www.ispor.org

ISPOR Improving healthcare decisions

ISPOR Special Interest Group

HANDLING MISSING DATA IN HEALTH ECONOMICS AND OUTCOMES RESEARCH - STATE OF THE ART GUIDANCE AND GOOD PRACTICES

Forum session by the ISPOR Statistical Methods in HEOR Special Interest Group

Virtual ISPOR 2021 | May 19, 2021

Introduction and History of the ISPOR Missing Data in HEOR Key Project

Rita Kristy, MS, Astellas Pharma Global Development, Northbrook, IL, USA

Past Chair, ISPOR Statistical Methods in HEOR Special Interest Group

Disclaimer

The speaker is a paid employee of Astellas. This presentation is intended for informational purposes only and does not replace independent professional judgment. This presentation is not intended to be legal advice. Statements of fact, positions taken and opinions expressed are those of the speaker individually and, unless expressly stated to the contrary, do not necessarily reflect the opinion or position of the speaker's employer, Astellas, or any of its subsidiaries and/or related entities.

Speakers:

- **Rita Kristy, MS,** Astellas Pharma Global Development, Northbrook, IL, USA
- **Gian Luca Di Tanna, PhD, MPhil, MSc,** The George Institute for Global Health, University of New South Wales, Sydney, NSW, Australia
- **Kumar Mukherjee, PhD,** Philadelphia College of Osteopathic Medicine, Georgia Campus, Suwanee, GA, USA

Agenda:

- History of the Missing Data Project
- Terminology on Missing Data
- Findings from Literature
- Guidance and Recommendations
- Next Steps
- Q&A

History

- Missing data needs to be critically evaluated in statistical analyses
- Mentioned as an important topic in the SIG proposal
- Formal proposal was in April 2017
- Approval by the Health Science Policy Council in November 2017

ISPOR Statistical Methods in HEOR Special Interest Group (SIG)

Mission: To provide statistical leadership for strengthening the use of appropriate statistical methodology in health economics and outcomes research and improve the analytic techniques used in real world data analysis.

Co-Chairs of SIG

- Gian Luca Di Tanna, Head of Statistics and Associate Professor, The George Institute for Global Health and UNSW, Sydney, NSW, Australia
- Helene Karcher, PhD, Global Head, RWE ophthalmology, Novartis, Basel, BS, Switzerland
- Rita Kristy, MS, Senior Director, Medical Affairs Statistics, Astellas Pharma Global Development, Northbrook, IL, USA

Co-Chairs of ISPOR Missing Data in HEOR Working Group

- Necdet Gunsoy, PhD, MPH, Head of HTA Analytics, Global Market Access and Pricing, AbbVie Ltd., England, United Kingdom
- Gianluca Baio, PhD, MSc, Professor of Statistics and Health Economics, University College London (UCL), England, United Kingdom

Project plan

- Two papers to be developed
  - Literature review and evaluation of current practices
  - Guidance and recommendations for handling missing data
- Detailed search criteria determined and refined (September 2018)

Literature review

- Abstracts were reviewed by working group and assessment of a full review made
  - Context
  - Type of missing data
  - How was it addressed
  - Limitations
- Literature review updated in August – November 2020

Conference presentations and article

- Presentations:
  - Forum presentation, May 2018: HANDLING MISSING VALUES IN REAL-WORLD DATA: ARE THERE CHALLENGES FOR REGULATORY DECISIONS FOR MEDICAL PRODUCTS?
  - Workshop, November 2018: ARE MISSING DATA PROPERLY ACCOUNTED FOR IN HEALTH ECONOMICS AND OUTCOMES RESEARCH?
  - Workshop, May 2019: HOW TO TACKLE THE ESTIMATION OF TREATMENT IMPACT IN THE PRESENCE OF DIFFERENTIAL WITHDRAWAL AND MISSING DATA AMONG STUDY ARMS?
- White Paper:
  - Are Missing Data Properly Accounted for in Health Economics and Outcomes Research? G Baio; N Gunsoy; N Onwudiwe; D Vanness, *Value and Outcomes Spotlight*, March/April 2020

Missing Data Terminology

Gian Luca Di Tanna, PhD, MPhil, MSc, The George Institute for Global Health, University of New South Wales, Sydney, NSW, Australia

Chair, ISPOR Statistical Methods in HEOR Special Interest Group

DO YOU KNOW WHAT THE ACRONYMS **MAR**, **MCAR**, **MNAR** STAND FOR AND THEIR MEANING?

