Supporting Information

Title: Interaction Networks of Trauma Providers Are Associated with Length of Stay

Authors: You Chen[1], Mayur B. Patel[2,3,4], Candace D. McNaughton[5], Bradley A. Malin[1,6,7]

Author Affiliations:
[1]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN
[2]Department of Surgery, Division of Trauma, Surgical Critical Care, and Emergency General Surgery, Vanderbilt University Medical Center, Nashville, TN
[3]Department of Neurosurgery, Vanderbilt University Medical Center, Nashville, TN
[4]Department of Hearing & Speech Sciences, Vanderbilt University Medical Center, Nashville, TN
[5]Department of Emergency Medicine, Vanderbilt University Medical Center, Nashville, TN
[6]Department of Biostatistics, School of Medicine, Vanderbilt University, Nashville, TN
[7]Department of Electrical Engineering & Computer Science, School of Engineering, Vanderbilt University, Nashville, TN

To Whom Correspondence Should be Addressed:

You Chen, Ph.D.
2525 West End Ave, Suite 1475
Department of Biomedical Informatics
Vanderbilt University Medical Center
Nashville, TN 37203 USA
Email: you.chen@vanderbilt.edu

Keywords: Interaction Network; Electronic Medical Records; Length of Stay; Network Analysis; Statistical Modeling; Trauma

S1: Spectral Co-Clustering Algorithm to Infer Groups of Patient Encounters and Operational Areas

Co-clustering aims to simultaneously cluster the rows and columns of an input data matrix. It seeks blocks of rows and columns that are inter-correlated as opposed to seeking similar rows or columns as transpires in traditional clustering methods. In the co-clustering, the row cluster prototypes incorporate column clustering information and vice versa [1]. Spectral co-clustering employs matrix decomposition techniques (e.g., eigenvectors selection) and formalizes co-clustering as a bipartite graph partition problem (e.g., finding minimum cut vertex partitions in the bipartite graph) [2-3]. There is evidence that shows spectral-based co-clustering can obtain an optimized co-clustering result [2-3].

We use a matrix *A* (of size of *m* and *n*) to store the interactions of healthcare professionals from *n* operational areas on the electronic medical records (EMRs) of *m* patient encounters. The matrix is shown on the right hand side of Equation 1. The cell value of the matrix, for instance $p_i o_j$, represents the number of actions the healthcare professionals from an operational area o_j performed on the EMRs of patient encounter p_i. We apply a spectral co-clustering algorithm to decompose the matrix and yield the clusters of rows (patient encounters) and clusters of columns (operational areas). For each patient encounter p_i, the algorithm will assign a cluster ID c_j; and for each operational area, the algorithm will assign a cluster ID c'_j. Patients with the same cluster ID are grouped into a cluster; and operation areas with the same cluster ID are also grouped into a cluster. The number of clusters is determined by the minimum cut of patients and operational areas. By applying this method to the VUMC data, we obtained three clusters of patient encounters and three clusters of operational areas as shown in Fig S1.

$$\left(\begin{bmatrix} p_1 & c_i \\ p_2 & c_j \\ p_3 & c_i \\ \vdots & \\ p_m & c_i \end{bmatrix}, \begin{bmatrix} o_1 & c'_i \\ o_2 & c'_j \\ o_3 & c'_i \\ \vdots & \\ o_n & c'_i \end{bmatrix}\right) = \text{Co-Clustering}\left(\begin{bmatrix} p_1o_1 & p_1o_2 & p_1o_i & p_1o_{i+1} & \cdots & p_1o_n \\ p_2o_1 & p_2o_2 & p_2o_i & p_2o_{i+1} & \cdots & p_2o_n \\ \vdots & & \vdots & & & \vdots \\ p_{m-1}o_1 & p_{m-1}o_2 & p_{m-1}o_i & p_{m-1}o_j & \cdots & p_{m-1}o_n \\ p_mo_1 & p_mo_2 & p_mo_i & p_mo_j & \cdots & p_mo_n \end{bmatrix}\right) \quad (1)$$

For each cluster of patient encounters, we investigate how the healthcare professionals from operational areas interact with each other to provide care to the patients. To do so, we build networks to

represent the interactions between the operational areas. The weight of the edge between a pair of operational areas is measured via cosine similarity. For instance, imagine there are 3 patient encounters (e.g., p_1, p_2, and p_3) in a cluster and two operational areas (e.g., o_1 and o_2) committed actions to their EMRs. Specifically, o_1 committed actions to p_1 and p_2 (represented as vector [1 1 0]) and o_2 committed actions to p_1, p_2 and p_3 (represented as vector [1 1 1]). In this case, the weight between o_1 and o_2 is the cosine similarity between vectors [1 1 0] and [1 1 1]. The number of actions for healthcare professionals from a cluster of operational areas on the EMRs of a cluster of patient encounters are stored in a matrix *B*, which is a reordered matrix of *A*. In *B*, the patient encounters or operational areas in a cluster are arranged together. An example of $\boldsymbol{P_1O_1}$, $\boldsymbol{P_1O_2}$, $\boldsymbol{P_1O_3}$ is shown in Fig S1.

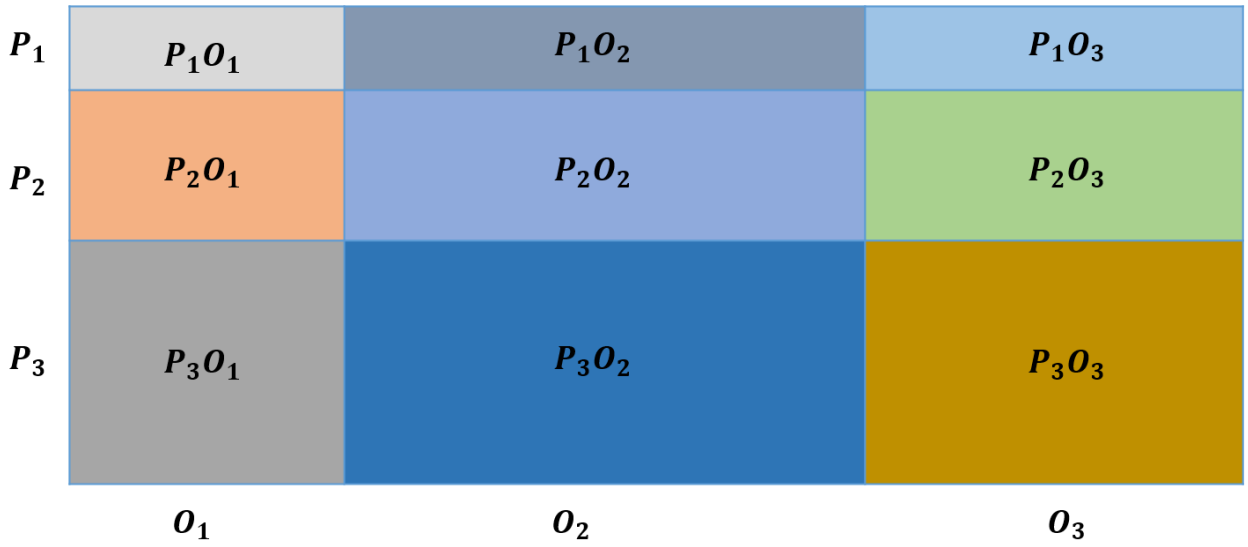

Fig. S1. After spectral co-clustering, patient encounters are clustered into three groups $\boldsymbol{P_1}$, $\boldsymbol{P_2}$, and $\boldsymbol{P_3}$; operational areas are clustered into three groups $\boldsymbol{O_1}$, $\boldsymbol{O_2}$, and $\boldsymbol{O_3}$.

