BMJ Open

Detection for pathway effect contributing to disease in systems epidemiology with a case–control design

Jiadong Ji,[1] Zhongshang Yuan,[1] Xiaoshuai Zhang,[1] Fangyu Li,[2] Jing Xu,[1] Ying Liu,[3] Hongkai Li,[1] Jia Wang,[4] Fuzhong Xue[1]

To cite: Ji J, Yuan Z, Zhang X, *et al*. Detection for pathway effect contributing to disease in systems epidemiology with a case–control design. *BMJ Open* 2015;**5**:e006721. doi:10.1136/bmjopen-2014-006721

► Prepublication history and additional material is available. To view please visit the journal (http://dx.doi.org/10.1136/bmjopen-2014-006721).

Received 24 September 2014
Revised 21 November 2014
Accepted 17 December 2014

[1]Department of Epidemiology and Biostatistics, School of Public Health, Shandong University, Jinan, China
[2]Department of Neurology, Capital Medical University, Xuanwu Hospital, Beijing, China
[3]Department of Public Health and Clinical Medicine, Umea University, Umea, Sweden
[4]School of Mathematics, Shandong University, Jinan, China

Correspondence to
Professor Fuzhong Xue;
xuefzh@sdu.edu.cn

ABSTRACT

Objectives: Identification of pathway effects responsible for specific diseases has been one of the essential tasks in systems epidemiology. Despite some advance in procedures for distinguishing specific pathway (or network) topology between different disease status, statistical inference at a population level remains unsolved and further development is still needed. To identify the specific pathways contributing to diseases, we attempt to develop powerful statistics which can capture the complex relationship among risk factors.

Setting and participants: Acute myeloid leukaemia (AML) data obtained from 133 adults (98 patients and 35 controls; 47% female).

Results: Simulation studies indicated that the proposed Pathway Effect Measures (PEM) were stable; bootstrap-based methods outperformed the others, with bias-corrected bootstrap CI method having the highest power. Application to real data of AML successfully identified the specific pathway (Treg→TGFβ→Th17) effect contributing to AML with p values less than 0.05 under various methods and the bias-corrected bootstrap CI (−0.214 to −0.020). It demonstrated that Th17–Treg correlation balance was impaired in patients with AML, suggesting that Th17–Treg imbalance potentially plays a role in the pathogenesis of AML.

Conclusions: The proposed bootstrap-based PEM are valid and powerful for detecting the specific pathway effect contributing to disease, thus potentially providing new insight into the underlying mechanisms and ways to study the disease effects of specific pathways more comprehensively.

Strengths and limitations of this study

- Powerful new statistical methods were proposed in detecting whether the pathway effect is significantly different between the case and control groups within the framework of systems epidemiology.
- Statistical simulations were conducted to assess their performance, and a real data set for acute myeloid leukaemia was further analysed to validate their practicability.
- Bootstrap-based pathway effect measures are valid and powerful for identifying the specific pathway contributing to disease, providing potential new insight into underlying mechanisms and more comprehensive ways to study the disease effects of specific pathways.
- The limitation of the proposed bootstrap-based methods is the computation burden on the bootstrap procedure used to evaluate the CI when dealing with big data.

INTRODUCTION

Epidemiology involves the research of disease prevalence, incidence and its risk factors. However, traditional epidemiology mainly focuses on a single risk factor related to a disease, and this simplicity of the single-level paradigm has serious limitations,[1] which have been increasingly criticised.[2] Recent advances in high-throughput-omics technologies include genomics, epigenomics, transcriptomics, proteomics and metabolomics. It offers the potential to provide new insight into the underlying mechanisms in breadth and depth. The integration of traditional epidemiology with various omics data promotes a novel epidemiology branch, systems epidemiology,[3 4] which is expected to create a network system to study health and disease at a human population level.[1] Under this framework, the focus has been shifted from identification of independent risk factors to exploration of network or pathway effects on specific diseases. Nevertheless, systems epidemiology also proved to be a great challenge in impeccably designed and well-powered epidemiological studies with powerful new statistical analysis methods.

Numerous network analytical methods, which have been applied in studying human behaviour, physiology, systems biology and modern drug development,[5–9] have provided the computational framework for data integration and biomarker selection in systems epidemiology.[10] Among these, pathway analysis is an essential task for network analysis in systems epidemiology. Several methods have been proposed, including but not limited to, a Gene Set Enrichment Analysis (GSEA) approach,[11] Prioritising Risk Pathways fusing single nucleotide polymorphisms (SNPs) and pathways,[12] Bayesian Pathway Analysis,[13] pathway analysis approaches based on the adaptive rank truncated product statistic,[14] and a sub-pathway-based approach to study the joint effect of multiple genetic variants.[15] Generally, these methods are suitable for various omics data in systems epidemiology. However, most of them attach little importance to the complex relationships (correlation and topological structure) between nodes in pathways, and only consider the probability of disease-related nodes co-occurring in pathways.

