Web Appendix 1

We prepared a linkage file with the following information from the administrative Department of Corrections (DOC) records: first and last name, middle initial, social security number (SSN), state and date of birth, race (black, white, or other), sex, marital status, all known aliases (up to 16 aliases were recorded per individual), and father's surname (for women). Prior research in Australia suggested that use of aliases increases the accuracy of linkage of criminal justice records with the national death index (1, 2). Given multiple records for individuals with multiple aliases or SSNs, a total of 383,970 records for 145,718 individuals were included. The National Death Index (NDI) searched for these records through December 31, 2012. NDI identified potential matches for 67,468 individuals, representing 170,209 records. Of these, 41,326 had only one record that matched to a NDI record and 26,142 individuals had two or more alias records that matched to at least one NDI record.

NDI calculated a probabilistic score for each potential match using an algorithm based on the extent of agreement between the information on the submission record and the corresponding NDI death record. NDI also categorized each potential match into one of five mutually exclusive classes based on how many identifying items match and how important those items are in determining true matches. Records coded as class 1 (the highest level of agreement) matched exactly on SSN, first name, middle initial, last name, sex, state of birth, birth month, and birth year, while other classes specify different types of inexact matches. Of the potential matches in our data, 4,021 records (2.4%) were "exact matches." For each class, NDI determined a probabilistic score threshold, above which the record is identified as a "true match" (3). Using

this approach, NDI classified 18,939 (11.1%) records corresponding to 7,864 unique individuals as true matches.

Web Appendix 2

We identified cohort members who died using data made available through public websites (SSDMF.info, cancelthesefunerals.com) in April 2015. The data file from this website was sourced from Social Security Administration prior to the removal of records from the Social Security Master Death File (SSDMF) in 2011, but subsequent updates through December 31, 2012 do not include deaths recorded by the states (4-6). Identifiers used for SSDMF linkage were identical to those sent to the NDI for matching, including multiple aliases per individual. To match records from the DOC database with SSDMF, we first selected all of the cases from the SSDMF file that matched on any of the SSNs in the DOC records, including multiple aliases. We then constructed a match-quality score by counting how many times each of the selected cases successfully matched on the following criteria: (1) identical match on surname and personal name (i.e., first name or given name); (2) first two characters of surname are identical; (3) first four characters of the surname are identical; and (4) first four characters of the personal name (i.e., first name or given name) are identical. The resulting match-quality score had a range from zero, indicating no successful matches on any of these criteria, to four, indicating successful matches on all four criteria. We considered all record pairs with scores of three or four to be legitimate matches (i.e., deaths according to SSDMF), which yielded 5,933 individuals with matching records in the SSDMF. We manually reviewed all of the cases that had matching social security numbers but did not have a sufficiently high match-quality score and checked for possible errors in name spellings or changes in names (e.g., for women) that may have prevented

