

## RESEARCH ARTICLE

# A national assessment of legacy versus new generation Medicaid data

Jessy K. Nguyen MPP | Prachi Sanghavi PhD 

Department of Public Health Sciences,  
Biological Sciences Division, The University of  
Chicago, Chicago, Illinois, USA

**Correspondence**

Prachi Sanghavi, Department of Public Health  
Sciences, Biological Sciences Division, The  
University of Chicago, 5841 S. Maryland  
Avenue, MC2000, Chicago, IL 60637, USA.  
Email: [psanghavi@health.bsd.uchicago.edu](mailto:psanghavi@health.bsd.uchicago.edu)

**Funding information**

This research was supported by a grant from  
the Agency for Healthcare Research and  
Quality (R01HS025720). Acquisition of  
Medicare and Medicaid claims data was  
sponsored by the University of Chicago  
Becker Friedman Institute and the University  
of Chicago Center for Health Administration  
Studies, respectively.

**Abstract**

**Objective:** To compare the performance of Medicaid legacy, Medicaid new generation, and Medicare claims on data analytic tasks.

**Data Sources:** Medicaid Analytic eXtract (MAX) claims (legacy) of 100% beneficiaries in 2011 (all states except Idaho), 2012 (all states), 2013 (28 states), and 2014 (17 states); 2016 Transformed Medicaid Statistical Information System Analytic Files (TAF) claims (new generation) of 100% beneficiaries from all states; Medicare claims of 20% beneficiaries in 2011–2014, 2016.

**Study Design:** We focused on the chain of events that starts with an out-of-hospital medical emergency and ends with hospital death or survival to discharge. We developed six data quality indicators to assess ambulance variables; linkage between claims; external cause of injury code reporting; and death reporting on hospital discharge status codes. For the latter, we estimated injury severity and modeled its association with death in the Medicare population. We used the model to compare reported versus expected deaths by injury severity in the Medicaid population. Datasets were compared by state and fee-for-service versus managed care.

**Data Extraction Methods:** Medicare and Medicaid beneficiaries with emergency ambulance transports.

**Principal Findings:** Medicare claims had high performance across indicators and states; MAX claims substantially underperformed on multiple indicators in most states. For example, most states reported external cause codes for over 90% of Medicare but less than 15% of Medicaid injury cases. Medicaid fee-for-service did not consistently perform better than Medicaid managed care. Compared with MAX, TAF claims performed significantly better on some indicators but continued to have poor external cause code reporting. Finally, MAX and TAF managed care records reported deaths at discharge in the range of expected deaths; however, fee-for-service claims might have underreported high-severity injury deaths.

**Conclusions:** New generation Medicaid claims performed better than legacy claims on some indicators, but much more improvement is needed to allow high-quality policy analysis.

**KEYWORDS**

data quality, death reporting, injuries, Medicaid, medical emergencies, Medicare

**What is known on this topic**

- A major impediment to using national Medicaid claims and encounter data for policy analysis and national-level estimation has been the concern that there may be substantial variation in data quality between states.
- To date, assessment of Medicaid data quality has largely occurred outside of a realistic research context, without reasonable benchmarks for comparison, and focused chiefly on encounter records over fee-for-service claims.
- Further, the new generation of Medicaid data, referred to as Transformed Medicaid Statistical Information System Analytic Files (TAF), has yet to be examined.

**What this study adds**

- We developed data quality indicators based on analytical tasks that are commonly performed in health services research, and compared performance of these across Medicaid legacy, Medicaid new generation, and Medicare claims data.
- Legacy Medicaid data performed poorly on multiple indicators in most states; Medicare performed highly across all indicators and states. Medicaid fee-for-service claims did not perform consistently better than Medicaid managed care encounter records.
- Medicaid TAF performed better than legacy Medicaid data on some indicators but needs further improvement.

## 1 | INTRODUCTION

Medicaid served 73.4 million people in the United States at an estimated cost of \$600 billion in 2017.<sup>1</sup> As a joint federal-state initiative, Medicaid program policy can differ across states, creating a unique opportunity for studying the impacts of varying policies.<sup>2</sup> To aid such research and allow for national-level estimation, for years 1999 to 2015, the Centers for Medicare and Medicaid Services (CMS) created the Medicaid Analytic eXtract (MAX) files, a national, beneficiary-level compilation of insurance claims, akin to Medicare claims files.<sup>3,4</sup> However, concern that there may be a substantial amount of variation in data quality between states has been a major impediment to using these files.

CMS acknowledged MAX data quality issues through anomaly tables by state and year and commissioned Mathematica Policy Research to write *Medicaid Issue Briefs*, which have documented missingness, inconsistent coding, and reporting errors. These analyses largely focused on managed care encounter records and identified these records to be altogether unusable in some states, based on a comparison with fee-for-service (FFS) claims.<sup>5-8</sup> However, FFS claims have received limited reviews and may have data quality issues too, and thus may not be appropriate benchmarks.<sup>9</sup>

Recently, in part to address these issues, CMS created a new generation of Medicaid data called the Transformed Medicaid Statistical Information System (T-MSIS) Analytic Files (TAF), available for all states from 2016 onwards.<sup>10-12</sup> Accompanying these are a few new data quality briefs, which, like the ones for MAX, provide high-level overviews of data fields and indicate at least some problems persist.<sup>13-17</sup> However, as these as well as the earlier analyses were conducted outside of actual research contexts, deeper investigations are needed to better understand the usability of these data for policy research.

In this study, we conducted the first academic assessment of TAF data as well as a parallel investigation of MAX data that addresses three limitations of past work. First, to study the usability of Medicaid data in a realistic research scenario, we created analytical datasets for studying the chain of events that starts with an out-of-hospital medical emergency and ends with either hospital death or survival to discharge. To allow deeper investigation, we focused about half of our analysis on trauma, the leading cause of death in the under 65 years of age population.<sup>18</sup> This required us to link claims to create episodes of care, use diagnosis codes to generate injury severity scores, and estimate mortality, much of which is relevant to other diseases too. Second, in addition to comparing MAX versus TAF, we also assessed Medicare fee-for-service claims, which are commonly used in high-quality studies and may indicate the data quality that can be achieved by claims.<sup>19-24</sup> Finally, we separately assessed both Medicaid fee-for-service claims and managed care encounter records and thus did not assume fee-for-service claims provide a benchmark.

## 2 | METHODS

### 2.1 | Data

We used the MAX other therapy and inpatient claims of a 100% sample of Medicaid beneficiaries for all states in 2011, except for Idaho, all states in 2012, and 28 and 17 states each in 2013 and 2014; we used TAF files for 100% of Medicaid beneficiaries in 2016 (see Appendix S1). We obtained demographic and enrollment information from the MAX and TAF personal summary files. We restricted our sample to those 18 years of age or older and flagged individuals as being enrolled in fee-for-service or

managed care programs based on plan type codes in the personal summary file.

