Statistical Analysis of Social Network Change

Teresa Danielle Schmidt
Portland State University

Follow this and additional works at: https://pdxscholar.library.pdx.edu/open_access_etds

Recommended Citation

This Dissertation is brought to you for free and open access. It has been accepted for inclusion in Dissertations and Theses by an authorized administrator of PDXScholar. For more information, please contact pdxscholar@pdx.edu.
Statistical Analysis of Social Network Change

by

Teresa Danielle Schmidt

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Systems Science

Dissertation Committee:
Martin Zwick, Chair
Neal Wallace
Thomas Kindermann
Jason Newsom

Portland State University
2020
Abstract

This project explores two statistical methods that infer social network structures and statistically test those structures for change over time: regression-based differential network analysis (R-DNA) and information theory-based differential analysis (I-DNA). R-DNA is adapted from bioinformatics and I-DNA employs reconstructability analysis.

This project applies both R-DNA and I-DNA to analyze Medicaid claims data from one-year periods before (May 2011 - Apr 2012) and after (Jan 2013 – Dec 2013) the formation of the Health Share of Oregon Coordinated Care Organization (CCO). The formation of CCOs was legislated by the state of Oregon in 2012 with the triple aim of improving health outcomes, reducing cost, and increasing patient satisfaction. We hypothesize that Health Share’s CCO formation would be followed by several changes in the healthcare delivery network.

Our primary aim is to contribute methodologically to the field of social network analysis by demonstrating and comparing these methods’ capacity for network inference and statistical testing. Our secondary aim is to contribute substantively to the field of health policy by identifying changes in the healthcare delivery network that followed Health Share’s CCO formation.

Application of R-DNA and I-DNA to claims data involves three steps: (a) the inference of billing provider networks, (b) measurement of a ‘distance’ between networks before and after Health Share’s CCO formation, and (c) statistical testing of this distance.
by resampling. Both methods afford what is akin to a network-level $t$ test for significant network difference between two time periods.

We use R-DNA and I-DNA to analyze three different billing networks: (1) the full network, including all 1,298 billing providers who operated within Health Share’s region and were actively billing during both time periods, (2) the network of the top 30 billing providers by patient volume, and (3) a care sector network defined in terms of primary, specialty, ancillary, mental/behavioral, facility, and ‘other’ billing provider types.

To address our primary aim, we demonstrate how four standard methods of data analysis can be used for network inference – correlation, multiple regression, partial least squares regression, and reconstructability analysis – and how three methods of data resampling can be used for statistical testing – permutation, one-sample bootstrapping, and two-sample bootstrapping. We conduct comparisons between these variants of R-DNA and I-DNA and make recommendations for how one might select from them in future social network analysis projects.

In terms of our secondary aim, applications of R-DNA and I-DNA reveal complementary insights and three specific patterns, all of which differ from our hypotheses. First, network connectivity consolidated after Health Share’s formation, producing stronger connections within the network’s core and weaker ones throughout its periphery. Changes in connectivity between primary and specialty care sectors reveal an increase in patients who received neither type of care, which may indicate increased obstacles to access. Changes between primary and mental/behavioral care
sectors reveal a decrease in the number of patients receiving both types of care, which may suggest that referrals did not increase as expected.

We conclude that both I-DNA and R-DNA are useful for inferring social network structures, for descriptively exploring the types of change that occur in them over time, and for testing whether those changes are statistically significant.
Table of Contents

Abstract........................................................................................................................................... i

List of Tables ....................................................................................................................................... vii

List of Figures ....................................................................................................................................... ix

Chapter 1: Introduction and Overview .............................................................................................. 1

  Method 1: Regression-based Differential Network Analysis ......................................................... 2

  Method 2: Information-Theory-Based Differential Network Analysis ......................................... 5

  Coordinated Care Organization (CCO) Insurance Claim Data ..................................................... 7

  Statistical Analysis .......................................................................................................................... 9

  Results ............................................................................................................................................. 11

  Aims and Contributions of this Work .............................................................................................. 12

Chapter 2: Background ................................................................................................................... 15

  Network Notation ............................................................................................................................ 16

  Measurement of Social Network Structure .................................................................................... 18

  Statistical Inference of Social Network Structure ........................................................................ 19

  Descriptive Analysis of Social Networks ...................................................................................... 23

  Statistical Analysis of Social Networks ........................................................................................ 25

  Methodological Contributions ....................................................................................................... 29

Chapter 3: Methods .......................................................................................................................... 30

  Permutation Tests .......................................................................................................................... 32

  One-Sample Bootstrap Testing ...................................................................................................... 34

  Two-Sample Bootstrap Testing ....................................................................................................... 34
Summary...................................................................................................................... 252

Chapter 7: Discussion and Summary................................................................. 254

Introduction ...................................................................................................... 254

Methodological Contributions .................................................................... 256

Substantive Contributions ........................................................................ 265

Limitations .................................................................................................... 271

Future Work ................................................................................................. 276

Conclusion ..................................................................................................... 280

References ...................................................................................................... 282

Appendix A: Categorization of NUCC Fields...................................................... 306

Appendix B: Relationships in the Top 30 Networks by RA Inference.............. 308

Appendix C: Resampling Distribution Information for Statistical Testing .......... 309
List of Tables

Table 2.1. Example of Weighted, Directed Network Data in Matrix Form ......................... 17
Table 3.1. Comparison of Traditional versus Bootstrapped t value Calculation .............. 36
Table 3.2. Arrangement of CCO data for statistical testing.......................................... 37
Table 3.3. Examples of Hypothesis Testing with I-DNA and R-DNA ............................. 42
Table 3.4. Comparison of Desirable Features via I-DNA and R-DNA ......................... 85
Table 3.5. Defining Associations AB, AC and BC from Observed Distribution ABC ....... 88
Table 4.1. Distribution of Patients and Providers across time periods ......................... 103
Table 4.2. Data in Claim Level Format ............................................................................ 106
Table 4.3. Patient by Billing Provider Format ............................................................... 106
Table 4.5. Primary taxonomies for performing providers by care sector ...................... 110
Table 4.6. Taxonomy, performing, and billing provider counts by care sector.......... 112
Table 4.7. Rates of claims and providers, before and after Health Share’s formation ... 114
Table 4.8. Counts of Patients and Claims by Care Sector ........................................... 115
Table 5.1. Summary of Changes in Full Network Visualizations ................................. 132
Table 5.2. Top 30 billing providers by patient volume ................................................. 136
Table 5.3. Summary of change in the top 30 networks as indicated by each method of inference................................................................................................................. 138
Table 5.4. Summary of best models for the top 30 network inferred by RA............... 142
Table 5.5. Top 30 Network edges found by RA, Correlation, and SBT ...................... 156
Table 5.6. Summary of Pairs in Visualized Networks ....................................................... 172
Table 5.7. Mean absolute differences in the care sector network after Health Share..... 200
Table 5.8. Largest changes in calculated probabilities for care sector network as inferred by RA
Table 5.9. Summary of best models for the care sector network inferred by RA
Table 5.10. Summary of Changes in Care Sector Network Visualizations
Table 5.11. Patient utilization of primary and specialty care
Table 5.12. Change in patient utilization of primary and specialty care
Table 5.13. Proportions of patients receiving mental health and primary care
Table 5.14. Observed probabilities for mental and primary care utilization (in bold), and their difference from expected probabilities (in red or blue)
Table 5.15. Summary of findings for care sector network
Table 6.1. Statistical tests of network change
Table 6.2. Top 30 reference distributions by inference and resampling method
Table 6.3. Care sector reference distributions by resampling method
Table 6.4. Primary and specialty sector reference distributions by resampling method
Table 6.5. Primary and mental/behavioral sector reference distributions by resampling method
Table 6.6. Top 30 and care sector reference distributions relative to observed distances
Table 6.7. Connection-specific reference distributions relative to observed distance
List of Figures

Figure 1.1. Change in Network Structure ................................................................. 1

Figure 3.1. Overview of the analytical procedures for R-DNA and I-DNA................. 30

Figure 3.2. Comparison of hypothetical RA network structures .............................. 48

Figure 3.3. Hypothetical Connectivity Scores by PC (Left) and MLR (Right) ............. 70

Figure 3.4. Illustration of Connectivity Score Calculation in PLS ............................. 76

Figure 3.5. Hypothetical Histogram of Connectivity Scores ...................................... 80

Figure 4.1. Performing and Billing Providers in a Hypothetical Coordinated Care Organization ........................................................................................................... 99

Figure 4.2. Timeline of Datasets from Before and After Health Share’s Formation..... 101

Figure 4.3. Histogram depicting how many patients had each number of billing providers before Health Share’s Formation ................................................................. 113

Figure 5.1. Full network of 1299 billing providers by correlation, before Health Share’s formation ........................................................................................................... 125

Figure 5.2. Full network of 1299 billing providers by correlation, after Health Share’s formation ........................................................................................................... 125

Figure 5.3. Full network of 1299 billing providers by SBT, before Health Share’s formation ........................................................................................................... 129

Figure 5.4. Full network of 1299 billing providers by SBT, after Health Share’s formation ........................................................................................................... 129

Figure 5.5. Top 30 billing provider network by RA ..................................................... 144

Figure 5.6. Complexity of Associations in the Before and After Network by RA ........ 145
Figure 5.7. Top 30 billing provider network by correlation (r≥.1).......................... 148
Figure 5.8. Correlation weights before and after Health Share’s formation................. 149
Figure 5.9. Absolute strengths of the strongest (left) and weakest (right) correlations.. 150
Figure 5.10. Top 30 billing provider network by SBT (SBT≥.001)............................. 153
Figure 5.11. Strengths of the strongest (left) and weakest (right) SBT connections..... 155
Figure 5.12. Top 30 billing provider network by RA, SBT, and correlation.............. 157
Figure 5.13. Strongest 200 connection weights for SBT and correlation............... 160
Figure 5.14. Scatterplot and contingency table for a connection found by information
theory but not found by correlation ................................................................. 163
Figure 5.15. Scatterplot and contingency table for a relationship found by correlation but
not by information theory ................................................................. 164
Figure 5.16. Network of connections found by RA, correlation and SBT .............. 166
Figure 5.17. Top 30 billing provider network by multiple regression (B ≥.145)....... 168
Figure 5.18. Top 30 billing provider network by PLS (s ≥.08)................................. 169
Figure 5.19. Strongest 200 connection weights for regression and PLS ............ 170
Figure 5.20. Top 30 billing provider network by correlation, multiple regression, and PLS
......................................................................................................................... 173
Figure 5.21. Percent of connections above threshold during only one period, full range
......................................................................................................................... 175
Figure 5.22. Percent of connections above threshold during one period, in restricted range
......................................................................................................................... 177
Figure 5.23. Amount of variance explained by correlation, multiple regression, and PLS, before Health Share’s formation (at left) and after it (at right)......................... 180

Figure 5.24. Absolute strengths of the strongest (left) and weakest (right) regression coefficients.............................................................................................................. 183

Figure 5.25. Absolute strengths of the strongest (left) and weakest (right) PLS weights ................................................................................................................................. 183

Figure 5.26. Percentages of patients’ claim frequencies in nominal categories .......... 195

Figure 5.27. Change as inferred in the care sector network.............................................. 196

Figure 5.28. Care sector network before and after Health Share’s formation as inferred through SBT.................................................................................................................. 198

Figure 5.29. Care sector network before and after Health Share’s formation as inferred through PLS ................................................................................................................. 199

Figure 5.30. Pairwise connections by correlation, before and after Health Share formed ........................................................................................................................................... 205

Figure 5.31. Pairwise connections by SBT, before and after Health Share Formed ...... 206

Figure 5.32. Pairwise relationships by PLS, before and after Health Share formed ...... 208
Chapter 1: Introduction and Overview

This dissertation brings two statistical approaches to the field of social network analysis, with the capacity for inferring network structures and testing them for significant change over time. Here, change in network structure includes the appearance or disappearance of connections, as well as the strengthening or weakening of connections that persist (see Figure 1.1). The two approaches applied here combine several existing methods that are currently used in other fields. Regression-based Differential Network Analysis (R-DNA) is borrowed from the field of bioinformatics, and Information-theory-based Differential Network Analysis (I-DNA) is developed from existing methods in information theory. Both approaches were applied to a healthcare delivery network to statistically test whether its structure changed following the legislation of a healthcare transformation in the state of Oregon.

Figure 1.1. Change in Network Structure

Network analysis (NA) is a general methodology for measuring, representing and analyzing structural connectivity among interacting elements in a system (Butts, 2008; Wasserman & Faust, 1994). In the social sciences, social network analysis is widely implemented as a research tool for a variety of social phenomena in psychology (Butts, 2008), sociology (Palloni, 1998), and program evaluation (Durland & Fredericks, 2006).
Across these application areas, network analysis has often relied on direct measurement of social relations, often requiring time intensive methods of data collection about each possible relationship in a network. But recent advances in technology (e.g., Bluetooth, smart phones, online forums) allow for many new types of automatic data collection that can be used to statistically infer social network structures.

Surprisingly, statistical inference of network structure is virtually absent from the field of social network analysis (see Kindermann & Gest, 2011; Ver Steeg & Galstyan, 2012 for two exceptions), as are statistical tests of difference between networks. The use of statistics in social network analysis is usually restricted to parameter estimation, either to characterize a network or the network-related attributes of its members (Carrington, Scott, & Wasserman, 2005; Fienberg, 2012). Several procedures exist for testing whether two networks are significantly similar (Krackhardt, 1987; Snijders, Dormaar, & Dijkstra, 1990), but these do not allow a researcher to assert whether they are significantly different. This distinction is analogous to being able to conduct a correlation between two variables but not a dependent-samples t test. Similarity tests do not allow one to investigate whether a social network changed significantly after an event or intervention.

**Method 1: Regression-based Differential Network Analysis**

Regression-based Differential Network Analysis (R-DNA) uses linear regression to derive network structures, a mean absolute difference metric to measure the distance between networks, and resampling techniques to test the significance of that difference. Its parent method, differential network analysis (DNA), is a descriptive method for comparing the connectivity of two networks (Horvath, 2011) that has been used primarily
in the field of bioinformatics to study gene and protein networks (Fuite, Vernon, & Broderick, 2008; Fuller et al., 2007; Weckwerth, Loureiro, Wenzel, & Fiehn, 2004). With DNA, pairwise relationships between connected members are defined with weighted ‘connectivity scores’ (e.g., correlations) and those scores are compared across datasets from different treatment conditions. These descriptive comparisons reveal connections that are stronger, weaker, present, or absent in a condition of interest.

Only recently has descriptive DNA been augmented with resampling techniques to allow for hypothesis testing (Gill, Datta, & Datta, 2010; Yates & Mukhopadhyay, 2013). This hypothesis testing relies on regression-based connectivity scores, such as correlations, partial correlations, or partial least squares, ¹ which is why we (here) refer to this method as Regression-based Differential Network Analysis, or R-DNA, when these types of network inference are coupled with the statistical comparison of network structures. Regression-based scores have the desirable feature of being standardized, so a set of differences in connectivity can be aggregated into a summarizing statistic and then subjected to a hypothesis test. With the additional capacity for statistical testing, R-DNA has been applied beyond the field of bioinformatics, to study change in environmental systems (Chen et al., 2013). This dissertation demonstrates the application of R-DNA to the field of social network analysis, by testing whether a healthcare delivery network changed significantly after a state-legislated change.

¹ Partial least squares may make network differences most apparent (Pihur, Datta, & Datta, 2008)
To use R-DNA for network comparison, network structures are first inferred by constructing datasets that define a continuous variable for each network member and a case for each observation of members’ behavior. For example, in this dissertation we define a continuous variable for each healthcare (billing) provider and a case for each patient indicating the number of claims that each provider billed for them. Regression based techniques are then used to infer the relationships between each pair of network members, generating a matrix of connectivity scores that indicate each network’s structure. Once network inference has been applied separately to the observations from two different time periods, connectivity scores can be compared to identify differences in network structure. These differences can be aggregated into a distance statistic that summarizes the extent to which the network changed.

By itself, the distance statistic is only a descriptive measure. This distance might be trivially different from zero (indicating no change) and might have arisen from sampling error or from chance variations in network members’ behavior. So, to test this distance for statistical significance, R-DNA employs one of several resampling methods: permutations, one-sample bootstrapping, or two-sample bootstrapping. Each resampling method enables us to estimate the probability that the distance could have resulted from chance alone, even if the null hypothesis were true and the two networks were not systematically different.

---

2 For the purposes of graphical network displays, a threshold for the magnitude of connectivity scores can be used to disregard weak connections between social entities.
Permutation testing (also called re-randomization) involves combining the cases from both datasets into a superset and then randomly sorting them into a new (permuted) pair of datasets. One-sample bootstrapping involves combining the cases from both datasets into a superset and then drawing cases randomly with replacement to create a new (bootstrapped) pair of datasets. And two-sample bootstrapping involves resampling with replacement from each dataset separately, without ever combining the two into a superset. With any of the above resampling techniques, many pairs of resampled datasets are created, and matrices of regression-based connectivity scores and distances are calculated for each pair. Resampled distances are then aggregated into a distribution.

In the case of permutations and one-sample bootstrapping, the distribution of resample-based distances can be directly used to calculate the probability of a Type I error. If fewer than 5% of resample-based distances are as great as or greater than the observed distance, the networks from before and after can be considered significantly different. In the case of two-sample bootstrapping, the set of resample-based distances can be used to calculate a standard error for the observed distance. Then one uses this standard error to compare the observed distance to a standard \( t \) distribution and determine statistical significance.

**Method 2: Information-Theory-Based Differential Network Analysis**

Information-theory-based Differential Network Analysis (I-DNA) uses reconstructability analysis (RA) to derive network structures, uses a sum of absolute differences in probabilities metric to measure the distance between networks, and uses any of the same resampling techniques as above to test the significance of that distance.
Based on information theory, RA is a data mining method that can detect deviations from mutual independence among variables (Klir, 1986; Krippendorff, 1986; Zwick, 2004). This method has been applied to a wide variety of fields, including bioinformatics (Kramer, Westaway, Zwick, & Shervais, 2012), environmental studies (Zwick, Shu, & Koch, 1995) and decision analysis (Johnson & Zwick, 2000), but to the author’s knowledge has not yet been used to investigate social network phenomena.

In using RA to infer a social network structure, each network member is represented as a nominal variable with some number of behavior states, and each case represents an observation of members’ behaviors. For example, in this dissertation we will define a nominal variable for each healthcare (billing) provider and a case for each patient for whom the provider might have billed no claims, one claim, or multiple claims for them. By analyzing data in this form, RA derives a ‘best model,’ or a calculated probability distribution that most accurately represents members’ observed behavior patterns while using minimal degrees of freedom. Associations in an RA model indicate the most salient relationships among members and provide a type of a network structure. But unlike standard network analysis methods, which are limited to linear, dyadic relationships, RA associations can be nonlinear and multi-way, identifying more complex relationships among members when present.

For the purposes of network comparison, RA can be applied separately to datasets from two time periods, and RA best models can be compared to measure a network

---

3 A best model is identified with one of several selection criteria, such as AIC or BIC.
distance. Best models identified by RA include a set of calculated probabilities, so two models can be compared by measuring the differences in their calculated probabilities. These differences can be aggregated into a distance measure that summarizes the extent to which the network changed following an event or intervention.

Like with R-DNA, the I-DNA distance measure is technically only descriptive. This distance might be trivially different from zero (indicating no change) and might have arisen from sampling error or from chance variations in network members’ behavior. So, to test this distance for statistical significance, I-DNA employs one of the same resampling methods that were used in R-DNA: permutations, one-sample bootstrapping, or two-sample bootstrapping. And, just like in R-DNA, each resampling method enables us to estimate the likelihood that the distance could have resulted from chance alone, even if the null hypothesis were true and the two networks were not systematically different.

**Coordinated Care Organization (CCO) Insurance Claim Data**

This dissertation applies I-DNA and R-DNA to the analysis of a healthcare billing provider network in a Coordinated Care Organization (CCO) in the state of Oregon. *Health Share of Oregon* ([http://healthshareoregon.org/](http://healthshareoregon.org/)), is one of sixteen CCOs created to fulfill a state legislated transformation in Oregon’s Medicaid health care delivery system (Joint Special Committee on Health Care Transformation, 2011; Senate Bill 1580, 2012). CCOs are subject to a fixed global budget, which motivates care delivery in a cost-efficient way: CCOs retain profits if they meet patients’ health needs under budget and suffer losses if they do not. Health Share of Oregon comprises eleven affiliate
healthcare organizations that must collectively provide physical, dental, mental and behavioral health services to all Medicaid members in the Tri-County area. Like other CCOs approved by the Oregon Health Authority (OHA, n.d.), Health Share’s goal is to facilitate care coordination among its providers and achieve what is known as the triple aim: improved health outcomes, improved experience of care, and reduced cost (Berwick, Nolan, & Whittington, 2008).

In order to fulfill state requirements, Health Share has encouraged its clinics to get certified as Patient-Centered Primary Care Homes (PCPCHs) and has prioritized PCPCH clinics when enrolling new patients (OHA, 2012). To fulfill one component of the PCPCH model, patients were assigned a primary care provider (PCP) and encouraged to access care primarily through his or her PCP. Medical home models have been shown to improve health outcomes and patient satisfaction (Rosenthal, 2008) and may also reduce costs by preventing duplicated services (Davis, Schoenbaum, & Audet, 2005) and hospitalizations (Bazemore et al., 2018). Under this model, patients can be expected to receive care more predominantly from their PCPs and specialists whom the PCP refers them to.

We expected that Health Share’s formation as a CCO, and particularly its assignment of patients to PCPs, would result in statistically detectable changes in the network of healthcare delivery. To test this, we used Medicaid claims data from two one-year time periods before and after Health Share’s formation on September 1, 2012. The ‘before period’ contained data from May 2011 through April 2012, and the ‘after period’
contained data from January through December of 2013.\textsuperscript{4} Both I-DNA and R-DNA were used to infer the healthcare delivery network during both time periods, to measure the extent to which the network changed, and to test that change for statistical significance.

**Statistical Analysis**

To apply R-DNA, before and after datasets were defined such that a continuous variable (i.e., column) represented each healthcare billing provider and a case (i.e., row) represented each patient who accessed care from at least one of Health Share’s providers during the given time period. Cases further indicated the frequency with which each billing provider billed claims for each patient, allowing regressions to infer the strength of connectivity between each pair of billing providers based on the amount of overlap in the sets of patients they billed for. Once network structures were inferred for each time period, a network distance was calculated, and then each of three resampling methods were used to test whether the distance was statistically significant.

To apply I-DNA, before and after datasets were defined in the same way as for R-DNA except that variables were nominalized to indicate whether billing providers billed claims for each patient zero times (0), one time (1), or multiple times (2). When applying I-DNA, before and after networks were inferred with information-theoretic approaches, a network distance was calculated, and then each of three resampling methods were used to test whether the distance was statistically significant.

\textsuperscript{4} Note we left a transitional “buffer” period starting four months before Health Share’s formation and continuing through four months after Health Share’s formation.
We defined three different billing provider networks for analysis with I-DNA and R-DNA. The first network was a ‘full network’, including all 1,298 billing providers who were active within the Health Share service region during both time periods, plus one node to represent all billing providers from outside the service region. The second network featured the top 30 billing providers by patient volume, who accounted for the lion’s share of claims (88.2%) and billed for the majority of patients (96.7%). And the third network grouped billing providers into six sectors of care, indicating the network of billing patterns among primary, ancillary, specialty, mental/behavioral, facility, and ‘other’ health care sectors.

Networks were inferred by three R-DNA inference methods (correlation, multiple regression, and partial least squares regression), and by two I-DNA methods (RA and a simplified version of RA called standardized bivariate transmission which is restricted to detect only pairwise connections). Following the application of each inference method, network structures from before and after Health Share’s formation were compared to calculate a general measure of network distance. This distance was then statistically tested by permutation, one-sample, and two-sample bootstrapping methods. Our joint application of both I-DNA and R-DNA allowed us to compare these methods on the

---

5 Billing providers were selected as the primary level of analysis rather than performing providers, because there was a more manageable number of them and because their behavior was expected to be more stable and less subject to random variation than the behavior of individual performing providers.

6 This classification was based on what kinds of performing providers they most often billed claims with, after sorting performing providers by their self-reported specialties in the National Uniform Claim Committee’s (NUCC) Taxonomy of Health Care Specialties.
networks they inferred, the types of changes they found, and whether or not they could statistically detect change over time.

Results

Statistical testing of the top 30 network and the care sector network indicated significant change after Health Share’s formation, and this finding held true regardless of which inference method or resampling technique was used for statistical testing. We did not test the full network of billing providers for significant change over time, as it was too large to even infer by the more computationally intensive inference methods (especially RA). The full network was only inferred by correlation – the simplest version of R-DNA – and standardized bivariate transmission – the simplest version of I-DNA – and it was not statistically tested for change by any method.

In addition to our statistical hypotheses that change (of some type) would occur, we had several substantive hypotheses regarding the nature of the changes we expected. Our hypothesis for the full network and top 30 network was that connectivity would increase over time. We thought that patients’ assignment to PCPs would lead them to move through the healthcare delivery system in increasingly predictable patterns\(^7\) and would produce stronger connections in the network after Health Share’s formation. Our results did not indicate support for this hypothesis. The full network by correlation and standardized bivariate transmission indicated weaker average levels of connectivity over time, and patterns in the top 30 network generally indicated that connectivity

\(^7\) We expected that connections would strengthen more often than weaken, that they would appear more often than disappear, and when applying RA that increased connectivity would result in connections becoming more complex (i.e., multiway) more often than less complex after Health Share’s formation.
consolidated, with its strongest connections even stronger after Health Share’s formation and with its weakest connections even weaker.

We also had hypotheses for two specific connections within the care sector network. First, we expected that PCP influence over patients’ utilization of specialty services would cause patients to ‘bounce back and forth’ less often between primary and specialty providers and would weaken the connection between those sectors. Second, we expected that Health Share’s incorporation of mental and behavioral services would lead to more referrals from the mental/behavioral to the primary sector and would strengthen the connection between those sectors. However, our results did not provide strong support for either hypothesis. The largest change in the relationship between primary and specialty sectors was an increase in the proportion of patients receiving neither type of care, and changes in the relationship between primary and mental/behavioral sectors indicated a decrease in the proportion of mental/behavioral patients who also received primary care. These findings indicate areas where additional research is warranted.

**Aims and Contributions of this Work**

Our primary aim for this dissertation is to contribute methodologically to the field of social network analysis by demonstrating and comparing R-DNA and I-DNA methods for their capacity to infer network structures and statistically test them for change over time. We are able to translate the R-DNA method from where it originated in bioinformatics to our target field of social network analysis. And we develop I-DNA as an analogous tool that can statistically detect network change but that also leverages the
benefits of RA and information theory to identify nonlinear and multiway patterns of network connectivity.

We find both R-DNA and I-DNA methods to show promise as social network analysis tools, and we argue that they contribute to the field on two specific fronts. First, we provide several examples of how one can infer network structures from data, rather than measuring network connections directly. This includes development of RA as a useful inference method for social network structures with complex interactions that cannot be captured well at the pairwise level. Second, our application of these methods provides a general framework for statistical detection of network-level change, which has previously been absent from this field. We demonstrate the versatility of this framework by applying it with several different inference methods and resampling techniques, and we make recommendations for method selection in future research studies.

Our secondary aim for this dissertation is to contribute substantively to the field of CCO health policy, by identifying changes in the healthcare delivery network that followed Health Share of Oregon’s formation as a CCO. Both I-DNA and R-DNA indicate that Health Share’s network of care delivery did change significantly over time. And although we did not find compelling support for any of our specific hypotheses regarding the nature of that change, the results generated by these methods still provide valuable insights into the types of changes that likely occurred, indicating several areas that warrant further investigation in future research.
We conclude that both I-DNA and R-DNA are useful approaches for inferring social network structures, for descriptively exploring the types of change that occur in them over time, and for testing whether those changes are statistically significant.
Chapter 2: Background

Several bodies of literature are relevant to this dissertation. Chapter 2 reviews current approaches to measuring, defining, and statistically testing social network structures, including the challenges that motivate the current work. Chapter 3 presents I-DNA and R-DNA, describing their information-theory-based and regression-based approaches to network inference, their measurements of network change, and their capacity for statistical testing. And Chapter 4 provides background and literature regarding Oregon’s legislative healthcare transformation and the Medicaid insurance claims data that are available before and after its implementation.

Network analysis (NA) is a general methodology for measuring, representing and analyzing structural connectivity among interacting elements in a system (Butts, 2008; Wasserman & Faust, 1994). In the social sciences, social network analysis (SNA) is widely implemented as a research tool for a variety of social phenomena, including relationship dynamics (Cartwright & Harary, 1956; Heider, 1946), the formation of groups (Breiger, 1974), accessibility of information and opportunities (Burt, 1987, 1992; Granovetter, 1973), preferences for connection with similar others (McPherson, Smith-Lovin, & Cook, 2001), and the spread of opinions (Katz & Lazarsfeld, 1955), behaviors (Rogers & Kincaid, 1981), and social change (Hamblin, Jacobsen, & Miller, 1973; Myers, 1997). Through these applications, SNA has contributed to a number of social science disciplines, including psychology (Butts, 2008), political science (Ward, Stovel, & Sacks, 2011), economics (Bögenhold, 2013), sociology (Palloni, 1998), and program evaluation (Durland & Fredericks, 2006).
Network Notation

The methods presented in this dissertation view network members as variables, borrowing basic notation from Gill and colleagues’ work with R-DNA for genetic research (Gill et al., 2010). Let a graph $\mathcal{G}$ contain the set of $g$ network members, which we will denote as

$$
\mathcal{G} = \{x_1, x_2, x_3, \ldots x_g\}.
$$

In this set, any two members might be referred to as $x_i$ and $x_j$, and a relationship between them is represented as $s_{ij}$. When focusing on a specific member of the network (say $x_i$), that member is referred to as an ‘ego’ and all other members are referred to as potential alters ($a$) that might have ties to the ego. In a complete network, where everyone is connected to everyone else, the set of alters will include all members except the ego.

$$
a = \{x_1, x_2, \ldots x_{i-1}, x_{i+1}, \ldots x_j\}
$$

For most social network analysis metrics, a network’s structure is captured in the form of a matrix. A ‘sociomatrix,’ as introduced by Moreno (1934) and Jennings (Moreno & Jennings, 1938), is an arrangement of structural measurements that indicate the relationships between each pair of network members (Wasserman & Faust, 1994). These measurements are arranged in a $g$ by $g$ matrix (illustrated below in Table 2.1), where the rows and columns list each network member in identical order. With this setup, diagonal cells, where $x_i = x_j$, indicate each member’s ‘relationship’ with his or herself.
(usually null or denoted with a 1), and the off-diagonal cells indicate relationships between each ego and all of his or her potential alters.

Relationships in a sociomatrix might be either dichotomous or valued, and might be either directed or non-directed (Wasserman & Faust, 1994). For a non-directed network, where \( s_{ij} = s_{ji} \), the matrix will be symmetric, with the same arrangement of relationships above and below the matrix diagonal. However, in the case of directed relationships, where \( s_{ij} \neq s_{ji} \), the matrix will be asymmetric. Here, the row member is understood to be the ‘sender’ of a relationship \((x_i)\), and the column member the ‘receiver’ \((x_j)\). The hypothetical example in Table 2.1 shows a weighted and directed network of social outings, where member \( x_1 \) spends 20% of his outings with member \( x_3 \) but member \( x_3 \) spends only 10% of his outings with member \( x_1 \).

<table>
<thead>
<tr>
<th></th>
<th>( x_1 )</th>
<th>( x_2 )</th>
<th>( x_3 )</th>
<th>( \ldots )</th>
<th>( x_g )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x_1 )</td>
<td>-</td>
<td>0</td>
<td>.2</td>
<td>( \ldots )</td>
<td>0</td>
</tr>
<tr>
<td>( x_2 )</td>
<td>0</td>
<td>-</td>
<td>.1</td>
<td>( \ldots )</td>
<td>.5</td>
</tr>
<tr>
<td>( x_3 )</td>
<td>.1</td>
<td>.3</td>
<td>-</td>
<td>( \ldots )</td>
<td>.1</td>
</tr>
<tr>
<td></td>
<td>( \vdots )</td>
<td>( \vdots )</td>
<td>( \vdots )</td>
<td>( \vdots )</td>
<td>( \vdots )</td>
</tr>
<tr>
<td>( x_g )</td>
<td>0</td>
<td>.2</td>
<td>.4</td>
<td>( \ldots )</td>
<td>-</td>
</tr>
</tbody>
</table>

The sociomatrix approach to data representation complements other methods well because it emphasizes relationships among variables, rather than attributes of the variables in isolation (Knoke & Yang, 2008). Sociomatrices are still limited to pairwise, or dyadic, relationships among members, but data in a matrix format can be subjected to
a number of useful algebraic operations, so nearly all descriptive network metrics are calculated with data in this form (Wasserman & Faust, 1994). Specific to this dissertation, sociomatrices are used for the R-DNA method and for comparative analyses between I-DNA and R-DNA.

**Measurement of Social Network Structure**

In the social sciences, network links have most often been measured directly, either through self-reports about relationships or observations of social interactions. This has posed a considerable obstacle for the field, because traditional methods are quite labor intensive. Questionnaires are very time consuming for participants, especially when they include a roster of all those whom a person might or might not be related to (Krackhardt & Stern, 1988). Similarly, interviews (Burt, 1984, 1985) and observations of social interaction (Bernard, Killworth, & Sailer, 1982; Galaskiewicz, 1985) are very time consuming for researchers (Carrington et al., 2005). These obstacles are compounded if a network is followed over time. Archival records have been the easiest way to acquire network data (e.g., Burt & Lin, 1977), but historically these have only been available for very few types of relations, such as financial transactions between organizations or instances of author co-citation (Carrington et al., 2005).

In addition to feasibility limitations, traditional methods of measuring networks have been challenged along the lines of accuracy, reliability, and validity (Wasserman & Faust, 1994). Individuals have documented limitations in their ability to recall social interactions (Bernard et al., 1982), and there are questions of reliability (Burt, Marsden, & Rossi, 1985; Mouton, Blake, & Fruchter, 1955a), construct validity (Burt et al., 1985;
Mouton, Blake, & Fruchter, 1955b), and measurement error (Holland & Leinhardt, 1973) for social networks just like for other types of data collection in the social sciences (Knoke & Kuklinski, 1982; Wasserman & Faust, 1994). For example, when a network is observed cross-sectionally, the stability of measured relationships over time is often unknown.

Measurement issues are especially problematic in network analysis, because many network metrics are highly sensitive to missing data (Knoke & Kuklinski, 1982). The missed observation of one highly connected member could hugely bias a large number of the network parameters, such as density, or average shortest path length (described below). For this reason, the boundaries of the network should be clear and non-arbitrary, as defined by either the researcher or the network members themselves (Knoke & Kuklinski, 1982; Laumann, Marsden, & Prensky, 1989; Wasserman & Faust, 1994). And when observations are missing – either because some members of a network are unmeasurable or because the population is too large to measure all members – special precautions should be taken to ensure that measurements are sufficient to estimate the structure of the network (Choudhury et al., 2010; Kurant et al., 2011; Smith & Moody, 2013).

**Statistical Inference of Social Network Structure**

To address the problems associated with missing data, multiple techniques have been developed for *topological inference*, the statistical inference of network structure (see Kolaczyk, 2009 for an overview). Data might be missing on relations, members, or both, and approaches have been developed for each situation. In the most extreme case,
where entire areas of a network are unknown, a set of measured areas are analyzed to infer the structure of the remaining network that connects them. This general approach is referred to as tomographic inference, for its similarity to the way two-dimensional x-ray and MRI images are used to define three-dimensional structures. Multiple methods have been used for this, including hierarchical clustering (Castro, Coates, & Nowak, 2004) and maximum likelihood methods (Coates et al., 2002). But tomographic inference has yet to be applied to the inference of social networks, and has been applied primarily in phylogenetics (biology) and computer science.

Perhaps the easiest case of missing data is when all members are known and when only some of the relations are missing. In this situation, link prediction is possible by analyzing the set of known relations and member attributes to determine what attributes are predictive of a relation between two members. These attributes might stem from measurements of the members themselves (e.g., place of work), but they might also stem from member-level network measurements, such as the number of mutual connections that a pair of members has. Once predictive attributes have been identified by the known relations, they can be used to predict which missing relations that are likely to exist. Logistic regression (Kolaczyk, 2009) and latent factor modeling (Hoff, 2009) have both been used to find missing links in social networks.

The third type of network inference, the inference of association networks (Kolaczyk, 2009), is most relevant to the current work. Here, a population of network members is known, but all relations are inferred from patterns in their attributes or behavior. This might be due to missing data for a large proportion of relations, there may
be limited confidence in the validity of relation measurements, or there may be interest in a type of relation that cannot feasibly be measured for a large network. In these situations, measures of members’ attributes or behaviors can be used to infer associations among them. Correlations have been used to infer network relations (Fuller et al., 2007; Weckwerth et al., 2004), as have partial correlations (Yates & Mukhopadhyay, 2013), partial least squares (Gill et al., 2010), mutual information (Fuite et al., 2008; Shalizi, Camperi, & Klinkner, 2006), and a variety of other approaches (see Oates & Mukherjee, 2012). These inference approaches are not yet used in social network analysis, but are common in bioinformatics where gene networks are inferred from patterns in gene expression.

There are several advantages to the statistical inference of network relations. The most obvious advantage is a practical one, because inference enables research on very large networks where it would not be feasible to directly measure all relationships. This is becoming especially true with recent advances in technology that allow for automatic collection of many individual behaviors via Bluetooth (Eagle & Pentland, 2006), and many other avenues for online behavioral expression that are easy to measure (Fienberg, 2012), such as Twitter (Krishnamurthy, Gill, & Arlitt, 2008) and Facebook posts (Clouston, Verderry, Amin, & Gauthier, 2009; Lewis, Kaufman, Gonzalez, Wimmer, & Christakis, 2008). Some online behaviors, such as posting content or downloading an app, are technically individual behaviors, but may be reflective of underlying social relationships.
A second advantage to network inference is that it can provide alternate definitions for relations between members. For example, individuals might have a great many ‘friends’ on Facebook, but patterns in their online posting behavior might indicate their behavioral association with only a small fraction of them. In this Facebook example, inference narrows the definition of a relation, but inferences can also broaden it. For example, when looking at two members’ attendance at various social events (e.g., Davis, Gardner, & Gardner, 1941) a correlation would reflect not only the events that both attended, but also events that only one member attended and those that neither member attended. With this larger set of information, inference could also identify the inversely associated individuals who attended opposite social events. With one of the methods used here (I-DNA), it is even possible for inference to capture three-way and higher-way associations (i.e., ‘hypergraphs’), expanding network analysis beyond its usual limitation to the study of graphs with only pairwise relationships.

The inference of relations does not avoid challenges of reliability or validity, as measurements of behavior or attributes might still be subject to measurement or sampling error, and the generality of an association may be so broad or complex that it does not map easily onto any theoretical understandings of a social relationship. However, inferential network mapping provides a third advantage in that inference allows specific relations to be tested for statistical significance (e.g., Steinhauser, Krall, Müssig, Büssis, & Usadel, 2008). For large networks with many relation inferences, extra precautions are necessary to protect against the probability of Type I errors (Dudoit, Shaffer, & Boldrick,
But statistical testing still provides some basic information regarding the certainty with which relations are known to exist.

Finally, network inference is advantageous in that it can sometimes allow for subsequent statistical comparison of network structures. This is the essence of both of the methods used in this dissertation. When the members of two networks are comparable, such as when the same network is inferred during two time periods, observations can be resampled to determine whether changes in the inferred networks are likely to be due to chance. In other words, the same data that allows for network inference can also be resampled to statistically test for differences in network structure. This will be described in Chapter 3.

**Descriptive Analysis of Social Networks**

Once a network’s structure is measured or inferred, it can be descriptively analyzed for a number of different characteristics. In the social sciences, the vast majority of network analyses are descriptive in nature (see Wasserman & Faust, 1994), and even when one or more aspects of a network are subjected to statistical testing, it is common to use visualizations and descriptive measures of network structure to provide general insights regarding the social network and phenomena of interest. Several common metrics are reviewed here and are applied in Chapter 5 to the health care networks that were inferred with I-DNA and R-DNA.

Many descriptive metrics are at either the member level or the network level. At the member level, the simplest metric is *degree* – the number of alters that a member is
connected to. For directed graphs, degree can be further specified as *in-degree, out-degree, or total degree*, referring to the number of directed relations that target a member, stem from a member, or do either. When standardized against the maximum degree possible, these degree metrics are one measure of a member’s *centrality* in a network. But many measurements of centrality exist, and a common alternative to the degree centrality measure is *betweenness centrality*, or the proportion of shortest paths (i.e., chains of relationships) that go through this member. Member-level metrics are often used to identify key members in the network, and they can also be re-interpreted as attributes of the individual members.

Many member-level metrics have corresponding metrics at the network-level. *Density*, the fraction of member pairs who are connected, can be thought of as a network-level aggregate of degree. Centrality measures also have network-level analogies in *centralization*, which indicates the heterogeneity of centrality scores among members. However, network-level metrics can illustrate properties of a network that are more meaningful than would be expected from an aggregate of member-level metrics. High centralization, for example, can indicate that the network has a core that is more densely connected than its periphery. A large *average shortest path length* can indicate that there are areas of the network that are relatively isolated from one another. And high *transitivity*, where members with a mutual connection tend to also be connected with each other, can indicate the presence of dense clusters throughout the network.

Both the member-level and network-level metrics above can be used to descriptively compare networks and to measure changes in a network over time. For
example, a member might cease to have the largest in-degree, or a network’s density might increase or decrease after an event or intervention. In addition to standard descriptive analyses, however, several metrics for set comparison have been adapted for the comparison of networks (see Snijders et al., 1990). The first of these is the Jaccard Index (Jaccard, 1901), which is a ratio of relations observed in both sets to relations observed in at least one set. When looking at a social network over two time points, the Jaccard index would measure network stability as the fraction of relationships from either time period that were present during both.

A variant of the Jaccard Index is ‘simple matching’ (Sokal & Michener, 1902). The simple matching ratio recognizes not only the relations observed in both sets but also the relations not observed in either set. In network analysis, simple matching would measure network stability as the fraction of possible relationships that were consistently present or absent across both time periods. Both Jaccard and simple matching produce similarity measures between zero and one, and can be subtracted from one to indicate the extent of difference between two networks. A variant of simple matching for weighted relations, called mean absolute difference (Gill et al., 2010), will be used when conducting R-DNA. However, these comparative measures must be coupled with other methods to allow for the statistical testing of difference.

**Statistical Analysis of Social Networks**

Nearly all statistical analysis of social networks revolves around parameter estimation. Parameters might be about network members, the formation or dissolution of connections, or the network as a whole – each of which is described further below. In
these applications, statistics is used to estimate the capacity for member-level network attributes (e.g., degree, centrality) to predict non-network attributes, such as attitudes (Burt, 1987), behaviors (Valente, 1996) or outcomes (Christakis & Fowler, 2007; Comola & Prina, 2013). Although less common, the inverse can also be investigated – members’ behaviors or non-network attributes can be used to predict network attributes (Comola & Prina, 2013; Lee, Scherngell, & Barber, 2009).

In parameter estimation for network members, network and non-network attributes are typically viewed as variables, and the network members are viewed as observations, or cases. For example, one might want to determine whether an individual’s location in a network is a good predictor of a subsequent outcome for them, such as having their book reviewed (Nooy, 2011) or emerging as a leader (Emery, 2012). However, by being members of the same network these observations are typically recognized as non-independent, violating an assumption made by virtually all traditional statistical tests (Kolaczyk, 2009; Robins & Pattison, 2005). So, in conducting member-level statistical analyses, efforts must be made to compensate or control for this dependency. Various approaches can test for and define dependency in network data (Robins & Pattison, 2005), and several approaches have been developed to control for it, including the quadratic assignment procedure (Krackhardt, 1987), Markov Chain Monte Carlo methods (Leenders, 1995), and actor-oriented modeling (Snijders, 2005).

Parameter estimation is also commonly used to investigate research questions about the formation or dissolution of social network ties. Some common approaches to this include stochastic actor-oriented models, most commonly with SIENA software
(Snijders, van de Bunt, & Steglich, 2010), exponential random graph models (ERGMs) (Kolaczyk, 2009; Pauksztat, Steglich, & Wittek, 2011), and Markov processes (Stadtfeld & Geyer-Schulz, 2011). These approaches allow one to estimate the extent to which different dynamic forces, such as social selection (homophily) and social influence (contagion; Snijders et al., 2010), influence the likelihood of network members forming a connection (Lazega, Mounier, Snijders, & Tubaro, 2012), maintaining connections (Schaefer, Light, Fabes, Hanish, & Martin, 2010), or engaging with those to whom they are connected (Pauksztat et al., 2011).

In addition to the analysis of members and connections, statistics can facilitate estimation of network-level parameters (see Kolaczyk, 2009). Here, basic parameters of an observed network (e.g., centralization, average degree) are tested for their probability of chance occurrence by comparing them to a large number of randomly generated networks (Kolaczyk, 2009, p.163). These tests can help to identify any non-random characteristics that appear to be present in a network structure (e.g., Apicella, Christakis, Fowler, & Marlowe, 2012; Moreno & Jennings, 1938). Alternatively, if a randomly connected network does not seem a fitting comparison, networks can also be analyzed for the presence or absence of power law degree distributions (Clauset, Shalizi, & Newman, 2009a) which characterize their scale-free properties (Kolaczyk, 2009, p.81-85).

Unfortunately, networks can only be compared to a few network types (e.g., random, scale-free), which limits the extent to which their parameters can be estimated. A researcher might be confident that a network is not random, and not scale-free, and yet all characterizing information beyond that is only descriptive. This limitation is a motivation
for research with exponential random graph models (ERGMs) (Kolaczyk, 2009). In the same way that a Gaussian distribution can be captured by the parameters of its mean and variance, ERGMs aim to capture the characteristic parameters of various other types of networks. The ERGM approach facilitates more types of parameter estimation, as well as the adaptation of other traditional statistical tools for analysis of network data (Fienberg, 2012).

Lastly, it should be noted that there are statistical techniques for estimating the degree of similarity between two networks. Both the Jaccard index and simple matching can be tested for statistical significance (Snijders et al., 1990), and the quadratic assignment procedure has also been shown to identify when network structures are ‘correlated’ (Krackhardt, 1987). These approaches are similar to the methods used in the current work in that they allow the researcher to compare two networks with the same set of members but with different sets of relations. However, estimation of similarity between two networks is different from testing the null hypothesis that they are the same. In similarity estimation, one might either conclude that two networks are significantly similar or refrain from conclusion. By contrast, in statistical testing of network equivalence, a researcher allows for the possibility of rejecting the null hypothesis and concluding that networks are significantly different. Hypothesis testing of differences in network structure appears absent from the field of social network analysis, but R-DNA performs this function in bioinformatics, and I-DNA is developed in this dissertation for the same purpose.
Methodological Contributions

By applying I-DNA and R-DNA to Medicaid insurance claims data, this dissertation contributes methodologically to the field of social network analysis on two fronts. First, statistical inference methods have not yet been used much to identify social network relations, so examples of these types of applications are expected to be valuable for the field. Advances in technology have made socially-relevant data much easier to collect, and they allow people to exhibit more and more behaviors online (Fienberg, 2012). In the interest of leveraging these new sources of data, this dissertation demonstrates the application of a regression-based method for network inference from the field of bioinformatics and develops an alternative approach to network inference based on existing methods in information-theory.

Second, this dissertation contributes to the growing set of statistical approaches for analyzing social networks. Many traditional statistical techniques have been difficult to adapt for network data, because they assume independence of observations (Kolaczyk, 2009) and often have distributional assumptions as well. Social network members are by definition non-independent, and many social network attributes are not normally distributed (Clauset et al., 2009a; Zhao & Ye, 2013). However, both I-DNA and R-DNA define network members as variables, and in doing so do not assume their independence. I-DNA is also non-parametric, which means it does not have distributional assumptions for network members’ behaviors when defined as variables. Both I-DNA and R-DNA will be shown in the following chapters to afford statistical testing of change in network structure, and to be useful for studying how a social network changes over time.
Chapter 3: Methods

The contributions of this work will stem from the application of two statistical methods to analyze network change over time. Information-theory-based Differential Network Analysis (I-DNA) and Regression-based Differential Network Analysis (R-DNA) both provide the capacity to (1) infer a network structure, (2) measure a difference between two networks, and (3) test that difference for statistical significance. These basic steps are summarized below in Figure 3.1 and in Table 3.3 on p.42. Both methods are expected to be useful for the field of social network analysis by enabling one to test whether a network’s structure has changed significantly between two time periods.