According to *B*, we use cosine similarity to calculate the weight of edges for each pairs of operational areas, and then construct an interaction network for each patient group. The networks for the patient group $\boldsymbol{P_1}$, $\boldsymbol{P_2}$, and $\boldsymbol{P_3}$ are depicted in Fig S2, Fig S3, and Fig S4, respectively. The network affiliated with each patient encounter group is considered to be its interaction pattern. Since the edge weight is based on the cosine similarity, it is a continuous value in the (0,1) range.

Since weak relations between operational areas may be an artifact of noise, so for each interaction pattern (network of operational areas), we filtered edges with weights smaller than 0.1. This threshold was determined empirically based on a prior study on healthcare team identification [6]. Specifically, it was shown that the retention of relations with weights greater than 0.1 leads to the inference of interaction structures whose plausibility are deemed to be plausible by clinical and administrative experts.

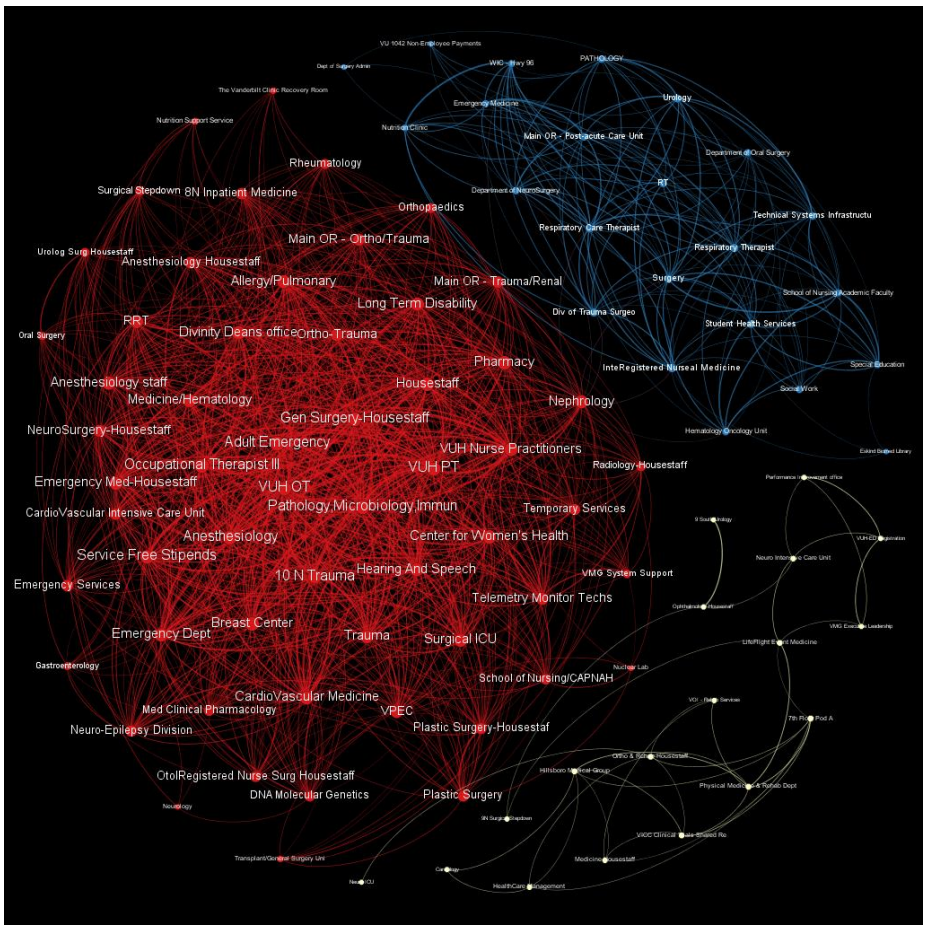

Fig. S2. A network of operational areas for patient group P_1. Each node is an operational area, which is labeled with an area name. The size of a node is determined by the connection degree of the node. The network is considered as an interaction pattern. There are three communities: i) acute- care; ii) post-acute care and iii) rehabilitation care in this interaction pattern.

For each inferred interaction pattern, we use a community detection algorithm [4] to infer communities of operational areas. We guided the algorithm using a heuristic that is based on the

optimization of modularity [4], which is efficient (in running time) and effective (in quality of communities) for weighted and undirected graphs. High modularity indicates dense connectivity of operational areas within communities and sparse connectivity across communities. The communities in each interaction pattern are represented using different colors (as shown in Fig S2, Fig S3 and Fig S4). At the same time, to identify the operational areas that have more connections (e.g., bridges to other operational areas) in the interaction pattern, we use connection centrality [5] to set the size of a node (operational area). Degree is a measure of connection centrality in a graph.

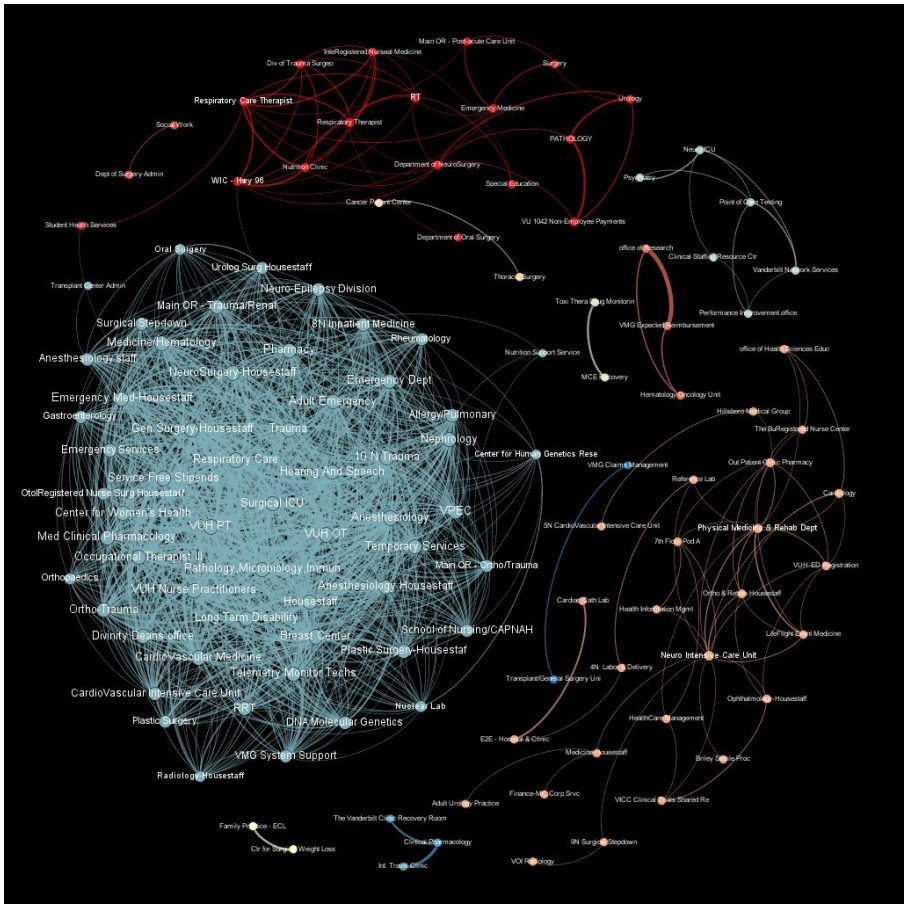

Fig. S3. A network of operational areas for patient group P_2. Each node is an operational area, which is labeled with an area name. The size of a node is determined by the connection degree of the node. The network is considered as an interaction pattern. There are three communities: i) acute- care; ii) post-acute care and iii) rehabilitation care in this interaction pattern.

