Published in final edited form as:

Psychol Assess. 2015 September ; 27(3): 975–984. doi:10.1037/pas0000094.

A Latent Transition Analysis for the Assessment of Structured Diagnostic Interviews

Pamela Scorza,

Department of Global Health and Population, Harvard School of Public Health

Katherine E. Masyn,

Harvard Graduate School of Education

Joshua A. Salomon, and

Department of Global Health and Population, Harvard School of Public Health

Theresa S. Betancourt

Department of Global Health and Population, Harvard School of Public Health

Abstract

Structured diagnostic interviews administered by lay people are commonly used to assess psychiatric disorders, including depression, in large epidemiologic studies. Many interviews utilize “gate” questions, such as screening questions, that allow interviewers to skip entire survey sections for a particular respondent, saving time and reducing respondent fatigue. However, most depression estimates based on these response data are predicated on the assumption that the gate questions function without measurement error or bias. The tenability of this assumption is questionable, and its violation could compromise the reliability and validity of those estimates of depression. In this study, we used a novel application of latent transition analysis to cross-sectional data, accounting for measurement error in different response pathways through the depression module in the World Mental Health Composite International Diagnostic Interview. The analysis included data from 19,734 participants 18 years of age in the Comprehensive Psychiatric Epidemiologic Surveys. The latent transition analysis, allowing for measurement error in screening questions and exclusion criteria, produced a higher estimate of the lifetime probability of experiencing depression than did the algorithm based on the Diagnostic and Statistical Manual for Mental Disorders, 4th Edition, Text Revision. This illustration of latent transition analysis applied to item-level data from a complex structured diagnostic tool with gate questions demonstrates the potential utility of an analytic approach that does not automatically assume gate questions function without measurement error. This model could also be used to probe for evidence of measurement bias in the form of differential item function when using structured diagnostic tools in different cultures and languages.

Keywords

depression; diagnostic interview; screening question; skip pattern; latent class transition

Depression prevalence at a population level is assessed primarily through structured diagnostic interviews that do not require clinician judgment and can be administered by lay interviewers. Although these interviews are based on criteria for depression from the *Diagnostic and Statistical Manual for Mental Disorders*, 4th Edition, Text Revision (*DSM-IV-TR*) and the International Classification of Disease, tenth edition (ICD-10), the translation of these criteria to structured diagnostic interviews may impact both the reliability and validity of these interviews as depression assessment instruments. In particular, gate questions could potentially compromise the quality of the assessment-based estimates. *Gate questions* are elements that control the respondent pathways through a survey based on responses given to those questions. Screening questions are a ubiquitous type of gate question used in structured diagnostic interviews that determine whether a respondent will “enter” or “skip” an entire section or module of the survey related to a particular diagnostic outcome. Within a given section, there may be further gate questions that “eject” a respondent from the remainder of that section or module. These gate questions save considerable time and guard against respondent fatigue, but they yield data sets with vast amounts of “missing” responses reflecting the complex skip patterns. Further, the diagnostic classifications made using interview responses are usually based on algorithms that must assume the gate questions are functioning without measurement error or bias.

The reliability of the overall depression estimates and the validity of interpreting those estimates as indicative of actual depression have been analyzed largely by intertemporal or interrater agreement, and by convergent validity with other accepted tools, including comparison with clinician diagnoses (W. W. Eaton, Neufeld, Chen, & Cai, 2000; Kessler et al., 2004; Robins, Helzer, Croughan, & Ratcliff, 1981). More recently, researchers have begun to use latent variable approaches, including factor analysis, item response theory, and latent mixture modeling, in order to obtain more accurate estimates of diagnostic scores and classifications. These approaches utilize item-level data and allow for imperfect item functioning (N. R. Eaton, Krueger, & Oltmanns, 2011; Faraone & Tsuang, 1994; Fossati et al., 2005; Garrett, Eaton, & Zeger, 2002; Krueger, 1999; Krueger, Caspi, Moffitt, & Silva, 1998; Krueger, Chentsova-Dutton, Markon, Goldberg, & Ormel, 2003; Krueger et al., 2011; Krueger & Markon, 2006; Lahey et al., 2004; Slade & Watson, 2006).

The complexity of data from structured diagnostic interviews for depression presents challenges for any analytic framework, and these challenges also need to be surmounted when applying latent variable approaches to such data. In particular, complex skip logic and the number of items in the structured diagnostic tools make model specification and estimation for an entire diagnostic module difficult or impossible. In tools with screening questions and exclusion criteria, a large number of participants may have missing data for all the main symptom diagnostic criteria questions at the core of the module, having been excluded based on their question responses in an earlier portion of the module. To address this challenge, many of the latent variable analyses utilize only the subset of the sample that

was not excluded from the core module, and include only the main symptom criteria questions embedded deep in the module as indicator items for the measurement model. This implicitly assumes that all respondents excluded from that section of the module have a score of “0” on the diagnostic continuum, with a corresponding classification of *not affected* (e.g., no depression present). Although using latent variable methods in this way is an improvement, in that it allows for imperfect item functioning for the core symptom criteria, it still fails to provide information on screening questions, the functioning of which could vary across populations.

The purpose of this article is to introduce an alternative latent variable approach to analyzing item-level data from a structured diagnostic interview with gate questions that addresses these current limitations by allowing for the inclusion of data on the full sample of respondents and imperfect functioning of both screening questions and exclusion criteria as well as core symptom criteria. The technique we use, latent transition analysis (LTA), is well-established in the realm of longitudinal latent variable modeling. The novelty of our approach lies not in the model itself but in the specification and application of this model to cross-sectional data reflecting response pathways through a structured interview for depression. In the remainder of the article, we briefly describe the conventional LTA model and the modification of LTA for use with structured interview data. We then apply this approach to data from a population-based study of mental health using the Composite International Diagnostic Interview (CIDI) depression module. We compare the results of our model-estimated depression classification rates using the LTA approach with results based on the *DSM-IV-TR*-based algorithm calculations, and with results based on a latent class analysis (LCA) of responses to core symptom criteria embedded in the CIDI depression module.

