



Published in final edited form as:

Psychiatr Ann. 2009 July 1; 39(7): 729–735. doi:10.3928/00485713-20090625-08.

Statistical Approaches to Modeling Multiple Outcomes In Psychiatric Studies

Armando Teixeira-Pinto, Ph.D.¹, Juned Siddique, Dr.PH.², Robert Gibbons, Ph.D.³, and Sharon-Lise Normand, Ph.D.⁴

¹ Department of Biostatistics and Medical Informatics, Faculty of Medicine, CINTESIS, University of Porto, Porto, Portugal

² Department of Health Studies, The University of Chicago, Chicago, U.S.A

³ Center for Health Statistics, University of Illinois at Chicago

⁴ Department of Health Care Policy, Harvard Medical School; Department of Biostatistics, Harvard School of Public Health, Boston, U.S.A

Abstract

Increasingly, multiple outcomes are collected in order to characterize treatment effectiveness or to evaluate risk factors. These outcomes tend to be correlated because they are measuring related quantities in the same individuals. While the analysis of outcomes measured in the same scale (commensurate outcomes) can be undertaken with standard statistical methods, outcomes measured in different scales (non-commensurate outcomes), such as mixed binary and continuous outcomes, present more difficult challenges.

In this paper we contrast some statistical approaches to analyze non-commensurate multiple outcomes. We discuss the advantages of a multivariate method for the analysis of non-commensurate outcomes including situations of missing data. A real data example from a clinical trial, comparing different treatments for depression in low-income women, is used to illustrate the differences between the statistical approaches.

1. Introduction

Multiple outcomes are increasingly collected both in randomized clinical trials and observational studies in order to characterize treatment or intervention effectiveness, or to investigate the association of the outcomes with other variables of interest. The desire to include more than one outcome arises for several reasons including a lack of consensus on the most important clinical outcomes or a desire to demonstrate effectiveness on more than one outcome. The inclusion of multiple outcomes is particularly common in psychiatric studies where disease complexity is often not adequately characterized by a single outcome measure. Depression, for example, is assessed by multiple instruments.

The collection of several outcomes in a study allows different analytical strategies for analysis. The outcomes can be combined into a single composite endpoint using a variety of pooling rules or scoring algorithms. Several types of composite endpoints exist such as taking a simple average of the outcomes or using conjunctive or compensatory rules (see the review by Neuhauser [1]). Another frequently adopted option is to consider each outcome separately by analyzing each independently of the others [2]. However, the situation of multiple outcomes fits perfectly in the framework of several statistical methods designated as multivariate methods, and in particular, multiple informant analyses.[3,4]

Pooling strategies have the disadvantage of reducing the information collected and potentially attenuate important features of the data. Also, any missing observation in one outcome may reduce the sample size if a complete-case analysis is adopted (although we could complement this approach with some sort of imputation technique (ref to the missing data paper in this series) or may produce biased estimates when using available-case analysis even if the missingness is completely at random [5] . Another major drawback of pooling is that it fails when the outcomes are of different natures or are measured on different scales, i.e., non-commensurate outcomes. For example, combining a binary outcome (such as the presence of a symptom) and a continuous outcome (such as a well-being score) requires additional decisions. Often, information is wasted by dichotomizing the continuous outcome so it is “poolable” with the binary outcome.

Analyzing the outcomes separately does not require that the outcomes are commensurate (measured on the same scale) because each outcome is treated as if the other outcomes were not observed. Although the simplicity of such an approach is appealing, the correlation between the outcomes is effectively ignored. This could result in a loss of efficiency in the analysis leading to less power to detect treatment effects (and larger confidence intervals for the estimates). If some outcomes are missing for individuals, separate analyses may produce biased estimates of the covariate effects on the outcomes. Finally, if primary interest is in testing for an overall treatment effect, separate analyses do not provide such an estimate without further work and individual tests for each outcome raises the issue of adjusting the p-values for multiple comparisons.[6]

In this article we present a multivariate method that (1) analyzes all the outcomes at the same time by taking into account their correlations and (2) allows mixtures of different types of outcomes (for example binary and continuous outcomes). Other approaches have been proposed to analyze non-commensurate outcomes in a multivariate framework but with some limitations regarding the settings where they can be applied. [7,8] In section 2 we introduce a real data example that will be used throughout the paper. In section 3 we contrast the finding using individual analysis of the outcomes with those using multivariate methods and interpretation of the results. We conclude in section 4, we conclude with some general recommendation.

