SOURCES OF SAFETY DATA AND STATISTICAL STRATEGIES FOR DESIGN AND ANALYSIS IN CLINICAL TRIALS

2017 JOINT STATISTICAL MEETINGS

Richard C. Zink, Ph.D.
Principal Research Statistician Developer
JMP Life Sciences, SAS Institute, Inc.
Adjunct Assistant Professor of Biostatistics
UNC - Chapel Hill

richard.zink@jmp.com

ASA BIOP SAFETY WORKING GROUP

- Series of papers to examine safety throughout medical product development and postmarket [1-4]
- **Sources of safety data and statistical strategies for design and analysis:**
  - Clinical trials
  - Postmarket surveillance
  - Real world insights
  - Transforming data into evidence
- *Therapeutic Innovation & Regulatory Science*

ASA BIOP SAFETY WORKING GROUP

- Rima Izem (FDA/CDER)
- Qi Jiang (Amgen)
- Michelle Krukas-Hampel (QuintilesIMS)
- Mark Levenson (FDA/CDER)
- Haijun Ma (Amgen)
- Olga Marchenko (QuintilesIMS)
- Estelle Russek-Cohen (FDA/CDER)
- Matilde Sanchez-Kam (Sanchez-Kam Consulting)
- Richard C. Zink (JMP Life Sciences, SAS Institute Inc)

OUTLINE

- Challenges of safety analysis
- Safety data
- Analysis strategies
  - A healthy dose of data visualization
- Pros and cons of clinical trials for assessing safety
- Primary focus is on adverse events (AEs)
  - Occurrences of clinically-relevant changes in other safety endpoints are often reported as AEs
  - Changes in alanine aminotransferase (ALT) can be represented by *Alanine aminotransferase abnormal*, *Alanine aminotransferase increased*, or *Alanine aminotransferase decreased*

INTRODUCTION

- Randomized clinical trials are the gold standard for evaluating the efficacy of a new intervention
  - Designed for primary endpoints
  - Population we can easily study
- Obtain an assessment of the safety profile
- AEs, laboratory abnormalities, vital signs, hospitalizations, QOL, electrocardiograms
- Increased emphasis on proactive and comprehensive evaluation of safety throughout the medical product life cycle [5-7]
- Methods are often descriptive due to a number of challenges

CHALLENGES WITH SAFETY ANALYSES

- Studied in limited population
- Rarity of many safety outcomes
- Enrichment with sicker population creates issues with generalizability
- Numerous safety endpoints repeatedly measured over time
- Characteristics including duration, severity, causal relationship
- Analyses by subgroup, duration of therapy, compliance

CHALLENGES WITH SAFETY ANALYSES

- Not all safety endpoints and analyses can be pre-specified
- Events occur spontaneously at any time
- Medical coding may be inaccurate or inconsistent
- Differential rates of drop out
- Benefit-risk assessment
- Trials for chronic indications are too short for assessment
- Individual versus collective ethics

DATA: NONCLINICAL DEVELOPMENT

- Pharmacology studies
- General toxicity studies
- Toxicokinetic and nonclinical pharmacokinetic studies
- Reproduction toxicity studies
- Genotoxicity studies
- Studies to assess carcinogenic potential

DATA: CLINICAL DEVELOPMENT

- Phase I: Safety and dosing in healthy volunteers (patients in oncology studies)
- Phase II: Efficacy and safety in patients with disease
- Phase III: Confirmatory for efficacy, additional safety

DATA: MEDICAL CODING

- Prevent loss or distortion of data
- Consistent terminology throughout stages of development
- Improved timeliness for analysis, exchange and decision making
- Facilitates electronic exchange of data
- **MedDRA**: Medical Dictionary for Regulatory Activities
  - Classifies AEs and groups terms for analysis
- **WHO-DD**: World Health Organization's Drug Dictionary
  - Analyses of active ingredient