Polling Question

Definitions of missing data mechanisms (Rubin, 1976)

MCAR Missing Completely At Random	Missingness does not depend on anything related to the substantive research question
MAR Missing At Random	Missingness might depend on its value, but this dependence is broken within the strata of fully observed variables
MNAR Missing Not At Random	Even within strata of observed variables, missingness still depends on the value itself

Definitions of missing data mechanisms (Rubin, 1976)

3 variables Y, X, Z with *Y and Z fully observed*, the probability of X being missing depends:

MCAR Missing Completely At Random	Missingness does not depend on anything related to the substantive research question	Neither on X, Y or Z
MAR Missing At Random	Missingness might depend on its value, but this dependence is broken within the strata of fully observed variables	Y and/or Z, but given (Y,Z) not on X
MNAR Missing Not At Random	Even within strata of observed variables, missingness still depends on the value itself	X, Y and/or Z

A dichotomized definition

MCAR
Missing Completely At Random

MAR
Missing At Random

Ignorable Missingness

Missing values occur independent of the data collection process

MNAR
Missing Not At Random

Non-Ignorable (Informative) Missingness

There is a structural cause on the missingness mechanism which depends on the unobserved predictors and/or missing value itself

(Some…) approaches available

MCAR
Missing Completely At Random

MAR
Missing At Random

Weighting

- Complete Case Analysis
- Available Case Analysis (Pairwise deletion)
- Inverse Probability Weighting

Single Imputation

- Last Observation Carried Forward
- Mean Imputation
- Regression Imputation
- Indicator Method
- Nearest Neighbour
- Random Forest
- Boosted Regression Trees

Multiple Imputation

- Multiple Imputation
  - Joint Models
  - Full Conditional Specification (FCS)
    - Multiple Imputation by Chained Equations
    - FCS – Twofold
    - FCS – Moving Time Window
    - FCS – Linear Mixed-Effects Model or GLM
- Bayesian MI

(Some…) approaches available

MCAR
Missing Completely At Random

MAR
Missing At Random

Weighting
- Complete Case Analysis
- Available Case Analysis (Pairwise deletion)
- Inverse Probability Weighting

Single Imputation
- Last Observation Carried Forward
- Mean Imputation
- Regression Imputation
- Indicator Method
- Nearest Neighbour
- Random Forest
- Boosted Regression Trees

Multiple Imputation
- Multiple Imputation
  - Joint Models
  - Full Conditional Specification
    - Multiple Imputation by Chained Equations
    - FCS – Twofold
    - FCS – Moving Time Window
    - FCS – Linear Mixed-Effects Model or GLM
- Bayesian MI

MNAR
Missing Not At Random

- Bayesian MI
- Pattern Mixture Models

Methodology and Results

Kumar Mukherjee, PhD, Philadelphia College of Osteopathic Medicine, Georgia Campus, Suwanee, GA, USA

Co-Chair, Member Engagement of the ISPOR Statistical Methods in HEOR Special Interest Group

Objective of the study

- To provide a critical review of handling missing data in HEOR studies on the bases of a Systematic Literature Review (SLR) aimed at identifying methods used to handle missing data within the HEOR field with a specific focus on cost, utility, and Patients Reported Outcomes (PROs)

Search Strategy

Search procedure

- PubMed search till August 2020
- *Focus area:*

  - Missing data or Incomplete data

  - Cost

  - Effectiveness: HRQoL measures, Utility
- *Study exclusion criteria*:

  - Endpoint was outside the scope

  - No discussion or review of methods to handle missing data

  - Full text in English language was unavailable
- No minimum sample size was set as an inclusion criteria.