For the same years, we used claims of a 20% simple random sample of Medicare fee-for-service beneficiaries. We used non-institutional claims, such as those from ambulance suppliers, from the Carrier file, outpatient claims, and admission data from the Medicare Provider Analysis and Review (MedPAR) file. Finally, we obtained demographic and enrollment data from the Master Beneficiary Summary File.

## 2.2 | Identification of out-of-hospital medical emergencies

We identified our sample of emergency ambulance services in the other therapy file for Medicaid and the Carrier file for Medicare using Healthcare Common Procedure Coding System (HCPCS)<sup>25</sup> codes A0427, A0429, and A0433 for all states. For California, we included the code X0030 for emergency transports in the Medicaid program.<sup>26,27</sup> Because some individuals might be enrolled in Medicaid for very brief periods, we required beneficiaries to be enrolled for at least 90 consecutive days following the date of the ambulance transport to make the Medicaid sample more comparable to the Medicare population. We required Medicare beneficiaries to be enrolled in Parts A and B during the month of their ambulance transport. Because Medicare is the primary payer for dual-eligible beneficiaries, we dropped the dual population from the Medicaid sample. Finally, for both Medicare and Medicaid, we dropped all ambulance claims within a day for beneficiaries who had multiple emergency modes of transport on that day to reduce complications with linkage to hospital claims (see Appendix S1).

## 2.3 | Data quality indicators

### 2.3.1 | Percent of emergency ambulance claims with pickup and drop-off modifier codes

Pickup and drop-off HCPCS modifier codes that identify categories of location types (residence, scene, hospital, etc.) are often

required in ambulance billing as only certain destinations are covered.<sup>28</sup> For research, these codes are particularly helpful for linkage to other health care services, such as nursing homes and emergency departments.<sup>29</sup> We created an indicator for each insurance program and state representing the percent of emergency ambulance service claims with a valid pickup and drop-off modifier (Table 1).

### 2.3.2 | Percent of emergency ambulance claims with mileage

The HCPCS code A0425, A0380 (Medicaid only), and A0390 (Medicaid only) are used for submitting mileage information on a claim and often required for mileage-based payment schedules, which are common.<sup>28,30,31</sup> In California, the Medicaid billing policy also used the code X0034 for mileage through 2016.<sup>26,27</sup> For research purposes, mileage is an important variable for studies on out-of-hospital medical emergencies like cardiac arrest, trauma, acute myocardial infarction, and stroke to assess or adjust for the distance traveled between pickup and drop-off locations.<sup>32</sup> Distance is also a common instrumental variable in health services research that aims to make a causal inference. This indicator shows the percent of emergency ambulance service claims that submitted mileage information for each insurance program and state.

### 2.3.3 | Percent of emergency ambulance claims that successfully linked to a hospital visit

Linking ambulance transports to hospital claims is crucial for obtaining diagnosis, procedure, and outcome information, as well as for studying hospital destination decisions. In Medicare and Medicaid claims, emergency department visits are rolled up into inpatient claims if the patient is admitted, or outpatient claims otherwise. This third indicator reported the percentage of emergency ambulance claims that successfully linked to a hospital for either admission or outpatient services, by state and insurance program. Transports were linked to hospital claims up to 2 days after the ride to allow for late-night transports and potential date errors, with prioritization given to admissions over outpatient

**TABLE 1** Definitions of data quality indicators

Indicator	Definition
Pickup/drop-off modifier codes	Percent of emergency ambulance claims with a valid pickup and drop-off modifier code
Ambulance mileage information	Percent of emergency ambulance claims with mileage information
Ambulance-hospital linkage	Percent of emergency ambulance claims that successfully linked to a hospital claim for either admission or outpatient services
External cause of injury reported on inpatient claim	Percent of linked ambulance and hospital admission trauma cases that reported an external cause of injury code (ICD-9CM or ICD-10CM e-code) in any diagnosis column
External cause of injury reported on outpatient claim	Percent of linked ambulance and outpatient visit trauma cases that reported an external cause of injury code (ICD-9CM or ICD-10CM e-code) in any diagnosis column
Reported versus expected in-hospital mortality	Comparison of the expected versus reported in-hospital deaths by injury severity score <sup>a</sup>

Abbreviations: ICD-9CM, International Classification of Diseases, Ninth Revision, Clinical Modification; ICD-10CM, Tenth Revision, Clinical Modification.

<sup>a</sup>Injury severity scores were calculated as the sum of squares of the three highest Abbreviated Injury Scale scores.

care. For the linked ambulance and hospital admission cases, we included Medicaid beneficiaries who were expired at discharge even though they were not enrolled for at least 90 consecutive days following the date of the ambulance transport, which we otherwise required for all observations.

### 2.3.4 | Percent of admissions with an external cause of injury code

The majority of states require the reporting of International Classification of Diseases (ICD) external cause of injury codes to their statewide hospital discharge data systems.<sup>33,34</sup> These codes are reported alongside the nature of injury codes (fractures, head trauma) to indicate the non-medical cause of injury (fall, car crash). External cause codes are crucial for public health research and policy because they identify targets for intervention.<sup>19</sup>

For ambulance transports that were successfully linked to hospital admission claims, we first identified hospitals claims with a valid nature of injury code in the first three diagnosis code columns (ICD-9CM for MAX or ICD-10CM for TAF) (see Appendix S1).<sup>35</sup> Then, we created an indicator that reports the percent of linked ambulance and hospital admission trauma cases that reported an e-code in any diagnosis column, by state and insurance program.

### 2.3.5 | Percent of outpatient claims with an external cause of injury code

Similar to the admissions indicator for external cause codes, this indicator reports the percent of linked ambulance and outpatient visit trauma cases that reported an e-code in any diagnosis column, by state and insurance program.

### 2.3.6 | Reported versus expected in-hospital mortality

Mortality is a primary outcome in many studies.<sup>36-38</sup> Medicare provides reliable beneficiary death information that is validated against the Social Security Administration's data, and in fact, accounts for over 99% of deaths among people 65 years of age and over.<sup>39</sup> In contrast, Medicaid death data in the personal summary file may be under-reported and unreliable,<sup>40</sup> making it difficult to study mortality. However, a potential second source of some death information is the hospital discharge status on admission claims. For Medicare, we found the hospital discharge status to be highly accurate; 94.2% of injury cases in our sample that had a death date during their hospital stay had a hospital discharge status indicating death (see Appendix S1). We assessed the usability of this status in Medicaid by first assigning each case an injury severity score, then estimating the proportion of expected deaths by injury severity score, and finally comparing the expected versus reported in-hospital deaths.