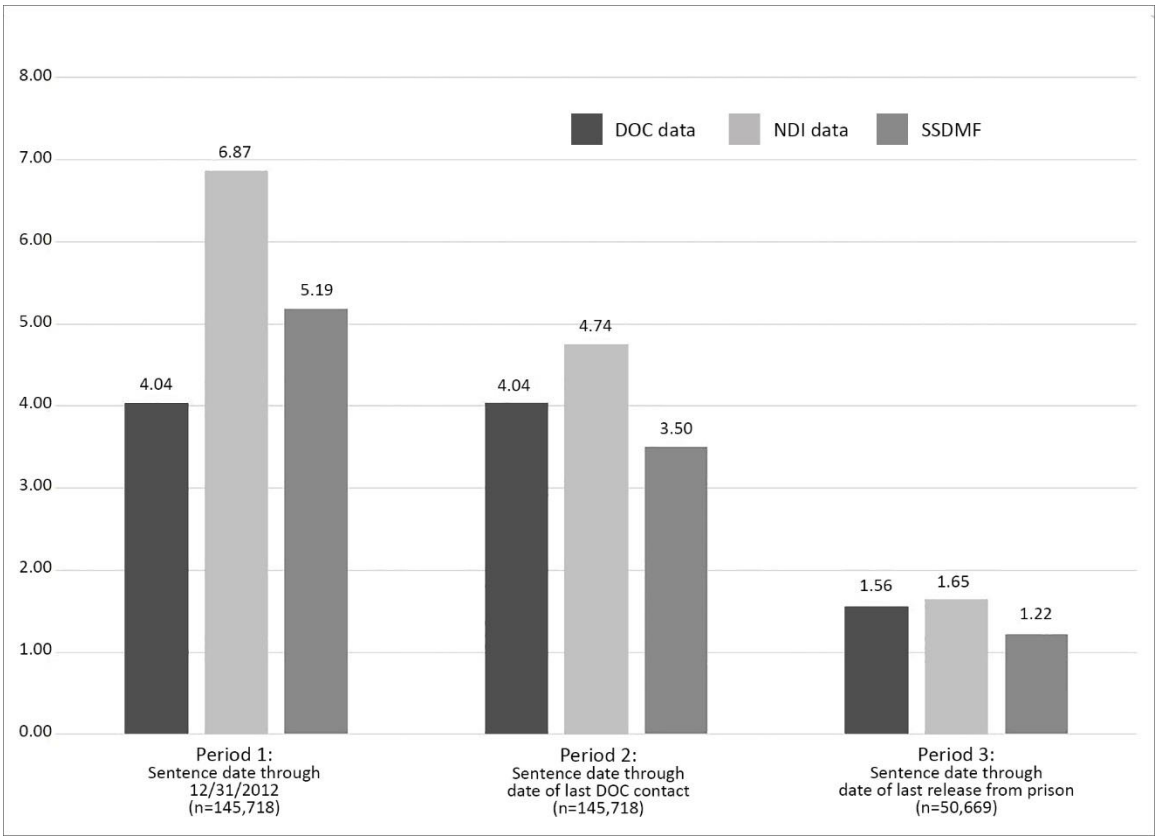
matching. Through this process, we identified 41 additional matches. In total, we identified 5,974 deaths using SSDMF through December 31, 2012.

Web Appendix 3

Because the DOC only recorded deaths that occurred while individuals were in state custody/supervision, it is important to account for the custody/supervision status of individuals when assessing agreement across data sources. To address this, we defined three observation periods that reflect different levels of certainty that the person was alive and under DOC custody/supervision. Each observation period began on the individual's date of sentencing; the earliest date was January 1, 2003. The first observation period ended on December 31, 2012, the cutoff date for NDI searches. This period provided the longest window of time for comparing deaths recorded by NDI and SSDMF, but many deaths that occur within this observation period were not recorded by the DOC because the individuals were no longer under custody/supervision. The second observation period was more restrictive, ending on the date of last DOC contact, defined by whichever of the following events occurred last: a DOC-recorded death in custody/supervision, release from prison without further supervision, discharge from parole or probation, or contact made by a corrections agent with an individual on parole/ probation. The third observation period was even more restrictive, covering only time during which a person was known to be in state prison during the study period. Individuals who never went to prison were not included in this observation period. For everyone else, this observation period ended on the date of their last release from prison (coded as December 30, 2012, for those still in prison on that date) or a DOC-recorded death in custody/supervision. The calculation of

exposed person time in the third observation period counted only time spent in prison; for those with multiple prison sentences, time out of prison was deducted from exposed person-time.

Web Figure 1. Unadjusted Mortality Rates (Deaths per 1,000 Person-Years) among Sentenced Individuals in a U.S. State Correctional System by Vital Status Data Source and Observation Period, 2003-2012



Web Appendix 4

In Web Table 1, we report subgroup specific results for the second observation period, which tracks individuals through the last date of recorded contact by the DOC. Narrowing the length of the observation period in this way diminished the discrepancies between the mortality rate estimates using the DOC records and the NDI (rate ratio [RR]=1.17). The disparity between mortality rates estimated by the SSDMF and the DOC (RR=0.78) became wider in this observation period, and the SSDMF ascertained fewer deaths (2,119) than the DOC (2,446) and the NDI (2,870). The restriction of the observation period did not change the differential between the mortality rate estimated by the NDI compared to SSDMF (RR=1.35), or the subgroup differences noted above.