DATA: DATA STANDARDS

- **CDISC**: Clinical Data Interchange Standards Consortium
- Global, platform-independent standards to improve data sharing
- Gains in efficiency, flexibility and savings in resourcing
- Numerous Standards
  - **SEND**: Standard for Exchange of Nonclinical Data
  - **SDTM**: Study Data Tabulation Model
  - **ADaM**: Analysis Data Model
  - Therapeutic area specific
- Combine data within and across sponsors to identify signals

ANALYSIS: SAFETY ANALYSIS PLANS

- **PSAP**: Program-wide Safety Analysis Plan [5]
- **SPERT**: Safety Planning, Evaluation and Reporting Team
- Not a current regulatory requirement
- Several sponsors have implemented PSAPs
- Document data and analyses to characterize the safety profile throughout the product life-cycle
- Facilitate interactions with regulators regarding safety strategies
- Aid in the evaluation of the benefit-risk profile in post-market
- **iSAP**: Statistical Analysis Plan for the ISS limited to development

ANALYSIS: SAFETY MONITORING

- Ongoing review of data collected to protect patient safety, trial credibility, and validity of results
- **DMC**: Data Monitoring Committee [8]
- Safety Assessment for IND Safety Reporting [9]
  - **SAC**: Safety Assessment Committee
  - Oversee “evolving safety profile… by evaluating… cumulative serious adverse events… in the development program, as well as other… safety information”
  - Meet more frequently than DMC
  - Review entire safety database
  - Recommend when to submit an IND safety report to FDA and investigators

ANALYSIS: REPORTING

- AEs that occur since previous study visit are reported by the patient or care-giver
- Additional AEs may be identified by the clinician through in-clinic or laboratory assessments that have worsened since baseline
- Verbatim text coded using MedDRA
- Traditionally summarized by preferred terms, grouped by system organ classes in order of decreasing frequency

ANALYSIS: REPORTING

- Three-tier approach for analysis of AEs [5]
  - Tier I
    - Pre-planned tests for expected or clinically relevant AEs
    - Often no adjustment for multiplicity unless numerous events to consider
  - Tier II
    - Unexpected common (≥ 4 patients in a single arm) events should consider multiplicity adjustments
    - FDR: False Discovery Rate provides more balanced approach to type I error and power [10]
  - Tier III
    - Rare events summarized in a listing

ANALYSIS: REPORTING

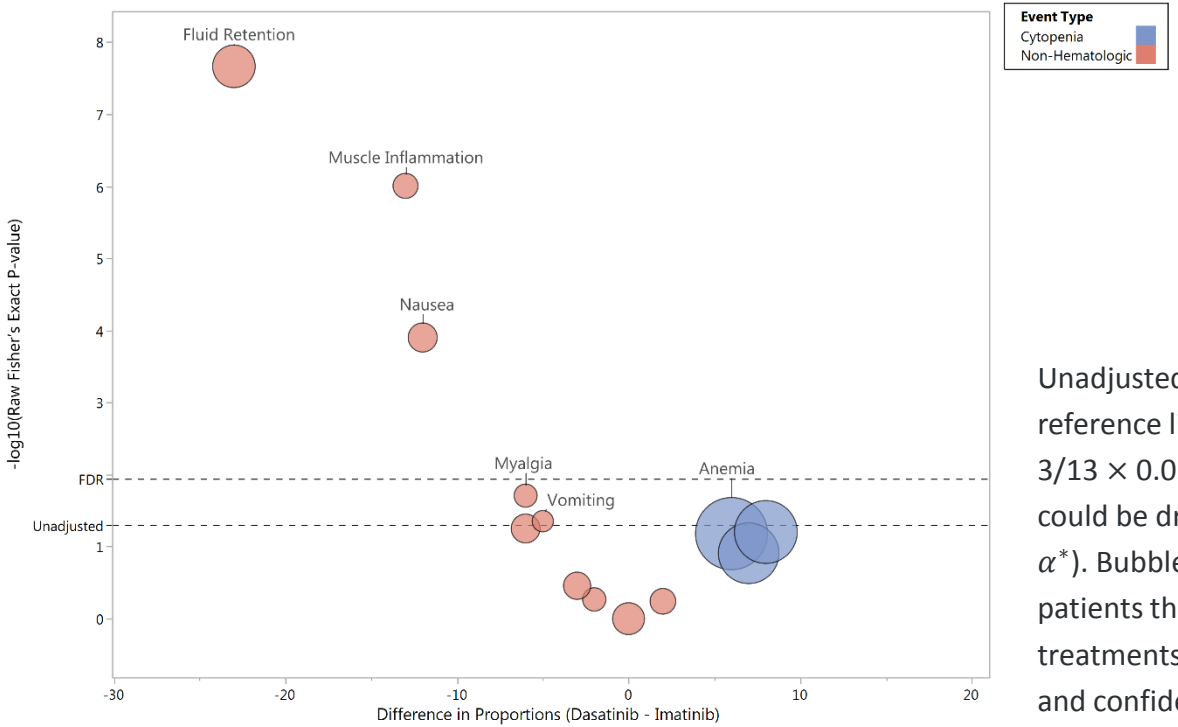
Table 6.3 Drug-Related Adverse Events that Occurred in at Least 10% of Treated Patients

