Supplemental Online Appendix: Modeling the association between physician risky-prescribing and the complex network structure of physician shared-patient relationships

Xin Ran[1,2], Ellen R. Meara[3], Nancy E. Morden[2,4], Erika L. Moen[1,2], Daniel N. Rockmore[5,6], and A. James O'Malley[1,2]

[1] Department of Biomedical Data Science, Geisel School of Medicine at Dartmouth, Lebanon, NH, USA
[2] The Dartmouth Institute for Health Policy and Clinical Practice, Geisel School of Medicine at Dartmouth, Lebanon, NH, USA
[3] Department of Health Policy and Management, Harvard T.H. Chan School of Public Health, Boston, MA, USA
[4] United HealthCare, Minnetonka, MN, USA
[5] Department of Mathematics and Department of Computer Science, Dartmouth College, Hanover, NH, USA
[6] The Santa Fe Institute, Santa Fe, NM, USA

1 Supplemental methods

1.1 Materials and overview of risky prescribing study

In this study, a 40% random sample of Medicare Part D claims (prescription drug events) from 2014 was used to retrieve beneficiaries' prescription fill records and their corresponding prescribers, for three classes of risky drugs: opioids, benzodiazepines, and sedative-hypnotics. The beneficiaries' prescription records including the physicians who prescribed their drugs from Medicare Part D claims were used to trace the trajectories of patients' prescriptions and to construct physicians' prescribing indexes. See the beginning of Section 2 and Figure S1 in for more details on the physician prescribing indexes and the study cohort definition and workflow. Separately, we used a 40% random sample of all Medicare fee-for-service claims from 2014 of beneficiaries residing in the state of Ohio to extract relevant physician-patient encounters for constructing the physician network. Because the complete physician-patient encounter data was only available for patients residing in the state of Ohio, we limited our shared-patient physician network to physicians caring for Ohio residents in 2014.

1.2 Construction of shared-patient physician network

A unipartite physician network is constructed from the ensemble of physician-patient encounters observed in the Medicare fee-for-service claims data. A visit to physician i followed by another visit to physician j by the same patient within

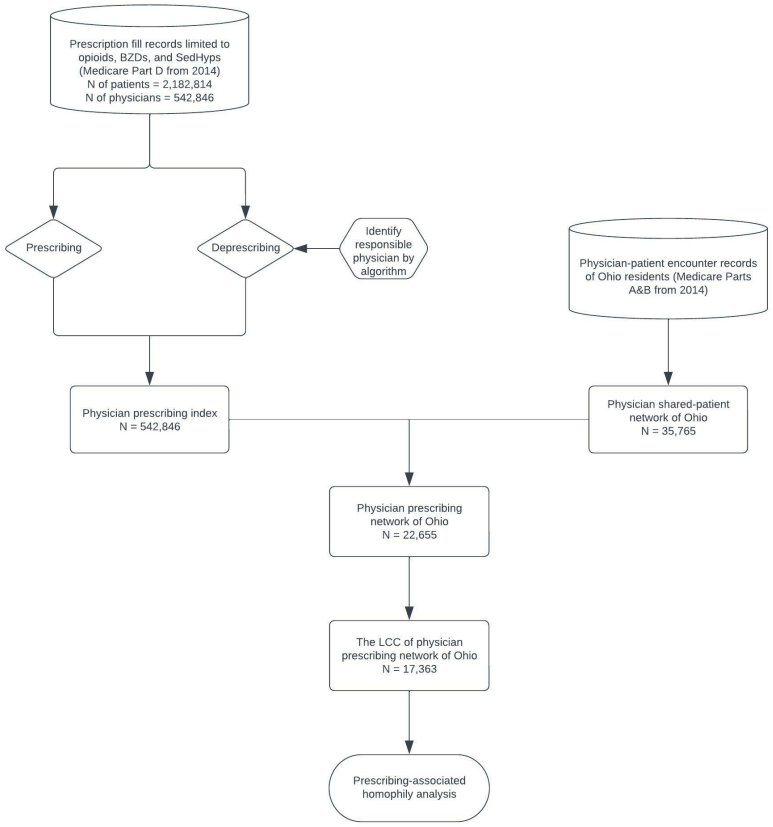


Fig. S1: Study cohort definition and workflow

Note: Beneficiaries were included in the study if they had at least 3 months of continuous coverage of Medicare Parts A, B, and D, and 2 years of continuous Parts A and B coverage prior to cohort entry. LCC = largest connected component.

a certain time window may provide evidence for a meaningful professional relationship("patient referral") from physician i to j [9,2,1]. A lengthy time window of 365 days is used to allow gaps between physician visits to span months, minimizing the chance that a true directed relationship is missed. However, because a dyad with a bidirectional relationship is most likely to involve physicians who know each other and who have made a deliberate choice to refer their patients to one another, we restricted the connections between physician pairs to only mutual ties (those dyads with directed edges in both directions). Next, we transformed the network to a binary network using a threshold of 0; thus, sharing at least one patient in each direction constitutes an edge in the network. We limit the network to physicians who had at least prescribed one drug in the drug

classes of interest, including opioids, benzodiazepines, or sedative-hypnotics in 2014 according to Medicare Part D data. Finally, we reduce the network to its largest connected component (LCC) to eliminate isolated dyads (pairs of physicians who only shared patients with each other and thus have a network degree of 1) as such physicians are likely practicing in a part-time or otherwise reduced or restricted manner.

