Translating Service Use to Cost

• Assuming we have collected information about medical service use, we need to translate this use into a cost that can be compared with the measure of study effectiveness
• Commonly this translation is done by identifying price weights and multiplying medical service use by price weights for those services
• Alternatively, when studies are conducted in health systems that have ready access to administrative data such as billing records, we may be able to use these systems to obtain measures of both medical service use and cost (or expenditure)

Study Perspective

• No matter which strategy we choose, appropriate price weights depend on the study’s perspective
  – From a social perspective, price weights should reflect social opportunity cost, which need not equal payment
  – From other perspectives, it may be more appropriate for price weights to reflect payment
• From all perspectives, price weights should reflect marginal not average cost
• It is rare that investigators can identify measures of social opportunity cost
• Common solution is to develop second-best approximations of this cost
Outline

• National tariffs and center-specific prices
• Accuracy of national tariffs
• Incomplete sets of price weights
• Sources of price weights
• Comparability of price weights from different jurisdictions and different times
• Center/country-specific or averaged price weights
• Program cost
• Administrative data

National Tariffs

• Wherever they are available, national tariffs, for example, diagnosis-related group (DRG) payments in Australia or the United States or health resource group (HRG) payments in Great Britain, are a commonly used and well accepted source of price weights for costing out medical service use

Advantages of National Tariffs

• Usually provide price weights for most if not all of the services that are measured in the study
• Are inexpensive to obtain
• Within individual countries, they are usually developed by use of a common methodology
• Make it difficult for investigators to pick and choose among price weights to make an intervention look more or less favorable than it should
• The fact that these tariffs represent what is spent by governments may also be considered an advantage, particularly by governmental decision-making bodies
Center-Specific Price Weights

- Center-specific price weights are also a common used source of price weights for costing out medical service use

Advantages of Center-Specific Price Weights

- If price weights come from the centers in which the study was conducted, they provide a more accurate estimate of the cost that was actually incurred within the study.
- Cost and medical service use can have important interactions that, if not accounted for, can lead to biased results.
  - Given that efficient producers use greater amounts of relatively less costly services and smaller amounts of relatively more costly services, use of a single set of price weights for all providers tends to overstate cost.
  - This principle holds for studies performed in a single center, in multiple centers in a single country, or in multiple centers in multiple countries.

Which Price Weights?

- The most appropriate source of price weights depends in part on whether the question we are asking is more national or center specific, but the question being asked cannot be our only consideration.
  - e.g., if our goal is to make national resource allocation decisions, we might conclude that use of a set of national price weights would yield data appropriate for making such decisions.
  - But, the representativeness of the resulting cost estimates depends on the representativeness of both medical service use and price weights.
  - Once one factor is unrepresentative, it is not clear whether ensuring that the other factor is representative necessarily yields the best estimates.
Accuracy of National Tariffs?

- Beck et al. compared British Trust tariffs with the results of a service-specific costing exercise and found substantial differences between the two.
- Heerey et al. compared results from microcosting versus Irish DRGs and found differences ranging between –9 and 66%.
- Shwartz et al. compared results derived by multiplying U.S. charges times Federal cost-to-charge ratios with estimates of cost derived by use of relative value units and found inaccuracy at the patient level, but the averages usually were within 10% of each other.

Possible Sources of Difference

- Some of these differences may have occurred simply because cost differs by center, in which case national tariffs need not agree with the cost in any single center.
- Others may have more to do with differences in the level of aggregation that are employed by the different data sources than they do with the fact that some estimates are national and some are local.

Do The Differences Affect Study Conclusions?

- Common sense suggests they should; Rittenhouse et al. and Reed et al. have argued that they may.
- However, there is at best limited evidence about whether or not they do.
  - Both Chumney et al. and Taira et al. found that costing method affected the estimate of cost by treatment group.
  - However, in both cases, the intervention was effective enough that the authors concluded that the differences did not lead to conflicting recommendations about adoption.
Incomplete Sets of Price Weights

- In many studies, price weights are derived from a selected set of centers that participated in the study.
- The analog in multinational studies is that price weights are often collected from a subset of participating countries.
  - Within each country, they are often derived from a single source, which in the same study can vary between center-specific price weights in some countries to national tariffs in others.

