Subgroup Analysis Findings In Safety Data With No Statistical Power: What Do These Really Mean?

Melvin S. Munsaka*

Abstract

Various regulatory guidelines require safety data to be presented by subgroups of interest, including gender, age, race, and other factors of interest. For the most part, subgroup analyses of safety data tend to be more appropriately performed in the context of integrated analyses of safety for the obvious reason of the availability of large amounts of data which are useful, for example, in getting better estimates of incidences of adverse events. Integrated analyses of safety also facilitate for assessment of trends in subgroups which may not be possible with study level data. The subgroup analyses are typically a repetition of the tabular presentations done for the entire data. It is often not clear as to the purpose of these subgroups analyses, but could include addressing questions such as, are adverse events the same across subgroups and within and between treatments; is the time of occurrence of adverse events the same across subgroups, and so on. In this discussion, we will consider the question of subgroup analyses within the context of safety data and highlight some of the approaches used. Challenges related to the interpretation of findings from subgroup analyses of safety data will also be discussed.

Key Words: Subgroup, Integrated Analysis of Safety

1. Introduction

A variety of regulatory guidance documents recommend safety data to be presented by subgroups of interest. These are often prespecified and include demographic factors such as gender, age, race, and other pertinent factors of interest such as baseline disease characteristics. In general, subgroup analyses of safety data tend to be more appropriately performed in the context of integrated analyses of safety for the obvious reason of the availability of large amounts of data which are useful, for example, in getting better estimates of incidences of adverse events. Integrated analyses of safety also facilitate the assessment of trends in subgroups which may not be possible with study level data. The subgroup analyses are typically a repetition of the tabular presentations done for the entire data. However, it is often not clear as to the purpose of these subgroups analyses, but could include addressing questions such as, are adverse events the same across subgroups and within and between treatments; is the time of occurrence of adverse events the same across subgroups, and so on. In subsequent sections, we consider the question of subgroup analyses in the context of safety data and highlight some of the approaches used. Challenges related to the interpretation of findings from subgroup analyses of safety data will also be discussed.

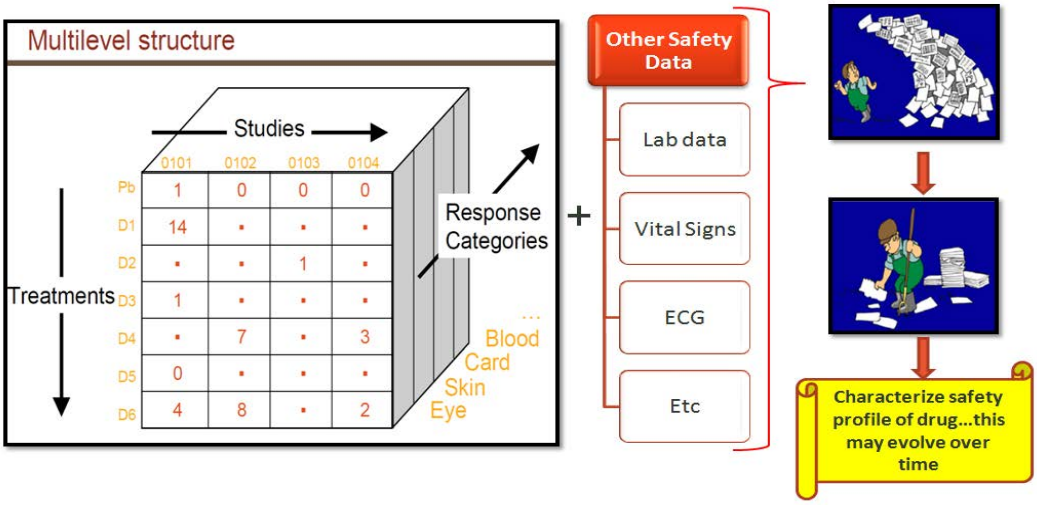
2. Safety Data in Clinical Trials

Compared to efficacy data, safety data tend to be much more complex and highly interrelated. Common safety data domains include adverse events (AEs), clinical laboratory data, vital signs data, and electrocardiogram (EKG or ECG) data. Various safety outcomes can also be derived from these core data domains along with other pertinent data, for example composite endpoints used in cardiovascular safety. Other relevant data used in safety in

*Takeda Development Center North America, Inc., One Takeda Parkway, Deerfield, IL 60015

assessing safety include demographics, drug exposure information, concomitant medications, concurrent disease, and so on. The challenges in the data and analysis include the multidimensional character of the data and their inter-relatedness, see Figure 1.