From the perspective of systems epidemiology, networks are abstract representations of biological systems at the population level, which have illustrated multilevel causes of the occurrence, development and prognosis of complex diseases. In the network, variables (risk factors) are represented by nodes, with their interactions or correlation by edges (or arrows). Perturbations in networks disrupt biological pathways and result in human diseases usually in the following situations: (1) the topological structure of pathways (or networks), but not pathway effects, is the same between different disease status; (2) the topological structure of the network changes under different disease status, including nodes or edges removal. Focusing on the former scenario, we attempt to develop a novel statistical method for detecting the pathway effect within a network between different disease status under a case–control design, in order to identify the specific pathway contributing to disease. To assess the performance of the proposed statistics Pathway Effect Measures (PEM), statistical simulations were conducted to evaluate type I error and power, and a real data set was further analysed to validate their practicability.

METHODS

Pathway effect and PEM statistics

The proposed PEMs were developed under the framework of a graph model.[16] Graph G is an ordered pair of disjoint sets (V, E), where V and E are finite sets. V=V(G) is the set of vertices and E=E(G) is the set of edges. A pathway is a graph P, where $V(P)=\{x_0, x_1,\ldots, x_K\}$ and $E(P)=\{x_0x_1, x_1x_2, \ldots, x_{K-1}x_K\}$, K denotes the number of edges in E(P) defined as the length of this pathway. Figure 1 shows a specific network with six nodes X_i (i=1, …, 6) and 10 edges. In this network, the effect of a specific pathway $X_1\rightarrow X_3\rightarrow X_5\rightarrow X_6\rightarrow X_4$ is $\beta=\prod_{k=1}^{K}\beta_k$, where β_k denotes the standardised regression coefficient between k^{th} and $(k+1)^{th}$ nodes in the pathway. In this paper, we mainly focus on detection of the pathway effect between different disease status (case and control) under the same topological structure of the network. The difference in pathway effect between case and control $D=\beta_D-\beta_C$ can be introduced as an estimate of the pathway effect contributing to the disease, where β_D is the pathway effect among cases and β_C the pathway effect among controls. Since the distribution of D is not available, two typical PEM statistics were proposed for detecting the pathway effect contributing to the disease.

Non-parametric bootstrap test

The statistic (PEM-D) is defined as

$$D=\beta_D-\beta_C=\prod_{k=1}^{K}\beta_k^D-\prod_{k=1}^{K}\beta_k^C \quad (1)$$

where β_k^D and β_k^C represent the standardised regression coefficient between k^{th} and $(k+1)^{th}$ nodes in the pathway from cases and controls, respectively. To test whether a pathway has an effect on the disease of interest, we conducted hypothesis testing with $H_0:\beta_D=\beta_C$ versus $H_1:\beta_D\neq\beta_C$. Bootstrap methods[17] were further employed to perform the hypothesis test. The percentile bootstrap CI approach was conducted as follows: (1) draw a large number of bootstrap samples (eg, 1000) and estimate D in each of them to form a bootstrap distribution; (2) define the limits of a 1-α CI as α/2 and 1-α/2 percentiles of the bootstrap distribution and (3) reject the null hypothesis ($H_0:\beta_D=\beta_C$) if the CI does not include zero.

Since the bootstrap distribution may fail to centre at the sample estimate of D, a bias-corrected bootstrap CI[18] was also adopted in this study. The detailed procedure is outlined as follows: (1) form bootstrap distribution as above; (2) find a z-score $z_\rho=\Phi^{-1}(\rho)$, where Φ^{-1} is the inverse cumulative distribution function for standard

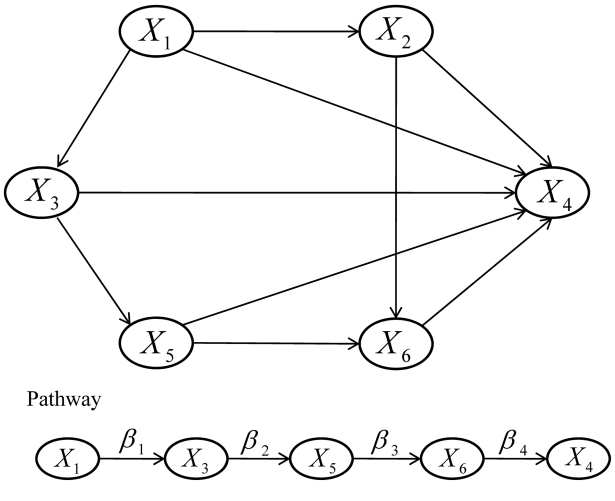

Figure 1 Network and one specific pathway.

normal distribution and ρ the proportion of the bootstrap distribution greater than the original sample D; (3) calculate $z_{lower}=z_{\alpha/2}-2z_{\rho}$ and $z_{upper}=-z_{\alpha/2}-2z_{\rho}$, the limits of a 1-α CI are percentile ranks from the bootstrap distribution for $\Phi(z_{lower})$ and $\Phi(z_{upper})$; and (4) make a decision using the bias-corrected CI of the test statistic D.