Incomplete Sets of Price Weights (II)

- The countries from which price weights are collected might be ones that:
  - Enroll a large number of participants in the study.
  - Represent the spectrum of economic development among countries that participated in the study.
  - Have readily available price weights, or
  - Have regulators that require a submission for reimbursement.
- In industry-sponsored studies, they also may include countries in which the study sponsor wishes to make economic claims.

Sampling Price Weights

- It may be costly to obtain weights for all services measured in the study in all participating centers and countries.
- May lead us to sample price weights for medical services that:
  - Are observed most frequently in the study.
  - Are considered likely to be affected by the intervention.
  - Have particularly high or low price weights.
Imputing Missing Price Weights

- When price weights are collected for a subset of the medical services that have been observed in a study, we must develop a method for imputing price weights for the remaining services
  - If little additional information is available about these services, may want to use the mean price weight for similar services
    - e.g., use the simple average of the price weight estimates from hospitalizations for which these estimates are available.
  - If additional information is available, e.g., relative values for a medical service, prediction models can be used to impute price weights

Implications for Number of Services Sampled

- If we can build such a prediction model, what do we know about the number of countries in which we should collect price weights and the number of price weights we should collect within each country?
- Glick et al. have reported that imputation error is minimized when we obtain estimates for fewer types of services from as many countries as feasible, rather than by obtaining estimates for more types of services in fewer countries
  - But reduction of imputation error may need to be traded-off against the cost of data collection

Sources of Price Weights

- Sources of price weights differ by country and by medical service
- For inpatient services, many countries now use patient classification systems like DRGs or HRGs
  - Some provide measures of both the relative cost of a hospital stay – a relative value or a relative weight – and the reimbursement/cost for a stay
  - Others provide measures of relative cost, and we must then independently identify the cost per relative weight
- For studies that enroll participants in the developing world, The WHO has developed estimates of price weights for inpatient care in at least 49 countries

### Sample Australian and U.S. DRG Weights

<table>
<thead>
<tr>
<th>Diagnosis/Procedure</th>
<th>Australia</th>
<th>U.S.</th>
<th>Relative Weight</th>
<th>ALOS</th>
<th>Cost ($AUD)</th>
<th>DRG Relative Weight</th>
<th>ALOS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chest pain F74Z</td>
<td>0.42</td>
<td>1.54</td>
<td>1296</td>
<td>143</td>
<td>0.57</td>
<td>2.1</td>
<td></td>
</tr>
<tr>
<td>Cholecystectomy w/common bile duct exploration, no co</td>
<td>2.47</td>
<td>5.77</td>
<td>7698</td>
<td>196</td>
<td>1.60</td>
<td>5.7</td>
<td></td>
</tr>
<tr>
<td>GI hemorrhage w/o cc G61A</td>
<td>0.79</td>
<td>3.49</td>
<td>2476</td>
<td>174</td>
<td>1.01</td>
<td>4.7</td>
<td></td>
</tr>
<tr>
<td>GI obstruction w/o cc G65A</td>
<td>1.54</td>
<td>5.96</td>
<td>4789</td>
<td>180</td>
<td>0.98</td>
<td>5.4</td>
<td></td>
</tr>
<tr>
<td>Heart failure and shock F64Z</td>
<td>1.20</td>
<td>5.40</td>
<td>3736</td>
<td>127</td>
<td>1.03</td>
<td>5.2</td>
<td></td>
</tr>
<tr>
<td>Hernia (&lt;10, AU; &lt;17 US) G10Z</td>
<td>0.84</td>
<td>1.33</td>
<td>2618</td>
<td>163</td>
<td>0.87</td>
<td>2.9</td>
<td></td>
</tr>
<tr>
<td>Liver transplant A01Z</td>
<td>26.57</td>
<td>29.27</td>
<td>82885</td>
<td>480</td>
<td>8.94</td>
<td>17.9</td>
<td></td>
</tr>
</tbody>
</table>