LTA for Structured Interviews

LCA

LTA is a longitudinal extension of LCA. In LCA, a measurement model is hypothesized in which a categorical latent variable underlies a set of manifest categorical indicator variables. As in a factor analysis, the shared variance among the manifest variables is explained by the underlying latent variable. The categories of the latent variable, also referred to as *mixture components* or *latent classes*, can be interpreted as unobservable subgroups of individuals in the population. The subgroups are characterized and distinguished by different response pattern probabilities on the observed items. The LCA model is parameterized in terms of the number of classes, the relationship between the latent class variable and the observed items (e.g., class-specific item response probabilities), and the latent class proportions. Specifically, the relationship between the observed responses, $(u_{1i}, u_{2i}, \dots, u_{Mi})$ for individual i on M items and the K -class latent variable, c , is expressed as

$$\Pr(u_{1i}, u_{2i}, \dots, u_{Mi}) = \sum_{k=1}^K [\pi_k \cdot \Pr(u_{1i}, u_{2i}, \dots, u_{Mi} | c_i = k)], \quad (1)$$

where π_k is the proportion of individual in the population with membership in Class k , and $\sum \pi_k = 1.0$. Imposing the assumption of conditional independence, that is, assuming that the underlying latent class variable explains all of the shared variance between the observed indicators and that the observed responses across the indicators within person are independent conditional on latent class membership, Equation 1 simplifies to

$$\Pr(u_{1i}, u_{2i}, \dots, u_{Mi}) = \sum_{k=1}^K \left[\pi_k \cdot \prod_{m=1}^M \Pr(u_{mi} | c_i = k) \right]. \quad (2)$$

This basic LCA model is represented in path diagram format in Figure 1. A detailed description of applications and recent developments in LCA analysis appears elsewhere (e.g., Hagnaars & McCutcheon, 2002; Masyn, 2013; Nylund, Asparouhov, & Muthén, 2007; Reinecke, 2010; Rindskopf, 2009; Vermunt & Magidson, 2004).

LTA

In a LTA, there are multiple time periods, with a latent class variable for each time period, and individuals in the population may have different patterns of latent class membership across time. The *transition* portion of the LTA describes the relationship between the latent class variables across time. In a first-order LTA, latent class membership at any given time period depends only on the latent class membership at the immediately preceding time period. The latent class variables across time need not have the same indicators, the same number of classes, nor the same substantive meaning. At each time period, t , the latent class measurement model describing the relationship between the observed responses, $(u_{t1i}, u_{t2i}, \dots, u_{tMi})$, for individual i on M_t items and the K_t -class latent variable, c_t is given by

$$\Pr(u_{t1i}, u_{t2i}, \dots, u_{tMi}) = \sum_{k=1}^{K_t} \left[\pi_{tk} \cdot \prod_{m=1}^{M_t} \Pr(u_{tmi} | c_{ti} = k) \right]. \quad (3)$$

Assuming a first-order transition model, the relationship between the latent class variables, $c_1, c_2, \dots, c_T$, corresponding to the T time periods, can then be described as a set of $(T-1)$ transition matrices, with K_t rows and K_{t+1} columns, consisting of conditional probabilities in the form

$$\begin{bmatrix} \tau_{1|1}^t & \tau_{2|1}^t & \dots & \tau_{K_{t+1}|1}^t \\ \tau_{1|2}^t & \tau_{2|2}^t & \dots & \tau_{K_{t+1}|2}^t \\ \vdots & \vdots & \ddots & \vdots \\ \tau_{1|K_t}^t & \tau_{2|K_t}^t & \dots & \tau_{K_{t+1}|K_t}^t \end{bmatrix}, \quad (4)$$

where

$$\tau_{w|v}^t = \Pr(c_{t+1}=w|c_t=v), \quad (5)$$

$$\text{and } \sum_{w=1}^{K_{t+1}} \tau_{w|v}^t = 1.0.$$

In other words, the transition probability corresponding to the $[v, w]$ element of the t th transition matrix is the probability of membership in Class w at Time $t+1$, given membership in Class v at Time t . This basic first-order LTA model is represented in path diagram format in Figure 2. In recent years, LTA models have gained much popularity for the analysis of discrete change in longitudinal data, most notably in prevention sciences research (Castellini et al., 2013; Hyppolite & Trivedi, 2012; Mackesy-Amity et al., 2013; Syed & Seiffge-Krenke, 2013; Turrise et al., 2013; Van den Akker, Dekovic, Asscher, Shiner, & Prinzie, 2012). More details about LTA and discrete-time Markov chain models (of which LTA is a special case) can be found elsewhere (e.g., Collins & Lanza, 2010; Lanza, Flaherty, & Collins, 2003; Rindskopf, 2010).

LTA for Structured Interviews

Using a measurement model based on one or more categorical latent variables, rather than a factor analytic model based on one or more underlying continuous latent variables, is appropriate for use with data from structured diagnostic interviews that are constructed from within a classification paradigm for which the “true” nature of the construct of interest is conceived of as categorical (as is often the case in psychiatry and psychiatric epidemiology).

In order to apply an LTA approach to a cross-sectional structured diagnostic interview, we recast a *time period* as a subsection of a diagnostic module with a sequence of items marked by a *gate* made up of one or more screening questions or exclusion criteria. For a given individual, a sequence of transitions across those periods reflects a pathway through the instrument and, consequently, a particular diagnostic status. The *periods* can be identified by examining the questions themselves, but also through examination of the data and the dominant skip patterns.

