



Designed missingness to better estimate efficacy of behavioral studies-application to suicide prevention trials

Ofer Harel^{1*}, Jeffrey Stratton¹ and Robert Aseltine²

*Correspondence: ofer.harel@uconn.edu



CrossMark

← Click for updates

¹Department of Statistics, University of Connecticut, 215 Glenbrook Rd. Unit 4120, Storrs, CT 06269, USA.

²Institute for Public Health Research, University of Connecticut Health Center, 263 Farmington Avenue, MC 3910, Farmington, CT 06030-3910, USA.

Abstract

Randomized trials of diverse behavioral interventions routinely observe declines in problem behavior among control subjects that cannot be attributed to flawed experimental design. Explanations for this pattern of effects generally focus on assessment reactivity. The research activities and procedures themselves may constitute an intervention of sorts, and control subjects may be better characterized as an “intervention lite” group rather than an untreated control group. Such conditions may lead to serious underestimates of the efficacy of behavioral interventions. One possible remedy for this problem is the use of “designed missingness” for collecting baseline or pretest data. This strategy intentionally collects data on only a subset of indicators and uses imputation techniques to address the resulting structured missingness. In this study, pretest questionnaires used in the evaluation of a large suicide prevention intervention were modified to reduce the amount of behavioral information collected at baseline among control subjects. Four versions of the pretest questionnaire were used: one full version and three truncated versions, each of which included a different subset of items from the full version. The completion rate increased from 12% for pretests to 19% for posttests. Additionally, treatment effects after imputation were larger for subjects who were assigned the truncated versions of the pretest than for subjects who were assigned the nontruncated questionnaire at pretest. Although more research is needed on this subject to establish optimal questionnaire configurations and study designs, “designed missingness” methods have the potential to improve the assessment of treatment effects in a broad range of efficacy studies.

MeSH Keywords: Questionnaires, statistical models

Keywords: Missing data, multiple imputation, questionnaire design, statistical models, suicide, randomized controlled trial

Introduction

Randomized trials of diverse behavioral interventions routinely observe declines in problem behavior among *control* subjects that cannot be attributed to flawed experimental design (e.g., contamination). For example, in nearly a dozen separate studies of risky drinking among adults reviewed by the U.S. Preventive Services Task Force the average decline in drinking from baseline to follow-up among treatment subjects was 28%, but was 16% for control subjects [1]. Explanations for this pattern of effects generally focus on *assessment reactivity*, which refers to changes in behavior that result from exposure to either intensive assessment protocols used to identify subjects for inclusion in the research study or routine baseline research assessments in a pretest-posttest control group design [2]. In

other words, the research activities and procedures themselves may constitute an intervention of sorts, and control subjects in this context may be better characterized as an “intervention lite” group as opposed to an untreated control group. Such conditions may lead to serious underestimates of the efficacy of behavioral interventions. An additional complication with long baseline assessments is incomplete data. Respondents may tire of answering questions and may not answer certain questions (item nonresponse) or may stop answering questions altogether (unit nonresponse).

One possible remedy for these problems is the use of “designed missingness” in the collection of baseline or pretest data. This strategy, which intentionally collects data on only a subset of cases and/or indicators and uses imputation techniques

to address the resulting structured missingness, has been employed to increase the efficiency and cost-effectiveness of data collection in large-scale epidemiologic studies [3]. Several different designed missingness strategies have been proposed in the past. Multiple matrix sampling asks every sampled unit a random sample of questionnaire items [4,5]. Split questionnaire designs, which restrict the assignment of blocks of questionnaire items to subgroups in the sample, constitute a similar approach [6]. Both these alternatives have historically been used to reduce respondent burden, but Chipperfield and Steel [7] have shown that such designs are more efficient in terms of a survey's cost assuming that constraints on the accuracy of specific estimates are met. Also, the correlation between items can be exploited to minimize the loss of information that results from not collecting complete observations from all individuals in the sample [7].