![Figure 3.1. Overview of the analytical procedures for R-DNA and I-DNA](image)

Central to the utility of both methods is their capacity for hypothesis testing (Step 3), which allows a researcher to establish whether network structures from two time periods are significantly different from one another. Our null hypothesis,
\( \mathcal{H}_0 : \text{Network}_{\text{before}} = \text{Network}_{\text{after}} \), \hspace{1cm} (1)

is that the two networks appear to belong to the same distribution of networks and any differences between them are likely due to chance alone. Our alternate hypothesis,

\( \mathcal{H}_1 : \text{Network}_{\text{before}} \neq \text{Network}_{\text{after}} \), \hspace{1cm} (2)

is that the networks belong to two different distributions of networks, and that differences between them are unlikely to be due to chance alone.

To test the hypothesis that two networks are different, the observed difference between the networks must first be boiled down to a single value, or test statistic. The classic statistical test of difference, Student’s \( t \) test, uses the difference in sample means as the test statistic. However, null hypothesis testing is not restricted to a comparison of means and can easily be extended to a comparison of models, or really any data representation, just so long as the difference between samples can be represented as a single value, \( \hat{\theta} \), that would be zero if the two samples (or networks) were equal. The absolute distance measure from I-DNA and the mean absolute difference measure from R-DNA both satisfy this criterion, allowing us to define the null hypothesis,

\( \mathcal{H}_0 : \hat{\theta} = 0 \) \hspace{1cm} (3)

It is always possible that a difference between \( \hat{\theta} \) and 0 could stem from chance variations in sampling, even if there’s no true difference between the networks. In that case, rejecting the null hypothesis would be incorrect and a Type I error. So, the value of \( \hat{\theta} \) must be compared to 0 along a reference distribution that summarizes the variation in \( \hat{\theta} \) that one would expect by chance alone. This allows us to estimate the likelihood that \( \hat{\theta} \)’s deviance from 0 could be accounted for by standard error, even if there were no
difference between the two samples (or networks). In the social sciences, the probability of a Type I error \( p \) must usually be less than or equal to 5\% for a finding to be declared statistically significant.

This dissertation demonstrates three resampling methods that can be used to estimate the probability of Type I error. These include permutation, one-sample bootstrapping, and two-sample bootstrapping. Once the difference between networks has been summarized with a statistic, such as a network distance \( \hat{\theta} \), any of these resampling methods might be used to statistically test that statistic’s deviance from 0.

**Permutation Testing**

Permutation testing is a statistical method that involves randomly shuffling observations between two datasets. This can also be thought of as resampling the data in full without replacement. The idea of using permutations was first developed by Fisher (1936) as a way to illustrate the accuracy of Student’s \( t \) test even when distributions were not strictly Gaussian. Shortly thereafter, Pitman (1937) suggested that the method could be generalized for comparing two populations, although at that time the computational demands were daunting. The gist of permutation testing has been described by Efron and Tibshirani (1993) as addressing the **two sample problem**: Two samples are suspected to be from different populations, and must be tested to determine whether they are indeed unlikely to be from the same population.

Statistical testing by permutation rests on the creation of a permutation-based reference distribution, and this distribution is created by generating a large number of
permutations in the data. For each permutation, all observations from both datasets (i.e., before and after) are pooled into a superset of cases, and then randomly sorted into a (permuted) pair of datasets. When observations contain multiple values (e.g., height and weight) the entire case is kept intact – randomization occurs at the case-level and not at the level of an individual datum. The number of cases in a permuted pair of datasets must also match the number of cases in the original pair. That is, if the original samples contained \( n \) and \( m \) cases, respectively, each permutation must also produce one sample with \( n \) cases and one sample with \( m \) cases.

When doing permutation testing, the distance statistic \( \hat{\theta} \) should be calculated with each pair of permuted datasets in the same way that it was calculated on the original datasets. The values of \( \hat{\theta} \) from permuted datasets are here denoted \( \hat{\theta}_{\text{perm}} \), and these are accumulated into a reference distribution \( N_{\text{perm}}(\theta) \). The reference distribution \( N_{\text{perm}}(\theta) \) illustrates the variance that would be expected in \( \hat{\theta} \) even if there were no systematic differences between the samples. In essence, it shows you how much variation you could expect in your distance statistic (\( \hat{\theta} \)) even when the null hypothesis is true. One can then estimate the probability of a Type I error by calculating the proportion of \( N_{\text{perm}}(\theta) \) values that are as far or further from zero than the observed \( \hat{\theta} \). For example, if only 1 in 100 pairs of permuted datasets produced a \( \hat{\theta}_{\text{perm}} \) as extreme as our observed \( \hat{\theta} \), we can say our \( \hat{\theta} \) has a \( p \) value equal to .01. If the \( p \)-value is less than or equal to .05, the difference between the inferred networks can be declared statistically significant.
One-Sample Bootstrap Testing

A simple adjustment can be made to the permutation approach to perform one-sample bootstrapping. As in permutation testing, cases from each dataset (observations from before and after) are still combined into a superset, but in one-sample bootstrapping, the cases are then randomly selected *with replacement* into two resampled datasets. This means that some cases might appear multiple times in either or both bootstrapped datasets, while other cases may not appear in either dataset. As in permutation testing, the number of cases in resampled datasets matches the number of cases in the original pair.

Statistical testing with one-sample bootstrapping is very similar to permutation testing because it relies on the creation of a resample-based reference distribution. For every pair of resampled datasets, the distance statistic $\hat{\theta}$ gets calculated in the same way that it was calculated on the original datasets. The values of $\hat{\theta}$ from are accumulated into a reference distribution ($N_{one-sample}(\theta)$), and one estimates the probability of a Type I error by calculating the proportion of $N_{one-sample}(\theta)$ values that are as far or further from zero than the observed $\hat{\theta}$. Ultimately, the one-sample bootstrapping technique is nearly identical to permutation testing, except that resampling is done with replacement.

Two-Sample Bootstrap Testing

Two-sample bootstrapping is a third approach to resampling that is different in spirit from the earlier two but still allows us to estimate the probability of a Type I error. With two-sample bootstrapping, cases from each dataset are not combined together into a superset of cases. Instead, each dataset is resampled *separately* with replacement to create
bootstrapped pairs of datasets. This involves the resampling of a pseudo ‘before’ dataset consisting only of observations from before, and a pseudo ‘after’ dataset consisting only of observations from after.

As in the other resampling methods above, two-sample bootstrapping requires that the distance statistic $\hat{\theta}$ be calculated on each pair of resampled datasets in the same way that it was calculated on the original datasets. These resampled distances are accumulated into a reference distribution ($\mathcal{N}_{\text{two-sample}}(\theta)$) and used to assist us calculating the probability of a Type I error. However, unlike the permutation and one-sample bootstrap techniques, this reference distribution does not directly tell us the $p$ value and some additional steps are needed to calculate it.

In order to establish whether our observed distance ($\hat{\theta}$) is statistically significant, we follow the standard protocol of a dependent-samples $t$ test, with three key differences. First, the $t$ value that is traditionally calculated in a paired-samples $t$ test uses an average difference between paired cases ($\overline{Diff}$) for the numerator, while in bootstrapping this difference is substituted by the measured distance between two networks ($\overline{Dist}$). This is a reasonable substitution because both metrics are measures of difference between datasets, and in both cases we want to establish that these differences are significantly different from zero.

Second, the $t$ value in a traditional paired-samples $t$ test uses an estimate of standard error that is taken from the standard deviation of differences between paired cases. Instead of this, bootstrapping substitutes the standard deviation of distances found
between bootstrapped samples. This is also a reasonable substitution because the
bootstrapped reference distribution effectively indicates how much the models might be
expected to differ if there were random error in the sampling of cases. Differences
between the two t value equations are illustrated in Table 3.1 below.

Table 3.1. Comparison of Traditional versus Bootstrapped t Value Calculation

<table>
<thead>
<tr>
<th></th>
<th>Traditional t Value</th>
<th>Bootstrapped t Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>t = ( \frac{\text{Diff}}{SD_{\text{Cases}}} )</td>
<td>t_{\text{Boot}} = ( \frac{\text{Dist}}{SD_{\text{Boot}}} )</td>
<td></td>
</tr>
</tbody>
</table>

The third difference between the traditional t test and the two-sample
bootstrapping approach is in the evaluation of the t value. With a paired-samples t test,
the t value is evaluated according to its degrees of freedom, which are calculated as one
less than the (paired) sample size (n). In our case, degrees of freedom are calculated as
one less than the number of (pairs of) bootstrapped samples created (b). This is a
reasonable substitution because the purpose of considering the degrees of freedom is to
qualify the estimate of standard error. In traditional applications of the t test, standard
error is based on the standard deviation of cases (n), but in our application it is based on
the standard deviation of bootstrapped samples (b). (For the present study there were 100
resamples by each method, so we have 99 degrees of freedom.)

Data Requirements

The data requirements for permutation and one-sample bootstrapping are
relatively simple. First of all, one needs to have observations of multiple samples. In the
analysis of network change over time, one needs observations of the network before and
after the event or intervention of interest. These data must be arranged so that a variable represents each network member and so a case represents each observation. For example, in the CCO data, billing providers (BPs) will be represented as variables, and patients will be represented as cases, indicating the number of times each patient was billed for by each provider. The CCO data format is illustrated below with fictitious data in Table 3.2.

**Table 3.2. Arrangement of CCO data for statistical testing**

<table>
<thead>
<tr>
<th></th>
<th>BP1</th>
<th>BP2</th>
<th>BP3</th>
<th>BP4</th>
<th>BP5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Before</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient 1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Patient 2</td>
<td>1</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Patient 3</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Patient 4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Patient 5</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Patient 6</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td><strong>After</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient 4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Patient 5</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Patient 6</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Patient 7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Patient 8</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Patient 9</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Note that the fictitious cases in Table 3.2 are divided into observations before and after the CCO intervention took place. Some patients were observed during both time periods (patients 4, 5, 6), while others were observed during only one time period. But all claims billed for each patient during a given time period are summarized in a single case for that time period. This arrangement of data will be used to infer network structures with I-DNA and R-DNA, with the only difference being that for I-DNA the data will be discretized into nominal variables.
The number of cases required depends on the number of permutations that will be generated. It is important that no two permutations are identical, and the total number of possible permutations is equal to the following ratio (Efrom & Tibshirani, 1993, p.207)

\[
\frac{N!}{n! \, m!}
\]

where \( N \) is the number of observations across both samples and \( n \) and \( m \) are the number of observations in each sample. This amount of data is relatively easy to come by – one could generate 12,870 unique permutations from two datasets with just 8 cases each.

**Assumptions**

Permutation and one-sample bootstrapping tests are non-parametric and have few statistical assumptions. They do not assume that variables have Gaussian distributions or that relationships among variables are linear. These types of tests do assume that observations are independent and identically distributed (or \( iid \)), which means that observations should be selected randomly from the population of interest and with no systematic relation to any of the variables. The validity of this \( iid \) assumption is central to making inferences, allowing one to approximate a population parameter from a sample statistic.

In applying permutation or one-sample bootstrapping tests to network data, network members (i.e., billing providers) should not be treated as \( iid \) observations (i.e., cases), because as members of the same network, they are by definition non-independent. However, network members can be represented as variables (which do not need to be \( iid \)), and patient-level observations of their behaviors can be represented as cases, so long
as it is reasonable to assume that observed attributes or behaviors are independent. Often the independence of observed behavior expressions is more plausible than the independence of network members – in the example of CCO data, patients are arguably more independent than medical providers. However, the validity of inferences can still be threatened if observations are systematically related to any of the variables. (In the example of the CCO data, it would be problematic to sample patients seen only by Dr.1, while ignoring those seen only by Dr. 2, Dr. 3, etc.) Cases should be sampled at random (or exhaustively), to allow for inference.

The accuracy of permutation and one-sample bootstrap testing increases with the number of resampling iterations completed. With these forms of testing, one uses a subset of all possible resamples to generate the reference distribution. There is some chance variation in which subset of permutations gets used, so in addition to the Type I error rate, there is also a ‘resampling error rate’ that can affect the hypothesis test. This error rate is higher for more stringent (i.e., lower) levels of $\alpha$, because resamples are more likely to produce values of $\mathcal{N}_{perm}(\theta)$ that are near the mean, leaving ambiguity about the distribution at extreme values. Assuming the desired Type I error rate ($\alpha$) is .05, Efron and Tibshirani (1993) suggest that 1901 permutations are necessary to reduce the permutation error to no more than $\pm$10%, effectively rendering the Type I error rate as somewhere between .045 and .055 (pp. 210-211). Although resampling testing with fewer iterations is not uncommon (e.g., Gill et al., 2010), especially when additional computations must be performed on every pair of permuted datasets.
Unlike permutation and one-sample bootstrapping, two-sample bootstrap testing relies on some additional statistical assumptions. Two-sample bootstrap testing relies on the \( t \) distribution, which assumes that the distance value is normally distributed. This may or may not be the case in a specific application, but we can get a good idea of whether this is distributed by looking at the distribution of two-sample bootstrapped differences. This should approximate normal and can be tested for whether it varies significantly from normal through the Shapiro-Wilk test of non-normality. When distributions are not reliably normal, permutations or one sample bootstrapping may be a better alternative.\(^8\)

The following sections describe the use of I-DNA and R-DNA for inferring networks at two time periods, measuring the distance between them, and testing that distance for statistical significance. A few examples of this methodological approach are first provided in Table 3.3, and stepwise instructions for applying the methods to Medicaid claims data will be described more in Chapters 4 and 5.

Note that with each method (I-DNA, R-DNA), hypothesis testing of network distances will be established through each of the three resampling methods that were described above. By testing the same hypotheses once with permutation, once with one-sample bootstrapping, and once with two-sample bootstrapping, we will be able to evaluate whether these different resampling techniques are differentially capable of statistically detecting network change. Chapter 6 will provide some comparisons between

---

\(^8\) Shapiro-Wilk tests retained (i.e., did not reject) normality of this distribution in all two-sample hypothesis tests in the current dissertation, see Appendix C on p.310.
these three resampling methods (which will be shown to perform virtually identically across all of applications in the current dissertation).
Table 3.3. Examples of Hypothesis Testing with I-DNA and R-DNA

<table>
<thead>
<tr>
<th>Step 1. Derive Network Structures</th>
<th>Step 2. Calculate Distance ($\hat{\theta}$)</th>
<th>Step 3. Test Null Hypothesis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>I-DNA</strong></td>
<td><strong>R-DNA</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Observed Data:</strong></td>
<td><strong>Observed Data:</strong></td>
<td></td>
</tr>
<tr>
<td>Nominal variable represents each network member</td>
<td>Continuous variable represents each network member</td>
<td></td>
</tr>
<tr>
<td>Cases represent observations</td>
<td>Cases represent observations</td>
<td></td>
</tr>
<tr>
<td>Cases are sorted into before and after datasets</td>
<td>Cases are sorted into Before and After datasets</td>
<td></td>
</tr>
<tr>
<td><strong>Resampled Data:</strong></td>
<td><strong>Resampled Data:</strong></td>
<td></td>
</tr>
<tr>
<td>Recreate pseudo before &amp; after datasets with permutation, single bootstrapping or double bootstrapping</td>
<td>Recreate pseudo before &amp; after datasets with permutation, single bootstrapping or double bootstrapping</td>
<td></td>
</tr>
</tbody>
</table>

**Step 1. Derive Network Structures**
- **I-DNA:** Perform RA* on Before and After datasets
- **R-DNA:** Perform PLS* on Before and After datasets
- Select best models by criterion (AIC, BIC, $\alpha_{inc}$)
- Associations indicate network structure

**Step 2. Calculate Distance ($\hat{\theta}$)**
- **I-DNA:** Calculate network distance by summing differences in calculated probabilities ($q$) from best models before (1) and after (2)
  $\hat{\theta}_{RA} = \sum |q^2 - q^1|$
- **R-DNA:** Calculate network distance as mean absolute difference (MDA), or the average amount of change per connection
  $\hat{\theta}_{PLS} = \frac{1}{g(g-1)} \sum_{i \neq j} |\hat{s}^1_{ij} - \hat{s}^2_{ij}|$

**Step 3. Test Null Hypothesis**
- **I-DNA:** Resampling generates 100 pairs of pseudo before and after datasets
- **R-DNA:** Resampling generates 100 pairs of pseudo before and after datasets
- Resampled distances are aggregated into a reference distribution $N(\theta)$
- Reference distribution is used to determine the probability a Type I error. If $< 5\%$, reject $H_0$.

*Alternatives to RA include SBT; Alternatives to PLS include correlation or multiple linear regression.
Method 1: Information Theory-based Differential Network analysis

I-DNA is one of two methods used in this dissertation to analyze network change. This approach has been developed from existing methods for a new application in social network inference and statistical testing. I-DNA uses reconstructability analysis (RA) to derive network structures as best models, uses a sum of differences in best models’ calculated probabilities to measure network distance, and uses resampling to test the significance of that distance. The I-DNA inference method, RA, is a data mining method that works well for data with many nominal variables, boiling complex behavior patterns down to a set of associations. When network members are represented as nominal variables, RA models indicate a type of network structure. And by dividing network observations into multiple datasets, such as those from before versus after an event or intervention, RA models can be compared to learn how the network changed over time.

Comparisons of RA models might be undertaken in a descriptive manner, if one simply observes associations appear, disappear, or change in structure between time periods. But when we use I-DNA for statistical testing, we must compare the RA best models from both time periods in a way that provides us with a measure of network distance ($\delta$). In this dissertation, I-DNA network distance is calculated by summing up the differences in calculated probabilities that underly the RA best models from before versus after the event or intervention. This distance measure is then tested for statistical significance through use of a resampling technique, like permutation or bootstrap resampling. If the probability of a Type I error is less than .05, the RA network can be said to have changed significantly over time.
Reconstructability Analysis

The basis for I-DNA is reconstructability analysis (RA; Klir, 1986; Zwick, 2004), which is a data mining methodology that detects deviations from mutual independence among nominal variables based on patterns in behavior. As a data mining methodology, RA can identify associations among a large number of variables by identifying prominent behavior patterns in a set of cases, or discrete observations of those variables. When variables are constructed to represent network members, such as individuals or organizations, RA associations can indicate relationships between them. Together, a collection of associations (called a model) provides a type of network structure, indicating who is associated with whom. If network members $x_1$ and $x_2$ are associated, the model would include association ‘$x_1x_2$’, separated from other associations with a colon, as in the model $x_1x_2 : x_3x_4$. However, unlike standard network analysis methods that are limited to linear and dyadic relationships between members, RA associations can be nonlinear, and might be dyadic, triadic, or even higher-way, depending on the complexity of the behavior patterns observed (Krippendorff, 1986).

Technically, an RA model is more than a specification of the associations among variables. The essence of an RA model is a set of calculated probabilities for every possible combination of variable behaviors, or states. In the CCO dataset, one combination would be the probability that a patient is billed for once by provider 1, multiple times by provider 2, and never by any other provider in the network. A full set of calculated probabilities, called a probability distribution is defined by the marginal probabilities of the associations included in the model. Put another way, every calculated
probability is a joint probability that takes the RA model associations (one-way, two-way, or higher) into account. Because an RA model can be boiled down to a calculated probability distribution, its accuracy, or goodness of fit, can be measured by comparing it with the probability distribution that was observed in the original dataset. The similarity of probability distributions that were observed (from the data) and calculated (from the model) indicates how well a model’s associations capture and replicate the underlying patterns in the data.

RA is like log-linear (LL) modeling in several ways. Both methods can derive multi-way associations among nominal variables, and both methods produce models in the form of calculated distributions. With both methods, the evaluation of a model’s goodness of fit usually entails a comparison of its calculated distribution against the observed distribution. The chief difference between RA and LL is that LL tends to be used for confirmatory analysis, to reject models that do not represent the data sufficiently well. By contrast, RA is commonly used for exploratory analysis (i.e., data mining) so RA methods include a set of techniques for searching through many different possible models (e.g., Zwick, 2001). RA also uses measures of entropy\(^9\) (uncertainty) which allow RA models to be described in terms of the amount of information captured (uncertainty reduced). This is an intuitive metric for nominal variables that is analogous to the amount of variance explained (Zwick, 2004).

\(^9\) Shannon entropy is a measure of uniformity in a probability distribution \((p)\), used for most applications of RA. Shannon entropy is defined as \(\sum -p \log_2 p\). See page 37.
Data Requirements

As a data mining methodology, RA does best with a very large number of cases. Specific guidelines can be compared to $\chi^2$ tests and LL, where the recommended number of cases is five times the ‘state space,’ or the Cartesian product of the cardinalities of each variable (Shannon & Zwick, 2004; Tabachnick & Fidel, 2006). These guidelines can be taken with a grain of salt, however, because technically they indicate the number of observations needed to demonstrate an all-way association among variables, and in network analysis it is unlikely that all network members would be associated with one another in a single relation of the highest order possible. So, while a network of 10 members contains a state space of over one thousand ($2^{10}$) even when they are defined as dichotomous variables, the actual data requirements are likely to be substantially less than 5,000 cases.\(^{10}\)

However, even moderated versions of the data requirements suggest that RA works best for analyzing networks that (a) are not too large, (b) have plenty of data available, and (c) include behaviors that can be reasonably captured in a small number of states. When member behaviors are potentially continuous (such as the frequency with which a provider bills for each patient), data must be discretized into a few nominal behavior states that can adequately capture the distribution of behavior for that member.

In addition to recommending a large number of cases overall, LL guidelines suggest that expected frequencies for all cells in the state space should be greater than 1,\(^{10}\)

---

\(^{10}\) It appears that a systematic analysis of RA’s sample size requirements (as a function of effect size) has not yet been conducted.
and at least 80% of cells should have an expected frequency greater than or equal to 5 (Tabachnick & Fidel, 2006). Analyzing data with expected frequencies lower than this has been observed to reduce statistical power in LL (Milligan, 1980), and would most likely have a similar effect with RA. When network members do not show much fluctuation in behavior (i.e., when variables consistently exhibit the same state), they can be excluded from the analysis or merged with a similar variable. For example, if several providers in a health clinic see very few patients, they will usually indicate ‘0’ claims and might do better if they were merged together into one variable for that clinic.

Similarly, when a nominal state is rarely observed throughout one or more variables, it can be recombined with other states. For example, four states of never, rare, medium, and frequent provider billing might be re-organized (rebinned) into three states: never, rare, and medium-to-high. It is permissible to combine states differently for different variables, because variables do not need to have the same states to be associated. However, exclusions and merges should be consistent across all datasets that will be directly compared. In this project, for the sake of simplicity, we bin all provider variables to indicate whether providers billed for each patient zero times (0), one time (1), or multiple times (2).

When comparing RA models, a final data requirement is that both datasets (from before versus after) need to include the same variables. This is for two reasons – to simplify our measurement of distance between networks, and to allow for resampling techniques like permutation testing which requires cases to be randomly shuffled between
When both datasets have the same variables, such as in the example dataset in Table 3.2 (p. 37), RA models can be compared to identify changes in network structure. An example is illustrated below in Figure 3.2, where data collected before an intervention suggests the model $x_1x_2x_3 : x_3x_4 : x_3x_5 : x_6x_7 : x_7x_8$, and data collected after the intervention suggests the model $x_1x_2x_3 : x_3x_4 : x_3x_5 : x_3x_6 : x_4x_5 : x_5x_7 : x_6x_7 : x_7x_8$. In this example, change in the network is obvious. The latter model, containing three more associations, would indicate that three relationships formed in the network following the intervention. Resampling techniques could be then used to test whether this change in network structure is statistically significant.

![Network Diagram](image)

**Figure 3.2. Comparison of hypothetical RA network structures**

**Assumptions**

To an extent, the assumptions of RA run parallel to the assumptions of permutation and one-sample bootstrap testing described above. Like those resampling

---

11 Cases from one dataset would systematically have ‘missing data’ or ‘nonsensical data’ if they were shuffled to a dataset that included different variables.
methods, RA is a ‘nonparametric’ technique, meaning it holds no a priori assumptions about population distributions (Krippendorff, 1986). The probability distribution observed in the original dataset might take any form, and so might the calculated probability that is offered by an RA model. This nonparametric feature makes RA applicable to the analysis of variables with non-Gaussian distributions, as is often the case with social network properties (Clauset, Shalizi, & Newman, 2009b; Zhao & Ye, 2013). By extension, the associations derived through RA do not need to be linear in nature but can also be nonlinear (Krippendorff, 1986).

Reconstructability analysis is not assumption-free, however. Also, like permutation testing, RA assumes observations are independent and identically distributed (iid), meaning cases should be selected randomly (or exhaustively) from a population of cases, and with no systematic relation to any of the variables. In the network of billing providers, it would not be acceptable to record each medical claim as its own case, because multiple claims by the same patient can hardly be assumed independent. Instead, a case should summarize all Medicaid claims involving a single patient. That way, we are only assuming independence between patients, which is more likely to be a reasonable assumption than independence between claims.12 The iid assumption is central to RA’s inference of a dataset’s underlying network structure, so observations should be defined as independently as possible.

12 We acknowledge that patients still likely have dependencies, such as biological relationships, geographic proximity, and similar socioeconomic factors pertaining to Medicaid eligibility, which we were not able to control for in the present study.
**I-DNA Step 1: Inferring Network Structure**

In using I-DNA for network analysis, an RA model is interpreted as a network structure. Recall that in its raw form an RA model is a distribution of calculated probabilities for every possible combination of variable states. So far, this distribution has been described as an indicator of the model’s accuracy, as it can be directly compared to the observed probability distribution from the original dataset. However, a model’s calculated distribution also contains information about the associations contained in the model. This information is made evident when comparing it to a ‘null’ distribution, or one where all combinations of states are equally likely (i.e., Shannon entropy is maximized). For example, in a network of three members \((x_1, x_2, x_3)\) who each have two behavior states (0 and 1, indicated by superscripts), there will be eight possible combinations of behavior \((x_1^0 x_2^0 x_3^0, x_1^0 x_2^0 x_3^1, x_1^0 x_2^1 x_3^0, x_1^0 x_2^1 x_3^1, x_1^1 x_2^0 x_3^0, x_1^1 x_2^0 x_3^1, x_1^1 x_2^1 x_3^0, x_1^1 x_2^1 x_3^1)\), or, more simply stated \((000, 001, 010, 011, 100, 101, 110, \text{ and } 111)\), and the null model would assume that each has a probability of 12.5%.

In virtually all datasets, the observed probability distribution will vary from this null distribution, meaning that some form of patterning exists in the data. An RA model, in trying to capture these patterns with a set of associations, will also vary from the null distribution, though usually not to the full extent of the observed distribution. This is because an RA model does not aim to replicate the data perfectly, but rather aims to capture the prominent associations while ignoring noise. In doing so, RA will assume the null distribution except for when associations dictate otherwise, a process referred to as ‘maximizing entropy subject to constraint.’ Ultimately, an RA calculated distribution
indicates what probabilities would be expected for each combination of behavior states knowing only those associations included in the model.

Technically, many of the associations in an RA model are simply “one-way associations” that impose nothing more than the marginal probabilities of each variable. In analyzing network data, network members will likely express one behavior state more frequently than another. For example, any given provider will bill for most patients zero times. These relative frequencies are captured in an RA model as marginal distributions, or what are called ‘main effects’ or ‘one-way associations’ in LL modeling. By themselves, one-way associations would indicate isolates in a network. If a member’s behavior appears to occur independently from all other members’ behaviors, an RA model will express this member as a one-way association. The model \( x_1: x_2: x_3 \) indicates that members \( x_1, x_2 \) and \( x_3 \) are all independent.

A model that contains all one-way associations (and no higher-way associations) is called the independence model. The independence model indicates the probability distribution that would be expected if all variables behaved independently. This is just like the expected distribution in a \( \chi^2 \) test of independence, consisting of joint probabilities of all main effects. In our case, the independence model would suggest a ‘network’ where all members were isolates. This model is more complex than the null model described previously. It contains basic information about the marginal distributions of members’ behaviors, so all probabilities are not equal. But in most applications of RA, these one-way associations are assumed to be a more reasonable starting point than the null model.
Reconstructability analysis will often derive two-way associations among variables. When a model includes two-way associations, such as the model $x_1x_2: x_3x_4$, it suggests that patterns in the data are more complex than can be accounted for by main effects. Using the independence model as reference, two-way associations are included only if they can provide information that goes sufficiently above and beyond the accuracy of the independence model. That is, two-way associations must maintain the marginal probabilities for each main effect and must specify additional information that substantially\textsuperscript{13} improves the accuracy of the model’s calculated distribution.

Two-way associations are akin to the dyadic relationships in standard NA. They indicate pairs of members whose behaviors are related in some way. These associations can stem from a wide variety of patterns in the data. Even in the simple case where variables have only two behavior states (e.g., 0=no claims billed, 1=claims billed), RA might identify an association based on behavior patterns that are consistent ($x_1^0x_2^0$ or $x_1^1x_2^1$) or opposite ($x_1^0x_2^1$ or $x_1^1x_2^0$) between members. When there are more than two behavior states (e.g., 0=no claims, 1=one claim, 2=multiple claims), two-way associations might stem from the predominance of any combination of behavior states among the two members (e.g., $x_1^0x_1^1$, $x_1^2x_1^1$, $x_1^3x_1^2$).

\textsuperscript{13} Determination of ‘substantial’ improvement will be discussed under Model Selection on p.54.
RA can also identify instances of three-way and even higher-way associations among members. This level of complexity goes beyond what is usually available with standard NA, indicating sets of three or more individuals whose behaviors are interrelated. The model $x_1x_2x_3: x_4$ indicates a three-way association between members $x_1$, $x_2$, and $x_3$, with member $x_4$ as an isolate. Like with two-way associations, a three-way association must maintain the marginal probabilities of all the associations it subsumes (one-way and two-way), adding additional information that substantially improves the model’s accuracy beyond those smaller associations.

Also like two-way associations, higher-way associations can stem from a wide variety of patterns in the data. In the simple case where variables have only two behavior states, a three-way association might stem from a behavior pattern where all members’ behaviors tend to match ($x_1^0x_2^0x_3^0$ or $x_1^1x_2^1x_3^1$), or where two members tend to match while the third behaves oppositely (e.g., $x_1^0x_2^0x_3^1$, $x_1^0x_2^1x_3^0$, $x_1^1x_2^0x_3^0$). With standard NA approaches, the connections among members can only be defined in terms of strength and direction (i.e., positive and negative), which is limiting. With RA, more nuanced connections can be quantified, such as ‘hypergraph’ associations between three or more members, and nonlinear relationships can be identified. This is possible in part because of RA’s treatment of data as nominal variables, which allows for a wider variety of patterns to be captured under a more general definition of relationship.
Model Selection

The technical process of RA modeling involves calculating the probability distributions for a large number of models and determining which of these alternatives provides the ‘best’ representation of the data. Determining the best model to represent an observed frequency distribution must take into consideration two desirable features: minimal complexity and maximal accuracy. First, a good model should be as simple as possible. The whole point of deriving a model is to reduce the complexity of the data to a calculated distribution that identifies prominent associations but ignores the noise. For this reason, an RA model cannot simply replicate the probability distribution that was observed in the data (sometimes called the fully saturated model), because it would most likely be overfitting, implying that there is an all-way interaction between all variables that cannot be simplified at all. This is rarely true.

Second, a good model should still include enough information to accurately represent the relationships that exist among variables. If the only concern were simplicity, as described above, one might just as well go with the independence model, the probability distribution that would be expected if all members were isolates. But this simple extreme is also unlikely to be accurate – a good RA model must still capture the prominent features in the data. It can’t just ignore them for the sake of simplicity. In most situations, neither the independence model nor the fully saturated model will be appropriate representations of the data. So model selection criteria are used to identify a model that is as accurate as possible but not too complex.
In RA, complexity is defined by the number of degrees of freedom \( df \) that are used up by a given model. For example, a dataset with three binary variables (and eight combinations of behavior states) can be represented as a model that uses anywhere from 3 to 7 \( df \). To minimize the complexity of an RA model is to minimize the number of degrees of freedom it requires. Actually, the null model, with equal probabilities for every combination of behaviors, requires 0 \( df \), but in standard applications of RA one-way associations are assumed from the start of the analysis, so the independence model is thought to require the fewest \( df \) possible (in this case 3). Models that are closer to the independence model can be said to be simpler (i.e., less complex).

In the most straight-forward application of RA, one starts with the independence model as a reference point and works “upward” toward the fully saturated model (i.e., the observed probability distribution). Models are identified at intermediate levels, each with higher complexity \( df \) requirements than the last. In a sense, the goal is to advance toward the saturated model so long as the benefits gained in accuracy outweigh the costs incurred in complexity. Incremental steps upward in complexity are governed by a lattice of structures (Zwick, 2001), which ensures that all models are hierarchically nested so they can be easily compared with a log-likelihood ratio (Neyman & Pearson, 1928a, 1928b).

There are at least three model selection criteria that can aid in identifying the best model: Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and incremental alpha \( \alpha_{inc} \). In the proposed dissertation, only BIC will be used because it is the most conservative and most likely to avoid a Type I error, but all three will be
described for the sake of completeness. The first two criteria, AIC and BIC, reward a model’s accuracy while also penalizing its complexity. With both of them, accuracy is defined as the similarity between a model’s calculated probability distribution and the data’s observed probability distribution, and complexity is defined by the number of degrees of freedom that a model uses. The equation for calculating a model’s AIC is

\[
AIC = -2N \sum p \ln q + 2df,
\]

where \( p \) refers to values in the data’s observed probability distribution, \( q \) refers to values in the model’s calculated probability distribution, and \( N \) refers to total the number of observations. Essentially, the summed terms \( \sum p \ln q \) indicate the extent of a model’s error, with greater error indicating a worse model. This penalty is weighted by the number of observations, as error is less forgivable with large samples, and is added to \( 2df \), a term which includes the model’s complexity.

The equation for a model’s BIC is quite similar,

\[
BIC = -2N \sum p \ln q + df \ln N,
\]

but here the penalty for complexity \( (df \ln N) \) scales up with the sample size, rather than acting as a fixed constant \( (2df) \). This scaling works to compensate for the increased risk of overfitting that is present when modeling datasets of large sample size. Because of its wariness against overfitting, BIC tends to be more conservative than AIC, and when the two disagree BIC will choose a simpler model.

The third selection criterion, \( \alpha_{inc} \), requires each increase in complexity (from the reference independence model) to produce a statistically significant improvement in the
calculated distribution. For example, the calculated distribution for a model with a three-way association \((x_1x_2x_3)\) must be significantly more accurate than a model with three two-way associations \((x_1x_2 : x_1x_3 : x_2x_3)\), which in turn is significantly more accurate than a model with two two-way associations \((x_1x_2 : x_1x_3)\), and so on for every step down to the independence model \((x_1 : x_2 : x_3)\). Significance testing for \(\alpha_{inc}\) is conducted with the standard log-likelihood ratio (LLR)

\[
LLR = \sum p \log \frac{q_{complex}}{q_{simple}}
\]

where \(p\) refers to the values of the data’s observed probability distribution, \(q_{simple}\) refers to values in the calculated distribution of the simpler model (e.g., \(x_1x_2 : x_1x_3\)), and \(q_{complex}\) refers to values in the calculated distribution of the more complex model under consideration (e.g., \(x_1x_2 : x_1x_3 : x_2x_3\)).

When using \(\alpha_{inc}\) as the model selection criterion, the value of the LLR is compared to a \(\chi^2\) distribution to test for significance. The \(\chi^2\) takes into account the sample size \(n\) and the difference in degrees of freedom (\(\Delta df\)) between the two models under comparison. If the LLR is larger than the critical \(\chi^2\) value, one rejects the null hypothesis (that the two models are equal) and concludes that the more complex model is a significant improvement over the simpler model. This hypothesis test is repeated for every increase in complexity, starting with the independence model and working toward the fully saturated model. Usually, a log-likelihood test will fail well before arriving at the fully saturated model. The simpler model in the failed test is declared the best model for that dataset, having championed the previous test and being determined as unbeatable.
I-DNA Step 2: Measuring Network Distance

When an RA model has been determined as ‘best’ by one or more criteria, its associations are indicative of the network structure underlying the dataset. And when observations have been categorized into two datasets, such as those collected before and after an intervention, best models can be derived for each dataset and then compared to identify network differences. For example, the two networks shown in Figure 3.2 (p.48) could indicate changes in the network structure for members $x_1$ through $x_8$.

Descriptively, one can directly compare two sets of associations, noting which ones appear or disappear after an intervention takes places. This can be displayed in a network graph with a different line styles for edges according to whether they appeared in the before network only, the after network only, or both networks. More detailed comparisons are also possible by evaluating changes in the strength of associations that persist over both time periods. Here, the strength of an association is defined as the reduction of uncertainty it provides. At each time period, and association’s strength is calculated with a log-likelihood ratio (LLR; also called ‘transmission’) that looks only at the probability distributions for that association, comparing the best model and the independence model for their similarity to the observed probability distribution. Two LLRs can be compared, indicating changes in the strength of associations which can help to describe an intervention’s likely effect (eqn 18, p.88).

Beyond descriptive analysis, a statistical comparison of two RA network models can be accomplished with a summary statistic that indicates the ‘distance’ between the
networks before versus after the target event or intervention. The equation for network distance by RA is

\[ \hat{\theta}_{RA} = \sum |q^2 - q^1| , \]  

(8)

where \( q^1 \) refers to values in the calculated distribution of the before model (time 1) and \( q^2 \) refers to values in the calculated distribution of the after mode (time 2). We can basically sum up the extent to which all calculated probabilities changed between time periods, into a distance measure that is inherently standardized because model probabilities always add up to 1.

Once we’ve calculated the distance between the before and after networks by RA, we can test that distance for statistical significance with any of the three types of data resampling that were described in the previous section (p.32). We can create multiple resampled pairs of datasets, including a pseudo before and pseudo after dataset, use RA to determine the best model for each one, and then compare each pair of resampled datasets to create a reference distribution of resampled network distances by RA. That distribution can then be used to help us determine the probability of a Type I error, and ultimately whether the observed network distance is statistically significant. If the type I error is sufficiently unlikely \((p < .05)\), we can reject the null hypothesis that the two network models are equal and conclude that a significant change occurred in the network’s structure.
**Method 2: Regression-based Differential Network Analysis**

R-DNA is the second method proposed for analysis of network change. The R-DNA approach parallels the I-DNA approach, consisting of three existing methods that, collectively, can test a network for significant change. R-DNA uses linear regression analyses to infer network structures, mean absolute difference (MDA) to measure the distance between two networks, and a resampling technique to test the distance for significance. The first of these methods, linear regression, is commonly used as a descriptive method for network comparison in bioinformatics (Horvath, 2011). It uses a set of network observations to infer *connectivity scores*, or linear relationship coefficients, between all pairs of network members. When datasets can be compiled for multiple time periods (e.g., before vs. after), connectivity scores can be compared to reveal changes in the connectivity of a network over time.

The descriptive insights are themselves useful, but a global test of network change requires all differences in connectivity to be summarized with a single value. In using network inference for statistical testing, differences in connectivity scores are aggregated into the *mean absolute difference*, or MDA ($\bar{\theta}$), which is a global measure of network distance. The MDA is then subjected to a resampling-based hypothesis test in R-DNA.

Statistical testing of network distance in R-DNA is analogous to statistical testing in I-DNA. Multiple pairs of resampled datasets are created, and a regression-based method is used to infer a network structure from each resampled dataset. We can then calculate the MDA distance on each pair of resampled datasets, and we can aggregate them into a distribution of resampled distances. This distribution can be used to
determine the statistical significance of the distance that we observed in our original pair of datasets. If the probability of a Type I error is sufficiently low ($p < .05$) we can reject the null hypothesis and claim that the network changed significantly over time.

(Descriptive) Differential Network Analysis

The basis for R-DNA is its parent method, Differential Network Analysis, or DNA. Unlike RA, which is a general method applied to many fields, DNA was developed in the field of bioinformatics, and has been used primarily to identify differences in networks of genes and proteins (Fuite et al., 2008; Fuller et al., 2007; Horvath, 2011; Weckwerth et al., 2004). In its field of origin, DNA is primarily used to compare genetic networks of differing types, such as those of lean and obese mice (Fuller et al., 2007) or individuals with and without type II diabetes (Yates & Mukhopadhyay, 2013) or chronic fatigue syndrome (Fuite et al., 2008). These studies use microarray data from a number of networks (i.e., organisms) of each type. However, DNA is commonly not coupled with statistical testing, and only descriptively compares two networks. To differentiate descriptive DNA from the two forms of statistical DNA used here (I-DNA and R-DNA), it will be referred to as D-DNA (Descriptive Differential Network Analysis) when no statistical comparison is being made.

In applying D-DNA, researchers first summarize the observations from each network type by deriving connectivity scores. A connectivity score is a value to indicate the strength of a relationship between two ‘nodes.’ In bioinformatics, connectivity scores are derived for pairs of genes, but in social networks they can be derived for pairs of network members. Connectivity scores are calculated separately for each network – in
our case the before and after networks – and are arranged in a connectivity matrix. Often, connectivity scores are undirected, so the connectivity matrix is a diagonally symmetric table with one row and one column for each member. However, if directed relationships are desired, a connectivity matrix will be asymmetric, indicating different values for the relationship from member $i$ to member $j$ than from member $j$ to member $i$. In either case, the matrix summarizes the strength of all pairwise relationships in a network type, and two connectivity matrices can be compared to identify differences in connectivity. In the context of network analysis, this comparison would indicate relationships that appear, disappear, strengthen, or weaken over time.

In bioinformatics, applications of D-DNA have defined genetic connectivity scores with a variety of metrics, including correlations (Fuller et al., 2007; Weckwerth et al., 2004), partial correlations (Yates & Mukhopadhyay, 2013), partial least squares (Gill et al., 2010), and even mutual information (Fuite et al., 2008). In using D-DNA without intention of statistical comparisons, it is even possible to base connectivity scores on direct network measurements, such as frequency of interaction or self-reports of affinity. However, in using D-DNA for statistical testing, differences in connectivity must be aggregated into a single measure of distance. With that purpose in mind, standardized measures of connectivity are best, such as those provided by regression-based approaches.

**Data Requirements**

There is basic overlap in the data requirements for I-DNA and R-DNA. Both approaches use variables to represent network members and cases to represent independent
observations of members’ behaviors. However, in I-DNA network members were represented as nominal variables, whereas in R-DNA network members are represented as continuous variables. This can be advantageous when continuous data is available on member behaviors, as R-DNA can make use of continuous information without requiring the researcher to discretize (i.e., bin) data into nominal categories. The example data shown in Table 3.2 (p. 37) – with variables to represent billing providers and cases to represent patients’ frequency in claims – is continuous in nature and can be used ‘as is’ to generate connectivity scores for the network of health care delivery.

The amount of data required for R-DNA depends on which method will be used to derive connectivity scores, but these methods tend to require less data than I-DNA. The simplest method, correlation, only considers two variables at a time and for that reason requires the least amount of data. The precise minimum number of cases recommended for a correlation depends on the expected strength of correlation and the desired levels of statistical power ($\beta$) and Type I error ($\alpha$). But with the standard levels for social sciences ($\beta = .80$, $\alpha = .05$), a correlation coefficient that is at least moderate in size ($r = .3$; Cohen, Cohen, West, & Aiken, 2003; Cohen, 1988) requires only 84 cases. Even correlations that are small in size ($r = .1$) require only 782 cases to have an 80% chance of detection.\textsuperscript{14} For reference, application of I-DNA to a network of only 10 members with binary behavior (0 or 1) would probably require at least hundreds of cases as well.

\textsuperscript{14} All power analyses were conducted using G*Power (Faul, Erdfelder, Buchner, & Lang, 2009)
A more nuanced method of deriving connectivity scores is through multiple linear regression. Here, a linear model is calculated separately for each network member \((i)\), predicting that member’s behavior from the behavior of all potential other members \((j)\). When network members are operationalized as variables (e.g., Table 3.2, p.37), this is essentially the same as the standard multiple linear regression equation, where a dependent variable is regressed on a set of independent variables. Multiple regression is preferable to correlation analysis when several alters (members who are connected to the focal member) are similarly correlated with the focal member. Here, all variance in the focal member’s behavior that is predicted equally well by multiple alters will be ignored, so the resultant connectivity scores can be said to indicate alters’ *unique* connectivity, after controlling for the connectivity of the other alters. When using multiple regression, connectivity scores are defined as a set of partial correlation coefficients.

More observations are required to calculate connectivity scores with multiple regressions than with correlations, because a multiple regression estimates multiple coefficients simultaneously. Specific case requirements increase with the number of network members, though at a diminishing rate. For a network with 10 members, the use of multiple regression would produce 10 regression equations, each with 9 partial correlation coefficients. If the coefficients predict the focal member’s behavior at least moderately, 197 cases are recommended, and if their predictive capacity is small in size, 1,818 cases (see Green, 1991). These recommendations are more than twice what would be required to use simple correlations for the same network, but they are a “good deal”
considering the number of independent variables is eight-fold higher. Multiple regression requirements for data are substantial, but are still likely to be lower than those of RA.

A third option for calculating connectivity scores is the use of partial least squares (PLS) regression. PLS is like multiple regression in that it is essentially a linear representation of relationships. However, in addition to estimating the partial correlation coefficients for each predictor variable (IV), latent factors are extracted from the dependent and predictor variables, and these factors are optimized to maximize covariance between the dependent factors and the predictive ones. By adding these latent factors to the PLS regression equation, PLS is often more sensitive at detecting relationships. This has been found to be especially true when predictors are non-independent and when sample sizes are small (Carrascal, Galván, & Gordo, 2009; Datta, 2001; Pihur, Datta, & Datta, 2008), a strength that is ideal for network analysis because members are non-independent by definition, and often there are more network members (i.e., variables) than there are observations.

Specific data recommendations, however, are not well established for PLS. While it has been around since the 1970s (Wold, 1974, 1985), it was originally proposed as an alternative to principal components analysis, and currently its most popular application is in path modeling (Morales, 2011). PLS has only recently appeared in bioinformatics as a way to define network structures (Datta, 2001), and there is ongoing discussion regarding the minimum sample size. Some authors have argued that PLS can be used even when variables outnumber observations (Carrascal et al., 2009; Chin & Newsted, 1999; Datta, 2001), but this claim has not gone unchallenged. At least when used for dimension
reduction or path modeling, PLS may not be any more robust to small sample sizes than multiple regression, and may be equally likely to suffer low statistical power (Goodhue, Lewis, & Thompson, 2012; Henseler, Ringle, & Sinkovics, 2009). Until more research is done on this issue, sample sizes comparable to multiple regression requirements (e.g., Green, 1991) may be advisable.

Like with I-DNA, R-DNA relies on resampling techniques for statistical testing, so the same variables must be present in both datasets. In applying R-DNA to social network analysis, the network should contain the same members during both time periods of interest. If network membership changes between the before and after networks, only members who were measured during both time periods should be included for analysis.

Assumptions

R-DNA methods for deriving connectivity scores have several statistical assumptions. Like RA and resampling methods, all three connectivity methods hold the basic assumptions that observations are independent and identically distributed (iid). However, R-DNA holds some additional assumptions, because R-DNA methods for defining connectivity scores are all forms of linear modeling. In defining relationships linearly, correlation, multiple regression, and partial least squares all hold an inherent assumption that the relationships between network members are linear in nature.

Correlation and multiple regression hold several additional assumptions that are common to standard statistical analysis. They assume normality in arrays, that the values of an independent variable are normally distributed for any value of the dependent
variable, and *homoscedasticity*, that the variance of independent variables are equal for every value of the dependent variable (Tabachnick & Fidel, 2006). These assumptions may not always be met, depending on the distribution of behaviors of participating network members. Medical providers, for example, will likely see most patients ‘never’ few patients ‘rarely’ and even fewer patients ‘often,’ resulting in a distribution that is highly positively skewed.

