存の因果効果推定法の問題点として、クラスター・施設への生徒や顧客、患者の“割り当て”が無作為ではなく(経済的余裕のある重症患者ほど遠方の大学病院に受診するなど)それを決定する要因が介入法選択にも影響を与えることが考慮されず大きなバイアスを生じる可能性がある。但しクラスターは(病院や学級など)数十以上が一般的であり単純にどのクラスターに所属するかの傾向スコアを計算するなどの方法を使うことはできず新しい発想が必要である。



図：術式の効果の施設異質性の例

は重要であるが、既存手法では集団全体での最適介入しかわからない。また通常の解析法や既存の因果効果推定法の問題点として、クラスター・施設への生徒や顧客、患者の“割り当て”が無作為ではなく(経済的余裕のある重症患者ほど遠方の大学病院に受診するなど)それを決定する要因が介入法選択にも影響を与えることが考慮されず大きなバイアスを生じる可能性がある。但しクラスターは(病院や学級など)数十以上が一般的であり単純にどのクラスターに所属するかの傾向スコアを計算するなどの方法を使うことはできず新しい発想が必要である。

(2) 因果メカニズムの理解

独立変数が従属変数に与える因果効果に異質性が存在するという事は、何らかの調整変数が調整効果を与えている、あるいは中間変数の媒介効果にさらに異質性が存在するなどのいくつかの可能性が存在するが、これらこそ因果メカニズムの理解に重要である。また(1)に記載した介入の最適化の分野で利用されつつある機械学習的な諸手法の問題点として、異質性が何に依存しているかのモデリングあるいは背後の因果メカニズムの理解が非常に難しいという問題点を統計モデルにより解決することが可能となる。

(3) 適応的デザインでの解析への応用

マーケティング分野ではECサイトやWeb広告などで、そして医学分野でも近年利用されつつある現状の適応的デザインは一部の条件の中止や途中参加の被験者の条件間の人数の変更などであり、異質性を考慮しない研究デザインである。これに対して本来意義があるのは適応的に異質性を検出しそれに応じて提示する条件を変更するデザインであり、そのためには異質性を明示的に考慮した因果効果モデルを開発する必要がある。

(4) 母集団情報やマクロデータ、代表性のあるデータを利用した補正

異質性の問題と標本の偏りによる選択バイアスの問題は密接に関連している。公的統計や医学データベース(特にレセプトデータ)など一部変数については母集団分布が分かる場合、あるいは代表性のある標本からの情報が変数によっては集計値で時系列的に得られる場合など、多様な情報源を利用した選択バイアスの補正について、異質性の存在する場合についてこれら外部データを利用した適切な補正法の研究を行う。

4. 研究成果

研究成果リストに記載されているように、3の研究内容(1)から(4)それぞれについて国内外の学会発表を行い、その一部については査読付き雑誌として掲載させ、また現在投稿中である。具体的には、3の(1)については、介入への割り当てだけでなくクラスターへの割り当てが無作為でないことを考慮した新しいタイプの因果効果推定を一般化モーメント法および準ベイズ推定の枠組みで開発した。(2)については応用研究においてメカニズムを考慮した因果効果推定の応用をマーケティングなどにおいて行った。(3)については複数時点かつ多段階のモデルにおける因果効果推定法の研究を行い、いくつかの仮定の下で一致推定を行うことを可能にする推定法の開発を行った。(4)についてはマクロ情報をモーメント制約とする一般化モーメント法および準ベイズ法の研究開発を行った。