Asymptotic normal distribution statistic (PEM-U_D)

From the definition of $D=\beta_D-\beta_C$, the asymptotic normal distribution statistic is defined as

$$U_D=\frac{\beta_D-\beta_C}{\sqrt{var(\beta_D)+var(\beta_C)}} \quad (2)$$

Where $var(\beta_D)$ and $var(\beta_C)$ denote the variance of β_D and β_C, respectively, which can be calculated by four different methods:

(1) the exact estimator[19]

$$var(\beta)_{exact}=\prod_{k=1}^{K}(s_{\beta_k}^2+\beta_k^2)-\prod_{k=1}^{K}\beta_k^2,$$

where s_{β_k} is the SE of β_k;

(2) the unbiased estimator[19]

$$var(\beta)_{unbiased}=\prod_{k=1}^{K}\beta_k^2-\prod_{k=1}^{K}(\beta_k^2-s_{\beta_k}^2);$$

(3) the multivariate Δ estimator[20]

$$var(\beta)_{mult-delta}=\Delta cov(\beta_1,\beta_2,\ldots,\beta_K)\Delta^T,$$

where

$$\Delta=\left[\frac{\partial\beta}{\partial\beta_1},\cdots,\frac{\partial\beta}{\partial\beta_K}\right];$$

(4) the bootstrap estimator.[17] The former three variance estimators were derived under the independent assumption of the standard regression coefficients β_k.

In summary, the significance test procedure was conducted as follows: (1) calculate $\beta_D-\beta_C$ from the original sample; (2) conduct the bootstrap procedure to estimate the 1-α percentile bootstrap CI, bias-corrected CI and calculate $var(\beta_D)$ and $var(\beta_C)$ by four different methods; and (3) reject the null hypothesis if the CI does not include zero or the p value less than the significance level α.

Simulation

Simulations were conducted to evaluate the performance of D and U_D under different sample sizes, pathway effect and pathway length. For the specific pathway with length K, the simulated (K+1)-dimensional variables (nodes) were generated from a multivariate normal distribution $N_{K+1}(\mu,\sum)$ with mean vector μ and covariance matrix $\sum$. In this article, we specified the mean vector $\mu=\mathbf{0}$ and covariance matrix.

$$\Sigma=\begin{pmatrix}1 & \beta_1 & 0 & \cdots & 0 & 0\\ \beta_1 & 1 & \beta_2 & 0 & \cdots & 0\\ 0 & \beta_2 & \ddots & \ddots & \ddots & \vdots\\ \vdots & 0 & \ddots & \ddots & \beta_{K-1} & 0\\ 0 & \vdots & \ddots & \beta_{K-1} & 1 & \beta_K\\ 0 & 0 & \cdots & 0 & \beta_K & 1\end{pmatrix}$$

Under the null hypotheses ($H_0:\beta_D=\beta_C$), the data were generated by setting β_k^D and β_k^C to suffice $\prod_{k=1}^{K}\beta_k^D=\prod_{k=1}^{K}\beta_k^C$. Under the alternative hypotheses ($H_1:\beta_D\neq\beta_C$), different correlation coefficient β_k, pathway effect contributing to the disease $\delta=\prod_{k=1}^{K}\beta_k^D-\prod_{k=1}^{K}\beta_k^C$ and pathway length K were considered. All simulation data were generated by the R "**mvtnorm**" package available from CRAN (http://cran.r-project.org/).

Under H_0, 1000 simulations, given the various sample sizes (N=50, 100, 200, 300, 500, 1000) and pathway length (K=2, 3, 4, 5), were conducted to assess the type I error of the above two typical PEM, including the non-parametric bootstrap test with CI estimated by the percentile bootstrap and bias-corrected bootstrap methods, and asymptotic normal distribution statistic with variances calculated via four approaches. Under H_1, given the various sample sizes, we repeated 1000 simulations to assess the power under a different correlation pattern, δ and K, respectively.