2. Treating Depression in Low-Income Women

The WECare Study investigated outcomes during a 12-month period in which 267 low-income mostly minority women in suburban Washington, DC were treated for major depressive disorder [9]. Participants were screened for depression at women, infant, and various pediatric clinics. Subjects were randomly assigned to one of three groups: Medication, Cognitive Behavioral Therapy, and care-as-usual which consisted of a referral to a community provider. The main objective of the primary study was determination of the benefit of medication and cognitive behavioral therapy relative to community referral. Participants were interviewed by phone at baseline, every month for 6 months, and then every other month for the duration of the study. Major clinical outcomes were depression score measured using the Hamilton Depression Rating Scale (HDRS), instrumental role functioning as measured by the Social Adjustment Scale, social functioning (SF) as measured by the Short Form 36-Item Health Survey, and depression remission defined as a HDRS score of 7 or less. Smaller values of the instrumental functioning score and larger values of the social functioning score correspond to better outcomes. Baseline information included age, ethnicity, income, marital status, number of children, health insurance, education, employment, and stressful life events (Table 1).

Outcomes for the first six months of the study were reported in Miranda, J et al. [9]. In this paper, we use depression remission, instrumental role functioning, and social functioning to

demonstrate statistical approaches for assessing the treatment effect on these three outcomes. For illustration purposes, several social functioning scores at six months were deleted to demonstrate problems with conventional methods when data are missing. The primary research question addressed in this paper is whether the Medication and Cognitive Behavioral treatment groups had better depression and functioning outcomes at 6-months as compared to the care-as-usual group (Community Referral).

3. Statistical Methods

3.1 Separate analyses of each outcome

A common approach used when analyzing multiple outcomes is to analyze each outcome separately by regressing each outcome on treatment indicators and additional covariates. In the WECare study, outcomes are adjusted for baseline depression in order to correct for chance initial differences among treatment groups. The regression models depend on the type of the outcome that is being modeled. For example, for continuous outcomes, a linear regression model is typically assumed, while for binary outcomes, logistic or *probit* regression models are used.

For depression remission, we use a probit regression model to estimate the treatment effect adjusted for baseline depression. A probit regression model gives very similar results to the logistic regression. The regression coefficients of the probit model are approximately 1.6 times the coefficients obtained from a logistic regression model. This does not mean that the estimated effects in the probit model are larger than the ones given by the logistic model. In fact they are very similar but measured in different scales. We use a probit model because it allows a direct comparison of treatment effects from a multivariate approach next discussed. For the two functioning scores, we estimate linear regressions models for each outcome with treatment indicators and the respective baseline measurements as covariates.

3.2 Multivariate approach using a latent variable

Rather than modeling each outcome separately, consider a multivariate approach that models the three outcomes in a similar way as the separate models but that additionally takes into account the correlation between the outcomes. Why would an investigator want to adopt this analytical strategy? When the study outcomes have no missing values (or they are missing completely at random), analyzing each outcome separately will provide unbiased estimates for the treatment effects, even if the outcomes are correlated. In this case, the separate models for each outcome will give correct treatment effect estimates but will have larger standard errors than if the correlations among outcomes were taken into account. With sufficiently large sample sizes, investigators may not be concerned so that the tradeoff between simplicity of the analysis procedure and larger errors might favor the simple one-outcome-at-a-time approach.