Figure 1: Multidimensional nature of safety data. Source: Modified from Kerman, et al 2007



Various, laboratory data can be a manifestation of potential safety concerns. For example, increased levels of alanine aminotransferase (ALT), aspartate aminotransferase (AST), total bilirubin, and other hepatic safety parameters can be indicative of potentially compromised hepatic safety. Pathological features of diseases often lead to asymmetric non-normal distributions, high variability in safety related measurements, and highly heterogeneous subpopulations. Thus, it may not be appropriate to analyze safety data using conventional methods. Additionally, key safety endpoints of concern may not be known a priori and it is not uncommon to observe unexpected safety occurrences.

There are also challenges in reporting and in general, typical study-level clinical trial data are generally not sufficient to conclusively assess safety. Another characteristic of safety data is that there is often large volumes of output which are time consuming to generate. This is often coupled with challenges in comprehension, interpretation, and communication. Often descriptive tabular outputs with lots of exploration and review of individual patient data are primary source for safety assessment. Rarely is there comprehensive analytical approaches and inference to better ascertain the safety profile of the drug and which can aid in decision making.

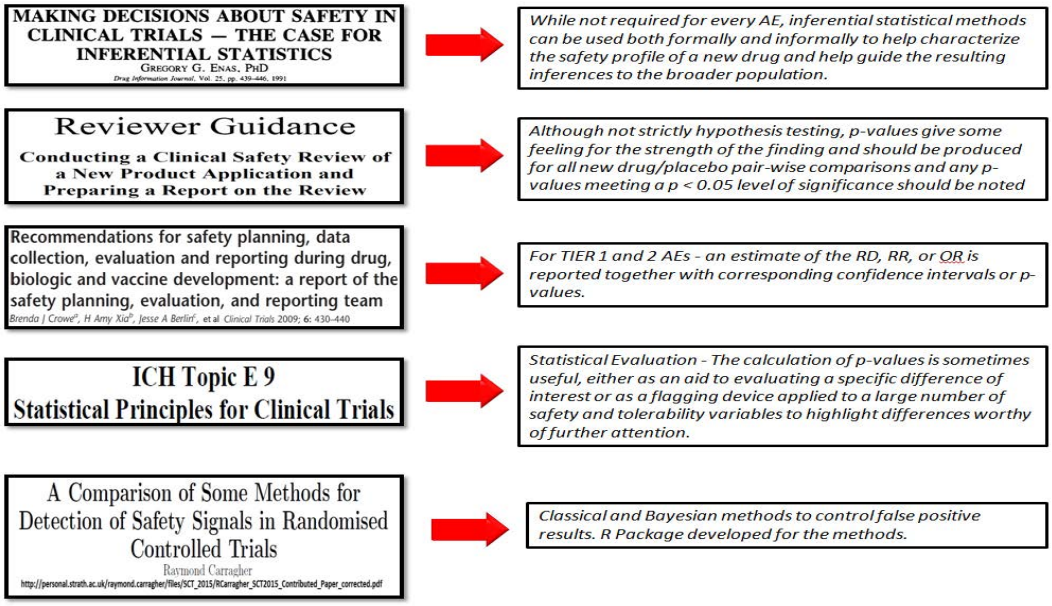
3. Statistical Inference of Safety Data

The question of statistical inference in safety data is one that is often considered controversial. There are many questions and challenges and controversies and varying opinions when it comes to statistical inference of safety data. There are questions on what, when, and how to perform inference in safety data coupled with challenges in interpretation given the many statistical challenges and in general a difficult statistical testing framework. The majority of clinical studies are often powered to assess efficacy, except in those cases where the primary outcome of interest is safety, such as in the case of cardiovascular outcome studies in diabetes. There are many arguments that have been put forward, see for example Huster (1991), against performing any form of inference when looking at safety data

in general. These include lack of well-defined and prespecified hypotheses, multiplicity issues, the question of no adjustment versus too much adjustment, false positive and false negative findings, multiple sources of data, low power, data complexity and lack of standard criteria for adverse event concerns. There is usually inadequate information to specify the safety endpoints, and to characterize the population at risk at the study level. Also, the cost of studying any specific adverse event definitively in a single study would be prohibitive. Thus, safety information is monitored throughout the efficacy trials and the safety profile of a drug is best described across trials. The manner in which safety information is collected often leads to questions of validity regarding formal statistical inference and can be difficult to tackle, especially with traditional frequentist approaches.