APPLICATION

The proposed two typical PEM were applied to acute myeloid leukaemia (AML) data, consisting of T helper type 17 (Th17) cells, regulatory T (Treg) cells and their related cytokine transforming growth factor β (TGFβ) in a bone marrow microenvironment from 98 patients with AML and 35 controls collected by the Qilu Hospital of Shandong University in China. Patients with AML were diagnosed according to the French-American-British (FAB) classification system. Patients with hypertension, diabetes, cardiovascular diseases, or chronic or active infection, or were pregnant were excluded. Individuals with a slight iron deficiency anaemia, having no immunological changes, were used as controls. Clinical characteristics of the participants are presented in table 1. Informed consent was obtained from all participants before enrolment in the study in accordance with the Declaration of Helsinki. TGFβ is a pre-requisite for the induction of CD4+ T-cell Foxp3 expression and differentiation into Treg cells. TGFβ is also critical for human Th17 cell differentiation.[21 22] All six methods were conducted to detect the pathway (Treg→TGFβ→Th17) effect contributing to AML.

Table 1 Clinical characteristics of patients grouped according to acute myeloid leukaemia (AML) status

	AML (n=98)	Control (n=35)	p Value
Gender (female)	47 (48.0%)	16 (45.7%)	0.819
Age (years)	42.36±13.89	39.63±13.03	0.313
Treg (%)	2.20 (1.51)	1.37 (2.04)	0.001
Th17 (%)	2.48 (3.08)	2.49 (2.73)	0.657
TGFβ (pg/mL)	3700.20 (5803.65)	9763.59 (6633.97)	<0.001

Data are presented as means±SDs, medians (IQRs); compared continuous variables with two sample t test or Wilcoxon rank-sum test, and categorical variables with a χ^2 test.

RESULTS

Simulation

Type I error rate

Table 2 shows the estimated type I error rates of the two proposed PEM D and U_D under different pathway lengths and sample sizes. The type I error rates of the non-parametric bootstrap test are close to the given nominal level (α=0.05) when the sample size reaches 300, and 500 for the asymptotic normal distribution statistic.

Statistical power

The power of the proposed PEM D and U_D under K=3 is shown in figure 2. It can be seen that the power of the proposed six methods monotonically increases with the sample size and δ. The bootstrap-based tests (percentile bootstrap, bias-corrected bootstrap, variance estimated via a bootstrap) perform better than the other three tests, with the bias-corrected bootstrap CI method having the highest power. The power can be considerable under a larger δ (figure 2C), even with the small sample size (200).

Figure 3 presents the power of the proposed PEM under three different correlation patterns (figure 3A–C) given same δ=0.1 with pathway length K=2. It suggests that the power of all six methods decreases when the correlation pattern increases, with figure 3A showing the

Table 2 Type I error for two typical PEM in different scenarios

Sample size	50	100	200	300	500	1000
Pathway length K=2			$(\beta_1^D, \beta_2^D) = (0.4, 0.2)$ $(\beta_1^C, \beta_2^C) = (0.2, 0.4)$			
Multivariate Δ*	0.027	0.039	0.045	0.032	0.040	0.051
Exact*	0.019	0.035	0.042	0.031	0.040	0.050
Unbiased*	0.033	0.043	0.045	0.033	0.040	0.051
Bootstrap*	0.040	0.052	0.057	0.042	0.049	0.059
Percentile bootstrap†	0.046	0.057	0.059	0.047	0.057	0.059
Bias-corrected bootstrap†	0.056	0.068	0.058	0.048	0.057	0.058
Pathway length K=3			$(\beta_1^D, \beta_2^D, \beta_3^D) = (0.1, 0.3, 0.5)$ $(\beta_1^C, \beta_2^C, \beta_3^C) = (0.5, 0.1, 0.3)$			
Multivariate Δ*	0.005	0.021	0.037	0.035	0.048	0.045
Exact*	0.003	0.015	0.033	0.035	0.047	0.044
Unbiased*	0.015	0.026	0.042	0.040	0.050	0.047
Bootstrap*	0.011	0.020	0.044	0.037	0.053	0.048
Percentile bootstrap†	0.029	0.034	0.049	0.047	0.054	0.050
Bias-corrected bootstrap†	0.059	0.057	0.059	0.053	0.058	0.057
Pathway length K=4			$(\beta_1^D, \beta_2^D, \beta_3^D, \beta_4^D) = (0.6, 0.5, 0.1, 0.3)$ $(\beta_1^C, \beta_2^C, \beta_3^C, \beta_4^C) = (0.1, 0.3, 0.5, 0.6)$			
Multivariate Δ*	0.004	0.012	0.022	0.030	0.035	0.047
Exact*	0.001	0.009	0.021	0.026	0.032	0.047
Unbiased*	0.010	0.019	0.032	0.035	0.037	0.047
Bootstrap*	0.014	0.020	0.034	0.037	0.039	0.054
Percentile bootstrap†	0.033	0.034	0.042	0.045	0.048	0.058
Bias-corrected bootstrap†	0.065	0.060	0.055	0.052	0.054	0.056
Pathway length K=5			$(\beta_1^D, \beta_2^D, \beta_3^D, \beta_4^D, \beta_5^D) = (0.1, 0.3, 0.4, 0.5, 0.6)$ $(\beta_1^C, \beta_2^C, \beta_3^C, \beta_4^C, \beta_5^C) = (0.6, 0.5, 0.1, 0.3, 0.4)$			
Multivariate Δ*	0.000	0.004	0.018	0.034	0.035	0.032
Exact*	0.000	0.003	0.014	0.031	0.033	0.029
Unbiased*	0.001	0.013	0.023	0.039	0.038	0.033
Bootstrap*	0.004	0.007	0.026	0.037	0.037	0.041
Percentile bootstrap†	0.022	0.035	0.043	0.049	0.049	0.043
Bias-corrected bootstrap†	0.066	0.061	0.059	0.057	0.056	0.046