However what happens in the more common case when data are not missing completely at random? In the WECare data we deleted the social functioning scores of several participants. Women missing these observations have higher instrumental role functioning scores than women with observed scores SF (Table 2). This suggests that women with missing SF are sicker than those with measured SF. Moreover, most women with missing SF belong to the Community Referral arm (71%) and only 8% of the missing SF arises from the Medication arm. Statistical theory tells us that a model for subjects with observed SF will produce biased estimates for the treatment effect. (ref to the missing data paper in this series)

There are several options available to researchers to model multiple outcomes when measured on the same scale, referred to as commensurate outcomes. For example, for normally distributed outcomes we can use a multivariate linear regression or for multiple binary

outcomes we can use a generalized linear mixed model [8,10,11]. With outcomes that are not measured on the same scale, non-commensurate outcomes, there is no simple multivariate distribution to use. The difficulty arises because there is no obvious way to express the multivariate distribution for the mixed type of outcomes.

What can be used instead? The main trick is to include a common unobserved (or latent) variable for all three regression equations. This latent variable establishes the link between the regression equations – the outcomes are measured on the same individuals so the latent variable induces the needed correlations among the outcomes. We assume the latent variable completely specifies the correlation among the outcomes, i.e., given the latent variable the outcomes are assumed to be independent. This permits examination of the outcomes as independent of each other by accounting for the correlation through the latent variable.

The latent variable is assumed to have a normal distribution with mean 0 and some variance, and is scaled (multiplied by a value that has to be estimated) to accommodate the different nature of each outcome. The only restriction regarding the correlation is that the outcomes have to be positively correlated. If some of the correlations are negative they can easily be changed to positive by inverting the outcome scale. This is accomplished by multiplying one of the negatively correlated outcomes by minus one. In the WECare example, the outcome instrumental role functioning is negatively correlated with depression remission and social functioning. We therefore multiply each participant's instrumental role functioning score by minus one, changing the correlation with the remaining outcomes to positive. The covariate effects for instrumental role functioning are afterwards multiplied again by minus one in order for them to be interpreted in the correct scale.

3.3 Interpretation of the regression parameters

The regression equations used in the latent variable approach are *conditional* on the latent variable. For this reason, the regression parameters in the latent variable model also have to be interpreted somewhat differently than when fitting separate models. However, investigators are most interested on the unconditional effects, similar to the usual interpretation in regression models. For the continuous outcomes, the regression parameters can be interpreted in the same manner as those from a separate analyses approach. For the binary outcomes, this is not the case. To compute the treatment effect that is comparable to that obtained from a separate analysis, the regression coefficients are divided by the square root of (one plus the variance of the latent variable). A more detailed discussion about this can be found in Teixeira-Pinto, A and Normand, S-L [12].

In Table 3 the treatment estimates for remission and instrumental role functioning, the outcomes that have no missing data, are virtually identical between the separate regression and the multivariate latent variable approaches. The Table 3 estimates have been transformed and are therefore directly comparable to those from the separate regression models. The coefficients from the probit model for the depression remission outcome can be approximated to odds ratios by taking the exponential of 1.6 times the coefficients. Using the coefficients in Table 3 from the latent variable model, the odds ratios of remission when assigned Medication and Cognitive Behavioral Therapy relative to Community Referral are $\exp(0.36 \times 1.6) = 1.78$ (95% Confidence Interval = (1.11; 4.55)) and $\exp(0.01 \times 1.6) = 1.02$ (95% Confidence Interval = (0.63; 2.56)), respectively.

The estimated treatment effects for social functioning are very different between the two analytical approaches. In separate analyses, the Medication arm has an average increase of 6.99 (se=4.86) points compared to the Community Referral but this effect is not statistically significant (p=0.153). In contrast, the multivariate regression approach using the latent variable yields a statistically significant benefit (p = 0.005) of 13.37 (se=4.70) points in the Medication

arm compared to the Community Referral arm. Why does this occur? Because the outcomes are correlated, information is “borrowed” from the other outcomes through the correlation in order to compensate for the missing outcome. This information is passed through the latent variable and consequently, the estimation of the regression parameters for social functioning should be less biased. For the other outcomes the estimates are identical to the separate analysis because they have complete data. In fact, the correlation between depression remission and the two other outcomes is -0.39 and 0.36 for instrumental role function and social function, respectively, and the correlation between these last two is -0.40 . The effect in the Cognitive Behavioral Therapy arm is not statistically significant in both approaches but the effect estimate obtained with the latent variable model is almost three times larger than the estimate obtained with the separate regressions (7.67 ($se = 4.73$) versus 2.74 ($se = 4.94$) points).

