ESTIMATING MORTALITY IMPACTS: A TALE OF TWO DATA SOURCES

Jennifer Schore, Mathematica Policy Research, Inc.

The evaluation of the National Long Term Care Demonstration[1] is estimating program impacts on mortality both because of its importance as an outcome and because of its indirect effect on the use of health care services. Mortality could be decreased by the demonstration's careful monitoring of health problems and its ability to improve access to services. Mortality could also be decreased if the demonstration enabled clients to remain outside of nursing homes, where they tend to deteriorate quickly and may lose their desire to live, or if it reduced the amount of time spent in the hospital, an environment in which the likelihood of contracting communicable diseases (such as pneumonia) is increased. On the other hand, if the demonstration substitutes community care for medically-oriented health care and institutional services, to the extent that community residence entails less immediate access to medical personnel, mortality rates could conceivably be increased.

To measure the demonstration's impacts on mortality, data were collected from two sources: interviews and client tracking forms[2] that collected information on sample members for twelve or eighteen months following random assignment[3] and death records collected from state offices of vital statistics at the conclusion of the evaluation. Interview data alone were not felt to be sufficient to estimate mortality impacts because it was believed that the information on deaths from interviews and client tracking records was likely to be more complete for treatments than controls. Such a difference would lead to bias in estimates of the demonstration's impact on mortality because the control group death rate would be underestimated to a greater degree than the treatment group rate.

State vital statistics offices were searched for death records for all sample members who did not respond to their last scheduled interview using identical procedures and data to determine mortality for treatments and controls, in order to have a source of data that was not differentially available for treatments and controls. However, for various reasons, which will be discussed in more detail below, death records data on deaths can undercount actual deaths though it is not expected that the degree of undercounting would differ between treatments and controls.

This problem is less serious than the problems with the interview/client tracking data because the degree of undercounting by death records is expected to be the same for the two groups. Thus, the death records will yield unbiased estimates of the ratio of treatment to control group death rates, or, equivalently, of the proportionate effect that channeling has on mortality. However the difference between treatment and control group death rates expressed in percentage points will be understated by the undercounting rate.

In this regard the death records data are superior to the interview/client tracking data as a source of information concerning mortality. The degree to which death records understate actual death rates is unknown, but at least it should be equal for the two groups. However, if the factor of proportional undercounting could be reduced for both groups by using additional data on deaths, without distorting the expected equivalence between the two groups in the degree of underreporting, then the estimated death rates and treatment/control difference would be biased by a smaller factor. That is the approach proposed here. By also counting as dead those individuals who were found to be deceased from the interview or client tracking. but not from the death records, some of the undercounting can be eliminated. If we can show that supplementing the death records data with these other data has the effect of increasing the observed death rates by the same factor for the two groups, then we can be confident that any remaining undercount continues to be equal for the two groups and lower than it would be if only death records were used. Below we first describe the data sources more completely, then present the formal statistical test that was conducted to ensure that adding the interview/tracking data on mortality does not distort the proportional nature of the undercount.

A. THE LIMITATIONS OF DEATH DATA FROM INTERVIEWS AND CLIENT TRACKING RECORDS AND FROM OFFICIAL DEATH RECORDS

Interviews and client tracking questionnaires (hereafter referred to collectively as "interview" data or "interviews") both collected information concerning the death of sample members; in most cases this was the date of death. It was hypothesized that these data would be more complete for treatments than controls because channeling program staff would be in closer and more frequent contact with treatment group members than interviewers would be with (treatment or) control group members. In addition, baseline attriters (17 percent of the control group but only 7 percent of the treatment group) would not be contacted by interviewers for follow-up. These two factors would cause more deaths to be missed for the control group than the treatment group and could result in an overestimate of treatment/control differences in mortality.

In order to avoid the bias due to differential measurement, a death records search was initiated for each sample member who did not respond (either completely or partially) to his/her last scheduled interview. In order to ensure that the search itself would not be biased, individuals searching for the death records were unaware of treatment/control status of listed sample members, and data used to match death records to sample members were restricted to those items available on the screening interview which was administered prior to random assignment. On the other hand, while the death records were expected to provide comparable data on mortality for treatments and controls, they were also expected to underestimate deaths for both groups and, thus, underestimate impacts. The underestimation was hypothesized to stem from three sources:

1. The administrative lag in the processing of the records. Death certificates are filled out by hospitals and local jurisdictions before being passed on to State Offices of Vital Statistics, in which our searches took place. This lag was estimated to range from 2 to 5 months among the states involved in the channeling demonstration.

