

Identifying Differences in Physician Communication Styles with a Log-Linear Transition Component Model

Byron C. Wallace and Issa J. Dahabreh and Thomas A. Trikalinos

Center for Evidence-Based Medicine, Brown University
{byron.wallace, issa.dahabreh, thomas.trikalinos}@brown.edu

M. Barton Laws and Ira Wilson

Health Services, Policy & Practice, Brown University
{michael_barton_laws, ira_wilson}@brown.edu

Eugene Charniak

Computer Science, Brown University
ec@cs.brown.edu

Abstract

We consider the task of grouping doctors with respect to communication patterns exhibited in outpatient visits. We propose a novel approach toward this end in which we model *speech act* transitions in conversations via a log-linear model incorporating physician specific components. We train this model over transcripts of outpatient visits annotated with speech act codes and then cluster physicians in (a transformation of) this parameter space. We find significant correlations between the induced groupings and patient survey response data comprising ratings of physician communication. Furthermore, the novel sequential component model we leverage to induce this clustering allows us to explore differences across these groups. This work demonstrates how statistical AI might be used to better understand (and ultimately improve) physician communication.

Introduction and Motivation

Physician-patient communication is a critical component of health-care (Ong et al. 1995). There is evidence that the relationship between the physician and the patient, and specifically the degree of patient-centeredness in communication, affects patient enablement, satisfaction, and burden of symptoms (Little et al. 2001). Several studies have reported an association between metrics of physician-patient communication quality and health outcomes (Epstein and Street 2007; Kaplan, Greenfield, and Ware Jr 1989; Oates, Weston, and Jordan 2000). Furthermore, a systematic review of studies investigating physician-patient communication found several verbal behaviors to be significantly associated with health outcomes (Beck, Daughtridge, and Sloane 2002).

To analyze, evaluate and improve physician communication skills, investigators have proposed annotation schemas that capture salient properties of physician-patient interactions (Roter and Larson 2002; Laws et al. 2011a). Here we use the Generalized Medical Interaction Analysis System (GMIAS) (Laws 2013), which is explicitly based in *speech act* theory (Searle 1969; Habermas 1984; Austin 1955).

Roughly, speech acts capture the social acts embodied in utterances, such as promising or asking a question.

Insofar as they capture salient communication patterns in consultations, such annotations may permit the development of evidence based interventions to improve communication quality. Ultimately, we would like to use quantitative methods to better understand how physicians differ in interaction styles and what makes for effective clinical communication. Toward this end, *we propose a novel approach to grouping physicians with respect their communication styles*, specifically as a function of how they use speech acts. The hope is that discovering latent communication types will allow us to link attributes of communication to clinically relevant outcomes; this may in turn inform interventions aimed at improving physician communication. The induced groupings might also be used prospectively, e.g. by assigning new (previously unobserved) physicians to existing groups and then targeting their training accordingly.

Our approach relies on a novel additive sequential component model that extends our recent work in this vein (Wallace et al. 2013) to incorporate physician specific components into the transition probability model. These capture, e.g., doctor tendencies to deviate from baseline dialogue patterns. We then represent physicians by estimates of these parameters, and cluster them in a reduced-dimensionality transformation of this space. Ideally, induced groupings of physicians would correlate with measures of care, including patient satisfaction regarding physician communication. To assess if this is the case, we perform a hierarchical regression analysis to test for correlation between physician cluster assignments and survey data comprising patient assessments of physician communication.

The specific contributions of this work as follows.

- So far as we are aware, this is the first work to incorporate interlocutor-level terms in a sequential (Markovian) model of conversation.
- We propose to use (estimates) of these terms to subsequently induce groupings of speakers (physicians). We show that this novel strategy discovers clusters of physicians that significantly correlate with patient ratings.
- We present clinically interesting model output regarding characteristics of physician-patient interactions, thereby

This component-based approach is attractive because it allows us to isolate different factors in speech act usage across physicians. For example, the λ_ρ^{dr} terms capture physician specific differences in speech act usage with respect to different turn-taking patterns, while the $\sigma_{s_{t-1}}^{dr}$ terms capture differences in the relative frequencies with which speech act transitions occur in visits involving different physicians.