To further study homophily in hospital referral regions (HRRs) and the possible variation of risky-prescribing-associated homophily across different HRRs, we divided the LCC of the physician prescribing network into HRR sub-networks according to where the majority of their patients reside based on their Medicare fee-for-service claims in 2014. Both physicians must belong to the same HRR in order for the edge between them to be included in that HRR sub-network.

1.3 Homophily statistics: Isolating their effect and associated identifiability results

Several ERGM terms are available to study the homophily of continuous and categorical attributes. When studying the homophily of an attribute, it is important to adjust for the main effect of the node attribute to ensure that homophily is a relative measure as opposed to being impacted by the prevalence of the attribute across the network. The network statistics associated with these main effects are often named *nodecov* (for continuous attributes) and *nodefactor* (for categorical attributes) in ERGMs. Therefore, homophily can be thought of as a within-dyad interaction between the two nodes comprising the dyad of the given attribute [6,8]. Table S1 shows the mathematical definitions of the ERGM terms used in this study and their respective interpretation. When estimating the homophily effect while adjusting for the node-level effect of the same attributes, only the effect of uniform homophily (the same effect of having the attribute in common across all levels of the attribute) can be estimated. Otherwise, the coefficients of the network statistics will be unidentifiable due to the linear dependency between the predictors. In the following, we demonstrate the mathematics of such linear dependency when estimating differential homophily while controlling for network density and the main effect of the node attribute. As displayed in Table S1, for a binary node attribute taking a value of 0 or 1, the network statistic added to the model by the *nodefactor* term is equivalent to,

$$\frac{1}{2}\sum_{ij}(x_i + x_j)a_{ij}. \tag{1}$$

The two statistics added by the differential homophily term *nodematch* are,

$$\frac{1}{2}\sum_{ij} x_i x_j a_{ij} \tag{2}$$

for the node attribute taking a value of 1, and

$$\frac{1}{2}\sum_{ij}(1 - x_i)(1 - x_j)a_{ij} = \frac{1}{2}\sum_{ij}(1 - (x_i + x_j) + x_i x_j)a_{ij} \tag{3}$$

for the node attribute taking a value of 0. For a binary node attribute, three predictors in Equations 1, 2, and 3 are linearly dependent and the model is unidentifiable when including the *edges* term, the *nodefactor* term, and two *nodematch* terms of differential network homophily statistics in an ERGM for an undirected network. The result is observed by the fact that the *nodematch* term (Equation 3) is the sum of the *edges* (Table S1), *nodefactor* (Equation 1), and the 1-level *nodematch* (Equation 2) terms. Therefore, when controlling for the network density with the *edges* term and the main effect of an attribute with the *nodefactor* term, the uniform homophily statistic $\frac{1}{2}\sum_{ij}((1-x_i)(1-x_j)+x_ix_j)a_{ij}$ (the sum of the two *nodematch* levels) can be identified but its components (the two differential homophily terms) cannot.

Table S1: Definitions of ERGM terms for undirected networks and interpretations

Terms	Math definition	Interpretation
edges	$m = \frac{1}{2}\sum_{ij} a_{ij}$	Number of edges in the network; controls for network density
Node attribute		
nodefactor	$Z_k = \frac{1}{2}\sum_{ij}(\mathbf{1}(x_i = k) + \mathbf{1}(x_j = k))a_{ij}$	Number of times a node possessing a categorical attribute of value k appears on an edge in the network
nodecov	$Z = \frac{1}{2}\sum_{ij}(x_i + x_j)a_{ij}$	For continuous attributes, the sum of the attribute across node pairs for all edges present in network
Homophily term		
nodematch	$S = \frac{1}{2}\sum_{l}\sum_{ij}(\mathbf{1}(x_i = l)\mathbf{1}(x_j = l))a_{ij}$	Uniform homophily; the number of edges when two nodes have the same categorical attribute
nodematch	$S_l = \frac{1}{2}\sum_{ij}(\mathbf{1}(x_i = l)\mathbf{1}(x_j = l))a_{ij}$	Differential homophily; the number of edges when two nodes have the same categorical attribute taking a value of l
absdiff	$S = \frac{1}{2}\sum_{ij}\lvert x_i - x_j\rvert a_{ij}$	For continuous attributes, the sum of absolute difference in the attribute within a dyad across all edges present in the network

Note: $\mathbf{A} = [a_{ij}]$ is the adjacency matrix of the binary-undirected network in this study, and $a_{ij} = 1$ if physician i and j shared patients during 2014. The variables x_i and x_j are the node attributes of physician i and j, and k and l denote the values of a categorical attribute held by the two actors comprising a dyad.

We focus on the homophily of prescribing behaviors and homophily of physician specialty in the risky-prescribing-specific network while adjusting for network density and the above-mentioned network statistics. To avoid collinearity between different prescribing measures, the uniform homophily terms involving them were added one at a time in the model, adjusting in each model for network density, the differential effect of physician specialty on network density (i.e., the physician specialty *nodefactor* term), and the uniform homophily of physician

specialty. The R package statnet was used to implement the ERGM containing the associated network statistics as predictors [6,5].

2 Supplemental application to study of physician homophily in prescribing and deprescribing

2.1 Data preprocessing

To quantify physicians' prescribing behaviors, our primary interest is new prescription fills instead of refills because they represent a definitive step towards increased polypharmacy or risky drug combinations. We implemented an empirical rule of 20% overlapping fill length, where a subsequent prescription fill of the same drug written by the same physician was merged to the preceding fill if they overlapped or the gap in between is less than 20% of the fill length of the preceding prescription. It is also highly likely that a subsequent prescription of the same drug signed by a different physician that overlaps with this 20% buffer zone is still a refill of the preceding prescription. Therefore, subsequent prescription fills satisfying the 20% buffer were joined to the preceding fill and attributed to the initializing physician. Such preprocessing enabled us to reduce false positive discontinuations by distinguishing the discontinuation of a prescription from a temporary stop prior to a refill, as described in the following sections.