For each period, excepting the first, there may need to be one latent class, the *exclusion* class, that is characterized by observed exclusion from the module (with presumed exact nonendorsement of all subsequent items). Only those previously excluded from the module, based on the observed data, may be members of the exclusion class in a given period, but not all those excluded in the observed data have membership in said class. The number of latent classes considered for each period are three or four (not counting the exclusion class). For example, in the CIDI depression module, the latent class variable of the first “period” was measured by the first four questions. Three of those questions were symptom questions and the fourth was a duration question. In the observed data, any respondent not endorsing at least one symptom item *and* the duration item was ejected from the remainder of the module. Thus, in the observed data, based on those four questions, there are two observed groups of respondents: (a) respondents who screen-in to the next period (section) of the depression module, and (b) respondents who screen-out, skip the remainder of the module,

and have missing data for all remaining items of the module. A deterministic algorithm would assume the group of screened-out individuals are known to have a “zero” value on the depression continuum and a *nonaffected* diagnostic status. By using not a perfectly measured two-class latent variable but an imperfectly measured three-class or four-class latent variable underlying those first four questions, it allows for the possibility that there is an unobserved subgroup of respondents who were incorrectly screened out of the rest of module or, alternatively, an unobserved subgroup of respondents who were incorrectly screened in. For members of the *correctly screened-out* latent class in the first period, the transition probability for transitions to the exclusion class of the second period is fixed at 1.0, and transition probabilities for all other classes of the second period are fixed at 0.0. For members of the *correctly screened-in* latent class in the first time period, or an *incorrectly screened-in* latent class, the transition probability for transitions to the exclusion class of the second period would be fixed at 0.0. For members of an *incorrectly screened-out* latent class in the first time period, the transition probability for transitions to the exclusion class would be freely estimated in the model and not fixed at 1.0. A similar logic would be applied to establish the item set and number of classes for the second period of the transition model and the transition probabilities from the second to third period. Any exclusion class specified would be an *absorbing state* in the transition model, meaning that the probability of ever transitioning out of that class in a subsequent period would be fixed at zero. The final period is marked by the end of the actual diagnostic module and not by a gate. The latent classes of the final period would number at least three and include an exclusion class as well as *true clinical depression* class.

The precise number of latent classes for each time period, within these constraints, is determined through a process of model estimation and model comparison using global and relative tests of fit and fit indices in standard use for class enumeration, for example, overall model chi-square goodness of fit and the Bayesian information criterion. In the transition model, only a first-order Markov chain need be considered because of the hierarchy of the gates through the structure interview, that is, latent class membership in each period only depends directly on latent class membership in the preceding, adjacent period.

Determination of each period for the transitions and the number and type of classes is both instrument- and data-dependent. Thus, a clear and generic prescriptive explanation of the model specification is not given. In the remainder of the article, we illustrate this LTA approach with data from the CIDI depression module to provide greater clarity about the application of this technique and its potential utility.

Data Example

Sample

The data used for this analysis come from the National Institute of Health Collaborative Psychiatric Epidemiology Surveys (NIMH-CPES; Alegria, Jackson, Kessler, & Takeuchi, 2001–2003). The NIMH-CPES is comprised of three nationally representative surveys: the National Comorbidity Survey Replication, the National Survey of American Life, and the National Latino and Asian American Study. Survey data were collected from 2001 to 2003 and were based on a multistage area probability sample conducted in a total of 252

geographic areas or primary sampling units across the United States. More detailed information on the sampling procedure is provided in the NIMH-CPES data user's guide available for download on the NIMH-CPES website (University of Michigan, 2011). The tool used to diagnose depression in the CPES is a modified version of the CIDI, called the World Mental Health Composite International Diagnostic Interview (WMH CIDI).

In all three studies, all interviews were conducted using laptop-computer-assisted personal interview methods in the homes of respondents, with the exception of 14% of interviews in the National Survey of American Life that were conducted either partially or entirely by telephone. In the National Comorbidity Survey Replication, 9,282 interviews were completed (response rate, 70.9%); in the National Survey of American Life, 6,199 interviews were completed (response rate, 71.5%); and in the National Latino and Asian American Study, 4,864 interviews were completed (response rate, 80.3%). The total sample size was 20,345. Additional information about NIMH-CPES methods and data can be found in Special Issue 2 of the *International Journal of Methods in Psychiatric Research* (The NIMH Collaborative Psychiatric Epidemiology Surveys Initiative: Designs, Methods, and Instrumentation, 2004). All participants with data on the depression screening questions and depression module were used, for a total sample size of 19,734.

Measures

This analysis focused on those questions in the WMH CIDI depression module that determine diagnosis of major depressive episodes (MDE) in the *DSM-IV-TR*-based algorithm. At the beginning of the interview, the WMH CIDI includes three screening questions about sadness or depressed mood, feelings of discouragement, and loss of interest lasting several days or longer. Respondents endorsing one or more of these screening questions were administered the WMH CIDI depression module, which consists of 89 questions with dichotomous or ordinal response scales, incorporating defined skip patterns based on endorsed questions. Questions in the CIDI module relate to the presence and history of lifetime and 12-month depressive episodes. A diagnostic algorithm based on *DSM-IV-TR* criteria was applied to responses on the CIDI module to classify individuals with lifetime MDE. This analysis focused on the three depression screening questions and the questions in the CIDI depression module that are part of the diagnostic algorithm for MDE, which comprise a subset of the 89 questions.

Based on the skip logic and dominant missing data patterns (results from exclusion criteria) as well as the *DSM-IV-TR*-based diagnostic criteria for MDE, we specified three "periods" or parts of the CIDI depression module for use in the transition model. Items from each part of the structured interview module that were used in the actual algorithm for CIDI MDE diagnosis or that functioned as exclusion criteria were included as indicators for the respective latent class variables. In the following paragraphs, we describe in greater detail the item sets used as indicators for the latent class variable for each of the three parts of the LTA.

Part 1—Individuals who endorsed at least one of three depression screening questions were screened into the depression module. Endorsing none of the three screening questions would

eject an individual from the rest of the depression module in the CIDI. The screening questions asked,

Have you ever in your life had a period lasting several days or longer when most of the day you [...]:

1. Felt sad, empty or depressed.
2. Were very discouraged about how things were going in your life.
3. Lost interest in most things you usually enjoy like work, hobbies, and personal relationships.