Such fine-grained parameters are crucial to interpreting model output. It is not enough to cluster physicians via some black-box mechanism, even if such clusters are found to correlate with outcome metrics: a useful model in this case must expose differences in communication styles between clusters. We will return to this point in the Results section, highlighting specific parameters of interest between physician groups.

Clustering Physicians

To cluster physicians, we construct a matrix of row-vectors representing physicians. Each row corresponds to a doctor and is composed of their corresponding physician specific parameter estimates. Hence we construct a row corresponding to a specific doctor (dr) as: $\{\pi^{dr}, \sigma_s^{dr}, \lambda_\rho^{dr}\}$. However we drop ρ terms corresponding to $p \rightarrow p$ turns, as (presumably) they don't directly reflect on physician communication attributes. We stack 41 of these (one per physician) to create a design matrix. We reduce the dimensionality of this matrix via PCA – keeping the first two dimensions – and then project the vector representation of each physician into this space.

To induce groupings in this space, we simply use k -means (Hartigan and Wong 1979). Because we have only 41 doctors in total, we set $k=2$. Very crudely, we might thus hope to find one group of ‘good’ communicators and another group of comparatively ‘poor’ communicators. This is obviously an over-simplification, but seems a reasonable aim given limited data. We next present results from this analysis that suggests that the discovered clusters indeed correlate with patient response data. We then show how group distributions of the physician specific parameters comprising our model can provide insights into differences between these groups.

Results

Our aim here is to evaluate the clustering induced using the approach outlined above. The relevant question is: do measurements of individual physician communication quality correlate with cluster assignments? To address this, we leverage patient feedback provided in response to the questions shown in Table 2. (Recall that this feedback was provided following each visit.)

Specifically, to assess the association between physician cluster assignments and patient feedback regarding communication, we used a three-level linear mixed effects model (Rabe-Hesketh and Skrondal 2008) to account for the nesting of patients within doctors and the multiple questions answered by each patient. This random intercept represents the combined effects of omitted doctor characteristics and heterogeneity that is unexplained by the clustering. We assume that ‘cluster’ effects vary across the three sets of questions

How is the provider who takes care of your HIV at ...

Overall

- Q1 *explaining the results of tests in a way that you understand?*
- Q2 *giving you facts about the benefits and risks of treatment?*
- Q3 *telling you what to do if certain problems or symptoms occur?*
- Q4 *demonstrating caring, compassion, and understanding?*
- Q5 *understanding your health worries and concerns?*

HIV-specific

- Q6 *talking with you about your sex life?*
- Q7 *asking you about stresses in your life that may affect your health?*
- Q8 *asking about problems with alcohol?*
- Q9 *asking about problems with street drugs like heroin and cocaine?*

Adherence

- Q10 *giving you information about the right way to take your antiretroviral medicines?*
- Q11 *understanding the problems you have taking your antiretroviral medicines?*
- Q12 *helping you solve problems you have taking your antiretroviral medicines the right way?*

Table 2: Survey questions regarding provider (physician) communication that patients were asked following visits. Responses were provided on an ordinal scale from 1 (*excellent*) to 5 (*poor*).

shown in Table 2. That is, we assume physician groupings will exhibit different degrees of correlation with questions around communication addressing *overall*, *HIV-specific* and *adherence* issues. These questions were structured in this way by the domain experts who formulated them.

We denote the question number (1-12) by k , the physician by j and the visit (equivalently, the patient) by i (recall that we have more than one visit for most physicians). Further, we define the function $g(k)$ which maps questions into the three groupings shown in Table 2. For example, $g(10) = 3$, as question 10 belongs to the third (*Adherence*) set of questions. Finally, we introduce the cluster indicator $c_0(j)$, which is 1 if and only if physician j has been assigned to cluster 0 (obviously the cluster ‘numbering’ is arbitrary). Putting these terms together, we model patient response to question k after visit i involving physician j , y_{ij}^k , as follows

$$y_{ij}^k = \beta_j^k + \beta_{c_0}^{g(k)} c_0(j) \quad (4)$$

Where

$$\beta_j^k \sim \text{Normal}(\mu_k, \sigma_{jk}) \quad (5)$$

We fit this model via gradient-based optimization to estimate coefficients and associated standard errors. We are primarily interested in the $\beta_{c_0}^{g(k)}$ terms; if these are non-zero, it implies that the clusters discovered from the data correlate with patient assessments of physician communication.

