

**科学研究費助成事業 研究成果報告書**

平成 29 年 6 月 8 日現在

機関番号：13301

研究種目：若手研究(B)

研究期間：2013～2016

課題番号：25860436

研究課題名(和文)疫学研究における血縁構造化対策ガイドラインの作成

研究課題名(英文)The confounding effect of cryptic relatedness in epidemiology.

## 研究代表者

柴田 恭子 (Shibata, Kyoko)

金沢大学・附属病院・特任准教授

研究者番号：90535600

交付決定額(研究期間全体)：(直接経費) 2,300,000円

研究成果の概要(和文)：血縁構造を除外した集団と除外しない集団で、表現型と環境因子との関連性の結果に違いがあるか、サンプルサイズ、血縁構造化の大きさ、表現型の有病率、易罹病性の遺伝率、リスク環境因子の効果サイズの5つの観点で検証し、血縁構造化の影響を調べた。環境因子と遺伝因子すべての効果は一つの変量、つまり易罹病性に集約されることを考慮し、血縁構造を除外した集団と除外しない集団で、それぞれの易罹病性の分布に違いがあるかを検定を行ったところ、有意な差はみられなかったが、家系内に遺伝力がある程度大きく、稀な疾患を想定した限られた条件下でのシミュレーション結果であることも考慮し、今後の更なる詳細な検討は必要である。

研究成果の概要(英文)：We investigated the confounding effect of unadjusted cryptic relatedness among a rural cohort in the relationship between environmental risk factors and quantitative traits in terms of sample size, the prevalence, the heritability of liability, the effect size of environmental risk factors. We found a confounding effect of cryptic relatedness in the relationship between quantitative trait and environmental risk factors was not negligible.

研究分野：遺伝疫学統計学

キーワード：血縁構造

1. 研究開始当初の背景

疾患発症リスクとなる環境要因を同定することが目的の伝統的なコホート研究では、対照集団の血縁構造が必ずしも十分には考慮されていない。しかし、極めて近い遺伝的背景をもつ血縁者が対象集団内に多数存在した場合には、環境因子と疾患罹患リスクの関連解析において、この遺伝的背景による誤った交絡を大きく受ける可能性がある。これまで日本で行われてきた古典的手法によるコホート研究では、主に地方の農村部を対象とするものであったが、特にこのような地域集団のコホート研究では、交絡因子として対象集団の血縁構造化は深刻な問題となる。リスク環境要因の発見を目的とする古典疫学の関連解析で同様の調整を試みた例はない。そこで、対象集団の血縁構造化の影響を明らかにしようとする本研究課題の着想に至った。本研究の先行研究として、平成 25 年度挑戦的萌芽研究で行った「循環器コホート分析における血縁構造化の影響の調査」がある。この研究では、実データとして地域集団サンプル 1617 人を用いて、一塩基多型を用いたゲノムワイド関連解析により血縁構造を検出し、極めて近い血縁者を除外した集団と、血縁集団を除外しない集団で、表現型と環境因子との関連性の結果に違いがあるか、表現型データとして血圧値、リスク環境因子として BMI、喫煙歴、飲酒歴を用いた。その結果（下記、表 1、2、3、4、5）が示すように、この実データを用いた検証では、交絡因子としての血縁構造化の影響が深刻な問題となる検証結果は少なかったが、遺伝率の大きさによりバイアスがかかる可能性が示唆された。このような背景をふまえ、本研究では、家系内の遺伝力がある程度大きく、稀な疾患を仮定したシミュレーションデータを作成し、さらなる詳細な検証を試みようとする本研究課題の着想に至った。

**Table 1.** Results of the regression coefficients between systolic blood pressure and environmental risk factors in the sample with and without cryptic relatedness (CR): sample size 1039, ratio of CR to the population 20.2%, population prevalence 25%, heritability for liability 14.3%, prevalence in the subpopulation without CR 26%, prevalence in CR 26%.

	Estimated	Standard error	t-value	Pr(> t )
Intercept				
Sample with CR <sup>1</sup>	82.52	5.15	16.03	<2e <sup>-16</sup>
Sample without CR <sup>2</sup>	77.11	5.49	14.06	<2e <sup>-16</sup>
BMI				
Sample with CR	1.31	0.14	9.30	<2e <sup>-16</sup>
Sample without CR	1.36	0.16	8.34	3.23e <sup>-16</sup>
Alcohol consumption				
Sample with CR	-0.80	1.51	-0.53	0.60
Sample without CR	-0.42	1.61	-0.26	0.79
Smoking status				
Sample with CR	-0.66	0.71	-0.93	0.35
Sample without CR	-0.02	1.05	-0.02	0.98
Age				
Sample with CR	0.38	0.07	5.43	6.86e <sup>-08</sup>
Sample without CR	0.41	0.05	7.99	4.62e <sup>-15</sup>
Gender				
Sample with CR	-3.09	1.04	-2.97	0.003
Sample without CR	-2.46	1.01	-2.44	0.01

<sup>1</sup>Size of sample with CR was 1039 subjects. From the sampling data with CR, equation (1) in Results were found. Adjusted R-squared = 0.14.