We started by assigning injury severity scores to all individuals in our linked ambulance-inpatient Medicare and Medicaid cases, using ICD Programs for Injury Categorization (ICDPIC) software.<sup>41,42</sup> Specifically, we computed New Injury Severity Scores (NISS), which are the sum of squares of the three highest Abbreviated Injury Scale (AIS) scores (i.e., the three most severe injuries).<sup>43-48</sup> The AIS scores the severity of injuries on a one to six scale by body region based on ICD codes; we used the ICDPIC option to convert AIS scores of six to five before calculating the NISS.<sup>49</sup>

We then used a logit model to regress death on injury severity, specified as a categorical variable (bins 1-8, 9-15, 16-24, 25-40, and 41+<sup>50</sup>), age, sex, race, and state and year fixed effects in the Medicare population. We limited the sample of cases in this model to include only individuals aged below 74 years to obtain a better estimation of the age coefficients in the younger Medicaid population. We conducted an in-sample and out-of-sample validation of this model within Medicare (see Appendix S1).

Finally, we used the fitted parameters from the Medicare model to predict mortality rates for the Medicaid population, separately for FFS and managed care beneficiaries. We limited the Medicaid population to only individuals between the ages of 50 and 64 years to minimize extrapolation of the age associations from the older Medicare population. Within each injury severity bin, we compared the proportion of expected deaths, predicted from the model, with the observed reported deaths.

## 3 | RESULTS

The total number of emergency ambulance transports in our final analytical samples for the MAX FFS, MAX managed care, TAF FFS, and TAF managed care datasets were 7,049,225, 8,593,538, 4,321,233, and 4,340,765, respectively. For Medicare FFS, there were 6,162,975 emergency ambulance claims for the years 2011-2014 and 1,597,341 in the year 2016. States were not included in the analysis if they submitted too few claims to meet the minimum cell size requirements of our data use agreement with CMS. See Appendix S1 for sample flowcharts.

### 3.1 | Percent of emergency ambulance claims with pickup and drop-off modifier codes

Among Medicare FFS claims, all states reported pickup and drop-off modifier codes in 100% of cases in our sample (Tables 2 and 3). Overall, 29 FFS and 26 managed care states in TAF had statistically significant improvements over MAX (Table 3). In MAX, 27 FFS and 21 managed care state programs reported modifier codes in 90% or more cases; in TAF, these numbers increased to 36 FFS and 29 managed care programs. Twenty states reported modifier codes in at least 90% of transports for both MAX FFS and managed care programs; this number increased to 27 states in the TAF dataset. Twenty states had reported below 50% in either

TABLE 2 MAX performance on data quality indicators by state and insurance program (percentages shown) [Color table can be viewed at wileyonlinelibrary.com]

State	Pickup/drop-off codes			Mileage information			Hospital linkage			Inpatient e-codes			Outpatient e-codes		
	MAX FFS	MAX MC	MCARE FFS	MAX FFS	MAX MC	MCARE FFS	MAX FFS	MAX MC	MCARE FFS	MAX FFS	MAX MC	MCARE FFS	MAX FFS	MAX MC	MCARE FFS
AL	100	**	100	100	**	99	96	**	95	32	**	87	28	**	88
AK	2	**	100	88	**	97	93	**	94	**	**	93	12	**	92
AZ	49	72	100	90	99	100	89	97	96	35	2	97	13	9	98
AR	0	**	100	100	99	99	96	88	95	**	0	95	12	12	91
CA	1	42	100	97	95	100	93	91	95	0	4	91	11	12	95
CO	96	92	100	99	100	100	96	**	96	**	**	97	9	**	97
CT	98	100	100	51	73	99	97	90	96	68	21	95	26	9	97
DC	94	79	100	99	99	100	89	93	91	0	**	75	13	9	79
DE	90	94	100	95	30	99	90	98	94	**	**	94	**	2	93
FL	100	99	100	0	7	100	95	83	95	30	29	95	40	27	95
GA	99	99	100	24	28	99	93	93	94	8	3	91	4	5	96
HI	47	**	100	4	**	100	96	**	87	0	**	95	5	**	92
ID	59	**	100	99	**	99	98	**	96	0	**	91	0	**	92
IL	95	95	100	97	98	97	96	80	96	0	0	92	0	0	93
IN	96	73	100	94	96	98	98	94	96	26	16	84	14	22	88
IA	34	56	100	INC	INC	100	98	95	96	11	17	88	9	11	91
KS	70	67	100	96	95	98	96	96	96	**	**	90	3	2	92
KY	100	100	100	97	98	100	97	97	95	81	48	86	31	26	86
LA	99	92	100	99	99	99	95	96	94	**	23	90	**	3	89
ME	90	**	100	98	**	99	83	**	96	27	**	95	6	**	95
MD	47	46	100	0	0	92	94	87	96	**	5	95	9	8	96
MA	97	99	100	100	100	100	97	96	97	8	10	94	10	10	96
MI	1	0	100	99	100	99	96	98	96	24	16	88	11	10	90
MN	100	99	100	100	100	99	98	96	96	1	5	93	5	8	91
MS	100	100	100	37	51	99	82	96	95	**	5	92	**	17	94
MO	3	44	100	88	93	99	100	98	96	1	6	97	10	10	98
MT	78	**	100	31	**	96	92	**	96	0	**	89	13	**	90
NE	98	0	100	40	98	97	96	97	97	**	0	96	2	0	98
NV	98	96	100	99	99	100	93	97	94	0	0	93	5	9	93
NH	1	**	100	96	**	99	98	**	96	**	**	89	2	**	89
NJ	0	**	100	98	97	98	91	93	96	**	3	93	12	6	90
NM	98	99	100	33	39	92	93	93	94	23	21	94	14	14	93

TABLE 2 (Continued)

State	Pickup/drop-off codes			Mileage information			Hospital linkage			Inpatient e-codes			Outpatient e-codes		
	MAX FFS	MAX MC	MCARE FFS	MAX FFS	MAX MC	MCARE FFS	MAX FFS	MAX MC	MCARE FFS	MAX FFS	MAX MC	MCARE FFS	MAX FFS	MAX MC	MCARE FFS
NY	91	85	100	57	76	99	94	94	95	4	5	94	10	14	95
NC	0	0	100	38	40	99	96	96	96	1	**	95	10	9	97
ND	1	**	100	100	100	99	92	89	96	**	**	84	11	**	90
OH	100	95	100	99	99	99	96	96	96	2	4	80	1	1	82
OK	100	100	100	98	98	99	97	98	95	45	37	91	13	13	92
OR	93	90	100	63	68	99	95	96	96	9	10	89	7	7	87
PA	100	95	100	34	67	99	100	96	95	23	23	92	13	15	92
RI	100	100	100	98	97	100	94	96	95	0	**	93	21	**	99
SC	100	37	100	97	97	98	98	91	95	0	0	96	0	**	99
SD	1	**	100	98	**	99	92	**	95	2	**	72	18	**	77
TN	**	95	100	**	96	99	**	97	95	**	27	94	**	20	96
TX	61	60	100	99	100	97	96	94	95	0	1	86	0	7	89
UT	100	100	100	100	99	99	76	75	96	19	**	95	19	27	94
VT	1	**	100	99	**	100	98	**	96	6	**	84	3	**	95
VA	25	31	100	100	64	100	96	94	96	73	23	94	39	25	92
WA	89	62	100	99	100	100	97	89	95	3	**	96	9	**	88
WV	100	100	100	100	98	99	97	85	95	81	**	79	14	**	82
WI	21	44	100	99	95	99	96	97	96	63	19	97	51	14	99
WY	4	**	100	97	**	98	97	**	96	**	**	85	**	**	88

Difference between Medicare and Medicaid reporting levels<sup>a</sup>

High	≤10%
Medium	10% < x ≤ 20%
Low	20% < x ≤ 50%
Very low	>50%

Note: \*\* is placed in cells that contained too few claims to meet the minimum cell size requirements of our data use agreement with the Centers for Medicare and Medicaid Services. INC's are placed in cells that reported a mileage of 0 or blank for all ambulance claims.