In the current study, we employed a designed missingness strategy for two reasons. The first reason was to mitigate the potential for assessment reactivity. We hypothesized that by reducing number of questions in the baseline assessment, potential changes in thoughts or behavior that are stimulated by exposure to the issues raised in this assessment would be minimized. The second reason for using this designed missingness strategy was to reduce the amount of missing data. We hypothesized that as the pretest questionnaire becomes shorter, respondent fatigue is reduced, improving response rates and reducing incomplete data.

Our study used multiple imputation to account for the missing data produced by the split questionnaire design. A split questionnaire design using multiple imputation was introduced by Raghunathan and Grizzle [6]. They split the questionnaire into different versions based on partial correlations of questionnaire items from a pilot study. Our study differed in that the questionnaire items were split to reduce assessment reactivity, and only a subset of the questionnaire items were split randomly into three questionnaire versions.

This paper begins with the presentation of our motivating example for this study. The next two sections present the multiple imputation method used and our results. The final section discusses our findings.

Methods

Motivating example

This strategy was implemented as part of the Connecticut Youth Suicide Prevention Initiative conducted from 2006–2009 by the Connecticut Department of Mental Health and Addiction Services and the University of Connecticut Health Center.

The target population of this study was the Connecticut Technical High School System (CTHSS) and Trumbull High School/Regional Agriscience and Biotechnology Program's ninth-grade classrooms [8]. Seventeen (CT) schools were included in the intervention, which featured the "Signs of Suicide" (SOS) prevention program, a brief school-based suicide prevention program produced by Screening for Mental

Health, Inc. The program was selected because it is the only school-based program to show a reduction in self-reported suicide attempts (by 40%) within three months following program completion in a randomized controlled study [9].

This study was completed in two waves during the 2007-2008 and 2008-2009 academic years. The program was presented in schools in the treatment group from November through January; during the same period, students in the control group completed the pretest questionnaires but did not participate in the program. The study used a randomized pretest-posttest experimental design, with outcomes assessed at baseline and at three months post-intervention using anonymous questionnaires administered during class. Four versions of the pretest questionnaire were used: one full version and three truncated versions, each of which included a different subset of items from the full version. The posttest questionnaire was the same for all subjects. Following post-test data collection, the program was presented to schools in the control group. Twelve of the participating schools completed the full version of the pretest questionnaire; in the other five schools, classes were randomly selected to receive one of the four versions of the questionnaire. **Table 1** shows the number of questionnaires administered to the schools in this study.

Prior to program presentation, all eligible students were given a permission slip (consent form) to be completed by their parents. Only students with signed permission slips were included in the evaluation. In order to encourage the

Table 1. Questionnaire totals for schools.

School	Version				Total
	0	1	2	3	
1	55	0	0	0	55
2	53	0	0	0	53
4	15	0	0	0	15
5	68	0	0	0	68
6	33	0	0	0	33
8	105	0	0	0	105
9	58	0	0	0	58
10	47	16	12	14	89
18	13	38	29	0	80
19	0	8	52	0	60
21	79	0	0	0	79
22	108	0	0	0	108
23	0	0	0	64	64
24	93	0	0	0	93
25	107	0	0	0	107
26	148	0	0	0	148
27	0	52	0	0	52
28	0	0	22	0	22
Total	982	114	115	78	1289

return of permission slips, each returned slip signed by the parent/guardian and child was entered in a drawing for an American Express gift card or a portable DVD player. Entry in the drawing was independent of research participation. Gift cards were distributed at the conclusion of the pretest data collection at each school. The drawing for the two DVD players occurred following the completion of the post-test data collection.

Table 2 specifies the questions asked of students on each of the four pretests. The truncated versions were modified to reduce the amount of behavioral information collected at baseline for control subjects. The questionnaires collected information about attitudes related to suicide as well as relationships with others and some demographic information. The first 12 items were related to attitudes about suicide, and it is this set of items that differed among the truncated versions. The remaining questions (Q13-Q28) remained the same for all four versions. The first 12 items were allocated randomly among the three truncated questionnaires.

The SOS instrument measured students' attitudes and knowledge about suicide. Attitude was measured with a 10 item scale, and knowledge with a 7 item scale [9]. The survey also measured help-seeking in the past three months:

- had they sought treatment (Y/N).
- had they talked to parent/guardian, sibling, teacher/guidance counselor, other adult, hotline (Y/N), and
- had they talked to adult about friend (Y/N).