We show the induced clustering in Figure 3 (a), and in Figure 3 (b) we show the coefficient estimates and associated 95% confidence intervals for patient responses to questions concerning physicians in the two induced clusters. Rows correspond to question groupings (Table 2). We report p -values for each group of questions (recall that we assume a different correlation for each, as per Equation 4) with re-

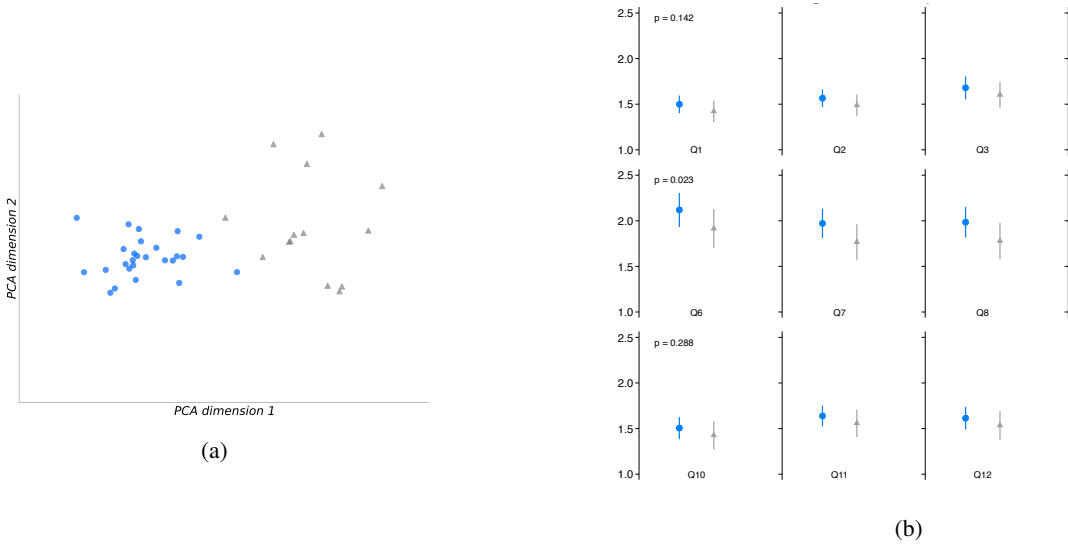


Figure 3: Discovered groups of physicians (a) and estimated differences between them (b), with respect to patient responses to questions surrounding physician communication. For the latter we show estimates and 95% confidence intervals for coefficients corresponding to the two induced clusters (Equation 4). Rows correspond to the three sets of questions in Table 2. Due to space constraints we do not show Q5 here (it is very similar to Q4). We report corresponding p -values for each of set of questions (first plot in each row). Physicians in the blue (circles) cluster receive consistently higher scores (worse reviews) than those in the gray (triangles) cluster, across all questions. We have omitted a single member of the latter group (far to the right of all plotted points) to ease visualization. This difference is significant ($p < .05$) for questions regarding communication around HIV-specific issues, and it is suggestive for the two other sets of questions.

Overall		
Q	Blue (●) cluster mean	Gray (▲) cluster mean
Q1	1.50 (1.40, 1.60)	1.42 (1.30, 1.54)
Q2	1.57 (1.47, 1.66)	1.49 (1.37, 1.61)
Q3	1.68 (1.55, 1.81)	1.60 (1.46, 1.74)
Q4	1.48 (1.38, 1.59)	1.41 (1.28, 1.53)
Q5	1.56 (1.44, 1.68)	1.49 (1.35, 1.62)
HIV-specific		
Q	Blue (●) cluster mean	Gray (▲) cluster mean
Q6	2.12 (1.93, 2.30)	1.91 (1.70, 2.13)
Q7	1.97 (1.81, 2.13)	1.77 (1.57, 1.96)
Q8	1.98 (1.82, 2.15)	1.78 (1.58, 1.98)
Q9	1.93 (1.76, 2.11)	1.73 (1.52, 1.94)
Adherence		
Q	Blue (●) cluster mean	Gray (▲) cluster mean
Q10	1.50 (1.38, 1.63)	1.43 (1.27, 1.58)
Q11	1.64 (1.52, 1.75)	1.56 (1.41, 1.71)
Q12	1.61 (1.49, 1.74)	1.53 (1.38, 1.69)

Table 3: Means and confidence intervals of patient responses to questions involving physicians comprising the respective clusters. Physicians in the gray cluster (▲) receive consistently better (lower) marks from patients.

spect to the null hypothesis that mean patient responses are the same between clusters for these questions.