Multiple regression, because of its inclusion of multiple predictor variables, is also assumed to contain *all* relevant predictor variables, an assumption referred to as *specification* (Tabachnick & Fidel, 2006). In the case of social network analysis, this assumption is easily met if all network members were measured during both time points. But it may be called into question if influential network members were not present at both time points. If membership changes drastically, it is possible that some relevant independent variables (i.e., central and influential members) would be removed from the analysis, which would bias estimates of the predictive power of alters (IVs) that remained.

Partial least squares (PLS) is commonly believed to have less stringent distributional assumptions than regression (Henseler et al., 2009), and to be more robust to problems of *multicollinearity*, where predictor variables are themselves correlated (Gustafsson, 2004; Haenlein & Kaplan, 2004). In support of these beliefs, PLS has outperformed multiple regression in several areas of application (Carrascal et al., 2009; Gustafsson, 2004; Pihur et al., 2008) and Monte Carlo simulation has supported its robustness to things like skewed distributions, multicollinearity, and model
misspecification (Cassel, Hackl, & Westlund, 1999). However, some comparative research suggests that other methods, including multiple regression and maximum likelihood estimation, are equally robust to diversions from normality (Goodhue et al., 2012; Vilares, Almeida, & Coelho, 2010), and that PLS regression coefficients may be biased in cases of multicollinearity (Henseler et al., 2009; Marcoulides, Chin, & Saunders, 2009), especially when sample sizes are small (Goodhue et al., 2012). Because of the concerns that have been raised, a conservative application of PLS would observe as closely as possible the assumptions and data requirements of multiple regression.

**R-DNA Step 1: Inferring Network Structure**

R-DNA uses linear regression approaches to define the structure of a network as a set of pairwise relationships among network members. Each of three methods – Pearson’s correlation (PC), multiple linear regression (MLR), and partial least squares (PLS) – can infer connectivity scores for all pairs of network members when they are defined as continuous variables (e.g., Table 3.2, p.37). Connectivity scores for each pair of network members can then be arranged in a connectivity matrix to indicate the full structure of a network. In the connectivity matrix, diagonal values will indicate a perfect relationship between each member and itself (1), and off-diagonal values will indicate the strength and direction (inverse or direct) of the linear relationship between each pair of members.

**Pearson’s Correlation**

Three regression-based techniques for inferring connectivity scores will be described here, although the proposed analyses will only use PLS. The simplest method
for deriving connectivity scores is to use Pearson’s Correlation (PC) to calculate correlation coefficients for each pair of network members \( i \) and \( j \).

\[
 r_{ij} = \frac{1}{n-1} \sum \left( \frac{x_i - \bar{x}_i}{s_i} \right) \left( \frac{x_j - \bar{x}_j}{s_j} \right) \tag{9}
\]

In this equation, terms \( \bar{x}_i \) and \( \bar{x}_j \) indicate the mean values observed for network members \( x_i \) and \( x_j \), terms \( x_i \) and \( x_j \) indicate individual measurements of those members, and terms \( s_i \) and \( s_j \) indicate the variances of their observed behavior distributions. Mathematically, a correlation amounts to the average distance of all observations from the means of both variables. For large coefficients, observations that are far from the mean in one variable also tend to be far from the mean in the other variable. This indicates a strong relationship between those variables, and in this case a strong connection between those two network members. Coefficients in the matrix might be either positive or negative, indicating whether two variables tend to behave similarly or inversely. Correlation matrices will always be symmetric, because a correlation between members \( x_i \) and \( x_j \) will always be identical to the correlation between \( x_j \) and \( x_i \).

Multiple Linear Regression

Multiple regression (MLR) is a second approach to deriving connectivity scores, and it may be preferred to correlation because it calculates alters’ unique predictive capacities, rather than their full (often duplicative) predictive capacities. In Figure 3.3 below, connectivity scores are shown for a hypothetical network where all members behave fairly similarly. On the left, scores have been calculated with PC. Because all members behave similarly, each of member \( x_3 \)’s alters \( (x_1, x_2, x_4, \text{ and } x_5) \) can predict
20-30% of $x_3$’s behavior. However, in the aggregate these scores do not make sense – their total value of 1.1 would suggest that the alters can collectively predict 110% of member $x_3$’s behavior. On the right, connectivity scores have been recalculated as partial correlation coefficients with MLR, and they are shown to be much smaller. Some predictive capacity was duplicated among $x_3$’s alters, and shared predictive power is ignored in MLR. So, the figure at right indicates only the unique predictive capacities of each alter.

![Hypothetical Connectivity Scores by PC (Left) and MLR (Right)](image)

**Figure 3.3. Hypothetical Connectivity Scores by PC (Left) and MLR (Right)**

Calculation of MLR follows the standard equation,

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p,$$

where $\hat{y}$ indicates the predicted value of the dependent variable, $\beta_0$ indicates the intercept of the linear solution that best fits the data, and the terms $\beta_1 x_1, \beta_2 x_2$, etc. indicate the partial correlation coefficients ($\beta$) assigned to each independent variable ($x$). In the case of analyzing networks, we might substitute the letters $j$ and $i$ so that the equation more intuitively describes the use of alters ($x_1$ through $x_j$) to predict the behavior of a focal member ($x_i$).
\[ \hat{x}_i = \beta_i^{(0)} + \beta_i^{(1)}x_1 + \beta_i^{(2)}x_2 + \cdots + \beta_i^{(j)}x_j \] (11)

In solving equation 11, the term \( \beta_i^{(1)} \) and \( \beta_i^{(2)} \) are taken as the connectivity scores between the ego member \( x_i \) and the alter variables \( x_1 \) and \( x_2 \), respectively. To use MLR in conducting R-DNA, one would calculate a regression equation for each member, considering all other members as potential alters in that equation.

Like correlation coefficients, connectivity scores from MLR will vary in strength and direction (i.e., positive or negative), but unlike correlation coefficients they will not necessarily be symmetric. This has to do with the way MLR ‘controls for’ the predictive capacity of other alters. For example, a regression of member \( x_1 \) on variables \( x_2 \), \( x_3 \), and \( x_4 \) will calculate B’s predictive power while controlling for the extent to which \( x_3 \) and \( x_4 \) also predict \( x_1 \). By contrast, the seemingly opposite regression of member \( x_2 \) on variables \( x_1 \), \( x_3 \), and \( x_4 \) will calculate \( x_1 \)’s predictive power while controlling for the extent to which \( x_3 \) and \( x_4 \) also predict \( x_2 \). At the crux of this difference, members \( x_3 \) and \( x_4 \) may have a different amount of predictive capacity for \( x_1 \) than they do for \( x_2 \). In controlling for their differential influence, MLR will often produce asymmetric connectivity scores.

Partial Least Squares

Partial least squares (PLS) is the third and most nuanced method for deriving connectivity scores for R-DNA. PLS has been found to be effective at identifying relationships in biological networks (Datta, 2001; Gill et al., 2010), and it has been shown to outperform correlation and pairwise mutual information measures in analyzing real
and simulated biological data (Pihur et al., 2008). This method is expected to be the most useful approach for inferring social network structure.

PLS can be understood best by comparison to standard multiple linear regression (MLR) and principal components regression (PCR; Stone & Brooks, 1990). In MLR, one measures the strength of each independent variable (IV) in predicting a dependent variable (DV). Terms are defined for each IV to maximize its correlation with the DV while also accounting for the predictive capacity of other IVs. These terms, called partial correlation coefficients, are defined directly from variance in the IVs, so they can be called "given regressors."

In PCR by contrast, variance is analyzed to identify latent factors among a set of IVs (Stone & Brooks, 1990). For example, several measurable variables regarding salary, tenure status, and volume of publications might collectively indicate an underlying, or latent factor of career achievement. PCR first pools the variance of all IVs and then finds the maximum amount of variance that can be accounted for by a single vector. This vector is identified as the first latent factor, and its variance is removed from the pool so that the second latent factor can be identified within the variance that remains. This process continues until latent factors can account for all of the variance among IVs. Note that in PCR, latent terms are not defined to predict a DV at all. Instead the purpose is to create latent variables that account for maximal variance among the IVs. These latent variables, composed of multiple IVs, can be called “constructed regressors.”

Partial least squares (PLS) is somewhere between MLR and PCR (Stone & Brooks, 1990). Like MLR, PLS builds a linear model to predict the DV from a set of IVs.
But like PCR, a PLS model constructs latent factors instead of relying on given regressors. These latent factors (which are indicated by superscripts in parentheses), are ordered based on their ability to maximize the amount of covariance between the DV and IVs. This gives PLS greater flexibility than MLR in creating a linear model that will best predict the DV. The basic equation for PLS (adapted from Pihur et al., 2008),

\[
\hat{x}_i = \beta_i^{(0)} + \beta_i^{(1)} t_i^{(1)} + \beta_i^{(2)} t_i^{(2)} + \cdots + \beta_i^{(\ell)} t_i^{(\ell)},
\]

looks very similar to the equation for MLR (11), as they are both equations for linear regressions. But here the terms \( \beta_i^{(1)} \) and \( \beta_i^{(2)} \) refer to the weights of constructed (latent) regressors \( t_i^{(1)} \) and \( t_i^{(2)} \) instead of partial correlation coefficients like in standard MLR.

In using PLS for R-DNA, a PLS regression equation is calculated for each network member \( (x_i) \), using latent terms from the variables of all other members \( (x_1, x_2, \ldots x_{i-1}, x_{i+1}, \ldots x_j) \). These latent regressors, \( t_i^{(1)} \) through \( t_i^{(\ell)} \), are first constructed in a manner similar to PCR. The covariance of all potential alters (IVs) is pooled, and the first latent factor is identified as the vector with the highest possible amount of covariance with the DV. This covariance is then removed so that the remaining covariance can be used to define subsequent terms.

A couple of additional constraints are present in the construction of latent variables. First, all latent terms must be orthogonal, or non-overlapping in their covariance with the DV. This ensures that all terms in the PLS equation represent unique predictive capacities. Second, the researcher defines a maximum number of latent regressors to be allowed, because the use of all possible latent factors would likely overfit
the data. Guidelines for the specific number of regressors are not well-established for PLS, but previous work suggests that very few are needed. In studying networks of 50 to several hundred genes, Pihur and colleagues (2008) found that PLS outperformed alternate methods with as few as three latent factors, and that it became less effective when more factors were used.

Once all latent terms have been defined (see Stone & Brooks, 1990 and Gill et al., 2010 for more mathematical detail), the PLS equation is put together through the standard method of ordinary least squares. Weights of the constructed regressors ($\beta_1^{(1)}, \beta_1^{(2)} ...$) are calculated to maximize the overall covariance with the DV, and individual latent terms ($t_1^{(1)}, t_1^{(2)} ...$) contain information regarding the IVs that contributed to them. In the equation below,

$$ t_1^{(f)} = c_{i1}^{(f)} x_1 + c_{i2}^{(f)} x_2 + ... + c_{i(i-1)}^{(f)} x_{i-1} + c_{i(i+1)}^{(f)} x_{i+1} + ... + c_{ij}^{(f)} x_j $$

the latent factor $t_1^{(f)}$ is shown to consist of the sum of contributions $c_{ij}^{(f)}$ from each of the alters (IVs) in the dataset ($x_1, x_2, ... x_{i-1}, x_{i+1}, ... x_j$). In using PLS for R-DNA, this means that a latent factor can contain some amount of predictive power from each potential alter in the network. This information holds the key for calculating connectivity scores with PLS, as it shows the magnitude of an alter’s predictive capacity for the focal member’s behavior. However, there are multiple latent factors that comprise the PLS equation, so the contributions must be aggregated to indicate an alter’s full predictive capacity for a focal member.
The formal equation for calculating a connectivity score is
\[ \hat{s}_{ij} = \beta^{(1)}_i c_{ij}^{(1)} + \beta^{(2)}_i c_{ij}^{(2)} + \cdots + \beta^{(\ell)}_i c_{ij}^{(\ell)}, \]
(14)
where an alter’s contributions to latent terms \( c_{ij}^{(\ell)} \) are multiplied by partial least squares coefficients \( \beta^{(\ell)}_i \), and then summed. A visual depiction of this process is also provided on the following page in Figure 3.4. Here, PLS is shown to provide connectivity scores from members \( x_2, x_3 \) and \( x_4 \) to the focal member \( x_1 \). A PLS regression equation contains several latent terms, each of which have contributions from alters \( x_2, x_3, \) and \( x_4 \). Those contributions are aggregated to determine each alter’s connectivity to member \( x_1 \).

As with MLR, PLS connectivity scores will often be asymmetric. In equation 14 above, \( \hat{s}_{ij} \) indicates the sum of member \( x_j \)’s contributions to the prediction of member \( x_i \), but the sum of contributions of member \( x_i \) to the prediction of member \( x_j \), will likely be a different value. If desired, directed connectivity scores can averaged to create undirected scores. One can sum \( \hat{s}_{ij} \) and \( \hat{s}_{ji} \) and then divide the sum in half (Gill et al., 2010). This simplifies the network and is desirable when predictive directions do not have any meaningful significance. But whether directed or undirected, PLS scores appear to maintain distinct advantages over scores derived by PC or MLR. They reflect only unique predictive capacities between pairs of network members and augment members’ ability to demonstrate connectivity via latent factors.
False Discovery Rate

In applications of D-DNA, considerable attention has been given to Type I errors, or the false discovery rate (FDR, see Dudoit, Shaffer, & Boldrick, 2003). When connectivity scores are calculated with statistical techniques, such as PC, MLR, or PLS, each one can be tested for significance. However, networks with more than a few members feature a great many connections—a network with 100 members can have as many as 4,950 connections when undirected, and 9,900 when directed. With a typical Type I error rate of .05, a single rejected hypothesis has only a 5% chance of being false. But when conducting hundreds or even thousands of hypotheses, false discoveries become a systematic problem.

Several techniques exist for addressing FDR in multiple hypothesis testing. The simplest and most conservative approach is the Bonferroni correction for multiple...
hypothesis tests (Cabin & Mitchell, 2000; Rice, 1989). Here, the researcher sets a familywise error rate (FWER) at the level desired for Type I error (e.g., $\alpha' = .05$) and divides that rate by the number of hypothesis tests that will be conducted. For each individual hypothesis test, the critical value is adjusted to reflect a probability that is a small fraction of the FWER, ensuring that total risk of Type I error will not advance beyond the desired rate.

The drawback to Bonferroni’s correction is that a study completing many hypothesis tests will need to employ a very small Type I error rate for each hypothesis. This is especially problematic in bioinformatics, where it’s not uncommon to test thousands of hypotheses simultaneously. The Type I error rate is linearly related to statistical power, so forcing it to be small can prevent researchers from identifying many (true) network connections. There have been several adaptations to Bonferroni’s correction, including procedures by Šidák (1967), Holm (1979), and Hochberg (1988), and there are also alternatives outside of Bonferroni (see Dudoit et al., 2003 for an applied review). Most recent developments in bioinformatics employ Bayesian inference with hierarchical modeling (Efron, 2004; Gómez-Villegas, Salazar, & Sanz, 2013; Yi, Xu, Lou, & Mallick, 2013) and resampling techniques (Dudoit, Gilbert1, & van der Laan, 2008; Li & Dye, 2013). But standards for are not well-established in bioinformatics yet and are beyond the scope of this work.

Fortunately, when using R-DNA for social network analysis, the statistical significance of individual connectivity scores is relatively unimportant. Recall that the primary aim of R-DNA is to test for network change at a global level. Changes in
individual connectivity scores are only of post hoc concern if the central hypothesis indicates change in the overall network structure. When global change is detected, there may be reason to look closer at specific members or specific connectivity scores. But even then, the researcher could test for significant change in a connectivity score without having established whether the ‘before’ and ‘after’ scores are themselves significant. These follow up tests would likely be reserved for the few connections that changed most drastically between time periods, and this would minimize the FDR problem considerably. With only a few post-hoc tests, simpler approaches to FDR, like Bonferroni’s correction, would likely be sufficient.

**R-DNA Step 2: Measuring Network Distance**

Once all connectivity scores have been derived, they can be arranged in a connectivity matrix to indicate the full structure of a network. In looking for network change, a connectivity matrix can be formed for each time period, and the values can be compared. Many software packages will be able to use the connectivity information in this matrix to produce network visualizations that will illustrate the network’s structure. However, some fine tuning may be necessary to assist descriptive comparison, and some additional calculations are necessary to test network change for statistical significance.

**Descriptive Comparisons**

Descriptive network analyses provide a basic way to illustrate network change over time. Individual connections may become stronger or weaker after an intervention, and individual members may be connected to different alters. Some basic network
metrics might also facilitate the interpretation of network change, including the density of connections, the average shortest path length, and the degree to which members appear to be centralized. With metrics such as these, networks can be examined to see if they have the properties of common network types, such as scale-free networks.

It may be necessary to filter out some connections before running descriptive network analyses. Regression-based techniques are likely to find connectivity scores between nearly all members, even if the majority of those connections are very weak. Members who share at least one patient will have a slightly positive correlation, and pairs who share no patients but serve at least one patient on their own will have a slightly negative correlation. There may be very few instances where two members’ behaviors are precisely independent ($s_{ij} = 0$), and weak connections may not be meaningful enough to pay attention to in analyzing the network. Visualizations, for example, are more likely to be illustrative of a network pattern if they include only the most important connections, rather than all connections among all members.

Many network metrics are also unable to take connection weights into account. Standard calculations for density, path length, and centralization, for example, consider connections to be present or absent and do not account for changes in strength over time. One approach is to test for the significance of each connection, excluding those that are ‘insignificant’. But there are issues associated with the rate of false discovery that are not entirely resolved (discussed above), and with extremely large datasets the vast majority of connections may still be significant. In the current project, the CCO claims dataset has
5.6 million claims and over 183 thousand patients. With sample sizes this large, even very weak connectivity scores may be found to be statistically significant.

When making visualizations or conducting descriptive analyses, there are a couple of options for limiting the inclusion criteria. The first is to define a minimum absolute connectivity value (Gill et al., 2010; Weckwerth et al., 2004; Zhang et al., 2009), and to include only the connections that surpass it. In bioinformatics this threshold is often informed by previous research on the network, but a researcher could also select the top 5% or 10% of connections for analysis. Histogram might also inform a threshold if discontinuities can be seen, like in Figure 3.5 below where a subset of connectivity scores deviates from Gaussian distribution at either tail.

![Hypothetical Histogram of Connectivity Scores](image)

**Figure 3.5. Hypothetical Histogram of Connectivity Scores**

Another option for visualization criteria is to include only positive connectivity scores. This appears to be common practice in bioinformatics (Gill et al., 2010; Weckwerth et al., 2004), and may be appropriate for studying some social networks. However, in the CCO dataset one might be interested to find pairs of medical providers who serve distinct (i.e., opposite) patient populations and are in a sense negatively associated. The inclusion or exclusion of inverse associations, then, will likely depend on the social network under focus. We include them in this dissertation.
Mean Absolute Difference

Exclusion criteria can be useful for visualizations and descriptive analyses, but they are not necessary to test a network for significant change over time. R-DNA uses the mean absolute difference (MDA) measure (Gill et al., 2010) to summarize changes in network structure, and pairs it with data resampling techniques to discern whether those changes are statistically significant. The MDA metric ($\hat{\theta}$), can make use of all connectivity scores derived in Step 1 – exclusion criteria do not need to be applied to connectivity scores derived from the dataset or from any permuted dataset.

To calculate MDA, one begins with a connectivity matrix for each time period. In these matrices, diagonal values indicate a perfect relationship between each member and itself (1), and off-diagonal values indicate the strength and direction (inverse or direct) of the linear relationship between each pair of members. The equation for MDA,

$$\hat{\theta} = \frac{1}{g(g-1)} \sum_{i \neq j} |\hat{s}_{ij}^1 - \hat{s}_{ij}^2|,$$

is essentially a summation of all (absolute) differences in these connectivity matrices. In the notation here (adapted from Gill et al., 2010), the terms $\hat{s}_{ij}^1$ and $\hat{s}_{ij}^2$ indicate the connectivity scores that were estimated from the ‘before’ and ‘after’ matrices, respectively. Their absolute differences are summed and then divided by the total number of possible connections among network members ($g(g-1)$).

For reference, the MDA equation is quite similar to a variant of the Jaccard Index called ‘simple matching’ (Sokal & Michener, 1902), which has been applied to network analysis (see Snijders, Dormaar, & Dijkman-Caes, 1990). Simple matching treats all
connections as present (1) or absent (0) and defines network similarity as the number of consistencies (i.e., connections present/absent in both networks), divided by the number of connections possible. With simple matching, the distance between two networks is one minus the degree of similarity, the same value that MDA would produce if connection weights were ignored. When connection weights are relevant, MDA provides a more nuanced version of simple matching that can take into account subtle changes in connectivity.

In R-DNA, the MDA statistic ($\hat{\theta}$) is subjected to data resample testing – such as permutation or bootstrapping – to determine whether changes in network structure could have been produced by chance alone. By following the same R-DNA procedure with many permuted pairs of datasets (PLS connectivity scores and MDA measurement), it is possible to examine how much variance would be expected in MDA values even if there were no systematic difference between the observations before and after the intervention. Through resampling techniques, a reference distribution ($\mathcal{N}_{perm}(\theta)$) can be custom made to test the observed MDA for statistical significance. If the probability of a Type I error is determined to be sufficiently low ($p < .05$), the null hypothesis can be rejected, and network change can be claimed with reasonable confidence.

Post Hoc Analyses

When resample-based testing indicates a significant MDA, there may be follow-up questions regarding what specific changes occurred in the network after an intervention. Fortunately, adjustments to MDA provide the capacity to test specific
members and even specific connections for change over time. When the connectivity of a particular member \((x_i)\) has changed substantially, that change can be measured by

\[
\hat{\theta}_{x_i} = \frac{1}{g - 1} \sum_{j \neq i} |\hat{s}_{ij}^1 - \hat{s}_{ij}^2|.
\]  

(16)

In this variant of the MDA, \(x_i\) is the focal member, and the terms \(\hat{s}_{ij}^1\) and \(\hat{s}_{ij}^2\) represent the difference in connectivity scores between that member and each alter at times 1 and 2. Identical to the hypothesis testing of the standard MDA (equation 15), this member-specific MDA value can be compared to a reference distribution that is created from permuted datasets. For each resampled pair of datasets, the mean absolute difference can be calculated for the same focal member. The \(p\) value for the observed MDA can then be calculated with reference to a distribution of resampled differences in connectivity for that member.

Although not described in the (2010) paper by Gill, Datta, and Datta this process could even be followed for significance testing of a specific connection weight. The equation

\[
\hat{\theta}_{s_{ij}} = |\hat{s}_{ij}^1 - \hat{s}_{ij}^2|
\]  

(17)

indicates \(\hat{\theta}_{s_{ij}}\) as the absolute difference in connectivity for a specific connection \((\hat{s}_{ij})\) as observed over two time periods. Once this observed difference is calculated, the same difference can be calculated for each permuted pair of datasets. Then, depending on how often the permuted values are at least as great as the observed difference, this change in connection strength may be found sufficiently rare to be claimed as statistically significant.
When conducting follow-up tests for changes in specific connectivity scores or specific members, one should be highly selective. Testing all members and all connections for significant change would elicit problems with the false discovery rate (FDR), and the validity of significant findings would be highly suspect. If the number of post hoc tests is relatively small (i.e., \( \leq 50 \)), and if the sample size is relatively large, Bonferroni’s correction, or other simple and conservative methods, may be able to adequately address FDR without sacrificing too much statistical power.

**Comparative Application of I-DNA and R-DNA**

By applying I-DNA and R-DNA to the same datasets, the two methods can be compared for their performance in evaluating network change. Both approaches have the capacity to test for statistical significance, but each method differs in its definition of network relationships and in its measurement of network distance. This means the methods may have different capacities to statistically detect network change, and they may also identify different network structures in each dataset.

Some of the features unique to each method are summarized below in Table 3.4. R-DNA holds advantages in its use of continuous data, the volume of data required, and the ability to investigate changes in specific network members. I-DNA holds advantages in its freedom from distributional assumptions, in its detection of nonlinear and higher-way associations, and in its ability to filter network relations without risking false discoveries (by using BIC model selection to determine which associations are important). The two methods can be expected to derive different network structures because of their different approaches to inference. So, comparisons of their networks are
expected to illustrate the benefits of each method for facilitating descriptive and post-hoc analyses.

**Table 3.4. Comparison of Desirable Features via I-DNA and R-DNA**

<table>
<thead>
<tr>
<th></th>
<th>I-DNA</th>
<th>R-DNA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DATA</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use of Continuous Data</td>
<td>Discretized to nominal</td>
<td>☑ Can be used as is</td>
</tr>
<tr>
<td>Distributional Assumptions</td>
<td>☑ None</td>
<td>Normality in arrays, homoscedasticity</td>
</tr>
<tr>
<td>Volume of Data Required</td>
<td>Large</td>
<td>☑ Medium</td>
</tr>
<tr>
<td><strong>STRUCTURE</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relationship Definition</td>
<td>Associations (information theory)</td>
<td>Connections (linear modeling)</td>
</tr>
<tr>
<td>Number of Relations Found</td>
<td>Minimal – only the strongest associations</td>
<td>Maximal – most pairs will be connected</td>
</tr>
<tr>
<td>Types of Relations Found</td>
<td>☑ Linear, nonlinear, dyadic, triadic, higher</td>
<td>Linear, dyadic</td>
</tr>
<tr>
<td><strong>DISTANCE</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance Measure</td>
<td>Sum of probability differences</td>
<td>Mean absolute difference</td>
</tr>
<tr>
<td>Relationship Filtration for Descriptive Analysis</td>
<td>☑ Not necessary with BIC model selection</td>
<td>Bonferroni, strength threshold or alternative</td>
</tr>
<tr>
<td>Post Hoc Analyses Possible</td>
<td>Associations only</td>
<td>☑ Connections, members</td>
</tr>
</tbody>
</table>

☑ *Highlights advantages held by one method over the other.*

The most obvious way to compare I-DNA and R-DNA is by the results of their statistical testing. Each of them will produce a $p$ value to indicate the probability of a Type I error – the likelihood that chance variations in billing provider behavior could produce a network distance as large as the one we observed. So, we could compare the I-DNA and R-DNA on the extent to which they are able to statistically detect network change after the formation of the Health Share CCO.

A more nuanced and descriptive way to compare I-DNA and R-DNA is by comparing the actual networks that they derive. These networks can be visualized for
general comparisons, which may indicate different degrees of change or stability over time. Note that I-DNA will produce hypergraphs with a fixed number of associations (some of them potentially three-way or higher-way) as indicated by the best models during each time period, and R-DNA will produce pairwise relationships between nearly all network members, many of which are very weak in strength. These comparisons can be furthered by implementing visualization thresholds, where one sets a limit on the number of R-DNA connections that is approximately equal to the number of pairs of connected providers by I-DNA. This avoids consideration of exceedingly small R-DNA connections and allows for fairer comparisons between methods on basic metrics like the proportion of connections that appear or disappear between time periods.

Even with a visualization threshold, however, the comparison of higher-way connections by I-DNA to a set of pairwise connections in R-DNA is still a bit like comparing apples to oranges. So more rigorous comparison of the I-DNA and R-DNA network structures will require some additional logistics, such as breaking higher-way I-DNA associations to be broken down into two-way associations. To address this issue, we develop an additional method of network inference, called standardized bivariate transmission.

Standardized Bivariate Transmission

An I-DNA model can be simplified by constructing a \( g \times g \) bivariate transmission matrix. Here, higher-way associations are broken down into two-way associations (\( x_1 x_2 x_3 \) becoming \( x_1 x_2, x_1 x_3, \) and \( x_2 x_3 \)), and one-way associations are ‘built up’ into two-way associations (\( x_4 \): \( x_5 \): \( x_6 \) becoming \( x_4 x_5, x_4 x_6, \) and \( x_5 x_6 \)). To
create this matrix, ‘transmission’ (see Krippendorff, 1986) is calculated for each pair of network members to indicate the strength of their association as observed in the data. To calculate transmission, the probabilities underlying each pairwise association must first be drawn from the probability distributions of the data and the independence model. These two sets of probabilities must ignore all other variables, which is done by summing across their marginal probabilities.

The process of drawing probabilities for a pairwise association is illustrated below in Table 3.5, where the observed probability distribution $x_1 x_2 x_3$ is broken down into associations $x_1 x_2$, $x_1 x_3$, and $x_2 x_3$. Here, the original dataset contains 8 probabilities for the binary variables $x_1$, $x_2$, and $x_3$

$$(x_1^0 x_2^0 x_3^0, x_1^0 x_2^0 x_3^1, x_1^0 x_2^1 x_3^0, x_1^0 x_2^1 x_3^1, x_1^1 x_2^0 x_3^0, x_1^1 x_2^0 x_3^1, x_1^1 x_2^1 x_3^0, x_1^1 x_2^1 x_3^1)$$

shown in the leftmost column. But the probabilities for association $x_1 x_2$ can ignore variable $x_3$. So the probabilities $x_1^0 x_2^0 x_3^0$ and $x_1^0 x_2^0 x_3^1$ are summed to define the probability of $x_1^0 x_2^0$, the probabilities $x_1^0 x_2^1 x_3^0$ and $x_1^0 x_2^1 x_3^1$ are summed to define the probability of $x_1^0 x_2^1$, and so on. This same process can be used to identify all of the probabilities for associations $x_1 x_2$, $x_1 x_3$, and $x_2 x_3$. 
In a pairwise association where both variables have only 2 states (such as in Table 3.5 above), there will be four probabilities of interest from the observed (p) distribution, \( p(x_1^0 x_2^0), p(x_1^0 x_2^1), p(x_1^1 x_2^0), \) and \( p(x_1^1 x_2^1) \) and four probabilities of interest from the calculated (q) distribution of the independence model, \( q(x_1^0 x_2^0), q(x_1^0 x_2^1), q(x_1^1 x_2^0), \) and \( q(x_1^1 x_2^1) \). Once these (8) probabilities are gathered, the transmission for a pair of members \( (T_{ij}) \) is calculated as

\[
T_{ij} = \sum_i p(x_i x_j) \log \frac{p(x_i x_j)}{q(x_i x_j)}. \tag{18}
\]

Here, transmission is defined as a ratio of observed \( (p_{ij}) \) and independent \( (q_{ij}) \) probability distributions, as measured against the standard of the observed distribution \( (p_{ij}) \).

\[
T_{\text{max}} = \min \left\{ -\sum_i p(x_i) \log_2 p(x_i), -\sum_j p(x_j) \log_2 p(x_j) \right\}. \tag{19}
\]

The bivariate transmission matrix will use these standardized transmission values to populate the off-diagonal cells, with each transmission value indicating the strength of

<table>
<thead>
<tr>
<th>ABC Probabilities</th>
<th>AB Probabilities</th>
<th>AC Probabilities</th>
<th>BC Probabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x_1^0 x_2^0 x_3^0 )</td>
<td>.04</td>
<td>( x_1^0 x_2^0 )</td>
<td>.10</td>
</tr>
<tr>
<td>( x_1^0 x_2^0 x_3^1 )</td>
<td>.06</td>
<td>( x_1^0 x_2^1 )</td>
<td>.20</td>
</tr>
<tr>
<td>( x_1^0 x_2^1 x_3^0 )</td>
<td>.09</td>
<td>( x_1^1 x_2^0 )</td>
<td>.30</td>
</tr>
<tr>
<td>( x_1^1 x_2^1 x_3^0 )</td>
<td>.14</td>
<td>( x_1^1 x_2^1 )</td>
<td>.40</td>
</tr>
<tr>
<td>( x_1^1 x_2^1 x_3^1 )</td>
<td>.19</td>
<td>( x_1^1 x_2^1 )</td>
<td>.40</td>
</tr>
</tbody>
</table>
association for a pair of members. Transmission values are always positive and can range from 0 to the minimum entropy of $x_i$ or $x_j$, as shown below in equation 19. Higher values correspond to stronger associations, and they are always symmetric, with $T_{ij}$ being equivalent to $T_{ji}$. So, the transmission matrix will be symmetric with values of 1 along the diagonal to indicate perfect ‘dependence’ between a member and its own behavior.

After an I-DNA bivariate transmission matrix has been composed, two final adjustments must be made before comparing it with a R-DNA connectivity matrix. Transmission scores can range from 0 to $T_{max}$, while connectivity scores can range from $-1$ to 1. So, transmission scores must be standardized, dividing each by $T_{max}$, to ensure that they range between 0 and 1. Then, if the direction of the connectivity scores is not important, one can simply ignore them, comparing transmission values to the absolute values (i.e., magnitudes) of connectivity scores. If direction is important, it may be possible to augment transmission values with signs, based on whether each association indicates behavior patterns that are congruent (positive) or opposite (negative). However, this is not always possible when variables have more than 2 states, because an association might stem from an abundance of congruent and opposite behaviors, like $x_1^2x_2^2$ and $x_1^1x_2^3$. In that case, signs cannot be attributed to associations, and directions must be ignored in both I-DNA and R-DNA matrices.

Once the issue of directionality is resolved, an initial comparison of the I-DNA and R-DNA network structures is possible by comparing their $(g \times g)$ relationship matrices. Transmission and connectivity scores are likely to run somewhat parallel, with stronger I-DNA associations corresponding to stronger R-DNA connections. This is
because the present use of transmission (eqn 18, p.88) is mathematically equivalent to information captured, which is the nominal version of ‘variance explained’. A scatter plot can show the general relationship between transmission and connectivity scores, and their correlation can be calculated.

To the extent that they are correlated, I-DNA and R-DNA can be said to agree about underlying network structure. This indicates validity in their findings and suggests that the two methods offer comparable value for defining the network structure. However, there may be some interesting discrepancies between I-DNA and R-DNA matrices under some circumstances. A scatter plot may reveal bivariate outliers – pairs of members whose transmission and connectivity scores appear to disagree. These discrepancies could stem from I-DNA’s nominal treatment of continuous data or from R-DNA’s inability to detect nonlinear relationships. Further investigation may be fruitful with these outliers, as they may be unique in the behavior patterns they contain.

A closer comparison of I-DNA and R-DNA structures is also possible by filtering out some of the network relationships. Both the transmission and connectivity matrix will include a relationship weight for every pair of members, and very few of these will be 0, implying that virtually everyone is connected to everyone else at least a little bit. This is especially misleading for the I-DNA network, because it is unlikely for a best model to derive associations between all network members. To correct for this, associations can be deleted from the $g \times g$ transmission matrix if they are not present or nested in the I-DNA best model. Then, the same number of connections can be removed from the $g \times g$ connectivity matrix, starting with the lowest magnitudes and working upward.
This will result in two filtered matrices with the same number of pairwise relationships in both of them. The I-DNA matrix will contain all pairwise relationships contained in the best RA model, and an R-DNA network structure will reflect the same number of strongest relationships derived through linear modeling. These filtered network structures can be compared more thoroughly, and relationships may be found to be present in one network but missing from the other. This type of comparison can likely be facilitated with a second graph visualization, where relationships assume one of three colors, depending on whether they are present in the I-DNA network, the R-DNA network, or both.

**Summary**

The methods described in this chapter will now be used to address our primary methodological aim as well as our secondary substantive aim. In the chapters that follow, we will demonstrate that I-DNA and R-DNA provide complementary approaches to infer social network structures and test them for change over time. We’ll demonstrate applications of I-DNA and R-DNA on a dataset of Medicaid claims, by inferring the structure of a billing provider network before and after its formation as a Coordinated Care Organization (CCO). We will explore how that network of healthcare delivery changed following CCO formation, which will address our secondary substantive aim. And we’ll have the opportunity to compare I-DNA and R-DNA approaches on the types of networks they infer, the types of changes they identify, and their capacity to statistically detect network change over time.
Chapter 4: Coordinated Care Organizations

Overview

This dissertation applies I-DNA and R-DNA methods toward the analysis of a healthcare delivery network in the state of Oregon. More specifically, we focus on the billing provider network in Health Share of Oregon (http://healthshareoregon.org/), which is one of sixteen Coordinated Care organizations (CCOs) that have been created to fulfill a state legislated transformation in Oregon’s Medicaid health system (House Bill 3650, 2011; Senate Bill 1580, 2012). Legislative change came about primarily because of the rising cost of the State’s Medicaid program, the Oregon Health Plan (OHP). Medical inflation had been increasing at an average annual rate of 7.6% for the past two decades (CMS, 2012a), and following the economic recession of 2008, OHP enrollment increased by 56% to approximately 671,000 members (Goldberg, 2013). When CCO legislation passed, OHP had become the fastest growing portion of Oregon’s budget (Goldberg & Kaufmann, 2012), costing $4.5 billion in 2012 (CMS, 2013).

Despite historic growth in spending, OHP payments to providers had not kept up with the rising costs of medical care. Low rates of provider payment shifted a considerable amount of cost to commercial insurers and was understood to cause Medicaid patients difficulty in accessing care (Lewin Group, 2003). Beyond financial concerns, OHP enrollment policies were criticized for creating obstacles to coverage (Carlson, DeVoe, & Wright, 2006; Wallace, McConnell, Gallia, & Edlund, 2010) that disproportionately affected vulnerable populations (Solotaroff et al., 2005; B. Wright, Edlund, & Gallia, 2005), and the range of services covered by OHP was criticized as
inconsistent and confusing for those enrolled (Oberlander, Marmor, & Jacobs, 2001). Fragmentation between different sectors of care (medical, mental, and substance abuse treatment) was also well documented at the time (CMS, 2012a; Cutler, McFarland, & Winthrop, 1998; Labby, 2006; Laws, Gabriel, & McFarland, 2002).

In light of these many challenges, Oregon was approved for a Section 1115 Medicaid waiver, which allowed them to embark on a five-year health care transformation to improve OHP’s efficiency, value, and health outcomes (CMS, 2012a). This transformation centered on the formation of Coordinated Care Organizations (CCOs), which provided several fundamental changes to OHP (see McConnell et al., 2014 for a review). First, all health sectors – medical, mental, addictions, and dental – were financially integrated in each CCO, all functioning under a collective global budget. Before, managed care organizations (MCO’s) had been formed separately for each health sector, so capitated budgets were specific to each sector (e.g., medical or mental health services), creating fragmentation. Second, global budgets were allocated according to what the average cost per person ought to be (‘per-member-per-month’). Before the transformation, MCO capitated budgets had been based on the volume of health services provided, implicitly incentivizing service delivery irrespective of health outcomes.

A third major change with Oregon’s health care transformation is that the state started using a number of additional metrics and measurements to monitor care quality, access, and value (CMS, 2012a). This monitoring was designed to make CCO efficiency transparent and to incentivize ongoing transformation and performance improvement. Monitoring included things like the rate of re-hospitalization and the Consumer
Assessment of Healthcare Providers and Services (CAHPS), an indicator of member satisfaction (Oregon Health Authority (OHA), 2013b). CCOs were also required to facilitate the formation of community advisory councils and to involve their input in board decisions regarding quality improvement and elimination of health disparities. A growing portion of all CCO funding has been set aside each year in a state-level quality incentive pool, awarded to CCOs who achieved state goals (‘benchmarks’) for quality and value (OHA, 2013a).

Oregon’s health care transformation plan is designed to address the state’s “triple aim” of improving health outcomes, improving quality of care (i.e., member satisfaction), and reducing cost (OHA, 2012). To date, CCOs have been able to keep and often exceed their promises to curb cost increases per capita (Goldsmith & Henderson, 2017; Kaufmann, 2017), and have also been observed to positively impact patients’ experiences of access to care, quality of care, and better connections to primary care (Wright, Royal, Broffman, Li, & Dulacki, 2019). For example, they have been found to increase timely prenatal care initiation (Muoto et al., 2016), to reduce emergency department visits and preventable hospital admissions, and to improve some measures of access and appropriateness of care ( McConnell et al., 2017).

Presently there are sixteen CCOs in the state of Oregon, all functioning within the same requirements and benchmarks. The Oregon Health Authority has requires that all CCOs (a) establish a governance structure that ensures transparency, accountability, and responsiveness to community advisory councils, (b) establish protocols for engaging patients and their families in the design and implementation of culturally appropriate
care, (c) develop a strategy and a set of best practices for identifying and eliminating health disparities, (d) develop payment methodologies that incentivize care coordination and promote the triple aim, (e) expand health information technology to participate in the statewide health information exchange, and, perhaps most importantly, (f) ensure patients’ access to patient-centered, high quality, and coordinated care through a network of providers and home- and community-based services (OHA, 2012).

To satisfy this last requirement, CCO’s had to encourage clinics to become state certified as Patient-Centered Primary Care Homes (PCPCHs), and had to prioritize PCPCHs for the enrollment of new patients (OHA, 2012). PCPCH is a model of primary care that emphasizes prevention services, primary care, and care coordination, particularly for individuals with chronic conditions and complex health care needs (PCPCH Standards Advisory Committee, 2012). The core attributes of PCPCHs include accessibility to needed care, accountability for health outcomes, comprehensiveness of whole-person care, continuity of care over time, coordination and integration of the health care system, and person- and family-centered care. Under the PCPCH model, each patient is assigned a primary care physician (PCP), and encouraged to access care primarily through his or her PCP (OHA, 2013d). In Health Share’s CCO, patients’ access to specialty care must go through a PCP’s referral or approval (Health Share of Oregon, 2013).

CCOs across the state of Oregon have been very successful at enrolling Medicaid members (patients) at PCPCH clinics, including 92% of members statewide and 96% of those at Health Share in 2017 (Oregon Health Authority, 2019b). Medical home models
have been shown to improve health outcomes and patient satisfaction (Rosenthal, 2008) and may also reduce costs by preventing duplicated services (Davis et al., 2005). Under the PCPCH model, Medicaid patients can be expected to receive care from fewer providers. Patients’ PCP assignments remain consistent unless they move, are discharged from a PCP clinic, or request to switch their PCP (Health Share of Oregon, 2013). With this policy, people are less likely to simply visit whichever provider is available at the time when care is needed. And they will likely see fewer specialists under the PCPCH model, because their PCPs will be involved in those referrals.

In the current dissertation project, the implementation of the PCPCH model of care – and more specifically the assignment of patients to PCPs – was expected to lead to visible changes in Health Share’s network of care delivery in two ways. First, we expected that increased reliance on primary care and PCPs would make referral patterns more pronounced following Health Share’s formation, leading to increased levels of billing provider connectivity throughout the network. Second, we expected patients would move back and forth between primary and specialty providers less often and rely more predominantly on primary care than specialty care after Health Share’s formation. Outside of the PCPCH model, we also expected that Health Share’s integration of mental/behavioral services into its CCO would result in a stronger connection between mental/behavioral and primary care sectors of the network.

**Medicaid Claims Data**

The following sections of this chapter describe the acquisition and preparation of data resources used in the current dissertation project. The primary data were a set of
Medicaid claims from before and after Health Share of Oregon’s formation as a Coordinated Care Organization (CCO). Claims were acquired through the Center for Outcomes Research and Education (CORE) at Providence Health & Services. Secondary data pertaining to providers’ medical specialties and practice locations were downloaded from publicly available sources through the National Plan and Provider Enumeration System’s (NPPES) NPI Registry. These data resources were used to generate analytic datasets that allow for network inference (Chapter 5) and statistical testing (Chapter 6) by Information theory-based Differential Network Analysis (I-DNA) and Regression-based Differential Network Analysis (R-DNA). They are used for the methodological aim of this project, where we demonstrate and compare the usefulness of I-DNA and R-DNA for network inference and statistical detection of network change, as well as for the substantive aim of this project, where we learn how Health Share’s network of healthcare delivery may have changed after its formation as a CCO.

The data used for this work requires a substantial amount of introduction, which is why we have dedicated an entire chapter to it. First, we should define the notion of a billing provider, which is an individual or organization that submits claims to the Centers for Medicare and Medicaid Services (CMS), for reimbursement of services rendered to Medicaid patients. In the networks to follow, networks members (or ‘nodes’) will represent billing providers who operate within Health Share’s network of healthcare delivery. These billing providers should be differentiated from performing providers – the individuals who ‘render’ or provide medical services to a patient. The relationship between these two types of providers is illustrated with artificial entities in Figure 4.1 on
the following page. In that figure, a performing provider might sometimes have their own private practice and bill for themselves (like Dr. 10), or they might work for a clinic/hospital (like Drs. 1-9) that has one or more billing entities to handle billing.
Figure 4.1. Performing and Billing Providers in a Hypothetical Coordinated Care Organization
Figure 4.1 also illustrates several types of healthcare organizations and the types of relationships that are possible among them. Clinics and hospitals are often subsidiaries or partners of larger organizations that might provide medical care, insurance coverage, or both. In the Health Share of Oregon CCO, several of these larger types of organizations work together as ‘affiliates’ under the same global budget toward achievement of the Triple Aim.\textsuperscript{15} The formation of Health Share changed the legal relationship between these affiliates, financially integrating them so that they all began functioning under a collective global budget. This is likely to have changed patterns of healthcare delivery among them, potentially at every level in Figure 4.1. This project analyzes changes at the billing provider level, basically using that level to explore how Health Share’s formation affected patterns of healthcare delivery to Medicaid patients.

\textbf{Primary Data}

The primary dataset used for this dissertation research is a body of deidentified insurance claims records covering two time periods – one before and one after the formation of Health Share on September 1, 2012. The “before” dataset contains claims between May 1, 2011 and April 30, 2012. The “after” dataset contains claims between January 1, 2013 and December 31, 2013. We left a four-month period on either side of Health Share’s formation date (May 1 - December 31, 2012) where no claims were

\textsuperscript{15} Note that the names in Figure 4.1 are fictitious – Health Share’s actual affiliate organizations include Adventist Health, CareOregon, Central City Concern, Kaiser Permanente, Legacy Health, Oregon Health & Science University, Providence Health & Services, Tuality Healthcare and Clackamas, Multnomah, and Washington County Health Departments.
analyzed (see Figure 4.2). This window was designed to allow for a short transition period while changes were likely underway.

![Figure 4.2. Timeline of Datasets from Before and After Health Share’s Formation](image)

Insurance claims data were accessed on March 30, 2015 through the Center for Outcomes Research and Education (CORE) at Providence Health & Services, which is in contract with Health Share to provide analytic services (Moss Adams LLP, 2014). CORE receives insurance claim data quarterly, and the data accessed were understood to be a complete account of all Medicaid insurance claims for health care provided through any of Health Share’s 11 affiliate organizations (Health Share of Oregon, 2015). There were 5,602,445 medical claims across the two time periods, including 2,949,815 claims before Health Share’s formation, and 2,652,630 claims after Health Share’s formation. Only 69 claims (0.00001%) were missing provider information, all from 2013, and those were dropped from the study (leaving 5,602,376 claims for analysis).

As of September 2012, Health Share had over 200,000 patients (Medicaid members) and over 18,000 performing providers (Health Share of Oregon, 2012) in the tri-county area. Health Share’s full CCO includes 11 major affiliate organizations -
Advocate Health, CareOregon, Central City Concern, Kaiser Permanente, Legacy Health, Oregon Health and Science University, Providence, Tuality Healthcare, and the Multnomah, Washington, and Clackamas County Health Departments – who generate millions of insurance claims per year. These claims are filed in Oregon’s All Payer All Claims format (APAC, OHA, 2013c), which includes many lines of information regarding the patient, performing and billing providers, and the type of medical care provided, including diagnostic and treatment codes, date of treatment, and a variety of other indicators (Office for Oregon Health Policy and Research, 2013).

Researchers at Providence’s CORE reorganized claims data into a simplified claim-level format, and provided us with claims records containing four variables: a deidentified ID for each patient, the NPI (National Provider Identification) ID for the performing provider who rendered the service on the claim, the NPI for the billing provider who submitted the claim for reimbursement, and the time period in which the care was delivered (before versus after Health Share’s formation). Claims data only included medical claims – pharmaceutical claims were excluded because they would be more indicative of a pharmaceutical network than a medical care network. The majority of medical claims (84% or 4,684,368) were professional claims, for services rendered by non-institutional providers, such as physicians and suppliers (Hicks, 2015). A smaller portion of the claims (16% or 918,008) were facility medical claims, for services rendered by institutions, such as use of equipment, supplies, or laboratory or radiology services.
Claims data indicated over 180 thousand patients, over 20 thousand performing providers, and over 9 thousand billing providers across the two datasets (see Table 4.1). There was high turnover in the membership of each set, with only 61% of patients, 40% of performing providers, and 25% of billing providers present during both time periods. However, those who were present during both periods comprised most claims: 84% of Medicaid claims listed patients who were present during both periods; 87% listed performing providers who were present during both periods; and 93% listed billing providers who were present during both periods. This suggests that most medical care was delivered to a relatively consistent set of patients through a consistent set of performing providers and billed for by a consistent set of billing providers over time. The many patients and providers who were only present during one time period had very low levels of activity during that period.