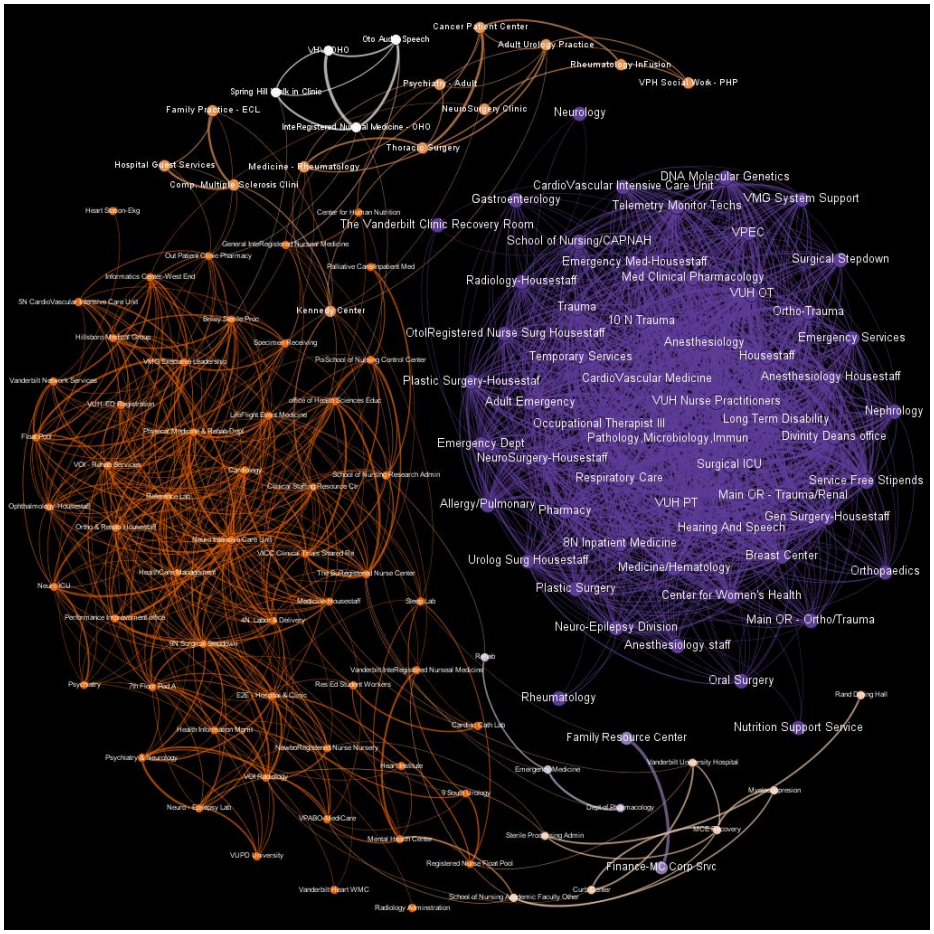

Fig. S4. A network of operational areas for patient group P_3. Each node is an operational area, which is labeled with an area name. The size of a node is determined by the connection degree of the node. The networks are considered as an interaction pattern. There are two communities: i) acute-care; ii) a combination of post-acute care and rehabilitation care in this interaction pattern.

S2: Measuring the Distribution of Potential Confounders

The 5,588 inpatient encounters were assigned 1,010 distinct PheWAS codes and 1,627 distinct procedure codes during their inpatient stays. They were affiliated with 8 insurance programs and 67 access action types. To evaluate if potential factors: 1) PheWAS codes, 2) procedure codes, 3) insurance programs, 4) age, 5) number of historical encounters, 6) admission months and 7) access action types were equally distributed on disparate patient encounter groups, we used a Pearson Correlation Coefficient (PCC) to measure the relationship between the distributions.

We list the measurement of the PheWAS codes as an example. To do so, we first calculate the distributions for each PheWAS code on the three patient encounter groups, the results of which are shown in Fig S5 to Fig S7. In Fig S5, each point represents a PheWAS code. Its *x*-axis value is the percentage of patient encounters in group P_1 assigned a specific PheWAS code, while its *y*-axis value is the percentage of patient encounters in group P_2 with the same PheWAS code. The distributions for procedure codes (Fig S8-S10), insurance programs (Fig S11-S13), age (Fig S14-S16), historical encounters (Fig S17-19), admission month (Fig 20-22) and access action types (Fig 23-25) over pairs of patient encounter groups is derived similarly. For the age distribution, we partitioned the population into 3-year age groups, beginning at 18 years old (minimum) and concluding at 102 years old (maximum). The PCC models were applied to obtain the corresponding P values.

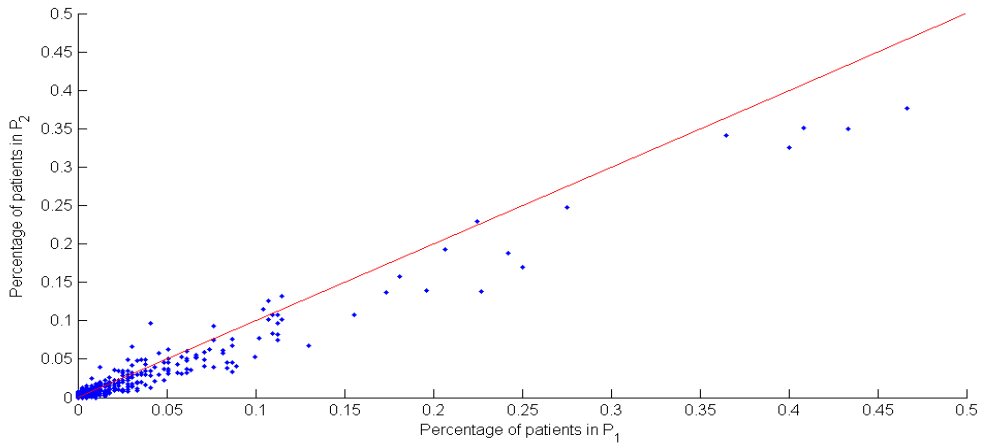

Fig. S5. Distributions of PheWAS codes between patient groups P_1 and P_2.

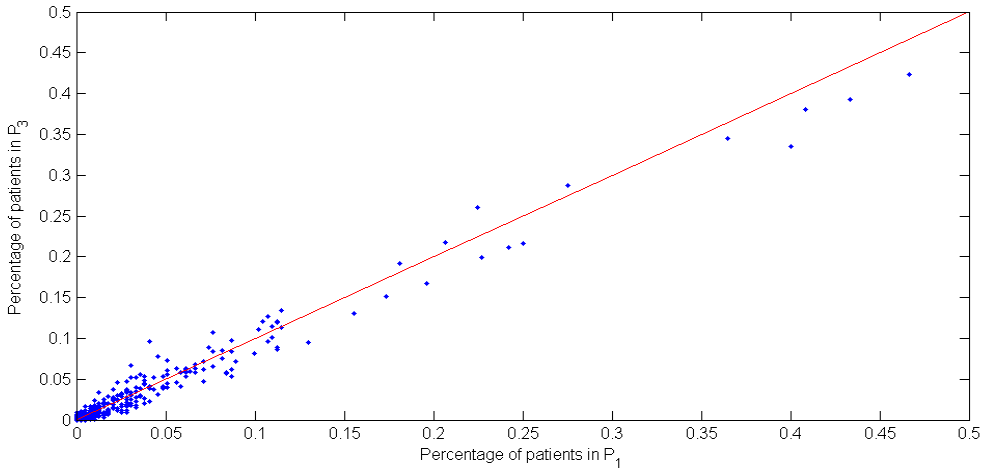

Fig. S6. Distributions of PheWAS codes between patient groups ***P1*** and ***P3***.