Abbreviations: MAX FFS, Medicaid Analytic eXtract managed care; MCARE FFS, Medicare fee-for-service; MAX MC, Medicaid Analytic eXtract managed care; MCARE FFS, Medicare fee-for-service; e-code, external cause of injury code.

<sup>a</sup>The data quality was indicated as high, medium, low, or very low if the percentage point difference with Medicare falls within the interval. We compared the same years (2011–2014) between MAX and Medicare.

TABLE 3 TAF performance on data quality indicators by state and insurance program (percentages shown) [Color table can be viewed at wileyonlinelibrary.com]

States	Pickup/drop-off codes			Mileage information			Hospital linkage			Inpatient e-codes			Outpatient e-codes		
	TAF FFS	TAF MC	MCARE FFS	TAF FFS	TAF MC	MCARE FFS	TAF FFS	TAF MC	MCARE FFS	TAF FFS	TAF MC	MCARE FFS	TAF FFS	TAF MC	MCARE FFS
AL	100	**	100	100	**	99	97	**	95	**	83	11	**	87	
AK	43	**	100	94	**	97	94	**	95	**	84	19	**	88	
AZ	63	99	100	98	99	100	44	93	96	1	89	15	13	91	
AR	1	**	100	100	**	99	89	**	96	**	87	4	**	95	
CA	4	44	100	78	96	100	85	86	96	6	93	17	18	96	
CO	95	100	100	100	100	100	93	87	96	22	88	16	24	92	
CT	100	**	100	**	**	99	89	**	96	**	90	10	**	96	
DC	100	96	100	100	96	100	87	62	92	5	91	11	12	90	
DE	93	100	100	3	2	99	95	97	96	**	82	**	6	93	
FL	100	100	100	76	41	100	89	97	96	5	87	4	6	94	
GA	100	100	100	96	92	100	96	89	95	3	88	5	9	94	
HI	100	100	100	100	100	100	88	94	92	**	87	**	10	88	
ID	95	100	100	100	100	99	90	94	96	0	88	23	29	92	
IL	93	100	100	99	93	98	90	94	96	5	87	11	13	89	
IN	97	85	100	99	99	99	96	97	96	62	84	0	11	89	
IA	100	100	100	100	100	100	95	96	97	**	85	19	19	88	
KS	68	86	100	98	99	99	83	94	96	**	86	**	14	91	
KY	100	100	100	100	100	100	96	96	95	22	84	15	12	86	
LA	99	100	100	100	100	98	90	93	95	**	86	0	4	90	
ME	96	**	100	100	**	100	87	**	96	**	93	12	**	92	
MID	7	**	100	1	100	100	90	86	96	**	90	**	0	91	
MA	98	98	100	100	100	100	92	93	96	14	90	16	9	95	
MI	2	1	100	100	100	99	95	98	96	29	86	12	10	86	
MN	100	100	100	100	100	99	92	62	97	3	90	11	9	89	
MS	100	100	100	26	41	99	76	85	96	**	88	**	13	93	
MO	3	59	100	99	99	100	100	97	96	13	94	15	16	97	
MT	82	**	100	83	**	99	90	**	96	**	88	18	**	90	
NE	99	100	100	86	98	97	92	95	97	**	93	**	**	96	
NV	100	100	100	99	100	100	94	95	95	4	86	14	10	91	
NH	11	98	100	85	87	100	84	89	97	74	87	12	16	91	
NJ	**	0	100	98	97	99	96	96	96	1	87	7	9	93	
NM	99	99	100	21	29	92	70	90	95	44	91	6	9	91	

(Continues)

TABLE 3 (Continued)

States	Pickup/drop-off codes			Mileage information			Hospital linkage			Inpatient e-codes			Outpatient e-codes		
	TAF FFS	TAF MC	MCARE FFS	TAF FFS	TAF MC	MCARE FFS	TAF FFS	TAF MC	MCARE FFS	TAF FFS	TAF MC	MCARE FFS	TAF FFS	TAF MC	MCARE FFS
NY	100	74	100	41	98	99	93	90	95	5	**	93	13	12	95
NC	100	**	100	37	**	99	99	**	95	**	**	89	4	**	95
ND	11	**	100	100	**	99	86	**	97	68	**	91	12	**	92
OH	100	100	100	100	100	99	93	96	96	**	0	81	11	11	81
OK	100	**	100	100	**	99	93	**	96	3	**	86	11	**	93
OR	99	97	100	98	100	100	87	97	95	**	**	86	7	6	91
PA	100	100	100	**	**	99	98	93	96	6	10	87	7	12	91
RI	100	100	100	99	100	100	90	95	95	0	0	94	19	1	97
SC	100	50	100	100	100	100	88	93	95	**	**	91	10	10	97
SD	90	**	100	98	**	99	79	**	96	**	**	82	**	**	85
TN	99	99	100	**	0	99	56	94	95	**	3	88	**	14	93
TX	100	100	100	100	100	99	84	93	95	**	1	85	10	10	91
UT	100	100	100	100	95	100	75	79	96	73	**	91	17	**	95
VT	28	**	100	98	**	100	93	**	96	**	**	91	18	**	85
VA	31	**	100	100	**	100	87	**	96	10	**	88	11	**	92
WA	100	100	100	99	100	100	92	92	96	**	1	91	16	14	90
WV	99	100	100	100	100	99	97	96	95	62	37	76	22	17	77
WI	100	100	100	100	100	99	96	96	96	**	**	89	8	9	95
WY	27	**	100	99	**	99	94	**	95	**	**	84	8	**	86

Difference between Medicare and Medicaid reporting levels<sup>a</sup>



Note: \*\* is placed in cells that contained too few claims to meet the minimum cell size requirements of our data use agreement with the Centers for Medicare and Medicaid Services. Abbreviations: MCARE FFS, Medicare fee-for-service; e-code, external cause of injury code; TAF FFS, Transformed Medicaid Statistical Information System (T-MSIS) Analytic Files fee-for-service; TAF MC, T-MSIS Analytic Files managed care.