On the other hand, there is a clear recognition to carry out some level of inference in safety data. Enas (1991) argued that while not required for every adverse event, inferential statistical methods can be used both formally and informally to help characterize the safety profile of a new drug and help guide the resulting inferences to the broader population. He suggested that the particular setting or phase of clinical drug development can dictate to some degree whether description with or without formal inference is appropriate. Enas also outlined some recommendations for safety model building. He also pointed out to the use of appropriate methods to look for treatment by subgroup interactions, especially when looking at the entire safety database combining all controlled trials. As a matter of fact, many regulatory documents and other publications on safety data, directly or indirectly, point to an expectation of some level of inference in safety data, see for example Figure 2.

Figure 2: Some Recommendations in the Statistical Inference of Safety Data



Additionally, there have been efforts to address some of the statistical challenges in statistical inference of safety data, see for example, Heyse and Mehrotra (2004) and Mehrotra and Adewale (2012). In this discussion, we take the position that statistical inference is a plausible thing to do when looking at safety data. Indeed, the use of inference appears to be the current norm in the analysis of safety data. However, caution should be exercised in the findings from such inference and must be balanced with clinical implications and plausibility.

4. Some Thoughts on Subgroups

4.1 General Observations

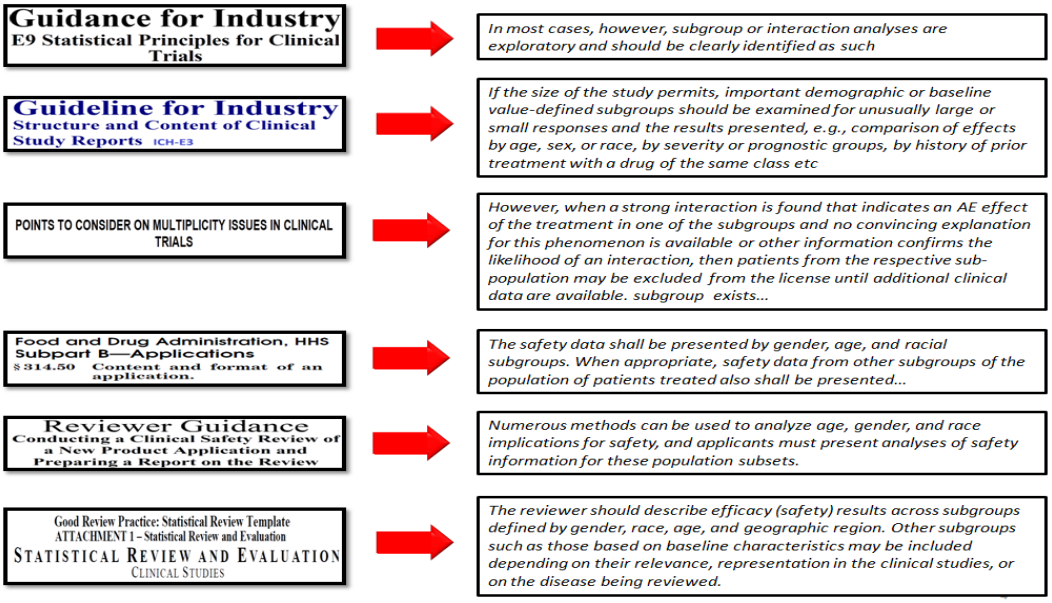
There is a lot of literature and somewhat vague universally agreed to guidance on how to perform subgroup analyses in practice. For the most part, the advice that one finds is with regards to what to and what not to do when dealing with subgroups along with proposals on how to address various statistical challenges. There is no consensus on how to statistically address the various challenges simultaneously and adequately. It is recognized however that findings from subgroup analyses can rarely be definitive. It is also recognized that there is general agreement that subgroup analyses should be done because subjects entering a clinical trial are not homogeneous and treatment for a subject with specific characteristics may differ from the average effect. Further, there may be specific subgroups of patients for whom treatment is more or less effective or harmful relative to the overall target population. Thus, there is need to assess the consistency or robustness of conclusions drawn among different subgroups defined by multiple characteristics and to address concerns on efficacy and/or safety in these subgroups. Subgroups analyses can be used to generate hypotheses on the drug effects for future studies in patients with different risk factors or other characteristics. Besides, there is clearly a recognized scientific and ethical obligation to identify such subgroups as this has a direct bearing on patients since doctors recommend treatment of individual patients on the basis of available evidence. More generally, it clearly recognized that there is need to improve how subgroups analyses are performed and interpreted, see for example Sargent, 2006.