2. Out-of-state deaths. Sample members who died outside the state in which they enrolled in the demonstration are unlikely to have death records in the demonstration state unless they maintained a residence in that state. The National Death Index (NDI) was considered as a source that would not suffer from this potential problem, however the administrative lag for the NDI averages 14 months.

3. Insufficient screen data to match death records to sample members. The screen data available included name, address, date of birth, and social security or Medicare number. Addresses sometimes changed and birth dates were corrected after the screen was administered but these changes were not available to this process in order not to introduce bias to the process. Social Security/Medicare numbers, if available on the screen, were sometimes missing on the death records, exacerbating the matching process.

B. TESTING THE FEASIBILITY OF A COMBINED DATA SOURCE STRATEGY

As explained above, the proportional nature of the error in the death records was viewed as less serious than the potential noncomparability of the interview data on mortality. However, supplementing the death records with the interview data on deaths could reduce this undercounting. If this supplementation did not introduce significant noncomparability where heretofore there had been none, the combined death data would provide a superior measure.

In order to test whether the addition of the interview data on mortality preserves the undercount presumed equivalent between treatments and controls in the proportion of deaths missed, we compare for treatments and controls the ratio of the mortality rate based on death records only to the mortality rate based on the combined data sources. In order to illustrate this, let:

N_D^* = the "true" number of deceased sample members (unobserved)

N_{DR} = the number of deceased sample members according to the death records

N_{DI} = the number of additional deceased sample members according to the "interview" data, i.e., those not counted in N_{DR}

N = the sample size

Then, the "true" mortality rate can be given by:

$$\frac{N_D^*}{N} = \frac{N_{DR}}{N} * \frac{N_D^*}{N_{DR}}$$

where the rightmost term is a scale factor that measures undercounting in the death records. We assume this factor to be equal for treatments and controls. This scale factor can be split into two components:

$$\frac{N_D^*}{N_{DR}} = \frac{N_{DI} + N_{DR}}{N_{DR}} * \frac{N_D^*}{N_{DI} + N_{DR}}$$

The first term to the right of the equal sign is the ratio of the mortality rate based on combined sources to the mortality rate based on records only. The second term, a scale factor describing the undercount of the death records and interview data combined, is unobservable, but smaller than the scale factor for records only. Since we assume that N_D^* / N_{DR} is equal for treatments and controls, if we can show that $N_{DI} + N_{DR}/N_{DR}$ is equal for both groups, then so must $N_D^*/N_{DI}+N_{DR}$ be equal for treatments and controls. If this is the case, our mortality impacts based on combined data will be underestimated by less than they were using death records only and will not be biased by noncomparable data for treatments and controls.

It is the inverse of the ratio of mortality rates that will be compared for treatments and controls:

$$P = \frac{N_{DR}}{N_{DI} + N_{DR}}$$

We will call this the "completeness" rate of the death records in identifying the total number of observable deaths. The treatment and control proportions are compared statistically using standard t-statistics.[4]

C. THE RESULTS

Table 1, a crosstabulation of mortality based on "interview" data with mortality based on the death records, gives an overview of the relative efficiency of each data source in reporting deaths, at 6, 12, and 18 months after random assignment. Mortality rates based on death records only are about 15 percent at six months, about 25 percent at 12 months, and for the 18-month cohort, about 31 percent at 18 months. Rates vary little between treatment and controls. Supplementing the death records data with "interview" death data raises the mortality rates to approximately 17 percent, 28 percent, and 36 percent at 6, 12, and 18 months, respectively.

