



## Estimating Inpatient Cancer Care Costs in California: Methodology Detail

This paper describes in detail the methods used by Deborah Schrag, MD, MPH, (Department of Medicine, Harvard Medical School/ Dana-Farber Cancer Institute, Boston, MA) and her research team to estimate the cost of inpatient care in California using California Office of Statewide Health Planning and Development (OSHPD) data that were linked to California Cancer Registry (CCR) data. The ability to link California Cancer Registry data to other databases is critical for studies involving cancer spending, as it allows researchers to assess spending by cancer stage.

### Table 1. Summary of Data Sources

- ▶ **CCR:** All patients diagnosed with breast, lung, colorectal, or prostate cancer
- ▶ **CCR-OSHPD linked data:** All hospitalizations in California acute care hospitals among patients diagnosed with cancer and reported to the CCR (hospitalizations in chronic care hospitals and hospitals outside of California are not represented)

### Overview

OSHPD data were linked to the California Cancer Registry data by the California Department of Public Health to facilitate estimation of the costs incurred by patients with cancer. Costs for inpatient care were estimated for cancer patients in relationship to the month and year of diagnosis. OSHPD data report hospital charges, not costs, so the charge data was converted to cost data by applying a year- and facility-specific ratio of cost to charge (RCC). The RCC was derived by examining each hospital's hospital financial disclosure report submitted to OSHPD for each year of interest. One generalized RCC was used per year per facility.

The spending estimates reflect care received by persons with the most common cancers diagnosed in California: breast, lung, colorectal, and prostate. Only patients with a first primary cancer are included in these analyses. Patients diagnosed with cancer at all ages are included.

The report represents patients diagnosed with cancer in California in 2007, 2008, and 2009. These patients are tracked for vital status through the end

of 2011. All spending estimates have been adjusted for inflation to 2013 constant US dollars.

The annual total cost of the initial year from diagnosis was calculated by summing (not averaging) the monthly costs across all patients for the initial 12-month period.

It is important to note that costs in the initial year of diagnosis and in the last year of life may be mutually exclusive or may be overlapping. For example, if a patient survives three years from cancer diagnosis, these costs are non-overlapping. However, if a patient survives exactly a year, then the patient's first-year costs and last-year costs will be identical.

The principal advantage of this data source is that it includes all payers. Because inpatient care is the most expensive component of cancer care, OSHPD data provide a way to estimate costs for all California patients across the spectrum of age and insurance. The main disadvantage of using these data is that they include only inpatient costs.

# Mean Costs During First Year of Cancer Diagnosis

## Calculate Inpatient Costs Using CCR-OSHPD

**Step 1a: Use CCR-OSHPD PPD data only.** Only OSHPD's Patient Discharge Data (PDD) was used. Patient copayments are not included.

Denominators reflect inpatient costs for the entire CCR population for a given subgroup. For example, if there are 100 colorectal cancer patients and only 85 are hospitalized in the first year of diagnosis, the denominator for estimating means and medians includes all 100 patients with colorectal cancer. Patients who are not hospitalized contribute no costs. All patients are censored in the month of death. All patients contribute costs for any month in which they were alive for at least part of the month. For all of the base case analyses, the denominators used to estimate the mean cost of hospital spending include the entire cancer population of California. In other words, many patients who are never hospitalized contribute zeroes and bring down the estimates of mean spending. Mean spending among people actually hospitalized is also available. These values can be obtained from the spreadsheet that breaks down cost estimates by payer (see next paragraph).

A key exception to this approach is the analysis that estimates costs based on patients' primary insurance type. This analysis considers costs for subgroups of patients for whom the payer is Medicare, Medicaid, or a commercial insurer. The health care payer information is available from the OSHPD hospitalization record, but is not available from CCR. For this reason, the denominators for the insurance-specific analyses do not include all patients, but rather represent exclusively hospitalized patients. This explains why the estimates are higher in this group of analyses: The denominators reflect only hospitalized patients.

