SEMINAR 4

**Seminar Report: Issues and Challenges in the Analysis of Biomedical Data**

**Introduction**

The fourth seminar in this series featured **Dr. Zhenzhen Jin** from the Department of Biostatistics at the Mailman School of Public Health, Columbia University. Dr. Jin, a fellow of prestigious organizations such as ASA, IMS, ISI, and ICSA, discussed the complexities involved in analyzing biomedical data. Her talk covered various research designs, data types, analytical challenges, and statistical approaches in biostatistics.

**Key Topics Covered**

**1. Study Designs in Biomedical Research**

Dr. Jin provided an overview of multiple study designs used in biomedical studies, including:

* **Randomized trials**
* **Case-control studies**
* **Cohort studies**
* **Phase I-IV trials**
* **Crossover studies**
* **Adaptive designs**

Each design plays a vital role in evaluating research questions in areas like drug development, clinical interventions, and disease prevention.

**2. Data Types and Sources**

Biomedical data can vary significantly in size and type. Dr. Jin highlighted:

* **Small and large sample datasets**
* **Big data** from electronic medical/health records (EMR/EHR), imaging, and omics (genomics, transcriptomics, proteomics) data
* **Examples:**
  + *mHealth sleep health mobile app study (SHMAS)*
  + *UK BioBank*, which tracks 500,000 participants and includes genomic data

**3. Example: Alzheimer's Disease Pathway**

A multi-omics approach is often employed to explore Alzheimer's disease progression. Dr. Jin illustrated how the pathway can include:

* **Genome → Transcriptome → Proteome → Metabolome → Imaging**

Specifically, her example involved studying Alzheimer’s in adults with Down syndrome through a **48-month longitudinal study**, covering:

* Health history and caregiver assessment
* Physical exams, MRI, and PET scans (amyloid and tau)
* Bloodwork and neurological assessments

Typical Course of a Patient Receiving Allogenic Stem Cell Transplant

Patient met an indication

Donor Search Assessment of patient organ function and disease status

Matched Unrelated unrelated

Family adult cord blood

Doner donor Admit for AlloHCT

Infections

2 weeks 2 months 2 weeks Hospital stay 1-3 months GVHD

4-6 weeks 5 weeks 8-12 weeks organ toxicities

6-9 months, off immunosuppression

Long term follow up

Allogeneic Stem cell transplantation

Non engraftment engraftment non- engraftment

(rejection) (rejection)

Infections graft versus host disease malignal relapse

Transplant related non- transplant

Mortality related mortality

What/ Why/ How ??

What : What is the problem (scientific question)?

Why : Why it is important ?

How : Hypothesis/ study design/ data collection/ data analysis/ summary

**4. Allogeneic Stem Cell Transplantation Process**

Dr. Jin discussed the sequence of events during a patient’s journey through stem cell transplantation. This process includes:

* **Donor Search:** Matching a donor through family, adult registry, or cord blood
* **Transplantation and hospital stay:** Involves risks of infection, graft-versus-host disease (GVHD), and long-term follow-up
* **Complications:** Risks of non-engraftment, organ toxicities, or relapse

**Statistical Challenges and Solutions**

Dr. Jin identified several challenges in biomedical data analysis, such as:

* **Missing data** (MCAR, MAR, MNAR)
* **Big data** complexities (volume vs. time)
* **High-dimensional data** and correlated outcomes in longitudinal studies

**Solutions to Missing Data:**

* Logistic regression models to assess missingness
* Multiple imputation or last observation carry-forward (LOCF)
* Sensitivity analysis to ensure robustness

**The Role of Statistics in Biomedical Research**

Dr. Jin emphasized that **statistics** plays a central role in:

* **Framing research questions and hypotheses**
* **Designing studies and data collection plans**
* **Analyzing data and interpreting results**

She stressed the importance of adhering to **statistical principles**, such as:

* **Sufficiency and likelihood principles**
* **Least squares estimation**
* **Bayesian methods and machine learning techniques** (for prediction and classification)

**Interpretability in Statistical Models**

One key takeaway from Dr. Jin’s seminar was the **trade-off between model complexity and interpretability**. Complex statistical methods are often difficult for non-statisticians to understand, while simpler tools (e.g., t-tests or regression) are more accessible. The **"art of statistics"** lies in balancing this complexity and making sound methodological decisions.

**Conclusion and Discussion**

The seminar concluded with a reflection on the **future of biomedical data analysis**:

* There is a growing need for **integrative research approaches** that combine statistical principles with robust methods to address new scientific challenges.
* Dr. Jin highlighted the importance of **careful planning** to avoid pitfalls such as data missingness and misinterpretation.
* Quoting Mosteller and Tukey (1977):  
  "*Every student of the art of data analysis must build on previous knowledge while reformulating insights for future applications.*"

**Key Takeaways**

1. **Biomedical research** requires thoughtful study design and robust statistical analysis to address the inherent complexities of the data.
2. **Missing data** remains a major challenge, with strategies like imputation and sensitivity analysis being crucial.
3. **Statisticians** must balance complex models with interpretability, ensuring that non-specialists can also understand the results.
4. The **art of statistics** involves applying simple tools effectively while being mindful of underlying assumptions and research questions.

This seminar provided valuable insights into the **role of statistics** in biomedical research, helping attendees understand how to navigate challenges such as missing data, longitudinal outcomes, and big data analysis.