Key search terms

- Missing data, Incomplete data
- Cost-effective, Cost-utility, Budget impact
- Health economics, Outcomes research, Health outcomes, Real world, Pragmatic trials, Registry, Claims / Administrative data
- Observational, Retrospective, Prospective, Comparative effectiveness
- Analyz, Methodol, Modelling, Models, Strateg
- Develop, Guidance, Best practice, Recommend

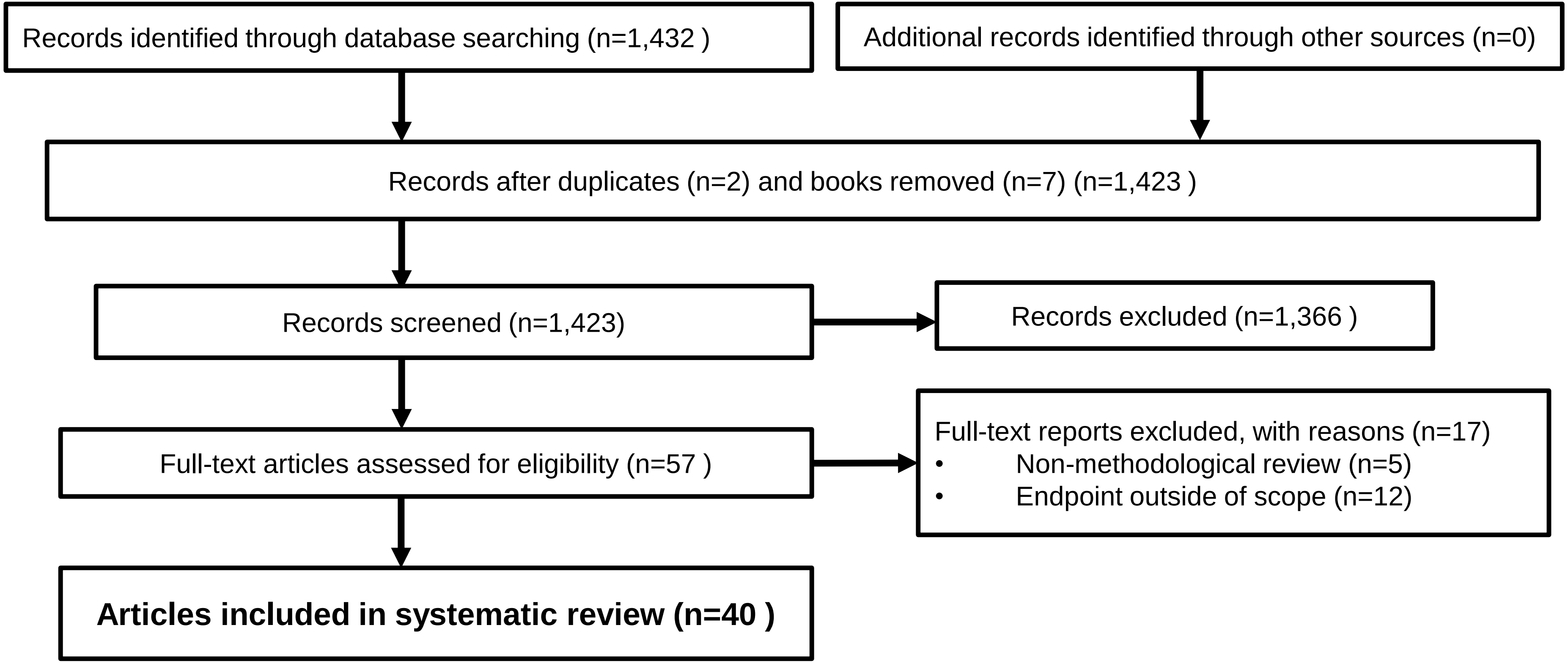
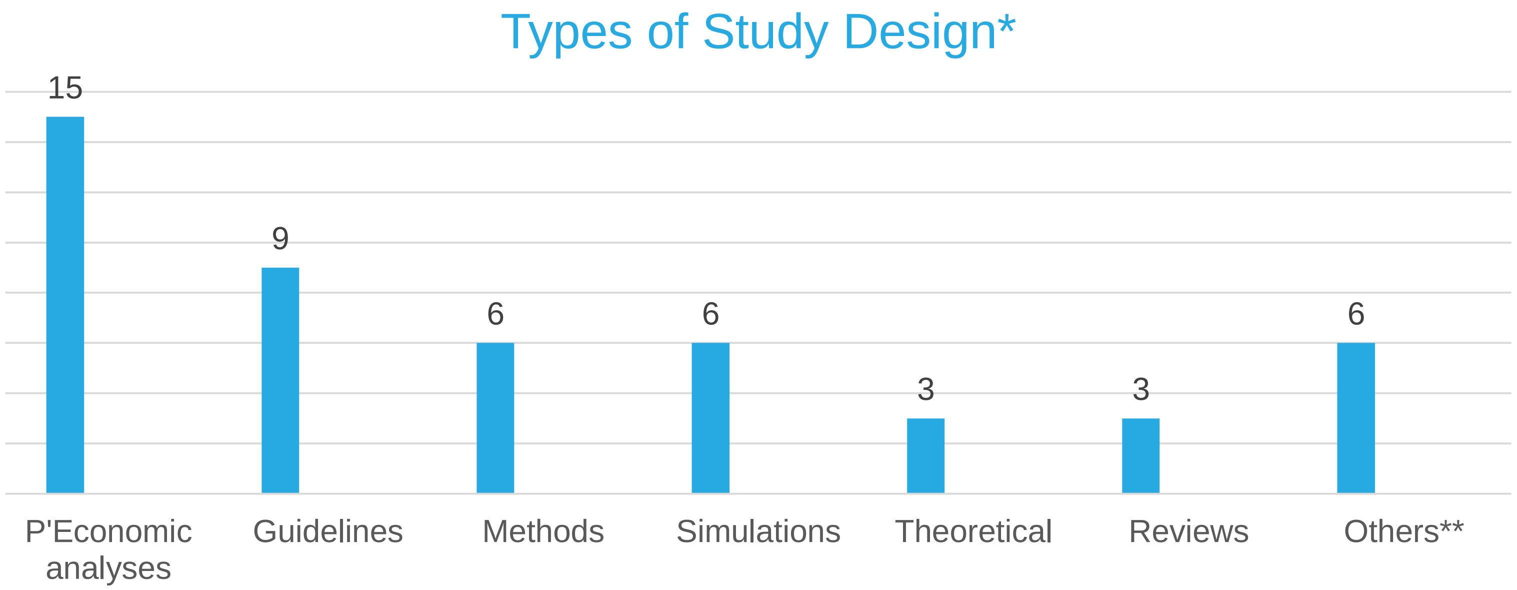