Table 4.1. Distribution of Patients and Providers across time periods

<table>
<thead>
<tr>
<th></th>
<th>Before Only</th>
<th>After Only</th>
<th>In Both</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients</td>
<td>29,315 (16%)</td>
<td>42,082 (23%)</td>
<td>112,561 (61%)</td>
<td>183,958</td>
</tr>
<tr>
<td>Performing providers</td>
<td>5,299 (26%)</td>
<td>6,739 (33%)</td>
<td>8,180 (40%)</td>
<td>20,218</td>
</tr>
<tr>
<td>Billing providers</td>
<td>2,729 (29%)</td>
<td>4,290 (46%)</td>
<td>2,327 (25%)</td>
<td>9,346</td>
</tr>
<tr>
<td>Claims</td>
<td>2,949,815 (53%)</td>
<td>2,652,561 (47%)</td>
<td>-</td>
<td>5,602,376</td>
</tr>
<tr>
<td>Claims for patients from each period</td>
<td>446,609 (8%)</td>
<td>449,421 (8%)</td>
<td>4,706,346 (84%)</td>
<td>5,602,376</td>
</tr>
<tr>
<td>Claims for performing providers from each period</td>
<td>218,655 (4%)</td>
<td>498,645 (9%)</td>
<td>4,885,076 (87%)</td>
<td>5,602,376</td>
</tr>
<tr>
<td>Claims for billing providers from each period</td>
<td>166,307 (3%)</td>
<td>210,830 (4%)</td>
<td>5,225,239 (93%)</td>
<td>5,602,376</td>
</tr>
</tbody>
</table>

**Supplementary Data**

A second, supplementary, dataset used in this dissertation contained information about billing and performing providers, downloaded on June 21st, 2015 from the National
Plan and Provider Enumeration System’s (NPPES) NPI Registry (see http://nppes.viva-it.com/NPI_Files.html). All providers are required to register for a National Provider Identification (NPI) number in order to bill for reimbursement from CMS, and their NPI registration includes some basic information about their areas of medical specialty and their locations of medical practice. From this supplementary data, zip code information was used to determine which billing providers had practice locations within Health Share’s region of coverage. And specialty information was used to identify which performing providers were working in primary, specialty, ancillary, mental/behavioral, facility, and ‘other’ sectors of the healthcare delivery network.

An additional supplementary dataset was available, but not ultimately used for analysis, which indicated whether each deidentified patient had been flagged within the last 36 months as having one or more chronic medical conditions, such as asthma, chemical dependency, congestive heart failure, depression, diabetes, emphysema, hypertension, obesity, and psychological or affective disorders. The list of patients with one or more chronic conditions was at one point intended for use in follow-up analyses to see how the network of healthcare delivery changed particularly for this subset of individuals with higher than average medical care needs. Across both datasets (before and after Health Share’s formation), 39.2% of Medicaid patients (72,184) were flagged as having at least one chronic medical condition during at least one time period, and these patients accounted for 69.6% of Medicaid claims across the two time periods. 59.3% of patients (109,313) were not flagged for any chronic conditions during either of the time periods of interest, and information was not available regarding the presence or absence
of chronic medical conditions for the last 1.3% of patients (2,461). Future analysis can be done with these data, but these analyses were not done as part of this dissertation.

**Dataset Preparation**

Some preliminary work was necessary to prepare Medicaid claims data for analysis by I-DNA and R-DNA. First, the set of billing providers was filtered down from 9,346 to 1,299. Most of this filtration was done to satisfy requirements by I-DNA and R-DNA that any networks compared must contain the same network members (as ‘nodes’). So, 7,019 billing providers who were only active during one time period were excluded from analysis. Additional filtration was based on geography: Of the 2,327 billing providers who were active during both time periods, 1,029 (44%) of them had a practice location outside of Health Share’s tri-county region of coverage (Oregon Health Policy Board, n.d.), many of which were outside the State of Oregon (522; 24%). Billing providers from outside the region were grouped together as though they were a single provider, labeled “outside providers.” The remaining 1,298 billing providers from within the Health Share region, plus this outsider group, were used to infer Health Share’s billing networks from before and after its formation\(^\text{16}\).

Claims data were then restructured from CORE’s claim-level format to a ‘patient by billing provider format,’ as shown below in Table 4.2 and 4.3. This format involved a large cross tabulation, where each billing provider NPI constituted a variable, each patient ID constituted a case, and the values in each cell indicated the frequency with

---

\(^{16}\) Billing providers from within the tri-county area accounted for 88% of the total claim volume (4,942,146 claims).
which each billing provider billed claims for each patient. The rationale for using this patient-level format was to address an assumption by I-DNA and R-DNA of independent observations. Individual claims are not independent observations of a healthcare provider network, because patients often received multiple treatments related to the same health care concerns. By rolling up claims data to the patient level, we achieved a higher level of independence among cases, although we acknowledge that patients are also likely to have dependencies, such as biological relationships and geographic proximity, which we were not able to control for in the present study.

<table>
<thead>
<tr>
<th>Claim#</th>
<th>Patient</th>
<th>Provider</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>1234</td>
<td>Patient A</td>
<td>NPI 3</td>
<td></td>
</tr>
<tr>
<td>1235</td>
<td>Patient B</td>
<td>NPI 2</td>
<td>2012</td>
</tr>
<tr>
<td>1236</td>
<td>Patient C</td>
<td>NPI 1</td>
<td>2012</td>
</tr>
<tr>
<td>1237</td>
<td>Patient C</td>
<td>NPI 2</td>
<td>2013</td>
</tr>
<tr>
<td>1238</td>
<td>Patient D</td>
<td>NPI 1</td>
<td>2013</td>
</tr>
<tr>
<td>1239</td>
<td>Patient E</td>
<td>NPI 3</td>
<td>2013</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Table 4.2. Data in Claim Level Format

After reformatting claims data to patient by billing provider format (as shown above in Table 4.3), the data were ready for analysis by R-DNA, as will be described in Chapter 5. For I-DNA, however, billing provider variables were also discretized, or made nominal. Instead of cells indicating the frequency with which providers billed for patients (values ranging from 0 to ∞), cells were re-coded to indicate whether each billing provider billed for each patient ‘never’ (0), once (1), or multiple times (2). Justification for defining these states came from initial explorations of the data, which suggested that

<table>
<thead>
<tr>
<th>NPI 1</th>
<th>NPI 2</th>
<th>NPI 3</th>
<th>NPI 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient A</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Patient B</td>
<td>1</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>Patient C</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Patient C</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Patient D</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Patient E</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Table 4.3. Patient by Billing Provider Format
variables were highly skewed. The most common frequency by far was zero, because most patients were never billed for by most billing providers. The second most common frequency by far was one, since patients who were billed for by a given provider were most often billed for only once.

After these data preparations, claims data were ready to be analyzed by I-DNA and R-DNA. These analyses are described more extensively in Chapter 5, but a few examples are still useful here: In the simplest application of these methods, each pair of billing providers (each pair of columns in Table 4.3) was used to calculate a correlation \( r \), indicating the extent to which those billing providers were ‘connected’ before Health Share’s formation versus after it. In a more complicated application, each billing provider was used as an outcome variable in a multiple regression equation, with all other billing providers serving as predictor variables, so that regression weights \( B \) could indicate pairwise levels of ‘connectivity’ between the focal provider (dependent variable) and each other provider. And in the most complicated application, discretized billing provider variables were submitted to an information-theory based data mining software (OCCAM) to identify patterns of association amongst billing providers which were interpreted as multiway connections.  

The applications of I-DNA and R-DNA described above were first used to infer patterns of connectivity amongst the ‘full’ network of Health Share’s 1,299 billing providers, using the full dataset containing 1,299 columns – or one for each billing

---

\(^{17}\) Two other variations were also applied, including partial least squares (PLS) regression and standardized bivariate transmission (SBT) which are described more fully in Chapter 5.
Second, the same applications were used to infer and test patterns of connectivity amongst the ‘top 30’ billing providers by patient volume, using a subset of 30 columns from that dataset (see p.136 in Chapter 5 for more information on this subset). Third, we applied I-DNA and R-DNA methods to a ‘care sector network’ with six sectors of healthcare delivery: primary, ancillary, specialty, mental/behavioral, facility, or ‘other’ care. Applications of I-DNA and R-DNA to this third network required some additional preparatory steps, so that billing frequencies could be summed into six columns, one for each care sector.

Aggregation of billing patterns into one network member (‘node’) for each care sector required a three-step process. We first sorted medical specialties from the NUCC Health Care Provider Taxonomy into our six care sectors (see Appendix A), then sorted performing providers according to which types of medical specialties they had listed in the NPI registry, and finally sorted billing providers according to the type of performing provider that they predominantly billed claims with.\textsuperscript{18} Categorization of NUCC taxonomies underwent several iterations, as several categories were found to contain very few billing providers. Dental care was ultimately merged with ‘other’ care, and all facilities (inpatient, outpatient, ‘other’) were merged into a single category. After NUCC taxonomies were categorized, performing providers were sorted according to the ‘primary’ taxonomy listed in the NPPES database when downloaded on June 21\textsuperscript{st}, 2015. Nearly all performing providers (97.5\%, or 19,715) had a taxonomy flagged as the

\textsuperscript{18} Billing providers have NPI taxonomies as well, but we could not directly sort billing providers by their taxonomies because most are technically facilities even if they specialize in something like primary care.
primary one, or, if their primary taxonomy was not flagged, had a majority of taxonomies that fell into a single type of care (1.6%, 314).

Some minor obstacles were encountered when classifying performing providers. There were 118 performing providers who had been deactivated in the NPPES database by the time of the download, so their taxonomy information was not available. Those providers were not classified into any type of care and were excluded from billing provider classification. Fortunately, they represented a few claims (44,668, 0.8%) and very few Medicaid patients were seen exclusively by them during the study (766, 0.4%).

There were also some performing providers who did not have a primary taxonomy flagged and who, according to their taxonomies listed, appeared evenly split between multiple types of care (70, .3%). For them, a hierarchy was used to break ties, with facility classifications having priority, followed by specialty, primary, mental health, ancillary, and ‘other’ care (resulting in 3, 30, 19, 13, and 5 providers respectively). Specific rankings in this hierarchy followed some basic rationale: A provider with a facility specialization (e.g., a hospital) is not likely be an individual, so should be classified as a facility rather than any other type of provider. A provider with specialty qualifications should be called out separately from primary care, as they likely provide specialty services and not only primary care. We selected primary over mental/behavioral care because we wanted the mental/behavioral sector to be predominantly counselors and psychologists and not to contain providers with primary care qualifications (e.g., clinicians). Lowest priority was given to the classification of ancillary (e.g., lab technicians) and ‘other’ unclassified taxonomies.
Table 4.4 shows the most common taxonomies for each care sector, weighted by patient volume. For example, patients who received primary care were most likely to receive it from someone whose primary taxonomy was ‘Family Medicine Physician’.

<table>
<thead>
<tr>
<th></th>
<th>Most Common Primary Taxonomy (# Patients)</th>
<th>2nd Most Common (# Patients)</th>
<th>3rd Most Common (# Patients)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary</td>
<td>Family Medicine Physicians (116,837)</td>
<td>Internal Medicine Physicians (68,400)</td>
<td>Pediatric Physicians (65,892)</td>
</tr>
<tr>
<td>Ancillary</td>
<td>Clinical Medical Laboratories (43,598)</td>
<td>Optometrists (13,780)</td>
<td>Respiratory, Developmental, Rehabilitative and Restorative Physical Therapists (1,174)</td>
</tr>
<tr>
<td>Specialty</td>
<td>Diagnostic Radiology Physicians (139,473)</td>
<td>Emergency Medicine Physicians (95,155)</td>
<td>Anesthesiologist Physicians (23,286)</td>
</tr>
<tr>
<td>Mental</td>
<td>Behavioral Health &amp; Social Service Counselors (21,291)</td>
<td>Psychiatric Physicians (9,325)</td>
<td>Psychiatric/Mental Health Nurse Practitioners (6,503)</td>
</tr>
<tr>
<td>Other</td>
<td>Other Service Provider Specialists (21,567)</td>
<td>Diagnostic Radiology Physicians (8,761)</td>
<td>Gastroenterology Physicians (851)</td>
</tr>
<tr>
<td>Facility</td>
<td>General Acute Care Hospitals (87,497)</td>
<td>Emergency Medicine Physicians (21,589)</td>
<td>Durable Medical Equipment Suppliers (20,539)</td>
</tr>
</tbody>
</table>

After performing providers were classified into types of care, billing providers were classified according to the type of performing provider they billed the most patients with. This classification was done with performing provider data across both time periods, so billing providers were classified consistently throughout the full study period 2011-2013. Classifications were weighted by patient volume, so performing providers who billed for more patients under a billing provider had more impact on that billing provider’s classification. They were not weighted by claim volume, however, because some healthcare services require a combination of professional and facility claims, making the total number of claims misleading as to the amount of medical care that was rendered.
There were eight billing providers who had “ties” based on the type of performing provider that they billed the most patients with. These ties were resolved by finding the identities of the providers online. Four behavioral health centers were classified as mental health billers, three general clinics (Portland Clinic & two Providence clinics) were classified as primary care, and one chiropractic practice was classified as ancillary care. Note that we did not classify billing providers from outside of the Health Share region. And we did not retain patients who were served exclusively outside the region. This resulted in the exclusion of 1,810 (1.3%) patients from the before period and 1,511 (1.0%) patients from the after period because it was not clear which healthcare sectors they received care from.

Table 4.5 provides an overview of care sector classification, including the number of NUCC Taxonomies, performing providers, and billing providers that were categorized into each care sector. The number of billing providers in each care sector was relatively even. Primary and specialty sectors contained the most billing providers (358, 357), followed by ancillary (197), facility (178), and mental/behavioral (146). However, care sectors varied quite a bit in terms of how many performing providers operated within them. For example, ancillary care had only about 8 performing providers per billing provider, whereas primary, specialty, and mental health care sectors had more than twice the average number of performing providers per billing provider. This suggests that billing practices may be more streamlined (i.e., orchestrated by fewer billing providers) in the ancillary sector, and may be more decentralized (i.e., distributed amongst more billing providers) in the primary, specialty, and mental/behavioral sectors.
Table 4.5. Taxonomy, performing, and billing provider counts by care sector

<table>
<thead>
<tr>
<th>NUCC Taxonomies</th>
<th>Performing Providers</th>
<th>Billing Providers</th>
<th>Avg Performing Providers per Billing Provider</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary</td>
<td>114</td>
<td>7,102</td>
<td>358</td>
</tr>
<tr>
<td>Ancillary</td>
<td>159</td>
<td>1,571</td>
<td>197</td>
</tr>
<tr>
<td>Specialty</td>
<td>157</td>
<td>6,030</td>
<td>357</td>
</tr>
<tr>
<td>Mental</td>
<td>69</td>
<td>2,702</td>
<td>146</td>
</tr>
<tr>
<td>Other</td>
<td>179</td>
<td>658</td>
<td>62</td>
</tr>
<tr>
<td>Facility</td>
<td>178</td>
<td>2,035</td>
<td>178</td>
</tr>
<tr>
<td>Total</td>
<td>856</td>
<td>20,216</td>
<td>1,298</td>
</tr>
</tbody>
</table>

Distributions and Change Over Time

Prior to applications of I-DNA and R-DNA, we looked at distributions of claims, patients, and providers, and we compared them across time periods to identify global changes in care delivery that occurred after Health Share’s formation. In these comparisons, we included claims with billing providers who were inactive during one time period or who were located outside of the Health Share region. This allowed us to capture changes in rates more comprehensively, without omitting information that pertained to providers who became more active or less active over time.

All frequency distributions for claims, patients, and providers were highly positively skewed, in what looked approximately like negative exponential curves. This is an intuitive trend – most patients were listed on very few claims, most performing providers were listed with very few patients, most billing providers were listed with very few performing providers, etc. Figure 4.3 provides an example histogram that illustrates the negative exponential appearance of all of these distributions. This figure shows the number of billing providers per patient before Health Share’s formation. Patients were most likely to have been billed for by only one billing provider (31,065; 22%) and were less and less likely to have been billed for by more and more billing providers. Note that
this distribution looked essentially the same after Health Share’s formation, and distributions of claims and performing providers exhibited the same shape as well.

![Figure 4.3. Histogram depicting how many patients had each number of billing providers before Health Share’s Formation](image)

We conducted rate comparisons before and after Health Share’s formation, for example to see whether patients demonstrated substantial changes in the number of claims, performing providers, or billing providers involved in their care delivery after Health Share’s formation. Results from these comparisons are summarized below in Table 4.6.
Table 4.6. Rates of claims and providers, before and after Health Share’s formation

<table>
<thead>
<tr>
<th>Rates per patient</th>
<th>Mean Before (Std Dev)</th>
<th>Mean After (Std Dev)</th>
<th>%Δ in Means</th>
</tr>
</thead>
<tbody>
<tr>
<td>Claims</td>
<td>20.8 (41.2)</td>
<td>17.2 (34.4)</td>
<td>-17.3</td>
</tr>
<tr>
<td>Performing providers</td>
<td>6.4 (7.2)</td>
<td>6.5 (7.7)</td>
<td>1.6</td>
</tr>
<tr>
<td>Billing providers</td>
<td>4.5 (3.9)</td>
<td>4.4 (3.9)</td>
<td>-2.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rates per performing provider</th>
<th>Mean Before (Std Dev)</th>
<th>Mean After (Std Dev)</th>
<th>%Δ in Means</th>
</tr>
</thead>
<tbody>
<tr>
<td>Claims</td>
<td>179.2 (2,160.0)</td>
<td>177.8 (1779.1)</td>
<td>-0.8</td>
</tr>
<tr>
<td>Patients</td>
<td>60.5 (333.2)</td>
<td>67.7 (391.3)</td>
<td>11.9</td>
</tr>
<tr>
<td>Billing providers</td>
<td>1.56 (1.4)</td>
<td>1.61 (1.4)</td>
<td>3.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rates per billing provider</th>
<th>Mean Before (Std Dev)</th>
<th>Mean After (Std Dev)</th>
<th>%Δ in Means</th>
</tr>
</thead>
<tbody>
<tr>
<td>Claims</td>
<td>2,297.4 (11,279.5)</td>
<td>2,069.1 (10,122.6)</td>
<td>-9.9</td>
</tr>
<tr>
<td>Patients</td>
<td>529.5 (2,386.5)</td>
<td>480.8 (2,158.8)</td>
<td>-9.2</td>
</tr>
<tr>
<td>Performing providers</td>
<td>23.9 (224.7)</td>
<td>23.0 (211.3)</td>
<td>-3.8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Total Counts</th>
<th>Mean Before</th>
<th>Mean After</th>
<th>%Δ in Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Claims</td>
<td>2,949,815</td>
<td>2,652,561</td>
<td>-10.6</td>
</tr>
<tr>
<td>Patients</td>
<td>141,876</td>
<td>154,543</td>
<td>4.4</td>
</tr>
<tr>
<td>Performing providers</td>
<td>13,478</td>
<td>14,918</td>
<td>12.9</td>
</tr>
<tr>
<td>Billing providers</td>
<td>5,056</td>
<td>6,617</td>
<td>30.9</td>
</tr>
</tbody>
</table>

* Note counts of claims, patients, performing and billing providers in all comparisons were limited to those with billing providers who active during both study periods (before and after) and located within Health Share’s region of care delivery.

The largest relative change observed in a rate was a decrease in the number of claims per patient, which decreased by 17.3% after Health Share’s formation. This change went in the same direction as % change in the total number of claims, which decreased by 10.6%, and was likely caused by a shift from the fee-for-service model during the before period to the CCO’s pay-for-performance model during the after period (Oregon Health Authority, 2019a). Once the CCO formed, billing providers did not have as many incentives to capture every single medical service as a distinct claim. So, it makes sense that the total number of claims would reduce over time.

The most peculiar change was an 11.9% increase in the number of patients per performing provider. This was difficult to make sense of, because the total number of performing providers increased by 12.9% after Health Share’s formation, and the total
number of claims decreased by 10.6%. We might have expected, then, that patients accessing less care (fewer claims) from a larger set of performing providers would result in fewer patients per performing provider on average. However, joint increases in the mean and standard deviation suggests that at least some performing providers demonstrated an increase in the number of patients they served. This foreshadows a main finding in Chapter 5, where billing patterns appear to consolidate amongst a subset of core providers who were more strongly connected after Health Share’s formation.

Following our comparisons of rates, we made some additional comparisons by care sector to identify any major shifts in utilization patterns. Table 4.7 below shows counts of patients served and claims billed by each sector, both before and after Health Share’s formation. The rank ordering of care sectors remained the same during both time periods – patients were most often billed by primary care providers, followed by facility, specialty, ancillary, mental, and ‘other’ providers, and claims were most often billed by facility providers, followed by primary, mental, specialty, ancillary, and other providers. But there were some differential changes in volume over time, with increases in ancillary care and decreases in specialty, facility, and other care, both by counts of patients and claims.

Table 4.7. Counts of Patients and Claims by Care Sector

<table>
<thead>
<tr>
<th>Patient Volume by Sector</th>
<th>Claim Volume by Sector</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before</td>
<td>After</td>
</tr>
<tr>
<td>Primary</td>
<td>116,350</td>
</tr>
<tr>
<td>Ancillary</td>
<td>31,109</td>
</tr>
<tr>
<td>Specialty</td>
<td>78,230</td>
</tr>
<tr>
<td>Mental</td>
<td>21,024</td>
</tr>
<tr>
<td>Other</td>
<td>10,785</td>
</tr>
<tr>
<td>Facility</td>
<td>82,765</td>
</tr>
<tr>
<td>Total</td>
<td>138,080</td>
</tr>
</tbody>
</table>
Several of the sector-specific trends in Table 4.7 agreed with our general expectations for CCO formation. Health Share’s implementation of a Patient-Centered Primary Care Home (PCPCH) model of care designated patients with a primary care provider (PCP) and required them to get a PCP referral in order to access specialty care services. So a decrease in specialty care utilization was as expected, and in fact 6.90% fewer patients accessed specialty care after Health Share’s formation. Health share’s incorporation of mental health services was not expected to change rates of care in that sector, and this rate increased by only about as much as one might have expected from Health Share’s overall rate of growth. It showed a 4.88% increase in mental health patients, as compared to a 4.38% increase in Health Share’s overall patient volume.

However, Table 4.7 also shows multiple trends that differed from expectations. The number of patients using primary care services basically did not change (increasing by only 0.34%), despite Health Share’s overall patient volume increasing by 4.38% and despite their implementation of the PCPCH model which should have encouraged patients to increasingly rely on primary care services over specialty care. This could be interpreted optimistically or pessimistically. Optimistically, Health Share’s encouragement of clinics to become PCPCHs could mean that patients started being able to access primary care services (and even visit their PCPs) within other sectors of the health care network and without needing to transfer to the primary sector. Pessimistically, however, this could indicate that patients encountered obstacles in meeting with their PCP or in gaining primary care services.
Finally, we examined changes in patients’ average number of primary care and specialty care billing providers. As expected, patients had fewer primary care billing providers and fewer specialty care billing providers after Health Share Formed. The average number of primary care providers dropped from 1.312 to 1.178 (a 10% decrease, with standard deviations of 1.028 and 0.930, respectively). And the average number of specialty care providers dropped from 1.163 to 1.021 (a 12% decrease, with SDs of 1.501, 1.426). These trends held true for subsets of patients with at least one claim of each type. So, among primary care patients, the average number of primary care providers dropped from 1.558 to 1.453 (by 7%, with SDs of 0.934, 0.817), and among specialty care patients, the average number of specialty providers dropped from 1.495 to 1.385 (by 7%, with SDs of 1.467, 1.417). These results were not statistically tested, but their direction of change provides loose support these trends moving in the anticipated direction after Health Share’s formation.

**Summary**

This chapter has described how Medicaid claims data were prepared into analytic datasets for analysis by I-DNA and R-DNA. Perhaps not surprisingly, there were elements of ‘messiness’ in these data resources, such as the high rates of turnover in the patients and providers during between time periods. And supplemental data sources were required to identify the billing providers who were operating within Health Share’s region of coverage and to categorize providers into sectors of care. However, we ultimately found claims to be amenable for network inference and statistical analysis. Analyses in the following chapters will utilize these claims data to develop
methodological contributions to the field of social network analysis and to provide insights regarding changes that occurred in Health Share’s network of care delivery.
Chapter 5: Network Inference Results

Overview

This chapter describes the application of five methods of network inference to insurance claim data. There are both substantive and methodological motivations for this. Substantively, each method allows us to infer the structure of a healthcare delivery network during two time periods, and to consider how it may have changed following Health Share’s CCO formation. Arguably, we can gain more confidence by applying multiple methods to the same datasets and determining what patterns ring true across them. These substantive motivations are complemented by methodological ones. Methodologically, the comparative application of multiple methods to the same datasets allows us to compare and contrast each method’s usefulness for network inference. In cases where methods produce differing results, we can consider the types of patterns that each is more and less able to capture in the form of a network structure.

To address both of the motivations above, three CCO networks are inferred in this chapter: a full network, a top 30 network, and a care sector network. Note that network inference is only the first step in conducting Information theory based Differential Network Analysis (I-DNA) or Regression based Differential Network Analysis (R-DNA), which culminates in a statistical test of change or difference. Following this chapter’s discussion of change in three CCO networks, chapter 6 will describe statistical tests of the significance of that change.
The first and largest CCO network contains 1,299 members, including 1,298 billing providers from within Health Share’s region, plus 1 member combining all billing providers from outside the region (aggregating their billing patterns together as though they were one single billing provider). The main hypothesis for the full network was that connectivity would increase over time. The PCPCH model of care was expected to encourage referrals and make billing patterns more pronounced. So, providers who previously saw similar sets of patients were expected to see even more similar patients after Health Share’s formation, and providers who previously saw fairly opposite sets of patients were expected to see even more opposite sets of patients.

The full network was inferred by only two methods, correlation and standardized bivariate transmission (SBT), which are the simplest forms of network inference for R-DNA and I-DNA, respectively. Due to computational limitations, the full network could not be inferred with other, more complex methods. Neither correlation nor SBT provided support for the hypothesis of increased connectivity, and in fact both methods showed small reductions in average connection strength after Health Share’s formation. But the methods also demonstrated a core-periphery network structure, with many members in the periphery who were largely inactive, sometimes billing for only one or two patients.

For more comprehensive analysis on the ‘core’ of the billing provider network, five methods of network inference were applied to the top 30 billing providers by patient volume. These top 30 billing providers were listed on claims with 96.7% of Health Share’s patients and accounted for 88.2% of all claims across the two periods. So, the top 30 network captured the bulk of Medicaid billing and was also small enough to be
analyzed with several inference methods for R-DNA (correlation, multiple regression, and partial least squares regression) and I-DNA (SBT and RA).

The hypothesis for the top 30 network was the same as the hypothesis for the full network – that connectivity would increase over time. Methods differed somewhat in their findings, but in general complemented one another to suggest that connectivity did not increase so much as it consolidated over time. After Health Share’s formation, connectivity strengthened within a subset of core connections and simultaneously weakened throughout the periphery of the top 30 network. RA and PLS regression methods were found to be the best at detecting change in this network by I-DNA and R-DNA, respectively, though it is important to note that PLS was only slightly more effective at detecting change than multiple regression.

The third network was defined around six care sectors, with nodes to represent billing providers in primary, ancillary, specialty, mental/behavioral health, facility, and ‘other’ sectors of the healthcare system. Billing providers were classified into these categories according to the type of performing provider they most often billed claims with,¹⁹ and performing providers were classified by their taxonomies as listed in the Centers for Medicare and Medicaid Services (CMS) National Plan and Provider Enumeration System (NPPES; see Appendix A). Once classified, billing providers in each sector were merged together, by summing patient-level billing patterns across all providers from that sector into a single network node.

¹⁹ Billing providers were assigned to care sectors by whichever type of performing provider was listed with them for the largest number of patients (not claims).
The care sector network was analyzed with four inference methods, based on findings from analysis of the top 30 network which indicated two methods as being most effective at finding network change (PLS and RA) and two additional methods as being useful for contextualizing the nature of that change (correlation and SBT). There were two hypotheses for the care sector network, both of which had very limited support. First, the connection between specialty and primary care was expected to weaken after Health Share formed, as the PCPCH model was expected to reduce primary care patients’ use of specialty care services. Primary care patients did use specialty care less by some measures, but a more predominant pattern was an increase in the number of people who did not receive either primary or specialty care, thus raising a question of whether folks were receiving primary and specialty care services within other care sectors – meaning they did not need care under the purview of a primary or specialty billing provider – or whether patients may have faced increased obstacles to accessing care.

The second hypothesis for the care sector network was that the connection between mental/behavioral and primary care would strengthen after Health Share formed. Inclusion of the mental/behavioral sector in Health Share’s CCO was expected to increase referrals from mental health to primary care providers. This appears to have happened for a subset of Health Share’s patients, as detected by PLS. But in general terms, Health Share’s formation was actually followed by a smaller proportion of mental health patients receiving care through the primary care sector, suggesting that referrals may not have happened as much as was expected. Hopefully this pattern indicates that patients were receiving some basic primary care through the mental/behavioral sector.
The comparative application of five inference methods highlighted several differences between them, most notably in their treatment of higher-way (e.g., three-way, four-way) behavior patterns. These were found to be pervasive in the top 30 and care sector networks during both time periods, as RA demonstrated in hyper-graph representations that were often difficult to interpret. Simple pairwise methods, like correlation and SBT, were essentially naïve to higher-way patterns, evaluating each pair without any regard for other providers in the network. And multiple regression and PLS excluded all higher-way patterns from pairwise representations by controlling for the behavior of other providers. Several examples from the top 30 analysis are used to illustrate pros and cons of each inference method, and to consider the conditions under which each might be optimal to use for network analysis research.

Network 1: The Full Network of Health Care Delivery

Health Share’s full billing provider network is shown below in Figures 5.1 through 5.4. This network was inferred during each time period for all billing providers who were active (billing at least one claim) during both time periods. There are 1,299 members in this network, including 1,298 billing providers whose office locations were within Health Share’s service region, plus 1 node representing all of the providers from outside the region. The nodes (i.e., billing providers) are located in the same place in each of the following figures, so network structures can be directly compared across time periods and across methods. Our hypothesis for this full network was that overall connectivity would increase after Health Share’s formation, including the formation of new connections and the strengthening of old connections over time. The full network
was first inferred with correlation, the simplest form of R-DNA, and then with SBT, the simplest form of I-DNA.

Figure 5.1 and Figure 5.2 depict the healthcare delivery network as inferred by correlation. Figure 5.1 shows what the network looked like before Health Share’s formation, and Figure 5.2 shows the network after Health Share’s formation. During both time periods, the absolute values of nearly all correlations among billing providers (>99.8%) were very small, between 0 and .1. A threshold of $|r| \geq .1$ was selected (arbitrarily) so that these weak connections could be filtered out for visualization purposes. The resultant figures (5.1 and 5.2) only contain 1,295 connections before Health Share formed, and 1,157 connections after, with the thickness of each connection corresponding to the absolute value of the correlation between each pair of providers. Figure 5.1 and Figure 5.2 omit about 500 billing providers who were not connected to anyone when the threshold was set to $|r| \geq .1$, including 505 billing providers in the before period and 516 billing providers in the after period. Nearly two thirds of the providers were ever disconnected by this threshold (313) were disconnected during both periods.

---

20 Spearman correlations were chosen over Pearson because the frequency data were highly skewed and thought to be better characterized as ordinal variables.
21 When rounded to 5 decimal points, there were only 26 edges in the before network with a weight of zero, and only 40 edges in the after network with a weight of zero.
Figure 5.1. Full network of 1299 billing providers by correlation, before Health Share's formation

Figure 5.2. Full network of 1299 billing providers by correlation, after Health Share's formation
Although connections and providers were filtered out of visualizations for the sake of interpretability, all providers and edges were included when measuring change in the network. On average, correlations among the 1,299 members decreased and correlation strengths weakened, indicating less connectivity after Health Share formed, counter to the main hypothesis. A decrease in correlations is evident by the change in raw correlation values (on a scale of +1 to -1), which changed on average by -0.0002. This was calculated by the following equation,

\[
\frac{1}{g(g-1)} \sum_{i \neq j} \hat{r}_{ij}^2 - \hat{r}_{ij}^1
\]

where \(\hat{r}_{ij}^1\) and \(\hat{r}_{ij}^2\) represent the correlation between nodes \(i\) and \(j\) during the first and second time periods (before and after Health Share formed), and the term \(g(g - 1)\) is the total number of members in the network (1,299). Correlation strengths – which are equivalent to the absolute values of correlations – also weakened over time, on average by 0.0003, providing fairly direct evidence that Health Share’s formation was followed by an overall decrease in connectivity. The decrease in correlation strengths is calculated by this equation below, which is identical to the equation above except for using absolute values of each correlation.

\[
\frac{1}{g(g-1)} \sum_{i \neq j} |\hat{r}_{ij}^2| - |\hat{r}_{ij}^1|
\]

A third variant of this equation is the mean absolute difference in correlations \((\bar{\theta}_{cor})\), which effectively aggregates of all amounts of change in either direction. This value was 0.0019, and can be interpreted on a scale of 0 to 2. (Theoretically, all correlations could have shifted from +1 to -1, or from -1 to +1, in both cases having an
absolute movement equal to 2). This mean absolute difference does not directly relate to our hypothesis, but it is the most inclusive way to aggregate network change over time. The MDA is the value that which could be statistical tested.

\[
\hat{\theta}_{\text{cor}} = \frac{1}{g(g-1)} \sum_{i \neq j} |\hat{r}_{ij}^2 - \hat{r}_{ij}^1|
\]  

(22)

All three the measures of change were relatively small, indicating that connectivity in the full network did not change very much when inferred by correlation. But the change that did occur went opposite the hypothesized direction, with connectivity decreasing over time. (This will be discussed further below.)

The full network of 1,299 members was also inferred with standardized bivariate transmission (SBT), as shown in Figure 5.3 and Figure 5.4. For this analysis, billing provider variables were converted from frequency variables (which list the exact number of claims billed for each patient by each provider) to nominal variables indicating whether a provider had billed for a patient zero times, one time, or two-or-more times during that period. Transmission was then calculated as a ratio of observed \((p(x_i, x_j))\) and independent \((q(x_i, x_j))\) probability distributions, weighted by the former, with the following equation.

\[
T_{ij} = \sum p(x_i, x_j) \log \frac{p(x_i, x_j)}{q(x_i, x_j)}
\]  

(23)

Here \(p(x_i, x_j)\) represents the set of observed joint probabilities for providers \(i\) and \(j\), for example the proportion of times patients were billed 0 times by provider \(i\) and 1 time by provider \(j\), or 1 time by provider \(i\) and 2 or more times by provider \(j\). Observed
joint probabilities $p(x_i, x_j)$ are compared against $q(x_i, x_j)$, the set of joint probabilities that one would expect if providers $i$ and $j$ were independent, or if $q(x_i, x_j)$ were to equal $p(x_i)p(x_j)$.

Once transmission is calculated, it is standardized to fall between 0 and 1 by dividing it by the maximum possible transmission,

$$T_{\text{max}} = \min\left\{ -\sum_i p(x_i) \log_2 p(x_i) , -\sum_j p(x_j) \log_2 p(x_j) \right\}$$

(24)

The maximum possible transmission is equivalent to the minimum entropy of either variable. This standardization allows all SBT connection weights in the network to be on the same scale and comparable with one another.
Figure 5.3. Full network of 1299 billing providers by SBT, before Health Share’s formation

Figure 5.4. Full network of 1299 billing providers by SBT, after Health Share’s formation
SBT connection weights during both time periods were much smaller than weights in the correlation network, so the minimum threshold for standardized transmission was set at $T_{std} \geq 0.00025$. This retained about the same number of edges in the transmission network as were retained in the correlation network at $r \geq 1$, excluding approximately 98% of edges during both periods\(^22\) and leaving only 1,477 edges visible before and 1,216 edges visible after Health Share’s formation. The threshold of 0.00025 also left many disconnected members hidden from view during each time period, including 858 members before and 890 after.

Although connections and providers were filtered out of visualizations for the sake of visual interpretability, all providers and edges were included when measuring change in the network. Like the correlation network, SBT values decreased after Health Share formed, suggesting an overall decrease in connectivity which ran counter to our hypothesis. Also like the correlation network, the amount of change was very small. The average change in SBT was only $-5 \times 10^{-7}$ as calculated by the following equation,

\[
\frac{1}{g(g - 1)} \sum_{i \neq j} SBT_{ij}^2 - SBT_{ij}^1
\]

where $SBT_{ij}^1$ and $SBT_{ij}^2$ represent SBT between nodes $i$ and $j$ during the first and second time periods (before and after Health Share formed), and the term $g(g - 1)$ is the total number of members in the network (1,299). Note that SBT values are inherently positive,

\(^{22}\)The SBT threshold led to the exclusion of 70,039 connections before Health Share’s formation (97.9%) and 61,241 connections after (98.1%). In addition, all pairs of providers with SBT < 0.00001 were treated as being not connected (or having a connection weight of zero). This included 771,535 (91.5%) pairs of providers before Health Share’s formation, and 780,584 (92.6%) pairs of providers after.
so there is no distinction between an average change in their raw or absolute values. A reduction in overall SBT values is equivalent to a reduction in the strength, and it suggests that connectivity decreased (and weakened) after Health Share’s formation, counter to the hypothesized direction of change, if only by a miniscule amount.

The equation for the mean absolute difference in SBT values is analogous to that for correlation, aggregating all amounts of change in either direction. The mean absolute difference for the full network by SBT was 0.0007, which is the value that could be statistically tested.

\[
\hat{\theta}_{SBT} = \frac{1}{g(g - 1)} \sum_{i \neq j} |SBT_{ij}^2 - SBT_{ij}^1|
\]  

Counts of connections and calculations of change in connection weights are summarized below in Table 5.1. Ultimately, both methods of inference found the average connection weight to decrease (weaken) after Health Share’s formation, and both methods also found fewer connections to fall above our visualization threshold after Health Share’s formation. This indicates agreement between these two methods that connectivity decreased over time, counter to the substantive hypothesis.
Table 5.1. Summary of Changes in Full Network Visualizations

<table>
<thead>
<tr>
<th>Method</th>
<th>Correlation</th>
<th>SBT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Threshold for visualization</td>
<td>0.1</td>
<td>0.00025</td>
</tr>
<tr>
<td>All connections found Before</td>
<td>1,295</td>
<td>1,477</td>
</tr>
<tr>
<td>All connections found After</td>
<td>1,157</td>
<td>1,216</td>
</tr>
<tr>
<td>Total edges in either period</td>
<td>1,878</td>
<td>1,812</td>
</tr>
<tr>
<td>Connections found in both periods</td>
<td>574 (31%)</td>
<td>881 (49%)</td>
</tr>
<tr>
<td>Connections found in one period</td>
<td>1,304 (69%)</td>
<td>931 (51%)</td>
</tr>
<tr>
<td>Connections found only before</td>
<td>721 (38%)</td>
<td>596 (33%)</td>
</tr>
<tr>
<td>Connections found only after</td>
<td>583 (31%)</td>
<td>335 (18%)</td>
</tr>
<tr>
<td>Changes in average connectivity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean change in connection weight</td>
<td>-0.0002 (eqn 20)</td>
<td>-5×10^-7 (eqn 25)</td>
</tr>
<tr>
<td>Mean change in absolute connection weight</td>
<td>0.0003 (eqn 21)</td>
<td>-5×10^-7 (eqn 25)</td>
</tr>
<tr>
<td>Mean absolute difference (θ)</td>
<td>0.0019 (eqn 22)</td>
<td>0.0007 (eqn 26)</td>
</tr>
</tbody>
</table>

When comparing the images for correlation and SBT, the correlation network appears to have more connections of high and very high strength. This is driven by what one might argue are spurious correlations among largely inactive billing providers. For example, in the before network there were 74 connection weights of \( r = 1.0 \) that were all caused by pairs of billing providers who each billed for only one patient but happened to have that one patient in common. Less extreme cases were also found throughout the correlation network, where providers had very strong connections despite only billing for several patients before or after Health Share’s formation.

The SBT inference method has protection against these arguably spurious connections, because transmission values are standardized (and effectively scaled) as a proportion of the maximum possible transmission (\( T_{\text{max}} \)). The maximum possible transmission is inherently large when a provider is mostly inactive,\(^{23}\) which ensures that SBT will be small. This scaling factor seems likely related to the most obvious

\(^{23}\) When providers are largely inactive they have a very low (Shannon) entropy, meaning a very uneven split in the proportion of patients they do and do not bill for. In information theory, a low entropy produces a large maximum possible transmission (\( T_{\text{max}} \)), ensuring that the SBT calculation (\( T/T_{\text{max}} \)) will be small.
differences in appearance between the correlation and SBT network figures. Most SBT connections appear thinner and fainter than correlation connections, whereas most correlation connections appear relatively thick. This is observable despite retaining a similar number of connections in SBT and correlation visualizations and despite scaling the visualized thickness to range between the heaviest and weakest connections included.

The visible differences between SBT and correlation networks illustrate the severity by which SBT values were scaled. The SBT network has fewer strong connections and has a wider gap between the strength of connections in its core and the weakness of the connections in its periphery. To illustrate this, note that all visualized correlations were at least 10% as strong as the strongest correlation (\(1 \leq r \leq 1.0\)), but only 2% of visualized SBT connections were at least 10% as strong as the strongest SBT. Similarly, the correlation images had only about 500 providers (40%) who were completely disconnected (and hidden from view), compared to over 800 disconnected providers (65%) in the SBT network. The vast majority of SBT edges were immensely weak compared to the strongest SBT, primarily due to scaling them by providers’ activity levels.

Depending on the research question, one may or may not find this scaling effect helpful. On the one hand, it may seem perfectly defensible to ignore connections that feel blatantly ‘spurious’ – where two providers only billed for one patient but happened to have that patient in common and thus achieved a perfect correlation despite having very little overlap in care delivery. But less extreme examples may be harder to exclude, for example if two providers billed for something like 200 patients in a year and had 150 of
them in common. This pattern feels much more relevant to care delivery, and an SBT network would downplay it as a weak connection whereas a correlation network would emphasize it as a stronger one.

In our case, we benefit from applying both methods to the same dataset and finding agreement in results by both types of inference. The full network appears to have become less connected after Health Share’s formation, whether it was inferred with correlations from claim frequencies or inferred with information theory from nominalized (binned) levels. By both methods, connectivity was relatively stable with a small general shift toward disconnection. It does not appear to matter whether you nominalize claim frequencies or whether you scale connection weights according to providers’ activity levels. The results refute our hypothesis either way.

When using I-DNA or R-DNA to detect change over time, there are two major obstacles to studying the full network of 1,299 providers. First, systematic or meaningful changes over time are not visually obvious, because the areas of the network with strongest connectivity look approximately the same during both time periods. Change may be especially problematic in the correlational network, because many of the most dramatic changes in connections occurred among nearly inactive billing providers, whose changes are not especially meaningful.

There are also computational limitations to measuring change in the full network over time. Although correlation matrices were calculated in minutes, and bivariate transmission matrices were calculated in just a couple of days, multiple regression, PLS regression, and reconstructability analysis are all very computationally intensive for a
network of this size. Network inference with multiple regression, for example, would require calculation of 1,299 equations, each with 1,298 predictor variables, for each of two time periods. Reconstructability analysis would require searching through a state space with hundreds of thousands of cells. These obstacles to network inference make the goal of statistical testing with I-DNA or R-DNA even more daunting, as that would require this large amount of effort for each of many pairs of resampled datasets.

Network 2: The Top 30 Billing Providers

To better observe systematic changes in Health Share’s billing provider network, more in-depth analyses were conducted on the inner core of the network, defined as the top 30 billing providers by patient volume. These top 30 providers accounted for a large bulk of claim activity over the study period, generating 88.2% (4,942,146) of all claims and billing for 96.7% (177,834) of Health Share’s patients. On average, these providers billed about 73,000 claims, with about 306 performing providers, and for about 16,765 patients during the study. These providers are listed below in Table 5.2, in descending order by patient volume. Their organizational names have been shortened for brevity, and duplicate shortened names have been numbered (e.g., Providence 1, Providence 2, etc.).

---

24 A trial run for RA on the full dataset, which was limited in scope to identify only one pairwise relationship, still took more than 45 hours to complete with OCCAM software.
Table 5.2. Top 30 billing providers by patient volume

<table>
<thead>
<tr>
<th>NPI</th>
<th>Name (# patients billed for)</th>
<th>NPI</th>
<th>Name (# patients billed for)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1003991845</td>
<td>Providence 1 (39,365)</td>
<td>1215910302</td>
<td>Adventist 1 (13,108)</td>
</tr>
<tr>
<td>1649339581</td>
<td>Multnomah County (38,534)</td>
<td>1265417174</td>
<td>Oregon Clinic (12,573)</td>
</tr>
<tr>
<td>1609824010</td>
<td>OHSU 2 (30,268)</td>
<td>1477527786</td>
<td>Oregon Anesthesiology (12,136)</td>
</tr>
<tr>
<td>1215983200</td>
<td>Northwest Acute (29,520)</td>
<td>1063648665</td>
<td>CEP America (11,812)</td>
</tr>
<tr>
<td>1376709535</td>
<td>OHSU 1 (26,825)</td>
<td>1023050457</td>
<td>Radiology Consultants (11,395)</td>
</tr>
<tr>
<td>1831112358</td>
<td>Legacy Emanuel 2 (23,734)</td>
<td>1447200167</td>
<td>AMR NW (11,185)</td>
</tr>
<tr>
<td>1699759498</td>
<td>Radiology Specialists (21,296)</td>
<td>1902827272</td>
<td>Legacy Clinics (10,854)</td>
</tr>
<tr>
<td>1114015971</td>
<td>Providence 3 (20,615)</td>
<td>1467771121</td>
<td>TRG LLC (10,491)</td>
</tr>
<tr>
<td>1114963956</td>
<td>Quest Diagnostics (19,138)</td>
<td>1467408013</td>
<td>Lifeworks NW (10,341)</td>
</tr>
<tr>
<td>1275591984</td>
<td>Tuality (18,331)</td>
<td>1679528012</td>
<td>Virginia Garcia (10,213)</td>
</tr>
<tr>
<td>1801887658</td>
<td>Adventist 2 (17,677)</td>
<td>1366536963</td>
<td>Providence 4 (10,193)</td>
</tr>
<tr>
<td>1083714141</td>
<td>Providence 2 (16,141)</td>
<td>1184786527</td>
<td>Kaiser (9,917)</td>
</tr>
<tr>
<td>1902059538</td>
<td>Legacy Lab Services (16,000)</td>
<td>1700983319</td>
<td>Epic Imaging (9,495)</td>
</tr>
<tr>
<td>1255354700</td>
<td>Legacy Mt Hood (13,671)</td>
<td>1700809027</td>
<td>Legacy Emanuel 1 (8,995)</td>
</tr>
<tr>
<td>1699870873</td>
<td>Providence 5 (13,379)</td>
<td>1053368563</td>
<td>Diagnostic Radiologists (8,757)</td>
</tr>
</tbody>
</table>

Our hypothesis for the top 30 network was essentially the same as our hypothesis for the full network – that connectivity would increase after Health Share’s formation. We expected that new connections would form more often than dissolve, and that existing connections would strengthen more often than weaken. But by limiting the analysis to only thirty providers, it became computationally feasible to use five inference methods. And the addition of RA allowed us to expand our hypothesis to address higher-way relationships, which we expected would become more complex rather than simplify. (For example, we expected that two-way connections would developing into three-way connections more often than three-way connections would dissolve into two-way.)