Table 2 presents treatment/control comparisons of the "completeness" rates of the death records in identifying deaths; this comparison was

TABLE 1

COMPARISONS OF DEATH RECORDS WITH INTERVIEW AND CLIENT TRACKING DATA ON NUMBER DECEASED AT 6-, 12-, AND 18-MONTH ANNIVERSARIES
(Research Sample)

Mortality Information from Death Records	Interview and Client Tracking Data: Treatment Group Alive	Treatment Group Dead	Treatment Group Unknown	Treatment Group Total	Control Group Alive	Control Group Dead	Control Group Unknown	Control Group Total
6 Months After Random Assignment								
No Death Record Found	2578	62	495	3135	1590	50	581	2221
Death Record Found	n.a.	539	28	567	n.a.	343	75	418
Total	2578	601	523	3702	1590	393	656	2639
12 Months After Random Assignment								
No Death Record Found	2264	110	425	2799	1366	86	519	1971
Death Record Found	n.a.	854	49	903	n.a.	542	126	668
Total	2264	964	474	3702	1366	628	645	2639
18 Months After Random Assignment								
No Death Record Found	874	84	302	1260	533	72	314	919
Death Record Found	n.a.	536	52	588	n.a.	322	82	404
Total	874	620	354	1848	533	394	396	1323

NOTES: Interview and client tracking information was used to construct sample members' "status" at 6, 12, and 18 months after random assignment. Individuals were classified as "alive" if they were known to be living in the community or in a hospital or nursing home on the particular date. Those not known to be alive or dead were classified as "unknown." These "unknown" groups are dominated by sample members who failed to complete a baseline interview.

Individuals for whom no death record was found include those not included in the search because they were known to be alive (respondents to their last scheduled interview) and those in the search but for whom no record was found.

described in the previous section. The t-statistics presented are those calculated assuming the proportions (that is, the completeness rates) compared are from populations with equal variances; the t-statistics under the unequal variance assumption were computed and do not differ much from those presented. A value for these proportions near unity indicates that the death records captured most of the death data available from the interviews; or equivalently, the interview data added little extra information concerning deaths to the death records. Thus, the hypothesis of interest, that the "interview" data offers more additional information concerning deaths of treatments than controls, would be manifested in a negative treatment/control difference in these completeness rates.[5]

We present results for the total sample first, followed by a brief discussion of the model and site results. At the six-month anniversary the death records identified about 90 percent of treatment and control deaths found on either the records or "interviews." At 12 months this "completeness" rate for the records drops slightly to about 89 percent for each group. At 18 months, for the 18-month cohort, the "completeness" rate drops to 88 percent for treatments and 85 percent for controls; this treatment/control difference is larger than at 6 or 12 months but is not statistically significant. The direction of this difference implies that, contrary to expectations, "interview" data available for controls added more information to the death records than did "interview" data for treatments. Although this result is somewhat puzzling--the cumulative effect of small positive differences in proportions across most of the sites--it refutes the hypothesis that "interview" data added more information for treatments than controls beyond that which was available from the records search.

We now turn briefly to model- and site-level comparisons of treatment and control death records "completeness" rates in order to determine whether differences in administrative lag across sites differentially affect site-level mortality estimates. At six, twelve and eighteen months treatment and control estimates of these rates are not statistically different from zero

TABLE 2

ESTIMATES OF COMPLETENESS RATES OF THE DEATH RECORDS IN IDENTIFYING OBSERVED DEATHS AT 6, 12, AND 18 MONTHS AFTER RANDOMIZATION
(Research Sample)