<sup>a</sup>The data quality was indicated as high, medium, low, or very low if the percentage point difference with Medicare falls within the interval. We compared the same year (2016) between TAF and Medicare. When comparing rates strictly between MAX and TAF, a cell with blue borders represents a statistically significant increase from MAX and a cell with red borders represents a statistically significant decrease from MAX, at an alpha level of 0.05.

[Correction added on 12 March 2022, after first online publication: Table 3 has been corrected in this version.]



MAX FFS or managed care; this number decreased to 13 states in TAF.

### 3.2 | Percent of emergency ambulance claims with mileage

The lowest reporting rate of mileage among Medicare FFS was 92%, with most states at 99% or higher. For MAX, the mileage reporting rates were within 10 percentage points of the state's Medicare reporting rate for 25 states overall, 35 state FFS programs, and 28 state managed care programs. These numbers improved in TAF to 27 states overall, 37 FFS programs, and 31 managed care programs. Eleven states had reporting rates below 50 percentage points of their Medicare rates in either their MAX FFS or managed care data. That number decreased to eight in TAF. Overall, 30 FFS and 20 managed care states in TAF had statistically significant improvements over their prior MAX rates.

### 3.3 | Percent of emergency ambulance claims that successfully linked to a hospital visit

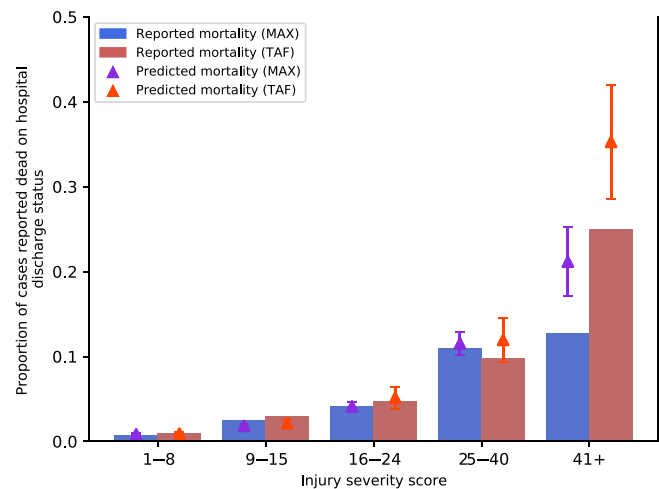
Medicare ambulance claims were linked to hospital claims in at least 95% of cases in almost all states. For MAX FFS claims, all but three states were within 10 percentage points of their state's Medicare linkage rates, and the lowest linkage rate was 76%. Similarly, all but three states' MAX managed care data had rates within 10 percentage points of their state's Medicare linkage rate. However, we found lower quality in the TAF FFS data. Many states in FFS and managed care had statistically significant decreases in linkage rates going from MAX to TAF, but most states in TAF still performed within 10 percentage points of Medicare.

### 3.4 | Percent of admissions and outpatient claims with an external cause of injury code

Among linked Medicare ambulance-hospital claims, most states had an 85% or higher reporting rate of the external cause of injury code on the linked admission or outpatient claim. However, across states, both FFS and managed care in MAX and TAF had extremely poor reporting rates of external causes on both admission data and outpatient claims, with many states reporting rates in the single digits.

### 3.5 | Reported versus expected in-hospital mortality

Figure 1 shows the comparison between expected mortality in the 50–64 years of age Medicaid FFS population, based on our model of death in younger Medicare beneficiaries, and reported mortality, by injury severity score band and MAX versus TAF. We found little



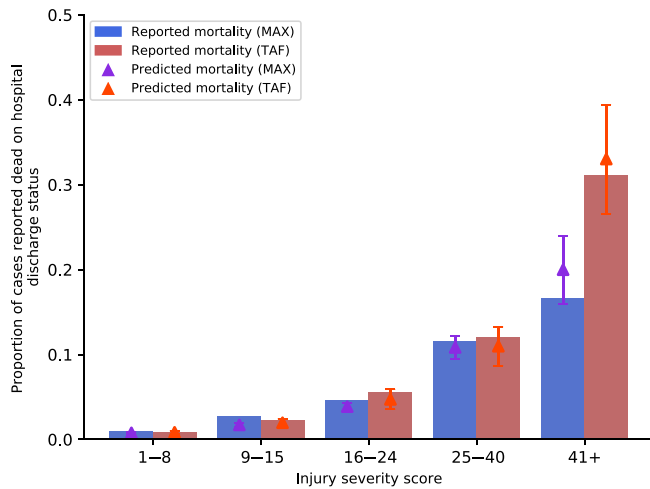
**FIGURE 1** Reported versus predicted mortality in Medicaid fee-for-service cases, by injury severity score (95% confidence intervals for predicted mortality shown). MAX, Medicaid Analytic eXtract, TAF, Transformed Medicaid Statistical Information System (T-MSIS) Analytic Files. The number of MAX fee-for-service observations in the new injury severity score bins 1–8, 9–15, 16–24, 25–40, and 41+ are 7641, 5499, 2155, 713, and 78, respectively. The number of TAF fee-for-service observations in the new injury severity score bins 1–8, 9–15, 16–24, 25–40, and 41+ are 4508, 1373, 606, 347, and 84, respectively. We trained the logit model using those aged 65–73 in the Medicare population and predicted mortality on those aged 50–64 in the Medicaid population. 95% confidence intervals for predictions in each bin are shown [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

difference between the expected and reported rates for both insurance programs in all bins except the highest. Specifically, the expected mortality rate was 8.3 percentage points higher in MAX and 10.3 percentage points higher in TAF than the reported mortality rate for cases with the most severe injuries (41+). Figure 2 shows a similar plot but for managed care cases. Here, we did not observe significant differences between the expected and observed mortality rates in any injury band of MAX or TAF data.

The Appendix S1 includes additional related analyses of our prediction model, including plots that demonstrate the model produced accurate in- and out-of-sample predictions.

## 4 | DISCUSSION

Concerns about the quality of national Medicaid data have been a longstanding hindrance to cross-state policy research. We conducted the first national comparison of legacy Medicaid data (MAX), new generation Medicaid data (TAF), and Medicare claims data on the performance of data analytic tasks within a realistic research scenario. Specifically, we created data quality indicators that assessed the feasibility of creating episodes of care that start with an emergency ambulance transport and end with hospital death or survival to discharge, which required linking information across claims, generating injury severity scores, and comparing expected versus reported deaths. Our findings challenge common



**FIGURE 2** Reported versus predicted mortality in Medicaid managed care cases, by injury severity score (95% confidence intervals for predicted mortality shown). MAX, Medicaid Analytic eXtract, TAF, Transformed Medicaid Statistical Information System (T-MSIS) Analytic Files. The number of MAX managed care observations in the new injury severity score bins 1-8, 9-15, 16-24, 25-40, and 41+ are 8619, 5256, 2040, 602, and 60, respectively. The number of TAF managed care observations in the new injury severity score bins 1-8, 9-15, 16-24, 25-40, and 41+ are 7132, 2123, 874, 471, and 106, respectively. We trained the logit model using those aged 65-73 in the Medicare population and predicted mortality on those aged 50-64 in the Medicaid population. 95% confidence intervals for predictions in each bin are shown [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

assumptions. First, though TAF is expected to address the shortcomings of MAX, and it has on some indicators, TAF exhibited serious data quality problems. Second, though Medicaid FFS claims are commonly used as quality benchmarks for managed care records,<sup>5-8,51</sup> we did not find that these performed consistently better than Medicaid managed care records. However, as the performance of Medicare claims in our analysis demonstrates, CMS claims data can achieve high quality and further improvements to TAF should be vigorously pursued.