The five methods of network inference differed in their findings on whether connectivity increased after Health Share’s formation, and additional analyses were required to be able to account for these differences. Table 5.4 contains the key findings.
from all five methods, which are summarized here before delving into the specific results by each method. Though results differed, they are ultimately taken to complement one another. And while some found connectivity to increase over time, the more general finding was that connectivity *consolidated* over time. After Health Share’s formation, connectivity increased for an inner core of connections, and weakened throughout the periphery of the top 30 network. This was most evident in the RA network, where higher-way associations strengthened while also simplifying to involve fewer billing providers. And it was also evident in the correlation network, where disparity in connection weights increased to produce a distribution where the strongest connections were even stronger, and the weakest connections were even weaker, after Health Share formed.

After identifying the more general pattern of consolidation, it became easier to understand how methods might differ in their determination of whether connectivity increased or decreased. Consolidation essentially means that connectivity is increasing among a subset of central provider pairs, while simultaneously decreasing among other providers. A simple aggregation of all changes, then, can easily wash out many of the effects in either direction, producing an average that falls just above or just below zero.
Table 5.3. Summary of change in the top 30 networks as indicated by each method of inference

<table>
<thead>
<tr>
<th>Method</th>
<th>Change in connectivity</th>
<th>Additional patterns observed</th>
<th>% stable pairs*</th>
</tr>
</thead>
<tbody>
<tr>
<td>RA</td>
<td>Connectivity increased (stronger associations, although among fewer providers)</td>
<td>When data were binned (made nominal) and when nonlinear and higher-way relationships were included, complex billing connections were found during both periods, often involving 3-5 providers. Complexity decreased over time to include fewer connections, and connections simplified to include fewer providers in each. But connectivity overall was stronger, based on a less uniform distribution of claims (i.e., lower entropy).</td>
<td>53% of connected pairs of providers were connected in both periods (47% in one period)</td>
</tr>
<tr>
<td>SBT</td>
<td>Connectivity decreased (average decrease in strength: −5E−7)</td>
<td>When data were binned (made nominal) and when nonlinear relationships were included but complexity was restricted to pairwise, the network showed very little change in connection weights and in the distribution of connection weights.</td>
<td>61% in both periods (39% in one period)</td>
</tr>
<tr>
<td>Correlation</td>
<td>Connectivity increased (average increase in strength: 0.00018)</td>
<td>When data were kept continuous, linear pairwise billing patterns also showed very little change in average connection weight (like SBT). But the distribution of connection weights shifted slightly, to produce a bigger gap between the strongest and weakest connections, indicating consolidation.</td>
<td>78% in both periods (23% in one period)</td>
</tr>
<tr>
<td>Multiple Regression</td>
<td>Connectivity decreased (average decrease in strength: 0.00145)</td>
<td>When data were kept continuous and when linear pairwise relationships were defined by unique predictive variance, connectivity became less consolidated, with less disparity between the strongest and weakest connections, implying that consolidation did not occur among unique pairwise relationships but instead occurred in multi-way relationships that involved three or more providers.</td>
<td>67% in both periods (33% in one period)</td>
</tr>
<tr>
<td>PLS</td>
<td>Connectivity increased (average increase in strength: 0.00105)</td>
<td>When data were kept continuous and when linear pairwise relationships were defined by their contributions to uniquely predictive latent terms, connectivity consolidated only slightly, if at all, with mostly the same distribution of connection strength during both time periods. This reiterates the notion that consolidation of connectivity occurred mainly in multi-provider relationships, remaining largely invisible at the pairwise level.</td>
<td>65% in both periods (35% in one period)</td>
</tr>
</tbody>
</table>

* The percentage of provider pairs who were connected during both periods was subject to visualization thresholds for all methods except for RA. Thresholds were designed to retain similar numbers of connections by each method, including 0.1 (correlation), 0.001 (SBT), 0.145
Not all five methods provided direct evidence for the consolidation of connectivity. But those that did not were still helpful to clarify the nature of this trend. For example, multiple regression and PLS – which define networks as unique relationships between providers – did not show consolidation of connectivity in the top 30 network. Instead PLS showed a mostly stable distribution of connectivity, and multiple regression actually showed less disparity in connection weights over time, as though connectivity became more evenly distributed and less concentrated. This result is counterintuitive, but ultimately suggests that that the type of connectivity that consolidated in this network was specific to ‘non-unique’ billing patterns that spanned three or more providers - for example, when the four-way relationship between Tuality, Adventist 1, CEP America, and AMR NW excluded Tuality after Health Share formed.

To understand how this works more clearly, note that network inference by multiple regression excludes multi-provider billing patterns from its estimation of connection strengths. In defining the relationship between any pair of providers, multiple regression filters out any covariance that could be doubly accounted for by another provider outside of the pair, and in that way it restricts its view to what is unique to that pair. As multi-provider connectivity consolidated, the providers involved in those denser, high-way relationships had a higher proportion of their billing patterns that were multiply-accounted for. This left less variability to be claimed by any ‘unique’ pairwise relationship. On the other hand, providers who become excluded from multi-provider relationships had more variability that could be attributed to unique pairwise

---

25 RA associations are listed Appendix B.
relationships. So, while central providers showed weaker unique connections over time, peripheral providers show stronger ones, causing regression’s apparent decrease in the disparity between the strongest and weakest connections.

Network inference by PLS regression is similar to multiple regression, in that it taps into unique relationships. But in this case PLS measures how much providers contribute to latent terms that are unique by definition (i.e., independent from one another). The use of latent terms allows PLS to adjust for multi-provider (non-unique) covariance to some extent, so its estimates of unique relationship strength can account for more variance in provider behavior than multiple regression connections can. PLS appeared relatively robust to the effects of multi-way consolidation of connectivity and did not show much of any change in the distribution of connection weights after Health Share formed.

The SBT network also did not show any evidence for connectivity consolidating. This was very counterintuitive at first but ultimately provided a unique insight that the types of consolidation found by RA and correlation were distinct types of consolidation. First, consider that SBT and RA are both based on measurements of transmission (i.e., association by information shared), with the main difference being RA’s ability to define higher-way relationships. The fact that SBT did not find any evidence of consolidation at the pairwise level tells us that RA’s evidence of consolidation was specific to the higher-way changes that occurred. When RA’s higher-way relationships are broken down into pairwise ones (equivalent to SBT), the evidence for consolidation disappears.
Now, consider that correlation and SBT and are both pairwise measurements of full (non-unique) connectivity, with the main difference being that correlation can utilize data in frequency form while SBT requires nominal data (binned into categories for 0, 1, or 2+ claims). The fact that SBT did not find evidence of consolidation with nominal data tells us that correlation’s consolidation was based on patterns of high-frequency billing.26

We might not have been surprised by either of these comparisons in isolation but taken together they are very interesting. For both RA and correlation to have found consolidation of connectivity, and for SBT to have been oblivious to it, RA’s evidence for consolidation would need to be specific to higher way relationships, and correlation’s evidence for consolidation would need to be specific to high-frequency patterns. Taken together, the results indicate that connectivity consolidated in at least two different ways. There was reduced complexity in multi-provider relationships, which involved fewer billing providers over time. And the overlaps in billing volume also became dominated by fewer providers. RA detects reductions in higher-way complexity, while correlation detects pairwise consolidation of billing volume in continuous form.

**Top 30 Network by Reconstructability Analysis**

Reconstructability analysis (RA) was the first of five inference methods applied to the top 30 billing providers. In this analysis, an ‘upward’ search was conducted separately for each of the datasets (before vs. after). The upward search began at the independence model, where no providers were connected, and progressed through models

---

26 Correlation analyses were actually replicated with binned data to confirm the disappearance of consolidation evidence when frequency data was recoded as 0, 1, or 2+ claims.
that had more and more connections. At each search level, we retained the 10 best models by BIC and used them to identify their ‘parent’ models at the next level – those equivalent except for the addition of one more connection (or higher-way connection). Searches ended after 200 levels, since no models had lower BICs beyond the 70th or 78th level of either search. Table 5.4 summarizes the main features of the two final best models (selected by BIC).

Table 5.4. Summary of best models for the top 30 network inferred by RA

<table>
<thead>
<tr>
<th></th>
<th>Before</th>
<th>After</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entropy in data</td>
<td>10.497</td>
<td>10.034</td>
<td>-4.4%</td>
</tr>
<tr>
<td>Entropy in best model</td>
<td>12.373</td>
<td>11.709</td>
<td>-5.4%</td>
</tr>
<tr>
<td>Change in degrees of freedom*</td>
<td>1,376</td>
<td>1,112</td>
<td>-19.2%</td>
</tr>
<tr>
<td>Search level where best model was identified</td>
<td>78</td>
<td>70</td>
<td>8</td>
</tr>
<tr>
<td>Percent of information captured by best model</td>
<td>60.8%</td>
<td>62.7%</td>
<td>3.1%</td>
</tr>
</tbody>
</table>

* Change is in reference to the independence model, as one moves toward a fully connected model (i.e., the data). Less change indicates a simpler (less complex) model structure.

The best model after Health Share formed had lower entropy, indicating stronger overall connectivity, but it also had less complexity, indicating fewer or simpler connections. This is somewhat strange to interpret, given that our hypothesis was for an increase in both the strength of connections as well as in the number and complexity of those connections. It forces us to unpack each of these trends separately. The increased strength of connectivity, which we expected, was visible even in RA datasets (prior to modeling), as the after dataset had less entropy than the before dataset. Less entropy means that the billing providers had less uniformity in billing, so they had stronger overall patterns. This decrease in entropy remained evident in the best models, which showed 5.4% less entropy (and stronger connectivity) after Health Share formed.
The decrease in complexity was not expected, but it was immediately evident from differences in both the search level and the change in degrees of freedom. The best model from after Health Share formed was found earlier in its search (at the 70th level), meaning it was closer to, and more similar to, the independence model. The after model also had less change in the number of degrees of freedom, meaning it needed fewer connections to achieve the best model with the lowest BIC. This is especially interesting given that the after model captures a slightly higher proportion of information than the before model (62.7% vs. 60.8%), so its relative simplicity cannot be due to a weaker fit to its data.27

Images of the top 30 billing provider network as inferred by RA can be seen in Figure 5.5. In that figure, each higher-way association is represented by a ‘hyper-edge node’ that connects to all members in that association. Grey edges indicate associations that were present during both time periods, and colored edges (blue or green) indicate memberships that were present in only one.

27 It is not appropriate to compare BIC values from both best models, since they have different reference points given that they are modeling different datasets. But the change in BIC, relative to the independence model, was also smaller after Health Share formed, indicating that the best model was more similar to the independence model, and was therefore simpler despite being the best fit in the search.
Figure 5.5. Top 30 billing provider network by RA
The provider memberships from each association are listed in Appendix B. There were fewer associations after Health share formed (25 versus 27), and those associations tended to have fewer providers in them. Figure 5.6 compares the complexity of associations during both time periods and illustrates how the network shifted from having predominantly 4-way associations to having predominantly 3-way associations.

![Figure 5.6. Complexity of Associations in the Before and After Network by RA](image)

As a whole, RA inference of the top 30 network showed stronger evidence for the *consolidation* of connectivity than it showed for an overall increase in it. While connectivity did strengthen, as evidenced by slightly lower entropy, it also became concentrated into a structure with fewer and simpler higher-way connections. This combination of effects – stronger connectivity but among fewer providers – provides direct evidence for consolidation. After Health Share’s formation, billing patterns among the top 30 billing providers became increasingly captured by a smaller subset of connections. This is somewhat different from what was hypothesized, which was not only for connection to strengthen, but for more of them to form and for them to become increasingly complex.
Having described the patterns observed in the RA networks, we can formally measure the amount of change that occurred. The measurement of network distance for RA is derived by comparing the contingency tables (i.e., probability distributions) that make up the best models from each time period. This is shown in the equation below

\[ \hat{\theta}_{RA} = \sum |q^2 - q^1| \]  

where \( q^1 \) and \( q^2 \) represent the calculated probabilities in the best models from before and after Health Share formed. Note that the distance measure for RA does not need standardized by the size of the network, \( g(g - 1) \), as it does for SBT and linear modeling methods (eqns. 26 and 22), because the sum of probabilities in an RA model will always be 1 and therefore the maximum difference across two sets of probabilities will always be 2. For the top 30 network of billing providers, the mean absolute difference (\( \theta_{RA} \)) was 0.43814, which is 21.9% of the theoretical maximum. This value does not directly pertain to our hypothesis for increased connectivity, but represents all change in the network, and will be tested for statistical significance in Chapter 6.

**Top 30 Network by Correlation**

The second method used to infer the top 30 network of billing providers was correlation. Spearman’s (rank-order) correlation method was first considered, because of its appropriateness for count data and for skewed distributions, but standard Pearson correlations were ultimately chosen so that findings from the correlation network could be compared more directly to findings from multiple regression and partial least squares (PLS) regression, which do not have rank-order counterparts. Connectivity in the
correlation network increased as hypothesized but only very slightly, by 0.00018 per correlation on average (see eqn 21 on p.126). Nearly all pairs of providers were correlated with an absolute value of at least 0.001 during both time periods.

Figure 5.7 illustrates the correlation network before and after Health Share’s formation. To generate the images, a weight threshold was set to $|r| \geq 0.1$. This filtered out all but 70 (16.1%) of the connections in the before network and 72 (16.6%) of the connections in the after network. The threshold of 0.1 was selected for a couple of reasons. First, it was one of the thresholds suggested in the bioinformatics literature (Pihur et al., 2008), and second, it produces approximately the same number of edges identified by RA if one were to break up higher-way associations into sets of pairwise ones (78 before and 70 after).
Figure 5.7. Top 30 billing provider network by correlation ($r \geq 1$)
The threshold of 0.1 also looks to be a reasonable cut-point in bar chart below (Figure 5.8), where the connection weights (i.e., the absolute values of correlations) have been plotted in order of strength. Figure 5.8 contains one column for each of the 435 correlations during each time period, in descending and ascending order for the before and after networks, respectively. Connections with weights of $|r| < 0.1$ were excluded from network visualizations (in Figure 5.7) and in this bar chart they are colored grey. Connections with weights $|r| \geq 0.1$ were included in network visualizations (Figure 5.7), and in this bar chart they have been colored blue or green.\(^\text{28}\) The threshold of 0.1 looks in this Figure 5.8 to be a reasonable cut-point, after which connection weights trail off.

![Figure 5.8. Correlation weights before and after Health Share’s formation](image)

The network images in Figure 5.7 illustrate a great deal of stability over time. As mentioned earlier, the average correlation strength only barely increased after Health Share’s formation (by 0.00018), and network images show that most visualized

\(^{28}\) Note that many of the connection weights that are blue or green in this bar chart were visualized in both the before and after network images, which means they appear grey in Figure 5.7.
connections were found during both time periods (62, 78%). There were ten connections that “appeared” above the visualization threshold, but they hardly outnumbered the eight connections that “disappeared”. Ultimately, it is difficult to argue that the correlation method found a substantial increase in connectivity after Health Share formed.

Despite the predominant finding of network stability, however, the correlation network offers some of its own evidence that connectivity consolidated a bit over time. This is illustrated in Figure 5.9, where the connection strengths of the strongest and weakest connections are plotted in descending order of strength. The single strongest connection in the after network was slightly stronger than the one in the before network (0.64641 versus 0.63836, see the left-hand side of the left-hand plot), as were most of the strongest 50 connections in the after network. Conversely, the single weakest connection in the after network was slightly weaker than the one in the before network (0.00031 versus 0.00005, see righthand side of the righthand plot), as were all of the 50 weakest connections in the after network.

![Figure 5.9. Absolute strengths of the strongest (left) and weakest (right) correlations](image)

---

29 Note that connection weights in the before and after period were rank-ordered separately, so the very strongest correlations from before and after can refer to different sets of providers.
This pattern seen in Figure 5.9 extends well beyond the subsets shown, and applies to most of the 435 provider pairs. The 272 weakest correlations after Health Share formed were weaker than the 272 weakest correlations before. And most of the 163 strongest correlations after Health Share formed (64%, 104) were even stronger than those before. The magnitudes of differences were all very small, but they still suggest an overall pattern of increased disparity between the billing providers with the strongest and the weakest connections over time. This consolidation means that there was slightly stronger connectivity among slightly fewer providers after Health Share’s formation.

The mean absolute difference in correlations ($\hat{\theta}_{corr}$, as calculated by eqn 22, p.127), which aggregates all (absolute) changes regardless of their direction, was 0.01841. This is only 0.92% of the theoretical maximum of two\(^{30}\), so it is much smaller than the amount of change identified by RA (21.9%) but can still be statistically tested in Chapter 6.

**Top 30 Network by Standardized Bivariate Transmission**

In comparing the billing provider networks inferred by RA and correlation, it is useful to consider what the RA network might have looked like if it had been constrained to identify only pairwise associations. To explore this, we calculated transmission for each pair of billing providers and then standardized it, dividing it by the maximum possible transmission (see eqns 23 and 24 on p.128). The average strength of connectivity via standardized bivariate transmission (SBT) actually decreased after Health Share’s

\(^{30}\) Theoretically all correlations could change from -1 to +1 between time periods, which is an absolute change of 2.
formation, which was contrary to our hypothesis and also contrary to the main results from RA and correlation. But it only changed by a miniscule amount, on average by $0.4e^{-5}$ as calculated by eqn 25 (see p. 130).

Values of SBT were very small (between 0.0001 and 0.04199) because most providers had highly skewed distributions, having billed for most patients zero times and billing for few patients two or more times. Skewed distributions caused the maximum possible transmission to be relatively large, so our division of $T_{ij}$ by $T_{max}$ results in a small SBT. To create the SBT visualizations shown in Figure 5.10, we set the connection threshold to SBT = 0.001. This retained a similar number of connections as were viewed in the RA and correlation networks – 75 edges before and 75 edges after Health Share’s formation.
Figure 5.10. Top 30 billing provider network by SBT (SBT ≥ .001)
The SBT network in Figure 5.10 appears to show more change than the correlation network did, in terms of who was connected to whom. There were twice as many connections appearing or disappearing by SBT (36) than there were by correlation (18). However, SBT showed basically the same level of connectivity during both time periods, with 75 connections consistently visualized and very little change in the average connective strength. This leaves no basis of support for the hypothesis that connectivity would increase. Even the mean absolute difference measure ($\theta_{SBT}$), which is based on the aggregate change in any direction, remained very small at 0.00029. (This is the value that will be statistically tested in chapter 6.)

Another interesting feature of the SBT network is that it does not provide direct evidence that connectivity consolidated over time. Where RA showed fewer and simpler connections after Health Share’s formation, SBT shows exactly the same number. And where correlation found increased disparity between the strongest and weakest connections, SBT shows almost perfect stability. Figure 5.11 plots the 50 strongest and weakest connections by SBT, illustrating that the strongest connections were no stronger after Health Share’s formation, and the weakest connections were not really any weaker. This means there is really no evidence for consolidation of connectivity by SBT.
Comparisons of RA, Correlation, and SBT

The top 30 networks inferred by all three methods so far – RA, correlation, and SBT - are shown below in Figure 5.12, and their connections are summarized below in Table 5.5. Despite not finding increased connectivity or consolidation, the SBT network still serves as a theoretical middle ground between correlation and RA networks. This is true both conceptually, for how relationships are defined, and also proves to be true visually in the amount of overlap between these three methods. As shown in Table 5.5, the visualized SBT network has higher overlap with other networks. It had 49% of connections in common with the RA network and 50% of connections in common with the correlation network, while RA and correlation networks only shared 37% of their connections, leaving 63% unique from one network or the other.
Table 5.5. Top 30 Network edges found by RA, Correlation, and SBT

<table>
<thead>
<tr>
<th></th>
<th>Before</th>
<th>After</th>
<th>Before or After (∪)</th>
<th>Before &amp; After (∩)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total edges found by any method (∪)</strong></td>
<td>114</td>
<td>107</td>
<td>139</td>
<td>82</td>
</tr>
<tr>
<td>Shared by all three methods (∩)</td>
<td>39 (34%)</td>
<td>40 (37%)</td>
<td>46 (33%)</td>
<td>33 (40%)</td>
</tr>
<tr>
<td>Shared by any two methods</td>
<td>31 (27%)</td>
<td>30 (28%)</td>
<td>27 (19%)</td>
<td>22 (27%)</td>
</tr>
<tr>
<td>Found only by RA</td>
<td>21 (18%)</td>
<td>16 (15%)</td>
<td>30 (22%)</td>
<td>11 (13%)</td>
</tr>
<tr>
<td>Found only by Correlation</td>
<td>15 (13%)</td>
<td>12 (11%)</td>
<td>20 (14%)</td>
<td>11 (13%)</td>
</tr>
<tr>
<td>Found only by SBT</td>
<td>8 (7%)</td>
<td>9 (8%)</td>
<td>16 (12%)</td>
<td>5 (6%)</td>
</tr>
<tr>
<td><strong>Total edges found by RA or Correlation (∪)</strong></td>
<td>106</td>
<td>98</td>
<td>135</td>
<td>69</td>
</tr>
<tr>
<td>Shared by RA &amp; Correlation (∩)</td>
<td>42 (40%)</td>
<td>44 (45%)</td>
<td>50 (37%)</td>
<td>36 (52%)</td>
</tr>
<tr>
<td>Found by RA but not Correlation</td>
<td>36 (34%)</td>
<td>26 (27%)</td>
<td>48 (36%)</td>
<td>14 (20%)</td>
</tr>
<tr>
<td>Found by Correlation but not RA</td>
<td>28 (26%)</td>
<td>28 (29%)</td>
<td>37 (27%)</td>
<td>19 (28%)</td>
</tr>
<tr>
<td><strong>Total edges found by RA or SBT (∪)</strong></td>
<td>99</td>
<td>95</td>
<td>137</td>
<td>57</td>
</tr>
<tr>
<td>Shared by RA &amp; SBT (∩)</td>
<td>54 (55%)</td>
<td>50 (53%)</td>
<td>67 (49%)</td>
<td>37 (65%)</td>
</tr>
<tr>
<td>Found by RA but not SBT</td>
<td>24 (24%)</td>
<td>20 (21%)</td>
<td>35 (26%)</td>
<td>9 (16%)</td>
</tr>
<tr>
<td>Found by SBT but not RA</td>
<td>21 (21%)</td>
<td>25 (26%)</td>
<td>35 (26%)</td>
<td>11 (19%)</td>
</tr>
<tr>
<td><strong>Total edges found by Correlation or SBT (∪)</strong></td>
<td>93</td>
<td>91</td>
<td>119</td>
<td>65</td>
</tr>
<tr>
<td>Shared by Correlation &amp; SBT (∩)</td>
<td>52 (56%)</td>
<td>56 (62%)</td>
<td>60 (50%)</td>
<td>48 (74%)</td>
</tr>
<tr>
<td>Found by Correlation but not SBT</td>
<td>18 (19%)</td>
<td>16 (18%)</td>
<td>25 (21%)</td>
<td>9 (14%)</td>
</tr>
<tr>
<td>Found by SBT but not Correlation</td>
<td>23 (25%)</td>
<td>19 (21%)</td>
<td>34 (29%)</td>
<td>8 (12%)</td>
</tr>
<tr>
<td><strong>Total edges found by RA (% of ∪)</strong></td>
<td>78 (90%)</td>
<td>70 (80%)</td>
<td>97</td>
<td>51 (59%)</td>
</tr>
<tr>
<td><strong>Total edges found by Correlation (% of ∪)</strong></td>
<td>70 (88%)</td>
<td>72 (90%)</td>
<td>80</td>
<td>62 (78%)</td>
</tr>
<tr>
<td><strong>Total edges found by SBT (% of ∪)</strong></td>
<td>75 (81%)</td>
<td>75 (81%)</td>
<td>93</td>
<td>57 (61%)</td>
</tr>
</tbody>
</table>
Figure 5.12. Top 30 billing provider network by RA, SBT, and correlation
Conceptually speaking, the SBT inference method can be thought of as the lowest common denominator between RA & correlation. Relative to RA, SBT is limited in that cannot detect higher way relationships – just like correlation cannot. And relative to correlation, SBT is limited in that must use binned variables and treat them nominally – just like RA must. Using binned variables means SBT cannot detect changes specific to high claim volumes, since they are grouped in the same bin for “2 or more” claims. And nominal analysis means that SBT cannot view treatment levels as ordinally related. In SBT, billing one claim for a patient is wholly different than billing two claims. It is not treated as falling ‘between’ the other categories of 0 and 2+ claims.

When framed as the lowest common denominator, SBT provides an interesting backdrop for the other methods’ substantive findings on connectivity and consolidation. Recall that both RA and correlation found increased connectivity and consolidation, while SBT found a decrease in connectivity and no consolidation. Since we know that RA and SBT both used the same (binned, nominal) variables, their differential findings suggest that RA’s connectivity & consolidation were driven predominantly by billing patterns that spanned 3 or more providers. And since correlation and SBT are both limited to pairwise analysis, their differential findings suggest that correlation’s connectivity & consolidation were driven primarily by the use of (un-binned) frequency data. This means that either the ordinality or the prevalence of high claim frequencies
(above 2) must be central to correlation’s ability to detect increased connectivity and consolidation patterns.\textsuperscript{31}

Using SBT as a midpoint between the other two methods also helps understand the relative capacities of RA and correlation for network inference. RA found connections between 35 pairs of providers whose SBT values did not exceed the visualization threshold. This suggests that many of RA’s connected pairs (35/97, or 36\%) were not strong enough to detect at a pairwise level but were embedded in strong complex (higher-way) connections. In a complementary manner, the correlation network identified 25 pairs of providers with connections above the correlation threshold but not above the SBT threshold. This implies that 31\% of correlations (25/80) were not easily detectable without the use of full frequency data (or at least higher-frequency bins).

Finally, using SBT as a midpoint between the other two methods helps to illustrate the amount of change that they can detect over time. As shown in the bottom rows of the right-most column of Table 5.5, the correlation network found the least amount of change, with 78\% of its connections persisting over both time periods. RA showed the largest amount of change, with only 59\% of its connections persisting, and SBT is between them, closer to RA, with stability in 61\% of its connections. This would suggest that RA was the most effective of these three methods for detecting change in the Health Share CCO network, arguably through its ability to detect higher-way connections between providers.

\textsuperscript{31}Correlation analyses were actually replicated with binned data to confirm the disappearance of consolidation evidence when frequency data was recoded as 0, 1, or 2+ claims.
There is an important caveat in visualized networks on all of these metrics, since the appearance and disappearance of connections is sensitive to the visualization threshold. With the exception of RA, where all associations from the best model have been visualized, all other networks are subject to connection weight thresholds, which have been approximated so as to capture a similar number of pairwise edges as RA generated. This means that correlation and SBT network images filtered out the majority of (very weak) connections during both time periods, leaving only the very strongest connections in Figure 5.7, Figure 5.10, and Figure 5.12.

**Figure 5.13. Strongest 200 connection weights for SBT and correlation**

Figure 5.13 illustrates the sensitivity of correlation and SBT to these visualization thresholds, by plotting the connection weights from either time period in descending order. The distribution of SBT looks especially sensitive, due to what appears to be a negative exponential shape in the connection weight distribution, and due to placement of the threshold near what looks like an asymptotic lower bound. This means that even miniscule changes in SBT values can move a connection above or below the threshold,
potentially exaggerating the amount of change in the visualized networks even if connections changed only slightly. The distribution of correlation weights appears less sensitive, by comparison, because the rate of decrease around the threshold is steeper. This means correlation weights have to change more substantially for their visualized edges to appear or disappear.

Threshold sensitivity is a limitation to keep in mind when comparing networks on their proportions of edges “shared” and the amount of change over time. This pertains to correlation and SBT, as well as the multiple regression and PLS methods that will follow. When virtually all pairs of network members are connected but many of them by tiny amounts, it can be difficult to determine which of the connections are meaningful. Especially for large networks, it is often infeasible or at least difficult to rigorously test all connections for statistical significance, due to the high likelihood of Type I errors. Fortunately, when the ultimate purpose of network inference is to allow for statistical testing of network change, all connection weights can be used for analysis, without any need to parse out those that are suspected unimportant.

**Types of Connections Found**

Having clarified the limitations imposed by threshold sensitivity, we can explore differences in the types of connections inferred by each method. Information theory has a more general definition of relationship than correlation, which allows RA and SBT to find nonlinear relationships. And RA’s inclusion of higher-way relationships allows it to find connections that may be invisible at the pairwise level. But information theory also requires that variables be discretized prior to analysis, while linear modeling methods can
utilize variables in continuous or frequency form. This difference prompts us to explore the nature of relationships that were only identified through either linear means (correlation) or information theoretic means (RA and SBT).

Figure 5.14 illustrates a nonlinear relationship – between Multnomah County and Providence 3 – that was detected by RA and SBT but was well below the correlation threshold \( r = -0.047 \). While standard correlations must show a linear relationship to appear connected, information theoretic relationships only have to contain an anomalously high or low probability in one or more state combinations. If you look at the contingency table at the right of Figure 5.14, you’ll see that some of the observed frequencies (bold) and expected frequencies (in parentheses) had large differences. Cells are shaded grey where the observed frequencies were smaller than would be expected by the marginal probabilities. You can see the relationship in the contingency table is not positive or negative but is a more complex type. The scatter plot at the left looks nonlinear and only crudely represented by the correlation of \(-0.047\).

32 A state is anomalously high or anomalously low when its observed frequency is markedly different from what you would have expected by the marginal probabilities from each variable.
Figure 5.14. Scatterplot and contingency table for a connection found by information theory but not found by correlation

Figure 5.15 illustrates a relationship – between AMR NW and OHSU 2 – that was detected in the before network by correlation but not by information theory (RA or SBT). This correlation was above the visualization threshold \( r = .12 \) and ‘positive’, meaning patients who were billed for by AMR NW with high frequency also tended to be billed for by OHSU 2 with high frequency. But when looking at the scatterplot on the left side of Figure 5.15, the relationship does not look very strong, or even very positive. The correlation, then, may have stemmed from high claim frequencies or ordinality that was inaccessible under information theory’s nominal treatment of binned data.

In the contingency table on the right side of Figure 5.15, cells are again shaded grey when the observed frequencies (in bold) were smaller than would be expected by marginal probabilities (in parentheses). This contingency pattern loosely follows a prototypical positive correlation, with higher frequencies than expected along the diagonal from the upper left to the bottom right corner. But in RA’s evaluation of overall patterns, the observed frequencies and expected frequencies were not different enough to
suggest a relationship between these providers. The largest difference was in cell 0,0, where the number of patients who were not billed for by either provider was 885 higher than one would expect by chance. But this difference still only amounts to .8% of the total patient volume and is less than a 1% increase over the expected volume for that cell.

Figure 5.15. Scatterplot and contingency table for a relationship found by correlation but not by information theory

Differences between the two information theory methods – RA and SBT – illustrate instances of holism in provider relationships. There were 30 pairs of providers who did not appear connected by SBT (or correlation) at the pairwise level, but who did appear connected by RA in higher-way associations. For example, Multnomah County and Lifeworks NW were in a three-way association with Tuality before Health Share’s formation, and a different three-way association with Providence 1 afterward. But as a pair these providers were not found to be connected by SBT during either time period – their SBT value was .0002 before Health Share and .00045 afterward, both times well
below the threshold of .001. This means that these two providers were key players in a higher-way relationship, but their direct relationship was virtually undetectable.

The inverse happened even more often. There were 35 pairs of providers who were connected by SBT at the pairwise level, but whose relatedness was not include in RA’s best model. For example, RA did not detect any connection that included both Radiology Consultants and Northwest Acute, during either time period, despite SBT values above the threshold (0.008 before; 0.009 after). It also did not find a connection between Radiology Consultants and Legacy Emanuel 1 (\(SBT = 0.003 \) both periods) or Providence 1 and AMR NW (\(SBT = 0.002 \) both periods). These pairwise relationships – although detectable by information theory, were dwarfed by the strength of other higher-way associations in RA’s determination of the best model. As the RA searched through more and more complex models, it found better success (by BIC) when adding more higher-way associations than these pairwise ones.

In looking for consistency between the networks inferred by correlation, RA, and SBT, we might have the most confidence about relationships, either before or after, among the 46 pairs of providers who were captured (and visualized) by all three methods of inference. These 46 pairs constitute 33% of the 139 pairs of providers who were found to be connected by any method. To some extent, the three methods also display similarities in the general structure of their networks, which can be seen below in Figure 5.16. All three methods indicate relatively high levels of interconnectivity within organizations, for example between Providence 1 through 5, and they indicate anesthesiology and ambulatory services as bridging between organizations.
Top 30 Network by Multiple Regression and Partial Least Squares

Two additional methods were used to infer the top 30 billing provider network: multiple linear regression and partial least squares (PLS) regression. Both of these methods estimate a ‘unique’ relationship between providers, meaning the amount of variance in one provider’s behavior that is predicted by another provider’s behavior and cannot be doubly accounted for by any third provider in the network. And for both of these methods, relationships are directional, so one provider’s billing patterns might strongly predict another’s, even if that predictive capacity is not reciprocated. The main difference between these two methods is that PLS estimates relationships indirectly, by the extent to which one billing provider contributes to latent terms that predict the
behavior of the other billing provider (denoted $s$). Multiple regression estimates, by contrast, are taken directly from covariance behavior patterns between billing providers (denoted $B$), without their predictive effect being relayed through any latent terms.

Regression and PLS networks are shown on the following pages in Figure 5.17 and Figure 5.18. For the sake of interpretability, their connections are visualized as non-directed links, based on a sum of weights from $i$ to $j$ and $j$ to $i$. Visualizations are also subject to thresholds – with directed regression weights $|B_{ij}| \geq 0.145$ and directed PLS weights $|s_{ij}| \geq 0.082$ – for comparability to RA, correlation, and SBT networks. The regression network features 70 edges above the threshold before Health Share’s formation and 75 after, and the PLS network features 71 PLS edges before and 74 after.
Figure 5.17. Top 30 billing provider network by multiple regression ($B \geq 0.145$)
Figure 5.18. Top 30 billing provider network by PLS ($s \geq 0.08$)
Like with correlation and SBT networks, the vast majority of pairs of providers were at least weakly connected, and adjustment of their visualization thresholds could have dramatically changed the number of connections included. Figure 5.19 plots the distributions of all regression and PLS weights, relative to the thresholds, illustrating their sensitivity. Many connections were only slight above or below the selected threshold.

![Figure 5.19. Strongest 200 connection weights for regression and PLS](image)

PLS and multiple regression found different results regarding our hypothesis for increased connectivity. Recall that we expected Health Share’s formation to be followed by stronger connections between providers and more connections between providers. Both methods had a few more connections appearing above the visualization threshold after Health Share formed (five and three more, for regression and PLS respectively), but multiple regression found the average connection strength ($|B|$) to weaken (by $|B| = -0.00145$, eqn 5.9) while PLS found it to strengthen ($|s| = 0.00105$, eqn 5.10). The equations for these metrics (below) are analogous to those used from previous methods, only with substitutions for the term representing connection weight (e.g., $B$, $s$).
\[
\frac{1}{g(g-1)} \sum_{i \neq j} |\bar{B}_{ij}^2| - |\bar{B}_{ij}^1|
\]  \hspace{1cm} (28)

\[
\frac{1}{g(g-1)} \sum_{i \neq j} |\hat{s}_{ij}^2| - |\hat{s}_{ij}^1|
\]  \hspace{1cm} (29)

While not directly relevant to our substantive hypothesis for increased connectivity, the mean absolute differences were $\hat{\theta}_{reg}=0.02456$ for multiple regression (see eqn 5.13) and $\hat{\theta}_{PLS} = 0.01560$ for PLS (see eqn 5.14). These values aggregate network change in all directions (increases and decreases), into a total that will be tested for statistical significance in Chapter 6.

\[
\hat{\theta}_{reg} = \frac{1}{g(g-1)} \sum_{i \neq j} |\bar{B}_{ij}^2 - \bar{B}_{ij}^1|
\]  \hspace{1cm} (30)

\[
\hat{\theta}_{PLS} = \frac{1}{g(g-1)} \sum_{i \neq j} |\hat{s}_{ij}^2 - \hat{s}_{ij}^1|
\]  \hspace{1cm} (31)

**Comparisons with Multiple Regression and PLS**

Having measured changes in connectivity for five different methods now, it is tempting to compare their values. For example, connections weakened by 0.00145 by multiple regression but only by $5E-7$ for SBT. And connections strengthened by 0.00145 for PLS, but by only 0.00018 for correlation. Unfortunately, fundamental differences in inference methods make these like comparisons of apples and oranges. It is inaccurate to compare an average change by correlation, where weights fall naturally between 0 and 1, to an average change by SBT, where weights were scaled by the activity levels of the providers involved (see p.132). Similarly, it is inaccurate to compare values with theoretical maximums (1 for absolute correlation and SBT), against values without
theoretical maximums, as regression coefficients and PLS contributions can be greater than 1.\textsuperscript{33}

Fortunately, we can still compare inference methods via network visualizations, so long as we keep in mind that these visualizations are subject to thresholds. Figure 5.20 compares regression, PLS, and correlation visualizations, and Table 5.6 compares all five methods for the number of connections found during each time period. In that table, the PLS network shows the most change of the linear modeling methods, with 35% of its edges unique to one time period, compared to 33% of regression edges and 23% of correlation edges. This suggests that PLS’ use of latent terms allowed it to be slightly more effective than multiple regression in detecting change in providers’ linear, pairwise relationships.

Table 5.6. Summary of Pairs in Visualized Networks

<table>
<thead>
<tr>
<th>Method</th>
<th>RA*</th>
<th>SBT</th>
<th>Correlation</th>
<th>Regression</th>
<th>PLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Threshold for visualization</td>
<td>none</td>
<td>0.001</td>
<td>0.1</td>
<td>0.145</td>
<td>0.082</td>
</tr>
<tr>
<td>All pairs found Before</td>
<td>78</td>
<td>75</td>
<td>70</td>
<td>70</td>
<td>71</td>
</tr>
<tr>
<td>All pairs found After</td>
<td>70</td>
<td>75</td>
<td>72</td>
<td>75</td>
<td>74</td>
</tr>
<tr>
<td>Total edges in either period</td>
<td>97</td>
<td>93</td>
<td>80</td>
<td>87</td>
<td>88</td>
</tr>
<tr>
<td>Pairs found in both periods</td>
<td>51</td>
<td>57</td>
<td>62</td>
<td>58</td>
<td>57</td>
</tr>
<tr>
<td>Pairs found in one period</td>
<td>46</td>
<td>36</td>
<td>18</td>
<td>29</td>
<td>31</td>
</tr>
<tr>
<td>Pairs found only before</td>
<td>27</td>
<td>18</td>
<td>8</td>
<td>12</td>
<td>14</td>
</tr>
<tr>
<td>Pairs found only after</td>
<td>19</td>
<td>18</td>
<td>10</td>
<td>17</td>
<td>17</td>
</tr>
</tbody>
</table>

* RA hyper-edges have been counted as distinct sets of pairs.

\textsuperscript{33} Even if we were to standardize billing provider variables prior to analysis, standardized regression coefficients can exceed one (Deegan, 1978).
Figure 5.20. Top 30 billing provider network by correlation, multiple regression, and PLS
Because Table 5.6 indicates PLS found only slightly more change than regression, it is worth investigating the possibility that this difference was only a function of our visualization threshold at $|s| \geq 0.082$ and $|B| \geq 0.145$. There were many connections close to the PLS threshold during at least one time period, including 57 connections within 20% of it ($0.0656 \leq s \leq 0.0984$) and 165 connections that were within 50% of it ($0.041 \leq s \leq 0.123$). And several of these connections moved above or below the threshold over time (4 of the connections from within 20% and 9 from within 50%). So, changing the PLS or regression threshold might have affected the relative amount of change found by either method.

To evaluate this more closely, we conducted a sensitivity analysis for all three linear methods, to evaluate the effect of varying visualization thresholds on the proportion of edges that changed over time. Each possible threshold was considered, from the highest level, which included only the strongest (absolute) edge weight, to the lowest level, which included even the weakest edge weight. Results are below in Figure 5.21. For each level (shown along the x-axis), we calculated the proportion of changing connections that were only above the threshold during one time period. The shaded portion of the graph shows the thresholds that are similar to those employed for other methods, including those that produced 60-90 connections during one period or 70-100 connections across both periods.
There are three technical notes when considering the distributions in Figure 5.21.

First, note that the threshold amount at each level (shown along the x-axis) was only rarely the same value across methods. For example, the 114th edge weight is $r = 0.12273$ for correlation, $\beta = 0.20274$ for multiple regression, and $s = 0.11498$ for PLS.

The distributions of edge weights were very different between the two methods, with regression weights being considerably larger. So, if the same threshold value were applied equally to both networks (e.g., 0.1), the visualized multiple regression network would have included many more edges than correlation or PLS networks, making methodological comparisons more difficult.\footnote{There were 80 correlation edges with weights $\geq 0.1$, followed by 127 regression edges and 76 PLS edges.}

The second technical note pertains to the large amount of ‘noise’ at higher threshold levels where fewer connections are included. (This is visible at the left-most edge of the plot in Figure 5.21.) To understand this, consider the most restrictive case, where the threshold includes only the strongest connection from either time period. In
that case, 100% of the connections were found during only one period, because the strongest connection weight was not identical during both periods. In the second most restrictive case, where the threshold includes the two strongest connections from either period, this value drops to 0% for multiple regression, because the two strongest weights belonged to the same pair of providers, one during each time period, so the before and after network were identical when including those two edges. These dramatic changes become less and less possible as the threshold is lowered to include more connection weights.

The third technical note pertains to the number of edges in each network, which affects the length of the curve for correlations when compared against PLS and multiple regression. Recall that correlations are bidirectional, meaning any given pair of providers has only one correlation. This makes for 435 total connections in a network of 30 members \((30 \times 29/2)\), measured at each of two time periods for a maximum of 870 unique correlation-based edge weights. By contrast, in PLS and multiple regression relationships are directional, meaning any pair of providers \((i\) and \(j\)) will have two connections, one from \(i\) to \(j\) and one from \(j\) to \(i\). This doubles the number of connections, from 870 to 1,740, and explains why the curves for PLS and multiple regression are twice as long as the curve for correlation in Figure 5.21.

Technical notes aside, there are some obvious patterns in Figure 5.21. By and large, PLS and multiple regression found very similar amounts of change in the top 30

\(^{35}\) In the case of PLS and correlation, the two strongest connections did not belong to the same pair of providers, so the proportion of connections that changed remained at 100%.
network, and both found reasonably higher rates of change than correlation. On average across all thresholds, PLS found 33.9% of its connections shift above or below the threshold over time (standard deviation =11.5%) compared to 32.5% for multiple regression (SD=11.5%) and 24.3% for correlation (SD=8.5%). The relative advantage of PLS was somewhat more pronounced for a subset thresholds approximately similar to RA, SBT, and correlation networks, as highlighted in the shaded portion of Figure 5.21 and expanded in Figure 5.22. For thresholds that produced 60-90 connections during one period or 70-100 connections across both periods, the average percentage of change was 38.0% of connections for PLS (SD=2.8%), 33.6% for multiple regression (SD=2.5), and 26.2% for correlation (SD=2.7%). However, even in this restricted range, PLS does not show much of an advantage over multiple regression in its detection of network change.

![Figure 5.22. Percent of connections above threshold during one period, in restricted range](image)

In considering Figure 5.21 and Figure 5.22, it is perhaps not surprising that the PLS and multiple regression networks showed more change than the correlation network. By definition, the correlation network encompasses all covariance between each pair of
providers, including billing patterns that are general throughout the network as well as the patterns specific to each pair. Unique relationships might change more often and more quickly than general, multi-provider billing patterns across the network. These general patterns appear to add noise to the estimate of each pairwise connection, and they enhance the appearance of stability over time.

It is somewhat surprising, however, that PLS and multiple regression networks were so similar in how much change that they detected. In other fields, PLS has been found to be more effective at detecting unique relationships when predictors are non-independent (Carrascal et al., 2009; Datta, 2001; Pihur, Datta & Datta, 2008). Interdependence of billing providers was certainly the case here, and multiple regression might have been expected to ignore a lot of connectivity between its providers when it was doubly accounted for across multiple pairs of providers. This is a problem that PLS should have been able to circumvent to a large extent by allowing providers to contribute to independent latent terms. However, the magnitude of this advantage appears to have been relatively small in this application.

Having considered the amount of change by each linear method, we might also compare methods for how well they can capture variance in provider behavior. Figure 5.23 compares this for correlation, PLS, and multiple regression, with one set of horizontal bars for each of the top 30 providers, listed at right. The length of the bars (along the x-axis) indicates the proportion of variance in each provider’s behavior that was accounted for (explained) by each method. For correlation (blue), this is equal to the largest absolute correlation. For multiple regression, this is equal to the $R^2$ of the model.
regressing that provider’s billing on all others’. For PLS regression, the $R^2$ has been partitioned into unique (light green) and shared (dark green) variance explained. Bars extend left for the network before Health Share’s formation, and bars extend right for the network after.
Figure 5.23. Amount of variance explained by correlation, multiple regression, and PLS, before Health Share’s formation (at left) and after it (at right)
A couple of patterns stand out in Figure 5.23. Perhaps most striking is the extent to which correlation outperforms both other methods in accounting for provider behavior—evidenced by the blue bars being longest. All providers had correlations that could account for more of their behavior than either of the other methods could. In theory, one might expect multiple regression $R^2$ values to at least match the largest correlation, because they encompass both unique and shared covariance. However, when predictor variables are highly correlated with one another—an issue called multicollinearity—they can actually suppress shared prediction of the dependent variable, which reduces the model’s $R^2$ (Velicer, 1978). The fact that multiple regression $R^2$ values for providers are consistently smaller than largest correlations, reiterates the extent to which patient billing is multiply-shared, and reflects the reality that multiply-shared patterns are not amenable to capture in pairwise relationships.

The second prominent pattern in Figure 5.23 is that multiple regression appears to outperform PLS in its prediction of provider behavior, as evidenced by how green bars (light + dark green) routinely surpass yellow bars. This comparison is somewhat misleading, because part of a multiple regression’s $R^2$ value includes covariance that is shared by all other providers (in dark green) even though this portion is not attributable to any unique pairwise relationship and is therefore absent from the inferred regression network. PLS, by contrast, is built on latent terms, which are all independent by definition, so its $R^2$ only encompasses unique prediction. A better comparison of PLS and multiple regression, then, is between the variance captured by PLS (in yellow) and the unique variance captured by multiple regression (in light green). These unique predictive
capacities are what were visualized in the network images in Figure 5.17, Figure 5.18, and Figure 5.20.

The comparison of unique variance captured by PLS (yellow) and multiple regression (light green) indicates that PLS routinely explains more variance in provider behavior with unique relationships than multiple regression can. Multiple regression faces a disadvantage here, because it can only attribute unique predictive capacity when it cannot be doubly accounted for by any other provider. PLS, through its use of latent terms, can capture more provider behavior patterns as unique pairwise relationships. So even if both methods showed similar amounts of network change, we might have more confidence in the PLS results because the PLS network accounts for more of providers’ behavior patterns than the regression network does.

Having compared the methods by how much change they reveal, and how much variance they can capture, we might also compare changes in their weight distributions, to determine whether they provided evidence for consolidation of connectivity. Recall that a primary finding from the correlation network was that connectivity appeared to consolidate over time, with its strongest connections even stronger after Health Share formed, and its weakest connections even weaker. PLS may have found this, to a much weaker extent, but surprisingly the multiple regression network showed less disparity over time, with its strongest connections slightly weaker after Health Share formed, and its weakest connections slightly stronger.
Why did the distribution of correlation weights show an increase in disparity, while the multiple regression distribution showed a decrease in disparity and the PLS distribution remained basically the same? Obviously, these changes are all very small and may be largely trivial in size. But to the extent that the trends are valid, we can explore the most likely explanation, which stems from change in billing patterns that are shared by multiple providers. These multiway patterns are captured in the correlation and RA networks, but not in multiple regression or PLS networks because those two methods look only at unique pairwise relationships between providers.
Let’s first consider the decrease in disparity by multiple regression. The fact that disparity decreased in the multiple regression network while increasing in the correlation network could be evidence that billing patterns among strongly connected providers became increasingly non-unique, or shared by sets of three or more providers. Patients had fewer billing providers after Health Share formed than they did before, going down from 3.77 providers per patient to 3.66, and given the consolidation found by RA and correlation, it seems unlikely that patients dropped billing providers at random. Instead, it seems plausible that patients’ sets of billing providers would have aligned around smaller sets of highly connected providers. If that were the case, these central providers would have increasingly similar sets of patients, causing greater overlap (i.e., larger correlations), but inversely leaving them with less room to be uniquely related.