The subgroup results for the third observation period are reported in Web Table 2. The difference between the NDI and DOC mortality rates was reduced considerably (RR=1.06) in this observation period, which covered only time that cohort members spent in prison, when DOC mortality records are likely to be most complete. The McNemar test still revealed a statistically significant difference in the NDI and DOC mortality rates, but we note that the sample size (*n*=50,674 in the third observation period) is large enough to reliably detect even small differences in mortality rates across data sources. Of the 50,674 people observed during this time period, the NDI recorded and 290 people as dying, 26 were identified as still alive by the DOC. We examined these records and found that 10 (38.5%) were deaths that occurred outside of the state. The discrepancy between the NDI and DOC death ascertainment was higher for blacks (RR=1.13) and people between the ages of 25 and 34 (RR=1.29). The SSDMF identified fewer deaths than either the NDI or the DOC in the third observation period.

Web Figure 1 summarizes the results of mortality rate differences across data sources and observation periods. Mortality rates were generally lower in the more constrained observation periods. The mortality rate estimated from the NDI data remained consistently higher than the SSDMF rate across all time periods, while the gap between the NDI and DOC narrowed considerably in observation periods 2 and 3, when the DOC records were more likely to detect deaths that occurred to people under their custody or supervision. Also, the SSDMF records yielded the lowest mortality rate estimates in observation periods 2 and 3.

We used Cohen's kappa to test inter-source agreement (7, 8). Web Table 3 presents cross-tabulations comparing how each pair of data sources recorded the vital status (living or deceased) of individuals over each time period. Consistent with the results from Web Table 1, the levels of agreement were highest between the NDI and DOC records, especially in period 3, when the two sources agreed on the vital status of all but 37 of the 50,674 individuals (kappa=0.93). The SSDMF produced lower levels of agreement with both DOC (highest kappa was .80 in period 3) and NDI (highest kappa was .83 in period 1).

Web Table 1. Comparison of Number of Deaths by Source for Observation Period 2: Sentence Date through Date of Last Department of Corrections Contact, U.S. State Correctional System, 2003-2012

	DOC		NDI		SSDMF		Source Comparisons: Mortality Rate Ratios & McNemar Tests		
	Number of Deaths	Crude Mortality Rate[a]	Number of Deaths	Crude Mortality Rate[a]	Number of Deaths	Crude Mortality Rate[a]	NDI/DOC	NDI/ SSDMF	SSDMF/ DOC
Full cohort	2,446	4.04	2,870	4.74	2,119	3.50	1.17[b]	1.35[b]	0.87[b]
Race									
Black	898	3.25	1111	4.03	686	2.48	1.24[b]	1.62[b]	0.76[b]
White	1539	4.74	1750	5.39	1429	4.40	1.14[b]	1.23[b]	0.93[b]
Other	9	2.04	9	2.05	4	0.91	1.00[b]	2.25[b]	0.45[b]
Sex									
Male	2112	4.03	2477	4.73	1827	3.49	1.17[b]	1.35[b]	0.87[b]
Female	334	4.12	393	4.85	292	3.60	1.18[b]	1.35[b]	0.874[b]
Age at study entry									
<25	532	2.43	629	2.88	386	1.76	1.18[b]	1.63[b]	0.73[b]
25-34	531	2.98	633	3.56	453	2.54	1.19[b]	1.40[b]	0.85[b]
35-44	582	4.32	664	4.93	526	3.91	1.14[b]	1.26[b]	0.91[b]
45-54	567	9.36	669	11.08	526	8.71	1.18[b]	1.27[b]	0.93[c]
≥55	234	17.63	275	20.78	228	17.23	1.18[b]	1.21[b]	0.98
Sentence Type									
Prison	722	4.29	764	4.54	559	3.32	1.06[b]	1.37[b]	0.77[b]
Probation	675	3.73	836	4.62	601	3.32	1.24[b]	1.39[b]	0.89[b]
Jail	125	3.42	153	4.19	110	3.02	1.23[b]	1.39	0.88[b]
Jail with probation	921	4.22	1113	5.10	845	3.87	1.21[b]	1.32[b]	0.92[b]
Fines and/or community service	3	2.79	4	3.72	4	3.72	1.33[b]	0.999[b]	1.33[b]