5. 主な発表論文等

〔雑誌論文〕 計12件（うち査読付論文 11件 / うち国際共著 1件 / うちオープンアクセス 0件）

| | |
|---|-------------------------------|
| 1. 著者名 Kato Ryo, Hoshino Takahiro | 4. 巻 59 |
| 2. 論文標題 Unplanned purchase of new products | 5. 発行年 2021年 |
| 3. 雑誌名 Journal of Retailing and Consumer Services | 6. 最初と最後の頁 102397 ~ 102397 |
| 掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.jretconser.2020.102397 | 査読の有無 有 |
| オープンアクセス オープンアクセスではない、又はオープンアクセスが困難 | 国際共著 - |
| 1. 著者名 Kato Ryo, Hoshino Takahiro | 4. 巻 10 |
| 2. 論文標題 Semiparametric Bayes instrumental variable estimation with many weak instruments | 5. 発行年 2021年 |
| 3. 雑誌名 Stat | 6. 最初と最後の頁 1-13 |
| 掲載論文のDOI (デジタルオブジェクト識別子) 10.1002/sta4.350 | 査読の有無 有 |
| オープンアクセス オープンアクセスではない、又はオープンアクセスが困難 | 国際共著 - |
| 1. 著者名 Mitsuhiro Masaki, Hoshino Takahiro | 4. 巻 3 |
| 2. 論文標題 Kernel canonical correlation analysis for data combination of multiple-source datasets | 5. 発行年 2020年 |
| 3. 雑誌名 Japanese Journal of Statistics and Data Science | 6. 最初と最後の頁 651 ~ 668 |
| 掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/s42081-020-00074-z | 査読の有無 有 |
| オープンアクセス オープンアクセスではない、又はオープンアクセスが困難 | 国際共著 - |
| 1. 著者名 齊藤勇樹・星野崇宏 | 4. 巻 - |
| 2. 論文標題 実店舗内における高額商品選択を促す事前の購買行動の検討 スマートカートデータを活用した実証分析 | 5. 発行年 2021年 |
| 3. 雑誌名 マーケティング・サイエンス | 6. 最初と最後の頁 - |
| 掲載論文のDOI (デジタルオブジェクト識別子) なし | 査読の有無 有 |
| オープンアクセス オープンアクセスではない、又はオープンアクセスが困難 | 国際共著 - |

| | |
|---|-----------------------|
| 1. 著者名 Kei Miyazaki, Takahiro Hoshino and Ulf Bockenholt | 4. 巻 39 |
| 2. 論文標題 Dynamic two stage modeling for category-level and brand-level purchases using potential outcome approach with semiparametric Bayes inference | 5. 発行年 2021年 |
| 3. 雑誌名 Journal of Business and Economic Statistics | 6. 最初と最後の頁 622-635 |
| 掲載論文のDOI (デジタルオブジェクト識別子) 10.1080/07350015.2019.1702047 | 査読の有無 有 |
| オープンアクセス オープンアクセスではない、又はオープンアクセスが困難 | 国際共著 該当する |

| | |
|--|--------------------|
| 1. 著者名 Keisuke Takahata and Takahiro Hoshino. | 4. 巻 e254 |
| 2. 論文標題 Parametric Identification of the Joint Distribution of the Potential Outcomes | 5. 発行年 2020年 |
| 3. 雑誌名 Stat | 6. 最初と最後の頁 e254 |
| 掲載論文のDOI (デジタルオブジェクト識別子) 10.1002/sta4.254 | 査読の有無 有 |
| オープンアクセス オープンアクセスではない、又はオープンアクセスが困難 | 国際共著 - |

| | |
|--|-----------------|
| 1. 著者名 Masaki Mitsuhiro and Takahiro Hoshino | 4. 巻 - |
| 2. 論文標題 Kernel Canonical Correlation Analysis for Data Fusion of Multiple Source Datasets | 5. 発行年 2020年 |
| 3. 雑誌名 Japanese Journal of Statistics and Data Science | 6. 最初と最後の頁 - |
| 掲載論文のDOI (デジタルオブジェクト識別子) なし | 査読の有無 有 |
| オープンアクセス オープンアクセスではない、又はオープンアクセスが困難 | 国際共著 - |

| | |
|---|-----------------------|
| 1. 著者名 Ryo Kato and Takahiro Hoshino | 4. 巻 48 |
| 2. 論文標題 The impact of competitors store flyer advertisement on EDLP/HiLo chain performance in highly competitive retail market: GPS information and POS data approach in Japan | 5. 発行年 2020年 |
| 3. 雑誌名 Journal of Advertising | 6. 最初と最後の頁 569-587 |
| 掲載論文のDOI (デジタルオブジェクト識別子) 10.1080/00913367.2019.1668892 | 査読の有無 有 |
| オープンアクセス オープンアクセスではない、又はオープンアクセスが困難 | 国際共著 - |