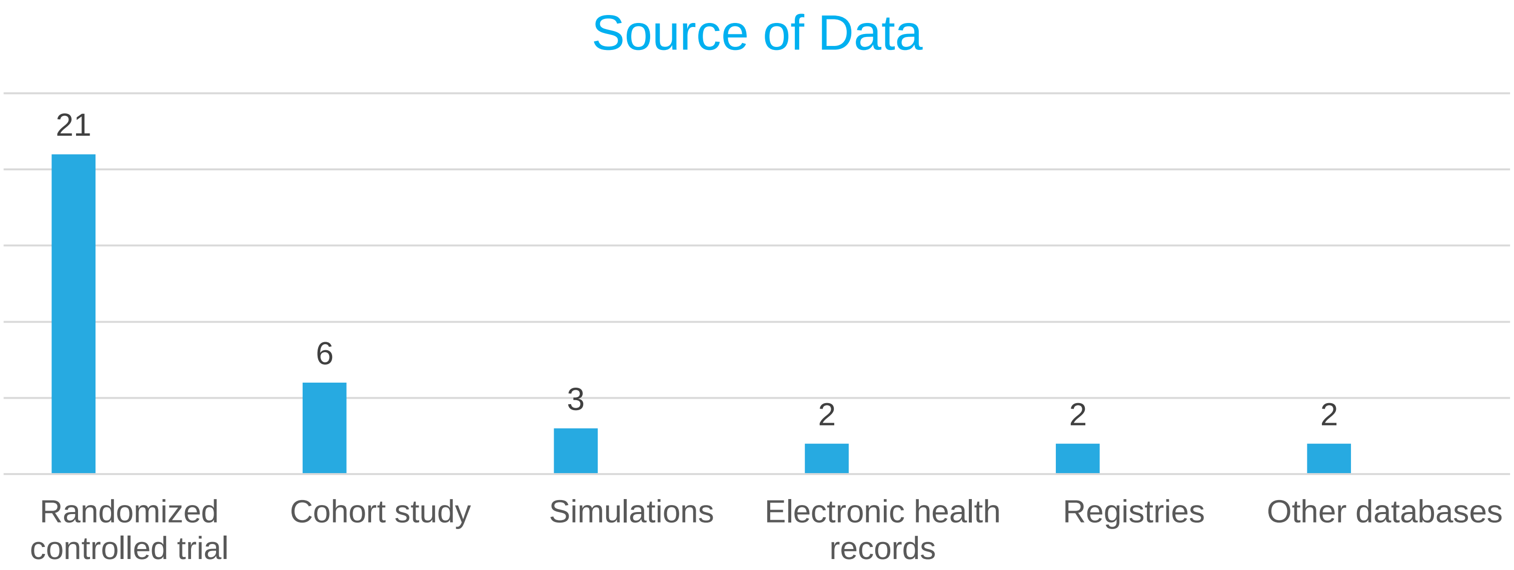
Figure 1: PRISMA Flow Diagram of Selection of Articles

