Use of Multiple Informant Data as a Predictor in Psychiatric Epidemiology

Nicholas J. Horton*
Nan M. Laird
Department of Biostatistics, Harvard School of Public Health, Boston, MA

Gwendolyn E.P. Zahner
Research Triangle Institute, North Carolina and
NHMRC Psychiatric Epidemiology Research Centre, Australian National University

Reprinted with permission, *International Journal of Methods in Psychiatric Research*, 1999; **8**(1):6–18.

Abstract

Multiple informant reports of psychopathology are often collected in studies of psychiatric epidemiology. Previous reports have primarily focused on methods for handling multiple informant information when it is used to measure a psychiatric outcome in a regression model. Here we deal with methods for incorporating multiple informant reports as predictor variables (covariates) in regression modeling. In general, there is no single appropriate analytic strategy in this setting. Approaches seen in the literature include separate analyses by informant, inclusion of both reports in the model, or combination of the reports (using the "OR" rule or use of concordant reports). Other approaches include use of measurement error models and latent class analysis. We review these approaches, discuss their relative advantages and disadvantages and illustrate them with an example where dichotomous reports of psychopathology are used to predict mental health service utilization in a community-based sample.

Key words: multiple informants, regression analysis, logistic models, latent class analysis, errors-in-variables models, mental-health service utilization

*Address for correspondence: Department of Biostatistics, Harvard School of Public Health, 655 Huntington Ave., Boston, MA 02115, USA, horton@hsph.harvard.edu

1 Introduction

Researchers in psychiatric epidemiology often rely on the use of multiple informants to determine psychopathology because assessment of psychopathology is inherently error prone. With some notable exceptions, most types of psychopathology are not easily described or classified into diagnostic categories (Verhulst 1995), and thus there is lack of reproducibility in classification. Continuous measures may more accurately report on an internal state, but these too are not perfect measures. These assessment problems are particularly pronounced for research involving children (Canino, Bird, Rubio-Stipec & Bravo 1995). By collecting reports from multiple informants, one expects that psychopathology (the underlying trait of interest) can be more accurately and reliably determined (Achenbach, McConaughy & Howell 1987). In the field of child psychiatric epidemiology, multiple informant data are routinely collected because children may be too young to provide reliable information on their cognitive state (Edelbrock, Costello, Dulcan, Kalas & Conover 1985, Breton, Bergeron, Valla, Lepine, Houde & Gaudet 1995). Parents and/or teachers are also asked to report on the child. Other potential informants include mental health workers, clinicians, observers, and peers.

While multiple informant data are most often used to determine an outcome in epidemiological studies, it is also commonly used as a predictor in models for some other outcome such as service use, or some other psychopathology or disorder. In this paper we discuss issues associated with the analysis of studies which utilize multiple informant data as a predictor in regression models. In this section we review how multiple informants are used in psychiatric epidemiology and give some examples of studies which utilize them. In section 2 we describe the analysis of multiple informant data. We will consider situations where one assumes that the informants can be thought of as *interchangeable* (Zahner, Pawelkiewicz, DeFrancesco & Adnopoz 1992, Bird, Gould & Staghezza 1992)—measures of the same underlying trait of interest—as well as the more common situation where the informants provide overlapping but distinct information. In section 3 we apply these methods to an example involving prediction of mental health service utilization using parent and teacher reports of psychopathology and review our recommendations in section 4.

An inherent feature of multiple informant data is that one anticipates discordant reports. If there is no discordance, the additional reports provide no new information. There are at least two explanations for this discordance. The first posits that the discordance is due to errors in measurement, in that psychopathology is difficult to gauge accurately and thus multiple measurements are needed to ascertain the true internal state of a child. This is viewed as a measurement error problem in the statistical literature. The second approach asserts that each measurement is tapping into a different aspect of the same trait, and each is providing unique information about the child. Verhulst (1995) suggests that the problem is particularly severe for children, as their behavior is greatly affected by their environment. To better understand these questions, researchers have studied agreement between informants (Fitzmaurice, Laird, Zahner & Daskalakis 1995, Edelbrock et al. 1985, Edelbrock, Costello, Dulcan, Conover & Kala 1986, Chilcoat & Breslau 1997, Jensen, Xenakis, Davis & Degroot 1988, Jensen, Traylor, Xenakis & Davis 1988, Zahner & Daskalakis 1998). Achenbach et al. (1987) performed a meta-analysis of 269 samples from the psychological literature which calculated a measure of agreement (Pearson r's) for ratings between different informant types. Table 1 displays the mean correlations

Table 1: Mean correlations between informants (meta-analysis)

	Parent	Teacher	MHW	Observer	Peer	Self
Parent	.59					
Teacher	.27	.64				
MHW	.24	.34	.54			
Observer	.27	.42		.57		
Peer		.44[a]			.73	
Self	.25	.20	.27		.26	.74[b]

a: pooled peers, b: test-retest

MHW: Mental Health Worker

Source: Achenbach et al., Psychological Bulletin, 1987

for each informant pair. Note that the correlations for what Achenbach calls "similar" pairs (those on the diagonal, i.e. two teachers) are much higher than those for dissimilar pairs (i.e. parent and teacher reports). The low correlations observed in dissimilar pairs may be the result of differences in how children behave in different settings (Zahner et al. 1992, Edelbrock et al. 1986). These results would suggest that informants might be considered interchangeable only if they are both of the same type, although for parent informants it has been shown that fathers and mothers differentially report a child's underlying state (Jensen, Traylor, Xenakis & Davis 1988).

There have been a number of papers which review methods to integrate reports from multiple informants (Canino et al. 1995, Bird et al. 1992, Offord, Boyle, Racine, Szatmari, Fleming, Sanford & Lipman 1996). One approach that is useful at the data collection stage is to resolve discordant reports using a consensus procedure. Several studies have utilized this technique (Schuckit & Smith 1996, Shaffer, Gould, Fisher, Parides, Trautman, Moreau, Kleinman & Flory 1996), either by involving a third informant if the first two disagree, or having the two informants participate in a consensus meeting.