Across all six indicators, Medicare claims exhibited high levels of complete information, including on pre-hospital care, external causes of injuries, and death, and also consistency between claims that allowed linkage. In contrast, in MAX data, states varied widely in reporting of pre-hospital information and almost all states had extremely poor reporting of external injury cause codes. However, the linkage between the ambulance and hospital claims was successful in most states. With TAF, reporting of pre-hospital information substantially improved, but states continued to have poor reporting of external injury cause codes. This is a major shortcoming of the TAF data because injuries are the leading cause of death among the under 65 years population and external causes of injury codes are crucial to injury prevention policy. In 2010, CMS resolved this issue with Medicare claims by including separate ICD diagnosis fields for external causes<sup>19,29</sup>; a similar change may help here.

Medicaid fee-for-service claims did not perform consistently better than managed care records on any indicator; future research

should not rely on this assumption. Furthermore, FFS claims in both MAX and TAF data appear to underreport in-hospital deaths in high-severity injury cases, while managed care records in both MAX and TAF report in-hospital deaths within the expected range. While CMS considers hospital discharge status in its reports, it has not flagged this variable as anomalous in any state.<sup>9</sup>

Our findings identify parts of national Medicaid data that need to be improved, but also provide a state-by-state and FFS versus managed care plan breakdown that can help determine segments of data that may be used for high-quality inference. Though much of the discussion on Medicaid data quality has focused on problems with managed care, our analysis shows that states often perform similarly in both managed care and FFS programs, indicating the underlying issues may be related to common state elements, including Medicaid program structure, billing and coverage policies, and information technology infrastructure. More generally, performance on some indicators was poor across states, suggesting some data issues might not be state-specific but rather are broader national issues, perhaps related to CMS data infrastructure.

Our analysis should not be viewed as an investigation of improper billing; our goal was to assess data usability for research. Thus, missing data, while problematic for research, may be consistent with state billing policy. Most of our study limitations pertain to the assessment of death reporting on the Medicaid hospital discharge status. First, we may have underestimated injury severity scores if diagnosis codes were not completely reported in the Medicaid inpatient data. Though the number of diagnosis codes reported may be associated with the type of Medicaid payment policy, the distributions of computed injury severity scores between FFS and managed care were similar (Appendix Figures 11 and 12 ). If we underestimated injury severity for individuals who died in the hospital, then some individuals who should have been in higher injury severity bands may be incorrectly placed in lower injury severity bands. This might explain part of the gap that we observe between predicted and reported deaths of individuals with the highest injury severity. Second, we extrapolated the role of age in the association between injury severity and mortality from 65-73 year-old Medicare beneficiaries to 50-64 year-old Medicaid beneficiaries. However, we explored this further through a sensitivity analysis in which we trained our same model on an older Medicare cohort (74-85 years of age) and predicted mortality for a younger Medicare cohort (65-73 years of age) (see Appendix S1). We found that our model predicted the reported rates for the younger population almost perfectly, providing evidence that the gaps found in the highest bin of Figure 1 may be due to underreported death information rather than an effect of age. Nonetheless, other studies suggest there may be a non-linear feature in the age effect that was unaccounted for and might have led us to overestimate the expected mortality rate in the younger population.<sup>52,53</sup> Finally, our model did not account for chronic conditions, as we expected high levels of incompleteness in these measures in the Medicaid population. This could also have led to an overestimation of the expected death rates.

About half of our analysis focused on injuries, a leading cause of death in the Medicaid population, but these findings

too have implications for other health conditions. For example, our findings on death reporting are unlikely to be specific to injuries. Also, we created injury severity scores using diagnosis codes; comorbidity scores also rely on the completeness of diagnosis reporting. We were unable to create injury severity scores for outpatient claims because too few diagnosis codes were provided, and this same issue would make it challenging to create comorbidity scores too.

Improvement efforts for Medicaid data could target three key areas that affect a wide range of research topics. First, complete diagnosis coding is crucial for almost any area of health research to identify clinical conditions, estimate comorbidities, measure disease severity, and study nonfatal outcomes. Currently, far fewer diagnosis code fields are included in the inpatient files of Medicaid MAX (nine codes) and TAF (admitting plus 12 codes) compared with Medicare's MedPAR file (admitting plus 25 codes). The number of diagnosis code fields is even lower in Medicaid outpatient claims and encounter records, which include emergency department visits (two codes vs. Medicare's principal plus 25 codes). One source of this issue for Medicaid MAX and TAF may be the lumping of many different types of non-inpatient services into a single file, which may prioritize fields that are common across services, as opposed to Medicare's separate institutional outpatient file.

Second, death is a key outcome in many studies but is incompletely reported in Medicaid data. Medicare, on the other hand, has complete death information, in part because it validates this data with the Social Security Administration, which has a financial incentive and reporting process to document beneficiary deaths. One option for Medicaid that is available at a high cost to researchers is the Centers for Disease Control and Prevention's National Death Index, a repository of the nation's death certificates. If CMS used these data to provide complete death information in the Medicaid files, the research potential of these datasets would significantly change.

Finally, though Medicaid state payment policies for FFS programs are available, such as from the Medicaid and CHIP Payment and Access Commission, a similar compilation for managed care plans in each state would provide researchers information on coverage and billing policies, which are crucial to understanding the incentives and rules that drive the data generation process.

Medicaid claims and encounter records have the potential to serve as the richest data available for studying health and health care for a large, young, and marginalized American population. They can uniquely provide information on patient history, utilization of a range of services including pre-hospital care and post-acute care, and long-term outcomes. Further, the federal-state structure of the Medicaid program creates a unique opportunity to compare the impacts of varying state policies. However, for robust work to happen, higher-quality research data is needed. Our analysis should serve as a useful tool for understanding some of the more and less reliable dimensions of Medicaid data within an actual research context, unlike the macroscopic reports that

currently exist. Good policy making requires a solid evidence base; improving Medicaid data quality should be a high priority for CMS and policy makers.

## ACKNOWLEDGMENTS

We thank Nadia Ghazali for programming and data management support.