### Fee Schedules

- **UK**:

- **US**
  - [www.cms.hhs.gov/FeeScheduleGenInfo/](http://www.cms.hhs.gov/FeeScheduleGenInfo/)
  - [www.cms.hhs.gov/ProspMedicareFeeSvcPmtGen/](http://www.cms.hhs.gov/ProspMedicareFeeSvcPmtGen/)
  - [www.cms.hhs.gov/Mcpartbdrugavgsalesprice/02.asp](http://www.cms.hhs.gov/Mcpartbdrugavgsalesprice/02.asp)

- **Australia**

- Fee Schedules’ Web Addresses Change Routinely

### Sample U.S. Outpatient Fees

- Non-facility fee (RVU * $37.34)

- **Office visit, new**
  - Problem-focused history & exam; straightforward medical decision making; typically spends 10 minutes face-to-face
  - RVU: 36

- **Clinical Diagnostic Laboratory Fee Schedule ($2003) (REF)**
  - National midpoint
  - Heparin assay
    - RVU: 24.72
  - Blood platelet aggregation
    - RVU: 40.56

  - Average of floor and ceiling cost
  - Drug infusion pump supplies
    - RVU: 43.02
  - Rose tinted glasses
    - RVU: 9.28
Detailed Costing Studies
- Dutch Health Insurance Board’s guidelines for standardizing the estimation of price weights
- General hospital cost studies (Garattini et al.; Oostenbrink et al.)
- Hospital pharmacy cost studies (Scullen et al.)
- Study-specific costing exercises
  - Analysis of accounting data
  - Use of the top-down cost-block program
  - Estimation of hospital cost functions
  - Time and motion studies

Time and Motion Studies
- Step 1: Enumerate services
  - Identify and define each activity (Iglar)
  - Determine the exact production process (Finkler)
- Step 2: Define a time standard for each service
  - Self estimation (Kinosian, TPN)
  - Self measurement (Iglar)
  - Observer measurement (Finkler, TPN)
- Step 3: Define a frequency measure for each service
- Step 4: Define the supplies and equipment used
- Step 5: Identify fixed and variable costs by service

Results of Time and Motion Study
Time and Motion Issues (I)

- Sample size
  - Estimating a confidence interval around a single mean (or set of means)
  - Potentially observing the same providers and the same patients multiple times
    - Two stage cluster sampling?
- Patient type
  - May want to define standards for different patient types
    - Iglar: seven patient types

Time and Motion Issues (II)

- Boundaries between activities
  - Different observers may disagree about when one activity has stopped and another has started
- Starting and stopping an activity
  - The production process is often thought of as being linear (one starts the task and continues until one completes it)
  - In practice, it is often interrupted
- Joint time
  - How should you time an activity when more than one task is being performed simultaneously?

Work Sampling

- Sample activities during the day

  "We used a small portable random alarm beeper (JD-7, Divilbliss Electronics, Champaign, IL) to determine the sampled time points. The beeper was set at a predetermined frequency (5.0/hr); beeps occurred at random so nurses could not anticipate the sampled time point. At each beep, the nurses recorded what they were doing (activity) and with whom (contact) in a small code book that they carried with the beeper."