*For PEM-U_D with different methods, estimate the variance.
†For PEM-D with different methods, estimate the CI.

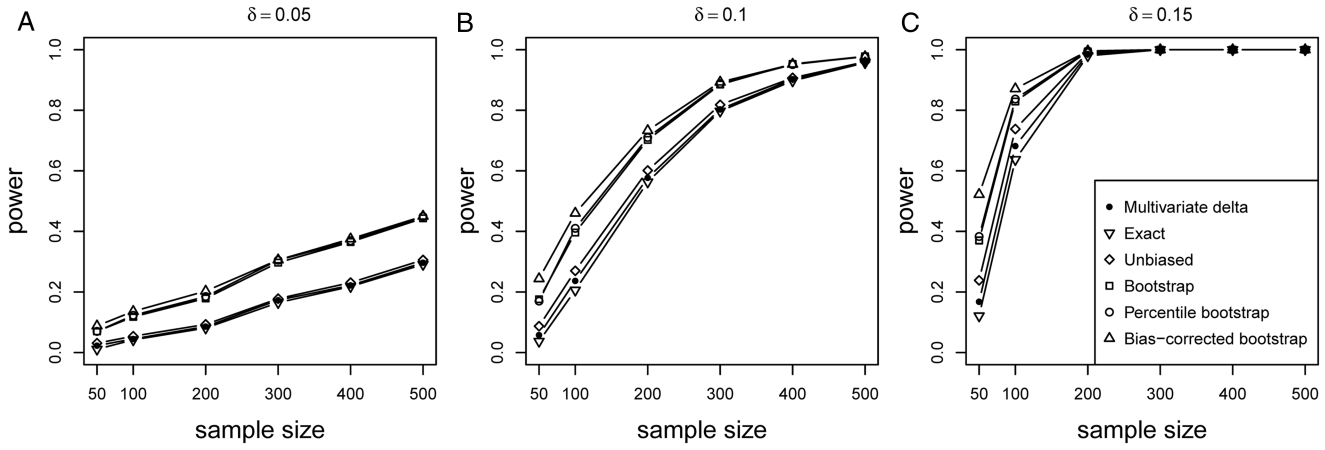

Figure 2 Power of PEM under different sample size and different d given pathway length K=3. Different pathway effect contributing to the disease d=0.05 (A), d=0.1 (B) and d=0.15 (C) were given, respectively.

highest power followed by figure 3B,C. Again, the bootstrap-based tests still have more advantageous performance than the other three.

The power of the proposed PEM under different pathway length K is shown in figure 4. Figure 4A shows that the power increases monotonically with K, given the same δ. Moreover, figure 4B shows the power with the same correlation pattern, though δ decreases with the increase in pathway length, the power may still increase. In addition, U_D with variance estimated via a bootstrap is more powerful than other variance-estimated approaches (exact, unbiased, multivariate Δ).

Application result

Table 3 shows the results of the proposed PEM for detecting the pathway (Treg→TGFβ→Th17) effect contributing to AML. This pathway effect can be detected significantly at α=0.05 by the percentile bootstrap and bias-corrected bootstrap CI method, as well as by PEM-U_D estimated the variance via the unbiased estimator,[19] the multivariate delta estimator,[20] the bootstrap estimator.[17]

DISCUSSION

Systems epidemiology couples traditional epidemiology with modern high-throughput technologies which seek to integrate pathway-based (or network-based) analysis into observational study designs to enhance the understanding of biological processes in the human organism. It provides a means to organise and study the interdependencies of factors (eg, genes, proteins, metabolites)[23–28] at a human population level. Within this framework, the identification of pathways effects responsible for specific diseases has been one of the essential tasks. In the framework of bioinformatics, various methods existed for inferring biological networks[29–31] aiming to mine underlying networks for identifying biological modules, clustering interactions, and topological features of the network such as degree and betweenness centrality.[32–34] Despite these procedures for distinguishing specific pathway (or network) topology between different disease status, statistical inference at a population level remains unsolved, and further development is still necessary In summary, both the specific pathway (or network) topology and their effect on phenotype (or disease) should be considered in systems