Preferred Term	Dasatinib (N = 258)	Imatinib (N = 258)	Unadjusted 95% Confidence Interval and P-value	FDR Confidence Interval and P-value
Anemia	232 (90)	217 (84)	6 (0.2, 11.8) 0.0661	6 (-1.5, 13.5) 0.1074
Thrombocytopenia	181 (70)	160 (62)	8 (-0.1, 16.1) 0.0628	8 (-2.5, 18.5) 0.1074
Neutropenia	168 (65)	150 (58)	7 (-1.4, 15.4) 0.1237	7 (-3.8, 17.8) 0.1787
Fluid Retention	49 (19)	108 (42)	-23 (-30.7, -15.3) <0.0001	-23 (-32.9, -13.1) <0.0001
Diarrhea	44 (17)	44 (17)	0 (-6.5,6.5) 1.00	0 (-8.4, 8.4) 1.00
Nausea	21 (8)	52 (20)	-12 (-17.9, -6.1) 0.0001	-12 (-19.6, -4.4) 0.0005
Rash	28 (11)	44 (17)	-6 (-12.0, 0.0) 0.0561	-6 (-13.7, 1.7) 0.1074
Musculoskeletal Pain	28 (11)	36 (14)	-3 (-8.7, 2.7) 0.3499	-3 (-10.4, 4.4) 0.4549
Headache	31 (12)	26 (10)	2 (-3.4, 7.4) 0.5746	2 (-5.0, 9.0) 0.6225
Muscle Inflammation	10 (4)	44 (17)	-13 (-18.2, -7.8) <0.0001	-13 (-19.7, -6.3) <0.0001
Fatigue	21 (8)	26 (10)	-2 (-6.9, 2.9) 0.5409	-2 (-8.4, 4.4) 0.6225
Myalgia	15 (6)	31 (12)	-6 (-10.9, -1.1) 0.0196	-6 (-12.3, 0.3) 0.0638
Vomiting	13 (5)	26 (10)	-5 (-9.5, -0.5) 0.0445	-5 (-10.8, 0.8) 0.1074

Values are N(%). N were derived from % and arm totals. Data for treatment sample sizes and percents from Table 4 of Kantarjian et al. (2010). Table is sorted by overall incidence of each event. Data on system organ classes were not available. Confidence intervals are based on the risk difference of dasatinib minus imatinib using a normal approximation. For the FDR intervals, $\alpha^* = 3/13 \times 0.05 = 0.0115$.