## ORCID

Prachi Sanghavi  <https://orcid.org/0000-0002-0738-3193>

## REFERENCES

1. Truffer CJ, Rennie KE, Wilson L, Eckstein ET. Actuarial report on the financial outlook for Medicaid. Centers for Medicare & Medicaid Services. 2018 2018:82.
2. State Overviews|Medicaid. Medicaid.gov - Keeping America Healthy. 2020. <https://www.medicaid.gov/state-overviews/index.html>. Accessed May 4, 2020.
3. Ruttner L, Borck R, Nysenbaum J, Williams S. Guide to MAX data. Mathematica; 2015. 10. <https://www.mathematica.org/our-publications-and-findings/publications/guide-to-max-data>. Accessed April 27, 2020.
4. Medicaid analytic EXtract files (MAX) user guide. Chronic Condition Data Warehouse. 2020. 49.
5. Byrd VLH, Dodd AH. Assessing the usability of encounter data for enrollees in comprehensive managed care 2010-2011. Mathematica. 2015:12. <https://www.mathematica.org/our-publications-and-findings/publications/assessing-the-usability-of-encounter-data-for-enrollees-in-comprehensive-managed-care-2010-2011>. Accessed June 15, 2020.
6. Dodd AH, Nysenbaum J, Zlatinov A. Assessing the usability of the MAX 2007 inpatient and prescription encounter data for enrollees in comprehensive managed care. Mathematica Policy Research. 2012. <https://ideas.repec.org/p/mprr/mprrres/d5680a1d9a11400da54941acef189dd4.html>. Accessed June 22, 2020.
7. Byrd VLH, Dodd AH. Assessing the usability of MAX 2008 encounter data for comprehensive managed care. Mathematica. 2013. 19. <https://www.mathematica.org/our-publications-and-findings/publications/assessing-the-usability-of-max-2008-encounter-data-for-comprehensive-managed-care>. Accessed June 15, 2020.
8. Byrd VLH, Dodd AH. Assessing the usability of encounter data for enrollees in comprehensive managed care across MAX 2007-2009. Mathematica. 2012:11. <https://www.mathematica.org/our-publications-and-findings/publications/assessing-the-usability-of-encounter-data-for-enrollees-in-comprehensive-managed-care-across-max-20072009>. Accessed June 15, 2020.
9. Medicaid Analytic eXtract (MAX) general information. Center for Medicare and Medicaid Services. 2020. <https://www.cms.gov/Research-Statistics-Data-and-Systems/Computer-Data-and-Systems/MedicaidDataSourcesGenInfo/MAXGeneralInformation>. Accessed February 15, 2020.
10. Transformed Medicaid Statistical Information System (T-MSIS) analytic files (TAF). Center for Medicare and Medicaid Services. <https://www.medicaid.gov/medicaid/data-systems/macbis/medicaid-chip-research-files/transformed-medicaid-statistical-information-system-t-msis-analytic-files-taf/index.html>. Accessed October 12, 2020.
11. Medicaid and CHIP T-MSIS analytic files data release. Center for Medicare and Medicaid Services. 2019. <https://www.cms.gov/newsroom/fact-sheets/medicaid-and-chip-t-msis-analytic-files-data-release>. Accessed October 12, 2020.
12. Additional Medicaid and CHIP T-MSIS analytic files data release. 2020. <https://www.cms.gov/newsroom/fact-sheets/>