Suicidal behavior in the past 3 months was also measured:

thoughts about suicide or ideation (Y/N), making a plan (Y/N), and attempting suicide (Y/N). The baseline study included 1,291 students. The sample was 58% male and 42% female. Ten percent of respondents spoke English as a second language. The students self-identified their race/ethnicity as: White non-Hispanic (60%), Black non-Hispanic (6%), Hispanic (23%), Multi-ethnic (9%), and other (2%).

Multiple imputation

The three truncated pretest questionnaires introduced "designed missingness" into our data set. Any statistical analysis would be complicated by the different patterns of missing responses across the three truncated versions. We used multiple imputation (MI) to overcome this difficulty; see e.g., [10,11]. We created m multiple complete data sets, filling in the missing observations in a principled way. Multiple imputation incorporates the uncertainty due to the missing data in the imputation process [12].

We performed a complete-data analysis on the m different data sets, and then combined the results using rules defined by Rubin [11]. The variability due to the missingness was added into the estimated variability of the statistic of interest. A good summary of multiple imputation as well as software that is available is provided by Harel and Zhou [12]. We used a SAS macro called *IVEware* [13]. This macro produced imputed values for each individual in the data set, conditional on all the values observed for that individual. Imputations were created using a sequence of multiple regressions, varying the type

Table 2. Questionnaire items by version.

Item	Versions	Content	Valid Values
Q1	0 1 -- --	People who talk about suicide don't really kill themselves.	Yes/No (1/5)
Q2	0 1 -- --	People who commit suicide are usually suffering from depression or some other mental illness.	Yes/No (1/5)
Q3	0 -- -- 3	Most suicide attempts occur without any warning signs or clues.	Yes/No (1/5)
Q4	0 1 -- --	Depression is an illness that doctors can treat.	Yes/No (1/5)
Q5	0 -- -- 3	The best thing to tell a suicidal friend is to "pull yourself together and things will get better."	Yes/No (1/5)
Q6	0 -- 2 --	If I talk to someone about their suicidal feelings, it may cause them to commit suicide.	Yes/No (1/5)
Q7	0 -- 2 --	Alcohol use is not related to suicidal behavior.	Yes/No (1/5)
Q8	0 -- 2 --	Sometimes young people have so many personal problems they have no other options besides suicide.	Likert (1/2/3/4/5)
Q9	0 1 -- --	If someone really wants to kill himself/herself, there is not much anyone can do about it.	Likert (1/2/3/4/5)
Q10	0 -- 2 --	It's none of my business if a friend says he/she wants to kill himself/herself.	Likert (1/2/3/4/5)
Q11	0 -- 2 --	If I were feeling really down, I would try to talk to a counselor or some other adult about my problems.	Likert (1/2/3/4/5)
Q12A	0 1 -- 3	If a friend told me ...: I wouldn't know what to do.	Likert (1/2/3/4/5)
Q12B	0 1 -- 3	If a friend told me ...: I would keep it to myself.	Likert (1/2/3/4/5)
Q12C	0 -- 2 3	If a friend told me ...: I would wish that I had not found out about it.	Likert (1/2/3/4/5)
Q12D	0 -- -- 3	If a friend told me ...: I would keep it a secret if my friend made me promise not to tell.	Likert (1/2/3/4/5)
Q12E	0 -- -- 3	If a friend told me ...: I would tell an adult at school about it.	Likert (1/2/3/4/5)
Q12F	0 -- -- 3	If a friend told me ...: I would tell a parent or some other adult outside of school about it.	Likert (1/2/3/4/5)

of regression model by the type of variable being imputed. This study used 100 multiple imputations based on a recommendation by Harel [14]. We analyzed the influence of pretest version on survey response for both the complete data and using multiple imputation. More details about our multiple imputation methodology can be found in the **Appendix 1**.