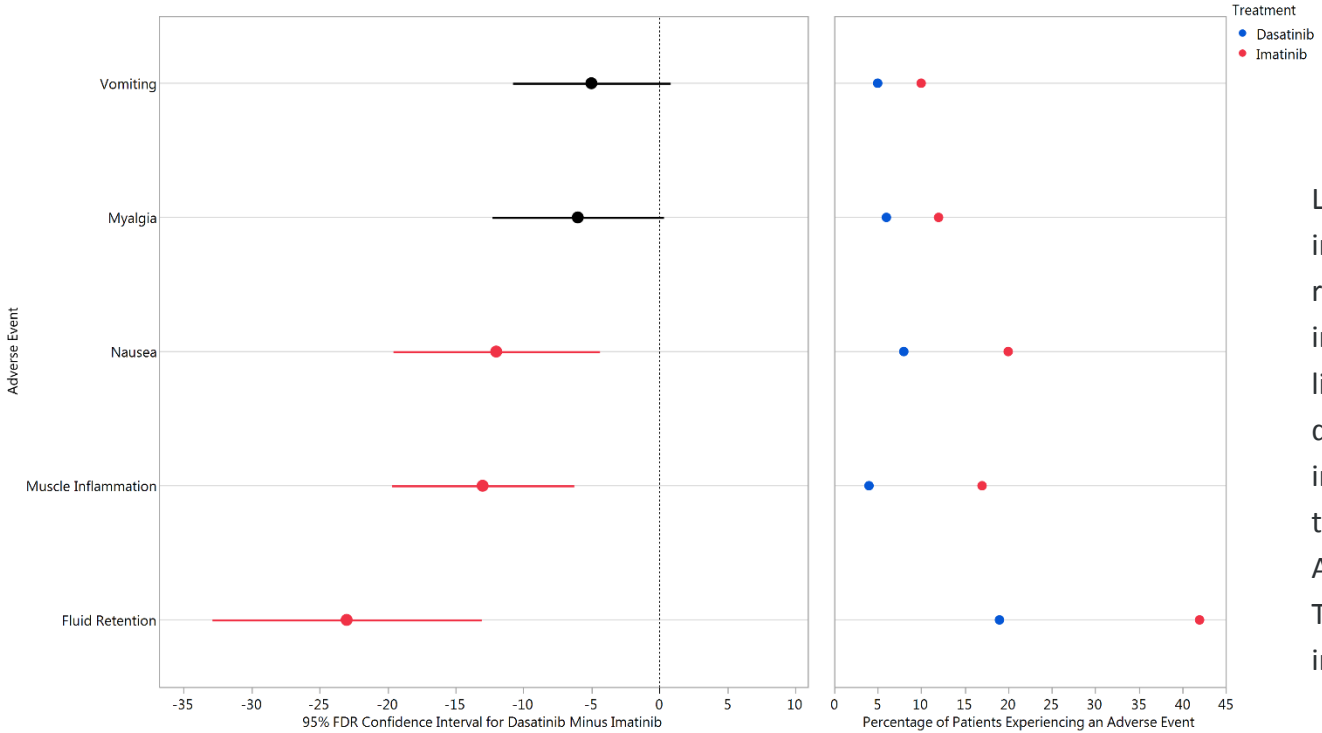
- Traditional summary of AEs
- Table from [11,12]