A clear pattern emerges: physicians in the blue (●) group consistently receive less favorable reviews than those in the gray (▲) group (recall that lower is better here). This difference is particularly pronounced for questions around HIV-specific communication ($p < .05$). For the other two sets

of questions, the difference is consistent and the p -value is suggestive, though not significant. Thus the discovered clusters correlate with patient feedback. But what is different between these groups that might account for this difference?

To explore this, we can consider specific parameter value distributions over the physicians comprising the respective clusterings. One phenomenon that might result in patient dissatisfaction is that of “advising without permission” (Gaume et al. 2009; Moyers, Miller, and Hendrickson 2005), in which physicians offer unsolicited advice. Relatedly, physicians may promise to take action (e.g., run a test) without first discussing it with the patient. In our model, such tendencies would manifest in the $\lambda_{d \rightarrow d, commissive}^{dr}$ and $\lambda_{d \rightarrow d, directive}^{dr}$ terms, which express the relative likelihood of the corresponding physician issuing commissives and directives (respectively), within a single ‘turn’ (i.e., not directly in response to a patient utterance). In Figure 4 we show histograms of these parameter values for the physicians comprising the two clusters, with the same color coding as in Figure 3. We have fit independent normals to the data from the respective clusters, denoted by the solid and dotted lines for the gray (▲) and blue (●) clusters, respectively.

The parameters are indeed different between physician groups, as we would expect under the hypothesis that advising without patient permission negatively correlates with patient satisfaction around communication. Indeed, for both $d \rightarrow d, commissive$ and the $d \rightarrow d, directive$, a t-test between group values suggests a difference with $p < .001$.² And the difference is in the expected direction: physicians

²We recognize the assumption of normality is not necessarily

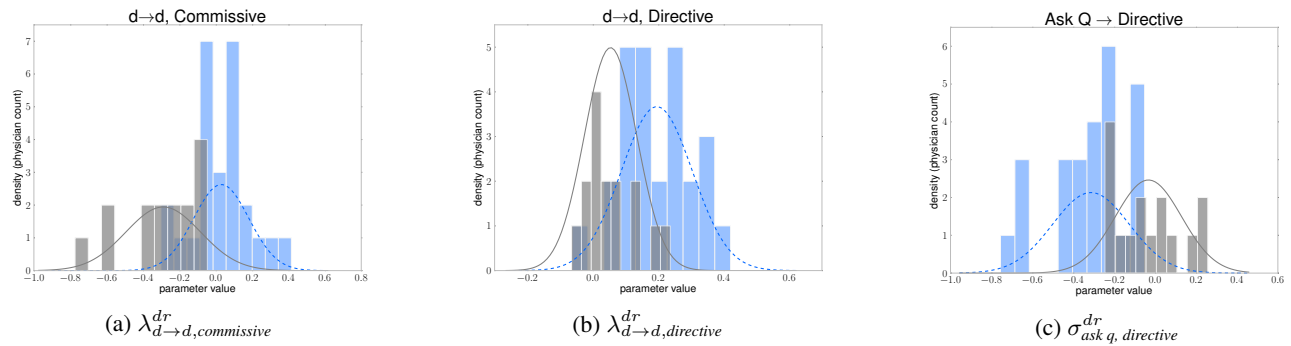


Figure 4: Histograms (and fitted normals) for model parameter estimates across the physicians comprising the two induced clusters of physicians. See text for discussion.

who seem to advise (or make decisions) without patient input receive poorer marks from patients. Interestingly, however, the trend is reversed when we consider the parameter corresponding to the transition from *Ask Q* to *Directive* (i.e., $\sigma_{ask\ q, directive}^{dr}$; Figure 4c). That is, directives – most of which are issued by physicians – follow questions more often in interactions involving doctors that belong to the favored gray (solid line) cluster. We might speculate that patients appreciate instruction (or action) when consulted, but not when unsolicited.

Related Work

To our knowledge, this is the first attempt to use natural language processing to cluster doctors with respect to how they communicate with their patients. We also believe this is the first work on clustering interlocutors with respect to *speech act* usage in conversation. However in interesting recent related work Howes et al. (2013) found that patterns of conversational topics seem to correlate with patient satisfaction and that features from dialogue transcripts may predict treatment adherence.