4.2 Subgroup Analysis in Safety Data and Guidance

In safety, the objective of the analysis is to get an understanding of subjects with desirable outcomes for one or more safety endpoints and also identify subjects with increased risk due to exposure to treatment. The primary purpose of subgroups analyses in safety is to detect unusual subgroups/safety signals, identify safety problems that are limited to a subgroup of patients, or that are more commonly observed in a subgroup of patients. Indeed, various guidance documents do point out to the need to perform subgroups analyses for safety data, see some examples in Figure 3.

Additionally, these analyses should provide a more informative summary of the observed data and assess treatment heterogeneity on safety. A major problem with subgroup analyses as noted earlier is that the typical single clinical trial is rarely appropriate for assessing safety in subgroups, even for large trials. Complete assessment of safety often turns to the use of pooled safety data in order to evaluate safety in subgroups. A major barrier in subgroup analyses in safety is that there is a lot of potential for misinterpretation of subgroup results and it is prudent that findings from subgroup findings be treated with caution.

Figure 3: Some Recommendations in the Statistical Inference of Safety Data



5. The Integrated Analysis of Safety (IAS)

The specific guidance for integrated analyses of safety are provided in CFR 21 314.50 (d) (5) (vi) (a). Of note, it is stated that:

CFR 21 314.50 (d) (5) (vi) (a)

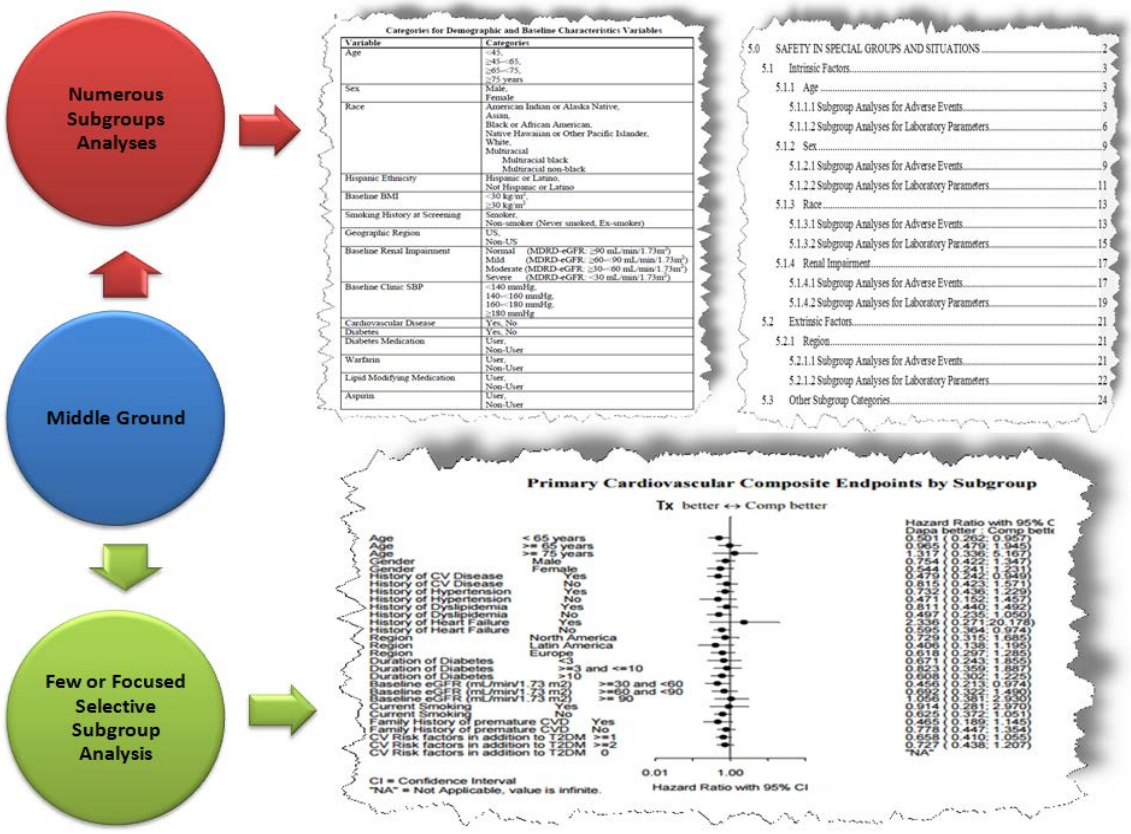
The applicant shall submit an integrated summary of all available information about the safety of the drug product, including pertinent animal data, demonstrated or potential adverse effects of the drug, clinically significant drug/drug interactions, and other safety considerations, such as data from epidemiological studies of related drugs. The safety data shall be presented by gender, age, and racial subgroups. When appropriate, safety data from other subgroups of the population of patients treated also shall be presented, such as for patients with renal failure or patients with different levels of severity of the disease. A description of any statistical analyses performed in analyzing safety data should also be included· · ·