ANALYSIS: REPORTING



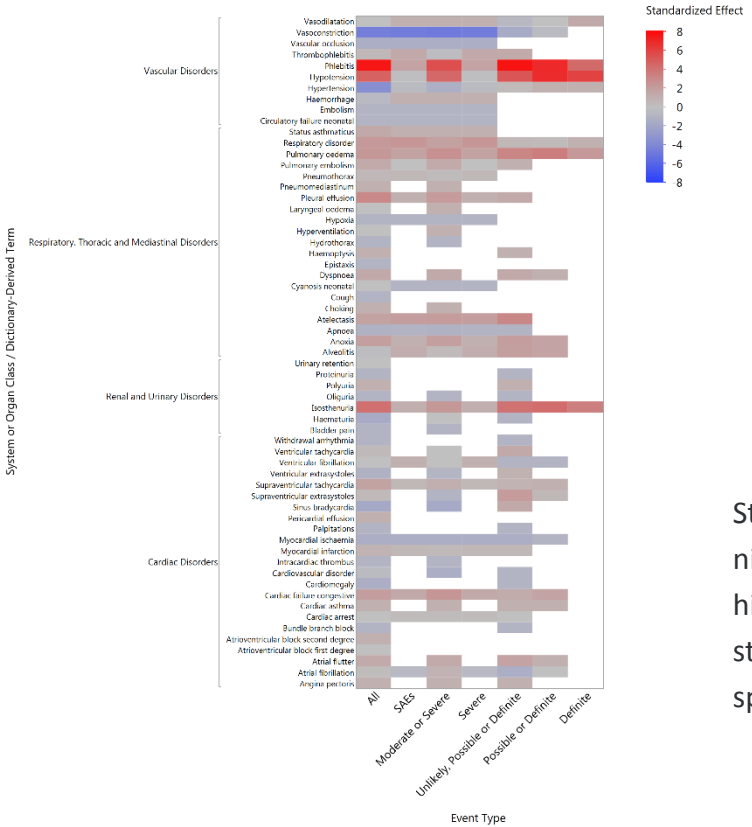
Unadjusted reference line drawn at –log10(0.05) = 1.3. FDR reference line drawn at –log10(0.0115) = 1.9393, where $\alpha^* = 3/13 \times 0.05 = 0.0115$. Alternatively, the FDR reference line could be drawn at -log10(maximum unadjusted p-value $\leq \alpha^*$). Bubble area is proportional to the total number of patients that experience an adverse event for both treatments combined. Data from Table 4 of [11]. P-values and confidence intervals computed in [12].

ANALYSIS: REPORTING



Left panel displays a forest plot of FDR intervals for dasatinib minus imatinib; red intervals indicate significantly increased risk for imatinib. Reference line is drawn at 0 to indicate no difference between dasatinib and imatinib. Right panel presents a dot plot to communicate the incidence of each AE for each treatment arm. Data from Table 4 of [11]. P-values and confidence intervals computed in [12].

ANALYSIS: REPORTING



Standardized effect is the risk difference for experiencing an adverse event for nicardipine minus placebo divided by its standard error. Darker red or blue indicates higher risk on nicardipine or placebo, respectively. Cells are white when the standardized effect cannot be calculated, most often when no events occur. Due to space limitations, a subset of system organ classes is presented. Data from [13].