A commonly used approach involves a separate analysis for each of the informants (Feldman

& Weinberger 1994, Bernstein, Cohen, Skodol, Bezirganian & Brook 1996, Gould, Fisher, Parides, Flory & Shaffer 1996). This approach is reasonable if one of the informants is considered "optimal", in which case it may be possible to collect information on only that informant (Bird et al. 1992). As a specific example, Boyle and Offord et al. used data from 1800 participants in the Ontario Child Health Study to predict substance use for adolescents using parent and teacher dichotomous ratings of psychopathology (Boyle, Offord, Racine, Fleming, Szatmari & Links 1993, Offord et al. 1996). The researchers analyzed each informant separately. Teacher reports of conduct disorder were significantly associated with use of alcohol and hard drugs. Parent reports also showed associations, but these were not statistically significant.

A third general approach used by many researchers in psychiatric epidemiology is to combine dichotomous informant ratings using the "OR" or "AND" rules. With the "OR" rule, a child is considered to have the trait of interest if any of the informants report its presence (Zahner et al. 1992, Kaufman, Jones, Stieglitz, Vitulano & Mannarino 1994). As one example, Beardslee, Wright, Salt, Drezner, Gladstone, Versage & Rothberg (1997) describe an intervention study involving a cohort of families with young children and at least one parent with an affective disorder. To judge the effect of the intervention on the well-being of the children in these families, they modeled child psychopathology using as predictors intervention status and changes in a set of parental behaviors and attitudes, where a change was considered present if either parent scored positively for the change. In this case the two parents are the two informants of change.

Other approaches to combine reports have been suggested. The "AND" rule considers a child to have the trait only if all informants report its presence. A related approach for continuous predictors involves averaging the separate continuous reports to form a single continuous report (Allen, Kuperminc, Philliber & Herre 1994). Some researchers have considered only the subjects with concordant reports (Marshall & Graham 1984), dropping those subjects with discordant reports from the analysis. Another approach includes both informant reports in the model in some situations (Bird et al. 1992). We will now review some of the advantages and disadvantages of these approaches.

2 Analyzing data with multiple informants as predictors

We will consider a simple yet common setting: we have a single dichotomous outcome Y for each subject, and two manifest dichotomous informant reports X and Z. We consider dichotomous reports because psychiatric epidemiologists often use diagnostic instruments which generate a binary indicator of *caseness*. Murphy (1995) discusses the relative merits of dimensional vs. diagnostic measurement.

Our example will involve parent reports and teacher reports as X and Z, respectively. There may also be a vector (denoted in bold) of length k of other covariates $\mathbf{W}$. In some settings it is appropriate to assume that there is a dichotomous latent (unobserved) variable corresponding to the true state of the child. We denote this true state as Q.

We can identify two classes of regression models for use with multiple informants:

1. We wish to make inferences about the conditional distribution of Y given the observed (manifest) informant reports, $f(Y|X, Z, \mathbf{W})$, or
2. We wish to make inferences about the conditional distribution of Y given the true (unobserved or latent) child psychopathology, $f(Y|Q, \mathbf{W})$.

The interpretation of these models is different. In the former (usual) case, we are interested in the actual report—even though it may be "noisy"—because we are interested in prediction using information available to clinicians and researchers. In the latent variable case, we want to gauge the effect of the true, but unobserved construct. Given this setting and these classes of regression models, we will now describe the advantages and disadvantages of the various approaches which we have cataloged.

2.1 Approaches using the manifest variables

We will begin by considering one of the simplest approaches: fitting multiple models, each using only one informant report. This approach may also be taken if there is particular interest in the different informants, if one informant is considered to be superior to other informants (an *optimal informant* (Loeber, Green, Lahey & Stouthamer-Loeber 1989)), or because of complications arising from missing data for other informants.

We may first wish to consider whether regression estimates are *sensitive* to a particular informant; i.e. is

$$E(Y|X, \mathbf{W}) = E(Y|Z, \mathbf{W})? \tag{1}$$

As one test for sensitivity, consider fitting the linear logistic model (though other models may be considered):

$$\text{logit}\left(E\left[\begin{pmatrix} Y \\ Y \end{pmatrix}\right]\right) = \begin{pmatrix} \beta_0 + \beta_1 X + \boldsymbol{\beta}_2 \mathbf{W} \\ (\beta_0 + \gamma_0) + (\beta_1 + \gamma_1) Z + (\boldsymbol{\beta}_2 + \boldsymbol{\gamma}_2)\mathbf{W} \end{pmatrix}$$

where the outcome is paired with the separate informant reports. We are interested in testing:

$$H_0 : \boldsymbol{\gamma} = \begin{pmatrix} \gamma_0 \\ \gamma_1 \\ \boldsymbol{\gamma}_2 \end{pmatrix} = \mathbf{0}. \tag{2}$$

If the null hypothesis of equal coefficients is rejected, then the linear logistic regression models are sensitive to the choice of informant: that is, the association between the informant reports and the other factors in the study—including the outcome—varies by informant. This would suggest that reporting the results based only on one informant may be misleading.

In the case where the test of sensitivity does not reject, it may be reasonable to consider the informants to be replicate measurements of an underlying trait. We are interested in the effect of the true underlying trait, but only observe that trait with some random misclassification. In simple situations, it can be shown that non-differential measurement error (i.e. the probability of mis-measurement does not depend on other study variables (Greenland 1980)) biases parameter estimates towards the null (Carroll, Ruppert & Stefanski 1995). Most analyses, however, are more complicated than these situations, and parameter estimates are not always attenuated to the null. In a multiple regression setting with multiple covariates measured with error, the bias may be in any direction (Carroll et al. 1995). In this situation, besides being an inefficient use of resources, utilizing only one informant report in the regression model does not take into account this bias due to measurement error. A variety of ad-hoc approaches have been suggested which attempt to address this problem.

Use of the "OR" rule, the "AND" rule or concordant pair analysis attempts to reduce measurement error by utilizing the information in the separate reports to construct a report that is presumed to be more accurate. The "OR" rule is thought to mimic the process used by clinicians to reconcile discordant reports, though clinicians differentially weight the contributions from each informant, and have the advantage of training, experience and intuition to inform their decisions (Bird et al. 1992). It is hoped that the combined report will have little or no residual measurement error. Brenner & Blettner (1993) performed a series of simulation studies to determine the performance of these combination rules under a common set of assumptions. They found that for the "OR" and "AND" rules, parameter estimates were attenuated towards the null, but that the combined reports had less attenuation than models using a single individual report. Besides being inefficient (Walter 1984) because data from discordant pairs are dropped from the analysis, Brenner & Blettner (1993) showed that an analysis using only concordant pairs yields an overestimate for the effect of that variable (i.e. the analysis was biased away from the null).