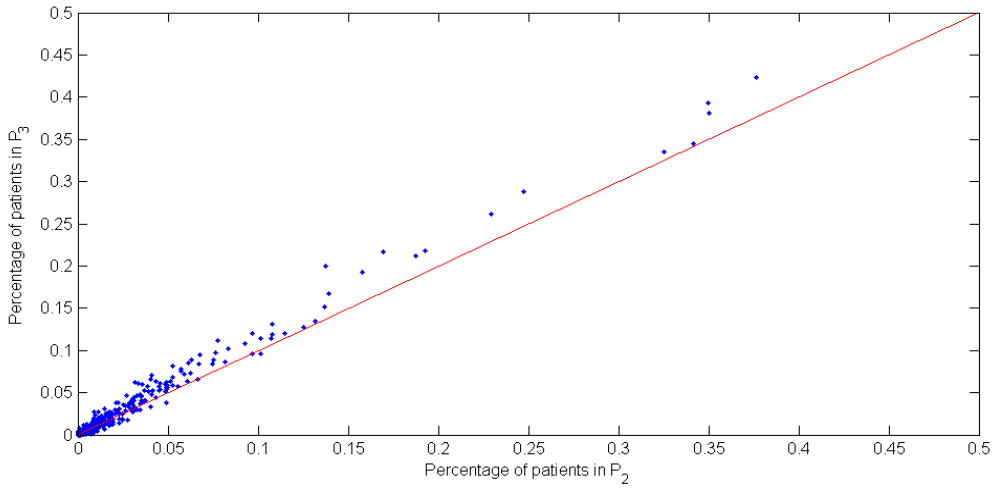

Fig. S7. Distributions of PheWAS codes between patient groups ***P2*** and ***P3***.

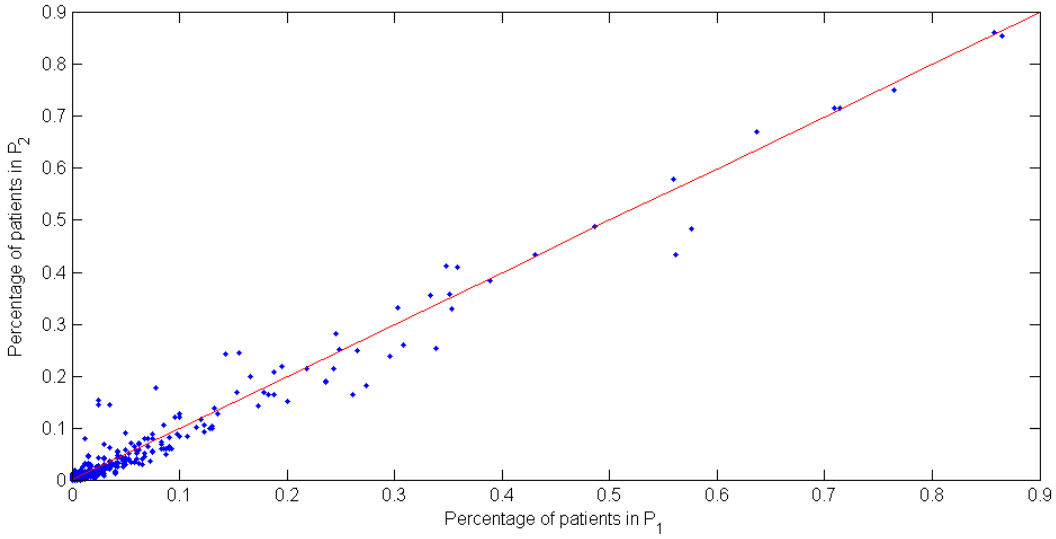

Fig. S8. Distributions of procedure codes between patient groups ***P₁*** and ***P₂***.

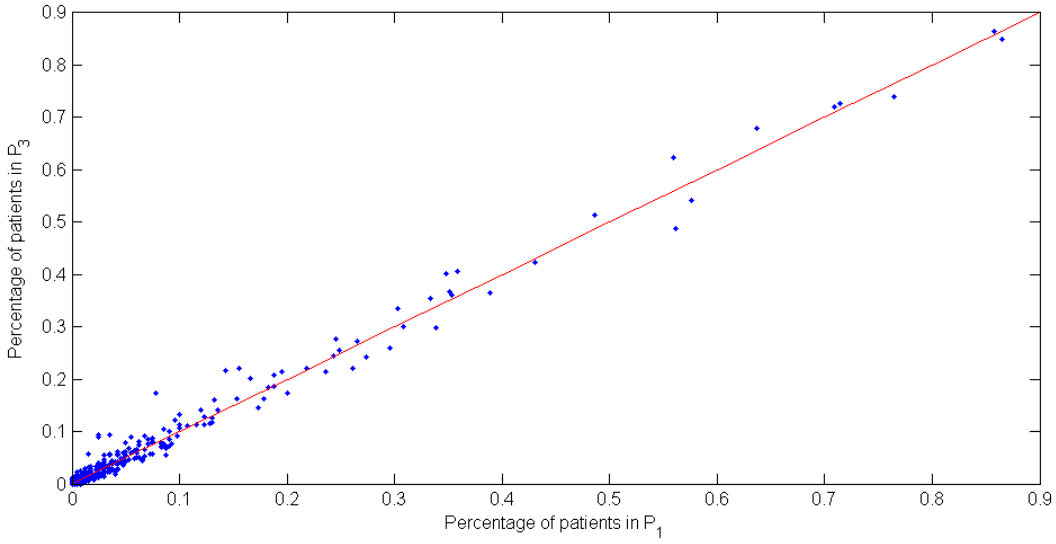

Fig. S9. Distributions of procedure codes between patient groups ***P₁*** and ***P₃***.