**Step 1b: Aggregate data by patient service date.** One individual patient may have multiple records from different dates within a hospitalization or for different hospitalizations. All services for a patient rendered on the same day were first aggregated.

**Step 1c: Aggregate data by months from cancer diagnosis date.** Patients who, according to the CCR, were diagnosed with cancer between January 1, 2007,

and December 31, 2009, inclusive, at any age comprised the analytic cohort for cost analyses. Cohort selection and analyses were restricted to primary cancers. When patients had a secondary cancer diagnosis, these patients (and the costs of hospitalization for any secondary cancer) were retained. The frequency of second cancers within the initial year of diagnosis is low.

Aggregated data from step 1b were allocated to each patient's "month from diagnosis" using the following method:

$$\frac{\text{Date of service} - \text{Date of diagnosis} + 15}{30 \text{ days}} = \text{Months from diagnosis date}$$

For example: If patient X was diagnosed on June 1, 2007, and was hospitalized on June 28, 2007, then:  $(28 - 1 + 15) / 30 = 1.4$  months

So the services rendered on June 28, 2007, would be attributed to month 1 (i.e., the month after the month of diagnosis).

To further illustrate:



**Step 1d: Convert charge data to cost data using hospital-specific ratios of cost to charge (RCCs).** OSHPD data report hospital charges, not costs. The ratios of real costs to hospital charges vary considerably from facility to facility and year to year. To estimate true costs, charges were converted to costs by applying a year- and facility-specific ratio of cost to charge (RCC). The RCC was derived by examining each facility's hospital financial disclosure report. These reports are submitted to OSHPD for each year of interest (i.e., 2006-2011) and were obtained from OSHPD. Estimates are obtained by calculating:

$$\frac{\text{Total operating expenses} - \text{Other operating revenue}}{\text{Total gross patient revenue}} = \text{RCC}$$

An excerpt is shown here:

### Hospital-Specific RCCs for Acute Care Facilities in California

FACILITY ID	NAME	2007
106010735	Alameda Hospital District	0.246940*
106010739	Alta Bates Medical Center	0.244557
106010844	Alameda County Medical Center	0.450173

\*One generalized RCC was used per year per facility; researchers are unable to further break down the RCC per procedure, as these data are unavailable in the annual hospital financial report. In other words, it is not possible to obtain RCCs specific to oncology units.

**Step 1e: Adjust for inflation.** All dollar amounts have been adjusted for inflation and are reported in 2013 USD using Part A price adjusters.

PRICE ADJUSTERS* 2013 DOLLARS	2006	2007	2008	2009	2010	2011
Part A	0.828081	0.858275	0.970938	0.976033	0.974523	0.987356

\*Source: 2012 Annual Report of the Boards of Trustees of the Federal Hospital Insurance and Federal Supplementary Medical Insurance Trust Funds, Federal Hospital Insurance and Federal Supplementary Medical Insurance Trust Funds, table V.D1, April 23, 2012, [www.treasury.gov](http://www.treasury.gov).

For example: If Part A Service X spent \$356, and the date of service was in 2008, then this would be adjusted as follows:  $\$356 / 0.970938 = \$367$ .

**Step 1f: Calculate mean and total month-specific costs for the first year of cancer diagnosis.** After completing steps 1a-1e, the costs in 2013\$ for each patient in each month relative to diagnosis are known. Denote by  $x_{ij}$  the cost of patient  $i$  in the  $j$ th month from diagnosis, and by  $I_{ij}$  the indicator whether patient  $i$  is still alive in the  $j$ th month. The mean month-specific spending is calculated as:

$$y_j = \frac{\sum_{i=1}^N x_{ij} I_{ij}}{\sum_{i=1}^N I_{ij}}, j = -1, \dots, 11$$

, where  $N$  is the total number of patients. For example, to find mean cost in month -1 (the month prior to dx during which patient underwent diagnostic work-up),

add each patient's month -1 costs together and divide by the number of patients who contributed to month -1. Repeat for month 0 (the month of diagnosis), month 1, ..., month 11. If a patient is alive but does not receive any services in a given month, then that patient is included in the denominator, and  $x_{ij} = \$0$  is set as the cost for that month. Mean costs were censored for death, meaning that patients stop contributing to both denominator and numerator after death. To find out total costs for the first year of cancer diagnosis,  $Y$ , sum the mean costs from month 0 through month 11, or  $Y = \sum_{j=0}^{11} y_j$ . See Table 2 on the following page for an example.