More generally, there has been related work on clustering speakers with respect to observed patterns in conversations. Grothendieck et al. (2009) clustered speakers by their turn-taking behavior and found that clusters correlate with population traits. Mairesse et al. (2006; 2007) argue that personality type affects language use, and show that these types can be recognized (see also (Oberlander and Gill 2006)). Other efforts have focussed on clustering and/or segmenting speakers in multi-party conversations based on auditory signals (e.g., (Jin and Schultz 2004)).

There has also been some work on modeling (and predicting) speech acts in conversations (Stolcke et al. 2000). Our own prior work has specifically considered doing so for physician-patient interactions (Wallace et al. 2013; 2014). And others have explored visual structures of patient-provider interactions to qualitatively assess communication (Cretchley et al. 2010; Angus et al. 2012). However, these works did not look to subsequently identify *groups* of physicians with respect to observed conversational patterns.

warranted here, but this provides a crude measure of difference.

Discussion, Future Directions and Limitations

We have introduced novel model of speech act use in physician-patient visits that factorizes the probability of a given speech act into shared and physician specific terms. We used estimates of this model’s parameters to represent physicians, and then clustered these representations. We found that the induced clusters significantly correlated with patient survey response data regarding physician communication. And we found (tentative) quantitative evidence for the hypothesis that the practice of ‘advising without permission’ correlates negatively with patient ratings of physician communication quality. In future work we hope to explore this and related phenomena more carefully. Eventually we hope such investigations lead to evidence-based interventions targeted at improving patient care.

We acknowledge that the model we have presented is limited in that it still makes the naïve first-order Markov assumption regarding dialogue, when in reality conversation exhibits longer-range and non-linear contingencies (e.g., conversational *threads*). However we believe even this simple model captures the most salient properties of conversation (and indeed, that we find correlations with patient marks would seem to support this). Moreover, while we have incorporated physician specific terms for speaker dyads (the ρ terms) and speech acts, we do not have terms that capture the interaction between physician, speaker *and* transition, as this would result in an unwieldy number of parameters. This means that even though we exclude the $p \rightarrow p$ speaker terms when we induce clusters, some of the signal we pick up on may still be due to differences in *patient* populations rather than to the physician.

References

- Angus, D.; Watson, B.; Smith, A.; Gallois, C.; and Wiles, J. 2012. Visualising conversation structure across time: Insights into effective doctor-patient consultations. *PLoS one* 7(6).
- Austin, J. 1955. *How to do Things with Words*, volume 88. Harvard University Press.
- Beck, R.; Daughtridge, R.; and Sloane, P. 2002. Physician-patient communication in the primary care office: a systematic review. *The Journal of the American Board of Family Practice* 15(1):25–38.