Results

The unit nonresponse rates and completion rates by version are summarized in **Table 3**. The non-truncated questionnaire (version 0) was complete if a respondent responded to all the non-demographic questionnaire items (Q1-Q28). The truncated questionnaires differed from version 0 by containing a subset of questionnaire items Q1-Q12. They were complete if respondents responded to those items relevant to their pretest version (**Table 2**) as well as the remaining items Q13-Q28. Cases were considered “unit nonresponses” if they did not even attempt the questionnaire; they were incomplete if the survey was attempted but some questions were left blank. As indicated by the contrast between completion rates for version 0 versus all other versions, we found that the response rate was larger for the truncated versions of the questionnaire compared to the full version. Results in **Table 3** indicate that the completion rate for the truncated versions was around 85% while the completion rate for the non-truncated version was 73% in pretest, and the posttest completion rate was 59% for the non-truncated version compared to 78% for the truncated versions.

In addition to the effects of version on completion rates, we also examined whether truncated pretest versions might reduce the potential for assessment reactivity. This would be indicated if version was predictive of knowledge of and attitudes toward suicide assessed at posttest. To address this question we regressed each knowledge and attitude item on version, controlling for race, gender, reduced lunch status (Lunch), grade point average (GPA), and mother’s education level (MomEd). We also included a variable indicating whether a student received the SOS intervention or not

(Treatment). We modeled an individual student’s response to a given variable using:

$$Q_i = \beta_0 + \beta_1 \text{Version}_i + \beta_2 \text{Race}_i + \beta_3 \text{Gender}_i + \beta_4 \text{Lunch}_i + \beta_5 \text{GPA}_i + \beta_6 \text{MomEd}_i + \beta_7 \text{Treatment}_i.$$

The dependent variable was the questionnaire item itself. Items Q1-Q7 were dichotomous responses, requiring the use of logistic regression. The remaining items Q8-Q12 were Likert scale response questions. We used a generalized logit model for questions Q8-Q12. We also chose to include the lunch status, race, and mother’s education status because these covariates had high rates of missingness. They were included to see if they were significant in the multiple imputation modeling. Gender and GPA were also included because differences in students’ attitudes toward suicide due to them is common. We assessed the quality of fit for the logistic regressions using the Hosmer-Lemeshow test, and the fit for the generalized logit fits was assessed using the proportional odds test.

Our analysis contrasted the results of complete case analysis and multiple imputation (MI) to determine if our split questionnaire/MI methodology reduced the potential for assessment reactivity. The regression coefficients for the test version along with their *p*-values for each item are presented in **Table 4**. Out of 17 questions, the MI analysis yielded five questions with a significant (*p*-value <0.10) version coefficient. The complete case analysis was similar. Questions 3, 6, 7, 9, and 12F had

Table 3. Unit nonresponse and completion rates by version.

Test	(n)	Version	Unit nonresponse rate	Completion rate
Pretest	1,291	All	4/1291=0.0031	984/1291=0.76
	982	0	4/982=0.0041	720/982=0.73
	114	1	0	99/114=0.87
	117	2	0	101/117=0.86
	78	3	0	64/78=0.82
Posttest	1,291	All	237/1291=0.18	818/1291=0.63
	982	0	212/982=0.22	577/982=0.59
	114	1	7/114=0.06	89/114=0.78
	117	2	16/117=0.14	89/117=0.76
	78	3	2/78=0.03	63/78=0.81

Table 4. Significance of VERSION in regressions.

Item	Versions	PostTest Complete*	PostTest MI†	Result
Q1	0,1	NS†	NS	--
Q2	0,1	NS	NS	--
Q3	0,3	0.40 (0.02)	0.44 (0.03)	More knowledge
Q4	0,1	NS	NS	--
Q5	0,3	NS	NS	--
Q6	0,2	0.42 (0.06)	0.38 (0.07)	More knowledge
Q7	0,2	-0.44 (0.04)	-0.55 (0.01)	Less knowledge
Q8	0,2	NS	NS	--
Q9	0,1	-0.33 (0.06)	-0.38 (0.03)	Better attitude
Q10	0,2	NS	NS	--
Q11	0,2	NS	NS	--
Q12A	0,1,3	NS	NS	--
Q12B	0,1,3	NS	NS	--
Q12C	0,2,3	NS	NS	--
Q12D	0,3	NS	NS	--
Q12E	0,3	NS	NS	--
Q12F	0,3	0.30 (0.06)	0.25 (0.09)	Better attitude