2.2 Modeling patient prescription states

For each of the patients, their prescription fills of the three-drug classes of interest were divided into discrete time intervals, capturing the initialization and discontinuation of a prescription, with each time interval reflecting the number and class of drugs they were exposed to. Based on the number and the class of drugs a patient was taking during each of the discrete time intervals, we assigned the patient to the $2^3 = 8$ combinations of prescription states. Figure S2 shows the workflow of modeling patient prescription states. Every initialization or discontinuation of a drug will lead to the changing of a patient's prescription state. State *zero* is a state of taking no drugs in the three-drug classes of interest. States $\{O, B, S\}$ correspond to taking at least one drug in precisely one of the opioids, benzodiazepines, or sedative hypnotics drug classes. States $\{OB, OS, BS\}$ correspond to taking at least two drugs concurrently involving two different drug classes. State OBS is to take at least three drugs of three different classes at the same time. Numbers from 1 to 8 were respectively assigned to these eight prescription states for the ease of mathematical notation in the following sections. A sequence of prescription states was thereby obtained for each patient, including both prescribing and deprescribing events.

2.3 Attributing physician responsibility to prescribing and deprescribing prescriptions

Several studies have quantified patients' receipt of polypharmacy and physicians' opioid prescribing patterns among different specialties [10,7]. However, current

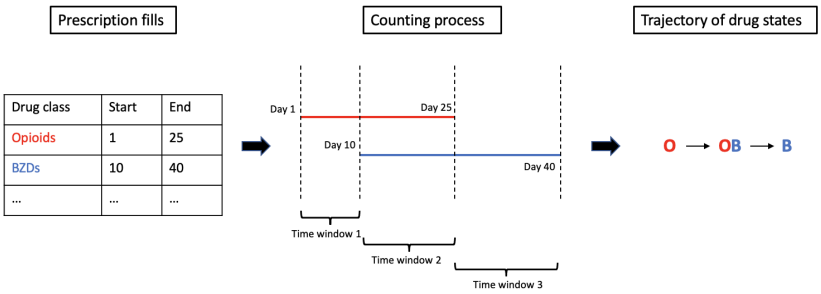


Fig. S2: Workflow of modeling patient prescription states.

Note: Panel (A) shows a made-up example of a patient's sequence of prescription fills with their corresponding drug class, the start and the end of the fill. Panel (B) shows the counting process to split the sequence of prescription fills into discrete exposure time intervals that reflect the initialization and the discontinuation of a prescription fill. The red line indicates the prescription fill length of the opioid in panel (A), and the blue line indicates the benzodiazepine fill length. Panel (C) shows the corresponding prescription state during each time interval in panel (B) and the transition between them, forming a trajectory of prescription states across time. "O" stands for filling an opioid, "B" stands for filling a benzodiazepine, and "OB" stands for filling an opioid and a benzodiazepine concurrently. BZDs = benzodiazepines

approaches for quantifying physicians' prescribing behaviors identify risky prescribing without accounting for the extent to which appropriate deprescribing occurs. Unlike prescribing, deprescribing often takes place in conversations during physician-patient encounters involving reviews of patients' medications [4] and triggers no insurance claim. This leads to challenges in identifying the physician or physicians responsible for deprescribing. An additional novel contribution of this paper is the development of heuristic algorithms for identifying likely instances of deprescribing in claims data. This quantitative framework can be generalized to comprehensively quantify any other prescribing behaviors by not only incorporating the number and classes of patients' drug exposures but also the changes in the drug exposures reflected by the time sequences of prescribing and deprescribing events.

The physician who initiated the prescription was attributed as the responsible physician. For example, the physician who prescribed an opioid to a patient who was already taking a benzodiazepine is the one responsible for the patient's prescription state transition from state B to state OB (or from state 3 to state 5 using numerical notation). Deprescribing often takes place in conversations during physician-patient encounters where the physician reviews and discusses medications with their patients [4]. A decision to terminate a prescription triggers no insurance claims. Therefore, there is no record of a responsible deprescribing physician in the Medicare data. However, when a patient no longer refills a long-term prescription following a physician-patient encounter, we assume that

a deprescribing conversation took place during the patient's most recent clinical visit in developing our deprescribing algorithm (Algorithms S1 and S2). Each prescription of each patient is initially treated as a target prescription that can potentially be discontinued. We first sought to exclude the prescriptions for acute conditions and thereby obtain a set of prescriptions for which intentional deprescribing by physicians could have occurred by requiring the target prescriptions to be longer than 30 days. The start date of a prescription fill is assumed to be the date when the patient visited the physician. The physician that the patient visited most recently before the end of the target prescription is selected as the candidate physician for deprescribing the target prescription. The patient has to discontinue filling the prescription within 30 days (inclusive) after visiting the candidate physician. Recall that any subsequent prescriptions are joined with the preceding prescription if they are likely to be refilled as described in Section 2.1. This preprocessing step protects against any temporary suspension of a prescription that a patient may refill later on, which can be a false positive deprescribing event by the identified physician. Finally, the patient prescription state transition associated with such termination was attributed to this identified physician with a contribution weight if multiple responsible physicians were identified. In this way, we obtained relatively balanced information regarding each physician's prescribing and deprescribing activities.

