Lessons from other disciplines: Analysis of data recorded from multiple informants

Nicholas Horton (Amherst College, nhorton@amherst.edu)
May 17, 2021
https://nhorton.people.amherst.edu/multinform_2021_05.pdf

Acknowledgements

- joint work with:
  - Nan Laird
  - Garrett Fitzmaurice
  - Sharon-Lise Normand
  - Constantine Daskalakis
  - Jane Murphy
  - Stuart Lipsitz
  - Frank Yoon
  - Heather Litman
  - Kate Aloisio
  - Kevin Roberts
  - Louise Ryan
- support from NIH R01MH054693
- see also references and https://web.archive.org/web/20151215212838/http://www.math.smith.edu/multinform

Intro to the analysis of multiple informant reports

- multiple informant reports are often collected in studies of children (because they are not reliable reporters)
- inherent feature of multiple informants is that we expect discordance (due to considerable measurement error)
- if no discordance, no need to collect the (redundant) information
- sources of discordance
  - measurement error
  - each informant is tapping into different aspects of the underlying construct
- missing data is common (since measures are being taken on others beyond those being directly studied
- “job security for statisticians”
- useful analogies for nutrition and supplement assessment (dietary intake)

Why not pool reports?

1. the optimal algorithm for combining reports depends upon the type of measurement error present;
2. pooling results in loss of information about source-specific effects;
3. pooling does not permit the examination of differences in risk-factor effects across sources; and
4. many pooling algorithms are not clearly defined in the presence of missing data.

Why not analyze separately?

1. separate analyses yield multiple (and often differing) sets of results for the different sources (difficult to interpret);
2. separate analyses provide no formal means of evaluating how similar or different the results are across the various sources (or to summarize them in a single set of results, if they are sufficiently similar); and
3. separate analyses may be based on different subsets of the data, if some subjects are missing one source and others are missing another.
4. neither pooling strategies nor separate analyses address the potential bias resulting from missing data in this setting.

Unified regression model

1. Fitzmaurice et al, *AJE* (1995) and Horton and Fitzmaurice, *SIM* (2004) describe a flexible model that incorporates all multiple-source reports in a single multivariate regression analysis;
2. Can test for source differences in outcome, and estimation of different source effects where necessary;
3. Can test if the effects of other risk factors on the outcome differ by source; and
4. Includes partially observed data from subjects with missing source observations.

Multiple informants as predictors

1. Horton and Fitzmaurice (SIM, 2004) also describe a flexible model that incorporates all multiple-source reports as predictors of an outcome not measured by multiple informants
2. The model can test for source differences in association with the outcome
3. The model can incorporate estimation of different source effects where necessary

Application: Child psychopathology

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

Application: Assessing linkage to primary care (EHR)

Table 1. Missingness of self-report and administrative report of linkage.

	Self report		
Admin. report	Observed	Missing	Total
Observed	307	140	447
Missing	10	13	23
Total	317	153	470

Horton, Saitz, Laird, and Samet (*Health Services & Outcomes Research Methodology*, 2002)

Application: Multiple measures of depression

- Siddique et al "Multiple imputation for harmonizing longitudinal non-commensurate measures in individual participant data meta-analysis" (*SIM*, 2015)
    - Some studies used the Children's Depression Rating Scale (CDRS)
    - Others used the Hamilton Depression Rating Scale (HDRS)
    - None used both
    - Are there ways to harmonize and map between measures and domains?

- Horton et al "Multiple Informants: Mortality associated with psychiatric disorders in the Stirling County Study" (*AJE*, 2001)
    - Changing measures of depression over time in a longitudinal study
    - How to calibrate back and forth from DPAX1 to DPAX2

Latent variable modeling (Horton et al, *SIM* 2008)