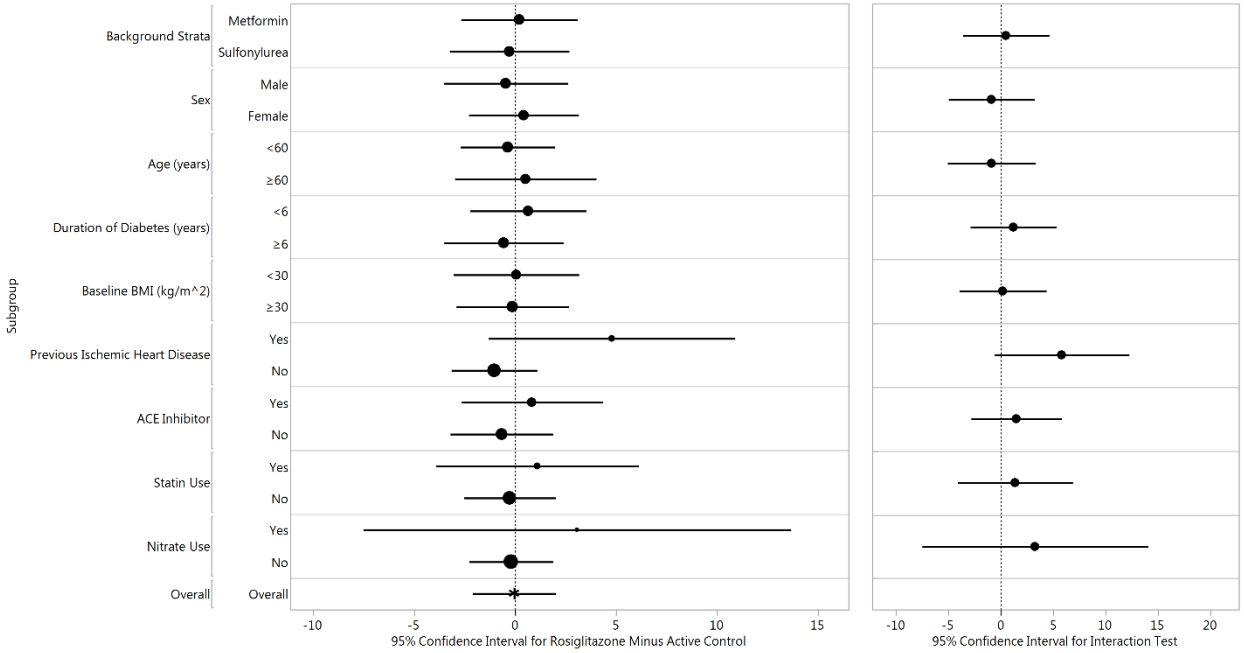
ANALYSIS: SUBGROUPS

- Differential treatment response within demographic, genetic, disease, environmental, behavioral or regional characteristics
- Assess consistency and robustness of results obtained for the entire study population
- Assess whether estimated overall effect is broadly applicable to patients across the proposed indication
- Generate hypotheses for future research

ANALYSIS: SUBGROUPS

- Transparency is key
  - Subgroup size
  - Number of subgroups assessed (not just reported)
  - Subgroups determined pre or post hoc?
  - Multiplicity adjustments applied?
  - Stratified randomization used?
  - Heterogeneity assessed?
- Regulatory guidance [14]
  - Prefers unadjusted p-values and confidence intervals
  - Encourages confidence intervals for tests of heterogeneity

ANALYSIS: SUBGROUPS



Unadjusted 95% confidence intervals are based on the risk difference of cardiovascular death or hospitalization for rosiglitazone minus active control using a normal approximation. Interaction tests are based on unadjusted 95% confidence intervals for the difference in treatment effects between the two subgroup levels (level 1 minus level 2). Bubble area in the left panel is proportional to the total number of patients within each subgroup level. Data from [15].

ANALYSIS: SUBGROUPS

- Rare endpoints will require meta-analysis
  - Sufficient power for meaningful inference
  - More precise estimates of response within subgroups
  - Assess consistency of subgroup response across studies (replication)
- Meta-analyses [16-18]
  - Pre-planned
  - Assess heterogeneity and poolability of the included trials
  - Include all appropriate studies to avoid biased conclusions

CLINICAL TRIALS: ADVANTAGES

- High quality, prospective and uniform data collection, fastidious review, and diligent cleaning
  - Coding and data standards
  - Centralized labs and event adjudication
- Rich and multifaceted
  - Can write detailed narratives of AEs
- More straightforward and reliable comparison of treatments and estimation of incidence and prevalence
  - Concurrent control groups
  - Randomization
  - Blinding

CLINICAL TRIALS: DISADVANTAGES

- Expensive
- Animal studies may not be predictive of the effect in humans
- Challenging to power studies for safety endpoints
  - Detecting rare events and/or moderate safety shifts
- Enrolled patients may not be representative of population
  - Limited concomitant therapies
  - Limited co-occurring disease
  - Compliance to medication

CONCLUSIONS

- Proactively plan for a comprehensive safety evaluation at the start of any development program
  - Distinguishing between anticipated and unanticipated events
  - Consider the effects of patient exposure
  - Utilize appropriate multiplicity adjustment
  - Utilize proper meta-analysis
    - Sufficient power
    - More precise estimates within subgroups
    - Examining the consistency of findings across subgroups and trials
  - Data visualization to efficiently review and summarize data
- Move from summarizing safety events as they occur to predicting their occurrence based on biomarkers [19]