Algorithm S1 Pseudo code: Crediting physicians for the deprescribing

Input: A set of n prescriptions patient h receives and their prescribing physicians.
Output: The responsible physician for deprescribing the prescription of interest.

for $i = 1, 2, \ldots, n$ **do**
  if length of prescription i >= 30 days **then**
    $\Delta_t \leftarrow$ empty list ▷ difference between two dates
    for $j = 1, 2, \ldots, n$ and $j \neq i$ **do**
      $s_j \leftarrow$ start date of prescription j
      $e_i \leftarrow$ end date of prescription i
      if $e_i - s_j > 0$ and $e_i - s_j <= 30$ **then** ▷ prescription i discontinues within 30 days of encountering a physician
        Add $e_i - s_j$ to Δ_t
      end if
    end for
    Find the prescription p with minimum Δ_t and corresponding prescriber k ▷ most recent encountered physician
    responsible physician for deprescribing prescription $i \leftarrow$ physician k
    if multiple responsible physicians **then**
      contribution weight $\leftarrow$ 1 / number of contributors
    end if
    if no physician is held accountable after applying above criteria **then**
      responsible physician for deprescribing prescription $i \leftarrow$ a pseudo-physician
    end if
  else
    responsible physician for deprescribing prescription $i \leftarrow$ a pseudo-physician
    $i \leftarrow i + 1$
  end if
end for

Algorithm S2 Pseudo code: Determining the deprescribing state transition a physician is responsible for

Input: (1) The responsible physician k for deprescribing target prescription i for patient h; (2) a set of n prescriptions patient h receives.

Output: The deprescribing state transition that physician k is responsible for.

$D_b \leftarrow$ empty list ▷ List of drugs patient h is taking before discontinuation of prescription i
$D_a \leftarrow$ empty list ▷ List of drugs patient h is taking after discontinuation of prescription i
for $j = 1, 2, \ldots, n$ **do**
    $s_j \leftarrow$ start date of prescription j
    $e_j \leftarrow$ end date of prescription j
    $e_i \leftarrow$ end date of prescription i
    if $e_i - 0.1 > s_j$ and $e_i - 0.1 < e_j$ **then**
        Add j to D_b
    end if
    if $e_i + 0.1 > s_j$ and $e_i + 0.1 < e_j$ **then**
        Add j to D_a
    end if
end for
determine prescription state m based on D_b
determine prescription state n based on D_a
physician k is responsible for patient h's deprescribing state transition from m to n

2.4 Physician transition responsibility matrix

Given the eight different prescription states patients can take, and the sequence of their prescription states, an 8 by 8 transition responsibility matrix was established for every physician to capture their contribution to prescribing and deprescribing across all of their patients' prescription state transitions.

The physician transition responsibility count matrix ($PTRCM$) is constructed for each physician by summarizing across all the patients to whom they've prescribed and deprescribed. The rows in the matrix $PTRCM$ correspond to patients' prescription states in the preceding prescription exposure time intervals, while the columns correspond to patients' prescription states in the subsequent time intervals. For a given physician k, the ij'th cell in the $PTRCM$ is the total number of patient prescription state transitions from state i to state j to which the physician contributed. The PTRCM is mathematically defined as,

$$PTRCM^{(k)} = \begin{bmatrix} C_{1,1}^{(k)} & C_{1,2}^{(k)} & \dots & C_{1,j}^{(k)} & \dots & C_{1,8}^{(k)} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ C_{i,1}^{(k)} & C_{i,2}^{(k)} & \dots & C_{i,j}^{(k)} & \dots & C_{i,8}^{(k)} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ C_{8,1}^{(k)} & C_{8,2}^{(k)} & \dots & C_{8,j}^{(k)} & \dots & C_{8,8}^{(k)} \end{bmatrix}. \tag{4}$$

Let h denote a patient and t the prescription exposure time interval. Then D_{ht} denotes patient h's prescription state during time intervals t and P_{ht} denotes the corresponding responsible physician for patient h's prescription state transition between time interval t and $t+1$. In addition, i and j are prescription states. Let $\mathbf{1}(event) = 1$ if $event$ is true and 0 otherwise. Then cell (i, j) in the $PTRCM$ for physician k is given by,

$$C_{i,j}^{(k)} = \sum_h \sum_t \mathbf{1}(D_{ht} = i, D_{h(t+1)} = j)\mathbf{1}(P_{ht} = k). \tag{5}$$

To account for the scenario when multiple physicians are responsible for a prescription state transition from i to j, suppose the total number of responsible physicians for patient h at time t is N_{ht}. Then compute $C_{i,j}^{(k)}$ as

$$C_{i,j}^{(k)} = \sum_h \sum_t \mathbf{1}(D_{ht} = i, D_{h(t+1)} = j)\frac{\sum_r \mathbf{1}(P_{htr} = k)}{N_{ht}}. \tag{6}$$

2.5 Quantitative measures of physician prescribing behavior

Difference between prescribing and deprescribing Given the $PTRCM$, we compute different prescribing indexes corresponding to various aspects of decision-making during physicians' prescribing practice. First, we compute the relative difference between the number of prescribing and deprescribing events. Notice that the cells above the diagonal of the $PTRCM$ are the prescribing transitions while cells below the diagonal are contributions to deprescribing transitions. By construction, there was no transition between the same prescription states (these were subsumed by the continuation of the prior fill event as described in the Materials section). The resulting measure for physician k is given mathematically as,

$$I_{base}^{(k)} = \frac{\sum_j \sum_{i<j} C_{i,j}^{(k)} - \sum_i \sum_{i>j} C_{i,j}^{(k)}}{\sum_j \sum_{i<j} C_{i,j}^{(k)} + \sum_i \sum_{i>j} C_{i,j}^{(k)}}. \tag{7}$$