Research Nurse’s Work Activities

<table>
<thead>
<tr>
<th>Work Activity</th>
<th>Mean proportion of workday (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient Care</td>
<td></td>
</tr>
<tr>
<td>Discussing patient care</td>
<td>9.7 (6.4-13.0)</td>
</tr>
<tr>
<td>Providing education</td>
<td>4.4 (3.4-5.4)</td>
</tr>
<tr>
<td>On phone with patient *</td>
<td>2.5 (1.4-3.5)</td>
</tr>
<tr>
<td>Phone, other patient care *</td>
<td>1.7 (0.8-2.6)</td>
</tr>
<tr>
<td>Computer</td>
<td>4.3 (2.2-6.4)</td>
</tr>
<tr>
<td>Other</td>
<td>6.2 (3.0-9.4)</td>
</tr>
<tr>
<td><strong>Subtotal</strong></td>
<td><strong>28.8 (24.1-33.2)</strong></td>
</tr>
<tr>
<td>Research related</td>
<td></td>
</tr>
<tr>
<td>Screening patients</td>
<td>10.2 (7.9-12.5)</td>
</tr>
<tr>
<td>Completing forms</td>
<td>11.2 (7.1-15.4)</td>
</tr>
<tr>
<td>Computer</td>
<td>4.1 (2.4-5.9)</td>
</tr>
<tr>
<td>Phone, research *</td>
<td>0.8 (0.3-1.3)</td>
</tr>
<tr>
<td>Other</td>
<td>16.2 (11.7-20.7)</td>
</tr>
<tr>
<td><strong>Subtotal</strong></td>
<td><strong>42.5 (38.1-46.7)</strong></td>
</tr>
<tr>
<td>Personal</td>
<td>16.4 (12.0-20.7)</td>
</tr>
<tr>
<td>In transit</td>
<td>12.5 (8.1-15.9)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>100.2 †</strong></td>
</tr>
</tbody>
</table>

Comparability of Price Weights from Different Jurisdictions

- Price weights collected from multiple countries may not be directly comparable
- Lack of comparability may be due to differing resource intensities
  - e.g., differing staff-to-patient ratios in units labeled “intensive care”
- Another source may be variation in countries’ handling of the cost of capital
  - e.g., Some countries may include it in hospital budgets; others may include it in public budgets
- Lack of comparability may also arise because of variation in countries’ accounting standards and practices, etc.

Impact of Lack of Comparability

- This heterogeneity in price weights:
  - May lead to problems of transferability of the study results
  - Should not bias the pooled result of the evaluation, so long as we multiply country-specific measures of medical service use times country-specific price weights
- We may want to adjust the observed price weights to make them comparable between countries.
Potential Adjustments

- Equating the methods different countries use to account for capital expenditure and overhead
  - e.g., Removal of overhead and capital costs when they were present in the price weights
- Translation of all price weights into a common currency
  - e.g., Purchasing power parity (PPP) statistics
    - Currency conversion factors that eliminate differences in price levels between countries

Purchasing Power Parity Statistics

<table>
<thead>
<tr>
<th>Country</th>
<th>PPP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>1.38</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>14.2</td>
</tr>
<tr>
<td>France</td>
<td>0.902</td>
</tr>
<tr>
<td>Hungary</td>
<td>126</td>
</tr>
<tr>
<td>Italy</td>
<td>0.85</td>
</tr>
<tr>
<td>Korea</td>
<td>7.94</td>
</tr>
<tr>
<td>Mexico</td>
<td>7.37</td>
</tr>
<tr>
<td>Netherlands</td>
<td>0.899</td>
</tr>
<tr>
<td>Sweden</td>
<td>9.06</td>
</tr>
<tr>
<td>Switzerland</td>
<td>1.68</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>0.627</td>
</tr>
<tr>
<td>United States</td>
<td>1.00</td>
</tr>
</tbody>
</table>


Inflation and Discounting

- Because costs and outcomes in different time periods are not directly comparable, their comparison requires conversion to a common time period
- One conversion factor reflects inflation
- Another conversion factor reflects time preference: we discount streams of cost and outcome to account for the fact that we value them differently, depending upon when they occur
Discounting Formula

\[
\sum_{t=0}^{N} \frac{B_t}{(1+R)^t}
\]

where \( t = 0 \) to \( N \); \( B_t \) equals the net benefits (benefits minus costs) in time \( t \); and \( R \) equals the discount rate

Inflation and Discounting (2)

- What is the relationship between inflation and discounting?
  - Real rate of discount \((1+r)\), where \( r \) = real rate of time preference
    - Used when costs are already inflation-adjusted (e.g., a common set of price weights used to value all service use)
  - Nominal rate of discount \([(1+r)(1+i)]\), where \( i \) equals the rate of inflation
    - Used when cost data come from different years
    - If sectors have relatively different inflation rates, need to use sectoral rates of inflation to evaluate changes in costs over time

What Discount Rate?