**Step 1g: Calculate mean and total month-specific costs for the first year of cancer diagnosis by cancer site and other demographic attributes.** See Table 3 on the following page for an example.

**Reporting costs by cancer site and stage iterates the basic method.** If patient A has stage I breast cancer and patient C has stage II breast cancer, then patient A would contribute to the mean and month-specific costs for the first year of cancer diagnosis for breast stage I and patient C for the stage II estimates. This method is repeated for each stage of breast, lung, colorectal, and prostate cancers, respectively, according to the AJCC 6.0 classification system as reported by CCR.

**Reporting costs by gender.** If patients A and C are female and patients B and D are male, then patients A and C would contribute to the mean and month-specific costs for the first year of cancer diagnosis for females, and patients B and D would contribute for males.

**Reporting costs by race/ethnicity.** If patient A and C are White and patients B and D are Asian, then patients A and C would contribute to the mean and month-specific costs for the first year of cancer diagnosis for Whites, and patients B and D would contribute for Asians.

**Reporting costs by county.** If patients A and C are from Napa County and patients B and D are not, then patients A and C would contribute to the mean and month-specific costs for the first year of cancer diagnosis for Napa County. Counties that contain fewer than 11 cancer patients have been masked to protect patient confidentiality and comply with data use agreements.

**Table 2. Mean Month-Specific and Annual Spending Estimates for the First Year of Cancer Diagnosis**

	MONTH(S) FROM DIAGNOSIS												
	-1*	0†	1	2	3	4	5	6	7	8	9	10	11
<b>Patient A</b>	\$100	\$150 (dx Feb 2007)	\$140	\$130	\$120	\$110	\$100	\$90	\$80	\$70	\$70	\$70	\$70
<b>Patient B</b>	\$150	\$200 (dx Mar 2008)	\$185	\$175	\$165	\$190 (died mo. 4)	n/a	n/a	n/a	n/a	n/a	n/a	n/a
<b>Patient C</b>	\$0	\$180 (dx Dec 2009)	\$175	\$160	\$150	\$140	\$130	\$120	\$115	\$115	\$115	\$115	\$115
<b>Patient D</b>	\$200	\$300 (dx Feb 2008)	\$275	\$250	\$240	\$230	\$200	\$175	\$175	\$160	\$160	\$160	\$150
<b>Calculating Mean Cost</b>	$(100 + 150 + 0 + 200) / 4$	$(150 + 200 + 180 + 300) / 4$	$(140 + 185 + 175 + 275) / 4$	$(130 + 175 + 160 + 250) / 4$	$(120 + 165 + 150 + 240) / 4$	$(110 + 190 + 140 + 230) / 4$	$(100 + 130 + 200) / 3$	$(90 + 120 + 175) / 3$	$(80 + 115 + 175) / 3$	$(70 + 115 + 160) / 3$	$(70 + 115 + 160) / 3$	$(70 + 115 + 160) / 3$	$(70 + 115 + 150) / 3$
Notes: The denominator represents the total patients contributing to mean spending each month. Even though patient C has spending of \$0 in month -1, patient C still contributes to mean spending. Because patient B died in month 4, patient B no longer contributes to the mean spending from month 5 onward.													
<b>Mean Cost</b>	<b>\$113</b>	<b>\$208</b>	<b>\$194</b>	<b>\$179</b>	<b>\$169</b>	<b>\$168</b>	<b>\$128</b>	<b>\$123</b>	<b>\$115</b>	<b>\$115</b>	<b>\$115</b>	<b>\$115</b>	<b>\$112</b>
<b>Mean Annual Cost</b>	$208 + 194 + 179 + 169 + 168 + 128 + 123 + 115 + 115 + 115 + 115 + 112 = \mathbf{\$1,741}$												

Note: Numbers shown in this table are only meant to illustrate the methods used; these are *not* actual spending estimates.