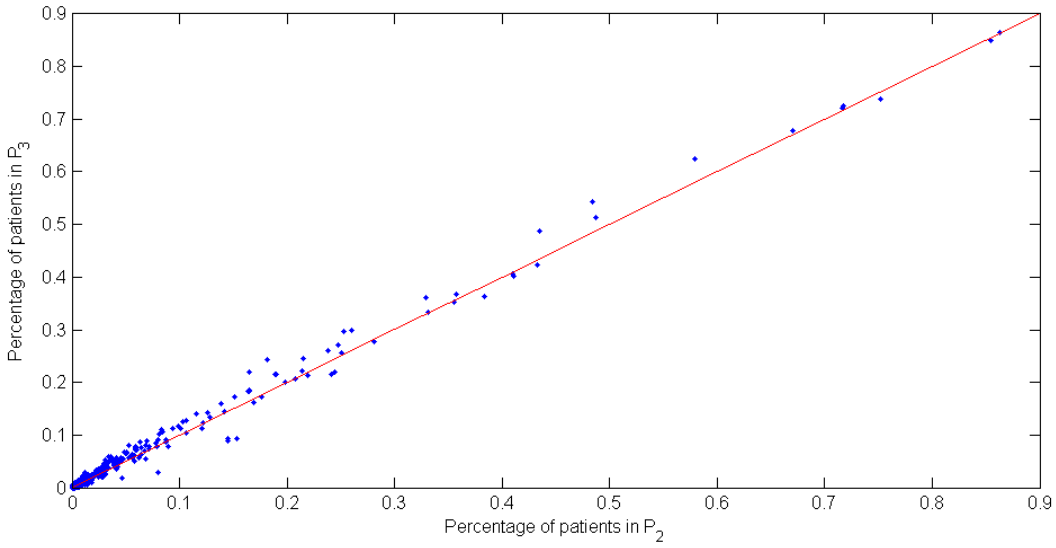

Fig. S10. Distributions of procedure codes between patient groups ***P₂*** and ***P₃***.

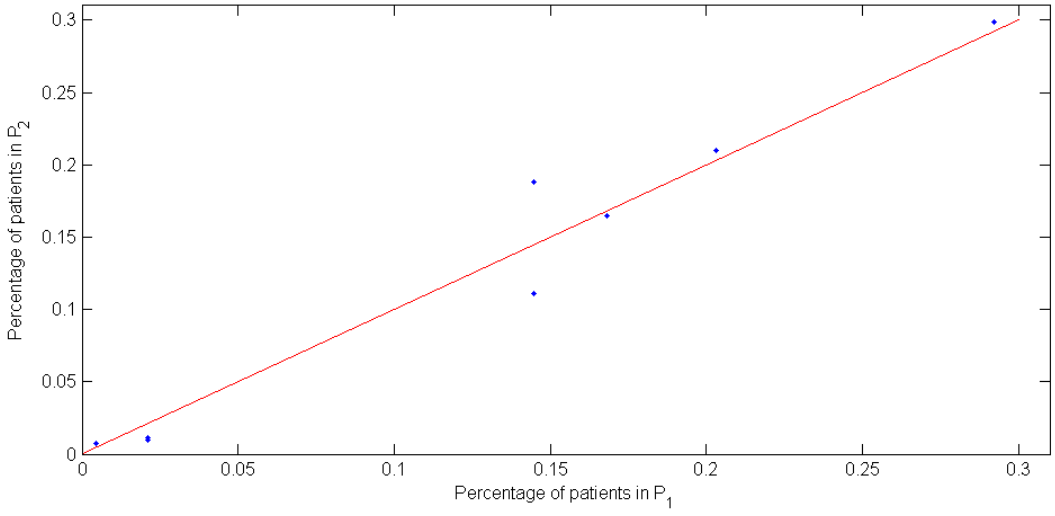

Fig. S11. Distributions of insurance programs between patient groups ***P₁*** and ***P₂***.

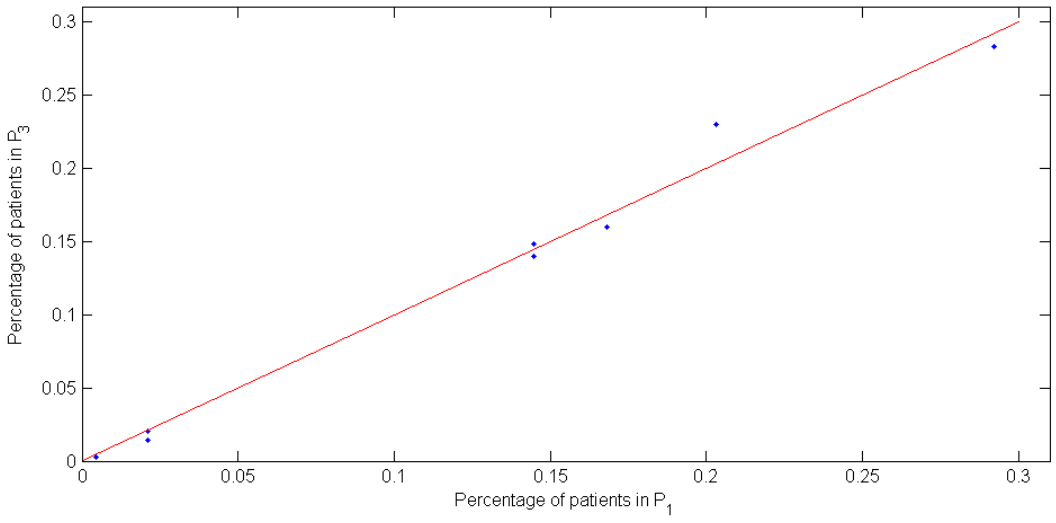

Fig. S12. Distributions of insurance programs on patient group P_1 and P_3.

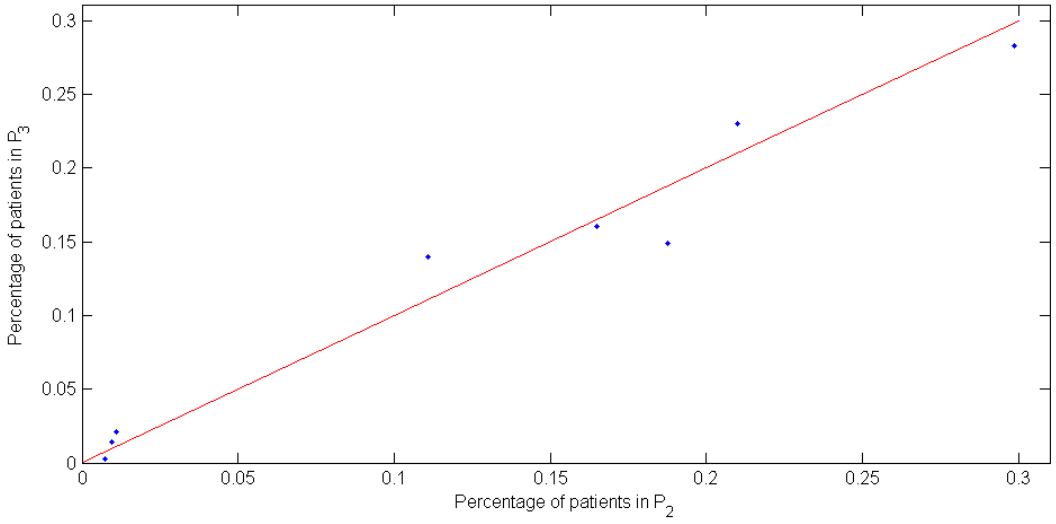

Fig. S13. Distributions of insurance programs between patient groups P_2 and P_3.

