

## Statistical Methods with Applications to DRG Analysis

**Presenters:** Prof. W. Sermeus (Center for Health Services & Nursing Research, Leuven University, Belgium) and Prof. Jason M. Sutherland (Center for Health Policy Research and Reform, Dartmouth College)

### Introduction

This workshop is relevant to professionals who routinely analyze data from case mix classification systems, such as DRG. The properties of the most common statistical methods will be discussed, assumptions explored, and alternatives debated. This workshop will provide a thorough grounding in the statistical methods used for DRG analyses and will review methods used.

This workshop is intended for quantitative professionals with experience analyzing data with some background in statistical methodology. All conference participants are welcome, though the material will be directed towards methodology.

### 1) Introduction to DRG analyses

Through the use of examples, sources of variability in medical practice will be linked to DRG cost and LOS analyses. Characteristics of summary statistics will be reviewed, including common measures of location and spread. The properties of the mathematical average are compared to most robust measures of location.

### 2) Trimpoint Definition and Outlier Identification

In DRG analyses, determination of inliers and outliers is an important consideration of statistical analyses. Advanced measures of DRG analyses are discussed with practical examples.

Model performance is an important objective when comparing classification systems. Common measures include  $R^2$  and MSE are sensitive to the definition of DRG trimpoints. This section will review outlier identification methods used by various countries and their resultant economic incentives.

### 3) Cost Modelling and Cost Weight Calculation

The distribution of patient costs (and LOS) within DRG is most often right-skewed. The most common approach is to log-transform cost and apply linear models. The statistical properties of this technique are investigated and results discussed. Alternative strategies are explored. Cost weights (and tariff levels) represent ratios of expected costs; the properties of the ratio estimate are developed and adjustments are proposed with examples.

- Log Transformations and Model Performance
- Other Non-linear Transformation Methods
- Cost Weight Calculation and the Ratio Estimator

Multivariate analyses are commonly used to model costs. Linear regression, weighted regression and non-linear models are explained, including error distributions and alternatives for link functions. Bayesian models are explored for incorporating prior information into cost modeling.

Cost weight calculation involves controlling for facility (hospital) level effects. We explore various alternatives, such as using fixed effects, random effects and the hospital specific relative value (HSRV) method.

Cost weights are point estimates; estimates of cost weight variability are rarely presented. Methods for variance calculation are discussed, including parametric and bootstrap estimates.

There are countries where the cost data is sparse or of poor quality. We discuss alternatives for developing cost weights in this environment, including use of proxies and assessing fitness of use of data from other countries.

#### **4) Robust Estimators**

Parametric models are often applied to analyzing DRG data. Since DRG data often fails the underlying assumptions, such as normality, robust estimators provide an alternative methodology. Different types of robust estimators are explored and their impact on measures of location discussed in the context of specific DRG examples. This section will also include data envelope analysis (extreme point method) in the context of efficiency measurement and will introduce robust regression.

#### **5) DRG and Survival Analysis Methods**

Length of stay data are typical time-to-event data. Common methods for time-to-event data are introduced, such as the Kaplan-Meier estimator and the proportional hazards model. There are situations, such as when patients are transferred or deaths, when patient data should be treated as censored data. The most common methods for analyzing censored data are discussed in the context of DRG analyses and their relative effects discussed.

#### **6) Design Considerations for Case Mix Classification Development**

Many countries have developed adaptations of DRG to suit local needs and priorities. However, methods for designing case mix systems employ a variety of statistical methods:

- Incorporating clinical input into case mix system design
- Criteria in developing grouper
- Considerations for data inclusion/exclusions
- Validation samples methods
- Statistical methods to assist in developing grouper cells (decision tree software, mixture distributions)
- Use of length of stay in the absence of cost data

**Duration:** All day