Because we deleted the missing data, we have the complete data and are able to calculate the bias of both approaches by fitting the model for the complete data. The treatment estimates obtained with the complete data for the social functioning are 12.07 for the Medication arm and 5.13 for the Cognitive Behavioral Therapy arm, clearly closer to the results obtained with the latent variable model.

4. Concluding Remarks

Historically, the use of multivariate methods for the general linear model required complete data and an assumption that the multiple outcomes of interest had jointly normal distributions (i.e., multivariate normality). While this restricted their application to continuous and normally distributed outcomes, the advantage of the multivariate approach was that it provided more parsimonious hypothesis tests and interval estimates than a series of univariate tests. With the development of generalized non-linear mixed-effects regression models that can accommodate missing data under fairly general statistical assumptions, the advantages of the multivariate approach can now be extended to reduction in bias produced by missing data, and as illustrated here, these methods can be applied to outcomes measured on different scales, including a mixture of discrete and continuous outcomes.

We have presented a multivariate strategy based on latent variable model to model mixed types of outcomes. This model is an alternative to analyzing each outcome separately which disregards the potential correlation between the outcomes and to pooling information into a composite endpoint, which loses information contained in the data. The dataset used as an example illustrates the advantage of the latent variable model in a situation when one of the outcomes is missing for some subjects and when the missing mechanism is not completely at random. In such cases, modeling the outcomes separately may produce biased estimates for regression parameters.

The latent variable approach has some disadvantages. While the model is not implemented in commercial software, a simple program can be written to provide estimates (see the Appendix). The latent variable model makes some assumptions that are not easily verifiable such as the assumption that the latent variable arises from a normal distribution. Because of the increasing frequency of multiple outcomes reported in psychiatric studies, however, the potential benefits in terms of increases in precision of estimation and power of testing by adopting a multivariate approach are extremely promising.

Acknowledgments

Dr. Normand's effort was supported by Grant MH54693 from the National Institute of Mental Health. The WeCare data were generously provided through the efforts of Dr. Jeanne Miranda.

References

1. Neuhauser M. How to deal with multiple endpoints in clinical trials. *Fundamental & Clinical Pharmacology* 2006;20:515–523. [PubMed: 17109645]
2. Pocock SJ, Geller NL, Tsiatis AA. The analysis of multiple endpoints in clinical trials. *Biometrics* 1987;43(3):487–498. [PubMed: 3663814]
3. Deskalakis C, Laird NM, Murphy JM. Regression analysis of multiple-source longitudinal outcomes: a “Stirling County” depression study. *American Journal of Epidemiology* 2002;155(1):88–94. [PubMed: 11772789]
4. Horton NJ, Fitzmaurice GM. Regression analysis of multiple source and multiple informant data from complex survey samples. *Stat Med* 2004;23(18):2911–2933. [PubMed: 15344194]
5. Li X, Caffo B, Scharfstein D. On the potential for illogic with logically defined outcomes. *Biostat* 2007;8(4):800–804.
6. Bland JM, Altman DG. Statistics notes: Multiple significance tests: the Bonferroni method. *BMJ* 1995;310(6973):170. [PubMed: 7833759]
7. Catalano PJ. Bivariate modelling of clustered continuous and ordered categorical outcomes. *Stat Med* 1997;16(8):883–900. [PubMed: 9160486]
8. Fitzmaurice GM, Laird NM. Regression models for mixed discrete and continuous responses with potentially missing values. *Biometrics* 1997;53(1):110–122. [PubMed: 9147588]
9. Miranda J, Chung JY, Green BL, Krupnick J, Siddique J, Revicki DA, Belin T. Treating depression in predominantly low-income young minority women: a randomized controlled trial. *JAMA* 2003;290(1):57–65. [PubMed: 12837712]
10. Liang K-Y, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986;73(1):9–22.
11. Hedeker, D.; Gibbons, RD. Longitudinal data analysis. New Jersey: John Wiley & Sons; 2006.
12. Teixeira-Pinto A, Normand S-L. Correlated Bivariate Continuous and Binary Outcomes: Issues and Applications. 2008