<sup>2</sup>Size of sample without CR was 829 subjects. From the sampling data without CR, equation (2) in Results were found. Adjusted R-squared = 0.15.

**Table 2.** Results of the regression coefficients between systolic blood pressure and environmental risk factors in the sample with and without cryptic relatedness (CR): sample size 400, ratio of CR to the population 52.5%, population prevalence 40%, heritability for liability 24.2%, prevalence in the subpopulation without CR 55%, prevalence in CR 26%.

	Estimated	Standard error	t-value	Pr(> t )
Intercept				
Sample with CR <sup>1</sup>	141.04	5.52	25.56	<2e <sup>-16</sup>
Sample without CR <sup>2</sup>	145.97	1.67	87.40	<2e <sup>-16</sup>
BMI				
Sample with CR	-0.12	0.15	-0.76	0.45
Sample without CR	0.04	0.05	0.79	0.43
Alcohol consumption				
Sample with CR	-6.11	1.81	-3.37	0.0008
Sample without CR	1.28	0.56	2.30	0.022
Smoking status				
Sample with CR	5.34	0.71	7.57	2.69e <sup>-13</sup>
Sample without CR	-0.39	0.25	-1.59	0.11
Age				
Sample with CR	-0.040	0.082	-0.48	0.63
Sample without CR	-0.13	0.02	-5.47	1.44e <sup>-07</sup>
Gender				
Sample with CR	-1.24	1.13	-1.10	0.27
Sample without CR	-0.92	0.32	-2.86	0.004

<sup>1</sup>Size of sample with CR was 400 subjects. From the sampling data with CR, equation (3) in Results were found. Adjusted R-squared = 0.25.

<sup>2</sup>Size of sample without CR was 190 subjects. From the sampling data without CR, equation (4) in Results were found. Adjusted R-squared = 0.16.

**Table 3.** Results of the regression coefficients between systolic blood pressure and environmental risk factors in the sample with and without cryptic relatedness (CR): sample size 400, ratio of CR to the population 52.5%, population prevalence 50%, heritability for liability 14.3%, prevalence in the subpopulation without CR 76.3%, prevalence in CR 26%.

	Estimated	Standard error	t-value	Pr(> t )
Intercept				
Sample with CR <sup>1</sup>	135.37	5.77	23.47	<2e <sup>-16</sup>
Sample without CR <sup>2</sup>	136.68	2.09	65.34	<2e <sup>-16</sup>
BMI				
Sample with CR	0.07	0.16	0.46	0.64
Sample without CR	0.06	0.05	1.11	0.27
Alcohol consumption				
Sample with CR	-12.34	2.00	-6.16	1.77e <sup>-09</sup>
Sample without CR	-7.75	0.81	-9.54	<2e <sup>-16</sup>
Smoking status				
Sample with CR	4.15	0.74	5.62	3.60e <sup>-08</sup>
Sample without CR	-2.32	0.32	-7.23	1.24e <sup>-11</sup>
Age				
Sample with CR	0.17	0.09	1.89	0.06
Sample without CR	0.27	0.03	7.84	3.56e <sup>-13</sup>
Gender				
Sample with CR	-1.13	1.18	-0.95	0.34
Sample without CR	0.11	0.41	0.27	0.79

<sup>1</sup>Size of sample with CR was 400 subjects. From the sampling data with CR, equation (5) in Results were found. Adjusted R-squared = 0.24.

<sup>2</sup>Size of sample without CR was 190 subjects. From the sampling data without CR, equation (6) in Results were found. Adjusted R-squared = 0.42.

**Table 4.** Results of the regression coefficients between systolic blood pressure and environmental risk factors in the sample with and without cryptic relatedness (CR): sample size 500, ratio of CR to the population 42%, population prevalence 40%, heritability for liability 22.1%, prevalence in the subpopulation without CR 50%, prevalence in CR 26%.