*Numbers presented are the regression coefficients for VERSION. The *p*-value for the coefficient is given in parenthesis.
 †A value of “NS” indicates regressions where VERSION is not significant (*p*-value >0.10).

questionnaire version as a significant predictor. With the exception of question 7, all version coefficients from the MI analysis and complete case analysis indicated that students who had the question in pretest had more knowledge or better attitude in posttest. This pattern of results persisted after controlling for multiple tests, suggesting that version had a significant impact on posttest knowledge and attitudes.

Model checking procedures under multiple imputation are an area of active and ongoing research. With 100 multiple imputations and 17 different questionnaire items, assessing how well the model fit was a difficult undertaking. We randomly selected three imputations (9, 32, and 64) and three questionnaire items (Q4, Q7, and Q12B) and looked at diagnostic plots and measures. We found no reason to reject the logistic model. The Q7 item may have an influential point, but it was evident on both the pre- and posttests, and both before and after imputation. We believe the model fit the data reasonably well.

In addition to regressions for each individual item, we also examined the impact of version on a summary scale comprised of the attitude questionnaire items. The variable SOS average (SOSavg) was a summary measure of student attitudes about suicide consisting of the average score of all or a subset of questionnaire items Q8-Q12F. Scores ranged from one to five for each question. Scores for questions Q11, Q12E, and Q12F were reversed to match the pattern of other questions, with higher scores representing more negative or less favorable attitudes toward suicide. Pretest version 0 subjects received all 10 questions, while subjects of the other pretest versions received a subset of those questions (See [Table 2](#) for details).

Figure 1 compares the mean SOS average for individuals

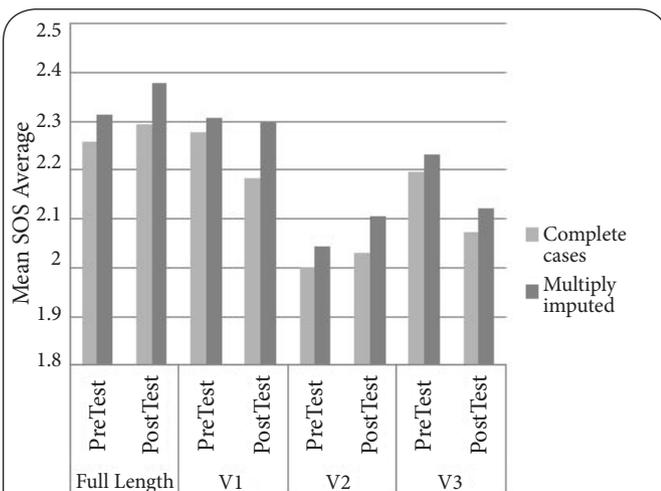


Figure 1. Mean SOS average by version and test. This bar chart shows the mean SOS average score for the pretest and posttest for each of the four types of questionnaires. The light gray bars represent scores for only the complete cases observed, and the dark gray bars show the multiply imputed scores. The left axis runs from 1.8 to 2.5 in increments of 0.1 and is the average of 10 questionnaire items that had values ranging from one to five.

who completed the questionnaires to the mean SOS average using multiple imputation. This is broken down further by questionnaire version. The results in [Figure 1](#) indicate that the mean SOS average produced by multiple imputation was slightly higher than that of complete responses regardless of questionnaire version. Attitudes were less favorable using multiple imputation. This could happen if nonrespondents to the questionnaires tended to be people with more negative attitudes about suicide.

For the full-length questionnaire, posttest scores showed an increase in SOS average, meaning an increase in negativity with regard to suicide attitudes. For version 1 of the questionnaire, the complete respondents showed a slight decrease between pretest and posttest SOS average. The multiple imputation yielded a posttest SOS average more similar to that of the pretest. For version 2, both the complete case results and multiple imputation results showed an increase from pretest to posttest. Both the complete responses and multiple imputation responses showed a decrease in negativity between the pretest and posttest for version 3.