Another approach suggests using both informants, possibly including an interaction between the informants. This model can be parametrized as:

$$\text{logit}\left(E\left[Y_i\right]\right) = \alpha_0 + \alpha_1 X_i + \alpha_2 Z_i + \alpha_3 X_i Z_i. \tag{3}$$

This is equivalent to fitting a saturated model for the $2\times2 = 4$ possible informant patterns. Although more complicated models might allow these informant effects to vary by other covariates, to simplify our exposition we do not further consider these models. If the goal is to control for psychopathology of the child, then fitting the saturated (four parameter) model is the best model using manifest (observed) covariates. This approach has the disadvantage of being complicated to interpret, as well as ignoring measurement error remaining after observing the multiple informant reports. Several authors have noted that such a model does not completely adjust for residual measurement error (Brenner 1992, Kaldor & Clayton 1985), and they suggest the use of an explicit model which incorporates measurement error.

It is possible to test whether it is reasonable to collapse over the discordant reports with the test:

$$H_0 : \alpha_1 = \alpha_2.$$

This test is similar to a test that $\gamma_1 = 0$. If this null hypothesis is not rejected, then it may be plausible to fit a model with two parameters (in addition to the intercept), one for a single report, and one for the group with two positive reports. Further simplifications are possible, since the "OR" and "AND" rules are special cases of this model with certain constraints. The "OR" rule is equivalent to setting $\alpha_1 = \alpha_2 = -\alpha_3$, and the "AND" rule is equivalent to setting $\alpha_1 = \alpha_2 = 0$. Each of these models has one parameter in addition to the intercept, and the adequacy of the "OR" or "AND" rules can be tested using the appropriate linear combinations of α_1, α_2 and α_3. For example, to test the adequacy of the "OR" rule, a 2 df Wald test or likelihood ratio test can be used.

2.2 Measurement error and latent class models

There is an extensive literature dealing with approaches to correct for the bias introduced by measurement error (see Bashir & Duffy (1997) for a comprehensive review). If auxiliary information is available about the extent and type of non-differential measurement error, it is possible to make a correction for this bias, and many authors have proposed techniques using differing approaches and assumptions (Rosner, Spiegelman & Willett 1990, Liu & Liang 1991, Rosner, Spiegelman & Willett 1992, Richardson & Gilks 1993, Kuha 1994, Haukka 1995, Forbes & Santner 1995). In general these approaches require auxiliary information, such as a validation study or a reliability study.

A validation study presupposes that there is some underlying truth, and there exists a "gold standard" measurement that may be expensive or time consuming to administer, and hence is not available for all subjects. For some subjects, data are collected on both the cheaper, less accurate measure as well as the "gold standard", and this information is used to correct for measurement error bias. This approach was used by Rosner et al. (1990), who used a first-order Taylor series approximation to correct the parameter estimates and variance estimates for mismeasured continuous covariates in a logistic regression. It is common for nutritional epidemiologists to consider a food diary to be such a "gold standard", as opposed to the easier to obtain but less accurate food questionnaire. Holford & Stack (1995) use the term "alloyed gold standard" to these all-too-common situations where we do not have a measure without fault. Wachholder, Armstrong & Hartage (1993) cautioned that measurement error corrections which assume a gold standard may yield substantial bias when there is no true gold standard.

In many settings—and in general in child psychiatric epidemiology—there may be no "gold standard", and hence it is not feasible to carry out a validation study. Instead, a reliability study may be performed. A reliability study involves the collection of multiple observations for each covariate measured with error, and uses this information to estimate the true relationship between the covariates and the outcome. This presupposes that if it were possible to observe an infinite number of observations of the same continuous trait, the average of these observations would yield a perfect estimate of the truth. Rosner et al. (1992) suggested an approach for this setting which generalized their previous work to a setting with a replication study.

There are problems with use of these measurement error models in child psychiatric epidemiology. These models generally assume that the measurement error variance is relatively small, and often make other assumptions that may not be realistic in practice. For example, the approach of Rosner et al. requires a rare disease assumption ($P(\text{outcome}) < 0.05$) as well as continuous measures. There are also design issues to be considered. For both the validation and replication models, it is generally assumed that multiple measurements will be taken on only a small subset of the study population. This is in contrast to the usual multiple informant setting, where in general multiple informants reports are gathered for all subjects, apart from unintended missing data. Collecting multiple measurements on all subjects may be an inefficient use of resources, since it is usually possible to estimate the attenuation parameters with auxiliary information from only a subset of subjects. More seriously, the assumption that the informant reports are interchangeable is not always tenable.

Latent class models (Goodman 1974, Formann & Kohlmann 1996) offer another approach to handling measurement error when multiple reports are available but no gold standard exists. These models assume that the true underlying value cannot be directly observed, but this value is related in some fashion to observable discrete quantities. Unlike latent variable models, which allow the existence of a continuous unobserved variable, latent class models presume that the unobserved variable takes on a discrete number of values, or classes. For example, a single dichotomous latent class variable corresponds to two latent classes. Given certain assumptions (to allow identifiability), latent class models allow the calculation of the probability that a subject is in a given latent class.

Latent class models have been adapted to the binary regression setting with multiple measurements of exposure and where interest centers around estimation of relative odds of *true* unobserved exposure (Walter 1984). Identifiability of these models is problematic; in general analytic solutions are not available. Formann & Kohlmann (1996) report that two classes are usually identifiable if four dichotomous manifest variables are observed, but that even in this setting identifiability depends on characteristics of the data set under investigation. Thus, when there are only two dichotomous observations on exposure per individual, a number of constraints are needed to identify the model (Walter & Irwig 1988). For the setting where two observations are taken on each individual, and there are two groups, Walter and Irwig assumed that the sensitivity and specificity were constant across populations. Drews, Flanders & Kosinski (1993) generalized this model to allow the errors of the tests to be dependent (though the parameters governing this association are assumed to be known), and proposed an EM algorithm (Dempster, Laird & Rubin 1977) to fit their model.