A second family of measures is instead based on the total number of drugs changed (not the number of transitions) to which the physician contributed, obtained by multiplying the count of transitions with the number of drug changes involved. For example, a transition from state B (or state 3) to state OBS (or state 8) involves a change of two drugs, as does the transition in the opposite

direction. Let $\mathbf{1}_{num}(\omega)$ quantify the number of drugs a patient is taking under the prescription state ω, given by,

$$\mathbf{1}_{num}(\omega) = \begin{cases} 0, & \omega = 1, \\ 1, & \omega \in \{2,3,4\}, \\ 2, & \omega \in \{5,6,7\}, \\ 3, & \omega = 8. \end{cases} \tag{8}$$

The resulting risky drug index of physician k is defined as,

$$I_{\alpha}^{(k)} = \frac{\sum_j \sum_{i<j} C_{i,j}^{(k)} |\mathbf{1}_{num}(i) - \mathbf{1}_{num}(j)|^{\alpha} - \sum_i \sum_{i>j} C_{i,j}^{(k)} |\mathbf{1}_{num}(i) - \mathbf{1}_{num}(j)|^{\alpha}}{\sum_j \sum_{i<j} C_{i,j}^{(k)} |\mathbf{1}_{num}(i) - \mathbf{1}_{num}(j)|^{\alpha} + \sum_i \sum_{i>j} C_{i,j}^{(k)} |\mathbf{1}_{num}(i) - \mathbf{1}_{num}(j)|^{\alpha}}. \tag{9}$$

Here α takes the value of 0 or 1 to define two measures distinguished by whether they account for the number of drug changes involved in the prescription state transition from state i to j or simply just count that some type of transition occurred. The index $I_{\alpha}^{(k)}$ is bounded between -1 (exclusive) and 1 (inclusive). The index I_{base} defined in Equation 7 can be obtained from Equation 9 when $\alpha = 0$.

Prescribing involving risky prescription state Our next measure quantifies the extent of physicians' involvement in the riskiest form of prescribing, state OBS (or state 8), given mathematically as,

$$I_{OBS}^{(k)} = \frac{\sum_j \sum_{i<8} C_{i,8}^{(k)}}{\sum_j \sum_{i<j} C_{i,j}^{(k)} + \sum_i \sum_{i>j} C_{i,j}^{(k)}} \tag{10}$$

where $C_{i,8}^{(k)}$ is the number of transitions physician k contributed to for which patients transitioned from state i to state 8 (state OBS). We also binarized the numerator to obtain a binary indicator of whether physician k has ever contributed to bringing a patient into state OBS, given by,

$$I_{everOBS}^{(k)} = \mathbf{1}(\sum_{i<8} C_{i,8}^{(k)} > 0). \tag{11}$$

Quantifying extent of prescribing and deprescribing We also developed several measures to quantify the intensities of physicians' prescribing and deprescribing. First, we quantified the percentage of patients' prescription state transitions a physician contributed to that involved two or more drug changes. For example, a transition from state $zero$ to state OBS, and a transition from state OBS to state B involves three and two drug changes respectively, the former by prescribing and the latter by deprescribing. For physician k, the proportion of physician's prescribing that contributes two or more drug changes

is,

$$I^{(k)}_{presc2mr} = \frac{\sum_j \sum_{i<j} C^{(k)}_{i,j} \mathbf{1}(|\mathbf{1}_{num}(i) - \mathbf{1}_{num}(j)| \geq 2)}{\sum_j \sum_{i<j} C^{(k)}_{i,j}}. \tag{12}$$

Likewise, the proportion of physician's deprescribing that involves changing the status of two or more drug classes is,

$$I^{(k)}_{deprescr2mr} = \begin{cases} \frac{\sum_i \sum_{i>j} C^{(k)}_{i,j} \mathbf{1}(|\mathbf{1}_{num}(i) - \mathbf{1}_{num}(j)| \geq 2)}{\sum_i \sum_{i>j} C^{(k)}_{i,j}} & , \sum_i \sum_{i>j} C^{(k)}_{i,j} \neq 0, \\ 0 & , \sum_i \sum_{i>j} C^{(k)}_{i,j} = 0. \end{cases} \tag{13}$$

The above measures may enter an ERGM as categorical or continuous node attributes to study the respective homophily effects in the network.

3 Supplemental results

We first present the descriptive results described in the Physician shared-patient networks subsection of the results in the main text. Table S2 includes the network statistics of each of the whole Ohio shared-patient physician network, the prescribing network, and the LCC of the prescribing network.

3.1 Prescribing and deprescribing measures

In the LCC of Ohio physician prescribing network, the distribution of indexes based on the difference between prescribing and deprescribing, i.e., I_0 and I_1, is skewed to the left (Table 1 in the main text). On average, among all the patient prescription state transitions a physician contributed to, around 0.9% of them involve bringing patients to the riskiest state OBS. Around 8.9% of physicians have ever contributed to patients' risky transition to state OBS. Among all the contributed transitions associated with prescribing, around 2.9% of them involved adding two or more drugs. Likewise, among all the contributed transitions associated with deprescribing, on average around 1.6% involved a reduction of two or more drugs.