**Table 3. Mean Month-Specific and Annual Spending Estimates for the First Year of Cancer Diagnosis, by Covariate (breast cancer shown as an example here)**

	MONTH(S) FROM DIAGNOSIS												
	-1*	0†	1	2	3	4	5	6	7	8	9	10	11
<b>Calculating Mean Cost</b>	$(100 + 0) / 2$	$(150 + 180) / 2$	$(140 + 175) / 2$	$(130 + 160) / 2$	$(120 + 150) / 2$	$(110 + 140) / 2$	$(100 + 130) / 2$	$(90 + 120) / 2$	$(80 + 115) / 2$	$(70 + 115) / 2$	$(70 + 115) / 2$	$(70 + 115) / 2$	$(70 + 115) / 2$
Notes: The costs are calculated using patients A and C from Table 2 above because patients A and C are in the breast cancer cohort. The denominator represents the total patients contributing to mean spending each month. Even though patient C has spending of \$0 in month -1, patient C still contributes to mean spending.													
<b>Mean Cost</b>	<b>\$50</b>	<b>\$165</b>	<b>\$158</b>	<b>\$145</b>	<b>\$135</b>	<b>\$125</b>	<b>\$115</b>	<b>\$105</b>	<b>\$98</b>	<b>\$93</b>	<b>\$93</b>	<b>\$93</b>	<b>\$93</b>
<b>Mean Annual Cost</b>	$165 + 158 + 145 + 135 + 125 + 115 + 105 + 98 + 93 + 93 + 93 + 93 = \mathbf{\$1,418}$												

\*Month prior to diagnosis or diagnostic “work up”. Month -1 does not contribute to mean annual spending calculation. †Month of diagnosis (dx) is defined as January 1, 2007 through December 31, 2009.

Note: Numbers shown in this table are only meant to illustrate the methods used; these are *not* actual spending estimates.

## Mean Costs During Last 12 Months of Life

All steps for estimating end-of-life (EOL) costs are analogous to the methods described above except steps 1c and 1f, redefined as:

**Step 1c (EOL): Aggregate data by months from death date.** Patients who, according to the CCR, were diagnosed with cancer between January 1, 2007, and December 31, 2009, inclusive, and who died between January 1, 2007, and December 31, 2011, composed the analytic cohort for the EOL cost analyses.

Each service record was allocated to each patient's "month from death" using the following method:

$$\frac{\text{Date of death} - \text{Date of service} + 15}{30 \text{ days}} = \text{Months prior to death}$$

The month of death was defined as the anchor point for each patient. The sample sizes drop when tracing backward from death because only patients who were diagnosed 2007-2009 and who died 2007-2011 are included in this analysis.

For example: If patient X was diagnosed on June 2, 2007; received services on June 29, 2007; and died on August 15, 2007, then:

$$(47^* + 15) / 30 = 2.06 \text{ months}$$

\*Calculated as 1 day in June + 31 days in July + 15 days in August.

So the services rendered on June 29, 2007, would be attributed to month 2 (i.e., two months prior to the month of death).