Kaldor & Clayton (1985) proposed a latent class model to address misclassification of exposure in epidemiological studies. Their approach assumes that conditional on the true exposure, the measurement errors of the two informants are independent, and that these errors are independent of the outcome of interest and the other covariates. More formally, let $m_{\mathbf{V}}$ be the actual number of individuals in a given category $\mathbf{v} = (y, z, x, \mathbf{w}, q)$ where y is the outcome, x and z are the informant reports, $\mathbf{w}$ is a vector of discrete covariates, and q is an unobserved latent class. In a setting with a dichotomous outcome, two dichotomous informant reports, k dichotomous covariates, and a single dichotomous latent variable (corresponding to two latent classes), $\mathbf{v}$ will take on $2^{1+2+k+1}$ different values. The observed data can be represented by a 2^{1+2+k} contingency table, since we are only able to observe $m_{yzxw.}$ (i.e. the counts summed over the unobserved variable). We can model these counts with a log-linear model, where $u_{\mathbf{V}} = E[m_{\mathbf{V}}]$ and $\log(u_{\mathbf{V}})$ is linear in parameters which describe main effects and interactions of the parameters. For the case-control setting with multiple measurements of an exposure and a single W, Kaldor and Clayton suggested models of the form:

$$\log(u_{\mathbf{V}}) = \theta + \theta_y^Y + \theta_w^W + \theta_x^X + \theta_z^Z + \theta_q^Q + \theta_{xq}^{XQ} + \theta_{zq}^{ZQ} + \theta_{wq}^{WQ} + \theta_{yq}^{YQ} + \theta_{yw}^{YW}.$$

We can represent this model using the following notation:

$$XQ/ZQ/WQ/YQ/YW$$

where the existence of an interaction presumes the inclusion of the main effects of variables in that interaction. This model assumes that conditional on the true latent class Q, X and Z are independent of both W and Y and each other. In these models, our primary interest relates to the parameters YQ (the interaction between the outcome and the latent class) and YW (the interaction between the outcome and the covariates). To fit models corresponding to logistic regression models for Y on W and Q, we must include the saturated WQ interaction. In this case we can interpret θ_{11}^{YQ} as the log odds for the relationship between the outcome and the latent measurement of exposure. Because of this saturated model for the covariates, it may be difficult to fit models for which the dimension of $\mathbf{w}$ is large or the latent class takes on more than two values (Dawid & Skene 1979). Another limitation of this approach is that the model requires discrete covariates since it utilizes a log-linear model for the cell counts.

Kaldor and Clayton proposed using the EM algorithm to fit the model by augmenting the observed data to include the latent dichotomous variable. This approach is equivalent to Ibrahim's EM by method of weights (Ibrahim 1990), where the missing covariate (the latent class variable) is never observed. At convergence, it is possible to estimate the probability that a subject is in a given latent class given their covariate values and outcome.

Examples of the use of latent class models in the literature include estimation of smoking experimentation in children (Fergusson & Horwood 1989), modeling of nicotine withdrawal in women (Madden, Bucholz, Dinwiddie, Slutske, Bierut, Statham, Dunne, Martin & Heath 1997), and creation of clusters of social, demographic and maternal characteristics which predict medical service use (Baker & Taylor 1997).

3 Example: Mental Health Service Utilization

3.1 Methods and Measures

We will illustrate these methods with a predictive model of service utilization for a sample of children in New Haven and Eastern Connecticut. The investigators collected information on psychopathology from both parents and teachers as predictors of use of school-based mental health services. A full description of the methods and dataset have been published previously (Zahner, Jacobs, Freeman & Trainor 1993, Zahner et al. 1992, Zahner & Daskalakis 1997). In 1986-87 a random sample of 6-11 year olds was taken from 54 schools in New Haven. In 1988-89 a multistage cluster sample was taken from 83 schools nested within strata consisting of small cities, suburban areas, or rural areas. Response rates from parents ranged from 70% of those with valid addresses in New Haven to 72% of those with valid addresses in eastern Connecticut. In this example, we will consider 1382 children with complete data from both parent and teacher informants.

The outcome of interest in this study is mental health service utilization in school-based settings. Service use was defined as a parental report that the child had ever seen a provider or been in a special program at school for a behavioral problem. If the service was used, the outcome was coded 1, and coded 0 otherwise.

In this example, predictors of service use include mismeasured variables of symptoms and impairment that were collected from multiple informants, as well as two variables (age and sex) that were obtained from administrative records and verified by parents and are assumed to be free of substantial measurement error.

We created a dichotomous report of psychopathology symptoms (SYMPAR and SYMTCH for parents and teachers, respectively) from the Child Behavior Checklist (CBCL (Achenbach 1991a)) and Teacher Report Form (TRF (Achenbach 1991b)) total problems score (set equal to 1 if the t-score is greater or equal to 60, the published cutpoint for borderline/clinical psychopathology). The discretized report is being used in an exploratory fashion to illustrate the example; Achenbach (1991a) discusses some of the advantages and disadvantages of using a specific categorical cutpoint. We note that there can be a loss of information from arbitrary dichotomization of continuous variables; we do not further consider this issue.

Dichotomous reports of impairment were collected from parents (IMPPAR) and teachers (IMPTCH) by asking whether current problems on the checklist prevented the child from doing other things that other children of the same age do. Other covariates of interest include age (OLD: 0=age 5-8, 1=age 9-11) and child gender (BOY: 0=girl, 1=boy). Table 2 displays the means for these covariates.

Table 2: Mean of covariates

Variable	Mean
OUTCOME	0.164
BOY	0.491
OLD	0.479
SYMPAR	0.187
SYMTCH	0.178
IMPPAR	0.030
IMPTCH	0.081

Table 3 displays the agreement (assessed by Cohen's κ) between the dichotomized CBCL

Table 3: Cross-tabulation of teacher and parent reports of symptoms and impairment

Symptoms (total problems)			Impairment		
PARENT	TEACHER REPORT		PARENT	TEACHER REPORT	
REPORT	0	1	REPORT	0	1
0	977	147	0	1242	28
1	159	99	1	99	13

and TRF reports for symptoms and impairment. The agreement between the symptom variables was fair (Landis & Koch 1977) ($\kappa = 0.26$). Zahner & Daskalakis (1998) have explored sources of informant variance for parent and teacher reports of symptoms in this sample. They found that age, sex, academic functioning, perceived need for treatment and mental health provider preference were predictors of agreement. The agreement between the impairment variables was slight ($\kappa = 0.13$). We note that the impairment measure is a low prevalence item, particularly for teachers.