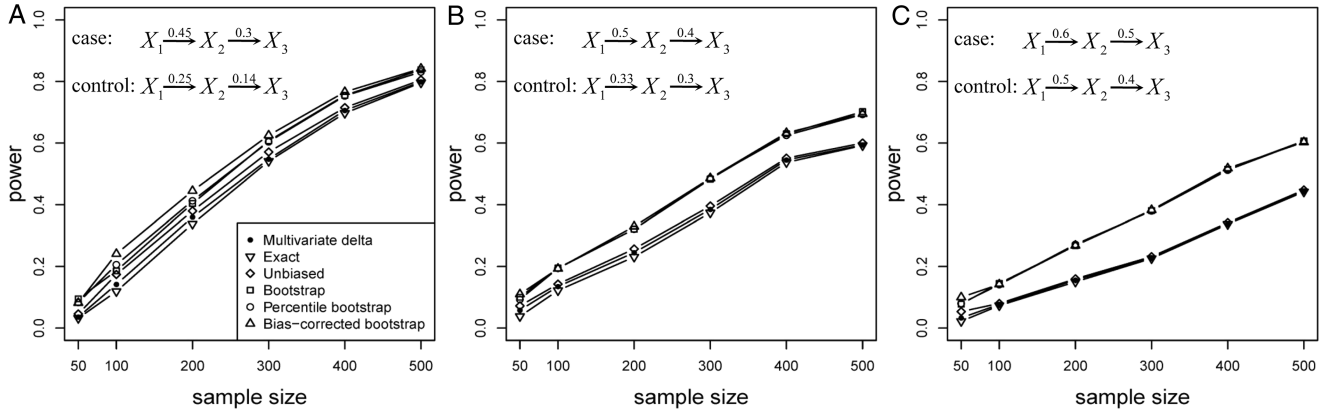

Figure 3 Power of PEM under different sample size and different correlation patterns given K=2 and d=0.1. (A) for correlation pattern given $(\beta_1^D, \beta_2^D) = (0.45, 0.3)$ and $(\beta_1^C, \beta_2^C) = (0.25, 0.14)$; (B) for correlation pattern given $(\beta_1^D, \beta_2^D) = (0.5, 0.4)$ and $(\beta_1^C, \beta_2^C) = (1/3, 0.3)$; (C) for correlation pattern given $(\beta_1^D, \beta_2^D) = (0.6, 0.5)$ and $(\beta_1^C, \beta_2^C) = (0.5, 0.4)$.

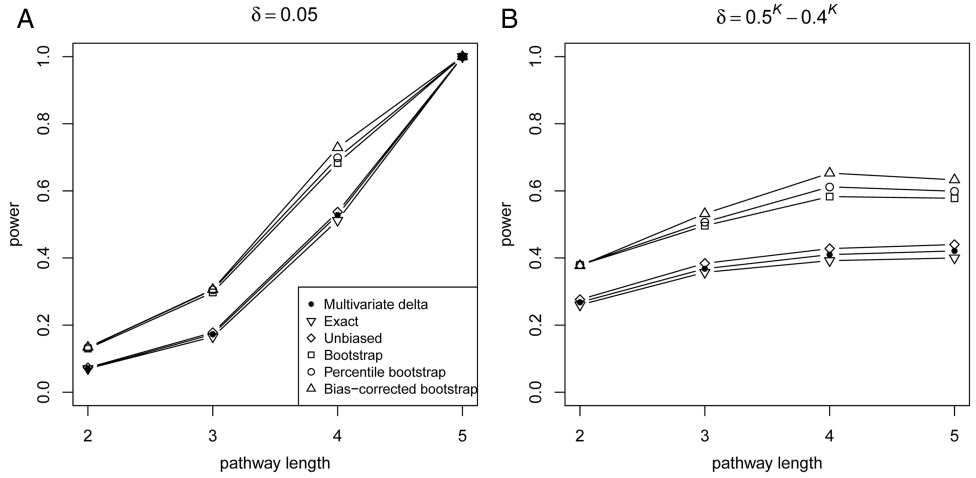

Figure 4 Power of PEM under different pathway length given sample size 300. (A) For different pathway length K given same d=0.05; (B) for different pathway length K given same $\beta_i^D = 0.5$ and $\beta_i^C = 0.4$, $(\delta = 0.5^K - 0.4^K)$ i = 1, . . . , K.

epidemiology data analysis. In this paper, we furnished two typical PEM to detect the pathway effect within a network between different disease status using the case–control design, expected to identify the specific pathway contributing to disease.