Out of a total of 19,734 individuals who were asked the screening questions, 8,704 (44%) did not endorse any screening item and were therefore excluded from receiving the CIDI depression module. The initial questions in the module then ask about the duration of experiencing the endorsed screening criteria. If this duration criterion (lasting 2 weeks or more) was not met, individuals were not asked any other questions in the module. An additional 6,389 individuals were excluded through this criterion. The latent class indicators for Part 1 of the transition module consisted of the three depression screening questions and the initial duration criterion, all with binary response scales. In total, 15,093 of 19,734 of individuals, or 76.7%, were excluded (correctly or incorrectly) in this first period.

Part 2—Individuals who proceeded through the module were then asked questions about three other criteria. If these three criteria were met, individuals were asked the rest of the questions in the module. The three criteria were (a) severity/impairment, (b) presence of any symptoms beyond the three screening questions, and (c) length of the worst episode experienced.

The questions eliciting these criteria formed four binary indicators for a latent class model labeled Part 2. Twenty-five percent of those who made it to Part 2 were excluded (correctly or incorrectly) from the rest of the module based on responses on items in this section.

Part 3—Individuals who were not screened out in Parts 1 and 2 proceeded through the remainder of the CIDI depression module. *DSM-IV-TR* diagnosis of a major depressive episode, based on the algorithm, requires that the respondent screen-in to Part 3 and then endorse at least five of nine symptoms. These nine symptoms comprise nine of the 10 indicators for the latent class variable in Part 3 of the LTA, coded dichotomously:

1. Depressed mood.
2. Markedly diminished interest or pleasure.
3. Significant weight change or change in appetite.
4. Insomnia or hypersomnia.
5. Psychomotor agitation or retardation as noticed by others.
6. Fatigue or loss of energy.
7. Feelings of worthlessness or excessive or inappropriate guilt.

8. Diminished ability to think or concentrate, or indecisiveness.
9. Recurrent thoughts of death, recurrent suicidal ideation without a specific plan, or a suicide attempt or a specific plan for committing suicide.

DSM-IV-TR diagnosis also requires that the symptoms are not the result of the direct physiological effects of a substance, or are not attributable to a general medical condition. This criterion was captured in the 10th indicator in Part 3. Figure 3 provides a visual heuristic of the response flow through the periods or parts of the transition model and the corresponding indicators for each part.

Analysis

LCA—Prior to specifying the transition model, separate latent class measurement models were established for the latent class variables for each period. For each of the three LCAs, a quasi-confirmatory approach was taken, using an a priori set of measurement models for each period, ranging from two to three latent classes, with constraints in the potential three class models placed on class-specific item endorsement probabilities. These constraints were specified to differentiate latent classes characterized by possible errors in the screening or exclusion criterion that marked the end of that part of the module. For example, a *correctly screened-out* class for Part 1 would have an endorsement probability for the duration item fixed at 0, a *correctly screened-in* class would have an endorsement probability for the duration item fixed at 1, and any classes with screening errors would have an endorsement probability for that same item between 0 and 1.

For Parts 1 and 2, we examined three alternative models for each part: (a) an unconstrained two-class model; (b) a three-class model with one class constrained such that the exclusion gate question operated with perfect sensitivity (i.e., class-specific item endorsement fixed at 1); and (c) a three-class model with one class constrained such that the gate question operated with perfect specificity (i.e., class-specific item endorsement fixed at 0). With only four binary items for Part 1 and four binary items for Part 2, no reasonably constrained four-class models were well-identified.

The class-specific item response probabilities, as well as the class proportion parameters, are estimated using full-information maximum likelihood, under the missing-at-random assumption with the expectation-maximization algorithm. All models were estimated using the Mplus statistical software package, Version 7.0 (Muthén & Muthén, 1998–2010), which allows for the use of random start value perturbations to guard against the expectation-maximization algorithm converging to a local rather than global optimum. The complex sampling design was taken into account—cluster, strata, and individual sampling weights were included by estimating standard errors and chi-square test statistics that were robust to non-normality (Muthén & Muthén, 1998–2012) and standard errors adjusted for nonindependence of clustered observations with a post hoc sandwich estimator.

Candidate LCA models were compared using standard fit indices and split-sample, dual cross-validation. Indices of relative model fit used were the consistent Akaike information criterion, the Bayesian information criterion, the Lo-Mendel-Rubin likelihood ratio test, and the bootstrapped likelihood ratio test. Overall model-data consistency was evaluated using

the chi-square goodness-of-fit test statistic and through examination of the corresponding standardized Pearson residuals. For the split-half, dual cross-validation, the sample was randomly divided in half. The set of candidate LCA models for each part were calibrated in both of the two random halves of the data and then cross-validated in the opposite half. For more details on these fit indices, model diagnostics, and the dual cross-validation, see Masyn (2013).

LTA—After selecting the final latent class measurement model for each part, the latent transition model was estimated using the measurement parameter estimates from the LCA models as starting values. As described previously, for all parts, excluding the first, we added one latent class to the final LCA for each part. This *exclusion* class has fixed measurement parameters corresponding to zero probability of item endorsement. The transition probabilities into the exclusion classes are fixed at 1.0 for the *correctly screened-out* classes and are fixed at 0.0 for all other classes.

We compared the measurement parameter estimates for each latent class variable of the LTA model with the parameter estimates of the separate LCA models to make sure the latent class solutions from each part were stable and not being influenced by the other parts in the transition model. One thousand sets random start value perturbations were used for each model.

In order to obtain the model-estimated prevalence of MDE from the LTA model, we summed the probabilities for each transition pathway through the three parts that terminated with membership in the clinically affected class in the final part.