| | |
|---|--------------------|
| 1. 著者名 Yuya Shimizu and Takahiro Hoshino | 4. 巻 e241 |
| 2. 論文標題 Doubly Robust-type Estimation of Population Moments or Parameters in Biased Sampling | 5. 発行年 2020年 |
| 3. 雑誌名 Stat | 6. 最初と最後の頁 e254 |
| 掲載論文のDOI (デジタルオブジェクト識別子) 10.1002/sta4.241 | 査読の有無 無 |
| オープンアクセス オープンアクセスではない、又はオープンアクセスが困難 | 国際共著 - |

| | |
|--|------------------------|
| 1. 著者名 Kato Ryo and Hoshino Takahiro | 4. 巻 in press |
| 2. 論文標題 Semiparametric Bayesian multiple imputation for regression models with missing mixed continuous?discrete covariates | 5. 発行年 2019年 |
| 3. 雑誌名 Annals of the Institute of Statistical Mathematics | 6. 最初と最後の頁 in press |
| 掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/s10463-019-00710-w | 査読の有無 有 |
| オープンアクセス オープンアクセスではない、又はオープンアクセスが困難 | 国際共著 - |

| | |
|---|------------------------|
| 1. 著者名 Kazuhiko Shinoda and Takahiro Hoshino | 4. 巻 36 |
| 2. 論文標題 Estimation of Local Average Treatment Effect by Data Combination | 5. 発行年 2022年 |
| 3. 雑誌名 Proceedings of the AAAI Conference on Artificial Intelligence | 6. 最初と最後の頁 in press |
| 掲載論文のDOI (デジタルオブジェクト識別子) なし | 査読の有無 有 |
| オープンアクセス オープンアクセスではない、又はオープンアクセスが困難 | 国際共著 - |

| | |
|---|----------------------|
| 1. 著者名 Kazuki Nishio and Takahiro Hoshino | 4. 巻 66 |
| 2. 論文標題 Joint Modeling of Effects of Customer Tier Program on Customer Purchase Duration and Purchase Amount | 5. 発行年 2022年 |
| 3. 雑誌名 Journal of Retailing and Consumer Services | 6. 最初と最後の頁 102906 |
| 掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.jretconser.2021.1 | 査読の有無 有 |
| オープンアクセス オープンアクセスではない、又はオープンアクセスが困難 | 国際共著 - |

〔学会発表〕 計15件（うち招待講演 0件 / うち国際学会 7件）

| |
|---|
| 1. 発表者名 Daisuke Moriwaki, Komei Fujita, Shota Yasui and Takahiro Hoshino |
| 2. 発表標題 Fatigue-Aware Ad Creative Selection |
| 3. 学会等名 WSDM 2020 : ACM International Conference on Web Search and Data Mining, Houston, TX, USA. (国際学会) |
| 4. 発表年 2020年 |

| |
|--|
| 1. 発表者名 Takayuki Toda, Ayako Wakano and Takahiro Hoshino |
| 2. 発表標題 Regression Discontinuity Design with Multiple Groups for Heterogeneous Causal Effect Estimation |
| 3. 学会等名 North American Summer Meeting of the Econometric Society (国際学会) |
| 4. 発表年 2019年 |

| |
|--|
| 1. 発表者名 Ryosuke Igari and Takahiro Hoshino |
| 2. 発表標題 A Data Fusion Approach for Interpurchase-Timing Model Using the Incomplete Purchase Histories |
| 3. 学会等名 ISMS Marketing Science Conference 2019 (国際学会) |
| 4. 発表年 2019年 |

| |
|--|
| 1. 発表者名 Ryo Kato and Takahiro Hoshino |
| 2. 発表標題 Measuring the Effect of Competitors' Store Flyers on EDLP and HiLo Store Performance: GPS Data Approach |
| 3. 学会等名 ISMS Marketing Science Conference 2019 (国際学会) |
| 4. 発表年 2019年 |

| |
|--|
| 1. 発表者名 齊藤勇樹・白石佳奈子・星野崇宏 |
| 2. 発表標題 スマートレジカートを利用した事前の購買商品の影響についての検討 |
| 3. 学会等名 日本マーケティング・サイエンス学会 第105回研究大会 |
| 4. 発表年 2019年 |