6. Appendix

```
/* SAS code for using the PROC NLMIXED to fit the latent model for the WeCare
data
```

```
Abbreviations:
```

```
dr - depression remission, irf - instrumental role functioning, sf - social
functioning
HRDSbline - Hamilton score at baseline, irfbline- IRF at baseline, sfbline -
SF at baseline
cbt - Cognitive Behavioral Therapy arm, medic - Medication arm
tau - latent variable
```

```
*/
```

```
PROC NLMIXED data = wecare_data_withmiss GCONV=1E-15;
```

```
parms a_dr=4 b_dr=3 c_dr=3 f_dr=0.03
      a_irf=1.7 b_irf=.76 c_irf=.87 f_irf=.21 sigma_irf =1
      a_sf=67 b_sf =-5 c_sf =-12 f_sf =0.25 sigma_sf =12
      sigmatau=1; *initial estimates obtained from the separated
regressions;
      bounds sigma_irf>0, sigma_sf>0, sigmatau>0.01; *constraining the std
deviations to be positive;
```

```

        stdconst = sqrt(1+sigmatau**2); *constant used in the probit model to
        obtain the marginal
        effects;

        *construction of the likelihood;
        p = a_dr*stdconst + b_dr*stdconst*cbt + c_dr*stdconst* medic +
f_dr*stdconst*HRDSbline + tau;
        mean_irf = a_irf + b_irf*cbt + c_irf*medic + f_irf*irfbline +
sigma_irf*tau;
        mean_sf = a_sf + b_sf* cbt + c_sf* medic + f_sf*sfbline + sigma_sf*tau;
        ll_dr = dr*log(PROBNORM(p)) + (1-dr)*log(PROBNORM (-p));
        ll_irf = -.5*((irf - mean_irf)/sigma_irf)**2 -log(sigma_irf);
        ll_sf = -.5*((sf-mean_sf)/sigma_sf)**2-log(sigma_sf);
        ll = ll_dr+ll_irf+ll_sf; *log likelihood for individual with complete
        observations;
        if missing (sf) then ll=ll1+ll2; *log likelihood for individual with
        missing SF;

        *model;
        model medic ~ general(ll);
        random tau ~ normal(0, sigmatau**2) subject=MID; *latent variable;
run;

```

Table 1

Social and demographic characteristics and baseline measurements of depression scores by treatment arm, for women enrolled in the WEcare study.