Different Types of Study Design



*Some studies used a combination of study designs
**Others include predictive modelling, comparative effectiveness research, survey research and cohort study using database

Different Types of Datasets Used



Note: Four studies did not use any data source

What Was Missing?

- Cost data
- Effectiveness data (HRQoL, QALY)
- Patient reported outcomes
- Covariates

Methodologies Used for Handling Missing Data

Methods*	Frequency
Multiple imputation^	28
Complete case analysis	11
Pattern mixture models	3
Last value carry forward (LVCF)	3
Item and aggregate imputation	1
Inverse probability weighting (IPW)	1
Semiparametric model with propensity score	1
Bayesian robust 2-stage causal modeling with instrumental variable	1
Additive least square support vector machine approach	1
Non-parametric test based on distance between observed data	1

*Some studies used multiple approach.
^ Among these 14 studies used chained equation. Four studies used Predictive Mean Matching (PMM).

Level of Addressing Missing Data Handling Issues

- *Categorized in 3 levels*:

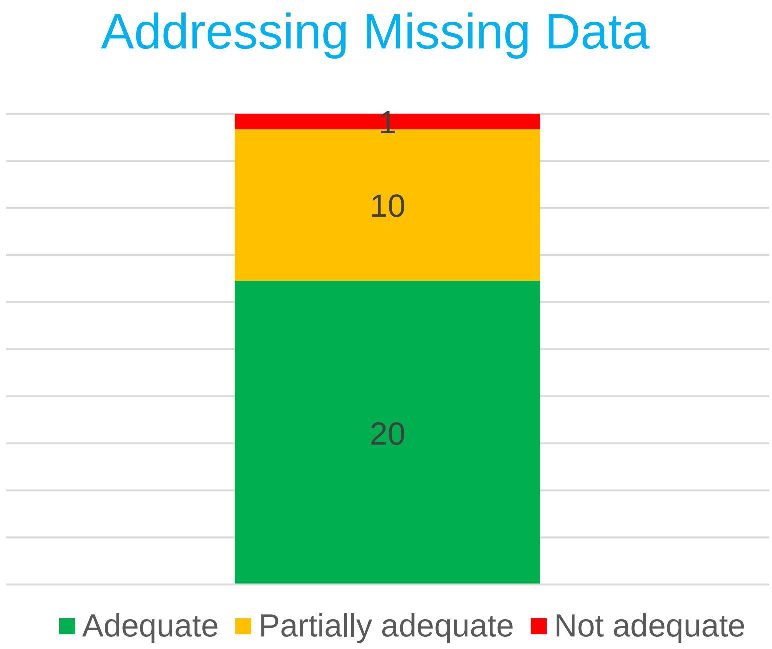
  - Adequately addressed (Article used multiple methods of handling missing data)

  - Partially adequate (Article used only one single method of handling missing data)

  - Not adequate (No missing data issues were addressed)

- For 9 studies, adequacy of methods could not be addressed as these were without any empirical illustration of methods or theoretical in nature.