- Cretchley, J.; Gallois, C.; Chenery, H.; and Smith, A. 2010. Conversations between carers and people with schizophrenia: a qualitative analysis using leximancer. *Qualitative Health Research* 20(12):1611–1628.
- Eisenstein, J.; Ahmed, A.; and Xing, E. 2011. Sparse additive generative models of text. In *Proceedings of ICML*, 1041–1048.
- Epstein, R., and Street, R. 2007. *Patient-centered communication in cancer care: promoting healing and reducing suffering*. National Cancer Institute, US Department of Health and Human Services, National Institutes of Health.
- Gaume, J.; Gmel, G.; Faouzi, M.; and Daepfen, J.-B. 2009. Counselor skill influences outcomes of brief motivational interventions. *Journal of substance abuse treatment* 37(2):151–159.
- Grothendieck, J.; Gorin, A.; and Borges, N. 2009. Social correlates of turn-taking behavior. In *Proceedings of ICASSP*, 4745–4748. IEEE.
- Habermas, J. 1984. *The Theory of Communicative Action*, volume 1. Beacon Press.
- Hartigan, J. A., and Wong, M. A. 1979. Algorithm as 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 28(1):100–108.
- Howes, C.; Purver, M.; and McCabe, R. 2013. Investigating topic modelling for therapy dialogue analysis. In *Proceedings of IWCS-CSCT*, 7–16. Potsdam, Germany: Association for Computational Linguistics.
- Jin, Q., and Schultz, T. 2004. Speaker segmentation and clustering in meetings. In *Proc. of ICASSP Meeting Recognition Workshop*.
- Kaplan, S.; Greenfield, S.; and Ware Jr, J. 1989. Assessing the effects of physician-patient interactions on the outcomes of chronic disease. *Medical Care* 1:110–127.
- Laws, M.; Bradshaw, Y.; Safren, S.; Beach, M.; Lee, Y.; Rogers, W.; and Wilson, I. 2011a. Discussion of sexual risk behavior in HIV care is infrequent and appears ineffectual: a mixed methods study. *AIDS and Behavior* 15(4):812–822.
- Laws, M.; Epstein, L.; Lee, Y.; Rogers, W.; Beach, M.; and Wilson, I. 2011b. The association of visit length and measures of patient-centered communication in HIV care: a mixed methods study. *Patient Education and Counseling* 85(3):e183–e188.
- Laws, M.; Beach, M.; Lee, Y.; Rogers, W.; Saha, S.; Korhuis, P.; Sharp, V.; and Wilson, I. 2012. Provider-patient adherence dialogue in HIV care: results of a multisite study. *AIDS and Behavior* 1–12.
- Laws, M. 2013. GMIAS coding manual and information (<https://sites.google.com/a/brown.edu/m-barton-laws/home/gmias>).
- Little, P.; Everitt, H.; Williamson, I.; Warner, G.; Moore, M.; Gould, C.; Ferrier, K.; and Payne, S. 2001. Observational study of effect of patient centredness and positive approach on outcomes of general practice consultations. *BMJ* 323(7318):908–911.
- Mairesse, F., and Walker, M. 2006. Automatic recognition of personality in conversation. In *Proceedings of the HLT-NAACL*, 85–88. Association for Computational Linguistics.
- Mairesse, F.; Walker, M. A.; Mehl, M. R.; and Moore, R. K. 2007. Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of Artificial Intelligence Research (JAIR)* 30:457–500.
- Moyers, T. B.; Miller, W. R.; and Hendrickson, S. M. 2005. How does motivational interviewing work? therapist interpersonal skill predicts client involvement within motivational interviewing sessions. *Journal of Consulting and Clinical Psychology* 73(4):590.
- Oates, J.; Weston, W.; and Jordan, J. 2000. The impact of patient-centered care on outcomes. *Family Practice* 49:796–804.
- Oberlander, J., and Gill, A. J. 2006. Language with character: A stratified corpus comparison of individual differences in e-mail communication. *Discourse Processes* 42(3):239–270.
- Ong, L.; de Haes, J.; Hoos, A.; and Lammes, F. 1995. Doctor-patient communication: a review of the literature. *Social science & medicine* 40(7):903–918.
- Rabe-Hesketh, S., and Skrondal, A. 2008. *Multilevel and longitudinal modelling using Stata*. STATA press.
- Roter, D., and Larson, S. 2002. The roter interaction analysis system (riass): utility and flexibility for analysis of medical interactions. *Patient education and counseling* 46(4):243–251.
- Searle, J. 1969. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press.
- Stolcke, A.; Ries, K.; Coccaro, N.; Shriberg, E.; Bates, R.; Jurafsky, D.; Taylor, P.; Martin, R.; Ess-Dykema, C.; and Meteer, M. 2000. Dialogue act modeling for automatic tagging and recognition of conversational speech. *Computational Linguistics* 26(3):339–373.
- Wallace, B. C.; Trikalinos, T. A.; Laws, M. B.; Wilson, I. B.; and Charniak, E. 2013. A generative joint, additive, sequential model of topics and speech acts in patient-doctor communication. In *Proceedings of EMNLP*, 1765–1775.
- Wallace, B. C.; Laws, M. B.; Small, K.; Wilson, I. B.; and Trikalinos, T. A. 2014. Automatically annotating topics in transcripts of patient-provider interactions via machine learning. *Medical Decision Making* 34(4):503–512.
- Wilson, I.; Laws, M.; Safren, S.; Lee, Y.; Lu, M.; Coady, W.; Skolnik, P.; and Rogers, W. 2010. Provider-focused intervention increases adherence-related dialogue, but does not improve antiretroviral therapy adherence in persons with HIV. *Journal of acquired immune deficiency syndromes* 53(3):338.