- Current practice, U.S. (and most, but not all, other developed countries): 3\% (U.S. Panel recommendation)
- Appropriate rate for less developed countries?
- Adjust for risk separately in the analysis
"When" to Discount

- The need to discount is not a function of the duration of the study; it is a function of the duration of follow-up per participant
  - For example, if you randomize over a 4 year period, but only follow participants for 60 days, you would adjust across the 4 years for inflation (i.e., use real dollars), but would not discount
  - Only if you followed each participant for more than a year would you both adjust for inflation and discount

When (II)

- You follow people for 4 years; at the end of follow-up you obtain price weights from the Federal government for the year 2004: Discount? Inflation adjust?
- You enroll people during a 6 month period and follow each of them for 6 months; either you collect bills or obtain price weights from the government for the year 2004 to estimate costs: Discount? Inflation adjust?
- You enroll people during a 3-year period, but follow each for only 1 year; you collect bills to estimate costs: Discount? Inflation adjust?
- You follow people for 4 years; you collect bills to estimate costs: Discount? Inflation adjust?

U.S. CONSUMER PRICE INDEX

<table>
<thead>
<tr>
<th>Year</th>
<th>All Items</th>
<th>Medical Care</th>
<th>Medical Care Services</th>
</tr>
</thead>
<tbody>
<tr>
<td>1990</td>
<td>130.7</td>
<td>162.8</td>
<td>162.7</td>
</tr>
<tr>
<td>1995</td>
<td>152.4</td>
<td>220.5</td>
<td>224.2</td>
</tr>
<tr>
<td>2000</td>
<td>172.2</td>
<td>260.8</td>
<td>266.0</td>
</tr>
<tr>
<td>2001</td>
<td>177.1</td>
<td>272.8</td>
<td>278.8</td>
</tr>
<tr>
<td>2002</td>
<td>179.9</td>
<td>285.6</td>
<td>292.9</td>
</tr>
<tr>
<td>2003</td>
<td>184.0</td>
<td>297.1</td>
<td>306.0</td>
</tr>
<tr>
<td>2004</td>
<td>188.9</td>
<td>310.1</td>
<td>321.3</td>
</tr>
<tr>
<td>2005</td>
<td>195.3</td>
<td>323.2</td>
<td>336.7</td>
</tr>
<tr>
<td>2006</td>
<td>201.6</td>
<td>336.2</td>
<td>350.8</td>
</tr>
</tbody>
</table>

http://data.bls.gov/cgi-bin/surveymosfcu
U.S. City Average, Not seasonally adjusted, base period: 1982-84 = 100
Medical Care=Drugs+supplies+MCS (professional and hospital services)
Center/Country-Specific Price Weights or Price Weights Averaged Among Centers/Countries

• When we have price weight data from more than one country, we also need to decide whether to apply country-specific price weights to medical service use observed in that country or whether to average the price weights from all study centers into a single set of price weights.

Center/Country-Specific or Averaged

• Ideal: Because relative prices can affect quantities of services provided, where feasible, multiply center/country-specific price weights times times center/country-specific counts of medical services.
  - For jurisdictions where price weights aren’t available:
    • Use (averages of) price weights from similar jurisdictions.
    • E.g., in a trial that enrolls patients in Western Europe, Asia and Latin America, one might average price weights from other Western European countries to value service use in Germany, but wouldn’t want to use this average for Asia or Latin America.