If this type of consolidation did occur, the opposite would have been true for providers who were not as central to the network. Those who were increasingly excluded from the network’s consolidated core would have had less similar sets of patients, causing less overlap with other providers after Health Share formed (i.e., lower correlations), but inversely allowing them more room for unique relationships. Thus, one can imagine how a trend of patient alignment around fewer providers (i.e., consolidation) could simultaneously lead to increased disparity in correlation-based connections and decreased disparity in regression-based connections.

Relative stability in the distribution of PLS edges likely speaks to its ability to ‘dig out’ unique relationships between network members. As patients’ billing provider networks became more aligned, and as multiply-shared billing patterns become stronger,
standard multiple regression was hindered in its ability to detect unique pairwise relationships among the most highly connected providers. But PLS appears to have overcome this limitation with using latent terms, as it found the distribution of connection weights to be approximately similar during both timer periods. This may give PLS a bit of an advantage when a research question pertains to pairwise relationships between providers and when multiply-shared behavior patterns are considered noise. Based on Figure 5.23, PLS appears more effective at discerning unique relationships between pairs of members in the context of multicollinearity, a feature that should lend itself well to many research questions in a traditional (pairwise) network context.

**Summary of the Top 30 Billing Providers**

As a set, the five methods of network inference provide some consistencies and some complementary information on the top 30 network of billing providers. From an organizational view, all methods provided evidence of strong within-organization connectivity, repeatedly identifying connections within Legacy, Providence, and Adventist. This offers face validity to all of the methods, since patients can be expected to see multiple providers within the same organization more than at competing ones. In addition, all methods found changes in connections between organizations. This also offers face validity, since Health Share’s formation would be expected to reorganize billing patterns between its member organizations as they come together under one global budget to serve the healthcare needs of the local Medicaid population.

Regarding our specific hypothesis for the top 30 network, methods differed on whether or not they found evidence of increased connectivity after Health Share’s
formation. This was somewhat complicated to evaluate, because we expected increases in connectivity to appear in the form of more connections, stronger connections, and in the case of RA, more complex (higher-way) connections. Two of the methods, correlation and PLS, found stronger connections and more connections, as hypothesized. But other methods had mixed findings. RA found stronger connectivity, but fewer connections that were less complex (lower-way), and multiple regression and SBT found weaker connectivity, with the same or fewer connections after Health Share formed.

Discrepancies in findings for the main hypothesis are reconciled somewhat by observing the consolidation of connectivity. This was most obvious in results from RA, where connectivity became stronger in general terms (i.e., with lower entropy) but more concentrated among fewer and simpler (lower-way) connections. The notion that connectivity may have consolidated over time, rather than simply increasing across the network, explains the near-zero methods of change in connectivity, and explains why some methods (e.g., correlation) could have found connections to strengthen, while other methods (e.g., multiple regression) might have found them to weaken.

Keep in mind that neither increased connectivity nor consolidation of connectivity were stated intentions of Oregon’s CCO legislation, so these findings do not evaluate any of its intended aims (e.g., reduced cost, better health outcomes, and higher satisfaction). We hypothesized that connectivity would increase as a natural consequence of CCO’s

---

36 Counts of connections were evaluated as counts of connections above a visualization threshold, with RA as the only exception. For all methods besides RA, virtually all pairs of providers were connected by at least a miniscule amount during both time periods, so visualization thresholds allow counts to refer to the appearance or disappearance of ‘reasonably’ strong connections between providers.
adoption of the Patient Centered Primary Care Home (PCPCH) model of care, where access to specialty services requires a referral from a primary care provider. It seemed reasonable to expect that adoption of the PCPCH would augment referral patterns, and that this would be measurable in a network as increased connectivity. However, we may have been wrong to expect this effect to be evenly distributed throughout the network of healthcare delivery, as referral patterns might have naturally consolidated around a central core of billing providers. This appears to have happened in the top 30 network.

Methodologically speaking, the biggest difference between the methods comes down to their treatment of multi-provider billing patterns. RA was the only method that formally captured these as higher-way relationships. It suggested that billing patterns were best represented as multi-provider patterns, and it gave us the most direct evidence of any method that connectivity consolidated after Health Share’s formation. The complexity captured by RA remains largely unpacked here. Theoretically, one could conduct a detailed comparison of higher-way connectivity during each period to identify which providers were gaining or losing importance in the network, and which specific billing patterns increased and decreased in frequency. But changes in RA probability distributions are rather nuanced and can be difficult to interpret.\(^\text{37}\)

All other methods either ignored or actively excluded multi-way relationships, so that they could infer the network in a strictly pairwise manner. These approaches have their own merits, as a reduction to pairwise connections can afford more accessible

\(^{37}\)One example of a change in RA probabilities would be an increase in the probability that a patient is billed for once by provider ‘A’, zero times by provider ‘B’, two or more times by provider ‘C’, etc., which is not straight-forward to interpret.
displays of overlap in patient sets and change over time. Among the three linear modeling methods applied here, PLS appeared to be the most effective method at distilling pairwise relationships from multiply-shared billing patterns. It found the most change in terms of connections appearing and disappearing, and its use of latent terms allowed it to account for more variance in provider behavior patterns than a simple multiple regression could.

Even if RA and PLS had the most interesting results, each of the other methods provided valuable secondary insights. Distributions of correlation and multiple regression weights both helped to validate the notion that consolidation was occurring in multi-way billing patterns. The patterns originally detected by RA are better understood after using these other methods to verify that a central core of providers developed stronger overall correlations (which include multi-provider patterns), and simultaneously weaker unique relationships (which exclude them). Together, these results point toward patients’ inclination to form smaller billing provider networks that still contained multiple providers and aligned with a central core within the top 30 network.

Finally, the absence of evidence for consolidation by SBT revealed that correlation’s evidence for consolidation was likely dependent on the use of continuous data. When restricting a method to pairwise relationships and using data in a binned, nominal form, evidence for consolidation disappeared. This suggests that the detection of consolidation is dependent on a method’s capacity either to detect higher-way relationships or at least to use data in continuous (or frequency) form.

Ultimately, this analysis of the top 30 providers points to RA as the best method for inferring higher-way billing patterns, and PLS as the best method for inferring unique
pairwise patterns. Contributions of the other methods were more indirect, in triangulating the general observation that multiway connectivity was consolidating. That said, all of the methods were useful to test the hypothesis that connectivity increased over time, and when results differed, all of the methods were useful to identify alternative patterns (i.e., consolidation) that could account for the discrepancies between results.

Selection of a method for network inference

When selecting a method to infer networks in other systems, one might consider whether network behavior can be represented reasonably well by nominal variables, and whether nonlinear and higher-way relationships are of interest. If nominal data representation is reasonable, and if these types of relationships are of interest, then RA provides a great advantage for network inference. In the top 30 network of billing providers, RA detected both nonlinear and higher-way relationships, and revealed a large degree of complexity in the healthcare network. It also appeared to detect the most change in the network of any method (with change in 47% of its provider pairs, as shown in Table 5.6, p.172).

If higher-way associations are too difficult to interpret, but nonlinear relationships are still expected, SBT could be a reasonable choice for network inference. In the top 30 network, however, SBT was not found to be directly insightful, and mostly served as a theoretical halfway point between correlation and RA. In the context of this project, SBT revealed that correlation’s evidence for consolidation of connectivity disappeared once frequency data was discretized, and RA’s evidence for consolidation of connectivity
disappeared once connections were forced to pairwise. This is indirectly very useful, but SBT results on their own may not have provided many insights.

If nonlinear relationships are not expected, or if nominal categories are not a reasonable representation of the data, then linear modeling is more appropriate, and one should consider whether unique or comprehensive relationships are desired. In the top 30 network of billing providers, correlations provided a more inclusive view of the relationships between billing providers, reflective of all patients that were overlapping between them. However, because many patients had overlap between many providers during both time periods, the correlation network showed relatively little change. If the interest is on purely pairwise aspects of a network, it may be wise to focus on unique relationships between members. This filters out what might be considered noise in a pairwise network – multiply-shared patterns that would be redundantly captured among many pairs – and may better identify the network features that change over time.

The choice between multiple regression and PLS comes down to an interest in using latent terms. The potential advantages of this were illustrated by the top 30 network, as PLS captured more provider variance in unique relationships than multiple regression did, and it found more change in the network than multiple regression or correlation. However, the amount of change that PLS captured was only barely larger than multiple regression, and PLS connections are more difficult to interpret. In PLS, a relationship between two network members reflects the extent to which one member’s behavior can contribute to latent terms that uniquely co-vary with the other network members’ behavior. This is much more abstract than simply calculating the extent to
which one member’s behavior uniquely predicts another’s. In some situations, the goal may be simply to identify the most prominent areas of network change, and PLS (or RA) may be ideal. But if it’s important to readily evaluate the nature of specific network changes, or to describe them in lay terms, multiple regression may be preferable.

**Network 3: The Care Sector Network**

The third network was a care sector network, with six nodes to represent primary, ancillary, specialty, mental/behavioral health, facility, and ‘other’ areas of billing in the healthcare system. The care sector network was inferred with RA and PLS, based on results from the Top 30 network that suggested these were the most directly informative methods by I-DNA and R-DNA, respectively. It was also inferred with correlation and SBT methods – the simplest forms of I-DNA and R-DNA – to provide some additional useful context. (This left multiple regression as the only method *not* used to infer the care sector network.)

There were two main hypotheses for the care sector network. The first (H₁) was that the relationship between primary and specialty care would ‘weaken’ after Health Share formed (i.e., that it would become less strongly positive). This hypothesis was informed by Health Share’s adoption of a Patient Centered Primary Care Home (PCPCH) model of care, which – among other things – involves assigning each patient to a primary care provider (PCP) and requiring that provider to direct the patient’s use of specialty care. Under the PCPCH, patients were expected to use primary care more predominantly and specialty care more sparingly, leading to less overlap in patients and weakening the relationship between the two sectors. As will be shown below, we found select evidence
to support this hypothesis, but the overwhelming pattern suggested that Health Share’s formation was followed by an increase in patients who did not have any claims billed for them by primary or specialty billing providers.

The second hypothesis for the care sector network (H2) was that the relationship between primary care and mental/behavioral health care would strengthen (i.e., that it would become more strongly positive). This was based on an expectation that the integration of mental health services in Health Share’s formation would lead to more mental health patients being referred for primary care services. More overlap in patient sets should have increased the strength of the relationship between mental health and primary care sectors. However, most results (detailed in subsequent sections) indicated that the relationship between these two sectors weakened after Health Share formed, caused by a smaller proportion of mental health patients who used primary care services.

Only one method, PLS, provided evidence of strengthening, in that case pointing to a unique predictive relationship from primary care to mental/behavioral care, after multi-provider billing patterns were filtered out and when remaining pairwise patterns were optimized through the use of latent terms.

We note in retrospect that both main hypotheses presumed connections would be ‘positive’ before Health Share’s formation, so our use of the term ‘weakening’ implied a relationship becoming less strongly positive, and our use of the term ‘strengthening’ implied a relationship becoming more strongly positive. A clearer statement of the hypotheses would have allowed for the possibility that the relationships between key sectors might have started out negative by some methods. If primary and specialty care
had started out negatively associated, our expected changes by the PCPCH model should have actually been for the connection to ‘strengthen’ by becoming increasingly negative. And if primary and mental/behavioral care sectors were negatively associated before Health Share’s formation, our expectation of increased referrals should have technically ‘weakened’ the relationship by moving it toward zero.

Fortunately for our interpretation, R-DNA inference methods found all health care sectors to be positively associated during both time periods, as patients receiving care in one sector were generally more likely to receive care in other sectors as well. This meant that in our applications of R-DNA, strengthening and weakening became synonymous with connections becoming ‘increasingly positive’ and ‘increasingly negative’. For RA and SBT methods, additional analyses were used to connect results with our hypotheses. Network inference by SBT does not differentiate between positive and negative types of connections, so contingency tables were needed to better evaluate the nature of changes that occurred. And the inference of higher-way connections by RA does not translate neatly into pairwise connection weights, so higher-way patterns were studied in greater detail to characterize the changes that occurred within them. These follow-up analyses will be described in the sections below.

In addition to the main network hypotheses, there were several smaller hypotheses assessed even though they did not translate neatly into a specific network effect. At the patient level, PCPCH adoption was expected to result in patients’ utilization of fewer primary and specialty care providers, on average. And, at the sector level, PCPCH adoption was expected to lead to decreased rates of specialty care utilization and
increased rates of primary care utilization. Rates of mental/behavioral health care utilization were expected to remain stable, because the integration of mental health into the CCO was not expected to increase the need for mental health and behavioral services so much as it was expected to increase the likelihood that mental health patients would get established with a PCP. Support was found for many of these secondary hypotheses, especially the reduction in specialty care utilization and lower counts of primary and specialty providers per patient.

**Care Sector Network Inference**

In order to infer the care sector network with I-DNA and R-DNA, datasets were defined for each time period with one row for each patient and with one column for each type of care (e.g., primary, ancillary). The value in each cell indicated the number of claims a patient had with all billing providers of that type (during that time period). For I-DNA analyses only, claim frequencies were then binned into three nominal categories, just as they were for the top 30 network. Each frequency was classified as representing 0 claims, 1 claim, or 2+ claims for a patient with billing providers of that type. Figure 5.26 illustrates the distribution of patients’ claim frequencies in each nominal category. For example, most patients (76-77%) had 0 ancillary claims, and among those with ancillary claims about half had 1 claim and the other half had 2 or more claims.
After datasets had been prepared, the care sector network was inferred with four methods, including two R-DNA methods (PLS and correlation) and I-DNA methods (RA and SBT). RA and PLS methods were selected based on results from the Top 30 network analysis in which those two methods were found to detect the most change over time. Correlation\textsuperscript{38} and SBT, which are the simplest forms of R-DNA and I-DNA, were selected to add additional context. This left multiple regression as the only method not applied. See Figure 5.27 – Figure 5.30 for networks inferred by each method.

\textsuperscript{38} Spearman (rank-order) correlations were used here because of their fitness for skewed and count data.
**Change in Correlation Network**

- **AUXILIARY**
- **MENTAL HEALTH**
- **SPECIALTY**
- **OTHER**

**Change in SBT Network**

- **ANCILLARY**
- **MENTAL HEALTH**
- **SPECIALTY**
- **OTHER**

**Change in PLS Network**

- **ANCILLARY**
- **MENTAL HEALTH**
- **SPECIALTY**
- **OTHER**

- **RA Network (no change)**

- **ANCILLARY**
- **MENTAL HEALTH**
- **SPECIALTY**
- **OTHER**

- □ connection decreased in strength  □ connection increased in strength  □ no change

* Note that connection weights are calculated as the weight ‘after’ minus the weight ‘before’.

Figure 5.27. Change as inferred in the care sector network
Figure 5.28. Care sector network before and after Health Share’s formation as inferred through correlation
Figure 5.28. Care sector network before and after Health Share’s formation as inferred through SBT
Figure 5.29. Care sector network before and after Health Share’s formation as inferred through PLS.
For the care sector network we did not calculate the change in average connection strength or count the number of changing connections (as was done in the top 30 network), because we did not have hypotheses related to those metrics. Network distance, however, was still calculated for each method (see Table 5.7) because it represents the total amount of network change and will be tested for statistical significance in Chapter 6. Note that while the RA network did not change in structure, indicating the same five-way and a four-way associations during each time period (see Figure 5.27), it did change its calculated probability distribution by a considerable amount. The mean absolute difference by RA ($\theta_{RA}$) was 0.15538, which is about 7.8% of the maximum change possible (0.15538/2.0) by that method.

<table>
<thead>
<tr>
<th>Method</th>
<th>Network Distance ($\theta$)</th>
<th>Maximum possible</th>
<th>% of Change Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation</td>
<td>0.04084</td>
<td>Out of 2</td>
<td>2.042%</td>
</tr>
<tr>
<td>PLS</td>
<td>0.01823</td>
<td>Unknown (≥ 2)</td>
<td>Unknown (≤0.912%)</td>
</tr>
<tr>
<td>SBT</td>
<td>0.01203</td>
<td>Out of 1</td>
<td>1.203%</td>
</tr>
<tr>
<td>RA</td>
<td>0.15538</td>
<td>Out of 2</td>
<td>7.769%</td>
</tr>
</tbody>
</table>

The amount of change by all other networks was considerably smaller. Change by correlation ($\theta_{corr}$) was about 2.0% of its total possible (0.04084/2) and change by SBT ($\theta_{SBT}$) was about 1.2% of its total possible. The amount of change by PLS is difficult to standardize, because the theoretical maximum for PLS coefficients (and therefore change in contributions to them) is unknown. But a maximum can be expected to be at least 2.0, since coefficients might have changed from at least -1 to at least +1 or vice versa. So total change in the PLS network appears to have been no greater than 0.9% of the total possible. The fact that RA found so much more change than the other methods suggests
that a lot of change in the care sector network involved multi-way relationships between sectors that were irreducible to pairwise relationships.

Care Sector Network by Reconstructability Analysis

The following sections describe the chief findings by each analysis method, followed by a specific discussion of results that pertained to each of the two main hypotheses. The best models by RA were very similar during both time periods, and they both had the same general structure with a four-way relationship between primary, specialty, facility and other care, and by a five-way relationship between primary, specialty, facility, ancillary, and mental/behavioral care. Despite the consistency in model structure, however, RA models did indicate a number of changes in their calculated probabilities.

The 20 largest probability changes found by RA are listed in descending order in Table 5.8. Each probability contains a set of states for each sector, for example the probability that a patient might have 2+ facility, primary, and specialty claims while having zero ancillary or mental health claims (see row 1). In general, the probabilities that decreased involved some combination of primary, specialty, and facility claims (e.g., rows 1, 4, 8, 9, 12, 18, and 19, shaded light blue). And the probabilities that increased tended to involve only one of those types of claims (primary or specialty or facility) without having a combination of those types (see rows 2, 6, 10, 11, 13, 14, 16, and 17, shaded peach).
Table 5.8. Largest changes in calculated probabilities for care sector network as inferred by RA

<table>
<thead>
<tr>
<th>Rank</th>
<th>Δ prob</th>
<th>Direction</th>
<th>State description*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.01827</td>
<td>decrease</td>
<td>2+ facility, 2+ primary, 2+ specialty</td>
</tr>
<tr>
<td>2</td>
<td>0.01264</td>
<td>increase</td>
<td>2+ primary</td>
</tr>
<tr>
<td>3</td>
<td>0.00848</td>
<td>increase</td>
<td>No claims of any type</td>
</tr>
<tr>
<td>4</td>
<td>-0.00650</td>
<td>decrease</td>
<td>2+ facility, 2+ primary, 1 specialty</td>
</tr>
<tr>
<td>5</td>
<td>-0.00537</td>
<td>decrease</td>
<td>2+ primary, 2+ specialty</td>
</tr>
<tr>
<td>6</td>
<td>0.00501</td>
<td>increase</td>
<td>1 primary</td>
</tr>
<tr>
<td>7</td>
<td>-0.00475</td>
<td>decrease</td>
<td>2+ facility, 2+ primary</td>
</tr>
<tr>
<td>8</td>
<td>-0.00402</td>
<td>decrease</td>
<td>2+ facility, 1 primary, 2+ specialty</td>
</tr>
<tr>
<td>9</td>
<td>-0.00401</td>
<td>decrease</td>
<td>1 facility, 2+ primary, 2+ specialty</td>
</tr>
<tr>
<td>10</td>
<td>0.00341</td>
<td>increase</td>
<td>1 ancillary, 2+ primary</td>
</tr>
<tr>
<td>11</td>
<td>0.00310</td>
<td>increase</td>
<td>2+ facility</td>
</tr>
<tr>
<td>12</td>
<td>-0.00277</td>
<td>decrease</td>
<td>2+ facility, 2+ mental, 2+ primary, 2+ specialty</td>
</tr>
<tr>
<td>13</td>
<td>0.00275</td>
<td>increase</td>
<td>2+ ancillary, 2+ primary</td>
</tr>
<tr>
<td>14</td>
<td>0.00267</td>
<td>increase</td>
<td>1 facility</td>
</tr>
<tr>
<td>15</td>
<td>0.00265</td>
<td>increase</td>
<td>1 facility, 2+ primary</td>
</tr>
<tr>
<td>16</td>
<td>0.00233</td>
<td>increase</td>
<td>2+ specialty</td>
</tr>
<tr>
<td>17</td>
<td>0.00196</td>
<td>increase</td>
<td>2+ mental, 2+ primary</td>
</tr>
<tr>
<td>18</td>
<td>-0.00187</td>
<td>decrease</td>
<td>2+ ancillary, 2+ facility, 2+ primary, 2+ specialty</td>
</tr>
<tr>
<td>19</td>
<td>-0.00186</td>
<td>decrease</td>
<td>1 ancillary, 2+ facility, 2+ primary, 2+ specialty</td>
</tr>
<tr>
<td>20</td>
<td>0.00176</td>
<td>increase</td>
<td>1 ancillary</td>
</tr>
</tbody>
</table>

*Care sectors not listed contained zero claims. For example, the first row had zero ancillary, zero mental health, and zero ‘other’ claims.

General patterns in Table 5.8 suggest that after Health Share formed, people became less likely than before to get care from all three of these sectors (primary, specialty, and facility). And they became more likely than before to get care from only one of those sectors. Even though primary, specialty, and facility sectors remained involved in a four-way and five-way association after Health Share’s formation, they had lower probabilities of an all-way overlap in patients they saw, and their relationship changed in such a way that they served increasingly distinct patient sets. This provides some initial evidence in support of the first main hypothesis that the relationship between primary and specialty care sectors would weaken over time.
Table 5.9 includes attributes of the care sector network as inferred by RA.

Networks during both time periods had the same model structure, which means they were found at the same search level and had the same number of degrees of freedom. They captured almost the exact same amount of information during both time periods. And their structure was identical, suggesting that the care sector network continued to be accounted for equally well by the same four-way and five-way relationships. The changes that occurred, then, were nuanced changes within those multiway relationships, for example in allowing for more distinct patient sets among primary, specialty and facility sectors.

<table>
<thead>
<tr>
<th>Table 5.9. Summary of best models for the care sector network inferred by RA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Before</strong></td>
</tr>
<tr>
<td>Entropy in data</td>
</tr>
<tr>
<td>Entropy in best model</td>
</tr>
<tr>
<td>Change in degrees of freedom*</td>
</tr>
<tr>
<td>Search level where best model was identified</td>
</tr>
<tr>
<td>Percent of information captured by best model</td>
</tr>
</tbody>
</table>

* Change is in reference to the independence model, as one moves toward a fully connected model (i.e., the data). Less change indicates a simpler (less complex) model structure.

Care Sector Network by Correlation

Most changes in the correlation network indicated weakened connectivity (shown by blue lines in Figure 5.27). The two biggest pairwise changes were decreases in the relationship between primary and ancillary care, which dropped by 0.099, and between primary and mental/behavioral care, which dropped by 0.085. Both of these decreases are especially interesting when recalling that there were increases in the number of patients.

---

39 Note that virtually all correlations were positive, so a decrease is equivalent to a weakening, where a correlation moves closer to zero (and not closer to negative 1).
receiving ancillary (12.45%) and mental health care (4.88%), while the number of patients receiving primary care remained basically the same (increasing by only 0.34%, see Table 4.7 on p.115). This suggests that increased utilization of ancillary and mental health services was driven by patients not concurrently receiving primary care.

There were few correlations that strengthened, and only two that strengthened by more than 0.01. These were the links between primary and other care (increasing by 0.013) and between primary and specialty care (increasing by 0.036). The strengthening between primary and specialty care is especially surprising, given that this relationship was hypothesized to weaken. (This will be discussed more later.) Absolute correlations during each period, and the amount of change over time are depicted below in Figure 5.30. In addition to the two largest changes described above, there were also noticeable decreases in the correlation between ancillary and mental health (by 0.068), between facility and mental health (by 0.060), and between ancillary and specialty care (by 0.054).
Figure 5.30. Pairwise connections by correlation, before and after Health Share formed Care Sector Network by Standardized Bivariate Transmission

Like with the top 30 network, pairwise connections inferred by SBT were nearly all very small in magnitude, as were the changes that occurred among them. A slight majority of pairs (eight out of fifteen) changed by less than 0.001, and most of the remaining pairs changed by hardly more than that. This reiterates the idea that much of the change in the care sector network occurred in higher-way relationships between sectors and was not detectable at a pairwise level. This looks to be especially evident when binning data into nominal levels (0, 1, or 2+ claims), such that patterns among higher levels of utilization are collapsed into a single level (2+ claims) and effectively hidden from view.
Figure 5.31. Pairwise connections by SBT, before and after Health Share Formed

SBT values during each time period, and the amount of change over time, are depicted above in Figure 5.31. The largest change was an increase between facility and specialty care (by 0.002). Before Health Share formed, this was already by far the strongest pairwise relationship in the network (SBT=0.023), and it became even stronger afterward (SBT=0.025). All other connections were less than 0.005 during both periods, and the only changes observed among them were slight decreases (by 0.001). These decreases were between mental health and each of ancillary, specialty, and facility care, as well as between ‘other’ care and each of facility and specialty care.

Care Sector Network by Partial Least Squares Regression

The PLS network is distinct from the others, in that it infers two relationships for each pair of sectors, one in either direction. All PLS connection weights (|s|) during each
time period, and the amounts of change over time, are depicted below in Figure 5.32. The two largest changes, by far, were decreases in the connection from ancillary to mental/behavioral (by 0.292), and from specialty to mental/behavioral (by 0.074). Interestingly, neither of those changes were reciprocated. The connection from mental/behavioral to ancillary decreased only slightly (by 0.008) and the connection from mental/behavioral to specialty care actually increased slightly (by 0.006).

Most connections in the PLS network were smaller than their correlation counterparts, despite PLS weights being able to take on values larger than the theoretical maximum of a correlation (1.0). PLS connection weights were also less evenly distributed than correlation weights, with a greater disparity between the strongest few connections and many weak connections. This difference is apparent when comparing the distribution of bar lengths in Figure 5.30 (for correlation) and Figure 5.32 (for PLS). And it suggests, again, that much of healthcare delivery patterns envelope multi-way relationships between care sectors. Billing patterns that were common to multiple care sectors could be captured (doubly) across pairwise correlations, but were filtered out of PLS connections, leaving a network that looks less connected overall. This allowed for less change to be observed in the connections that pertained to our two care sector network hypotheses.
Table 5.10 provides a summary of changes regarding which pairs of sectors were connected by each of the four inference methods applied. Because there were only six nodes in this network, we did not really apply any visualization threshold besides what was inadvertently applied because edge weights were only calculated out to five decimals. This meant that virtually all care sectors were connected to one another during
both time periods, and that there was very little change between time periods regarding which sectors were connected.

Table 5.10. Summary of Changes in Care Sector Network Visualizations

<table>
<thead>
<tr>
<th>Method</th>
<th>RA</th>
<th>Correlation</th>
<th>SBT</th>
<th>PLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>All pairs found Before</td>
<td>13</td>
<td>15</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>All pairs found After</td>
<td>13</td>
<td>15</td>
<td>14</td>
<td>15</td>
</tr>
<tr>
<td><strong>Total edges in either period</strong></td>
<td>13</td>
<td>15</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>Pairs found in both periods</td>
<td>13 (100%)</td>
<td>15 (100%)</td>
<td>14 (93%)</td>
<td>15 (100%)</td>
</tr>
<tr>
<td>Pairs found in one period</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>1 (7%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Pairs found only before</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>1 (7%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Pairs found only after</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td><strong>Total Network Distance (θ)</strong></td>
<td>0.15538</td>
<td>0.04079</td>
<td>0.00054</td>
<td>0.01807</td>
</tr>
</tbody>
</table>

On the Relationship between Primary and Specialty Care (H₁)

Across the methods applied here, most evidence countered our first care sector network hypothesis, that the relationship between primary and specialty care would weaken over time (i.e., become less strongly positive). Recall that this expectation stemmed from our understanding of the Patient Centered Primary Care Home (PCPCH) model of care, which required patients to get approval from a primary care provider (PCP) before accessing specialty care. Under the PCPCH model, PCPs were expected to direct patients to more appropriate specialty care, and in some cases prevent patients from using unnecessary specialty care. Patients who might have used many primary and specialty care services were expected to use less specialty care and rely more solely on primary care. This should have weakened the relationship between those two sectors.

However, the actual changes observed went counter to those expectations. There was basically no change observed by SBT (or RA) – which found the same amount of transmission between primary and specialty care sectors during both time periods. And
correlation and PLS methods found evidence that this relationship actually *strengthened*. The correlation increased by 12.5%, from \( r = 0.288 \) to 0.323. And PLS regression found increases both in specialty care’s unique prediction of primary care utilization (from \( B = 0.391 \) to 0.417, a relative increase of 6.6%), and in primary care’s unique prediction of specialty care utilization (from \( B = 0.173 \) to 0.186, a relative increase of 7.5%).

To better understand these counterintuitive findings, contingency tables were constructed to compare the proportions of patients who had one of several levels of utilization within each healthcare sector: 0, 1, 2-3, or 4+ claims. This table (Table 5.11) provides a modest amount of evidence for expected patterns of change, and it also illustrates why the overall correlation strengthened instead of weakening as expected. First, the proportion of people with multiple claims for both care types did decrease as expected: from 26.7% to 23.4% (the sum of the values in the bottom right of each table, in red font). And the number of people who received only primary care increased as expected, from 37.5% to 39.0% (the sum of the values at left, in purple font). These changes, in isolation, would have weakened the relationship between primary and specialty care, as we expected. But their impact was dwarfed by a more dramatic change in the network, the increase in people who did not receive either type of care. That proportion jumped from 5.8% to 10.4%, nearly doubling in size.
Table 5.11. Patient utilization of primary and specialty care

<table>
<thead>
<tr>
<th>Before Health Share</th>
<th>Specialty care utilization (# of claims)</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Primary care utilization (# of claims)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>5.8%</td>
<td><strong>3.2%</strong></td>
</tr>
<tr>
<td>1</td>
<td>11.7%</td>
<td>2.5%</td>
</tr>
<tr>
<td>2-3</td>
<td>12.9%</td>
<td>4.1%</td>
</tr>
<tr>
<td>4+</td>
<td>12.3%</td>
<td>6.2%</td>
</tr>
<tr>
<td></td>
<td>43.3%</td>
<td>16.0%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>After Health Share</th>
<th>Specialty care utilization (# of claims)</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Primary care utilization (# of claims)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>10.4%</td>
<td><strong>2.9%</strong></td>
</tr>
<tr>
<td>1</td>
<td><strong>12.7%</strong></td>
<td>2.3%</td>
</tr>
<tr>
<td>2-3</td>
<td><strong>13.6%</strong></td>
<td>3.7%</td>
</tr>
<tr>
<td>4+</td>
<td><strong>12.7%</strong></td>
<td>5.8%</td>
</tr>
<tr>
<td></td>
<td>49.5%</td>
<td>14.7%</td>
</tr>
</tbody>
</table>

In viewing the tables above, consider a strong linear relationship to be one with a high degree of alignment along the diagonal, starting from the top left and moving toward the bottom right. A perfect correlation between primary and specialty care utilization would have all patients situated along the diagonal, either with 0 claims from either sector, 1 claim from both, 2 claims from both, etc. A weaker relationship, then, would diverge from this pattern, with more people situated outside of that diagonal, either above or below it. In our hypothesis for primary and specialty care, we expected Health Share’s formation to result in fewer patients along the diagonal (i.e., fewer high utilizers of both types of care), and more patients below the diagonal (i.e., more patients predominantly using primary care). This would have weakened the correlation.

With that notion of a linear relationship in mind, consider Table 5.12 below, which has subtracted the amount of change in each percentage after Health Share’s
formation. Of central importance, note that the proportion of patients along the diagonal actually increased (by a net gain of 1.95%), rather than decreasing as expected. Because percentages must add to 100, a net increase along the diagonal equates to a net decrease in off-diagonal values. This global shift, from off-diagonal proportions to on-diagonal proportions, is at the crux of why the correlation strengthened.

Table 5.12. Change in patient utilization of primary and specialty care

<table>
<thead>
<tr>
<th>Change (% After—% Before)</th>
<th>Specialty care utilization (# of claims)</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary care utilization (# of claims)</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>4.59%</td>
<td>-0.31%</td>
</tr>
<tr>
<td>1</td>
<td>0.80%</td>
<td>-0.25%</td>
</tr>
<tr>
<td>2-3</td>
<td>0.54%</td>
<td>-0.38%</td>
</tr>
<tr>
<td>4+</td>
<td>0.20%</td>
<td>-0.43%</td>
</tr>
</tbody>
</table>

An important thing to note, however, is that the net increase in diagonal percentages was driven entirely by the increase in patients who received 0 claims from either care sector, and it strengthened the overall relationship despite decreases in patients receiving 1 of each, 2-3 of each or 4+ claims of each type of care. So, several reductions along the diagonal occurred as hypothesized, in that Health Share’s formation was followed by fewer patients with multiple claims in both sectors. The unexpected pattern was the large increase in patients who received neither type of care.

A couple of additional observations can be made, in changes that occurred above and below the diagonal. Health care patterns above the diagonal (in white cells in Table 5.12) decreased by a total of 1.94% after Health Share formed, indicating fewer patients who predominantly used specialty care. This type of effect also went as hypothesized, as
PCPs were expected to direct (and sometimes reduce) patients’ use of specialty care services. Additional support can be seen in health care patterns below the diagonal, because Health Share’s formation was followed by an increase in the number of patients with only primary care claims (by 1.54%, see the sum of white numbers in the leftmost column of Table 5.12).

However, despite select patterns in support of the first hypothesis, the general direction of change went counter to expectations. The reduction in patients receiving predominantly specialty care (above the diagonal), or equal amounts of primary and specialty care (along the diagonal), was not met with an increase in patients receiving predominantly primary care (below the diagonal), which would have decreased the strength of a linear relationship. Instead, those reductions shifted overwhelmingly to an increase in the number of patients who did not receive either type of care after Health Share formed.

It is worth considering whether patients who did not receive care from either the primary or specialty sector should be given so much emphasis in our interpretation of the change in this connection. After all, if we were to restrict Table 5.12 (p.212) to only look at patients who received care from at least one of these two sectors (primary or specialty), we would see the types of change we anticipated – a general increase in patients reliance on primary care relative to specialty care. Those types of patterns would also align better with our general observations of sector-level trends from Chapter 4, where we noted that the proportion of patients receiving primary care increased by 0.34% after Health Share’s formation and the proportion receiving specialty care decreased considerably by 6.90%
(see Table 4.7, p.115). An optimistic interpretation of these results would be that a larger proportion of patients did not need either primary or specialty care after Health Share’s formation, and among patients who needed care from either or both sectors we saw a greater reliance on primary over specialty care.

It is also worth considering whether other aspects of the PCPCH model of care could have prevented us from detecting the types of connection-level changes we anticipated. In addition to PCP assignments and PCP referrals to specialists, the PCPCH model involves coordination and integration of different health care sectors to achieve comprehensive whole-person care. In many cases, this may have resulted in the integration of primary care services at other types of clinics. Theoretically, patients may have been assigned to a PCP at a specialty or mental/behavioral clinic and begun receiving primary care services at that location. These types of organization-level adaptations that followed CCO formation may have obscured or confounded the types of changes we expected, based on how we defined the care sector network for this project.

Despite these nuances, our general finding still calls into question some aspects of Health Share’s impact, and specifically its successful utilization of PCPs in the PCPCH model of care. Our expectation was for PCPs – in the primary care sector – to be assigned to each patient and heavily involved in referring them to specialty care. And while evidence suggests this may have been achieved in some cases, a more predominant pattern was the increase in the number of people who were did not appear to receive specialty nor primary care in this network. If patients were not assigned to PCPs in other care sectors, then it would seem possible that the assignment and the requirement for PCP
approval may have created obstacles to access. If patients seeking specialty care were not successfully redirected to a PCP, they may have instead avoided or postponed both primary and specialty care utilization. This possibility warrants further research.

On the Relationship between Primary and Mental/Behavioral Care (H2)

Methods varied in their results for the second connection-level hypothesis (H2), that primary and mental/behavioral sectors would become more strongly connected over time. This hypothesis was based on an expectation that the integration of mental health services into Health Share’s CCO would lead to more mental health patients being referred for primary care services. More overlap in patient sets would have strengthened the relationship between these two sectors. However, most general patterns (shown below) suggested that this relationship weakened over time, even if it may have strengthened for select sets of patients. A smaller proportion of mental health patients used primary care services after Health Share formed (decreasing from 84.02% to 73.13%, see Table 5.13 below), suggesting that referrals to primary care may not have happened as much as expected, or that fewer patients completed referrals by the end of the 12-month post-period.

Table 5.13. Proportions of patients receiving mental health and primary care

<table>
<thead>
<tr>
<th>Mental Health Claims</th>
<th>Before</th>
<th>After</th>
<th>Primary Claims</th>
<th>Before</th>
<th>After</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>84.82%</td>
<td>86.19%</td>
<td>0</td>
<td>15.98%</td>
<td>26.87%</td>
</tr>
<tr>
<td>1</td>
<td>2.08%</td>
<td>2.00%</td>
<td>1</td>
<td>10.32%</td>
<td>10.38%</td>
</tr>
<tr>
<td>2-3</td>
<td>2.23%</td>
<td>2.02%</td>
<td>2-3</td>
<td>18.37%</td>
<td>17.29%</td>
</tr>
<tr>
<td>4+</td>
<td>10.87%</td>
<td>9.80%</td>
<td>4+</td>
<td>55.33%</td>
<td>45.46%</td>
</tr>
<tr>
<td>Any</td>
<td>15.18%</td>
<td>13.81%</td>
<td>Any</td>
<td>84.02%</td>
<td>73.13%</td>
</tr>
</tbody>
</table>
Before delving into the results from each specific method, it is worth considering some contextual factors that could have affected these global trends in care delivery. As we discussed earlier in our dissection of change between primary and specialty sectors, there are some general aspects of CCO formation and PCPCH implementation that may have expanded the types of care that were delivered in specific clinics. Integration of mental and behavioral health services into Health Share’s CCO, along with the encouragement of CCO clinics to work toward PCPCH certification, likely facilitated the incorporation of primary care services into mental and behavioral clinics, and incorporation of primary care into mental/behavioral health clinics. If this integrative care model were successful, patients would appear to move less often between primary and mental/behavioral care sectors at the billing level, and might very easily result in the patterns we observed above in Table 5.13.

With that context in mind, we can consider the results from each network inference method. Not all methods refuted our second hypothesis, and PLS and SBT provided modest evidence for the relationship strengthening. But, given the general trends outlined above, results from correlation analysis are the most intuitive. The correlation network found a substantial reduction in the relationship between mental/behavioral and primary care after Health Share’s formation, from $r = 0.118$ to 0.033. This reduction (by .085) was the second largest change observed in the correlation network (see Figure 5.30), following closely behind ancillary and primary care, which reduced by .099. It was also the single largest relative change observed in the network, constituting a reduction of that relationship’s strength by nearly 72%. So, the change
inferred through correlation was not only opposite of the hypothesized direction of change but also strongly so.

Results from the SBT network look at face value to be in contrast with correlation, because they suggest that the relationship between mental health and primary care strengthened over time, as we hypothesized, increasing from 0.00060 to 0.00090 (see Figure 5.31). However, as we dig into the patterns detected by SBT, we’ll find that its detection of increased constraint was driven by patterns which reiterate that these health care sectors served increasingly distinct patient sets. The nature of this strengthening, then, may not be in-line with the hypothesized direction of change.

Table 5.14 below shows the pairwise contingency table that was evaluated by SBT. Recall that SBT and RA require frequency counts to be discretized, so patients have been recoded as having 0, 1, or 2+ claims in each health care sector. In this table, observed probabilities (in large, bold font) can be compared to the probabilities that we would expect if the two health care sectors operated independently. The difference between each observed and expected probability is listed directly below the observed probability, either in red (when less than expected) or in blue (when greater than expected). Probabilities on the left-hand side of the table are from before Health Share’s formation (in blue-shaded cells), and probabilities on the right-hand side of the table are from after Health Share’s formation (in green-shaded cells).
In comparing the tables from before and after Health Share’s formation, we can observe that probabilities moved further from independence. The sum of absolute differences from expected probabilities (i.e., the non-bolded values) increased after Health Share’s formation, from 4.662% to 4.853%. This indicates that there was a larger gap between the probabilities that were observed and what we would have expected if the sectors operated independently. This is not technically how SBT is calculated, but this essential pattern is what allowed SBT to detect a strengthening in this relationship over time.

If we look more closely at Table 5.14, however, we can identify the specific cells where care patterns changed most dramatically. The largest increases in probability, relative to independence (in non-bolded values), were in the proportion of people with either (A) 2+ mental health claims and 0 primary care claims or (B) 2+ primary care claims and 0 mental health claims. The proportion of patients having those combinations increased beyond what one would have expected if the sectors were independent, pushing the observed probabilities away from independence by 1.03% and 1.12%, respectively.

<table>
<thead>
<tr>
<th>Primary Care</th>
<th>Before</th>
<th></th>
<th></th>
<th>After</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mental Health</td>
<td>Sum</td>
<td></td>
<td>Mental Health</td>
<td>Sum</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1</td>
<td>2+</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>13.30%</td>
<td>0.34%</td>
<td>2.09%</td>
<td>15.7%</td>
<td>14.88%</td>
</tr>
<tr>
<td>1</td>
<td>16.40%</td>
<td>0.23%</td>
<td>1.34%</td>
<td>18.0%</td>
<td>16.62%</td>
</tr>
<tr>
<td>2+</td>
<td>55.07%</td>
<td>1.52%</td>
<td>9.70%</td>
<td>66.3%</td>
<td>53.20%</td>
</tr>
<tr>
<td>Sum</td>
<td>84.8%</td>
<td>2.1%</td>
<td>13.1%</td>
<td>100.0%</td>
<td>84.7%</td>
</tr>
</tbody>
</table>
Conversely, the largest decreases in probability, relative to independence, were in the proportion of people with either (C) 0 claims of either type or (D) 2+ claims of both types. The proportion of patients having those combinations decreased below what one would have expected if the sectors were independent, pushing the observed probabilities away from independence by 1.16% and .98%. Patterns in the cells that increased and decreased most dramatically suggest that the ‘strengthening’ measured by SBT reflects a shift from patients receiving care at both sectors (or neither sector) toward patients receiving care from one sector or the other. This is similar to the idea of a ‘negative’ relationship between two variables, or a correlation that is increasingly inverse in nature. It refutes the essence of our hypothesis, by indicating a disassociation between primary and mental/behavioral care sectors.

In contrast to correlation and SBT, PLS pointed to a fundamental strengthening between these two care sectors. Although the relationship from mental/behavioral to primary care did not really change, weakening negligibly from $s=0.02295$ to $0.02289$ (a relative decrease of 0.3%), the relationship from primary care to mental health care increased considerably, from $s=0.01860$ to $0.04116$ (a 121% increase). This increase (by $s=0.02256$) was the PLS network’s fifth largest change by magnitude and was its very largest change in relative terms. It suggests that the use of primary care did become a stronger predictor of mental health utilization, after controlling for utilization patterns in other sectors and when allowing latent terms to maximize its unique predictive capacity.

To understand why the PLS results were so different from correlation and SBT results, it is important to remember how PLS network inference is fundamentally
different from the other two types of network inference. Recall the complexity of the care sector network, as originally inferred by RA. The RA network suggests that health care delivery is driven by four- and five-way relationships that, together, involve all healthcare sectors. Given that structure as a baseline, any efforts to define the network as a set of pairwise relationships will require some oversimplification. In reality, much of healthcare delivery by any pair of sectors is not really specific to those two sectors and is instead part of a larger and more complex healthcare delivery pattern.

Correlation and SBT can be said to basically ignore the issue of higher-way complexity, measuring the relationship between each pair of sectors as though they exist in isolation. There is benefit to including all associativity between each pair of sectors – for example, it is worth considering that Health Share’s formation was followed by a smaller proportion of mental health patients who also received primary care. But these methods cannot differentiate whether aspects of a pairwise relationship are unique to that pair or are merely reflective of larger patterns general to the network.

RA and PLS respond to this complexity in opposite (and complementary) ways. RA *captures* higher-way associations in its inference of network structure, therefore avoiding the oversimplification that would be necessary to define the network as a set of pairwise relationships. PLS – and multiple regression – do the opposite, *disregarding* all higher-way associativity and filtering each pairwise relationship down to what is unique to that pair. By disregarding patterns of association whenever they could also be accounted for by another sector, PLS and regression drill down to the unique, directed
relationships in each pair, based solely on patterns that are independent from other sectors.

It is interesting that the unique, directed relationships between primary and mental health care are so different than the full, undirected relationships as measured by correlation and SBT. General patterns pointed to less overlap between the two sectors after Health Share’s formation, but the addition of PLS results indicates that those trends may be specific to overarching patterns that involve other health care sectors as well. Once we remove those overarching patterns from the network, PLS shows primary care to dramatically increase in its prediction of mental/behavioral health utilization over time. This suggests that there may be select pathways through the healthcare delivery network, unique to primary and mental health sectors, which strengthened after Health Share’s formation. There may be subsets of patients or medical conditions for whom these two sectors of care had increased overlap in patients, and where patients were often referred between sectors.

**Summary of Care Sector Network**

The main results for the care sector network are summarized below in Table 5.15. In general, the methods used here did not provide strong support for either of the main hypotheses, which may give some cause for concern. The connection between primary and specialty care did not weaken as expected (H$_1$) and was actually found to strengthen by two of the four methods used. This strengthening stemmed from an unexpected increase, after Health Share’s formation, in the number of patients who did not receive care from either sector. Similarly, the connection between primary and mental health care
did not strengthen as expected (H_2), as Health Share’s formation was followed by a smaller proportion of patients who received both primary and mental health services. The bulk of evidence suggests that Health Share may not have led to an increase in referrals from mental health to primary care billing providers.

Table 5.15. Summary of findings for care sector network

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Inference Method</th>
<th>Before Value</th>
<th>After Value</th>
<th>After- Before</th>
<th>Percent Change</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>H_1: Weaker connection between Primary &amp; Specialty</td>
<td>RA</td>
<td>NA</td>
<td>NA</td>
<td>✓ Billing patterns showed fewer patients with primary and specialty claims; more patients with one sector or the other</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Correlation</td>
<td>0.28754</td>
<td>0.32342</td>
<td>0.03588</td>
<td>12.5%</td>
<td>✓ Strengthened</td>
</tr>
<tr>
<td></td>
<td>SBT</td>
<td>0.00344</td>
<td>0.00344</td>
<td>0</td>
<td>0.0%</td>
<td>× Did not change</td>
</tr>
<tr>
<td></td>
<td>PLS (P→S)</td>
<td>0.17341</td>
<td>0.18645</td>
<td>0.01304</td>
<td>7.5%</td>
<td>× Strengthened</td>
</tr>
<tr>
<td></td>
<td>PLS (S→P)</td>
<td>0.39094</td>
<td>0.41678</td>
<td>0.02584</td>
<td>6.6%</td>
<td>× Strengthened</td>
</tr>
<tr>
<td>H_2: Stronger connection between Primary &amp; Mental Health</td>
<td>RA</td>
<td>NA</td>
<td>NA</td>
<td>× Billing patterns did not illuminate any specific changes</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Correlation</td>
<td>0.11765</td>
<td>0.03311</td>
<td>-0.08454</td>
<td>-71.9%</td>
<td>× Weakened</td>
</tr>
<tr>
<td></td>
<td>SBT</td>
<td>0.00060</td>
<td>0.00090</td>
<td>0.0003</td>
<td>50.0%</td>
<td>✓ Strengthened (but became more inverse)</td>
</tr>
<tr>
<td></td>
<td>PLS (P→M)</td>
<td>0.01860</td>
<td>0.04116</td>
<td>0.02256</td>
<td>121.3%</td>
<td>✓ Strengthened</td>
</tr>
<tr>
<td></td>
<td>PLS (M→P)</td>
<td>0.02295</td>
<td>0.02289</td>
<td>-6E-05</td>
<td>-0.3%</td>
<td>× Weakened</td>
</tr>
</tbody>
</table>

Aside from the main hypotheses, however, there were several trends that provided modest support for Health Share changing in some of the anticipated ways. Of the subset of patients who received primary or specialty care, there was an increase in the proportion who received only primary and no specialty (see left-hand column of Table 5.12, p.212). The largest changes in RA probabilities (shown in Table 5.8, p.202) indicated less overlap between several sectors (primary, specialty, and facility), and the majority of SBT and correlation connections weakened over time, implying that patient sets became more distinct between sectors. And while many of the PLS connections increased in strength, the two largest changes by far in the PLS network were both decreases in connectivity.
It is interesting to consider the possibility that several healthcare sectors – at least as they were defined here – may have had less overlap in patients after Health Share’s formation. One might initially worry that Health Share’s integration of services and organizations resulted in a seemingly more fragmented system, where patients moved less between sectors of care. Pessimistically, it could be that Health Share’s formation created some additional barriers, such as administrative ones, that affected patients’ navigation between different sectors of the health care system. However, optimistically, this could indicate that Health Share facilitated patients to receive more seamless provisions of care (at least at the billing level), without needing as many referrals or transfers to other clinics or departments. This style of integrated care would be well in line with the PCPCH model of care.