To summarize, treatment effects differed between the questionnaire versions. Versions 2 and 3 resulted in mean SOS averages that were lower (more favorable) than that of the full length questionnaire. However, version 2 showed an increase in mean SOS average from the pretest to the posttest, while version 3 showed a *decrease* in SOS average from pretest to posttest. The full-length questionnaire group showed an increase from pretest to posttest. Version 1 showed a slight decrease in mean SOS averages from pretest to posttest, but with scores more similar to the full-length questionnaire group.

To simplify the presentation of these effects, we present in [Figure 2](#) the comparison between the multiply imputed mean

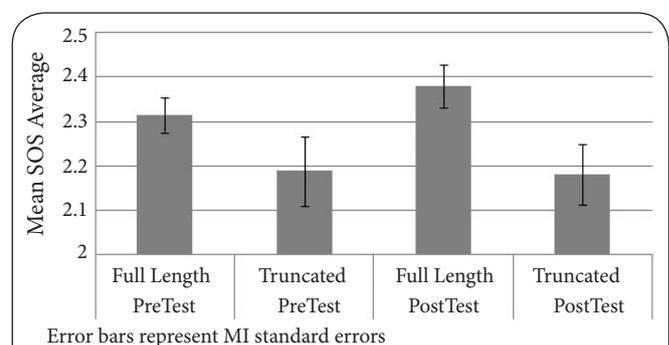


Figure 2. MI mean SOS average by test control versus truncated versions.

This bar chart shows the mean SOS average score for the pretest for the full questionnaire version and the combined truncated questionnaires. The gray bars represent scores that are the average of 10 multiply imputed questionnaire items that had values ranging from one to five. The multiple imputation standard deviations of these scores are marked by the error bars within each gray bar. The left axis runs from 2 to 2.5 in increments of 0.1.

SOS average for the full-length questionnaire versus the all truncated versions for both the pre- and posttest. **Figure 2** shows that the multiply imputed mean SOS average is higher for respondents receiving the full-length pretest questionnaire than for those receiving a truncated pretest questionnaire. This is true for both the pre- and the posttest.

Discussion

The objective of this study was to test the hypothesis that the length and content of pretest questionnaires may affect responses to posttest questionnaires, potentially undermining the ability to detect treatment or intervention effects in pretest-posttest study designs. Our results indicated that a designed missingness strategy allowed us to gather sufficient information concerning pre-intervention knowledge and attitudes while simultaneously reducing respondent burden and improving completion rates. The unit nonresponse rate at posttest was significantly lower among those completing truncated pretest versions (8%) compared to those completing the full version of the pretest (22%). In addition, our results revealed that posing a particular question at pretest altered responses to that question at posttest, such that treatment effects might be more difficult to detect. Pretest version (truncated vs. full) was associated with differences in responses to 6 items at posttest, and for 5 of those 6 items the direction of effects indicated that seeing the question for the second time in 3 months was predictive of more accurate knowledge and more favorable attitudes. This suggests that there was some learning happening as a result of exposure to the pretest, a significant concern when we have to “test” people prior to an intervention to assess its effects.

Although more research is needed on this subject to establish optimal questionnaire configurations and study designs, “designed missingness” methods have the potential to improve the assessment of treatment or intervention effects in a broad range of efficacy studies. Statistical power and the generalizability of results are likely to be enhanced by the reduced nonresponse associated with shorter baseline assessments. Moreover, assessment reactivity, where (pretest) testing itself can produce changes in outcomes irrespective of effects of the intervention, may be minimized by asking fewer questions at pretest and using multiple imputation to construct comparable baseline measures on the outcomes of interest. By ensuring that testing and instrumentation do not artificially raise the bar for detecting treatment effects, multiple imputation may provide a very efficient approach to improving trial designs.