In Figure S3, different prescribing indexes are displayed by physician specialty group for the LCC of the shared-patient prescribing physician network of Ohio in 2014; each physician was assigned to either the primary care physicians, medical specialist, and surgeon specialist group based on their lookup information in the National Plan and Provider Enumeration System (NPPES) [3]. In terms of overall prescribing and deprescribing reflected by I_0, there was no substantial difference across specialties, although surgeons and medical specialists appeared to have slightly higher prescribing index I_0 than other specialties. Other prescribing measures reflected physicians' prescribing behavior with more granularity. Primary care physicians and physicians of hospital-based services have a higher likelihood of bringing patients to state OBS, prescribing two or more drugs, and deprescribing two or more drugs, compared to medical specialists and surgeons.

Table S2: Network statistics of the largest connected component of the shared-patient physician prescribing network (specific to prescriptions of opioids, benzodiazepines, and sedative-hypnotics) for Ohio in 2014.

		Shared-patient physician network	
	Whole net	Prescribing net	LCC of prescribing net
Network statistics			
Number of nodes	35765	22655	17363
Number of ties	494462	265112	261816
Density	0.0008	0.0010	0.0010
Number of components	3002	2056	1
Size of LCC	27503	17363	17363
Degree (mean, IQR, SD)	27.7 (n/a, 44.0) (37.5)	23.4 (n/a, 38.0) (28.7)	30.2 (n/a, 45.0) (29.6)
Global clustering	0.168	0.171	0.171
Average path length	4.663	4.599	4.599
Prescribing statistics			
I_0 (mean, IQR)		0.871 (0.843, 1.0)	0.876 (0.875, 1.0)
I_{OBS} (mean, IQR)		0.009 (0.0, 0.0)	0.009 (0.0, 0.0)
$I_{everOBS}$ (# of 1, # of 0)		(1972, 20683)	(1412, 15951)
$I_{presc2mr}$ (mean, IQR)		0.030 (0.0, 0.0)	0.029 (0.0, 0.0)
$I_{depresc2mr}$ (mean, IQR)		0.017 (0.0, 0.0)	0.016 (0.0, 0.0)
Volume (mean, IQR, SD)	62.7 (n/a, 96.0) (86.3)	70.6 (n/a, 112.0) (87.8)	91.6 (18.0, 136.0) (90.3)

Note: The physician network is constructed based on the overlap of patient care at any point during 2014 between physician pairs treating patients residing in Ohio. The prescribing network is a subset of the whole network where its physicians have prescribed at least one opioid, benzodiazepine, or sedative-hypnotic during 2014. Volume is the number of Ohio Medicare fee-for-service beneficiaries a physician encountered throughout 2014. The values of n/a are suppressed to meet data suppressing rules designed to protect patient privacy by the Center for Medicare and Medicaid Services. LCC = largest connected component.

Figure S4 shows the homophily patterns in terms of prescribing and deprescribing in the ego network. An important observation is that central physicians with higher patient volume and higher node degrees have lower risky prescribing intensity than peripheral physicians. In addition, there are clusters of physicians with similar prescribing intensity and behavior.

3.2 ERGM Models and Triadic Homophily Analysis

In this subsection we first present the results for the 12-HRR stratified analyses referred to in the ERGMs for adjust homophily subsection of the results in the main paper (Table S3). The purpose of stratifying the ERGM models was to evaluate the level of heterogeneity in the level of homophily with respect to the risky prescribing index, I_{OBS}, and the index quantifying prescribing intensity, $I_{presc2mr}$, across regions.