- additional-medicaid-and-chip-t-msis-analytic-files-data-release. Accessed October 12, 2020.
13. Mangum A, Proctor K, Parker J. Missing and invalid diagnosis codes in 2017. *Mathematica*. 2019. 13.
  14. Baller J, Proctor K, Parker J. Usability of procedure codes in 2017. *Mathematica*. 2019. 14.
  15. Liu S, Nakajima R, Proctor K, Parker J. Missing eligibility group codes in 2017. 2019. <https://www.mathematica.org/our-publications-and-findings/publications/missing-eligibility-group-codes-in-2017-brief>. Accessed October 12, 2020.
  16. Khan T, Weizenegger E, Proctor K, Parker J. Completeness of the CHIP and dual status codes in 2017. 2019. <https://www.mathematica.org/our-publications-and-findings/publications/completeness-of-the-chip-and-dual-status-codes-in-2017-brief>. Accessed October 12, 2020.
  17. Baller J, Arguello A, Geibel MA, Natzke B, Proctor K, Parker J. The volume of encounter claim records from comprehensive managed care organizations in 2017. *Mathematica*. 2019. 21.
  18. 10 leading causes of death by age group, United States - 2018. Centers for Disease Control and Prevention. 2018. <https://www.cdc.gov/injury/wisqars/LeadingCauses.html>. Accessed April 28, 2020.
  19. Sanghavi P, Pan S, Caudry D. Assessment of nursing home reporting of major injury falls for quality measurement on nursing home compare. *Health Serv Res*. 2020;55(2):201-210. doi:10.1111/1475-6773.13247
  20. Gorges RJ, Sanghavi P, Konetzka RT. A national examination of long-term care setting, outcomes, and disparities among elderly dual eligibles. *Health Aff*. 2019;38(7):1110-1118. doi:10.1377/hlthaff.2018.05409
  21. Konetzka RT, Stuart EA, Werner RM. The effect of integration of hospitals and post-acute care providers on Medicare payment and patient outcomes. *J Health Econ*. 2018;61:244-258. doi:10.1016/j.jhealeco.2018.01.005
  22. Ayanian JZ, Landon BE, Newhouse JP, Zaslavsky AM. Racial and ethnic disparities among enrollees in Medicare advantage plans. *N Engl J Med*. 2014;371(24):2288-2297. doi:10.1056/NEJMsa1407273
  23. Death information in the research identifiable Medicare data. Research Data Assistance Center. 2018. <https://www.resdac.org/articles/death-information-research-identifiable-medicare-data>. Accessed October 13, 2020.
  24. Valid date of death switch. Research Data Assistance Center. <https://www.resdac.org/cms-data/variables/valid-date-death-switch>. Accessed October 13, 2020.
  25. HCPCS - General Information. Center for Medicare and Medicaid Services. 2020. <https://www.cms.gov/Medicare/Coding/MedHCPCSGenInfo/index?redirect=/MedHCPCSGenInfo>. Accessed May 13, 2020.
  26. HIPAA: Medical Transportation Code Conversion. Department of Healthcare Services Medi-Cal. 2016. [https://files.medi-cal.ca.gov/pubsdoco/hipaa/hipaaqa\\_code\\_conversions.asp](https://files.medi-cal.ca.gov/pubsdoco/hipaa/hipaaqa_code_conversions.asp). Accessed May 19, 2020.
  27. California medical transportation code conversion table. Department of Healthcare Services. 2016. [https://files.medi-cal.ca.gov/pubsdoco/hipaa/tables/Medical\\_Transportation\\_Code\\_Conversion\\_Table\\_24524\\_08.pdf](https://files.medi-cal.ca.gov/pubsdoco/hipaa/tables/Medical_Transportation_Code_Conversion_Table_24524_08.pdf). Accessed January 12, 2020.
  28. Medicare claims processing manual. Center for Medicare and Medicaid Services. <https://www.cms.gov/Regulations-and-Guidance/Guidance/Manuals/Internet-Only-Manuals-IOMs-Items/CMS018912>. Accessed January 5, 2020.
  29. Sanghavi PB, Jena AP, Newhouse JM, Zaslavsky A. Outcomes of basic versus advanced life support for out-of-hospital medical emergencies. *Ann Intern Med*. 2015;163:681-690. doi:10.7326/M15-0557
  30. Medi-Cal Code Conversions NewsFlash: Medical Transportation Code Conversion: Policy Overview. [https://files.medi-cal.ca.gov/pubsdoco/hipaa/articles/codeconversionsnews\\_24524\\_07.aspx](https://files.medi-cal.ca.gov/pubsdoco/hipaa/articles/codeconversionsnews_24524_07.aspx). Accessed July 14, 2021.
  31. CMS Manual System. 2009. <https://www.cms.gov/Regulations-and-Guidance/Guidance/Transmittals/downloads/R1821CP.pdf>. Accessed July 14, 2021.
  32. Sanghavi P, Jena AB, Newhouse JP, Zaslavsky AM. Outcomes after out-of-hospital cardiac arrest treated by basic vs advanced life support. *JAMA Intern Med*. 2015;175(2):196-204.
  33. Annett JL, Conn JM, Kohn M, Abellera J. How states are collecting and using cause of injury data: 2004 update to the 1997 report. American Public Health Association. 2005. 46.
  34. What are E-codes and why are they important? 2014. [https://ncdetect.org/files/2016/12/CCHI\\_E\\_CodeFactSheetJan2014.pdf](https://ncdetect.org/files/2016/12/CCHI_E_CodeFactSheetJan2014.pdf)
  35. HCUP external cause of injury code (E-code) evaluation report (Updated with 2013 HCUP Data). US Department of Health and Human Services. 2016. 13. <https://www.hcup-us.ahrq.gov/reports/methods/2016-03.pdf>
  36. MacKenzie EJ, Rivara FP, Jurkovich GJ, et al. A national evaluation of the effect of trauma-center care on mortality. *N Engl J Med*. 2006;354(4):366-378. doi:10.1056/NEJMsa052049
  37. Clay Mann N, Mullins RJ, Hedges JR, Rowland D, Arthur M, Zechnich AD. Mortality among seriously injured patients treated in remote rural trauma centers before and after implementation of a statewide trauma system. *Med Care*. 2001;39(7):643-653. doi:10.1097/00005650-200107000-00001
  38. Nathens AB, Jurkovich GJ, Cummings P, Rivara FP, Maier RV. The effect of organized systems of trauma care on motor vehicle crash mortality. *JAMA*. 2000;283(15):1990-1994. doi:10.1001/jama.283.15.1990
  39. Strengths and limitations of CMS administrative data in research. Research Data Assistance Center. 2018. <https://www.resdac.org/articles/strengths-and-limitations-cms-administrative-data-research>. Accessed October 2, 2020.
  40. Date of death (from Medicare EDB). Research Data Assistance Center. <https://www.resdac.org/cms-data/variables/date-death-medicare-edb>. Accessed October 2, 2020.
  41. Clark DE, Osler TM, Hahn DR. ICDPIC: Stata module to provide methods for translating international classification of diseases (Ninth Revision) Diagnosis Codes into Standard Injury Categories and/or Scores. Boston College Department of Economics. 2010. <https://ideas.repec.org/c/boc/bocode/s457028.html>. Accessed June 16, 2020.
  42. Clark DE, Black AW, Skavdahl DH, Hallagan LD. Open-access programs for injury categorization using ICD-9 or ICD-10. *Inj Epidemiol*. 2018;5(1):11. doi:10.1186/s40621-018-0149-8
  43. Osler T, Baker SP, Long W. A modification of the injury severity score that both improves accuracy and simplifies scoring. *J Trauma*. 1997;43(6):922-925; discussion 925-926. doi:10.1097/00005373-199712000-00009
  44. Brennenman FD, Boulanger BR, McLellan BA, Redelmeier DA. Measuring injury severity: time for a change? *J Trauma*. 1998;44(4):580-582. doi:10.1097/00005373-199804000-00003
  45. Sacco WJ, MacKenzie EJ, Champion HR, Davis EG, Buckman RF. Comparison of alternative methods for assessing injury severity based on anatomic descriptors. *J Trauma*. 1999;47(3):441-446; discussion 446-447. doi:10.1097/00005373-199909000-00001
  46. Stevenson M, Segui-Gomez M, Lescohier I, Di Scala C, McDonald-Smith G. An overview of the injury severity score and the new injury severity score. *Inj Prev*. 2001;7(1):10-13. doi:10.1136/ip.7.1.10
  47. Sears JM, Blannar L, Bowman SM. Predicting work-related disability and medical cost outcomes: a comparison of injury severity scoring methods. *Injury*. 2014;45(1):16-22. doi:10.1016/j.injury.2012.12.024
  48. Meredith JW, Evans G, Kilgo PD, et al. A comparison of the abilities of nine scoring algorithms in predicting mortality. *J Trauma*. 2002;53(4):621-628; discussion 628-629. doi:10.1097/00005373-200210000-00001
  49. Petrucelli E, States JD, Hames LN. The abbreviated injury scale: evolution, usage and future adaptability. *Accid Anal Prev*. 1981;13(1):29-35. doi:10.1016/0001-4575(81)90040-3
  50. Copes WS, Champion HR, Sacco WJ, Lawnick MM, Keast SL, Bain LW. The injury severity score revisited. *J Trauma Acute Care Surg*. 1988;28(1):69-77.
  51. Li Y, Zhu Y, Chen C, et al. Internal validation of Medicaid analytic eXtract (MAX) data capture for comprehensive managed care plan

- enrollees from 2007 to 2010. *Pharmacoepidemiol Drug Saf.* 2018; 27(10):1067-1076. doi:[10.1002/pds.4365](https://doi.org/10.1002/pds.4365)
52. Fatovich DM, Jacobs IG, Langford SA, Phillips M. The effect of age, severity, and mechanism of injury on risk of death from major trauma in Western Australia. *J Trauma Acute Care Surg.* 2013;74(2):647-651. doi:[10.1097/TA.0b013e3182788065](https://doi.org/10.1097/TA.0b013e3182788065)
53. Chiang W-K, Huang S-T, Chang W-H, Huang M-Y, Chien D-K, Tsai C-H. Mortality factors regarding the injury severity score in elderly trauma patients. *Int J Gerontol.* 2012;6(3):192-195. doi:[10.1016/j.ijge.2012.01.016](https://doi.org/10.1016/j.ijge.2012.01.016)

## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

**How to cite this article:** Nguyen JK, Sanghavi P. A national assessment of legacy versus new generation Medicaid data. *Health Serv Res.* 2022;57(4):944-956. doi:[10.1111/1475-6773.13937](https://doi.org/10.1111/1475-6773.13937)