Methodologically, our analysis of the care sector network allowed us to explore the utility of I-DNA and R-DNA for the analysis of change in specific pairwise connections. For our specific hypotheses, between primary, specialty, and mental/behavioral care sectors, the R-DNA methods were the most useful because they could directly measure whether connections became increasingly or decreasingly ‘positive’ after Health Share’s formation. SBT and RA were less ideally suited, as they define associativity in general terms without differentiating between relationships that are positive, negative, or even nonlinear in nature. And RA was perhaps the most limited, because its best model included only multi-way connections, which did not tell us anything specifically about pairwise relationships. This illustrates the importance of
selecting an inference method that aligns well with one’s research question and specific hypothesis.

Lastly, we note the utility of contingency tables, which were found to complement and extend both of our tests of pairwise change. Even in our applications of R-DNA, where measurements were a natural fit to our research question, we often sought clarification through analysis of pairwise frequency tables and through comparisons of observed and expected probabilities. These additional tools were essential to understand the types of changes that occurred in each pairwise connection and to evaluate whether those changes constituted support for our substantive hypotheses. The necessity of those additional tools suggests that our application of I-DNA and R-DNA to analysis of change in pairwise connections might best serve as an initial and exploratory step.

**Chapter 5 Summary**

This chapter defined three different networks of health care delivery, each of which provided a different vantage point to consider the impact of Health Share’s CCO formation. The full network, including 1,298 billing providers from within Health Share’s tri-county region, showed very little change over time. We hypothesized that connectivity would increase, and actually found a decrease in connectivity, although by a miniscule amount. Visualizations of the full network, however, pointed us toward an inner core of central billing providers, who were much more strongly connected to one another than they were to network’s larger, sparsely connected periphery.
The top 30 network, including the top 30 billing providers by patient volume, allowed for a deeper analysis of the full network’s core. We again hypothesized that connectivity would increase, and again found results that were largely contrary—or at least orthogonal—to what we expected. Results from the top 30 network suggested that connectivity did not increase uniformly so much as it became consolidated within its most central providers. This caused a greater disparity between the strongest and weakest connections, with fewer and stronger strong connections, and with more and weaker weak connections over time. This may reflect a more centralized restructuring of billing practices across organizations and services.

Results from the third network, of healthcare sectors, had the most direct relevance for healthcare delivery because it could indicate changes in patterns of care utilization by care type. And while some results provided limited support for hypothesized changes, such as a reduction in specialty care utilization and a greater reliance on primary over specialty care, they also give reason to question whether patients may have faced increased obstacles to access. After Health Share’s formation, the percentage of patients who did not utilize primary or specialty care nearly doubled, increasing from 5.8% to 10.4%, and the proportion of mental/behavioral patients who utilized primary care decreased from 84.0% to 73.1%. Generally speaking, there seemed to be less patient overlap between the different health care sectors after Health Share’s formation, which is a counterintuitive result following their integration in a single CCO.

Optimistically, these findings in the care sector network may indicate successful implementation of the PCPHC model of care. It is possible that the encouragement of
clinics to become certified as PCPCHs facilitated integration of care delivery within each care sector. For example, primary care clinics may have started doing routine mental health screenings and mental health clinics may have started offering more primary care services in-house. This would have meant patients did not need to transfer between sectors as often, and they might have been receiving more types of care without it being visible in our network. Pessimistically, however, it also seems possible that Health Share may not have been wholly successful in establishing patients with PCPs in the primary sector, or in aiding referrals between mental/behavioral and primary care sectors during the study period. Patients could have been technically assigned to PCPs without ever meeting them, and they may have been unable to get referrals or approval to access specialty providers. Additional research on potential obstacles seems warranted.

Many insights regarding the healthcare delivery network were made possible through comparative application of several methods of inference. In our investigation of connectivity in the top 30 network, some methods (RA and PLS) provided more direct insight than others, but all methods contributed meaningfully to our understanding of the consolidation phenomenon. And in our investigation of specific connections between primary, specialty, and mental/behavioral care sectors, a comparison of findings from several methods allowed us to build a more complete understanding of how these connections changed over time. Results by one method alone would have provided far fewer insights on how this healthcare network changed after Health Share’s formation.

Along with the substantive insights gained for CCO research, this comparative analysis has illustrated the merits of five methods of network inference. By applying each
of these methods to the same datasets, the comparison of results has allowed us to characterize the types of relationships that each method can infer, and to identify many of their pros and cons. The most notable differences found between methods centered on their treatment of higher-way patterns. These patterns were formally defined by RA, ignored by correlation and SBT, and excluded by PLS and multiple regression. Depending on one’s research question, the capture, exclusion, or ignorance of higher-way patterns could each be beneficial in its own way. But for the networks analyzed here, the greatest insights were gained through a comparative application of several methods.
Chapter 6: Statistical Testing Results

Overview

This chapter describes statistical tests of change in the top 30 and care sector networks of healthcare delivery that were inferred and described in Chapter 5. Fifty-one tests were conducted, as each network was inferred by several inference methods (e.g., correlation, reconstructability analysis), and each inference method was tested with three resampling techniques: permutation, single (one-sample) bootstrapping, and double (two-sample) bootstrapping. This large volume of statistical test results was conducted to facilitate comparisons between inference and resampling methods on their ability to detect network change. This was done in an exploratory manner, with no specific hypotheses for the relative performance of any inference or resampling methods.

An overview of statistical results is provided below in Table 6.1. All tests for change in the top 30 and care sector networks were statistically significant, indicating that all inference methods demonstrated a similar capacity to detect network change after Health Share’s formation. Within the care sector network, however, inference methods differed on whether they could detect change in the connection between primary and specialty care or between primary and mental/behavioral health care. No tests were conducted on the full network, due to computational limitations⁴⁰, or on other pairs of sectors in the care sector network (e.g., the connection between mental health and

---

⁴⁰ Statistical testing on the full network would have required computation of connections among 1,299 variables, for the original datasets as well as each of 100 pairs of resampled datasets, a process that would have taken weeks or months to calculate for most inference methods.
specialty care), as these connections were not specifically examined in Chapter 5. The choice of resampling technique did not appear to affect the statistical results in any of the tests conducted.
Table 6.1. Statistical tests of network change

<table>
<thead>
<tr>
<th>Network</th>
<th>Observed distance ($\hat{\theta}$)</th>
<th>Permutation $p$-value</th>
<th>Single Bootstrap $p$-value</th>
<th>Double Bootstrap $p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Top 30 Network</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RA$^a$</td>
<td>0.43814</td>
<td>&lt;0.01*</td>
<td>&lt;0.01*</td>
<td>&lt;0.01*</td>
</tr>
<tr>
<td>Correlation$^b$</td>
<td>0.01841</td>
<td>&lt;0.01*</td>
<td>&lt;0.01*</td>
<td>&lt;0.01*</td>
</tr>
<tr>
<td>SBT$^c$</td>
<td>0.00029</td>
<td>&lt;0.01*</td>
<td>&lt;0.01*</td>
<td>&lt;0.01*</td>
</tr>
<tr>
<td>Mult Regression$^d$</td>
<td>0.02456</td>
<td>&lt;0.01*</td>
<td>&lt;0.01*</td>
<td>&lt;0.01*</td>
</tr>
<tr>
<td>PLS$^e$</td>
<td>0.01560</td>
<td>&lt;0.01*</td>
<td>&lt;0.01*</td>
<td>&lt;0.01*</td>
</tr>
<tr>
<td><strong>Care Sector Network</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RA$^a$</td>
<td>0.15538</td>
<td>&lt;0.01*</td>
<td>&lt;0.01*</td>
<td>&lt;0.01*</td>
</tr>
<tr>
<td>Correlation$^b$</td>
<td>0.04079</td>
<td>&lt;0.01*</td>
<td>&lt;0.01*</td>
<td>&lt;0.01*</td>
</tr>
<tr>
<td>SBT$^c$</td>
<td>0.00054</td>
<td>&lt;0.01*</td>
<td>&lt;0.01*</td>
<td>&lt;0.01*</td>
</tr>
<tr>
<td>PLS$^e$</td>
<td>0.01807</td>
<td>&lt;0.01*</td>
<td>&lt;0.01*</td>
<td>&lt;0.01*</td>
</tr>
<tr>
<td><strong>Primary &amp; Specialty Care Sectors</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SBT$^c$</td>
<td>0.00000</td>
<td>0.49</td>
<td>0.44</td>
<td>0.52</td>
</tr>
<tr>
<td>Correlation$^b$</td>
<td>0.03588</td>
<td>&lt;0.01*</td>
<td>&lt;0.01*</td>
<td>&lt;0.01*</td>
</tr>
<tr>
<td>PLS (P→S)$^f$</td>
<td>0.01304</td>
<td>0.1</td>
<td>0.14</td>
<td>0.18</td>
</tr>
<tr>
<td>PLS (S→P)$^g$</td>
<td>0.02584</td>
<td>0.32</td>
<td>0.2</td>
<td>0.21</td>
</tr>
<tr>
<td><strong>Primary &amp; Mental Care Sectors</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SBT$^c$</td>
<td>0.0003</td>
<td>&lt;0.01*</td>
<td>&lt;0.01*</td>
<td>&lt;0.01*</td>
</tr>
<tr>
<td>Correlation$^b$</td>
<td>-0.08454</td>
<td>&lt;0.01*</td>
<td>&lt;0.01*</td>
<td>&lt;0.01*</td>
</tr>
<tr>
<td>PLS (P→M)$^h$</td>
<td>0.02256</td>
<td>0.02*</td>
<td>0.01*</td>
<td>0.01*</td>
</tr>
<tr>
<td>PLS (M→P)$^i$</td>
<td>-6E-05</td>
<td>0.39</td>
<td>0.48</td>
<td>0.49</td>
</tr>
</tbody>
</table>

$^a$ $\hat{\theta}_{RA}=\sum|q^2_i - q^1_i|$, where $q^1_i$ and $q^2_i$ are probabilities calculated from best models before (1) and after (2) Health Share’s formation

$^b$ $\hat{\theta}_{COR} = \frac{1}{g(g-1)} \sum_{i \neq j} |\hat{\rho}_{ij} - \tilde{\rho}_{ij}|$, where $\hat{\rho}_{ij}$ and $\tilde{\rho}_{ij}$ are correlations between each pair of network members $(i, j)$ during each time period (1,2) and $g(g-1)$ is the total number of connections in the network

$^c$ $\hat{\theta}_{SBT} = \frac{1}{g(g-1)} \sum_{i \neq j} |SBT^2_{ij} - SBT^1_{ij}|$, where $SBT^1_{ij}$ and $SBT^2_{ij}$ are standardized bivariate transmission values for each pair of network members $(i, j)$ during each time period (1,2)

$^d$ $\hat{\theta}_{OLS} = \frac{1}{g(g-1)} \sum_{i \neq j} |\hat{B}_{ij}^2 - \tilde{B}_{ij}^1|$, where $\hat{B}_{ij}^1$ and $\hat{B}_{ij}^2$ are regression weights for each pair of network members $(i, j)$ during each time period (1,2)

$^e$ $\hat{\theta}_{PLS} = \frac{1}{g(g-1)} \sum_{i \neq j} |\hat{s}_{ij}^2 - \tilde{s}_{ij}^1|$, where $\hat{s}_{ij}^1$ and $\hat{s}_{ij}^2$ are contributions to latent terms for each pair of network members $(i, j)$ during each time period (1,2)

$^f$ Identical to footnote ‘e’ but specific to contributions from primary care to specialty care

$^g$ Identical to footnote ‘e’ but specific to contributions from specialty care to primary care

$^h$ Identical to footnote ‘e’ but specific to contributions from primary to mental/behavioral care

$^i$ Identical to footnote ‘e’ but specific to contributions from mental/behavioral to primary care
Statistical Testing with I-DNA and R-DNA

To demonstrate a range of applications of Information theory based Differential Network Analysis (I-DNA) and Regression based Differential Network Analysis (R-DNA), each of three resampling techniques were used in conjunction with each of several network inference methods to test for significant change in Health Share’s CCO network. Central to each application of I-DNA and R-DNA was the calculation of a distance measure ($\hat{\theta}$) to serve as a test statistic. Distance measures, such as the mean absolute difference, were described in detail throughout Chapter 5 (e.g., eqn 22, p.127), and are summarized above in Table 6.1. These values are indicators of how much change occurred after Health Share formed, where change is defined in the most inclusive sense possible. In this chapter, each distance measure is tested for statistical significance, under the null hypothesis that no change occurred and the true distance is zero.

In order to statistically test each measure of network distance ($\hat{\theta}$), a customized reference distribution was first generated through one of three resampling techniques. Resampling entailed the application of an algorithm to randomly switch, add, or omit observations from each dataset, so that one can create many iterations of pseudo “before” and “after” dataset pairs. Each pseudo pair can then be used to infer pseudo network structures and measure resampled (pseudo) distances. It is the comparison of the originally observed distance against this type of reference distribution that allows us to determine if the observed distance is significantly different from zero.
Resampling Techniques

Three different resampling techniques were applied here, so that their results could be compared. Each resampling technique helps to estimate the probability of a type I error. This is the probability that we might have still measured the amount of change we did even if the null hypothesis (of no change) were true, based only on chance variations in the system’s behavior. Networks have variability in their behavior, so theoretically a network could show minor changes in billing patterns even if there were no systematic differences between the two time periods. Our intention with each resampling method was to determine whether the billing patterns before Health Share’s formation were systematically different than the billing patterns after it.

We conducted statistical tests with permutation resampling, single (one-sample) bootstrapping, and double (two-sample) bootstrapping. Permutation resampling essentially means combining cases from before and after datasets into one common pool, and then randomly sorting cases, without replacement, into pseudo ‘before’ and ‘after’ datasets of the same size. If our two original datasets were not systematically different, then shuffling cases between them should often produce about the same amount of difference as we observed in our original pair of datasets. Alternatively, if our datasets were systematically different, the difference observed in our original pair should be consistently larger than the differences in these randomly shuffled datasets.

Single (one-sample) bootstrapping is very similar to permutation, in that it essentially means shuffling cases between datasets at random and with replacement. The rationale here is the same as it was for permutation resampling. If our two datasets were
not systematically different, then bootstrapping should produce about the same amount of difference as we observed in our original pair of datasets. Alternatively, if our datasets were systematically different, the difference observed in our original pair should be consistently larger than the differences in bootstrapped datasets.

With both permutation and single bootstrapping, the probability of a type I error is calculated simply, as the proportion of times our resampled datasets produced a difference at least as large as the observed difference. If this occurred rarely (in fewer than 5% of resampling iterations), we can claim that the magnitude of change that we originally observed was statistically significant or unlikely to have stemmed from chance variations in billing patterns. However, if this occurred somewhat often (5% or more of resampling iterations), we retain the null hypothesis that the observed change might have simply been the product of random fluctuations in the network’s behavior.

Double (two-sample) bootstrapping is a little different from permutation and single bootstrapping, as the before and after datasets are kept separate while resampling with replacement. This follows a different rationale than the previous two techniques, because if our two datasets were systematically different, we should expect that their differences remain evident (rather than disappearing) after making random adjustments to the cases in each dataset. We don’t mix cases from the before and after datasets in any way, so resampled datasets should continue to show a similar amount of change as the original pair. Rather than checking whether resampled datasets consistently show less difference than our original pair, we check whether they are consistently greater than zero.
With double bootstrapping, the probability of a type I error is calculated with a variation of the paired samples $t$ test, after three modifications. In a traditional paired samples $t$ test, one first calculates a $t$ value by dividing the average difference among sample pairs by the standard deviation of differences among pairs (an indication of standard error). But in our applications of I-DNA and R-DNA, we divide our network distance measure ($\hat{\theta}$) by the standard deviation of bootstrapped distances. Both are reasonable substitutions, as our numerator remains an indicator of overall difference between two datasets, which we want to establish as different from zero, and our denominator indicates how much the numerator can be expected to vary with some random error in the sampling of cases.

Once the $t$ value is calculated, we can use a standard $t$ distribution and the number of degrees of freedom to calculate the $p$ value, or the probability of a type I error. The degrees of freedom in a traditional paired samples $t$ test is one less than the (paired) sample size, but in our case it is one less than the number of (pairs of) bootstrapped samples. This third substitution is basically an extension of our substitution for standard error. In traditional applications of the $t$ test, standard error is based on the standard deviation of cases, but in our application it is based on the standard deviation of bootstrapped samples. It makes better sense, then, to base our calculation of degrees of freedom off the bootstrapped sample volume rather than the case volume.

Statistical testing with any of the resampling techniques above assumes that the outcome variable (in our case distance) is continuous in nature and that observations (i.e. patients) are independent. There may be many instances where patients are not
completely independent, such as when they are members of the same family or from the same geographical location. But due to limitations in the data available, we cannot control for this dependency, so we cautiously assume that each observation of a patient’s interactions with the Health Share network can serve as an independent observation of the relationships between different billing providers and sectors in that network.

In addition to the assumptions above, the double bootstrap assumes that the focal outcome measure (in this case distance $\theta$) will be normally distributed. This may or may not be the case in all applications, but we can evaluate the distribution of bootstrapped distances and use a Shapiro-Wilk test to determine whether its shape is a significant deviation from normal. In cases where the Shapiro-Wilk test does not reject normality, we can cautiously assume that the distribution is normal enough, and that the $t$ and $p$ values are accurate. (Shapiro-Wilk test results and other distributional indicators are available in Appendix C.) Permutation resampling and single bootstrapping, by contrast, do not assume normality of the distribution in the outcome variable.

**Statistical Testing Methods**

Each of the three resampling techniques was applied to datasets from (1) the top 30 network, (2) the care sector network, (3) the connection between primary and specialty sectors, and (4) the connection between primary and mental/behavioral sectors. This totaled twelve resampling applications (of three techniques on each of four hypotheses). In each application, 100 pairs of pseudo “before” and “after” datasets were created, all of which matched the original datasets in the number of observations. This
meant that resampled pairs of datasets were identical to the original pair of datasets, except for adjustments that stemmed from the resampling process.

We note that an ideal application of I-DNA or R-DNA would use many more than 100 resampled datasets to create a reference distribution. It is not uncommon for resampling techniques to employ hundreds of thousands of iterations, affording more statistical power as well as specificity in estimating the likelihood of a type I error. However, distance measures by several of the inference methods employed here, in particular reconstructability analysis (RA) and partial least squares (PLS) regression, take a considerable amount of time to calculate. We estimated that 1,000 resampled datasets would take months to calculate by RA, or would potentially require additional computational resources, such as parallel or cluster computing. Given time constraints, we used only 100 resampling iterations for this demonstrative and comparative analysis.

Once resampled datasets were created, they were analyzed by each inference method and used to create all relevant reference distributions of resampled distances. For example, the same 100 permutation-resampled dataset pairs for the top 30 network were inferred by each of RA, SBT, correlation, multiple regression, and PLS, producing separate reference distributions of resampled distance measures (e.g., $\theta_{perm}$) by each inference method. This approach ensured that any differences in the resultant resampling distributions were due to differences in the inference methods themselves and could not be due to chance variations in resampling. If we had conducted separate resampling processes to generate reference distributions for correlation, multiple regression, etc.,
then differences in statistical results (including statistical significance) may have stemmed from chance variations in the resampling process.

**Results**

All statistical tests for change in the top 30 network and the care sector network were significant, indicating that all inference methods and resampling techniques were comparable in their ability to statistically detect network change after Health Share’s formation (see Table 6.1, p.230). Further, across all applications of I-DNA and R-DNA on these two networks, the $p$ value was less than .01, suggesting that if the null hypothesis were true (and no change occurred), there would be less than 1% chance for us to have measured as much change as we observed ($\hat{\theta}$). This implies a strong likelihood that the observed distance reflects actual, systematic changes in the system’s behavior over time.

Observed distance measures can be visually compared against resampled reference distributions for the top 30 and care sector networks in Table 6.2 and 6.3. These tables include frequency distributions to demonstrate variation in resampled distance measures as determined by each resampling method. For permutation (yellow) and single bootstrapping (blue), where cases from before and after were shuffled randomly into pseudo “before” and “after” datasets, the statistical significance is quite apparent because resampled distances were consistently smaller than the observed distance.\(^{41}\) For double

\(^{41}\) Out of 100 iterations, not even one randomly sorted pair of datasets produced as much change as we originally observed, indicating systematic differences between our original datasets ($p<.01$)
bootstrapping (green), where cases from before and after went through separate random resampling processes, statistical significance is visible in a different manner because resampled distances remained approximately the same as the originally observed distance and were consistently greater than zero. (More information on reference distributions is available in Appendix C.)
Table 6.2. Top 30 reference distributions by inference and resampling method

RA
$\theta = 0.43814$

Correlation
$\theta = 0.01841$

SBT
$\theta = 0.00029$

Multiple Regression
$\theta = 0.02456$

PLS
$\theta = 0.01560$

* x-axes indicate the magnitude of (resampled) distance measures, y-axes indicate the frequency of resampled datasets, and red arrows indicate the magnitude of the observed distance
Table 6.3. Care sector reference distributions by resampling method

<table>
<thead>
<tr>
<th>Method</th>
<th>RA</th>
<th>Correlation</th>
<th>SBT</th>
<th>PLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta$</td>
<td>0.15538</td>
<td>0.04079</td>
<td>0.00054</td>
<td>0.01807</td>
</tr>
</tbody>
</table>

* x-axes indicate the magnitude of (resampled) distance measures, y-axes indicate the frequency of resampled datasets, and red arrows indicate the magnitude of the observed distance.

In viewing the resampling distributions above, note that our statistical tests of network change do not align cleanly with our substantive hypotheses from Chapter 5, and therefore cannot be interpreted as validating any descriptive findings regarding the nature
of that change. Substantively, we hypothesized connectivity would increase in the top 30 network, and through our investigation found descriptive evidence for the consolidation of connectivity. But our statistical tests here use only the most general measure of network distance, which encompasses all changes in connectivity, including instances of strengthening, weakening, and changes in direction (e.g., from a negative to positive correlation). So statistical significance only provides evidence for change in a general sense and does not support any observations of connectivity increasing or consolidating.

In contrast to the network-level statistical tests, however, our pairwise connection-level statistical tests align better with substantive hypotheses from Chapter 5. Within the care sector network, we previously hypothesized that that the connection between primary and specialty care would weaken, and that the connection between primary and mental/behavioral health care would strengthen. Here we statistically test those connections for change, noting that our measures of change in each connection can only strengthen, weaken, or change direction, and cannot encompass multiple types of change like the distance measures for entire networks can. The nature of these connection-level distance measures is what allows our statistical results to potentially reinforce the descriptive conclusions reached in Chapter 5.

There are a couple of caveats to our interpretation of these statistical results, even at the connection-level. These tests of significant change are technically two-tailed tests – sensitive to significant change in either direction (strengthening or weakening), which differs from our substantive hypotheses which were expecting change in a particular direction. In addition, we’ve simultaneously minimized our statistical power and inflated
our family-wise error rate by running only 100 resample iterations and conducting 12 separate statistical tests on each connection. Both of these limitations necessitate cautious interpretation of statistical results in relation to our descriptive findings from Chapter 5.

Caveats aside, the majority of statistical tests on the connection between primary and specialty care were not statistically significant, with the exception of correlation (see Table 6.4). This connection was hypothesized to weaken after Health Share’s formation (i.e., become less strongly positive), as its Patient Centered Primary Care Home (PCPCH) model of care was expected to reorient care more predominantly within the primary sector, reducing the extent to which patients bounced around between primary and specialty care providers. Contrary to this substantive hypothesis, however, most methods of network inference found no change in that connection (SBT) or increases in strength (correlation, PLS). The relative change in connectivity was greatest by correlation (which strengthened by 12.5%, from $r = 0.288$ to 0.323), so it makes sense that correlation was the only inference method to reject the null hypothesis. But note its direction of change (strengthening) went counter to our expectations in Chapter 5.
Table 6.4. Primary and specialty sector reference distributions by resampling method

<table>
<thead>
<tr>
<th>Resampling Method</th>
<th>Correlation</th>
<th>Primary → Specialty</th>
<th>Specialty → Primary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Permutation</td>
<td>p=0.40</td>
<td>0.00000</td>
<td>0.03588</td>
</tr>
<tr>
<td>Single Bootstrap</td>
<td>p=0.43</td>
<td>0.01304</td>
<td>0.02584</td>
</tr>
<tr>
<td>Double Bootstrap</td>
<td>p=0.52</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observed Distance</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*x*-axes indicate the magnitude of (resampled) distance measures, *y*-axes indicate the frequency of resampled datasets, and red arrows indicate the magnitude of the observed distance.
The unique capacity of correlation to detect a significant increase between primary and specialty care sectors may speak to the nature of the change that occurred. As we learned in Chapter 5, SBT is somewhat limited compared to correlation, as it analyzes nominalized data (0, 1, or 2+ claims) while correlation can make full use of continuous claim frequencies. PLS can also be described as somewhat limited compared to correlation, in that it analyzes only the unique relationships between network members, disregarding billing patterns that can be doubly accounted for by other members. It seems possible, then, that the strengthening between primary and specialty care was part of larger, multi-sector patterns in the network (not unique to this pair of sectors), and that it was most predominantly visible among patients with higher claim frequencies. Those factors would explain why the relative amount of change by correlation was larger than the other inference methods, and why it was the only method to produce significant results.

Among statistical tests of change in the relationship between primary and mental/behavioral care sectors, all inference and resampling methods detected significant change between them. However, in PLS regression, which produces two directed relationships, only the connection from primary to mental/behavioral care was found to be significant, while the connection from mental/behavioral to primary care was not significant.\textsuperscript{42} As described in Chapter 5, the lack of change in connectivity from

\textsuperscript{42} The connection from the mental/behavioral care sector to the primary care sector was defined by a PLS regression that (a) used billing patterns of the primary care sector as the outcome variable, (b) defined independent ‘latent term’ predictor variables from all other care sectors, through a process much like principal components analysis, and (c) measured the extent to which billing patterns in the mental/behavioral sector contributed to those latent terms. The connection from primary to
mental/behavioral to primary care was a substantive surprise, because we expected that the incorporation of mental/behavioral healthcare would increase referrals from mental/health to primary care and would strengthen the relationship in that direction. However, given the very small amount of change measured by PLS (-6E-05), statistical non-significance is not especially surprising.

mental/behavioral care was defined inversely, using mental/behavioral billing patterns as the outcome variable and assessing the contributions of primary care to latent terms that predicted it.
Table 6.5. Primary and mental/behavioral sector reference distributions by resampling method

<table>
<thead>
<tr>
<th>Resampling Method</th>
<th>Primary to Mental</th>
<th>Mental to Primary</th>
<th>Observed Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Permutation</td>
<td>p&lt;0.01</td>
<td>p&lt;0.01</td>
<td>0.0003</td>
</tr>
<tr>
<td>Single Bootstrap</td>
<td>p&lt;0.01</td>
<td>p&lt;0.01</td>
<td>0.0003</td>
</tr>
<tr>
<td>Double Bootstrap</td>
<td>p&lt;0.01</td>
<td>p&lt;0.01</td>
<td>0.0003</td>
</tr>
</tbody>
</table>

*X-axes indicate the magnitude of (resampled) distance measures, y-axes indicate the frequency of resampled datasets, and red arrows indicate the magnitude of the observed distance.
Precision of estimates in change magnitude

In looking across the statistical results (i.e., p values) in Tables 6.1-6.5, a noticeable pattern is that resampling techniques behaved very similarly in their detection of significant results. This is somewhat striking given that the three resampling techniques were conducted independently and with only 100 iterations each. The consistency of p values across resampling methods, then, provides solid evidence for interchangeability among them. It doesn’t appear to matter whether you use permutation, single, or double bootstrapping to determine whether a network or connection has changed significantly over time.

The same thing cannot be said for inference methods, as they sometimes disagreed on whether a connection changed significantly, and they showed different levels of variability throughout the resampling process. This variability suggests that some inference methods may have offered more precision in estimates of distance, while others may have exhibited more noise. Differences in variability can be seen in Table 6.6 and Table 6.7, where resampling distributions were rescaled relative to the originally observed distance measures\(^ {43}\). This scaling allows the resampling distributions to be compared more directly to one another on the same x-axis, even though the observed distances they are being compared against varied widely in magnitude. Across the

---

\(^{43}\) Rescaling of each resampled distance was done by simply dividing it by the observed distance measure. Values greater than the observed distance measure have a value larger than 1.0, whereas values smaller than the observed distance measure have a value smaller than 1.0.
standardized distributions, some reference distributions are visibly wider (i.e., noisier), while other are narrower (i.e., more precise).

A good example of precision and noise is available in the top row of Table 6.6. When we resampled top 30 datasets by permutation and inferred networks by SBT (light green), resampled distances were about 15% lower than the original SBT distance and did not show much variation around the mean (StdDev=2.9%). By contrast, when permutation resampled datasets are inferred by RA, resampled distance show more variability (approximately 25-40% of the observed distance, with StdDev=5.5%). The relatively narrow distribution by SBT implies greater precision in our estimate of network distance by SBT, because the resampled distances remained relatively stable no matter how we shuffled the datasets. Higher variability by RA, by contrast, implies less precision in estimating distance, and implies that our observed distance measure might be a rougher estimate for the amount of change that occurred.
Table 6.6. Top 30 and care sector reference distributions relative to observed distances

Top 30 Reference Distributions

<table>
<thead>
<tr>
<th>Method</th>
<th>Permutation</th>
<th>Single Bootstrap</th>
<th>Double Bootstrap</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiple Regression</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PLS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SBT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RA</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* x-axes indicate the ratio of the resampled distances measures to the observed distance, y-axes indicate the frequency of resampled distances, and red arrows indicate the observed distance.
Table 6.7. Connection-specific reference distributions relative to observed distance

<table>
<thead>
<tr>
<th>Reference Distributions</th>
<th>Permutation</th>
<th>Single Bootstrap</th>
<th>Double Bootstrap</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary and Specialty Care Sector</td>
<td><img src="image1.png" alt="Graph" /></td>
<td><img src="image2.png" alt="Graph" /></td>
<td><img src="image3.png" alt="Graph" /></td>
</tr>
<tr>
<td>Primary and Mental/Behavioral Health Sector</td>
<td><img src="image4.png" alt="Graph" /></td>
<td><img src="image5.png" alt="Graph" /></td>
<td><img src="image6.png" alt="Graph" /></td>
</tr>
</tbody>
</table>

*x-axes indicate the ratio of the resampled distances measures to the observed distance, y-axes indicate the frequency of resampled distances, and red arrows indicate the observed distance.
When comparing inference methods on the variability of their reference distributions (relative to observed distance measures), remember that the same resampled datasets were used for inference by each method. So, in the top row of Table 6.6, the same 100 pairs of permutation resampled datasets were used to generate reference distributions by all inference methods shown in that chart (RA, SBT, correlation, multiple regression, and PLS). This means differences between reference distributions cannot be due to chance variations in the resampling process, and instead stem from distance estimation properties that are specific to the inference methods themselves.

That said, we still face a substantial limitation in this observation, due to the relatively small number of resampled datasets by any technique. We only conducted 100 resampling iterations in each case, so a certain amount of noise in the reference distributions can be expected. Differences between inference methods in their resampling variability were fairly small in most cases, with standard deviations usually within 1% - 8% of the observed distance among statistically significant tests. And there do not appear to be consistent patterns in terms of which inference methods demonstrate more precision. Future work is needed, with many more resampling iterations, to investigate whether some inference methods offer greater precision in estimating network change.

**Distributional Assumptions**

Various measurements were taken on each resampling distribution, including its mean, standard deviation, skewness, kurtosis, and Shapiro-Wilk test for normality. These measurements are included in Appendix C. Of particular interest, Shapiro-Wilk tests did not reject normality for any of the double bootstrapped resampling distributions,
indicating that it was reasonable to use those distribution in conjunction with a bootstrapping variation of a paired samples $t$ test. Interestingly, Shapiro-Wilk tests did occasionally reject normality for some of the permutation and single bootstrapped resampling distributions. Fortunately, those resampling methods of statistical analysis do not employ distributional assumptions. Their estimation of the probability of a type I error is based strictly on the proportion of resampled distances that are as large or larger than the originally observed distance.

**Summary**

This chapter has illustrated statistical testing of network change by I-DNA and R-DNA through a comparative application of three resampling techniques and five methods of network inference on two representations of a billing provider network. We find general evidence to suggest that Health Share’s network of healthcare delivery changed significantly after it formed as a CCO, both in terms of network connectivity among its top 30 billing providers and in pairwise connectivity between its sectors of care.

In addition to network-level applications of I-DNA and R-DNA, this chapter has demonstrated the methods’ utility for follow-up or post-hoc analyses at the connection level. In these applications we find general evidence to suggest that the connection between primary and mental/behavioral health sectors weakened significantly over time, but for the most part we do not find evidence to suggest a significant change between primary and specialty care sectors after Health Share’s formation (see Table 6.1, p.230).
Our comparative analysis of network-level and connection-level change by these inference and resampling methods demonstrate their utility for statistical testing of network change. All three resampling methods were found to be extremely comparable in their ability to statistically detect network change, as they consistently agreed across 17 hypothesis tests on whether the probability of a type I error was < .05. Inference methods were also found to be quite comparable in the detection of network-level change, although they sometimes disagreed on the detection of connection-level change. This suggests that one might exercise caution when selecting an inference method, to ensure that its definition of connectivity aligns well with the type of change that is theoretically anticipated.
Chapter 7: Discussion and Summary

Introduction

This dissertation project has addressed two aims, including a primary methodological aim and a secondary substantive aim. The methodological aim of the project was to apply Regression-based Differential Network Analysis (R-DNA) and Information-based Differential Network Analysis (I-DNA) to a social network context and to demonstrate and compare their capacities for network inference and statistical testing of network change. The substantive aim of the project was to apply and assess R-DNA and I-DNA by using them to (1) infer the structure of a healthcare delivery network from Medicaid claims data, to (2) explore how it changed following the formation of the Health Share of Oregon Coordinated Care Organization (CCO), and (3) to test whether the amount of change was statistically significant.

To address these aims, healthcare delivery networks were inferred from Medicaid claims data for healthcare organizations that formed the Health Share CCO on September 1, 2012. In their raw form, claims data listed only a deidentified patient ID, a billing provider ID, and a period of time. These data were cross-tabulated into analytic datasets of claim frequencies, with one row for each patient and one column for each network node (e.g., provider), indicating the number of claims filed for each patient by each node. Analytic datasets were prepared separately for claims filed during one-year periods before (May 1st, 2011 – April 30th, 2012) versus after (January 1, 2013 – Dec 31, 2013) Health Share’s formation. This format allowed for the inference of network structures by
treating variables (columns) as nodes and estimating their connectivity with other nodes through the patterns of association (e.g., correlation) in the patients they billed claims for.

Three types of networks were inferred to address the project’s aims. The first network was a ‘full network’, including all 1,298 billing providers who were active within the Health Share service region during both time periods, plus one node to represent all billing providers from outside the service region. The second network featured the top 30 billing providers by patient volume, who accounted for the lion’s share of claims (88.2%) and billed for the majority of patients (96.7%). The third network grouped billing providers into six sectors of care, indicating the network of billing patterns among primary, ancillary, specialty, mental/behavioral, facility, and ‘other’ health care sectors.

Both I-DNA and R-DNA were successfully applied to the analysis of the CCO network, demonstrating the methods’ capacity for network analysis and providing some insights into changes in healthcare delivery following Health Share’s CCO formation. Both aims were furthered by inferring the networks with multiple methods and comparing the network structures they produced. Where networks differed, we could see the types of network patterns that each method was more and less able to capture – for example methods differed in how they responded to multiway patterns of connectivity. Where networks structures were similar, we were able to gain more confidence in our conclusions regarding the nature of the changes that occurred in the Health Share

---

44 Billing providers were selected as the primary level of analysis rather than performing providers, because there was a more manageable number of them and because their behavior was expected to be more stable and less subject to random variation than the behavior of individual performing providers.
network – for example the top 30 network was found to demonstrate consolidation of connectivity, whereby a set of core connections became stronger while connections in the periphery became weaker. Our methodological aim was also furthered by comparative analysis with multiple resampling techniques, all of which produced very similar statistical results (i.e., similar p values).

Methodological Contributions

Application of R-DNA to a social network

To address the methodological aim, both I-DNA and R-DNA were used to statistically test change in the Health Share’s billing provider network following their formation as a CCO. The application of R-DNA was relatively straightforward, as this method has already been developed in the field of bioinformatics and only needed to be abstracted for successful application in a social network context (e.g., Pihur, Datta, & Datta, 2008). R-DNA was originally developed for use with microarray data, to statistically test for differences in genetic networks among subpopulations of species with different expressed traits (e.g., overweight versus normal-weight mice). In our application here, we extended this type of application to the analysis of Medicaid claims data, substituting billing providers as network nodes (instead of genes) and cross-tabulating claims data so that patients could be used as observations of connectivity between billing providers (instead of biological specimens).

Although theoretically simple, the application of R-DNA to a social network context is still a useful contribution for the field of social network analysis (SNA). Often
network connectivity is measured directly, through self-reports by network members or through pairwise measurements of connectivity, and this is hugely time-consuming as well as statistically limiting. When social network structures can be inferred from multiple observations, rather than measured directly from a single observation on each pair of members, we are able to statistically test networks for significant differences over time or under different conditions. With all of the big data resources available today, network inference is much more accessible, and it affords many advantages in terms of its capacity for statistical testing.

In our application of R-DNA, we used three different methods of network inference: correlation, multiple regression, and partial least squares (PLS) regression. We compared the networks they inferred and the amount of change they could detect in the network over time. This comparison of inference methods within R-DNA illuminates differences among them regarding what types of relationships and change they can detect and how these methods might address research questions relating to social phenomena. Of special note, previous research in bioinformatics has emphasized the advantages of PLS over correlation and multiple regression for their ability to tease out unique relationships between nodes and detect differences between networks. However, our application to the billing provider network did not find PLS to be very different from multiple regression in its ability to tease out unique relationships between providers or detect change over time. It is possible, then, that the advantages to using PLS over simpler, more accessible inference methods do not carry over into all applications of
social network analysis, for example in situations where there are not particularly high levels of multicollinearity.

**Development of I-DNA for statistical testing of network change**

This dissertation drew from information-theory to develop I-DNA as a second method for statistical detection of network change. This is a main contribution of the dissertation, because to date information theory has only rarely been used for inference of network structures (Ver Steeg & Galstyan, 2012), and to the author’s knowledge has not yet been used to statistically compare networks in any field. Information theory offers a unique formalization of associativity between variables, and reconstructability analysis (RA) in particular offers many desirable features for network inference – such as multiway association detection, best model searching, and capacity to detect nonlinear relationships. We expected that information theory-based measures of network inference and network comparison would be useful in the field of social network analysis, and in this project it was indeed found to be useful for understanding and testing social network change.

In order to develop the I-DNA method, we defined a measure of network distance that could be used on the type of network structures produced by RA (i.e., calculated probability distributions). This distance measure underwent a couple of iterations, because the best models identified by RA for any two periods will often be non-nested models and therefore difficult to directly compare (e.g., Lewis, Butler, & Gilbert, 2011). We ultimately settled on a relatively simple measure of network difference which summed the absolute differences of the calculated probabilities in each best model. Once
this measure was developed, we coupled it with the same methods of statistical testing (i.e., bootstrapping procedures) from R-DNA and demonstrated its capacity to test network change over time.

The methodological capacities of I-DNA complement those of R-DNA. Like R-DNA, I-DNA allows for statistical testing of network differences and network change over time. But unlike R-DNA, I-DNA can capture nonlinear and multiway associations among billing providers. This provides additional insights into the nature of a social network and its change over time that are unavailable through R-DNA, which is restricted to detecting linear and pairwise associations. In our application of I-DNA to CCO network inference and statistical testing, we demonstrated how a couple of information theory-based metrics could help identify patterns of connectivity and network change. Changes in the entropy of best models can be taken as an indication of changes in the level of complexity (i.e., interconnectivity) of the target system, since entropy is a measure of (the lack of) constraint within the system. And examination of the probabilities with the greatest change can point toward specific network members as well as specific patterns of change in network behavior.

**Methodological comparisons of I-DNA and R-DNA**

Because both R-DNA and I-DNA were applied to the same CCO network, we were able to compare them on their capacities for network inference and statistical detection of network change. And because we applied several variations of I-DNA and R-DNA, with each of several methods of network inference (e.g., correlation, RA) and data resampling (e.g., bootstrapping, permutation), many methodological comparisons were
possible. Ultimately, our comparisons between resampling methods showed virtually no difference in statistical results, suggesting that permutation, single (one-sample) bootstrapping, and double (two-sample) bootstrapping are all highly comparable in their ability to statistically test the distance between two networks. By contrast, comparisons of network inference methods showed many distinctions in the types of network connections they inferred, and in their ability to detect change over time. This led to some recommendations for selection of an inference method which will be discussed below.

We recognized that reconstructability analysis was very different from regression-based methods of network inference, and that direct comparisons between them would be like comparing apples to oranges. So, we developed a simplified pairwise version of reconstructability analysis, standardized bivariate transmission (SBT), to serve as a midway point between the multiway, nonlinear network structure produced by RA and the pairwise, linear network structure produced by correlation. SBT is a nonlinear pairwise method that is like correlation in its inability to detect multiway relations and is like RA in its use of nominal (binned) data to infer connections. Development of this method enabled us to better understand distinctions between the other inference methods that were applied, for example it allowed us to see which patterns of network change were specific to the multiway patterns as captured by RA, or specific to linear analysis of continuous variables as captured by correlation.

The fullest comparative analysis was conducted on a network of the top 30 billing providers, because it was small enough to be computationally feasible by all inference methods, and because we had a network-level hypothesis for it that each method could
theoretically address. All inference methods were successful in identifying a set of connections amongst the top 30 billing providers, and each of them detected some amount of change in connectivity over time. However, methods often disagreed about the strength of connectivity between specific providers, and even disagreed on whether the main network hypothesis (of increased connectivity) was supported or refuted. Our efforts to reconcile differences between methods led to insights regarding the types of connectivity patterns that could be detected through each inference method, to some substantive insights regarding consolidation that will be discussed further below, and to some recommendations for selection of an inference method in future studies.

Recommendations for selection of an inference method

Based on the findings described throughout Chapter 5, the largest differences between methods of network inference stem from their treatment of multiway patterns of connectivity – those involving three or more network members – and from their ability to make full use of continuous data. If one’s research question can be sufficiently addressed by a nominal (binned) version of the data, and if nonlinear or multiway patterns of connectivity are of specific interest, then there are significant advantages to conducting I-DNA with RA network inference. However, multiway associations can sometimes be difficult to interpret,\(^45\) so if they are not of specific interest one might not opt for RA. Conducting I-DNA with SBT can be another reasonable choice, when nonlinear patterns of connectivity are expected, and one is only interested in pairwise associations. But SBT

\(^{45}\) An example of difficulty interpreting a multiway connection would be to ‘make sense’ of why so many people had one claim with provider A, two or more with provider B, and zero with provider C.
also requires nominalized data, which sometimes results in a substantial loss of information.

If nonlinear and multiway relationships are not of interest, and if nominal categories are not a good representation of the data, then R-DNA methods of network inference will be preferable. Here one should consider whether general patterns in the network, which involve multiple network members, should be ignored or deliberately excluded. Correlation will look at each pair of members independently, as though no other members exist, so it is unable to distinguish between patterns of connectivity that are local (pairwise) and those that are symptomatic of a larger pattern across the network. By contrast, multiple regression and PLS infer ‘unique’ pairwise relationships between members, in a manner that deliberately excludes all multiway connectivity. In networks with strong multiway patterns of connectivity, this can result in the exclusion of a substantial amount of information. However, if one is more sincerely interested in defining the unique relationships between each pair of members, and if multiway patterns are thought of as noise, then multiple regression and PLS would both be good choices for network inference.

The choice between PLS and multiple regression comes down to a willingness to define connectivity through the use of abstract latent terms. In cases of multicollinearity, PLS is thought to hold an advantage over multiple regression, because predictor variables (nodes) can contribute to multiple latent terms, and in that way predictor variables can carve out additional capacity to be uniquely connected to one another (Gustafsson, 2004; Haenlein & Kaplan, 2004). However, this advantage has been contested in the literature.
(Marcoulides, Chin & Saunders, 2009), and in our applications PLS did not show much advantage over multiple regression, at least in terms of detecting network change.\textsuperscript{46} In many cases, multiple regression may be sufficient to identify unique relationships among network members, and it holds an advantage over PLS by being a little more interpretable in lay terms.

It is worth noting that all methods of network inference contributed meaningfully to our understanding of the top 30 network. The reconciliation of different results of each method, although difficult, produced a clearer picture of the types of changes that occurred in the Health Share network following its formation as a CCO. If one is not sure whether a research question is best addressed by unique, nonlinear, or multiway patterns of connectivity, then in some cases it may be useful to infer the network by several methods and compare their results.

**Demonstration of hypothesis testing on specific network connections**

In addition to network level hypothesis testing with R-DNA and I-DNA, we conducted two hypothesis tests on specific connections within the care sector network. Applications of R-DNA in bioinformatics have already demonstrated post-hoc analyses for specific connections and even for specific nodes (e.g., genes) which may have different patterns of overall connectivity within a network. We extended the notion of connection-specific R-DNA from bioinformatics to the CCO setting and explored how I-

\textsuperscript{46} PLS network inference accounted for substantially more variance in provider behavior, as shown in Figure 5.23 (p.177), but ultimately did not detect much more change in connectivity than multiple regression.
DNA might be developed to serve the same purpose. We found these applications to be good exploratory starting points in identifying patterns of change that occurred in specific connections, although additional analyses were necessary – for example examination of contingency tables – to further our understanding of those changes.

In our extension of R-DNA to hypotheses on specific connections, we found that simple measurements of the magnitude and direction of change were insufficient to evaluate our hypotheses. This was in part because R-DNA connections (e.g., correlations) can be either be positive or negative during each time period, which generated complications for the interpretation of change. A net increase in a connection’s weight might indicate that it became increasingly positive, decreasingly negative, or switched from negative to positive. Likewise, an increase in a connection’s absolute strength might mean it became increasingly positive, increasingly negative, or switched directions. We found that contingency tables were essential to understand the specific patterns that changed over time, and we recommend their use for future applications of R-DNA to social network phenomena.

Our application of I-DNA to specific connections was also somewhat complicated, and it necessitated the use of additional analytic steps. Network inference by RA produced only multiway connections in the care sector network, which meant that any insights regarding specific pairs of sectors were buried within multiway patterns. And network inference by SBT did not differentiate between positive, negative, or even nonlinear associations, so summary measurements of change provided very little information regarding the nature of that change. Again, we augmented this to a large
extent with contingency tables, and for RA we considered how focal healthcare sectors were involved in the calculated probabilities that changed the most (see Table 5.8, p.202). But the necessity of these additional steps highlights the exploratory nature of I-DNA for examination of specific connections and suggests that I-DNA may not be well-suited for analysis of specific pairwise connections.