	Death Records Completeness Rate				Sample Size	
	Treatment Group Rate	Control Group Rate	T/C Difference	t-statistic	Treatment	Control
Six Months Random Assignment						
Basic Model						
Baltimore	.889	.880	.009	(.15)	417	275
E. Kentucky	.879	.875	.004	(.05)	246	242
Houston	.940	.979	-.039	(-.98)	401	273
Middlesex County	.886	.791	.095	(1.63)	451	299
Southern Maine	.938	.911	.027	(.56)	264	260
Model Average	.906	.881	.025	(.96)	1779	1349
Financial Model						
Cleveland	.952	.972	-.020	(-.48)	388	198
Greater Lynn	.833	.846	-.013	(-.17)	309	308
Miami	.873	.854	.019	(.28)	450	300
Philadelphia	.901	.942	-.041	(-.87)	581	288
Rensselaer County	.907	.941	-.034	(-.55)	195	196
Model Average	.898	.907	-.009	(-.34)	1923	1290
Total Sample	.902	.893	.009	(.49)	3702	2639
Twelve Months After Random Assignment						
Basic Model						
Baltimore	.867	.867	.000	(.00)	417	275
Eastern Kentucky	.865	.846	.019	(.28)	246	242
Houston	.914	.963	-.049	(-1.33)	401	273
Middlesex County	.890	.851	.039	(.90)	451	299
Southern Maine	.917	.903	.014	(.32)	264	260
Model Average	.892	.888	.004	(.19)	1779	1349
Financial Model						
Cleveland	.934	.928	.006	(.14)	388	198
Greater Lynn	.861	.865	-.004	(-.07)	309	308
Miami	.835	.824	.011	(.19)	450	300
Philadelphia	.904	.903	.001	(.03)	581	288
Rensselaer County	.907	.920	-.013	(-.25)	195	196
Model Average	.891	.884	.007	(.32)	1923	1290
Total Sample	.892	.886	.006	(.40)	3702	2639
Eighteen Months After Random Assignment						
Basic Model						
Baltimore	.867	.794	.073	(1.15)	227	148
Eastern Kentucky	.871	.824	.047	(.53)	108	107
Houston	.933	.929	.004	(.09)	217	149
Middlesex County	.861	.794	.067	(1.08)	228	151
Southern Maine	.912	.855	.057	(.95)	142	142
Model Average	.892	.838	.054	(1.94)	922	697
Financial Model						
Cleveland	.905	.885	.020	(.29)	170	85
Greater Lynn	.767	.861	-.094	(-1.26)	166	167
Miami	.840	.805	.035	(.48)	229	152
Philadelphia	.872	.902	-.030	(-.56)	277	140
Rensselaer County	.917	.857	.060	(.72)	84	82
Model Average	.862	.863	-.001	(-.03)	926	626
Total Sample	.875	.849	.026	(1.27)	1848	1323

None of the treatment control differences on this table are statistically different from zero at the 5 percent level, using a two-tailed test.

in either model or in any site. At six months the "completeness" rate of the records ranges from about 79 percent in Middlesex County to about 98 percent in Houston. At twelve months the rates range from 82 percent in Miami to 96 percent in Houston and for the eighteen-month cohort at 18 months, the rates range from 77 percent in Greater Lynn to 93 percent in Houston. The model-and site-level treatment/control comparisons reflect the comparisons for the overall sample, and imply that no bias would be introduced to mortality impact estimates based on death records data by supplementing it with "interview" data.

In addition to comparing completeness rates of the death records in accounting for observed deaths for treatments and controls, we also compared mortality impact estimates based on death records only with those based on combined sources to ensure that using the combined data would not lead to different conclusions than using the records data alone. Therefore, we examine the increases observed in the mortality rates when the death records are supplemented with "interview" data and examine the effect of these increases on treatment/control differences.[6] (See Table 3.) Our conclusion is that the magnitude, direction and significance of treatment/control differences were seldom affected by data source. Therefore, we will discuss this investigation only briefly.

As mentioned earlier, supplementing death records with "interview" data caused mortality rates to increase from two to five percentage points for both treatments and controls in the overall sample. Treatment/control differences in mortality rate based on the death records were essentially unaffected by the addition of interview data. Differences were small and negative for months six and twelve, and small and positive for month eighteen, regardless of whether mortality was measured with death records only or death records supplemented by "interview" data.

It was also the case at the site-level that the magnitude, direction and significance of treatment/control differences were essentially unaffected by the choice of data source. An exception to this was the treatment/control difference for Houston at six and twelve months. (We note that in the comparison of death records "completeness" rates, the largest negative t-statistics were observed for Houston for these two time periods.) At six months, the mortality rate for treatments is 11.7 percent based on records only and 12.5 percent based on combined sources; for controls it is 17.2 percent based on records only and 17.6 percent based on combined sources. The treatment/control difference in mortality rates using only the records, at just under -5.5 percentage points, is significant at the 5 percent level, whereas the difference in rates based on combined sources, at just over -5.1 percentage points, is not significant at that level. A similar situation occurred at twelve months. However, despite the change in significance, conclusions regarding channeling's impact on mortality at that site are not affected by the choice of data source: the negative treatment/control differences are more likely due to differences between the two groups that occurred in spite of random assignment than to the impact of the demonstration.