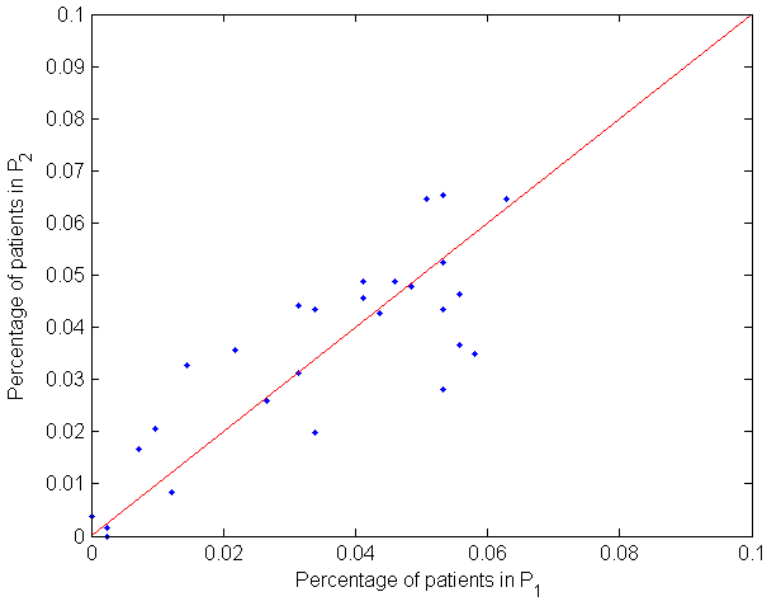

Fig. S14. Distributions of 3-year age groups for patient groups ***P₁*** and ***P₂***.

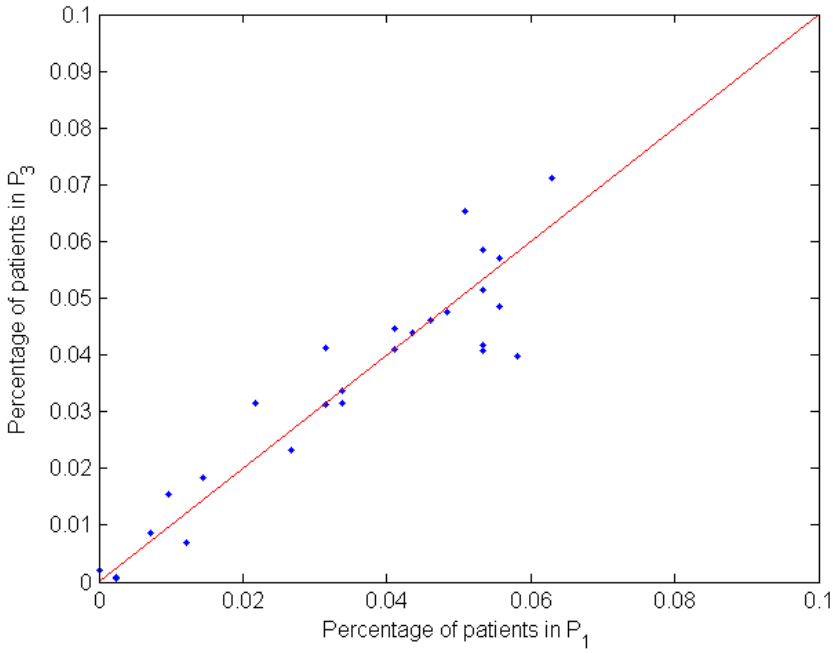

Fig. S15. Distributions of 3-year age groups for patient groups ***P₁*** and ***P₃***.

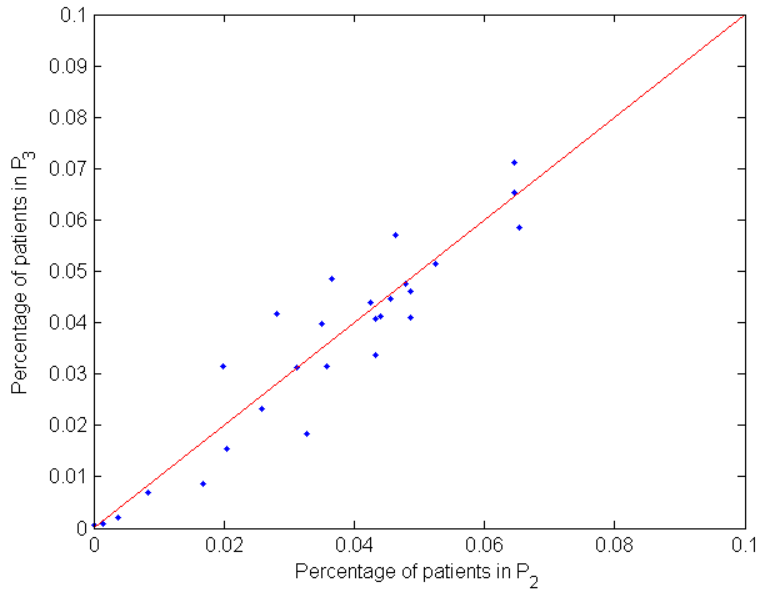

Fig. S16. Distributions of 3-year age groups for patient groups ***P₂*** and ***P₃***.

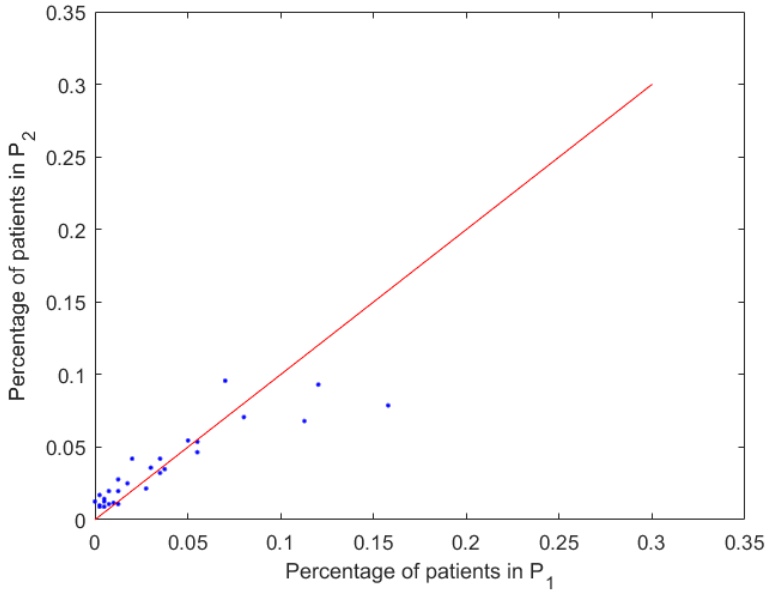

Fig. S17. Distributions of number of encounters per patients for patient groups ***P₁*** and ***P₂***.

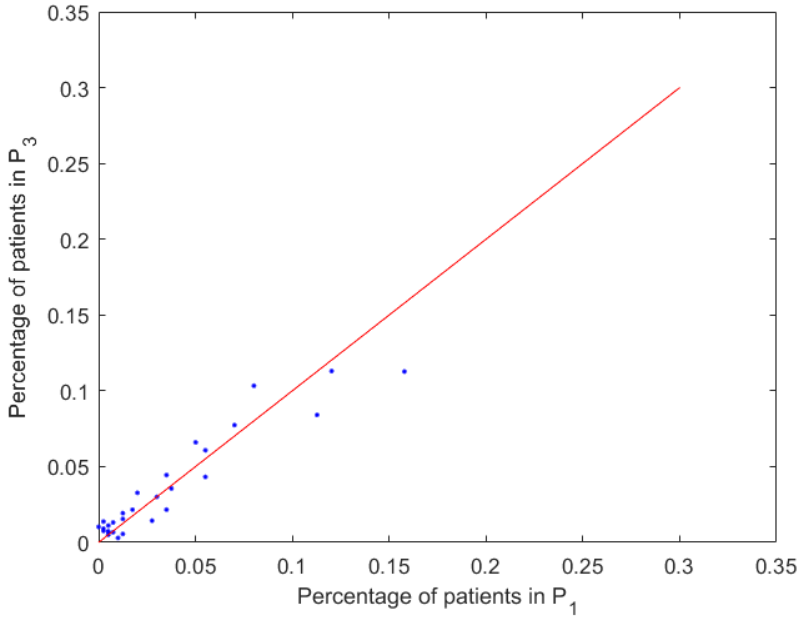

Fig. S18. Distributions of number of encounters per patients for patient groups ***P₁*** and ***P₃***.