	Total (n=267)	Medication (n=88)	Cognitive Behavioral Therapy (n=90)	Community referral (n=89)
	<i>mean (SD)</i>			
Age	29.3 (7.9)	28.7 (6.6)	29.8 (7.9)	29.5 (9.1)
Number of children	2.3 (1.4)	2.2 (1.2)	2.2 (1.5)	2.4 (1.6)
Baseline for Social Functioning	57.7 (25.3)	56.5 (24.6)	56.5 (23.9)	60.0 (27.5)
Baseline for Instrumental Role Functioning	3.5 (1.2)	3.6 (1.3)	3.5 (1.2)	3.3 (1.2)
Baseline Hamilton Score	16.9 (5.2)	17.9 (5.1)	16.3 (5.1)	16.5 (5.2)
	<i>n (%)</i>			
Employment				
<i>Working or looking for work</i>	219 (82.0)	69 (78.4)	76 (84.4)	74 (83.2)
<i>Not working or disabled</i>	48 (18.0)	19 (21.6)	14 (15.6)	15 (16.9)
Education				
<i>Less than high school</i>	99 (37)	37 (42)	27(30)	35 (39)
<i>High school</i>	87 (33)	31 (35)	29 (32)	27 (30)
<i>Some trade or college</i>	63 (26)	15 (17)	26 (29)	22 (25)
<i>College graduate</i>	18 (7)	5 (6)	8 (9)	5 (6)
Marital Status				
<i>Married or living with partner</i>	124 (46)	43 (49)	40 (44)	41 (46)
<i>Widowed or separated</i>	52 (20)	17 (19)	22 (24)	13 (15)
<i>Never married</i>	91 (34)	28 (32)	28 (31)	35 (39)
Ethnicity				
<i>Black</i>	117 (44)	34 (39)	41 (46)	42(47)
<i>White</i>	16 (6)	6 (7)	6 (7)	4 (4)
<i>Latina</i>	134 (50)	48 (55)	43 (48)	43 (48)
Schooling				
<i>Less than high school</i>	99 (37)	37 (42)	27 (30)	35 (39)
<i>High school or GED</i>	87 (33)	31 (35)	29 (32)	27 (30)
<i>Some trade or college</i>	63 (23)	15 (17)	26 (29)	22 (25)
<i>College graduate</i>	18 (7)	5 (6)	8 (9)	5 (6)

SD= Standard deviation

Table 2

Comparison of Instrumental Role Functioning, Remission, and Treatment Arms Stratified by Missingness of 6-Month Social Functioning Score.

Characteristic	Social Functioning missing (n=83)	Social Functioning observed (n=184)	p-value
Instrumental Role Functioning, <i>mean (SD)</i>	4.0 (1.5)	2.3 (1.2)	<0.001
Depression Remission (HDRS of 7 or less), <i>n (%)</i>	33 (40)	74 (40)	0.944
Treatment, <i>n (%)</i>			
Medication	7 (8)	81 (44)	<0.001
Cognitive Behavioral Therapy	17 (21)	73 (40)	
Community Referral	59 (71)	30 (36)	

SD = Standard deviation; HDRS = Hamilton depression rating scale

Table 3

Regression coefficients associated with the cognitive behavioral therapy, medication and community referral (adjusted for the respective baselines) for the three scores: Remission, Instrumental Role Functioning, and Social Functioning. Comparison between the approach using separate regressions for each outcome and the multivariate approach using a latent variable model.

6-Month Outcome	Separate Regressions			Multivariate Latent variable model		
	Coef.	(SE)	p-value	Coef.	(SE)	p-value
Depression Remission (HDRS of 7 or less)						
	N = 267 participants					
Hamilton score at baseline	-0.03	(0.02)	0.551	-0.02	(0.01)	0.146
Cognitive Behavioral therapy	-0.01	(0.19)	0.944	0.01	(0.19)	0.960
Medication	0.38	(0.19)	0.050	0.36	(0.19)	0.062
Community Referral				Reference Group		
Instrumental Role Functioning						
	N = 267 participants					
Instrumental Role Functioning at baseline	0.21	(0.07)	0.003	0.11	(0.06)	0.080
Cognitive Behavioral therapy	-0.11	(0.22)	0.606	-0.09	(0.22)	0.691
Medication	-0.87	(0.22)	<0.001	-0.84	(0.22)	<0.001
Community Referral				Reference Group		
<i>SD of the error term</i>	1.42	(0.06)	<0.001	1.04	(0.06)	<0.001
Social Functioning						
	N = 184 participants					
Social Functioning at baseline	0.26	(0.07)	<0.001	0.27	(0.06)	<0.001
Cognitive Behavioral therapy	2.74	(4.94)	0.581	7.67	(4.73)	0.107
Medication	6.99	(4.86)	0.153	13.37	(4.70)	0.005
Community Referral				Reference Group		
<i>SD of the error term</i>	22.48	(1.17)	<0.001	16.94	(0.93)	<0.001
<i>SD of the latent variable</i>				0.94	(0.09)	<0.001

SE = Standard error; SD = Standard deviation; HDRS = Hamilton depression rating scale.