	Estimated	Standard error	t-value	Pr(> t )
Intercept				
Sample with CR <sup>1</sup>	137.18	4.91	27.92	<2e <sup>-16</sup>
Sample without CR <sup>2</sup>	141.55	2.99	47.26	<2e <sup>-16</sup>
BMI				
Sample with CR	0.04	0.14	0.29	0.77
Sample without CR	0.09	0.08	1.11	0.27
Alcohol consumption				
Sample with CR	-9.47	1.59	-5.94	5.47e <sup>-09</sup>
Sample without CR	-5.34	1.00	-5.35	4.26e <sup>-07</sup>
Smoking status				
Sample with CR	3.47	0.64	5.45	8.10e <sup>-08</sup>
Sample without CR	-2.30	0.44	-5.18	4.26e <sup>-11</sup>
Age				
Sample with CR	0.10	0.07	1.34	0.18
Sample without CR	0.11	0.05	2.39	0.018
Gender				
Sample with CR	-0.95	1.10	-0.94	0.35
Sample without vCR	-1.10	0.60	-1.83	0.07

<sup>1</sup>Size of sample with CR was 500 subjects. From the sampling data with CR, equation (7) in Results were found. Adjusted R-squared = 0.19.

<sup>2</sup>Size of sample without CR was 290 subjects. From the sampling data without CR, equation (8) in Results were found. Adjusted R-squared = 0.16.

**Table 5.** Results of the regression coefficients between systolic blood pressure and environmental risk factors in the sample with and without cryptic relatedness (CR): sample size 500, ratio of CR to the population 42%, population prevalence 50%, heritability for liability 31.7%, prevalence in the subpopulation without CR 67.2%, prevalence in CR 26%.

	Estimated	Standard error	t-value	Pr(> t )
Intercept				
Sample with CR <sup>1</sup>	130.15	5.51	23.62	<2e <sup>-16</sup>
Sample without CR <sup>2</sup>	128.16	3.92	32.73	<2e <sup>-16</sup>
BMI				
Sample with CR	0.17	0.15	1.13	0.26
Sample without CR	0.22	0.10	2.13	0.03
Alcohol consumption				
Sample with CR	-16.61	1.67	-9.95	<2e <sup>-16</sup>
Sample without CR	-8.11	1.24	-6.53	2.99e <sup>-10</sup>
Smoking status				
Sample with CR	2.58	0.71	3.63	0.0003
Sample without CR	-2.87	0.65	-4.40	1.53e <sup>-05</sup>
Age				
Sample with CR	0.39	0.08	4.91	1.21e <sup>-06</sup>
Sample without CR	0.42	0.05	7.72	1.99e <sup>-13</sup>
Gender				
Sample with CR	-1.44	1.09	-1.33	0.19
Sample without CR	-1.29	0.72	-1.79	0.08

<sup>1</sup>Size of sample with CR was 500 subjects. From the sampling data with CR, equation (9) in Results were found. Adjusted R-squared = 0.21.

<sup>2</sup>Size of sample without CR was 290 subjects. From the sampling data without CR, equation (10) in Results were found. Adjusted R-squared = 0.24.

## 2. 研究の目的

本研究では、シミュレーションデータを作成し、血縁者を除外した集団と、除外しない集団で、表現型と環境因子との関連性の違いがあるか検証し、血縁構造化の影響を明らかにする。

## 3. 研究の方法

正規分布に従う量的表現型を仮定し、血縁構造を除外した集団と、血縁構造を除外しない集団で、表現型と環境因子との関連性の結果に違いがあるか、サンプルサイズ、血縁構造化の大きさ、表現型の有病率、易罹病性の遺伝率、リスク環境因子の効果サイズの5つの観点で検証し、血縁構造化の影響を調べた。検証用データとして、家系内の遺伝力がある程度大きく、稀な疾患を仮定したシミュレーションデータを生成した。一般集団での表現型発症頻度を0.1%、血縁構造の程度を1親等、2親等、3親等を仮定して、それぞれの遺伝力の条件下で、血縁構造ごとに異なる表現型発症確率を与えた。血縁構造と一般集団による混合集団のそれぞれの混合割合について、一般集団と血縁構造の比、4対6、2対8、6対4と変化させて、全サンプルサイズを100、500、1000、5000生成し、混合正規分布に従う量的表現型を想定した。環境因子と遺伝因子すべての効果は、一つの変量、つまり易罹病性に集約されることを考慮し、血縁構造を除外した集団と血縁構造を除外しない集団で、それぞれの易罹病性の分布に違いがあるかt検定を行った。

#### 4. 研究成果

検定の結果、有意な差は見られず、きわめて大きな影響が懸念されることは示唆されなかった。しかし、今回のように、家系内の遺伝力がある程度大きく、稀な疾患を想定して限られた条件下でのシミュレーション結果であることも考慮し、今後のさらなる詳細な検討が必要である。

#### 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計 0 件)

〔学会発表〕(計 0 件)

#### 6. 研究組織

##### (1) 研究代表者

柴田 恭子 (SHIBATA, Kyoko)  
金沢大学・附属病院・特任准教授  
研究者番号：90535600