Difference between algorithm analysis and LTA—Depression prevalence estimated from the LTA model was then compared with prevalence computed using the *DSM-IV-TR*-based algorithm. Classification of MDE in the CPES was made by applying the *DSM-IV-TR*-based algorithm to responses on the WMH CIDI. The algorithm follows *DSM-IV-TR* criteria, such that an individual must have at least one of two key criteria (depressed mood or loss of interest/pleasure) persisting for 2 weeks or more as well as at least five other symptoms, which could include weight loss/gain, insomnia/hypersomnia, psychomotor retardation/agitation, fatigue, feelings of worthlessness/guilt, diminished ability to think/concentrate, and suicidal thoughts.

Comparison with a standard LCA model using only diagnostic criteria from Part 3—We estimated an unconstrained two-class LCA model using only the 10 diagnostic symptom criteria in the final portion of the CIDI depression module (the Part 3 item set for the LTA) and using only observed data for those individuals not excluded from that portion of the module. One latent class was interpreted as the “MDE” class. To estimate the overall MDE prevalence using this approach, we applied the assumption that all individuals excluded from that final part of the module had a known zero probability of being in the MDE class.

Results

Descriptives

Univariate descriptive statistics for all items used in the three parts are shown in Table 1.

LCA

Restricted three-class models, both with *correctly screened-out* classes, fit the data well for both Part 1 and Part 2, and were better than other two candidate models considered for each part (unconstrained two-class and restricted three-class with *correctly screened-in* class). An unconstrained two-class model fit the data well for Part 3. Although well-identified in Part 3, unconstrained LCA models with more than two classes were not favored by the relative fit indices. Each of the final LCA models had acceptable model-data consistency, with less than 10% of the response pattern Pearson standardized residuals with absolute values above 3.0, a marker for reasonable goodness-of-fit, given the large sample size and large number of response patterns. These models also cross-validated in opposite halves of the data. Class 1 for both Parts 1 and 2 are the restricted *correctly screened-out* classes. Class 3 for both Parts 1 and 2 would be considered *affected* classes that are the classes most likely to be traversed by respondents with a latent MDE status. Of the two classes in Part 3, Class 2 is characterized by higher rates of depressive symptom endorsement and could be interpreted as the clinical MDE class. Class 1 is characterized by relatively lower rates of symptom endorsement and could be interpreted as a class with subclinical or no clinical depression symptomology.

Latent Transition Model

For the LTA, we added an *exclusion* class to the latent class variables for Parts 2 and 3. The *correctly screened-out* class in Part 1 had a transition probability into the Part 2 exclusion class fixed at 1. The other two classes in Part 2 had transitions probabilities into the Part 2 exclusion class fixed at 0. Similar constraints were imposed for the Part 2 to Part 3 transitions. Additionally, the exclusion class is treated as an *absorbing state*, such that the probability of ever transitioning out of the exclusion class is fixed at zero for all subsequent periods. In the LCAs for Parts 1 and 2 of the module, Class 1 was specified as the *correctly screened-out* class. Thus, in the LTA, the transition probabilities for Class 1 in each part to the exclusion class of the consecutive part were fixed at 1, and the transition probabilities for the other class to the exclusion class were fixed at 0. The model-estimated transition probabilities are given in Table 2. The transition probabilities for Part 1 to Part 3 were obtained by multiplying the model-estimated transition matrices for Parts 1 to 2 and Parts 2 to 3.

To obtain the LTA model-estimated prevalence of MDE, the probabilities for all pathways leading to the MDE class of Part 3 must be computed and then summed. Figure 4 displays the model-estimated transition probabilities as a probability tree, with the possible pathways to the final MDE class in bold. In this model, there are a total of four latent pathways leading to “MDE.” Those chains and their corresponding probabilities are as follows:

$$\text{Start} \rightarrow \text{C1.2} \rightarrow \text{C2.2} \rightarrow \text{C3.2} = (.31)(.81)(.36) = .09$$

$$\text{Start} \rightarrow \text{C1.2} \rightarrow \text{C2.3} \rightarrow \text{C3.2} = (.31)(.07)(1.00) = .02$$

$$\text{Start} \rightarrow \text{C1.3} \rightarrow \text{C2.2} \rightarrow \text{C3.2} = (.20)(.35)(.36) = .03$$

$$\text{Start} \rightarrow \text{C1.3} \rightarrow \text{C2.3} \rightarrow \text{C3.2} = (.20)(.65)(1.00) = .13$$

The sum of those four pathway probabilities is .27. In comparison, the estimated prevalence based on the *DSM-IV-TR* algorithm applied to the same data was .14. Fitting an LCA to only Part 3—assuming all those in the original sample with missing responses to the Part 3 items are known for certain *not* to be in the MDE class—the LCA estimated proportion of those in the MDE class of the LCA multiplied by the probability of having not been excluded from that final part of the CIDI depression module was .13.

Discussion

The structure of the diagnostic interview prevents the interviewer from probing to understand an individual's experience and assess whether the individual meets the criteria; and once an individual is excluded, further questions cannot be asked. If there is a nonzero probability that individuals who do not endorse screening questions or specific questions that constitute criteria in the algorithm actually fall into a more depressed category, not considering it could compromise the effectiveness of the CIDI in identifying individuals with depression. This is plausible if, for example, individuals do not feel comfortable endorsing the screening questions or if they misjudge or misremember the duration of their symptoms—both issues that Kessler et al. (1998) cited as problems in the CIDI. These problems could also differ across subgroups based on cultural or language group, educational level, or literacy (Alegria et al., 2009), and this could impact cross-population estimates of depression.

This article proposes and illustrates with real data how an LTA can be used to model respondent pathways through a structured interview with multiple screening and gate questions with complex skip patterns. This approach is distinct from algorithm-based diagnostic determinations and conventional single latent variable measurement models, in that the LTA approach allows for imperfect functioning of the gates in the instrument, rather than assuming, as the other two conventional approaches do, that the gates function perfectly and that the diagnostic status of those excluded from the later sections of an interview is known with absolute certainty based on the observed responses to the exclusion criteria.