Table 4 displays the correlation between the informant reports and the covariates. Note

Table 4: Correlations between predictors

	OLD	BOY	SYMPAR	SYMTCH	IMPPAR
BOY	-0.034				
SYMPAR	0.005	0.024			
SYMTCH	-0.003	0.088	0.258		
IMPPAR	0.037	0.042	0.190	0.108	
IMPTCH	-0.014	0.059	0.164	0.340	0.151

that the highest correlation (0.340) is between the teacher reports of impairment and symptoms, and that none of the correlations is large in magnitude.

3.2 Regression using manifest outcomes

We will begin by considering models utilizing the manifest variables. If we believe that the informants are measuring different constructs, then it may be appropriate to fit separate models, or include both informants in the model. If we believe that the informants are measuring the same construct, then a reasonable approach would involve a combination of the informant reports (using for example the "OR" and "AND" rules). Table

5 displays the parameter estimates and standard errors for both types of models, along

Table 5: Parameter estimates (and standard errors) for logistic regression models using manifest variables

Parameter	Parent	Teacher	P+T	OR rule	AND rule
BOY	0.50 (0.15)	0.43 (0.15)	0.45 (0.16)	0.41 (0.16)	0.52 (0.16)
OLD	0.42 (0.15)	0.50 (0.15)	0.48 (0.16)	0.52 (0.16)	0.44 (0.15)
SYMPAR	1.37 (0.16)		1.13 (0.17)		
SYMTCH		1.09 (0.18)	0.82 (0.19)		
SYMOR				1.21 (0.16)	
SYMAND					1.88 (0.23)
IMPPAR	1.29 (0.34)		1.00 (0.37)		
IMPTCH		0.89 (0.23)	0.71 (0.25)		
IMPOR				0.98 (0.21)	
IMPAND					2.19 (0.75)
adj.pseudo-R^2	0.14	0.12	0.18	0.15	0.13

with the generalized coefficient of determination (pseudo-R^2) (SAS Institute 1996). The generalized coefficient of determination is defined as:

$$R^2 = 1 - \left(\frac{L(\mathbf{0})}{L(\widehat{\boldsymbol{\beta}})} \right)^{2/n}$$

where $L(\mathbf{0})$ is the likelihood for the intercept only model, $L(\widehat{\boldsymbol{\beta}})$ is the likelihood for the model of interest, and n is the number of observations. The adjusted coefficient of determination is normalized such that it can achieve a maximum value of 1, and represents the proportion of variation accounted for by the model. We note that models with parent information have slightly higher pseudo-R^2 values than those with teacher information, and that the model with both sets of reports has a higher value than the models which use the “OR” rule or the “AND” rule. We also note that each informant remains a significant predictor even with the other informant included in the model, suggesting that informants provide different information. The coefficients for informant using the “AND” rule is higher than for the other models, but this coefficient also has a larger standard error than in the other models.

We also fit a model which included the interaction between parent and teacher reports of symptoms (SYMPxT) and parent and teacher reports of impairment (IMPPxT), but neither of these interactions were statistically significant. From this model, however, we can statistically test whether the “OR” rule fits by applying the constraints:

$$\text{SYMPAR} = \text{SYMTCH} = -\text{SYMPxT},$$

$$\text{IMPPAR} = \text{IMPTCH} = -\text{IMPPxT},$$

or whether the "AND" rule fits by applying the constraints:

$$\text{SYMPAR} = \text{SYMTCH} = \text{IMPPAR} = \text{IMPTCH} = 0.$$

The test of the "OR" rule yielded a test statistic of 32.06, which under the null hypothesis is distributed as a $\chi^2_4,\ p < 0.0001$. The test of the "AND" rule yielded a test statistic of 39.28, which under the null hypothesis is also distributed as a $\chi^2_4,\ p < 0.0001$. Neither of these combinations rules appear to be a good fit to the observed data, though the pseudo-R^2 values for these combination rule models are comparable to those of the models which use just a single informant.

3.3 Sensitivity of regression models to choice of informant

We can examine sensitivity of the regression estimates to the different informants. We constructed a dataset with two records per child with the same outcome for each child, but different predictors. We tested the null hypothesis that $\boldsymbol{\gamma} = \mathbf{0}$ using a generalized estimating equation approach with an independence working covariance structure and a "sandwich" variance estimator (Liang & Zeger 1986) to account for the association between the outcomes. The parameter and standard error estimates for this model can be found in Table 6. Our test of $\boldsymbol{\gamma} = \mathbf{0}$ yielded a test-statistic $\chi^2_5 = 8.26,\ p = 0.143$. This

Table 6: Test of Sensitivity of Informant reports

	Parameter	Estimate	SE	Z-score	p-value
β_0	INTERCEPT	-2.554	0.156	-16.35	0.0001
$\boldsymbol{\beta}_1$	SYMPTOM	1.372	0.167	8.20	0.0001
	IMPAIR	1.180	0.400	2.96	0.003
$\boldsymbol{\beta}_2$	OLD	0.463	0.155	2.99	0.003
	BOY	0.510	0.155	3.30	0.001
γ_0	INFORMANT	0.067	0.076	0.88	0.38
$\boldsymbol{\gamma}_1$	INF*SYMPTOM	-0.282	0.214	-1.31	0.19
	INF*IMPAIR	-0.292	0.439	-0.67	0.51
$\boldsymbol{\gamma}_2$	INF*OLD	0.034	0.052	0.65	0.51
	INF*BOY	-0.075	0.052	-1.44	0.15

does not provide much evidence to reject the hypothesis that the regression equation is insensitive to the choice of informant. This agrees with our finding that the coefficients of OLD and BOY are not sensitive to choice of informant model. However, the power of this test may be low, so subject area knowledge should also be used in determining the appropriate analysis.