Abbreviation: DOC, Department of Corrections database; NDI, National Death Index; SSDMF, Social Security Death Master File.

[a] Unadjusted mortality rates calculated as (#deaths/person-years*1,000) ; Not adjusted for age or other demographic characteristics

[b] $P < 0.001$ for McNemar test for equivalence of proportions from related/paired samples

[c] $P < 0.01$ for McNemar test for equivalence of proportions from related/paired samples

Web Table 2. Comparison of Number of Deaths by Source for Observation Period 3: Sentence Date through Date of Last Release from Prison, U.S. State Correctional System, 2003-2012

	DOC		NDI		SSDMF		Source Comparisons: Mortality Rate Ratios & McNemar Tests		
	Number of Deaths	Crude Mortality Rate[a]	Number of Deaths	Crude Mortality Rate[a]	Number of Deaths	Crude Mortality Rate[a]	NDI/DOC	NDI/ SSDMF	SSDMF/ DOC
Full cohort	275	1.56	290	1.65	215	1.22	1.06[d]	1.35[b]	0.78[b]
Race									
Black	108	1.22	122	1.38	85	0.96	1.13[b]	1.44[b]	0.79[b]
White	167	1.93	168	1.95	130	1.51	1.01	1.29[b]	0.78[b]
Other	0	0.00	0	0.00	0	0.00	NA	NA	NA
Sex									
Male	262	1.56	275	1.64	206	1.23	1.05[d]	1.33[b]	0.79[b]
Female	13	1.50	15	1.73	9	1.04	1.16	1.66	0.69
Age at study entry									
<25	21	0.34	23	0.37	21	0.34	1.10	1.10	1.00
25-34	28	0.53	36	0.68	20	0.38	1.29[c]	1.80[b]	0.72
35-44	58	1.41	63	1.53	41	1.00	1.09	1.54[b]	0.71[b]
45-54	106	6.31	108	6.46	86	5.14	1.02	1.26[b]	0.82[c]
≥55	62	15.59	60	15.12	47	11.85	0.97	1.28[b]	0.76[b]
Sentence Type									
Prison	218	1.83	226	1.91	174	1.47	1.04	1.30[b]	0.80[b]
Probation	12	0.65	14	0.75	11	0.59	1.16	1.27	0.91
Jail	22	2.51	22	2.50	14	1.59	1.00	1.57[d]	0.63[c]
Jail with probation	23	0.77	28	0.93	16	0.53	1.21	1.76[d]	0.69[b]
Fines and/or community service	0	0.00	0	0.00	0	0.00	NA	NA	NA

Abbreviation: DOC, Department of Corrections database; NDI, National Death Index; SSDMF, Social Security Death Master File.