In the data illustration, there was a potential 14-percentage-point (approximately twofold) difference in the MDE prevalence estimates comparing the LTA results with the two conventional approaches. It is noteworthy that the probability of the most severe pathway in the LTA, leading ultimately to the MDE class (.13) is approximately the same as the prevalence estimates from the *DSM-IV-TR* algorithm and selective LCA. In this case, the difference driving the higher overall prevalence estimate from the LTA analysis is that the model, as specified, allows for the possibility that not all respondents excluded from the module were “correctly” excluded. The results from the model suggest that, based on the other three possible pathways to the MDE class in the final part, anywhere from 2% to 14% of the total respondents, assumed to be unaffected by depression, might be incorrectly classified by techniques that assume perfect sensitivity and specificity of the gates for all

individuals. The four possible pathways leading to the “MDE” class in the LTA may be substantively and clinically different. Further examination using auxiliary information is needed to validate whether the full 14% not diagnosed by the response-based algorithm represent misdiagnoses.

The threat of measurement error and differential item functioning to reliability of scores and summary classifications, and the validity of the interpretation thereof, is well-documented in the measurement and assessment literature. There is no revelation in our data illustration that accounting for imperfect item performance in a diagnostic model can yield different results than approaches that treat some or all items as error-free. And it is not our intent to label the estimate based on our approach as “right” or “true.” We have not conducted an elaborate criterion-related validity analysis, comparing the diagnostic classifications from the different analytic methods with other correlated outcomes such as functional impairment, nor are there “gold standard” clinician evaluations of the respondents available for comparison. However, viewed as a sensitivity analysis, our approach does suggest that there could be a non-negligible impact of assuming error-free functioning of screening items and gate questions on the resultant summary estimates. The LTA approach presents the possibility of investigating the performance of those gates rather than being forced to assume they are unimpeachable. The LTA approach further offers the possibility, not demonstrated here, for empirical investigation of sources of differential item function for those gates, for example, gender or cultural group membership.

As to the veracity of the LTA model-based estimates in the data illustration, several matters specific to this particular application should be noted. The division of the module was not arbitrary, but was subjective, based on decision points in the CIDI module. The separate LCA models could only be estimated using data from those respondents not excluded from the module prior to that point. We imposed a conditional independence assumption on the items for each measurement model, and although we did not evaluate the tenability of this assumption directly or consider models that relaxed this assumption, the overall goodness-of-fit assessments we did do, and the adequate model-data consistency for our final measurement models, suggest this assumption was not inconsistent with the observed data. Although more confirmatory than exploratory, the selection of the final LCA model for each part was based on inferential and information heuristic criteria as well as substantive criteria, rendering the model selection process as subjective as it is in conventional exploratory mixture modeling. For the LTA model estimation, because of the complexity of the latent transition contingencies ($3 \times 4 \times 3$) and the amount of missing data, simultaneous estimation of the full latent transition model was not possible. Therefore, the Part 1 to Part 2 transitions and the Part 2 to Part 3 transitions were estimated separately.

Using this research as a proof-of-concept, there is much future simulation-based research that could be done to quantify the impact of measurement error and bias in structured interview screening items and gate questions on overall diagnostic outcomes estimates. The current data illustration only supported measurement models for each part that allowed for imperfect gate sensitivity, with results suggesting previous prevalence estimates may have been too low. However, it may be that imperfect gate specificity could lead to overestimation of prevalence. And it is not known what sensitivity–specificity trade-off would be for these

types of instruments. The LTA approach presented here provides the foundation for those future investigations that simply could not be done with the conventional approaches.

Mental disorders are the leading cause of years lived with disability worldwide, and depression accounts for the largest part of this disease burden (Whiteford et al., 2013). Thus, the need to track levels of depression and their sensitivity to various multisectoral interventions is great. Given that structured diagnostic interviews for mental disorders are increasingly being used in large epidemiologic studies across culturally and linguistically diverse groups (Bromet et al., 2011), the need for flexible measurement models that can effectively handle data from such instruments is critical. The LTA approach presented here offers advantages over conventional techniques for obtaining population estimates from structured interviews, principally, the ability to account for measurement error and bias that could threaten the reliability and validity of those estimates, and this approach should be further developed and applied across a range of structured instruments to understand its full utility and potential.

Acknowledgments

This research was supported in part by a Presidential Fellowship from the Harvard School of Public Health.

References

- Alegria M, Jackson JS, Kessler RC, Takeuchi D. Collaborative Psychiatric Epidemiology Surveys (CPES). 2001–2003
- Alegria M, Shrout PE, Torres M, Lewis-Fernández R, Abelson JM, Powell M, Canino G. Lessons learned from the clinical reappraisal study of the Composite International Diagnostic Interview with Latinos. *International Journal of Methods in Psychiatric Research*. 2009; 18:84–95. <http://dx.doi.org/10.1002/mpr.280>. [PubMed: 19507168]
- American Psychiatric Association. Diagnostic and statistical manual of mental disorders, text revision (DSM–IV–TR). Author; 2000.
- Bromet E, Andrade LH, Hwang I, Sampson NA, Alonso J, de Girolamo G, Kessler RC. Cross-national epidemiology of *DSM–IV* major depressive episode. *BMC Medicine*. 2011; 9:90. <http://dx.doi.org/10.1186/1741-7015-9-90>. [PubMed: 21791035]
- Castellini G, Fioravanti G, Lo Sauro C, Rotella F, Lelli L, Ventura L, Ricca V. Latent profile and latent transition analyses of eating disorder phenotypes in a clinical sample: A 6 year follow-up study. *Psychiatry Research*. 2013; 207:92–99. <http://dx.doi.org/10.1016/j.psychres.2012.08.035>. [PubMed: 23021908]
- Collins, LM., Lanza, ST. Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences. Vol. 718. New York, NY: Wiley; 2010.
- Eaton NR, Krueger RF, Oltmanns TF. Aging and the structure and long-term stability of the internalizing spectrum of personality and psychopathology. *Psychology and Aging*. 2011; 26:987–993. <http://dx.doi.org/10.1037/a0024406>. [PubMed: 21728443]
- Eaton WW, Neufeld K, Chen LS, Cai G. A comparison of self-report and clinical diagnostic interviews for depression: Diagnostic interview schedule and schedules for clinical assessment in neuropsychiatry in the Baltimore epidemiologic catchment area follow-up. *Archives of General Psychiatry*. 2000; 57:217–222. <http://dx.doi.org/10.1001/archpsyc.57.3.217>. [PubMed: 10711906]
- Faraone SV, Tsuang MT. Measuring diagnostic accuracy in the absence of a “gold standard”. *The American Journal of Psychiatry*. 1994; 151:650–657. <http://dx.doi.org/10.1176/ajp.151.5.650>. [PubMed: 8166304]
- Fossati A, Beauchaine TP, Grazioli F, Carretta I, Cortinovis F, Maffei C. A latent structure analysis of Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Narcissistic Personality