**Substantive Contributions**

To address the substantive aim, both I-DNA and R-DNA were used to evaluate some specific hypotheses pertaining to changes in Health Share’s billing provider network following its formation as a CCO. Most hypotheses stemmed from one aspect of CCO formation, the adoption of a Patient Centered Primary Care Home (PCPCH) model of care, and even more specifically on a couple aspects of the PCPCH model, including the assignment of each Medicaid member (i.e., patient) to a primary care provider (PCP) and the requirement for patients to get a referral from their PCP in order to access specialty care. These features were expected to reduce ‘noise’ in patients’ trajectories through the healthcare system, as patients would receive primary care services predominantly from their assigned PCP, and the PCPs referrals to specialty providers would likely be more efficient (i.e., necessitating fewer specialists) and consistent (i.e., to the same fellow providers) than patients’ self-selection of specialty care had been.

**Observations of network-level change**

Our expectations for the full network and top 30 network of billing providers was that connectivity would increase. That is, with increased involvement from PCPs in
guiding patients throughout the healthcare system, we expected that providers who had previously billed for a lot of patients in common (i.e., those who were positively associated) would bill for even more patients in common. And providers who had previously billed for relatively distinct sets of patients (i.e., those who were negatively associated) would bill for even more distinct sets of patients. We expected a reduction in the randomness or ‘noise’ of patient movement through the healthcare system, which would augment consistent patterns. In specific terms, our hypothesis was that connections would strengthen more often than weaken, that they would ‘appear’ (or become detectable) more often than they would disappear. And when applying RA that increased connectivity would result in connections becoming more complex (i.e., more multiway) after Health Share’s formation.

We applied only two methods of network inference to the full network – correlation and SBT\(^{47}\) – but both refuted our hypothesis, indicating very slight decreases in connectivity (\(\Delta r = -0.0003\); \(\Delta SBT = -5 \times 10^{-7}\)) and fewer connections after Health Share’s formation. We then applied five methods of network inference to the top 30 network – correlation, SBT, RA, multiple regression, and PLS regression – and those results were mixed with some supporting and some refuting our hypothesis. Reconciliation of discrepant results revealed a pattern of \textit{consolidation} in connectivity. Consolidation meant that a central core formed stronger connections while an outer periphery became more weakly connected over time. This caused visible increases in

\(^{47}\) We encountered computational limitations in applying RA, multiple regression, and PLS regression to a network with 1,299 nodes. Calculations would have taken weeks to run and required cluster computing.
disparity between strongest and weakest connections, but measures of average change in connectivity mostly evened out, resulting in very small values just above or just below zero ($\Delta r = -0.00018; \Delta SBT = -5 \times 10^{-7} \Delta B = -0.00145; \Delta s = -0.00105$)\textsuperscript{48}.

While we did not find the type of change that we hypothesized, our network-level finding of consolidation could theoretically stem from the same causal mechanism that we anticipated – Health Share’s implementation of the PCPCH model of care. If our underlying expectation was true, that the PCPCH caused referral patterns became more pronounced after Health Share’s formation, then it seems in retrospect quite plausible that the referral patterns among the largest billing providers would strengthen at the expense of referral patterns among smaller billing providers. Stated another way, if a relatively constant number of patients navigated the healthcare delivery network during both time periods, but PCPs guided them through a more consistent subset of pathways after Health Share’s formation, it stands to reason that some connections would strengthen while others would weaken. This might have been a more reasonable expectation rather than our hypothesis that connections would strengthening uniformly throughout the network.

Alternatively, it is also possible that the consolidation of connectivity was driven by other factors outside of the PCPCH. Health Share’s formation included partnerships with many organizations and involved the integration of care services that were previously separate (e.g., dental and mental health). It is reasonable to assume that these organizational changes were accompanied with some restructuring of billing practices,

\textsuperscript{48}$\Delta r$ is the average change in correlation-based connectivity, $\Delta B$ is the average change in multiple regression-based connectivity, and $\Delta s$ is the average change in PLS-based connectivity. Note that reconstructability analysis cannot provide direct measurements of change in pairwise connections.
which might very well have become more centralized or streamlined. A top-down restructuring of billing practices might well have produced the consolidation that we observed and would have been completely independent from Health Share’s implementation of the PCPCH model of care.

**Observations of connection-level change**

To infer the care sector network, billing providers were sorted into one of six types of care: primary, specialty, mental/behavioral, ancillary, facility, and ‘other’ care. This classification was based on what kinds of performing providers they most often billed claims with, after sorting performing providers by their self-reported specialties in the National Uniform Claim Committee’s (NUCC) Taxonomy of Health Care Specialties. Previous classifications of NUCC taxonomies into healthcare sectors could not be found in the existing literature, so we sorted taxonomies by hand for the purposes of this analysis. This classification went through several iterations, and some initial categories (e.g., dental, inpatient facilities) were merged with others to avoid the creation of sector nodes with very few billing providers. The final classification scheme is provided in appendix A.

We did not have a network-level hypothesis for the care sector network, but we did have hypotheses for two specific connections within it. Our first hypothesis, based on Health Share’s implementation of the PCPCH model of care, was that PCP influence over patients’ utilization of specialty and primary care services would cause the connection between those two care sectors to become less strongly positive over time (with patients bouncing back and forth less between primary and specialty providers). Our second
hypothesis, based on the integration of mental and behavioral health care into Health Share’s care delivery system, was that patients who received mental and behavioral care services would be more often referred to primary care, and that this would cause the connection between primary and mental/behavioral sectors to become more strongly positive.

Some results provided limited support for changes that were related to our hypotheses, such as a reduction in specialty care utilization and a greater reliance on primary over specialty care. However, the percentage of patients who did not utilize primary or specialty care nearly doubled after Health Share’s formation, increasing from 5.8% to 10.4%, and the proportion of mental/behavioral patients who utilized primary care decreased from 84.0% to 73.1%. These patterns indicate that patients might have either been receiving primary care services within other care sectors – meaning they did not need primary care services under the purview of a primary care billing provider – or that patients may have faced increased obstacles to accessing care. Future research may be needed to investigate patients’ access to primary care and the accessibility of PCPs within other sectors after Health Share’s formation.

Our dissection of changes in connectivity between care sectors brings up some interesting evaluative questions for networks of healthcare delivery. While we did not find any of the changes that we hypothesized, our post-hoc analyses of changes in the primary-specialty connection and the primary-mental/behavioral health sector connection made it apparent that some of the unanticipated changes might also be signs of increased effectiveness in care delivery. Future network evaluations of CCO legislation might do
well to consider how different sectors of care ‘should’ be connected to one another, and how those types of connections could be measured. It is unclear, for example, whether patient movement between care sectors would indicate an effective referral process or an inability of clinics to provide a sufficiently comprehensive range of care services.

**Observations of multiway connectivity**

Comparative analyses of the top 30 and care sector networks gave many indications of multiway connectivity. This was most obvious in networks inferred by RA, which indicated many multiway connections involving three to five nodes. But it was also subtly evident in comparisons between R-DNA inference methods, as the connection weights inferred by correlation – which include of patterns that are generally present across a network – were substantially larger than weights by multiple regression and PLS, which are restricted to patterns that are unique to each pair (see Figure 5.23 on p.180 for reference). This suggests that Health Share’s billing provider network has some patterns that are inherently complex and multidimensional and implies that attempts to boil down billing patterns to a set of pairwise relationships is to some extent an oversimplification.

A plain observation of complexity in a healthcare delivery network is hardly surprising, and its direct implications for healthcare delivery may not be entirely obvious. However, we note that our consideration of multiway connections greatly enhanced our understanding of the network changes that followed Health Share’s formation, and particularly for the top 30 network. Among the top 30 billing providers, RA’s inference of multiway connectivity provided the most direct evidence of consolidation, by demonstrating fewer connections that contained more providers after Health Share’s
formation. This notion of consolidation was reinforced by other methods, although more indirectly, and was ultimately the key to our reconciliation of inference methods on discrepancies in their results.

Understanding the context of multiway connectivity was also helpful in our comparison of inference methods on how much change occurred in the top 30 network. Correlation showed substantially less change in connectivity than multiple regression and PLS\textsuperscript{49}, by allowing multiway patterns to be doubly captured across each involved pair. Inversely, PLS and multiple regression showed more changes in connectivity than correlation, when exclusively measuring patterns that were unique to each pair. These differences suggest that there was more change in \textit{unique} pairwise connections than there was in global pairwise relationships. And while there may have also been meaningful changes in multiway patterns of connectivity (as evidenced by RA), these multiway connections appeared to remain very stable when boiled down to the pairwise level.

**Limitations**

This dissertation project contains both methodological and substantive limitations. Substantively, this project is limited from an evaluation standpoint, because our research only investigated if and how billing provider networks might have changed. We did not analyze data that pertained to whether patients became healthier, more satisfied with their care, or received services in a more cost-efficient manner. This means that our project cannot really evaluate Health Share’s effectiveness in achieving the triple aim of CCO

\textsuperscript{49} When considering the volume of appearance and disappearance of pairwise connections above a threshold; see Table 5.6 on p.172.
legislation including improved health outcomes, experience of care, and reduced cost (e.g., Berwick, Nolan & Whittington, 2008), and means that our results have limited capacity to evaluate this health policy.

Another substantive limitation of this work is that our research design does not allow for causal inference. We did not conduct a randomized controlled trial or compare Health Share’s network to a counterfactual network where a CCO did not form. We also did not measure changes in the Health Share network over a series of time periods to determine whether changes that immediately followed CCO formation were larger or different than routine changes that have occurred outside the context of CCO formation. These limitations in our study design make it inappropriate to attribute any observed changes to Health Share’s formation. Even though there are theoretical mechanisms that might explain a causal pathway between Health Share’s formation and the types of changes we observed, it is also very possible that those changes happened independently.

A third substantive limitation of this work is that our analysis assumed a predominant consistency in the attributes of billing providers over time. This assumption is not entirely warranted because CCO legislation included many incentives that would have shifted the nature of the organizations that these billing providers operated within. Many clinics became PCPCHs, for example, which would have expanded the range of services that their providers billed for. And the general shift from what used to be a fee-for-service model to the new CCO pay-for-performance model (Oregon Health Authority, 2019a) might have changed the way many claims were submitted, for example by removing incentives to capture every single medical service as a distinct claim.
We observed in Chapter 4 that the total number of claims submitted decreased by 10.6% despite the total number of Medicaid patients increasing by 4.38% (see Table 4.7, p.115). This indicates that there were likely some general shifts in billing, and in the behaviors of individual billing providers, followed Health Share’s formation. We also observed in Chapter 4 that only 61% of Medicaid patients were present during both time periods (see Table 4.1, p.103). This indicates the possibility that patient populations could have been substantially different in terms of basic demographics and healthcare needs, information about which we did not have access to in this study. By undertaking this project from a network analysis perspective, our focus was primarily on changes in the connections between billing providers, and we did not make attempts to control for changing attributes of the individual providers, the patient populations they served, or changes the volume or style of claim submission under a different payment model.

Finally, our project is substantively limited in scope because the bulk of analyses focused mainly on the core of the network – the top 30 billing providers – and on two specific connections between healthcare sectors – the connection between primary and specialty care and the connection between primary and mental/behavioral care. Other areas of these networks were not explored thoroughly, so we remain largely unaware of the types of network change that may have occurred among them. And it is possible that the changes we observed within the core and between specific sectors may not generalize to other areas of these networks. Connectivity throughout the periphery of the billing provider network may not have exhibited consolidation like we observed in the networks’ core, and changes in other connections – such as the connection between specialty and
mental/behavioral care could add insights and context to the changes we observed between those sectors and primary care.

This dissertation project also contains several methodological limitations. Perhaps the largest one is that we have only explored the usefulness of R-DNA and I-DNA for one social network. A better demonstration these methods’ utility for social network analysis should have at minimum several applications to different types of social networks. Because we only focused on detection of change in Health Share’s billing provider network, it is possible that some of our methodological findings would not generalize well to other contexts. For example, we made recommendations for selection of a network inference method based on the researcher’s interest in nonlinear relationships and multiway patterns of connectivity. Those themes may be less important for some social network applications, and there may be additional relevant considerations that we have missed.

This project was also methodologically limited because we only explored five methods of network inference. We compared change detectable through correlation, multiple regression, PLS regression, SBT, and RA, but there are many more approaches to network inference are worth exploring. Multiple regression models can be constructed in ways that include interaction terms, which would theoretically provide R-DNA with an approach to formalizing multiway patterns of connectivity. And there are additional

50 An obstacle to using interaction terms in a regression network is that they would require prior selection and formal definition of all interaction terms at one or more levels of (each two-way, three-way, and four-way interaction, etc.). This prior specification of every possible multiway interaction is not necessary in I-DNA, which can still identify multiway patterns of connectivity among network members.
approaches outside of I-DNA and R-DNA that may be worth exploring, for example
there are hierarchical Bayesian approaches to network inference (e.g., Gomes, Rao, &
Neville, 2018), and network comparisons are possible by using Laplacian-based measures
(Ginestet, Li, Balachandran, Rosenberg, & Kolaczyk, 2014). We do not consider this
project to have achieved a comprehensive comparison of network inference techniques
and note that there are many more methods available.

In a similar manner, this project is methodologically limited because we only
explored very generic measurements of network distance. The mean absolute difference
measure (from R-DNA), and the sum of differences in calculated probabilities (from I-
DNA), were both distance measures that aggregated all types of change in any direction.
This made them very inclusive and comprehensive measurements of network change, and
good candidates for the starting point of our methodological work, but they also did not
provide much of any information regarding what type(s) of change occurred. Even in our
most general hypothesis – that connectivity would increase – we needed to do additional
network calculations to assess whether that type of change occurred. This indicates that
our general distance measurements may have limited utility in the social sciences.

Finally, we encountered computational obstacles in this project that limited our
methodological contributions. Some methods of network inference were found to be
computationally very time intensive, particularly RA, PLS, and multiple regression. Not
only was network inference and distance measurement time consuming for the original
datasets (e.g., ~20 minutes for the top 30 network), but each resampled pair of networks
needed to be inferred and measured in the same manner. This meant that even a minimal
100 resamples on the top 30 network took multiple days to calculate distance measures for. Most inference methods were not practical to apply to the full network of 1,299 nodes, and 100 resamples is really a pitiful number by most standards. These limitations would likely hold true for applications of I-DNA and R-DNA to other large networks, and they may need to be addressed by more sophisticated computational resources.

**Future Work**

Future research can address many of the methodological limitations described above. Most obviously, we could apply I-DNA and R-DNA to different types of social networks besides billing provider networks, and we could use them to investigate different types of social phenomena. This is a critical next step in establishing whether these approaches are generally useful in the field of social network analysis. We can also expand the scope of our statistical methods beyond estimates of Type I error (p values) to include measures of effect size and confidence intervals. And we can apply techniques like parallel processing to overcome computational limitations and conduct statistical tests with many more resampling iterations than were used here (100).

Within the CCO analysis project, we could enhance our applications of I-DNA and R-DNA by tailoring our measurements of network change more closely to our research questions. For example, we could substitute our generic measurement of network distance for a measurement of change in eigenvector centrality (Bonacich & Lloyd, 2001), perhaps with the expectation that the primary care sector and PCPs would take on ‘higher status’ in the network after Health Share’s formation. We could also tailor our methods of network inference to be more intuitive for considering changes in a
network of healthcare delivery. For example, in our calculation of the connection between primary and specialty care sectors, we might exclude ‘non-patients’ who did not receive care from either sector. That would prevent large changes in the volume of ‘non-patients’ from overwhelming our measurements of change.

Some of the larger substantive limitations would be harder to address. It is not likely feasible to implement CCO legislation as a randomized controlled trial, and comparison of the changes in Health Share to a non-CCO billing network may be accompanied by confounding variables such as geographic variations in healthcare needs, organizational priorities, state policies, and public health initiatives. These constraints make it difficult to design a study that would ascertain beyond reasonable doubt that the changes we observed in the Health Share network were truly caused by its formation as a CCO.

That said, some natural extensions of this project would be to explore shifts in billing provider behaviors following CCO formation. Our assessment of change in the connection between primary and mental/behavioral sectors, and between primary and specialty care sectors, could be augmented considerably if we had a fuller appreciation for the nature of changes that occurred within these sectors. Chiefly, it would be very helpful to know whether patients who stayed within specific care sectors were accessing a broader range of care services after Health Share’s formation. For example, if patients within the mental/behavioral care sector had access to PCPs, and if patients within the primary sector had access to basic mental health services, we could adopt a more optimistic appraisal of the reduction in patient overlap between those two sectors.
Additional efforts could be made to hold some time-varying aspects of the healthcare system constant, to better tease out the impact of CCO formation. For example, we could subset our analyses to patients who were present during both time periods, which would reduce the possibility that the populations of Medicaid patients were qualitatively very different between time periods. Some patient attributes, like age and medical complexity, could still change between time periods, but we could be more confident that any changes in network structure were not be driven by attributes of a largely different patient population.

We could also recreate the care sector network based on performing providers who were active during both time periods rather than billing providers. It seems reasonable to expect that performing providers would change less in the range of services they offered than billing providers would change in the range of services they billed for – especially at clinics who became certified PCPCHs after Health Share’s formation. Based on this assumption, it seems likely that creation of a care sector network from the performing provider level might provide a more accurate partitioning of care delivery into our healthcare sectors of interest.

Another extension of this project would be to look further into the past and into the future, or to break up our claims data into smaller time segments, so that we could evaluate the magnitude and nature of network change that was occurring outside the context of CCO formation. For example, if we observed similar trends of connectivity consolidating at the start of the pre-period (late 2011) and at the end of the post period (late 2013), that should make us skeptical that it was caused by Health Share’s formation.
in 2012. On the other hand, if consolidation does look to be temporally associated with CCO formation, we might look further into the future to see whether its effect was temporary or long lasting.

Substantive contributions can also be extended in future work by digging deeper into different areas of the billing provider network. For example, future work can measure change in connectivity among a subset of billing providers in the network’s periphery to complement our analysis of change within the core top 30 billing providers. We can also explore changes among billing providers who were grouped in specific health care sectors, for example to see whether implementation of the PCPCH was followed by changes in billing provider connectivity within primary or specialty care sectors.

Lastly, future work could benefit by examining only a subset of our claims data which is restricted to patients of specific interest, such as those with chronic conditions or those with especially high levels of healthcare utilization. Patients with chronic conditions were flagged in the data resources that were acquired from Providence CORE (see p.103), but due to time constraints they were not utilized to infer and test additional definitions of the healthcare delivery network. Future research using only claims from this subpopulation could augment our substantive aims by indicating whether the billing provider network changed differentially for these subpopulations, and it could also augment our methodological aims by indicating whether inference methods respond differently to datasets that are more or less sparse. Simulation techniques, such as Monte Carlo simulation, could even be used with random subsets of patients to help us measure
how data sparsity may impact the performance of each network inference and resampling technique.

**Conclusion**

This project has developed the I-DNA method and has successfully demonstrated the use of I-DNA and R-DNA methods to statistically test change in a network of billing providers after Health Share of Oregon formed as a CCO. These applications have demonstrated the usefulness of I-DNA and R-DNA for social network change detection and have generated recommendations for their application with five different methods of network inference as well as three statistical resampling techniques. The research project has also produced statistical evidence to suggest that Health Share’s network of healthcare delivery changed significantly after it formed as a CCO, in network-wide patterns of connectivity as well as in specific pairwise connections between sectors of care.

In addition to statistical detection, our applications of I-DNA and R-DNA provided several insights regarding the nature of changes that occurred in Health Share’s billing provider network. Connectivity in the top 30 network was found to consolidate over time, producing a core that was more strongly connected and a periphery that was more weakly connected after Health Share’s formation. Changes in the relationship between primary and mental/behavioral health sectors indicated that referrals from the mental/behavioral sector to primary care providers may not have occurred as much as expected. And changes in the relationship between primary and specialty care sectors indicated an increase in the proportion of patients who were not billed for by providers in
either sector, calling into question whether patients were accessing primary care within other sectors or whether they may have faced increased obstacles to access.

In conclusion, this project has shown that I-DNA and R-DNA are effective tools for statistical detection of social network change. Both methods demonstrated success at inferring network structures, measuring change over time, and testing that change for statistical significance. And their joint application to the same datasets was found to be useful for building a more complete picture of the network changes that occurred. These methods are recommended for use in future research projects on social network phenomena, where they are expected to provide statistical capacities as well as descriptive insights regarding change in social networks.
References


https://doi.org/10.1137/070710111


https://doi.org/10.1145/511334.511337


Oregon Health Authority. (n.d.). Coordinated Care Organizations. Retrieved December 31, 2013, from Coordinated Care Organizations website:


Oregon Health Policy Board. (n.d.). Providers services contact information. Retrieved December 11, 2015, from Coordinated Care Organizations website:


https://doi.org/10.1016/j.socnet.2011.10.001


https://doi.org/10.1108/03684920410534010


https://doi.org/10.1016/j.socnet.2013.09.003


https://doi.org/10.1093/bioinformatics/btn660


Appendix A: Categorization of NUCC Fields

Categorization of NUCC Grouping, Classification, and Specialization fields into Primary, Specialty, Mental/Behavioral, Facility, and ‘Other’ provider types

<table>
<thead>
<tr>
<th>Primary Care</th>
</tr>
</thead>
<tbody>
<tr>
<td>Some Allopathic &amp; Osteopathic Physicians when practicing in Family Medicine, Internal Medicine, Pediatrics, or General Practice</td>
</tr>
<tr>
<td>Some Other Service Providers, when Naturopath, Midwife, or Midwife, Lay</td>
</tr>
<tr>
<td>Some Physician Assistants &amp; Advanced Practice Nursing Providers when Physician Assistant (unless specializing in Psych/Mental Health), Nurse Practitioner, Advanced Practice Midwife, or Clinical Nurse Specialist who specializes in Acute Care, Adult Health, Critical Care medicine, Community Health/Public Health, Chronic Care, Emergency, Ethics, Family Health, Gerontology, Home Health, Holistic, Informatics, Long-Term Care, Medical-Surgical, Neonatal, Neuroscience, Pediatrics, Perinatal, Perioperative, Rehabilitation, School, Transplantation, Women’s Health, Occupational Health, Oncology, Oncology (Pediatrics), or with blank specialization</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ancillary Care*</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Chiropractic Providers</td>
</tr>
<tr>
<td>All Dietary &amp; Nutritional Service Providers</td>
</tr>
<tr>
<td>All Eye and Vision Service Providers</td>
</tr>
<tr>
<td>Some Other Service Providers when in Medical Genetics (PhD), Genetic Counselor (MS), Military Health Care Provider, Acupuncturist, Case Manager/Care Coordinator, Community Health Worker, Sleep Specialist (PhD), Health Educator, Lactation Consultant, Clinical Ethicist, or Homeopath</td>
</tr>
<tr>
<td>All Pharmacy Service Providers</td>
</tr>
<tr>
<td>All Speech, Language, and Hearing Service Providers</td>
</tr>
<tr>
<td>All Respiratory, Developmental, Rehabilitative and Restorative Service Providers</td>
</tr>
<tr>
<td>All Laboratories</td>
</tr>
</tbody>
</table>

* Defined as Care that supports the services of a Physician

<table>
<thead>
<tr>
<th>Specialty Care</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Emergency Medical Service Providers</td>
</tr>
<tr>
<td>Some Allopathic &amp; Osteopathic Physicians when Independent Medical Examiner, Phlebology, Neuromusculoskeletal Medicine, Oral &amp; Maxillofacial Surgery, Transplant Surgery, Electrodagnostic Medicine, Allergy &amp; Immunology, Anesthesiology, Dermatology, Emergency Medicine, Medical Genetics, Neurological Surgery, Nuclear Medicine, Obstetrics &amp; Gynecology, Ophthalmology, Orthopaedic Surgery, Otolaryngology, Pathology, Physical Medicine &amp; Rehabilitation, Plastic Surgery, Preventive Medicine, Radiology, Surgery, Urology, Colon &amp; Rectal Surgery, Thoracic Surgery, Hospitalist, Clinical Pharmacology, Pain Medicine, or Legal Medicine</td>
</tr>
<tr>
<td>Some Physician Assistants &amp; Advanced Practice Nursing Providers when a Nurse Anesthetist (Certified Registered), Anesthesiologist Assistant</td>
</tr>
<tr>
<td>All Podiatric Medicine &amp; Surgery Service Providers</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mental/Behavioral Care</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Behavioral Health &amp; Social Service Providers (includes Counselor, Psychoanalyst, Poetry Therapist, Clinical Neuropsychologist, Behavioral Analyst, Psychologist, Social Worker, and Marriage &amp; Family Therapist)</td>
</tr>
<tr>
<td>Some Allopathic &amp; Osteopathic Physicians, when in (Psychiatry &amp; Neurology)</td>
</tr>
<tr>
<td>Some Physician Assistants &amp; Advanced Practice Nursing Providers when a Nurse Practitioner or Clinical Nurse Specialist who specializes in Psych/Mental Health</td>
</tr>
<tr>
<td>Facility Care</td>
</tr>
<tr>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>All Hospital Units</td>
</tr>
<tr>
<td>All Ambulatory Health Care Facilities</td>
</tr>
<tr>
<td>All Hospitals</td>
</tr>
<tr>
<td>All Residential Treatment Facilities</td>
</tr>
<tr>
<td>All Respite Care Facilities</td>
</tr>
<tr>
<td>All Agencies</td>
</tr>
<tr>
<td>All Managed Care Organizations</td>
</tr>
<tr>
<td>All Nursing &amp; Custodial Care Facilities</td>
</tr>
<tr>
<td>All Suppliers</td>
</tr>
<tr>
<td>All Transportation Services</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Other Care</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Nursing Service Providers</td>
</tr>
<tr>
<td>All Dental Providers</td>
</tr>
<tr>
<td>Some Other Service Providers when Interpreter, Contractor, Driver, Mechanotherapist, Naprapath, Legal Medicine, Reflexologist, Meals, Veterinarian, Funeral Director, Lodging, or Specialist (Graphics Designer, Prosthetics Case Management, Research Study, Research Data Abstract)</td>
</tr>
<tr>
<td>All Technologists, Technicians &amp; Other Technical Service Providers</td>
</tr>
<tr>
<td>All Nursing Service Related Providers</td>
</tr>
<tr>
<td>All Student, Health Care</td>
</tr>
<tr>
<td>All Group</td>
</tr>
</tbody>
</table>
### Appendix B: Relationships in the Top 30 Networks by RA Inference

#### RA Associations Before Health Share Formed

<table>
<thead>
<tr>
<th>Rank</th>
<th>Relationship</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Providence 1 – Quest Diagnostics – Multnomah County – Providence 5</td>
</tr>
<tr>
<td>2</td>
<td>Providence 1 – Quest Diagnostics – Multnomah County – Epic Imaging</td>
</tr>
<tr>
<td>3</td>
<td>Providence 1 – Kaiser – Tuality – Multnomah County – Virginia Garcia</td>
</tr>
<tr>
<td>4</td>
<td>Providence 1 – Kaiser – Tuality – Multnomah County – Providence 5</td>
</tr>
<tr>
<td>5</td>
<td>Providence 1 – Northwest Acute – Tuality – AMR NW</td>
</tr>
<tr>
<td>6</td>
<td>Providence 1 – Northwest Acute – AMR NW – Oregon Anesthesiology</td>
</tr>
<tr>
<td>7</td>
<td>Providence 1 – Oregon Clinic – Radiology Specialists</td>
</tr>
<tr>
<td>8</td>
<td>Providence 1 – Tuality – Providence 4 – Multnomah County</td>
</tr>
<tr>
<td>9</td>
<td>Providence 1 – Tuality – Providence 4 – Radiology Specialists</td>
</tr>
<tr>
<td>10</td>
<td>Providence 1 – Tuality – AMR NW – Radiology Specialists</td>
</tr>
<tr>
<td>11</td>
<td>Radiology Consultants – Northwest Acute – Legacy Emanuel 2 – Legacy Clinics</td>
</tr>
<tr>
<td>12</td>
<td>Diagnostic Radiology – CEP America – Adventist 1 – Adventist 2</td>
</tr>
<tr>
<td>13</td>
<td>CEP America – Adventist 1 – Tuality – AMR NW</td>
</tr>
<tr>
<td>14</td>
<td>Providence 2 – Multnomah County – Providence 5</td>
</tr>
<tr>
<td>15</td>
<td>Providence 3 – Kaiser – Multnomah County – Virginia Garcia</td>
</tr>
<tr>
<td>16</td>
<td>Providence 3 – TRG LLC – Multnomah County</td>
</tr>
<tr>
<td>17</td>
<td>Kaiser – OHSU 2 – Multnomah County – Providence 5</td>
</tr>
<tr>
<td>18</td>
<td>Adventist 1 – Northwest Acute – Tuality – AMR NW</td>
</tr>
<tr>
<td>19</td>
<td>Northwest Acute – Legacy Mt Hood – Legacy Emanuel 2 – Legacy Clinics</td>
</tr>
<tr>
<td>20</td>
<td>Northwest Acute – AMR NW – Oregon Anesthesiology – Legacy Clinics</td>
</tr>
<tr>
<td>21</td>
<td>Northwest Acute – Oregon Anesthesiology – Legacy Emanuel 2 – Legacy Clinics</td>
</tr>
<tr>
<td>22</td>
<td>Legacy Mt Hood – Legacy Emanuel 2 – Legacy Lab Services – Legacy Clinics</td>
</tr>
<tr>
<td>23</td>
<td>Tuality – Lifeworks NA – Multnomah County</td>
</tr>
<tr>
<td>24</td>
<td>OHSU 1 – OHSU 2</td>
</tr>
<tr>
<td>25</td>
<td>Legacy Emanuel 1 – Legacy Emanuel 2 – Legacy Lab Services</td>
</tr>
</tbody>
</table>

#### RA Associations After Health Share Formed

<table>
<thead>
<tr>
<th>Rank</th>
<th>Relationship</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Providence 1 – Providence 2 – Quest Diagnostics – Multnomah County</td>
</tr>
<tr>
<td>2</td>
<td>Providence 1 – Providence 2 – Tuality – OHSU 2 – Multnomah County</td>
</tr>
<tr>
<td>3</td>
<td>Providence 1 – Quest Diagnostics – Multnomah County – Epic Imaging</td>
</tr>
<tr>
<td>4</td>
<td>Providence 1 – Oregon Clinic – Oregon Anesthesiology</td>
</tr>
<tr>
<td>5</td>
<td>Providence 1 – Oregon Clinic – Multnomah County – Radiology Specialists</td>
</tr>
<tr>
<td>6</td>
<td>Providence 1 – Tuality – Providence 4 – Multnomah County</td>
</tr>
<tr>
<td>7</td>
<td>Providence 1 – Providence 4 – Multnomah County – Radiology Specialists</td>
</tr>
<tr>
<td>8</td>
<td>Providence 1 – Lifeworks NA – Multnomah County</td>
</tr>
<tr>
<td>9</td>
<td>Providence 1 – Oregon Anesthesiology – Legacy Emanuel 2</td>
</tr>
<tr>
<td>10</td>
<td>Radiology Consultants – Legacy Emanuel 2 – Legacy Clinics</td>
</tr>
<tr>
<td>11</td>
<td>Diagnostic Radiology – CEP America – Adventist 1 – Adventist 2</td>
</tr>
<tr>
<td>12</td>
<td>CEP America – Adventist 1 – AMR NW</td>
</tr>
<tr>
<td>13</td>
<td>Providence 2 – Providence 3 – TRG LLC</td>
</tr>
<tr>
<td>14</td>
<td>Providence 2 – Providence 3 – OHSU 2 – Multnomah County</td>
</tr>
<tr>
<td>15</td>
<td>Kaiser – Providence 4</td>
</tr>
<tr>
<td>16</td>
<td>Adventist 1 – AMR NW – Oregon Anesthesiology</td>
</tr>
<tr>
<td>17</td>
<td>Northwest Acute – Legacy Mt Hood – Legacy Emanuel 2 – Legacy Lab Services – Legacy Clinics</td>
</tr>
<tr>
<td>18</td>
<td>Northwest Acute – AMR NW – Oregon Anesthesiology – Legacy Clinics</td>
</tr>
<tr>
<td>19</td>
<td>Northwest Acute – Oregon Anesthesiology – Legacy Emanuel 2 – Legacy Clinics</td>
</tr>
<tr>
<td>20</td>
<td>Tuality – OHSU 1 – OHSU 2</td>
</tr>
<tr>
<td>21</td>
<td>Tuality – Multnomah County – Virginia Garcia</td>
</tr>
<tr>
<td>22</td>
<td>Providence 4 – Providence 5</td>
</tr>
<tr>
<td>23</td>
<td>Legacy Emanuel 1 – Legacy Emanuel 2 – Legacy Lab Services</td>
</tr>
</tbody>
</table>
## Appendix C: Resampling Distribution Information for Statistical Testing

### Change in the Top 30 Network

<table>
<thead>
<tr>
<th></th>
<th>Correlation</th>
<th>Multiple Regression</th>
<th>PLS</th>
<th>SBT</th>
<th>RA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed distance ($\theta$)</td>
<td>0.01841</td>
<td>0.02456</td>
<td>0.01560</td>
<td>0.00029</td>
<td>0.438142</td>
</tr>
</tbody>
</table>

#### Permutation resampled reference distribution

<table>
<thead>
<tr>
<th></th>
<th>$&lt;.01^*$</th>
<th>$&lt;.01^*$</th>
<th>$&lt;.01^*$</th>
<th>$&lt;.01^*$</th>
<th>$&lt;.01^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.00730</td>
<td>0.01225</td>
<td>0.00617</td>
<td>4.13E-05</td>
<td>0.13190</td>
</tr>
<tr>
<td>Standard deviation (SD)</td>
<td>0.00054</td>
<td>0.00078</td>
<td>0.00067</td>
<td>3.22E-06</td>
<td>0.02424</td>
</tr>
<tr>
<td>Ratio of SD to $\theta$</td>
<td>0.02900</td>
<td>0.03200</td>
<td>0.04300</td>
<td>0.01100</td>
<td>0.05500</td>
</tr>
<tr>
<td>Median</td>
<td>0.00728</td>
<td>0.01218</td>
<td>0.00612</td>
<td>4.09E-05</td>
<td>0.14712</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.21547</td>
<td>0.42670</td>
<td>0.86833</td>
<td>0.58531</td>
<td>-0.26311</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>3.04967</td>
<td>3.03105</td>
<td>4.07046</td>
<td>3.72929</td>
<td>1.33297</td>
</tr>
<tr>
<td>Shapiro-Wilk (SW)</td>
<td>0.98958</td>
<td>0.98191</td>
<td>0.95327</td>
<td>0.97467</td>
<td>0.83799</td>
</tr>
<tr>
<td>p value for SW$^1$</td>
<td>0.63054</td>
<td>0.18672</td>
<td>0.00137*</td>
<td>0.05087</td>
<td>4.38E-09*</td>
</tr>
</tbody>
</table>

#### Single bootstrap resampled reference distribution

<table>
<thead>
<tr>
<th></th>
<th>$&lt;.01^*$</th>
<th>$&lt;.01^*$</th>
<th>$&lt;.01^*$</th>
<th>$&lt;.01^*$</th>
<th>$&lt;.01^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.00713</td>
<td>0.01187</td>
<td>0.00609</td>
<td>4.15E-05</td>
<td>0.10844</td>
</tr>
<tr>
<td>Standard deviation (SD)</td>
<td>0.00061</td>
<td>0.00077</td>
<td>0.00064</td>
<td>3.48E-06</td>
<td>0.02393</td>
</tr>
<tr>
<td>Ratio of SD to $\theta$</td>
<td>0.03300</td>
<td>0.03100</td>
<td>0.04100</td>
<td>0.01200</td>
<td>0.05500</td>
</tr>
<tr>
<td>Median</td>
<td>0.00712</td>
<td>0.01193</td>
<td>0.00607</td>
<td>4.12E-05</td>
<td>0.10010</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.16334</td>
<td>-0.28648</td>
<td>0.28983</td>
<td>0.09511</td>
<td>0.12981</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>3.37695</td>
<td>3.69609</td>
<td>2.45057</td>
<td>2.86950</td>
<td>1.44092</td>
</tr>
<tr>
<td>Shapiro-Wilk (SW)</td>
<td>0.99350</td>
<td>0.97769</td>
<td>0.98152</td>
<td>0.98308</td>
<td>0.89034</td>
</tr>
<tr>
<td>p value for SW$^1$</td>
<td>0.91656</td>
<td>0.08764</td>
<td>0.17421</td>
<td>0.22902</td>
<td>5.10E-07*</td>
</tr>
</tbody>
</table>

#### Double bootstrap resampled reference distribution

<table>
<thead>
<tr>
<th></th>
<th>4.54E-41*</th>
<th>4.65E-41*</th>
<th>1.10E-23*</th>
<th>1.29E-85*</th>
<th>3.42E-67*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.02030</td>
<td>0.02832</td>
<td>0.01705</td>
<td>0.00030</td>
<td>0.43927</td>
</tr>
<tr>
<td>Standard deviation (SD)</td>
<td>1.10300</td>
<td>1.15300</td>
<td>1.09300</td>
<td>1.02100</td>
<td>1.00300</td>
</tr>
<tr>
<td>Ratio of SD to $\theta$</td>
<td>0.00082</td>
<td>0.00109</td>
<td>0.00119</td>
<td>4.26E-06</td>
<td>0.00994</td>
</tr>
<tr>
<td>Median</td>
<td>0.04400</td>
<td>0.04400</td>
<td>0.07600</td>
<td>0.01500</td>
<td>0.02300</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.02030</td>
<td>0.02829</td>
<td>0.01677</td>
<td>0.00030</td>
<td>0.44123</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>0.16918</td>
<td>0.08078</td>
<td>0.68354</td>
<td>0.32251</td>
<td>-0.56565</td>
</tr>
<tr>
<td>Shapiro-Wilk (SW)</td>
<td>2.84066</td>
<td>2.85024</td>
<td>2.78251</td>
<td>2.90656</td>
<td>3.09513</td>
</tr>
<tr>
<td>p value for SW$^1$</td>
<td>0.99426</td>
<td>0.99364</td>
<td>0.94421</td>
<td>0.98785</td>
<td>0.97135</td>
</tr>
</tbody>
</table>

$^*$ indicate where the probability of a type I error was less than 0.05

$^1$ statistical significance of the Shapiro-Wilk test indicates a deviation from a normal distribution.
## Change in the Care Sector Network

### Permutation resampled reference distribution

<table>
<thead>
<tr>
<th></th>
<th>Correlation</th>
<th>PLS</th>
<th>SBT</th>
<th>RA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed distance ($\theta$)</td>
<td>0.04079</td>
<td>0.01807</td>
<td>0.00054</td>
<td>0.15538</td>
</tr>
</tbody>
</table>

### Single bootstrap resampled reference distribution

<table>
<thead>
<tr>
<th></th>
<th>&lt;.01*</th>
<th>&lt;.01*</th>
<th>&lt;.01*</th>
<th>&lt;.01*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.00300</td>
<td>0.00738</td>
<td>0.00007</td>
<td>0.03184</td>
</tr>
<tr>
<td>Standard deviation (SD)</td>
<td>0.00065</td>
<td>0.00203</td>
<td>0.00002</td>
<td>0.00226</td>
</tr>
<tr>
<td>Ratio of SD to $\theta$</td>
<td>0.01605</td>
<td>0.11234</td>
<td>0.02911</td>
<td>0.01457</td>
</tr>
<tr>
<td>Median</td>
<td>0.00286</td>
<td>0.00743</td>
<td>0.00007</td>
<td>0.03167</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.66028</td>
<td>0.30672</td>
<td>0.42812</td>
<td>0.31830</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>2.98698</td>
<td>2.71049</td>
<td>2.47823</td>
<td>2.32230</td>
</tr>
<tr>
<td>Shapiro-Wilk (SW)</td>
<td>0.95747</td>
<td>0.98749</td>
<td>0.97120</td>
<td>0.97605</td>
</tr>
<tr>
<td>Mean for SW¹</td>
<td>0.00267*</td>
<td>0.47142</td>
<td>0.02733*</td>
<td>0.06517</td>
</tr>
</tbody>
</table>

### Double bootstrap resampled reference distribution

<table>
<thead>
<tr>
<th></th>
<th>5.74E-56*</th>
<th>1.13E-06*</th>
<th>1.26E-37*</th>
<th>5.66E-70*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.04076</td>
<td>0.02089</td>
<td>0.00054</td>
<td>0.18156</td>
</tr>
<tr>
<td>Standard deviation (SD)</td>
<td>0.00122</td>
<td>0.00360</td>
<td>0.00003</td>
<td>0.00376</td>
</tr>
<tr>
<td>Ratio of SD to $\theta$</td>
<td>0.02997</td>
<td>0.19914</td>
<td>0.04912</td>
<td>0.02418</td>
</tr>
<tr>
<td>Median</td>
<td>0.04081</td>
<td>0.02068</td>
<td>0.00054</td>
<td>0.18136</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.15749</td>
<td>0.56227</td>
<td>-0.22961</td>
<td>0.23578</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>3.80937</td>
<td>3.75530</td>
<td>2.91339</td>
<td>4.15752</td>
</tr>
<tr>
<td>Shapiro-Wilk (SW)</td>
<td>0.99064</td>
<td>0.97530</td>
<td>0.98642</td>
<td>0.97876</td>
</tr>
<tr>
<td>Mean for SW¹</td>
<td>0.71607</td>
<td>0.05693</td>
<td>0.39956</td>
<td>0.10645</td>
</tr>
</tbody>
</table>

* indicate where the probability of a type I error was less than 0.05
¹ statistical significance of the Shapiro-Wilk test indicates a deviation from a normal distribution
<table>
<thead>
<tr>
<th></th>
<th>Correlation</th>
<th>PLS P→S</th>
<th>PLS S→P</th>
<th>SBT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed distance (θ)</td>
<td>0.03587</td>
<td>0.01305</td>
<td>0.02585</td>
<td>-4.5E-06</td>
</tr>
</tbody>
</table>

**Permutation resampled reference distribution**

<table>
<thead>
<tr>
<th></th>
<th>p value</th>
<th>0.49</th>
<th>0.29</th>
<th>0.1</th>
<th>0.32</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>-6.68E-06</td>
<td>-0.00153</td>
<td>0.00728</td>
<td>-7.59E-06</td>
<td></td>
</tr>
<tr>
<td>Standard deviation (SD)</td>
<td>0.00328</td>
<td>0.01139</td>
<td>0.03475</td>
<td>0.00010</td>
<td></td>
</tr>
<tr>
<td>Ratio of SD to θ</td>
<td>0.09131</td>
<td>0.87331</td>
<td>1.34441</td>
<td>-21.90756</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>-0.00021</td>
<td>-0.00162</td>
<td>0.00108</td>
<td>-0.00001</td>
<td></td>
</tr>
<tr>
<td>Skewness</td>
<td>0.11094</td>
<td>0.03960</td>
<td>0.29051</td>
<td>-0.07474</td>
<td></td>
</tr>
<tr>
<td>Kurtosis</td>
<td>2.39261</td>
<td>3.14680</td>
<td>2.49179</td>
<td>2.45944</td>
<td></td>
</tr>
<tr>
<td>Shapiro-Wilk (SW)</td>
<td>0.98783</td>
<td>0.99082</td>
<td>0.98120</td>
<td>0.98547</td>
<td></td>
</tr>
<tr>
<td>p value for SW¹</td>
<td>0.49595</td>
<td>0.73066</td>
<td>0.16480</td>
<td>0.34295</td>
<td></td>
</tr>
</tbody>
</table>

**Single bootstrap resampled reference distribution**

<table>
<thead>
<tr>
<th></th>
<th>p value</th>
<th>0.44</th>
<th>0.2</th>
<th>0.14</th>
<th>0.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>-0.00037</td>
<td>-0.00046</td>
<td>0.00502</td>
<td>-1.77E-05</td>
<td></td>
</tr>
<tr>
<td>Standard deviation (SD)</td>
<td>0.00408</td>
<td>0.01342</td>
<td>0.03322</td>
<td>0.00010</td>
<td></td>
</tr>
<tr>
<td>Ratio of SD to θ</td>
<td>0.11360</td>
<td>1.02852</td>
<td>1.28550</td>
<td>-22.81884</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>-0.00069</td>
<td>-0.00026</td>
<td>0.00278</td>
<td>-1.79E-05</td>
<td></td>
</tr>
<tr>
<td>Skewness</td>
<td>0.32101</td>
<td>0.09870</td>
<td>0.24330</td>
<td>-0.17453</td>
<td></td>
</tr>
<tr>
<td>Kurtosis</td>
<td>3.34001</td>
<td>2.79114</td>
<td>3.12554</td>
<td>2.97301</td>
<td></td>
</tr>
<tr>
<td>Shapiro-Wilk (SW)</td>
<td>0.98927</td>
<td>0.99274</td>
<td>0.98818</td>
<td>0.99169</td>
<td></td>
</tr>
<tr>
<td>p value for SW¹</td>
<td>0.60575</td>
<td>0.87163</td>
<td>0.52133</td>
<td>0.79852</td>
<td></td>
</tr>
</tbody>
</table>

**Double bootstrap resampled reference distribution**

<table>
<thead>
<tr>
<th></th>
<th>p value</th>
<th>8.33E-17</th>
<th>0.17934</th>
<th>0.20717</th>
<th>0.51575</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.03626</td>
<td>0.01398</td>
<td>0.02987</td>
<td>9.14E-06</td>
<td></td>
</tr>
<tr>
<td>Standard deviation (SD)</td>
<td>0.00362</td>
<td>0.01415</td>
<td>0.03153</td>
<td>0.00011</td>
<td></td>
</tr>
<tr>
<td>Ratio of SD to θ</td>
<td>0.101</td>
<td>1.084</td>
<td>1.220</td>
<td>-25.253</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>0.03612</td>
<td>0.01282</td>
<td>0.03143</td>
<td>5.63E-06</td>
<td></td>
</tr>
<tr>
<td>Skewness</td>
<td>0.25134</td>
<td>0.32640</td>
<td>-0.12510</td>
<td>0.18116</td>
<td></td>
</tr>
<tr>
<td>Kurtosis</td>
<td>3.39148</td>
<td>3.02853</td>
<td>2.61816</td>
<td>3.83191</td>
<td></td>
</tr>
<tr>
<td>Shapiro-Wilk (SW)</td>
<td>0.98871</td>
<td>0.98965</td>
<td>0.99040</td>
<td>0.98565</td>
<td></td>
</tr>
<tr>
<td>p value for SW¹</td>
<td>0.56201</td>
<td>0.63617</td>
<td>0.69698</td>
<td>0.35336</td>
<td></td>
</tr>
</tbody>
</table>

* indicate where the probability of a type I error was less than 0.05

¹ statistical significance of the Shapiro-Wilk test indicates a deviation from a normal distribution
Connectivity between Primary (P) and Mental/Behavioral (M) Sectors

<table>
<thead>
<tr>
<th></th>
<th>Correlation</th>
<th>PLS P→M</th>
<th>PLS M→P</th>
<th>SBT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed distance (θ)</td>
<td>-0.08454</td>
<td>0.02256</td>
<td>-5.353E-05</td>
<td>0.0030</td>
</tr>
</tbody>
</table>

### Permutation resampled reference distribution

<table>
<thead>
<tr>
<th></th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Standard deviation (SD)</td>
<td></td>
</tr>
<tr>
<td>Ratio of SD to θ</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Median</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Skewness</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Shapiro-Wilk (SW)</td>
<td>&lt;.01</td>
</tr>
</tbody>
</table>

### Single bootstrap resampled reference distribution

<table>
<thead>
<tr>
<th></th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Standard deviation (SD)</td>
<td></td>
</tr>
<tr>
<td>Ratio of SD to θ</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Median</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Skewness</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Shapiro-Wilk (SW)</td>
<td>&lt;.01</td>
</tr>
</tbody>
</table>

### Double bootstrap resampled reference distribution

<table>
<thead>
<tr>
<th></th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>1.27E-39</td>
</tr>
<tr>
<td>Standard deviation (SD)</td>
<td></td>
</tr>
<tr>
<td>Ratio of SD to θ</td>
<td>1.27E-39</td>
</tr>
<tr>
<td>Median</td>
<td>1.27E-39</td>
</tr>
<tr>
<td>Skewness</td>
<td>1.27E-39</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>1.27E-39</td>
</tr>
<tr>
<td>Shapiro-Wilk (SW)</td>
<td>1.27E-39</td>
</tr>
</tbody>
</table>

* indicate where the probability of a type I error was less than 0.05

Statistical significance of the Shapiro-Wilk test indicates a deviation from a normal distribution.