[a] Unadjusted mortality rates calculated as (#deaths/person-years*1,000); Not adjusted for age or other demographic characteristics

[b] $P < 0.001$ for McNemar test for equivalence of proportions from related/paired samples

[c] $P < 0.01$ for McNemar test for equivalence of proportions from related/paired samples

[d] $P < 0.05$ for McNemar test for equivalence of proportions from related/paired samples

Web Table 3. Cross-Tabulation of Mortality and Data Source by Observation Period, U.S. State Correctional System, 2003-2012

Observation Period 1: Sentence date through Dec 31, 2012 [a]

	NDI		
DOC	Living	Deceased	Total
Living	137,731	5,541	143,272
Deceased	93	2,353	2,446
Total	137,824	7,894	145,718
Rate of Agreement		96.13%	
Kappa		0.441[b]	

	SSDMF		
DOC	Living	Deceased	Total
Living	139,086	4,186	143,272
Deceased	659	1,787	2,446
Total	139,745	5,973	145,718
Rate of Agreement		96.68%	
Kappa		0.410[b]	

	SSDMF		
NDI	Living	Deceased	Total
Living	137,631	193	137,824
Deceased	2,114	5,780	7,894
Total	139,745	5,973	145,718
Rate of Agreement		98.42%	
Kappa		0.825[b]	

Observation Period 2: Sentence date through date of last DOC contact [a]

	NDI		
DOC	Living	Deceased	Total
Living	142,738	534	143,272
Deceased	110	2,336	2,446
Total	142,848	2,870	145,718
Rate of Agreement		99.56%	
Kappa		0.877[b]	

	SSDMF		
DOC	Living	Deceased	Total
Living	142,903	369	143,272
Deceased	696	1,750	2,446
Total	143,599	2,119	145,718
Rate of Agreement		99.27%	
Kappa		0.763[b]	

	SSDMF		
NDI	Living	Deceased	Total
Living	142,766	82	142,848
Deceased	833	2,037	2,870
Total	143,599	2,119	145,718
Rate of Agreement		99.37%	
Kappa		0.813[b]	

Observation Period 3: Sentence date through date of last release from prison [a]

	NDI		
DOC	Living	Deceased	Total
Living	50,373	26	50,399
Deceased	11	264	275
Total	50,384	290	50,674
Rate of Agreement		99.93%	
Kappa		0.934[b]	

	SSDMF		
DOC	Living	Deceased	Total
Living	50,381	18	50,399
Deceased	78	197	275
Total	50,469	215	50,674
Rate of Agreement		99.81%	
Kappa		0.799[b]	

	SSDMF		
NDI	Living	Deceased	Total
Living	50,371	13	50,384
Deceased	88	202	290
Total	50,459	215	50,674
Rate of Agreement		99.80%	
Kappa		0.799[b]	

Abbreviation: DOC, Department of Corrections database; NDI, National Death Index; SSDMF, Social Security Death Master File.

[b] $P < 0.001$ for McNemar test for equivalence of proportions from related/paired samples

Web References

1. Kariminia A, Butler T, Corben S, et al. Mortality among prisoners: how accurate is the Australian National Death Index? *Australian and New Zealand Journal of Public Health* 2005;29(6):572-575.
2. Larney S, Burns L. Evaluating Health Outcomes of Criminal Justice Populations Using Record Linkage: The Importance of Aliases. *Evaluation Review* 2011;35(2):118-128.
3. National Death Index. *National Death Index Plus: Coded Causes of Death; Supplement to the NDI User's Manual*. Hyattsville, Maryland: National Death Index. Department of Health and Human Services. National Center for Health Statistics; 1999.
4. da Graca B, Filardo G, Nicewander D. Consequences for healthcare quality and research of the exclusion of records from the Death Master File. *Circ Cardiovasc Qual Outcomes* 2013;6(1):124-128.
5. Jacobs JP, O'Brien SM, Shahian DM, et al. Successful linking of the Society of Thoracic Surgeons Database to Social Security data to examine the accuracy of Society of Thoracic Surgeons mortality data. *J Thorac Cardiovasc Surg* 2013;145(4):976-983.
6. Blackstone EH. Demise of a vital resource. *J Thorac Cardiovasc Surg* 2012;143(1):37-38.
7. Agresti A. *An Introduction to Categorical Data Analysis* Hoboken, NJ: Agresti A. John Wiley & Sons; 2007.
8. Cohen J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 1960;20:37-46.