- Disorder criteria. *Comprehensive Psychiatry*. 2005; 46:361–367. <http://dx.doi.org/10.1016/j.comppsy.2004.11.006>. [PubMed: 16122536]
- Garrett ES, Eaton WW, Zeger S. Methods for evaluating the performance of diagnostic tests in the absence of a gold standard: A latent class model approach. *Statistics in Medicine*. 2002; 21:1289–1307. <http://dx.doi.org/10.1002/sim.1105>. [PubMed: 12111879]
- Hagenaars, JA., McCutcheon, AL., editors. *Applied latent class analysis*. Cambridge, UK: Cambridge University Press; 2002. <http://dx.doi.org/10.1017/CBO9780511499531>
- Hyppolite J, Trivedi P. Alternative approaches for econometric analysis of panel count data using dynamic latent class models (with application to doctor visits data). *Health Economics*. 2012; 21(Suppl. 1):101–128. <http://dx.doi.org/10.1002/hec.2813>.
- Kessler RC, Abelson J, Demler O, Escobar JI, Gibbon M, Guyer ME, Zheng H. Clinical calibration of *DSM-IV* diagnoses in the World Mental Health (WMH) version of the World Health Organization (WHO) Composite International Diagnostic Interview (WMHCIDI). *International Journal of Methods in Psychiatric Research*. 2004; 13:122–139. <http://dx.doi.org/10.1002/mpr.169>. [PubMed: 15297907]
- Kessler RC, Wittchen HU, Abelson JM, McGonagle K, Schwarz N, Kendler KS, Zhao S. Methodological studies of the composite international diagnostic interview (CIDI) in the US national comorbidity survey (NCS). *International Journal of Methods in Psychiatric Research*. 1998; 7:33–55. <http://dx.doi.org/10.1002/mpr.33>.
- Krueger RF. The structure of common mental disorders. *Archives of General Psychiatry*. 1999; 56:921–926. <http://dx.doi.org/10.1001/archpsyc.56.10.921>. [PubMed: 10530634]
- Krueger RF, Caspi A, Moffitt TE, Silva PA. The structure and stability of common mental disorders (*DSM-III-R*): A longitudinal-epidemiological study. *Journal of Abnormal Psychology*. 1998; 107:216–227. <http://dx.doi.org/10.1037/0021-843X.107.2.216>. [PubMed: 9604551]
- Krueger RF, Chentsova-Dutton YE, Markon KE, Goldberg D, Ormel J. A cross-cultural study of the structure of comorbidity among common psychopathological syndromes in the general health care setting. *Journal of Abnormal Psychology*. 2003; 112:437–447. <http://dx.doi.org/10.1037/0021-843X.112.3.437>. [PubMed: 12943022]
- Krueger RF, Eaton NR, Clark LA, Watson D, Markon KE, Derringer J, Livesley WJ. Deriving an empirical structure of personality pathology for DSM-5. *Journal of Personality Disorders*. 2011; 25:170–191. <http://dx.doi.org/10.1521/pedi.2011.25.2.170>. [PubMed: 21466248]
- Krueger RF, Markon KE. Reinterpreting comorbidity: A model-based approach to understanding and classifying psychopathology. *Annual Review of Clinical Psychology*. 2006; 2:111–133. <http://dx.doi.org/10.1146/annurev.clinpsy.2.022305.095213>.
- Lahey BB, Applegate B, Waldman ID, Loft JD, Hankin BL, Rick J. The structure of child and adolescent psychopathology: Generating new hypotheses. *Journal of Abnormal Psychology*. 2004; 113:358–385. <http://dx.doi.org/10.1037/0021-843X.113.3.358>. [PubMed: 15311983]
- Lanza ST, Flaherty BP, Collins LM. Latent class and latent transition analysis. *Handbook of psychology*. 2003:663–685. Four:26. Published Online: 15 APR 2003.
- Mackesy-Amiti ME, Finnegan L, Ouellet LJ, Golub ET, Hagan H, Hudson SM, Garfein RS. Peer-education intervention to reduce injection risk behaviors benefits high-risk young injection drug users: A latent transition analysis of the CIDUS 3/DUIT study. *AIDS and Behavior*. 2013; 17:2075–2083. <http://dx.doi.org/10.1007/s10461-012-0373-0>. [PubMed: 23142857]
- Masyn, K. Latent class analysis and finite mixture modeling. In: Little, T., editor. *The Oxford handbook of quantitative methods in psychology*. Vol. 2. New York, NY: Oxford University Press; 2013. p. 551–611. <http://dx.doi.org/10.1093/oxfordhb/9780199934898.013.0025>
- Muthén, LK., Muthén, BO. *Mplus user's guide*. 7th. Los Angeles, CA: Author; 1998–2012.
- Nylund KL, Asparouhov T, Muthén BO. Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling*. 2007; 14:535–569. <http://dx.doi.org/10.1080/10705510701575396>.
- Reinecke, J. Corsini *Encyclopedia of Psychology*. Vol. 4. John Wiley & Sons; 2010. Latent class analysis.
- Rindskopf, D. *The Sage handbook of quantitative methods in psychology*. London: SAGE Publications Ltd; 2009. Latent class analysis; p. 199–218.