At eighteen months the largest negative t-statistics for the comparison of death records "completeness" rate was observed in Greater Lynn, however this difference had no effect on magnitude, direction or significance of the treatment/control difference in mortality rates at that site. We concluded, therefore, that, apart from increasing treatment and control mortality rates slightly, the addition of "interview" data to death records data had virtually no effect on estimates of treatment/control differences in mortality rate.

D. CONCLUSIONS

We conclude from the analysis described above that combining interview data with death records reduces the extent of underreporting substantially and therefore reduces the proportional bias in estimated impacts on mortality, without introducing new bias due to the differential availability of interview data on mortality. Treatment/control comparisons of ratios of the number of deaths from the death records to deaths on interviews and records combined showed no statistically significant differences when measured for the overall sample or for any of the individual models or sites. This was due largely to the fact that death records captured a large proportion of the observed number of deaths. Thus, even if the interview data alone miss fewer treatment group than control group deaths, there is little treatment/control difference in the number of interview-reported deaths in excess of those captured by the death records. Therefore, for the purposes of estimating the demonstration's impact on mortality, we combined death records with interview and client tracking data.

FOOTNOTES

[1]The evaluation of the National Long Term Care Demonstration was performed by Mathematica Policy Research, Inc., under contract to the Department of Health and Human Services, contract no. HHS-100-80-0157. For more information concerning the demonstration, please contact the DHHS project officer, Ms. Mary Harahan, Office of the Secretary, Department of Health and Human Services, Room 447F, Hubert H. Humphrey Building, Washington, DC 20201. For a list evaluation publications, contact Publications Department, Mathematica Policy Research, Inc., P.O. Box 2393, Princeton, NJ 08540.

[2]Client tracking forms were used by site program staff to record the completion of events associated with program participation such as: completion of a screening interview, completion of a baseline assessment and care plan sign off. Deaths were also noted here. Client tracking forms are available for treatment group members only.

[3]The half of one sample enrolling earliest in each of the 10 demonstration sites were eligible for an interview at eighteen months. This subsample is referred to as the 18-month cohort

TABLE 3

ESTIMATED MORTALITY IMPACTS USING DEATH RECORDS ONLY AND DEATH RECORDS COMBINED WITH INTERVIEW/CLIENT TRACKING DATA 6, 12, AND 18 MONTHS AFTER RANDOMIZATION
(Research Sample)