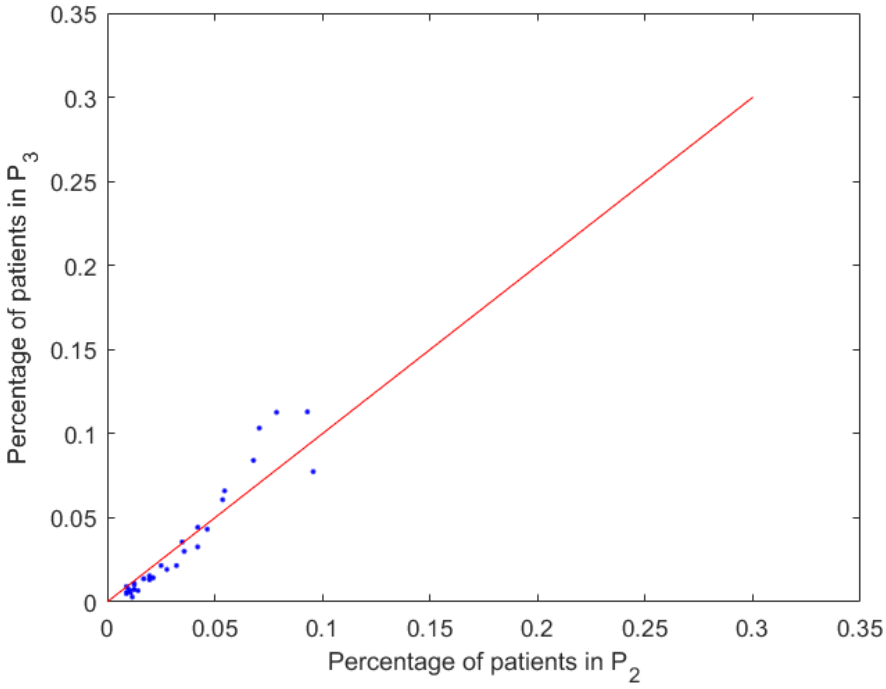

Fig. S19. Distributions of number of encounters per patients for patient groups ***P₂*** and ***P₃***.

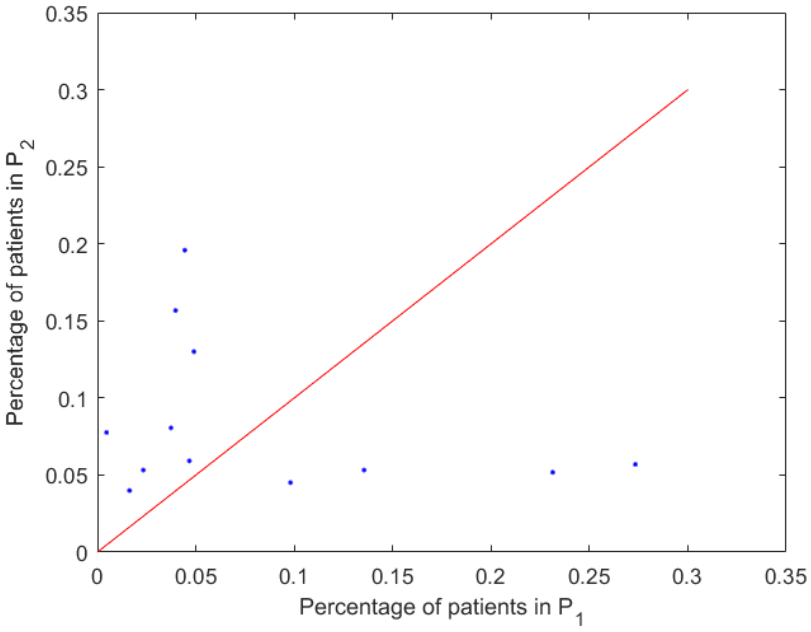

Fig. S20. Distributions of admission months for patient groups ***P₁*** and ***P₂***.

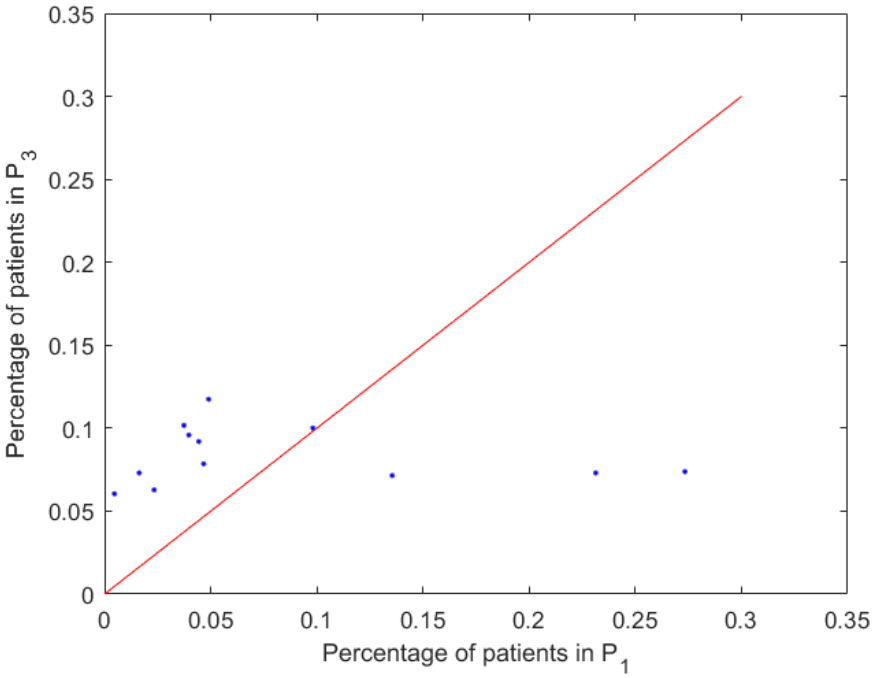

Fig. S21. Distributions of admission months for patient groups ***P₁*** and ***P₃***.

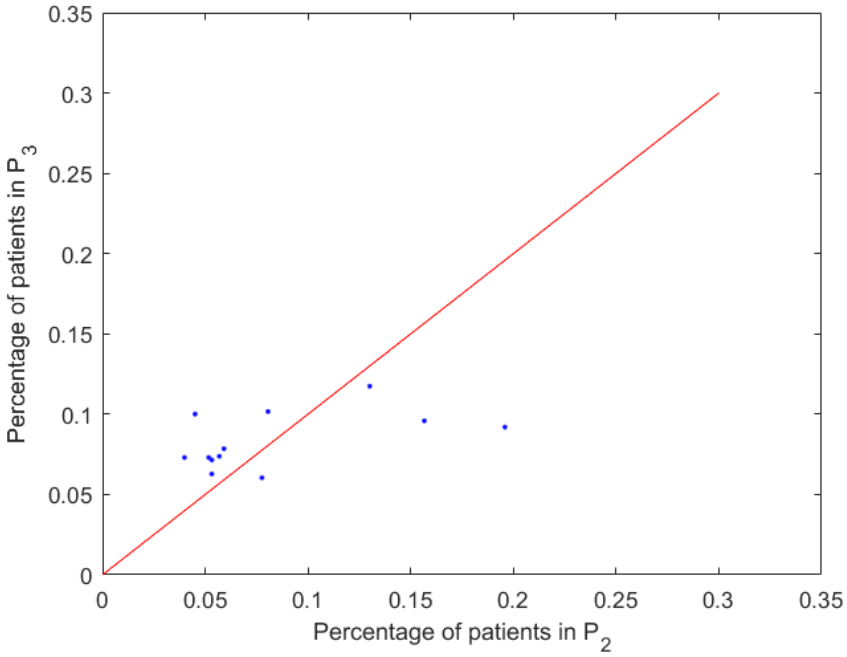

Fig. S22. Distributions of admission months for patient groups ***P₂*** and ***P₃***.