- Rindskopf, D. The Reviewer's Guide to Quantitative Methods in the Social Sciences. Routledge; 2010. Latent transition analysis; p. 199
- Robins LN, Helzer JE, Croughan J, Ratcliff KS. National Institute of Mental Health Diagnostic Interview Schedule: Its history, characteristics, and validity. *Archives of General Psychiatry*. 1981; 38:381–389. <http://dx.doi.org/10.1001/archpsyc.1981.01780290015001>. [PubMed: 6260053]
- Slade T, Watson D. The structure of common *DSM-IV* and ICD-10 mental disorders in the Australian general population. *Psychological Medicine*. 2006; 36:1593–1600. <http://dx.doi.org/10.1017/S0033291706008452>. [PubMed: 16882356]
- Syed M, Seiffge-Krenke I. Personality development from adolescence to emerging adulthood: Linking trajectories of ego development to the family context and identity formation. *Journal of Personality and Social Psychology*. 2013; 104:371–384. [PubMed: 22984832]
- The NIMH Collaborative Psychiatric Epidemiology Surveys Initiative: Designs, methods, and instrumentation. *Journal of Methods in Psychiatric Research*. 2004; 13(2, 4)
- Turrisi R, Mallett KA, Cleveland MJ, Varvil-Weld L, Abar C, Scaglione N, Hultgren B. Evaluation of timing and dosage of a parent-based intervention to minimize college students' alcohol consumption. *Journal of Studies on Alcohol and Drugs*. 2013; 74:30–40. <http://dx.doi.org/10.1037/a0030070>. [PubMed: 23200148]
- University of Michigan. Collaborative Psychiatric Epidemiology Surveys (CPES), 2001–2003 [United States]. 2011. Retrieved from <http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/20240>
- Van den Akker AL, Dekovic M, Asscher JJ, Shiner RL, Prinzie P. Personality types in childhood: Relations to latent trajectory classes of problem behavior and overreactive parenting across the transition into adolescence. *Journal of Personality and Social Psychology*. 2013; 104:750–764. <http://dx.doi.org/10.1037/a0031184>. [PubMed: 23276273]
- Vermunt, JK., Magidson, J. The Sage Encyclopedia of Social Science Research Methods. Sage Publications; 2003. Latent class analysis; p. 549–553.
- Whiteford HA, Degenhardt L, Rehm J, Baxter AJ, Ferrari AJ, Erskine HE, Vos T. Global burden of disease attributable to mental and substance use disorders: Findings from the Global Burden of Disease Study 2010. *The Lancet*. 2013; 382:1575–1586. [http://dx.doi.org/10.1016/S0140-6736\(13\)61611-6](http://dx.doi.org/10.1016/S0140-6736(13)61611-6).
- World Health Organization. The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines. Geneva: World Health Organization; 1992.

Figure 1.
Path diagram for basic latent class analysis model with M observed indicator items and underlying categorical latent variable, c .

Figure 2.

Path diagram for basic first-order latent transition analysis model for T time periods with $M(t)$ observed indicator items at each time period, t , measuring the categorical latent variable, c_t .

Figure 3. Heuristic diagram of the response flow through the “periods” or parts of the Composite International Diagnostic Interview depression module using the item subsets selected as latent class indicators for each part.

Figure 4.

Latent transition probability tree with pathways to “MDE” classification in bold. MDE = major depressive episode.

Table 1

Univariate Descriptive Statistics for All Latent Class Variable Indicators Grouped by Structured Interview “Periods” or Parts

Item	% Endorsing of those who were asked the question	% Not asked the question	% Endorsing out of the original <i>N</i> = 19,734
Part 1			
Depressed mood	47.12	0	47.12
Feeling discouraged	45.31	0	45.31
Anhedonia	36.0	0	36.0
Duration criterion	38.48	40.92	22.73
Part 2			
Severity ^a	96.85	76.84	22.43
Inability to be cheered up ^a	97.52	76.84	22.58
Impairment ^a	94.55	76.84	21.90
Length of episode	79.97	76.78	18.57
Part 3			
Depressed mood	99.06	81.66	18.17
Loss of interest	83.03	81.66	15.23
Appetite change	80.80	81.66	14.82
Sleep disturbance	89.0	81.66	16.33
Psychomotor retardation/agitation	48.93	81.66	8.98
Concentration problems	85.41	81.66	15.67
Fatigue	80.51	81.66	14.77
Feeling worthless	40.89	81.66	7.50
Suicidality	66.85	81.66	12.26
Physical exclusion	11.45	81.66	2.10

^aOriginal item had ordinal response options; however, if the response was the least severe option, the item is considered *not endorsed*.

Table 2

Latent Transition Analysis Model-Estimated Transition Probabilities Between Parts of the Composite International Diagnostic Interview Module

Part 2				
Exclusion		C2.1	C2.2	C2.3
Part 1	C1.1	1 ^a	0 ^a	0 ^a
	C1.2	0 ^a	.11	.81
	C1.3	0 ^a	0	.35
Part 3				
Exclusion		C3.1	C3.2	
Part 2	C2.1	1 ^a	0 ^a	0 ^a
	C2.2	0 ^a	.64	.36
	C2.3	0 ^a	0	1
Part 3				
Exclusion		C3.1	C3.2	
Part 1	C1.1	1 ^a	0 ^a	0 ^a
	C1.2	0 ^a	.52	.37
	C1.3	0 ^a	.22	.78

^aTransition probabilities that were fixed.