	Mortality Rates from Records				Mortality Rates from Combined Sources				Sample Size	
	Treatment Group Rate	Control Group Rate	T/C Difference	t-statistic	Treatment Group Rate	Control Group Rate	T/C Difference	t-statistic	Treatment	Control
6 Months After Random Assignment										
Basic Model										
Baltimore	15.35	16.00	-.65	-.23	17.27	18.18	-.91	-.31	417	275
E. Kentucky	11.79	11.57	.22	.07	13.41	13.22	.19	.06	246	242
Houston	11.72	17.22	-5.50*	-2.02	12.47	17.58	-5.11	-1.85	401	273
Middlesex County	17.29	17.73	-.44	-.15	19.51	22.41	-2.90	-.96	451	299
Southern Maine	22.73	19.62	3.11	.87	24.24	21.54	2.70	.74	264	260
Financial Model										
Cleveland	15.46	17.68	-2.22	-.67	16.24	18.18	-1.94	-.58	388	198
Greater Lynn	11.33	14.29	-2.97	-1.10	13.59	16.88	-3.29	-1.14	309	308
Miami	12.22	11.67	.55	.23	14.00	13.67	.33	.13	450	300
Philadelphia	17.21	17.01	.20	.07	19.10	18.06	1.04	.38	581	288
Rensselaer County	20.00	16.33	3.67	.94	22.05	17.35	4.70	1.17	195	196
Total Sample	**15.32**	**15.84**	**-.52**	**-.57**	**16.99**	**17.73**	**-.74**	**-.65**	**3702**	**2639**
12 Months After Random Assignment										
Basic Model										
Baltimore	23.50	26.18	-2.68	-.79	27.10	30.18	-3.08	-.87	417	275
Eastern Kentucky	18.29	18.18	.11	.03	21.14	21.49	-.35	-.09	246	242
Houston	21.20	28.57	-7.37*	-2.16	23.19	29.67	-6.48	-1.86	401	273
Middlesex County	27.05	28.76	-1.71	-.51	30.38	33.78	-3.40	-.98	451	299
Southern Maine	29.55	28.85	.70	.18	32.20	31.92	.28	.07	264	260
Financial Model										
Cleveland	25.52	25.76	-.24	-.06	27.32	27.78	-.46	-.12	388	198
Greater Lynn	18.12	25.00	-6.88*	-2.08	21.04	28.90	-7.86*	-2.26	309	308
Miami	20.22	18.67	1.55	.53	24.22	22.67	1.55	.49	450	300
Philadelphia	27.71	28.82	-1.11	-.34	30.64	31.94	-1.30	-.39	581	288
Rensselaer County	34.87	23.47	11.40*	2.49	38.46	25.51	12.95*	2.76	195	196
Total Sample	**24.39**	**25.31**	**-.92**	**-.84**	**27.36**	**28.57**	**-1.21**	**-1.06**	**3702**	**2639**
18 Months After Random Assignment										
Basic Model										
Baltimore	28.63	33.78	-5.15	-1.05	33.04	42.57	-9.53	-1.87	227	148
Eastern Kentucky	25.00	26.17	-1.17	-.20	28.70	31.78	-3.08	-.48	108	107
Houston	32.26	35.57	-3.31	-.65	34.56	38.26	-3.70	-.72	217	149
Middlesex County	32.46	33.11	-.65	-.13	37.72	41.72	-4.00	-.87	228	151
Southern Maine	36.62	33.10	3.52	.62	40.14	38.73	1.41	.24	142	142
Financial Model										
Cleveland	33.53	27.06	6.47	1.07	37.06	30.59	6.47	1.02	170	85
Greater Lynn	19.88	33.53	-13.65*	-2.84	25.90	38.92	-13.02*	-2.55	166	167
Miami	29.69	21.71	7.98	1.77	35.37	26.97	8.40	1.72	229	152
Philadelphia	39.35	32.86	6.49	1.31	45.13	36.43	8.70	1.70	277	140
Rensselaer County	39.29	21.95	17.34*	2.45	42.86	25.61	17.25*	2.37	84	82
Total Sample	**31.82**	**30.52**	**1.28**	**.77**	**36.31**	**35.98**	**.38**	**.19**	**1848**	**1323**

NOTE: Treatment/control differences are simple comparisons of means, not regression-adjusted estimates.

*Different from zero statistically at the 5 percent significance level, using a two-tailed test.

of the sample. The demonstration operated in two different forms, referred to as the basic model and the financial control model. Each model operated in five sites.

[4]The appropriate t-statistic for this test is:

$$t = (P_T - P_C) \ / \ \sqrt{s^2 \ (1/N_T + 1/N_C)}$$

where

$s^2 = ((N_T-1) \ P_T Q_T + (N_C-1) P_C Q_C) \ / \ (N_T+N_C-2)$

P_T, P_C = values of P for treatments and controls

$Q_T, Q_C = 1-P_T, \ 1-P_C$,

N_T, N_C = values of $N_{DI}+N_{DR}$ for treatments and controls,

in the case where P_T and P_C come from populations with equal variances. For the case where P_T and P_C come from populations with unequal variances:

$$t = (P_T - P_C) \ / \ \sqrt{(P_T Q_T/N_T) + (P_C Q_C/N_C)}$$

using Satterthwaite's approximation for degrees of freedom.

[5]This may imply that a one-tailed test would have been more appropriate than the two-tailed test presented for consistency with the rest of the evaluation. However, inspection of the t-statistic column and the small size of any negative numbers in that column, implies that a less conservative test would not have altered our conclusions.

[6]For the purposes of this investigation we use a simple comparison of mean mortality rates. Regression-adjusted treatment/control differences were used to estimate the actual impacts of the demonstration.

[7]In fact, when impacts are estimated controlling for baseline and site characteristics, these differences are no longer significant.