3.4 Latent class models

We next considered models which condition on the latent (unobserved) reports of symptoms and impairment. Kaldor & Clayton (1985) assumed that conditional on the true latent class, the informant reports are independent of each other, the outcome, and the other covariates. These model were fit using a SAS macro (SAS Institute 1996) which calculated the expected value of the latent variable and maximized the resulting log-linear model (using the augmented data) at each step of the iteration.

We were interested in a model with two latent classes as predictors: one for symptoms and one for impairment, but this model cannot be fit without assuming unrealistic constraints given only two informants. Instead, we fit a model which included SYMOR and which modeled impairment with a latent variable, and a parallel model which included IMPOR and modeled symptoms as a latent variable. Neither of these models reliably converged, possibly due to the low prevalence of impairment reports and their correlation with the symptom scores. We next fit a model which included age and sex as predictors, along with a latent variable with 2 classes for symptoms. Table 7 displays the parameter estimates for this model (with 2000 bootstrap samples used to estimate standard errors (Efron &

Table 7: Selected parameter estimates with bootstrap standard error estimates for latent class model for symptom scores

Parameter	Estimate	StdErr
BOY	0.42	0.21
OLD	0.43	0.22
SYMLATENT	2.78	0.36

Tibshirani 1993)). The results in Table 7 are consistent with those of Table 5. The SYMLATENT parameter (the log odds ratio of the relationship between the outcome and the latent informant report) is larger, though all the estimated standard errors for these coefficients are also larger.

Table 8 displays the probability of being in the latent class associated with psychopathology for each combination of outcome and covariate. We note that for subjects with no service use and no reports of symptoms, the estimated proportion in the latent class associated with psychopathology is low (ranging from 0.008 to 0.011). For subjects who

Table 8: Estimated probability of being in latent class associated with psychopathology

OUTCOME	OLD	BOY	SYMPAR	SYMTCH	Probability
0	0	0	0	0	0.008
0	0	0	0	1	0.114
0	0	0	1	0	0.148
0	0	0	1	1	0.741
0	0	1	0	0	0.009
0	0	1	0	1	0.126
0	0	1	1	0	0.163
0	0	1	1	1	0.762
0	1	0	0	0	0.008
0	1	0	0	1	0.114
0	1	0	1	0	0.147
0	1	0	1	1	0.740
0	1	1	0	0	0.011
0	1	1	0	1	0.151
0	1	1	1	0	0.193
0	1	1	1	1	0.798
1	0	0	0	0	0.106
1	0	0	0	1	0.661
1	0	0	1	0	0.723
1	0	0	1	1	0.977
1	0	1	0	0	0.117
1	0	1	0	1	0.686
1	0	1	1	0	0.746
1	0	1	1	1	0.980
1	1	0	0	0	0.105
1	1	0	0	1	0.660
1	1	0	1	0	0.723
1	1	0	1	1	0.977
1	1	1	0	0	0.140
1	1	1	0	1	0.729
1	1	1	1	0	0.783
1	1	1	1	1	0.983

used services but had no report of symptoms, the proportions ranged from 0.106 to 0.140. These estimated proportions give some sense of the magnitude of measurement error estimated in this model. We can consider how to combine symptom reports by considering the proportions in the latent class given parent and teacher reports of symptoms. A single teacher report had lower estimated proportion than a single parent report, but this difference was small in magnitude. We can consider the effect of using the "OR" rule or the "AND" rule from this table as well. For subjects with no report of service use, the "AND" rule seems appropriate, since the estimated proportion for subjects with 2 positive reports of symptoms is much higher than that for subjects with just a single positive report. Similarly, for subjects with a report of service use, the "OR" rule seems appropriate. Since different rules are needed for the different values of the outcome, neither of these combination rules seems appropriate in this example.

4 Discussion

In this paper we have reviewed a number of approaches for the use of multiple informants as dichotomous predictors in regression models. The general lack of agreement between multiple informants highlights the need for a variety of approaches to analysis. The researcher must explore the sensitivity to underlying assumptions of different methods which take into account the measurement error inherent in assessing psychopathology. This measurement error tends to weaken observed associations by attenuating parameter estimates towards the null, but the bias is not always predictable. Attempts to combine multiple informants using ad-hoc methods such as the "OR" or "AND" rules tend to decrease, but not eliminate, this bias. These rules have the advantage of yielding more parsimonious models, at the cost of some loss of predictive ability. For our example, we note that each informant remains a strong predictor even in the presence of the other in the model, which suggests that they provide separate information.

Approaches such as Kaldor and Clayton's latent class model provide a conceptually attractive method for dealing with multiple informants in the measurement error setting. Fergusson & Horwood (1989) studied maternal and child reports of maternal and child smoking. Use of latent class models for the correction of measurement errors in those reports resulted in a substantial increase in the strength of association between these variables. We believe that latent class approaches are the preferred approach when the underlying assumptions are reasonably met, but in general this requires more data (i.e. more than two informants) than is generally available in most investigations. These models are not panaceas: they add computational difficulty to model-fitting, complexity to model-interpretation and in our example, did not always converge (perhaps due to the low base rate of impairment). The former problem, while a nuisance, is alleviated by faster computers, though the number of nuisance parameters to be fit may limit the number of covariates that can be included in a model. In our example, this limitation precluded our ability to fit models with both reports of symptoms and impairment. We concur with the suggestion of Formann & Kohlmann (1996) that perhaps the best use of latent class analysis may be to understand misclassification rates and shed light on how best to combine information from multiple informant reports.

The assumptions underlying the latent class models—like any modeling assumptions—must be carefully considered to ensure valid results. Brenner (1992) found situations where positive error correlation between the dual measurements led to biased latent class estimates. We believe that in our setting, with parents and teachers providing separate reports, the conditional independence assumption is plausible, but this question needs to be considered in each analysis.