**Step 1f (EOL): Calculate mean and total month-specific costs for the last year of life.** After completing steps 1a-1e, the costs in 2013\$ for each patient in each month relative to death month are known. To find mean cost in month 0 (the month of death), add each patient's month 0 costs together

**Table 4. Mean Month-Specific and Annual Spending Estimates for the Last Year of Life**

	MONTH(S) PRIOR TO DEATH											
	0*	1	2	3	4	5	6	7	8	9	10	11
<b>Patient A</b>	\$300 (died May '08)	\$250	\$225	\$200	\$175 (dx Jan 2008)	\$170	\$160	\$150	\$150	\$150	\$145	\$145
<b>Patient B</b>	\$150 (died Sep '07)	\$100	\$90	\$80	\$70	\$60	\$50	\$40	\$40	n/a	n/a	n/a
<b>Patient C</b>	\$400 (died Aug '11)	\$375	\$350	\$325	\$300	\$280	\$270	\$260	\$250	\$240	\$240	\$0
<b>Calculating Mean Cost</b>	(300 + 150 + 400) / 3	(250 + 100 + 375) / 3	(225 + 90 + 350) / 3	(200 + 80 + 325) / 3	(175 + 70 + 300) / 3	(170 + 60 + 280) / 3	(160 + 50 + 270) / 3	(150 + 40 + 260) / 3	(150 + 40 + 250) / 3	(150 + 240) / 2	(145 + 240) / 2	(145 + 0) / 2
<b>Mean Cost</b>	<b>\$283</b>	<b>\$242</b>	<b>\$222</b>	<b>\$202</b>	<b>\$182</b>	<b>\$170</b>	<b>\$160</b>	<b>\$150</b>	<b>\$147</b>	<b>\$195</b>	<b>\$193</b>	<b>\$73</b>
<b>Mean Annual Cost</b>	283 + 242 + 222 + 202 + 182 + 170 + 160 + 150 + 147 + 195 + 193 + 73 = <b>\$2,219</b>											

\*Month of death is defined as January 1, 2007 through December 31, 2011.

Notes: Diagnosis noted as "dx". Numbers shown in this table are only meant to illustrate the methods used; these are not actual spending estimates.

and divide by the number of patients who contributed to month 0 (i.e., the denominator, which is specific to patient subgroup). Repeat for month 1 (the month prior to death), month 2, ..., month 11. If a patient is alive but has a cost of \$0 in a given month (during the period from which we have OSHPD data [January 1, 2007 to December 31, 2011]), then that patient is included in the denominator, and \$0 is set as the cost for that month. Mean costs were not censored for diagnosis, meaning that patients contribute to the denominator and the mean costs in the month of death and in all 11 months preceding their death, irrespective of when they were actually

diagnosed with cancer. Mean costs were censored before January 1, 2007 due to data availability. To find total costs for last year of life, sum the mean costs from month 0 through month 11. See Table 4 on the previous page for an example.

## Annual Total Costs

Using the results from above, the annual grand total costs of the initial year from diagnosis, as well as the annual grand total costs of the last year of life, can be calculated by summing (not averaging) all month-specific costs across patients. Following

the notation used in step 1f of “mean spending in the first year of diagnosis,” the annual grand total cost of the initial year from diagnosis is calculated as  $\sum_{j=1}^{12} \sum_{i=1}^N r_{ij} / 3$ . The summation is divided by three because there are three years of data (2007 to 2009) and the aim is to estimate the annual grand total. A similar process is used for the annual grand total spending of the last year of life. See the example below. These totals represent 12-month periods. See Table 5 for an example.