Additional files

Appendix 1

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

Authors' contributions	OH	JS	RA
Research concept and design	✓	--	✓
Collection and/or assembly of data	✓	--	✓
Data analysis and interpretation	✓	✓	✓
Writing the article	✓	✓	--
Critical revision of the article	✓	--	✓
Final approval of article	✓	--	✓
Statistical analysis	✓	✓	--

Acknowledgement

This project was partially supported by Award Number K01MH087219 from the National Institute of Mental Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Mental Health or the National Institutes of Health.

Publication history

EIC: Jimmy Efir, East Carolina University, USA.
Received: 17-Jul-2015 Final Revised: 24-Aug-2015
Accepted: 11-Sep-2015 Published: 19-Sep-2015

References

- Whitlock EP, Polen MR, Green CA, Orleans T and Klein J. **Behavioral counseling interventions in primary care to reduce risky/harmful alcohol use by adults: a summary of the evidence for the U.S. Preventive Services Task Force.** *Ann Intern Med.* 2004; **140**:557-68. | [Article](#) | [PubMed](#)
- Jenkins RJ, McAlaney J and McCambridge J. **Change over time in alcohol consumption in control groups in brief intervention studies: systematic review and meta-regression study.** *Drug Alcohol Depend.* 2009; **100**:107-14. | [Article](#) | [PubMed](#)
- Strauss WJ, Ryan L, Morara M, Iroz-Elardo N, Davis M, Cupp M, Nishioka MG, Quackenboss J, Galke W, Ozkaynak H and Scheidt P. **Improving cost-effectiveness of epidemiological studies via designed missingness strategies.** *Stat Med.* 2010; **29**:1377-87. | [Article](#) | [PubMed Abstract](#) | [PubMed Full Text](#)
- Shoemaker DM. **Principles and Procedures of Multiple Matrix Sampling.** 1973. | [Pdf](#)
- Munger GF and Lloyd BH. **The Use of Multiple Matrix Sampling for Survey Research.** *J Exp Educ.* 1988; **56**:187-191. | [Article](#)
- Raghunathan TE and Grizzle JE. **A Split Questionnaire Survey Design.** *J Am Stat Assoc.* 1995; **429**:54-63. | [Article](#)
- Chipperfield JO and Steel DG. **Efficiency of split questionnaire surveys.** *J Stat Plan Inference.* 2011; **141**:1925-1932. | [Article](#)
- University of Connecticut Institute for Public Health Research. **Connecticut Youth Suicide Prevention Initiative Local Evaluation – Final Report.** 2010. | [Pdf](#)
- Aseltine RH, Jr. and DeMartino R. **An outcome evaluation of the SOS Suicide Prevention Program.** *Am J Public Health.* 2004; **94**:446-51. | [PubMed Abstract](#) | [PubMed Full Text](#)
- Rubin DB. **Multiple imputations in sample surveys - a phenomenological Bayesian approach to nonresponse.** *Proceedings of the Survey research Methods Section of the American Statistical Association.* 1978; 20-34. | [Pdf](#)
- Rubin DB. **Multiple Imputation for Nonresponse in Surveys.** 1987. | [Website](#)
- Harel O and Zhou XH. **Multiple imputation: review of theory, implementation and software.** *Stat Med.* 2007; **26**:3057-77. | [Article](#) | [PubMed](#)
- Raghunathan TE, Solenberger PW and Van Hoewyk J. **IVeWare: and**

Variance Estimation Software User Guide. Survey Methodology Program, Survey Research Center, Institute for Social Research, University of Michigan. 2002. | [Website](#)

14. Harel O. **Inferences on missing information under multiple imputation and two-stage multiple imputation.** *Stat Methodol.* 2007; **4**:75-89. | [Article](#)
15. Raghunathan TE, Lepkowski JM, Van Hoewyk J and Solenberger P. **A Multivariate Technique for Multiply Imputing Missing Values Using a Sequence of Regression Models.** *Survey Methodology.* 2001; **27**:85-95. | [Pdf](#)

Citation:

Harel O, Stratton J and Aseltine R. **Designed missingness to better estimate efficacy of behavioral studies-application to suicide prevention trials.** *J Med Stat Inform.* 2015; **3**:2.
<http://dx.doi.org/10.7243/2053-7662-3-2>