Level of Adequacy of Addressing missing data in HEOR literature



Nature of Guidance / Recommendations from Literature

- Economic evaluation based on Registry – MI is preferred where data are MAR
- RCT based CEA studies – report information collection, assumptions and limitations of choice of methods, sensitivity analysis
- MI models for handling missing PROMs data – consider prevalence of missing data in the study, sample size, the number of items and numbers of levels within the individual items, analysis plan
- In PROMs data

  - in small sample, imputation at score or subscale level may outperform imputation at item level

  - in case of high proportions of item-nonresponse imputation at the item level may perform better
- Whether the *Missingness Pattern Approach* (MPA)'s assumptions are reasonable and deciding appropriateness of the method

Warnings and Myths

Gian Luca Di Tanna, PhD, MPhil, MSc, The George Institute for Global Health, University of New South Wales, Sydney, NSW, Australia

Chair, ISPOR Statistical Methods in HEOR Special Interest Group

1 Assumptions & Complete Case Analysis (CCA)

- All approaches have assumptions (including CCA)
- CCA can be valid but it is commonly not efficient
  - It is valid when the probability of each record being complete depends on the covariates only (not on the outcome)
  - It is efficient when all the missing values are in the outcome (or if records with missing covariate(s) also have a missing outcome)

Ignorable?

- MAR holds (Missing data ignorable): this does not mean we can sleep well and over-rely on CCA by default

3 When an approach to missing data is needed

- Unless the overall proportion of missing data records is really minimal (< *X*%) we should use (more than) one of the proposed approaches
- The proportion of missing data usually refers to % of data missing in outcome and covariates (exposure, confounders) but non on auxiliary variables
- The fraction of missing information (FMI) measure can give precious information on choosing auxiliary variables to be used in imputation models

4 MAR vs MNAR

- We cannot define a strategy according to MAR vs MNAR: this is untestable
- MI with specific adjustments (ie Pattern Mixture Models - PMM), Bayesian approaches can handle both

5 Multiple Imputation (MI)

- MI does not recover and replace missing values but attempts to produce valid and more precise results
- MAR holds:
  - MI is preferred to CCA as it reduces bias and/or improve precision (esp. when using incomplete cases and auxiliary variables)
- MNAR holds:
  - even CCA can be preferred to MI (if not using appropriate auxiliary variables)
  - Modify MI approach by PMM / tipping point analyses

6 On the imputation model

- Extra care to the regressions used to impute the missing values as costs are skewed and utilities bounded
- Correlation mechanisms to consider for costs & utilities

- Fundamental to **conduct sensitivity analysis** to assess the robustness of mechanisms also under plausible MNAR assumptions in trial-based CEA
- No one-size-fits-all models in MI regression equations: usage of as many predictors as possible so that MAR is reasonable
- MI is not necessarily the best option but it is for sure attractive given its flexibility: it can be confidently used under MAR and can be implemented under different MNAR scenarios
- Suggestions to **impute by randomized groups** and use of *reference-based/controlled MI*
- Increasing availability of methods to assess various situations of departures from MAR easy to implement (ie. *Bayesian approaches, PMM*)

- Many methods available, some are seldomly used
- Improve methods proposals, better conducted and more comprehensive simulations
- Make methods more accessible (ie. tutorials papers, code sharing, open access examples)

AFTER TODAY'S PRESENTATION, WHEN YOU NOTICE THAT YOUR DATASET CONTAINS MISSING DATA WHAT WILL YOU DO?

Polling Question

Next Steps

Ways to Get Involved

- Review manuscripts 1 and 2 this summer
  - If you are a SIG member, you will receive a draft of the manuscripts in your inbox with 3-4 weeks to review and provide feedback
- Our email is always open! Send us ideas for topics and methods you would like more information on statisticalmethodssig@ispor.org
- Connect with your peers and submit an abstract for ISPOR Europe 2021
  - Deadline June 10, 2021 for Issue Panel and Workshop Abstracts and June 29, 2021 for Research Abstracts

ISPOR Special Interest Group

Sign up as a Special Interest Group Member

1. Visit ISPOR webpage www.ispor.org
2. Select “Member Groups”
3. Select “Special Interest Groups”
4. On left side bar, click “Statistical Methods in HEOR”
5. Click button to “Join This Special Interest Group”

For more information, e-mail statisticalmethodssig@ispor.org