| |
|--------------------------------------|
| 1. 発表者名 光廣正基・星野崇宏 |
| 2. 発表標題 ガウス過程潜在変数モデルを用いたデータ融合法の提案 |
| 3. 学会等名 2019年度統計関連学会連合大会 |
| 4. 発表年 2019年 |

| |
|---------------------------------------|
| 1. 発表者名 慶野有輝・星野崇宏・柴田卓也・清水祐弥 |
| 2. 発表標題 欠測データ解析を用いた複数の政府統計調査の統計的融合 |
| 3. 学会等名 2019年度統計関連学会連合大会 |
| 4. 発表年 2019年 |

| |
|---|
| 1. 発表者名 光廣正基・星野崇宏 |
| 2. 発表標題 NMARな欠測でのガウス過程潜在変数モデルを用いたデータ融合法の提案 |
| 3. 学会等名 第22回情報論的学習理論ワークショップ (IBIS 2019) |
| 4. 発表年 2019年 |

| |
|--|
| 1. 発表者名 星野崇宏 |
| 2. 発表標題 店舗内での選択商品の事後の選択への影響 - スマートカートデータを用いた逐次選択の研究 - |
| 3. 学会等名 行動経済学会第13回大会 |
| 4. 発表年 2019年 |

| |
|---|
| 1. 発表者名 Kato ryo, Takahiro Hoshino |
| 2. 発表標題 Semiparametric Bayes model for multidimensional instrumental variables |
| 3. 学会等名 2018 Joint Statistical Meetings (国際学会) |
| 4. 発表年 2018年 |

| |
|---|
| 1. 発表者名 Keisuke Takahata, Takahiro Hoshino |
| 2. 発表標題 Relaxation of Ignorability and Independence Assumptions Under the Availability of Auxiliary Moment Conditions: Application to Data Fusion. |
| 3. 学会等名 2018 Joint Statistical Meetings (国際学会) |
| 4. 発表年 2018年 |

| |
|--|
| 1. 発表者名 高畑圭祐、星野崇宏 |
| 2. 発表標題 外部周辺情報を利用した異質性のある因果効果の識別と推定 |
| 3. 学会等名 2018 年度 統計関連学会連合大会 |
| 4. 発表年 2018年 |

| |
|------------------------------------|
| 1. 発表者名 加藤諒、星野崇宏 |
| 2. 発表標題 母集団情報を利用した欠測を含む操作変数への対処 |
| 3. 学会等名 2018 年度 統計関連学会連合大会 |
| 4. 発表年 2018年 |

| |
|---|
| 1. 発表者名 清水祐弥、星野崇宏 |
| 2. 発表標題 Biased sampling での母集団モーメントや母数の推定 |
| 3. 学会等名 2018 年度 統計関連学会連合大会 |
| 4. 発表年 2018年 |

| |
|---|
| 1. 発表者名 Kazuhiko Shinoda and Takahiro Hoshino |
| 2. 発表標題 Estimation of Local Average Treatment Effect by Data Combination |
| 3. 学会等名 the 36th AAAI Conference on Artificial Intelligence (国際学会) |
| 4. 発表年 2022年 |

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

| | 氏名 (ローマ字氏名) (研究者番号) | 所属研究機関・部局・職 (機関番号) | 備考 |
|---------------|--|---|----|
| 研究 分担 者 | 繁樹 算男 (Shigemasu Kazuo) (90091701) | 慶應義塾大学・社会学研究科(三田)・訪問教授 (32612) | |

6. 研究組織（つづき）

| | 氏名 (ローマ字氏名) (研究者番号) | 所属研究機関・部局・職 (機関番号) | 備考 |
|-------|--|--|----|
| 研究分担者 | 猪狩 良介 (Igarı Ryosuke) (00824468) | 法政大学・経営学部・准教授 (32675) | |
| 研究分担者 | 加藤 諒 (Kato Ryo) (30823843) | 神戸大学・計算社会科学研究センター・准教授 (14501) | |

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

| 共同研究相手国 | 相手方研究機関 | | | |
|---------|-------------------------|--|--|--|
| 米国 | Northwestern University | | | |