REFERENCES

1. Zink RC, Marchenko O, Sanchez-Kam M, Ma H & Jiang Q. Sources of safety data and statistical strategies for design and analysis: Clinical trials. Submitted to *Therapeutic Innovation & Regulatory Science*.
2. Izem R, Sanchez-Kam M, Ma H, Zink RC & Zhao Y. Sources of safety data and statistical strategies for design and analysis: Post-market surveillance. Submitted to *Therapeutic Innovation & Regulatory Science*.
3. Marchenko O, Russek-Cohen E, Levenson M, Zink RC, Krukas-Hampel M & Jiang Q. Sources of safety data and statistical strategies for design and analysis: Real world insights. Submitted to *Therapeutic Innovation & Regulatory Science*.
4. Ma H, Russek-Cohen E, Izem R, Jiang Q & Marchenko O. Sources of safety data and statistical strategies for design and analysis: Transforming data into evidence. Submitted to *Therapeutic Innovation & Regulatory Science*.
5. Crowe BJ, Xia HA, Berlin JA, et al. (2009). Recommendations for safety planning, data collection, evaluation and reporting during drug, biologic and vaccine development: a report of the safety planning, evaluation, and reporting team. *Clinical Trials* 6: 430-440.
6. Gould AL, ed. (2015). *Statistical Methods for Evaluating Safety in Medical Product Development*. Chichester, United Kingdom: John Wiley & Sons Ltd.
7. Jiang Q & Xia HA, eds. (2014). *Quantitative Evaluation of Safety in Drug Development: Design, Analysis and Reporting*. Boca Raton, Florida: CRC Press.
8. US Food and Drug Administration. (2006). Guidance for clinical trial sponsors: Establishment and Operation of Clinical Trial Data Monitoring Committees.
9. US Food and Drug Administration. (2015). Guidance for industry: Safety assessment for IND safety reporting (draft).

REFERENCES

10. Benjamini Y & Hochberg Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B* 57: 289–300.
11. Kantarjian H, Shah NP, Hochhaus A, et.al. (2010). Dasatinib versus imatinib in newly diagnosed chronic-phase chronic myeloid leukemia. *New England Journal of Medicine* 362: 2260-2270.
12. Ivanova A, Marchenko O, Jiang Q & Zink RC. (2017). Safety monitoring and analysis in oncology trials. In: Roychoudhury S & Lahiri S, eds. *Statistical Challenges in Oncology Clinical Development*. Boca Raton, Florida: CRC Press.
13. Haley EC, Kassell NF & Torner JC. (1993). A randomized controlled trial of high-dose intravenous nicardipine in aneurysmal subarachnoid hemorrhage. *Journal of Neurosurgery* 78: 537-547.
14. Committee for Medicinal Products for Human Use (CHMP). (2014). Guideline on the investigation of subgroups in confirmatory clinical trials (draft). European Medicines Agency.
15. Home PD, Pocock SJ, Beck-Nielsen H, et al. (2009). Rosiglitazone evaluated for cardiovascular outcomes in oral agent combination therapy for type 2 diabetes (RECORD): A multicentre, randomised, open-label trial. *Lancet* 373: 2125–2135.
16. US Food and Drug Administration. (2008). Guidance for industry: Diabetes mellitus — Evaluating cardiovascular risk in new antidiabetic therapies to treat type 2 diabetes.
17. Marchenko O, Jiang Q, Chakravarty A, et al. (2015). Evaluation and review of strategies to assess cardiovascular risk in clinical trials in patients with type 2 diabetes mellitus. *Statistics in Biopharmaceutical Research* 7: 253-266.
18. Marchenko O, Jiang Q, Chuang-Stein C, et al. (2017). Statistical considerations for cardiovascular outcome trials in patients with type 2 diabetes mellitus. *Statistics in Biopharmaceutical Research*. DOI:10.1080/19466315.2017.1280411.
19. US Food and Drug Administration. (2015). Drug safety priorities: Initiatives and innovation (2015-2016).