**Table 5. Total Annual Spending Estimates for the First Year of Cancer Diagnosis**

	MONTH(S) FROM DIAGNOSIS												
	-1*	0†	1	2	3	4	5	6	7	8	9	10	11
<b>Patient A</b>	\$100	\$150 (dx Jan 2007)	\$140	\$130	\$120	\$110	\$100	\$90	\$80	\$70	\$70	\$70	\$70
<b>Patient B</b>	\$150	\$200 (dx Mar 2008)	\$185	\$175	\$165	\$190 (died mo. 4)	n/a	n/a	n/a	n/a	n/a	n/a	n/a
<b>Patient C</b>	\$0	\$180 (dx Dec 2009)	\$175	\$160	\$150	\$140	\$130	\$120	\$115	\$115	\$115	\$115	\$115
<b>Patient D</b>	\$200	\$300 (dx Feb 2008)	\$275	\$250	\$240	\$230	\$200	\$175	\$175	\$160	\$160	\$160	\$150
<b>Grand Total (07-09)</b>	<b>450</b>	<b>830</b>	<b>775</b>	<b>715</b>	<b>675</b>	<b>670</b>	<b>430</b>	<b>385</b>	<b>370</b>	<b>345</b>	<b>345</b>	<b>345</b>	<b>335</b>
<b>Grand Total Per Year</b>	450 / 3 = <b>\$150</b>	830 / 3 = <b>\$277</b>	775 / 3 = <b>\$258</b>	715 / 3 = <b>\$238</b>	675 / 3 = <b>\$225</b>	670 / 3 = <b>\$223</b>	430 / 3 = <b>\$143</b>	385 / 3 = <b>\$128</b>	370 / 3 = <b>\$123</b>	345 / 3 = <b>\$115</b>	345 / 3 = <b>\$115</b>	345 / 3 = <b>\$115</b>	335 / 3 = <b>\$112</b>

\*Month prior to diagnosis or diagnostic “work up”. Month -1 does not contribute to total spending calculation.

†Month of diagnosis (dx) is defined as January 1, 2007 through December 31, 2009.

Notes: Using these methods, if breast cancer spending was an average of \$35,000 in the first year of diagnosis and \$53,000 in the last 12 months of life, it is not reasonable to add these two numbers together because they are not mutually exclusive. For example, if a patient was diagnosed January 1, 2007, and died April 31, 2008, then their January 2007 to December 2007 spending would be captured in their “first year of diagnosis,” and their May 2007 to April 2008 spending would be captured in their “last 12 months of life.” Their spending during May 2007 to December 2007 is captured in both their “first year of diagnosis” and their “last 12 months of life.” Numbers shown in this table are only meant to illustrate the methods used; these are *not* actual spending estimates.

## About the Author

Deborah Schrag, MD, MPH, is a medical oncologist; chief, Division of Population Sciences, Department of Medical Oncology, Dana-Farber Cancer Institute; and professor of medicine at Harvard Medical School. Schrag's team at Dana-Farber/Harvard Cancer Center includes Kun Chen, PhD, biostatistician and statistical programmer; Ling Li, MA, statistical programmer; and Jennifer Wind, MA, PMP, senior project manager.

The interpretation and reporting of these data are the sole responsibility of the author.

## Acknowledgments

The collection of cancer incidence data used in this study was supported by the California Department of Public Health as part of the statewide cancer reporting program mandated by California Health and Safety Code Section 103885; the National Cancer Institute's Surveillance, Epidemiology and End Results Program under contract HHSN261201000140C awarded to the Cancer Prevention Institute of California, contract HHSN261201000035C awarded to the University of Southern California, and contract HHSN261201000034C awarded to the Public Health Institute; and the Centers for Disease Control and Prevention's National Program of Cancer Registries, under agreement # U58DP003862-01 awarded to the California Department of Public Health. The Patient Discharge Data (PDD) used in this study

was collected by the State of California's Office of Statewide Health Planning and Development (OSHPD). The ideas and opinions expressed herein are those of the authors and endorsement by the State of California Department of Public Health, the National Cancer Institute, and the Centers for Disease Control and Prevention or their contractors and subcontractors is not intended nor should be inferred.

## About the Foundation

The California HealthCare Foundation (CHCF) is leading the way to better health care for all Californians, particularly those whose needs are not well served by the status quo. We work to ensure that people have access to the care they need, when they need it, at a price they can afford.

CHCF informs policymakers and industry leaders, invests in ideas and innovations, and connects with changemakers to create a more responsive, patient-centered health care system.

For more information, visit [www.chcf.org](http://www.chcf.org).

©2015 California HealthCare Foundation