We suggest the following guidelines for use of multiple informants as a predictor:

- It is possible to indirectly test whether informants are reporting on the same underlying construct, but these tests may not have much power, and subject area knowledge should be brought to bear on this question. The answer to this question will help to guide the analysis.
- If the informants provide distinct information (i.e. may be measuring different underlying constructs) then separate models may be appropriate to report if the effect of the individual informant is of primary interest. If the goal is to control for the variable reported by the multiple informants, then the model with both sets of reports works well.
- For settings where the informants are both reporting on the same underlying construct, it is appropriate to consider combining the reports using the "OR" rule (an ad-hoc approach) or to use latent class models (which require more assumptions).
- Latent class models are particularly well suited to settings where interest revolves around a single unobserved construct and there are many manifest reports. With only two informants, many assumptions are needed to fit these models.

Acknowledgments

We are grateful for the support provided by NIMH grants T32-MH17119, R01-MH54693-02 and U01-MH51465. We would also like to thank Constantine Daskalakis and the reviewers for their helpful comments. The Connecticut child surveys were conducted under contract to the Connecticut Department of Children and Youth Services while Dr. Zahner was on the faculty of the Yale University Child Study Center. Table 1 is reprinted with permission of the American Psychological Association and Dr. Thomas M. Achenbach.

References

Achenbach, T. M. (1991a). *Manual for the Child Behavior Checklist/4-18 and 1991 Profile*, University of Vermont: Department of Psychiatry.

Achenbach, T. M. (1991b). *Manual for the Teacher's Report Form and 1991 Profile*, University of Vermont: Department of Psychiatry.

Achenbach, T. M., McConaughy, S. H. & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: implication of cross-informant correlations for situational specificity, *Psychological Bulletin* **101**: 213–232.

Allen, J. P., Kuperminc, G., Philliber, S. & Herre, K. (1994). Programmatic prevention of adolescent problem behaviors: the role of autonomy, relatedness, and volunteer service in the teen outreach program, *American Journal of Community Psychology* **22**(5): 617–638.

Baker, D. & Taylor, H. (1997). Inequality in health and health service use for mothers of young children in southwest England, *Journal of Epidemiology and Community Health* **51**(1): 74–79.

Bashir, S. A. & Duffy, S. W. (1997). The correction of risk estimates for measurement error, *Annals of Epidemiology* **7**(2): 154–164.

Beardslee, W. R., Wright, E. J., Salt, P., Drezner, K., Gladstone, T. R. G., Versage, E. M. & Rothberg, P. C. (1997). Examination of children's response to two preventive intervention strategies over time, *Journal of the American Academy of Child Adolescent Psychiatry* **36**(2): 196–204.

Bernstein, D. P., Cohen, P., Skodol, A., Bezirganian, S. & Brook, J. S. (1996). Childhood antecedents of adolescent personality disorders, *American Journal of Psychiatry* **153**(7): 907–913.

Bird, H. R., Gould, M. S. & Staghezza, B. (1992). Aggregating data from multiple informants in child psychiatry epidemiological research, *Journal of the American Academy of Child Adolescent Psychiatry* **31**(1): 78–85.

Boyle, M. H., Offord, D. R., Racine, Y. A., Fleming, J. E., Szatmari, P. & Links, P. S. (1993). Predicting substance use in early adolescence based on parent and teacher assessments of childhood psychiatric disorder: results from the Ontario child health study follow-up, *Journal of Child Psychology and Psychiatry* **34**(4): 535–544.

Brenner, H. (1992). Use and limitations of dual measurements in correcting for nondifferential exposure misclassification, *Epidemiology* **3**: 216–222.

Brenner, H. & Blettner, M. (1993). Misclassification bias arising from random error in exposure measurement: Implications for dual measurement strategies, *American Journal of Epidemiology* **138**: 453–461.

Breton, J. J., Bergeron, L., Valla, J. P., Lepine, S., Houde, L. & Gaudet, N. (1995). Do children aged 9 through 11 years understand the DISC version 2.25 questions?, *Journal of the American Academy of Child Adolescent Psychiatry* **34**(7): 946–956.

Canino, G., Bird, H. R., Rubio-Stipec, M. & Bravo, M. (1995). Child psychiatric epidemiology: What we have learned and what we need to learn, *International Journal of Methods in Psychiatric Research* **5**: 79–92.

Carroll, R. J., Ruppert, D. & Stefanski, L. A. (1995). *Measurement Error in Nonlinear Models*, Chapman & Hall.

Chilcoat, H. D. & Breslau, N. (1997). Does psychiatric history bias mothers' reports? an application of a new analytic approach, *Journal of the American Academy of Child Adolescent Psychiatry* **36**(7): 971–979.

Dawid, A. P. & Skene, A. M. (1979). Maximum likelihood estimation of observer error-rates using the EM Algorithm, *Applied Statistics* **28**(1): 20–28.

Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society, Series B* **39**(1): 1–22.

Drews, C. D., Flanders, W. D. & Kosinski, A. S. (1993). Use of two data sources to estimate odds ratios in case-control studies, *Epidemiology* **4**(4): 327–335.

Edelbrock, C., Costello, A. J., Dulcan, M. K., Conover, N. C. & Kala, R. (1986). Parent-child agreement on child psychiatric symptoms assessed via structured interview, *Journal of Child Psychology and Psychiatry* **56**: 265–275.

Edelbrock, C., Costello, A. J., Dulcan, M. K., Kalas, R. & Conover, N. C. (1985). Age differences in the reliability of the psychiatric interview of the child, *Child Development* **56**: 265–275.

Efron, B. & Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*, Chapman & Hall.

Feldman, S. S. & Weinberger, D. A. (1994). Self-restraint as a mediator of family influences on boys' delinquent behavior: a longitudinal study, *Child Development* **65**: 195–211.

Fergusson, D. M. & Horwood, L. J. (1989). A latent class model of smoking experimentation in children, *Journal of Child Psychology and Psychiatry* **30**(5): 761–773.

Fitzmaurice, G. M., Laird, N. M., Zahner, G. E. P. & Daskalakis, C. (1995). Bivariate logistic regression analysis of childhood psychopathology ratings using multiple informants, *American Journal of Epidemiology* **142**(11): 1194–1203.

Forbes, A. B. & Santner, T. J. (1995). Estimators of odds ratio regression parameters in matched case-control studies with covariate measurement error, *Journal of the American Statistical Association* **90**: 1075–1084.

Formann, A. K. & Kohlmann, T. (1996). Latent class analysis in medical research, *Statistical Methods in Medical Research* **5**: 179–211.