Our simulation showed that both D and U_D kept stable under the null hypothesis with a large sample size. It indicated that the power of the proposed six methods increased monotonically with sample size, δ and K (figure 2 and 4A), and decreased when the correlation pattern increased (figure 3). Even though δ decreased, the power still increased with the increase in the pathway length under a fixed correlation pattern. Overall, the bootstrap-based tests (percentile bootstrap, bias-corrected bootstrap, variance estimated via a bootstrap) perform better than the other tests, with the bias-corrected bootstrap CI method having the highest power. Additional simulation also showed that all trends remained the same, regardless of the pathway length (see online supplementary figures S1 and S2). A significant pathway (Treg→TGFβ→Th17) effect contributing to AML has been detected in our real data (table 3). Not only does a functional antagonism exist between Th17 and Treg cells, but there is also a dichotomy in their generation,[35] and Treg, TGFβ and Th17 have been confirmed to be associated with AML.[36] Our results further demonstrated that the Th17–Treg correlation balance was impaired in patients with AML, suggesting that the Th17–Treg imbalance potentially plays a role in the pathogenesis of AML. In summary, the bootstrap-based methods are preferred for identification of the pathway effect contributing to disease.

Table 3 The results of the pathway (Treg→TGFβ→Th17) effect contributing to acute myeloid leukaemia using six different methods

	p Value or 95% CI of D
Multivariate Δ	0.048
Exact	0.091
Unbiased	0.014
Bootstrap	0.034
Percentile bootstrap	(−0.202 to −0.011)
Bias-corrected bootstrap	(−0.214 to −0.020)

A reviewer suggested that we show the conventional association parameters, for example OR. It is indeed important to obtain some association or effect parameters such as OR. However, unlike one single factor, it is extremely hard to define the pathway levels since one pathway usually refers to many factors with a specific topology structure. The aim of our study is to develop a novel statistical method for detecting the pathway effect within a network between different disease status using the case–control design; thus, the exposure unit is the pathway rather than one single factor.

Although PEMs were proposed under a case–control design, they can also compare the difference between any two groups (different times or treatment). For instance, the main problem for drug developers is that they have to determine which one can be chosen as a priority to be a therapeutic target when faced with many disease-specific pathways involved in complex networks. PEM can provide the researchers with one guide to choose the pathway that most likely contributes to the disease from a statistical perspective. Our results also highlight the great potential of the proposed PEM usage in systems epidemiology in advancing medicine research, and Th17–Treg balance may be a promising therapeutic approach in patients with AML. Although our proposed PEMs can be extended to a matched and nested case–control design, the distribution of PEM-D is still difficult to determine. The reason why the distribution of the difference in pathway effect between cases and controls is unknown is that our proposed method is derived from the multiplication of some correlated standardised coefficients. There seems to be little correlation with the epidemiological design.

One possible drawback of the proposed bootstrap-based methods is the computation burden on the bootstrap procedure used to evaluate the CI and SD of D, and theoretical justification work is highly desirable in future studies.

CONCLUSIONS

In this paper, we proposed two typical PEM to detect the pathway effect within a network between different disease status under a case–control design within the framework of systems epidemiology. Bootstrap-based PEM are valid and powerful for identifying the specific pathway contributing to disease, thus potentially providing new insight into the underlying mechanisms and more comprehensive ways to study the disease effects of specific pathways.

Acknowledgements The authors wish to acknowledge their colleagues for their invaluable work and the participants who agreed to participate in the data collection.

Contributors All authors conceptualised the study, acquired and analysed the data and prepared the manuscript, and read and approved the final manuscript.

Funding This work was supported by grants from National Natural Science Foundation of China (grant number 31200994 and 31071155).

Competing interests None.

Patient consent Obtained.

Ethics approval Medical Ethical Committee of Qilu Hospital, Shandong University, China.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