Multiple informant reports of exposure to violence

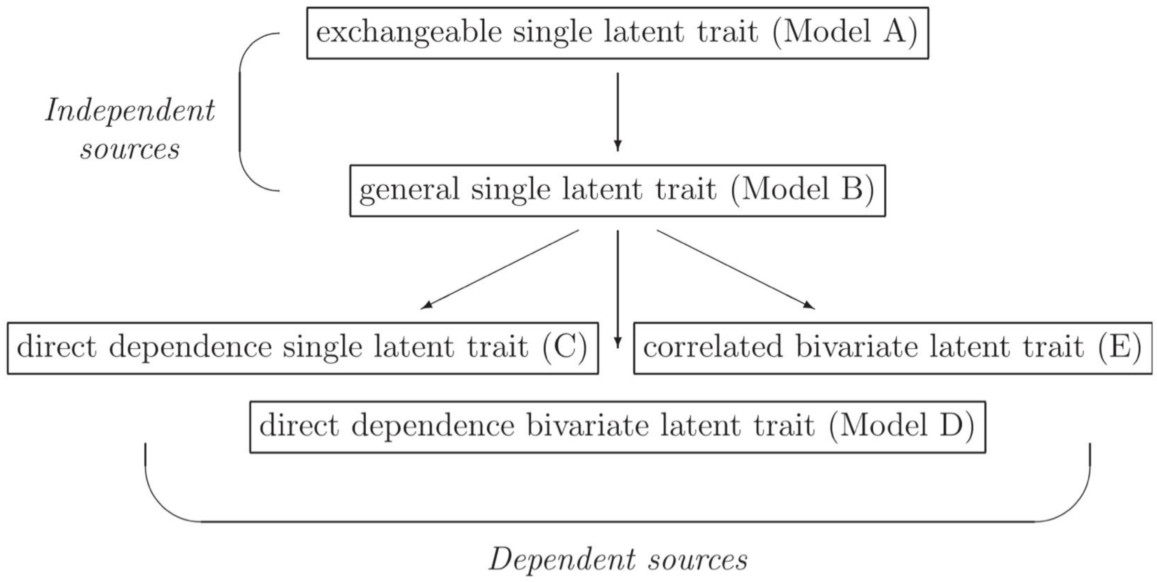

Figure 1. Nesting of models.

Application: Identifying triggers for HIV testing

Examination by Specific HIV Triggers in Medical Encounters Between 1994 and 2001 Where HIV Testing Was Recommended or Considered by the Clinician

HIV Trigger	% (Number of visits with HIV testing recommended or considered/total number of visits with triggers)
Men sex with men	71 (32/45)
Weight loss	68 (54/80)
Injection drug use	54 (91/167)
Hepatitis B and/or C	50 (51/103)
Community acquired pneumonia	50 (66/132)
Sexually transmitted disease	46 (100/217)
Crack/cocaine	42 (98/235)
Alcoholism/alcohol abuse	35 (49/139)
Homelessness	34 (57/167)
Herpes Zoster	21 (19/90)

Liddicoat et al "Assessing Missed Opportunities for HIV Testing in Medical Settings" (JGIM, 2004)

Assessed individual association of indicator variables to predict HIV testing

Back to assessment of nutrition and supplements

Accounting for missing data (MAR) in creative and efficient ways

- Rescue question
- Combining parallel measures (when appropriate)
- Food photography
- Incorporate direct observation on a very small subset
- Matrix sampling (to minimize respondent burdens)

Data fusion opportunities for non-commensurate measures

- Food intake vs. supplement intake
- 24 hour recall vs. food diary
- Addressing respondent bias and systematic sampling limitations
- Incorporating found (observational) data
- Future opportunities: metabolomics
- Lessons from MEMS (medication event monitoring system) caps

Open questions

What is the question of interest?

What quantity are we trying to measure?

What needs to be measured?

How well do we want to measure? How much effort/resources to assess it?

Takehome message: collecting and analyzing multiple informant measures can help to augment primary assessment and deepen understanding of nutrition and supplement use

Lessons from other disciplines: Analysis of data recorded from multiple informants

Nicholas Horton (Amherst College, nhorton@amherst.edu)
May 17, 2021
https://nhorton.people.amherst.edu/multinform_2021_05.pdf

References

Fitzmaurice et al. “Bivariate logistic regression analysis of childhood psychopathology ratings using multiple informants”, 1995, AJE, 142(11):1194-203.

Horton NJ and Fitzmaurice GM. “Regression analysis of multiple source and multiple informant data from complex survey samples”. Statistics in Medicine, 2004; 23(18):2911-2933.

Horton NJ, Saitz R, Laird NM and Samet JH. “A method for modeling utilization data from multiple sources: application in a study of linkage to primary care.” Health Services and Outcomes Research Methodology, 2002; 3(3):211-223.

Horton NJ, Laird NM, Murphy JM, Monson RR, Sobol AM, and Leighton AH. “Multiple informants: mortality associated with psychiatric disorders in the Stirling County Study”. American Journal of Epidemiology, 2001; 154(7):649-656.

Horton NJ, Roberts K, Ryan L, Suglia S, and Wright RJ. “A maximum-likelihood latent variable model for multiple informants”. Statistics in Medicine, 2008; 27:4992-5004.

Liddicoat RV, Horton NJ, Urban R, Maier E, Christiansen D, and Samet JH. “Assessing missed opportunities for HIV testing in medical settings.” Journal of General Internal Medicine, 2004; 19(4):349-356.

Siddique et al “Multiple imputation for harmonizing longitudinal non-commensurate measures in individual participant data meta-analysis” 2015, SIM, 34(26):3399-3414.