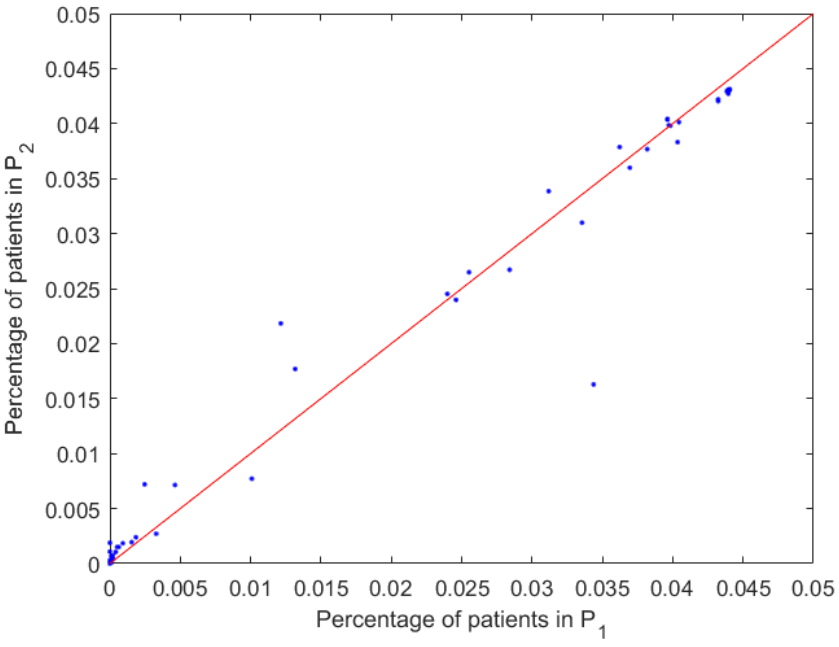

Fig. S23. Distributions of access action types for patient groups ***P₁*** and ***P₂***.

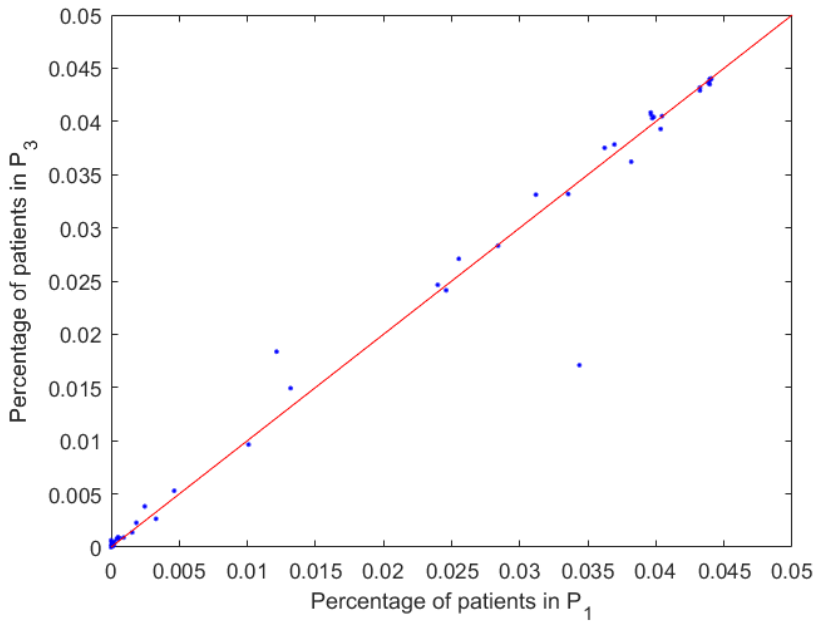

Fig. S24. Distributions of access action types for patient groups ***P₁*** and ***P₃***.

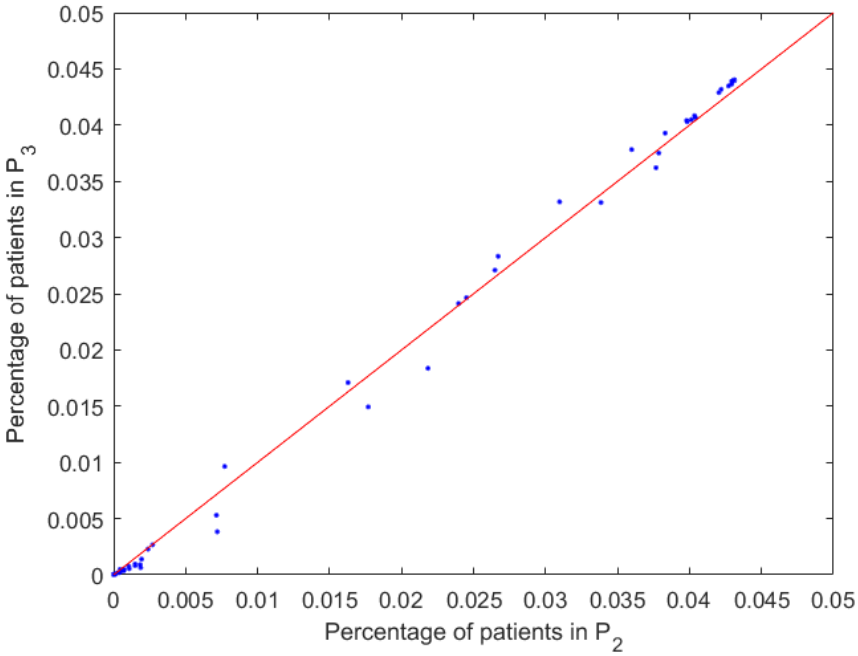

Fig. S25. Distributions of access action types for patient groups ***P₂*** and ***P₃***.

References

1. Labiod L, Nadif M. A unified framework for data visualization and coclustering. IEEE Trans Neural Netw Learn Syst. 2015; 26(9): 2194-2199.
2. Huang S, Wang H, et. al. Spectral co-clustering ensemble. Knowledge-Based Systems. 2015; 84: 46-55.
3. Kluger Y, Basri R, Chang JT, Gerstein M. Spectral biclustering of microarray data: coclustering genes and conditions. Genome Res. 2003; 13(4): 703-716.
4. Blondel V, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment. 2008; (10): P1000.
5. Yamagishi T, Gillmore MR, Cook KS. Network connections and the distribution of power in exchange networks. American Journal of Sociology. 1988; 93(4): 833-851.
6. Chen Y, Lorenzi NM, Sandberg WS, Wolgast K, Malin BA. Identifying collaborative care teams through electronic medical record utilization patterns. J Am Med Inform Assoc. 2017; 24(e1): e111-e120.