Goodman, L. A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models, *Biometrika* **61**(2): 215–231.

Gould, M. S., Fisher, P., Parides, M., Flory, M. & Shaffer, D. (1996). Psychosocial risk factors of child and adolescent completed suicide, *Archives of General Psychiatry* **53**(12): 1155–1162.

Greenland, S. (1980). The effect of missclassification in the presence of covariates, *American Journal of Epidemiology* **112**(4): 564–569.

Haukka, J. K. (1995). Correction for covariate measurement error in generalized linear models – A bootstrap approach, *Biometrics* **51**: 1127–1132.

Holford, T. R. & Stack, C. (1995). Study design for epidemiologic studies with measurement error, *Statistical Methods in Medical Research* **4**: 339–358.

Ibrahim, J. G. (1990). Incomplete data in generalized linear models, *Journal of the American Statistical Association* **85**(411): 765–769.

Jensen, P. S., Traylor, J., Xenakis, S. N. & Davis, H. (1988). Child psychopathology rating scales and interrater agreement: I. parents' gender and psychiatric symptoms, *Journal of the American Academy of Child and Adolescent Psychiatry* **27**(4): 442–450.

Jensen, P. S., Xenakis, S. N., Davis, H. & Degroot, J. (1988). Child psychopathology rating scales and interrater agreement: II. child and family characteristics, *Journal of the American Academy of Child and Adolescent Psychiatry* **27**(4): 451–461.

Kaldor, J. & Clayton, D. (1985). Latent class analysis in chronic disease epidemiology, *Statistics in Medicine* **4**: 327–335.

Kaufman, J., Jones, B., Stieglitz, E., Vitulano, L. & Mannarino, A. P. (1994). The use of multiple informants to assess children's maltreatment experiences, *Journal of Family Violence* **9**(3): 227–248.

Kuha, J. (1994). Corrections for exposure measurement error in logistic regression models with an application to nutritional data, *Statistics in Medicine* **13**: 1135–1148.

Landis, J. R. & Koch, G. G. (1977). The measurement of observer agreement for categorical data, *Biometrics* **33**: 159–174.

Liang, K.-Y. & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models, *Biometrika* **73**: 13–22.

Liu, X. & Liang, K. Y. (1991). Adjustment for non-differential misclassification error in the generalized linear model, *Statistics in Medicine* **10**: 1197–1211.

Loeber, R., Green, S. M., Lahey, B. B. & Stouthamer-Loeber, M. (1989). Optimal informants on childhood disruptive behavior, *Development and Psychopathology* **1**: 317–337.

Madden, P. A. F., Bucholz, K. K., Dinwiddie, S. H., Slutske, W. S., Bierut, L. J., Statham, D. J., Dunne, M. P., Martin, N. G. & Heath, A. C. (1997). Nicotine withdrawal in women, *Addiction* **92**(7): 889–902.

Marshall, J. R. & Graham, S. (1984). Use of dual responses to increase validity of case-control studies, *Journal of Chronic Disease* **37**(2): 125–136.

Murphy, J. M. (1995). Diagnostic schedules and rating scales in adult psychiatry, *Textbook in Psychiatric Epidemiology*, John Wiley & Sons, pp. 253–271.

Offord, D. R., Boyle, M. H., Racine, Y., Szatmari, P., Fleming, J. E., Sanford, M. & Lipman, E. L. (1996). Integrating assessment data from multiple informants, *Journal of the American Academy of Child Adolescent Psychiatry* **35**(8): 1078–1085.

Richardson, R. & Gilks, W. R. (1993). Conditional independence models for epidemiological studies with covariate measurement error, *Statistics in Medicine* **12**: 1703–1722.

Rosner, B., Spiegelman, D. & Willett, W. (1990). Correction of logistic regression relative risk estimates and confidence intervals for measurement error: the case of multiple covariates measured with error, *American Journal of Epidemiology* **132**: 734–745.

Rosner, B., Spiegelman, D. & Willett, W. C. (1992). Correction of logistic regression relative risk estimates and confidence intervals for random within-person measurement error, *American Journal of Epidemiology* **136**: 1400–1413.

SAS Institute (1996). *SAS/Stat Software: Changes and Enhancements through Release 6.11*.

Schuckit, M. A. & Smith, T. L. (1996). An 8-year follow-up of 450 sons of alcoholic and control subjects, *Archives of General Psychiatry* **53**(3): 202–210.

Shaffer, D., Gould, M. S., Fisher, P., Parides, M., Trautman, P., Moreau, D., Kleinman, M. & Flory, M. (1996). Psychiatric diagnosis in child and adolescent suicide, *Archives of General Psychiatry* **53**(4): 339–348.

Verhulst, F. C. (1995). Recent developments in the assessment and diagnosis of child psychopathology, *European Journal of Psychological Assessment* **11**(3): 203–212.

Wachholder, S., Armstrong, B. & Hartage, P. (1993). Validation studies using an alloyed gold standard, *American Journal of Epidemiology* **137**: 1251–1258.

Walter, S. D. (1984). Commentary on use of dual responses to increase validity of case-control studies, *Journal of Chronic Disease* **37**(2): 137–139.

Walter, S. D. & Irwig, L. M. (1988). Estimation of test error rates, disease prevalence and relative risk from misclassified data: a review., *Journal of Clinical Epidemiology* **41**(9): 923–937.

Zahner, G. E. P. & Daskalakis, C. (1997). Factors associated with mental health, general health and school-based service use for psychopathology, *American Journal of Public Health* **87**(9): 1440–1448.

Zahner, G. E. P. & Daskalakis, C. (1998). Modeling sources of informant variance in parent and teacher ratings of child psychopathology, *International Journal of Methods in Psychiatric Research* **7**(1): 3–16.

Zahner, G. E. P., Jacobs, J. H., Freeman, D. H. & Trainor, K. (1993). Rural-urban child psychopathology in a northeastern U.S. state: 1986-1989, *Journal of the American Academy of Child Adolescent Psychiatry* **32**: 378–387.

Zahner, G. E. P., Pawelkiewicz, W., DeFrancesco, J. J. & Adnopoz, J. (1992). Children's mental health service needs and utilization patterns in an urban community, *Journal of the American Academy of Child Adolescent Psychiatry* **31**: 951–960.