REFERENCES

1. Haring R, Wallaschofski H. Diving through the "-omics": the case for deep phenotyping and systems epidemiology. *OMICS* 2012;16:231–4.
2. Fallin MD, Kao WH. Is "X"-WAS the future for all of epidemiology. *Epidemiology* 2011;22:457–9; discussion 467–8.
3. Lund E, Dumeaux V. Systems epidemiology in cancer. *Cancer Epidemiol Biomarkers Prev* 2008;17:2954–7.
4. Hu FB. Metabolic profiling of diabetes: from black-box epidemiology to systems epidemiology. *Clin Chem* 2011;57:1224–6.
5. Leung EL, Cao ZW, Jiang ZH, *et al*. Network-based drug discovery by integrating systems biology and computational technologies. *Brief Bioinform* 2013;14:491–505.
6. Berg EL. Systems biology in drug discovery and development. *Drug Discov Today* 2014;19:113–25.
7. Zhang X, Wang W, Xiao K, *et al*. Translational medicine: application of omics for drug target discovery and validation. In: William CS, ed. *An omics perspective on cancer research*. Springer: The Netherlands, 2010:235–47.
8. Wu X, Jiang R, Zhang MQ, *et al*. Network-based global inference of human disease genes. *Mol Syst Biol* 2008;4:189.
9. Yates PD, Mukhopadhyay ND. An inferential framework for biological network hypothesis tests. *BMC Bioinformatics* 2013;14:94.
10. Adourian A, Jennings E, Balasubramanian R, *et al*. Correlation network analysis for data integration and biomarker selection. *Mol Biosyst* 2008;4:249–59.
11. Wang K, Li M, Bucan M. Pathway-based approaches for analysis of genomewide association studies. *Am J Hum Genet* 2007;81:1278–83.
12. Chen L, Zhang L, Zhao Y, *et al*. Prioritizing risk pathways: a novel association approach to searching for disease pathways fusing SNPs and pathways. *Bioinformatics* 2009;25:237–42.
13. Isci S, Ozturk C, Jones J, *et al*. Pathway analysis of high-throughput biological data within a Bayesian network framework. *Bioinformatics* 2011;27:1667–74.
14. Yu K, Li Q, Bergen AW, *et al*. Pathway analysis by adaptive combination of p values. *Genet Epidemiol* 2009;33:700–9.
15. Li C, Han J, Shang D, *et al*. Identifying disease related sub-pathways for analysis of genome-wide association studies. *Gene* 2012;503:101–9.
16. Bollobás B. *Modern graph theory*. Spinger-Verlag, 1998.
17. Efron B, Tibshirani RJ. *An introduction to the bootstrap*. New York: Chapman & Hall, 1993.
18. Efron B. *Better bootstrap confidence intervals*. *J Am Stat Assoc* 1987;82:171–85.
19. Goodman LA. The variance of the product of K random variables. *J Am Stat Assoc* 1962;57:54–60.
20. Sobel ME. Asymptotic confidence intervals for indirect effects in structural equation models. *Sociol Methodol* 1982;13:290–312.
21. Volpe E, Servant N, Zollinger R, *et al*. A critical function for transforming growth factor-beta, interleukin 23 and proinflammatory cytokines in driving and modulating human T(H)-17 responses. *Nat Immunol* 2008;9:650–7.
22. Yang L, Anderson DE, Baecher-Allan C, *et al*. IL-21 and TGF-beta are required for differentiation of human T(H)17 cells. *Nature* 2008;454:350–2.
23. Wang X, Gulbahce N, Yu H. Network-based methods for human disease gene prediction. *Brief Funct Genomics* 2011;10:280–93.
24. Saha S, Roman T, Galante A, *et al*. Network-based approaches for extending the Wnt signalling pathway and identifying context-specific sub-networks. *Int J ComputBiol Drug Des* 2012;5:185–205.
25. Ideker T, Sharan R. Protein networks in disease. *Genome Res* 2008;18:644–52.
26. Taylor IW, Wrana JL. Protein interaction networks in medicine and disease. *Proteomics* 2012;12:1706–16.
27. Jones RB, Gordus A, Krall JA, *et al*. A quantitative protein interaction network for the ErbB receptors using protein microarrays. *Nature* 2006;439:168–74.
28. Baranzini SE, Galwey NW, Wang J, *et al*. Pathway and network-based analysis of genome-wide association studies in multiple sclerosis. *Hum Mol Genet* 2009;18:2078–90.
29. Miller MA, Feng XJ, Li G, *et al*. Identifying biological network structure, predicting network behavior, and classifying network state with high dimensional model representation (HDMR). *PLoS ONE* 2012;7:e37664.
30. Di CB, Falda M, Toffolo G, *et al*. SimBioNeT: a simulator of biological network topology. *IEEE/ACM Trans Comput Biol Bioinform* 2012;9:592–600.
31. Kim DC, Wang X, Yang CR, *et al*. Learning biological network using mutual information and conditional independence. *BMC Bioinformatics* 2010;11(Suppl):S9.
32. Barabasi AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. *Nat Rev Genet* 2011;12:56–68.
33. Barabasi AL, Oltvai ZN. Network biology: understanding the cell's functional organization. *Nat Rev Genet* 2004;5:101–13.
34. Spirin V, Mirny LA. Protein complexes and functional modules in molecular networks. *Proc Natl Acad Sci USA* 2003;100:12123–8.
35. Bettelli E, Carrier Y, Gao W, *et al*. Reciprocal developmental pathways for the generation of pathogenic effector TH17 and regulatory T cells. *Nature* 2006;441:235–8.
36. Tian T, Yu S, Wang M, *et al*. Aberrant T helper 17 cells and related cytokines in bone marrow microenvironment of patients with acute myeloid leukemia. *Clin Dev Immunol* 2013;2013:915873.