Center/Country-Specific or Averaged (II)

• Corollary: If one has a set of price weights for each country that participated in the trial, one should not average them and use this average for all services measured in the trial.
  - The most common reasons suggested for such a strategy are that 1) reducing variability in the price weights will reduce variability in the estimated costs and 2) an average set of price weights may be more representative.
Center/Country-Specific or Averaged (III)

- However:
  - Empirically, use of a single set of price weights need not reduce variance
  - If substitution effects are strong, this strategy may introduce bias in the estimates of cost differences
  - Why is it more “representative” to use a set of price weights that no one faces?

Specific Recommendations, Valuing Service Use

- Identify a set of countries from which price weight data will be collected
- Identify a set of services for which price weights are needed
- Approximately 12-18 months before the end of the trial, identify an organization / group of investigators who will provide the price weights
- Approximately 6-12 months before the end of the trial, initiate price weight data collection
- Develop methods of imputation both for service use and countries for which price weights are unavailable

Program Cost

- Some studies, for example those evaluating novel behavioral interventions, assess programs that provide bundles of services that may not have price weights associated with them
- In these cases, there is a need to evaluate the cost of these programs and to assess how this cost changes as the number of participants changes
General Instruments Meant for Costing Particular Types of Programs

• Example: Drug Abuse Treatment Cost Analysis Program (DATCAP) and the Brief DATCAP
  – The DATCAP captures the cost of personnel, supplies and materials, major equipments, contracted services, buildings and facilities, miscellaneous resources, and those not recorded elsewhere
  – Various members of the treatment program’s staff, such as administrators, therapist coordinators, and accounting/finance personnel, are asked to identify specific services, and cost data are usually collected from general ledgers, personnel reports, expenditure reports, and inventory reports

Time Analyses

• Example #1: Anderson et al.’s site-specific diaries
  – All staff members who provided any of 94 different treatment and nontreatment services were asked to track cost over the course of 1 week
  – The staff indicated the amount of time and materials used in providing these services during each day of the week
• Example #2: Barrett et al.’s use of data from patient files and electronic databases in the development of the Secure Facilities Service Use Schedule

Study-Specific Instruments

• Examples:
  – Diabetes Prevention Program (DPP): DPP staff responded to questionnaires that elicited information about the types of personnel that participated in the intervention; the amount of time they participated; the health education materials distributed; etc.
  – A Stop Smoking In Schools Study (ASSIST): ASSIST staff responded weekly to standardized forms that asked about staff time, travel time and distance, consumables, accommodation, and vouchers for peer supporters who implemented the intervention
Administrative Data

- In cases where study participants are treated in health care systems that maintain administrative databases that quantify either medical service use, cost/payment, or both service use and cost/payment, data from these systems may be used to quantify service use and cost
- Examples in the U.S. include U.S. Medicare records, U.S. Veterans Administration (VA) records, and health maintenance organization records

Advantages of Use of Administrative Data

- They usually capture service use independent of site of care
- They are independent of fallible participant recall

Potential Limitations

- Some administrative databases aggregate service use, making it difficult to identify services and cost during a specified period (e.g., 3 months from some index date)
- The listed payments need not represent cost from the perspective adopted by the study
- The data may not be updated continuously (e.g., postings may be made every 6 months or every 1 year)
- The data may report covered services only
- If participants are enrolled in different administrative databases, it may be difficult to pool data because of differences in systems of care, costing, and coding
Summary

• We commonly estimate the cost outcome of a trial by multiplying medical service use by a set or sets of price weights for these services
• Two of the most common sources for price weights are national tariffs and center-specific price weights
• National tariffs have a number of strengths, including that they represent what is spent by governments
• Center-specific price weights, on the other hand, may give a more accurate estimate of the cost that was actually incurred in the trial

Summary (cont.)

• If for reasons of data collection cost we identify price weights for a selected set of services, it is incumbent that we develop a method for imputing price weights for the other services that were observed in the trial
• One approach is to use readily available price weights, such as those from DRG or HRG payment systems, to develop a prediction model for available price weights
• If for reasons of data collection cost we identify price weights for a selected set of centers or countries that participated in the trial, we should try to maximize the number of centers or countries in which price weights are collected rather than trying to maximize the number of price weights that are collected per center or country