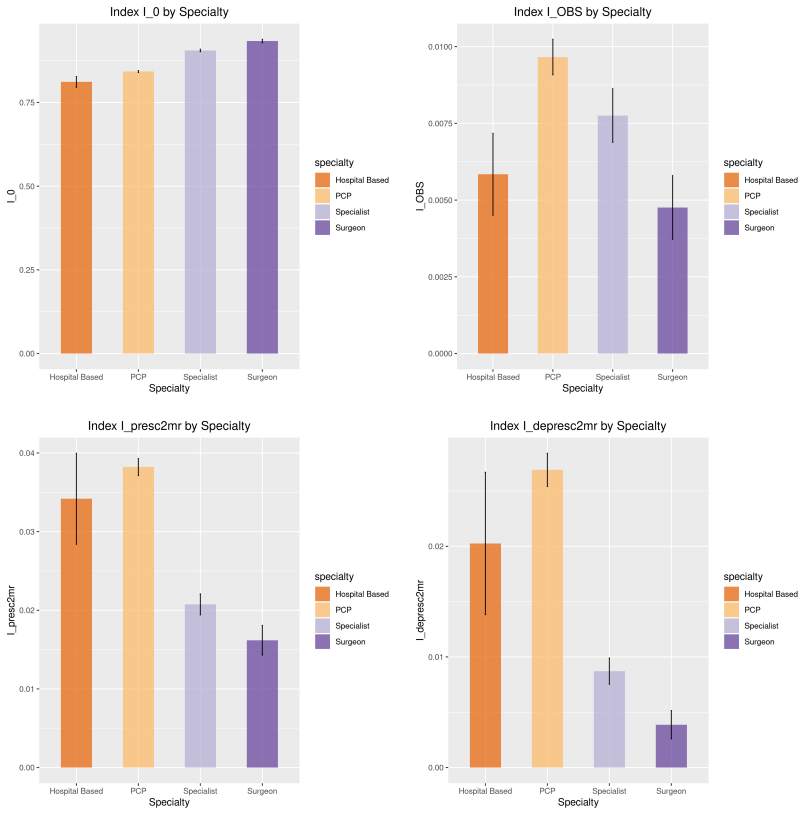


Fig. S3: Prescribing measures by specialty of physicians in the largest connected component of the shared-patient prescribing physician Ohio network in 2014. Prescribing measures by specialty of physicians in the largest connected component of the shared-patient prescribing physician Ohio network in 2014. Specialists are medical specialists other than surgeons. Hospital-based services include anesthesiology, radiology, and pathology. PCP denotes primary care physicians. I_{OBS} is the prescribing index based on a physician's contribution to bringing patients to prescription state OBS, $I_{presc2mr}$ is the prescribing index based on a physician's contribution to prescribing two or more drug types to patients, and $I_{deprescr2mr}$ is the deprescribing index based on a physician's contribution to deprescribing two or more drug types for patients. Error bars show the standard errors of the respective measures.

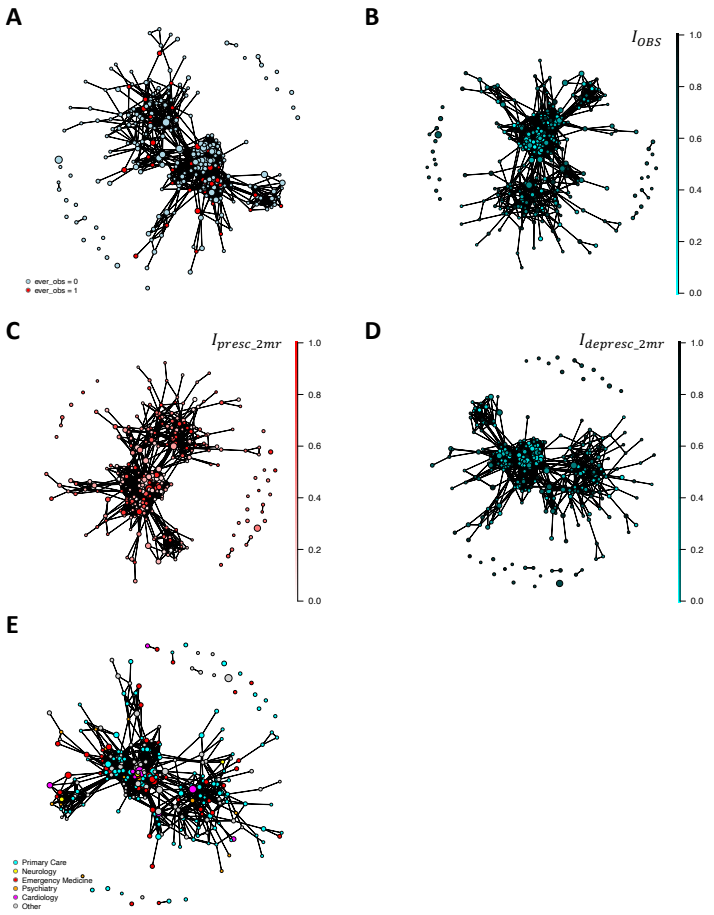


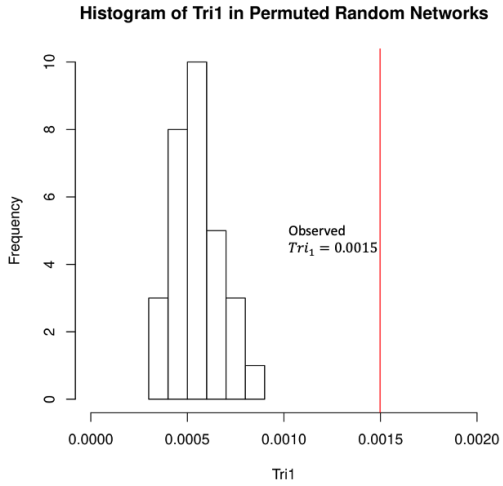
Fig. S4: Egocentric network of the physician with maximum node degree (N = 276). The egocentric network of the physician with the highest node degree in the LCC of the prescribing network. The ego physician was removed from the plot for the clarity of presentation. The ties shown in the plots are among the peers of the ego physician. The nodes are sized by physician annual volume (the number of distinct patients treated throughout the year). The colors of nodes correspond to their prescribing behavior or specialties. A) This shows the connections among physicians distinguished by whether they have ever contributed to bringing patients to the riskiest prescription state (the OBS state). B) This shows the connections among physicians where the node color represents the value of I_{OBS}, the proportion of times they bring their patients to prescription state OBS. C) This shows the connections among physicians where the node color represents I_{presc_2mr}, the proportion of prescribing events when two or more drugs are prescribed at once to the patients. D) This shows the connections among physicians where the node color represents $I_{depresc_2mr}$, the proportion of contributions to deprescribing events at which two or more drugs are deprescribed at once to the patients. E) This shows the connections among physicians colored by their specialties.

Table S3: ERGM adjusted homophily effects in HRR shared-patient sub-networks in 2014.

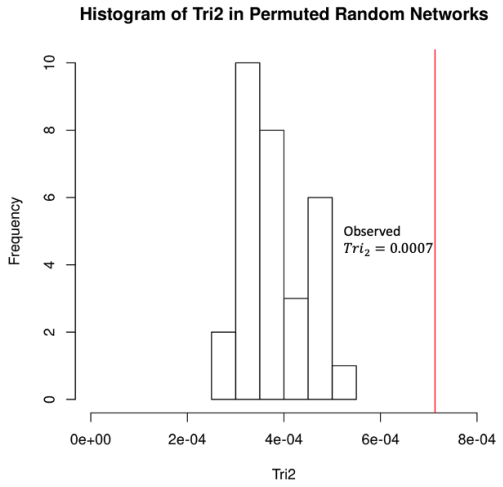
			Homophily effects of indexes							
	Descriptive Stats		absdiff(I_{OBS})		absdiff($I_{presc2mr}$)		absdiff($I_{depresc2mr}$)		$I_{everOBS}$	
HRR	N	Density	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
180	129	0.116	-3.998	3.662	0.004	1.623	1.052	1.898	0.026	0.133
357	193	0.106	0.810	2.707	-1.541*	0.714	0.467	0.775	-0.268	0.149
331	256	0.128	-5.156	2.675	-1.792*	0.697	-0.325	0.773	0.047	0.117
332	415	0.048	-1.765*	0.745	-0.863***	0.250	0.299	2.410	0.056	0.079
335	550	0.060	-1.887*	0.920	-0.881**	0.305	-0.545	0.416	0.063	0.070
326	648	0.050	-1.281	0.795	-1.066***	0.306	44.210	280.321	-0.080	0.056
325	750	0.030	-0.469	0.814	-0.917**	0.279	0.496	0.699	0.0002	0.089
334	1039	0.029	-0.532	0.416	-1.190***	0.226	-0.301	0.209	0.018	0.045
330	1164	0.024	-1.205**	0.419	-0.339	0.196	-0.071	0.319	0.0002	0.037
327	1760	0.015	-1.711**	0.584	-0.783***	0.179	-0.362*	0.171	0.120**	0.044
328	2623	0.010	-1.603***	0.370	-0.897***	0.142	-0.193	0.227	0.060	0.034
329	3101	0.008	-1.181***	0.234	-0.754***	0.109	-0.327*	0.157	-0.018	0.025

Note: The HRR sub-networks were partitioned from the largest connected component of the Ohio 2014 shared-patient physician prescribing network. The HRR sub-networks were not restricted to their respective largest connected components and so may not be fully connected. Significance levels: $***p < 0.001$; $**p < 0.01$; $*p < 0.05$.

The next results output (Figure S5) is referred to in the Triadic-level hyper homophily subsection in the results in the main text. The purpose of this analysis was to 1) evaluate whether the occurrence of three physicians sharing patients among themselves is associated with greater likelihood of them all being involved in risky prescribing than by chance; and 2) to evaluate whether the occurrence of two physicians sharing patients with a third common physician when all three of them have been involved in risky prescribing is associated with greater likelihood of these two physicians sharing patients between them than by chance. The statistically significant test results demonstrate that homophily and triadic level clustering interact in the network and thus that homophily is not just a dyadic-level phenomenon in relation to risky prescribing.



(a) Triadic homophily statistic Tri_1



(b) Triadic homophily statistic Tri_2

Fig. S5: Histogram of triadic homophily network statistics generated by the triadic homophily non-parametric test. The triadic homophily statistic Tri_1 is the proportion of closed triangles with the $I_{everOBS}$ node attribute (whether a physician has ever contributed to bringing patients to the riskiest prescription state OBS) in the network. The triadic homophily statistic Tri_2 is the proportion of open two-paths with all nodes having the same attribute that are closed in the network. Panel (a) is the histogram of Tri_1 and panel (b) is the histogram of Tri_2 calculated from 30 networks with randomly shuffled node attributes under the null hypothesis of no homophily with respect to the given prescribing index. The red vertical lines denote the values in the observed network.

References

1. C An, A J O'Malley, D N Rockmore, and C D Stock, *Analysis of the u.s. patient referral network*, Stat Med. **37** (2018), no. 5, 847–866.
2. Chuankai An, A James O'Malley, and Daniel N Rockmore, *Referral paths in the us physician network*, Applied network science **3** (2018), no. 1, 1–24.
3. Centers for Medicare & Medicaid Services and Centers for Medicare & Medicaid Services, *National plan and provider enumeration system*, https://download.cms.gov/nppes/NPI_Files.html.
4. Barbara Farrell and Dee Mangin, *Deprescribing is an essential part of good prescribing*, American family physician **99** (2019), no. 1, 7–9.
5. Mark S Handcock, David R Hunter, Carter T Butts, Steven M Goodreau, and Martina Morris, *statnet: Software tools for the representation, visualization, analysis and simulation of network data*, Journal of statistical software **24** (2008), no. 1, 1548.
6. David R Hunter, Mark S Handcock, Carter T Butts, Steven M Goodreau, and Martina Morris, *ergm: A package to fit, simulate and diagnose exponential-family models for networks*, Journal of statistical software **24** (2008), no. 3, nihpa54860.
7. Benjamin Levy, Leonard Paulozzi, Karin A Mack, and Christopher M Jones, *Trends in opioid analgesic–prescribing rates by specialty, us, 2007–2012*, American journal of preventive medicine **49** (2015), no. 3, 409–413.
8. Martina Morris, Mark S Handcock, and David R Hunter, *Specification of exponential-family random graph models: terms and computational aspects*, Journal of statistical software **24** (2008), no. 4, 1548.
9. A James O'Malley, Xin Ran, Chuankai An, and Daniel N Rockmore, *Optimal physician shared-patient networks and the diffusion of medical technologies*, Journal of Data Science (2022).
10. Katie J Quinn and Nigam H Shah, *A dataset quantifying polypharmacy in the united states*, Scientific data **4** (2017), no. 1, 1–11.