Integrated analyses of safety differ from study level analysis due to large amounts of data. These analyses are usually performed on predefined groupings (pooling) of studies with common elements. This is often detailed in the integrated analysis of safety statistical analysis plan where the pooling strategy is discussed. The pooling strategy needs to take into account various factors, such as the designs (e.g., double-blind versus open label), treatment, and duration of exposure, and so on. The basic idea is to pool data from relevant/similar studies and summarize the data as if they came from one source. Pooling data can help improve the precision of incidence estimates especially for rare adverse events. It also enables assessment of trends in small subgroups of patients, such as the elderly, that may not be possible with study-level data. Outputs from pooled analyses are used to populate various sections of the common technical document (CTD), including sections 5.3.5.3, 2.7.4, and 2.5 and the label. It is important to exercise caution when looking at analyses of safety based on pooled data and results should be cautiously interpreted and they can lead to challenges in conclusions drawn that are based on naive cumulative information.

6. Common Practice in Subgroup Analyses in Integrated Analysis of Safety

Within the the integrated analysis of safety, there two broad approaches. The first approach is the extreme scenario where numerous subgroup analyses are performed using a wide range of factors, both intrinsic and extrinsic. In this extreme case, every tabular output for the entire safety population will have corresponding tabular outputs for each subgroup factor considered which leads to massive amounts of outputs. On the other extreme, only few or focused selective subgroup analyses are done. One can also take a middle ground where a few required or standard subgroups analyses are performed along with those pertaining to safety concerns of special interest, see Figure 4 below.

Figure 4: Some Recommendations in the Statistical Inference of Safety Data



7. Issues and Challenges of Subgroup Analysis in Safety Data

Clearly, in the extreme case, the volume of tabular outputs can be such that it becomes challenging to make sense of the subgroup information since the resulting outputs will be numerous spanning across various safety domains. Besides, there are unclear expectations from guidance documents as to what to make of these subgroups. Even with prespecification, this cannot be done with completeness since it can never completely know in advance some of the safety issues. Further, subgroup findings without inference concluding that safety in subgroups is consistent with overall result can be questionable and interpretation of subgroup findings difficult.

Often when subgroup analyses are reported, the tendency is to report those findings exhibiting the largest treatment effects. Additionally, subgroup findings are often over interpreted or not adequately acknowledged which can lead to further research that is misguided

or, worse, to suboptimal patient care. It is also evident that subgroups analyses are often not definitive about implications on patient management and which of them have a direct bearing on the label. Another notable problem is the lack of coherence and consistence in reporting and what to do with the expected versus the unexpected finding. A further complicating factor is that there can be many subgroup sets and combinations thus making it even more challenging particularly with sparse data as often noted in adverse events.

Further, biological phenomenon may be poorly understood to make sense of the observed results from subgroup analysis. For continuous variables, the use of cut-off points can be arbitrary and different cut-off points could lead to different subgroups results. Another concern is that comparisons are often not consistently done, e.g., male versus female within treatments or male versus male across treatments. In reports, the verbiage used often compounds the problem further with the use of terms such as *consistent findings among subgroups, similar incidence, same incidence, no notable differences, no clinically meaningful or relevant differences , and identical incidence rates* and so on.

8. What Can be Done Better

8.1 Analysis and Presentation

Subgroup analysis of safety data should aim to identify patient needs by applying effective methods of analysis and clear reporting of subgroups and sensible interpretation. It is ideal to focus on a small number of clinically relevant subgroups. In safety, subgroups are often defined by factors that are prognostic for safety risks, e.g., elderly patients, patients with renal or hepatic impairment, diabetics, etc. Justification for their relevance should be related to therapeutic class or the mechanism of action. Focused subgroups analysis can, for example, be based on adverse events of special interest or one can base this on TIER 1 and TIER 2 adverse events, see for example, Crowe 2009. The use of graphical methods can be used to depict consistency of results in subgroups. It is also recommended to perform subgroup analyses to identify and quantify predictable and new safety problems commonly observed in subgroups of patients in a therapeutic area. Subgroups analysis of safety data must also be accompanied by correct interpretation coupled with appropriate inference and a systematic discussion from a statistical inference point of view. It is also important that the correct questions about subgroups are being asked and addressed (see for example, Chow and Liu, 2013), such as:

- Are subgroup findings consistent with overall results?
- Are the adverse event rates the same across a subgroup for patient taking the drug?
- Within subgroup levels, are adverse event rates the same across treatment groups?
- Is there a consistent association between the treatment group and the adverse event response across levels of a subgroup?
- Does the subgroup or factor predict an adverse event response?
- Is the time of occurrence of the adverse event the same across levels of a factor?

8.2 On Methodological Approaches

Regarding methodological approaches, one consideration is perhaps ask a different question, namely, who are the subjects experiencing the events (e.g., Harrel, 2005, Zink, 2010, Su et al, 2009, Docampo, E. et al, 2013) and apply appropriate methodology to identify

the subjects. There is a good level of mathematical sophistication for handling challenges in subgroup analyses. Hence, efforts should be made apply these methods in the analysis of subgroups in safety data. Every attempt should be made to use appropriate methods to address statistical challenges (e.g., address multiplicity problem via Bayesian approaches and other approaches, see for example, Jones et al, 2011, Callagher, 2014, Jones H.E. et al., 2011, Mehrotra and Adewale, 2012, Ohlssen, 2009, 2013). The use of informative graphical approaches should also be used to better depict subgroups in safety data (see for example CTSPedia Website, Gersonides Website). Another potential approach to address subgroup analyses would be to use meta-analysis based subgroup analysis which can be used to account for cross-study heterogeneity and which also take into account prognostic covariates of interest, see for example Stein-Cheung et al, 2015.

9. Summary and Conclusion

In some of the current practices, the value of subgroups analysis within the integrated analysis of safety is more or less insufficient and incomplete from a patient perspectives and there is room for improvement. The subgroup analyses as currently done in the integrated analysis of safety may have no bearing on patient management and care and may not provide guidance to physicians. There are viable ways to make subgroup analysis reporting in the integrated analysis of safety better and more useful and informative from a patient care perspective. Also, as noted by O'Neill (2008), there is a good amount of sophistication out there some of which can be used to address various challenges in the analysis of safety data and in subgroup analysis within safety data.

REFERENCES

Assmann, S. F., Pocock, S. J., Enos, L. E., and Kasten, L. E., (2000), "Subgroup analysis and other (mis)uses of baseline data in clinical trials," *Lancet*, 355, 1064–1069.

Bornkamp, B., Ohlssen, D., and Schmidli, H. (2014), "Estimating the treatment effect for an already selected subgroup," Workshop Verbundprojekt BIMIT.

Carragher, R. (2014), "A Comparison of some methods for detection of safety signals in randomised controlled trials," `http://personal.strath.ac.uk/raymond.carragher/files/SCT_2015/RCarragher_SCT2015_Contributed_Paper_corrected.pdf`

Chow, S-C. and Liu, J-P (2013), "Design and Analysis of Clinical Trials: Concepts and Methodologies, 3rd Edition", Wiley.

Crowe, B., 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.

CTSPedia Website - `https://www.ctspedia.org/do/view/CTSpedia/StatGraphHome`

Dmitrienko, A. and Lipkovich, I. (2014), "Exploratory subgroup analysis: Post-hoc subgroup identification in clinical trials,". EMA Expert Workshop.

Docampo, E. et al, (2013), "Cluster Analysis of Clinical Data Identifies Fibromyalgia Subgroups", *PLOS ONE*, 8, e74873.

EMA - "Points to consider on multiplicity issues in clinical trials."

Enas, G.G. (1991). "Making decisions about safety in clinical trials - the case for inferential statistics". *Drug Information Journal*, 25, 439–446.

FDA - HHS, 314.50 "Content and format of an application."

FDA - "Good Review Practice: Statistical Review Template."

FDA - "Reviewer Guidance - Conducting a Clinical Safety Review of a New Product Application and Preparing a Report on the Review".

FDA - "Guidance for Industry - E9 Statistical Principles for Clinical Trials."

FDA - "Guidance for Industry ICH-E3 Harmonized tripartite guideline structure and content of clinical study reports."

Foster, J., Taylor, J. and Na, B. (20XX), "Subgroup Identification in Randomized Clinical Trial Data Using Random Forests and Regression Trees."

Gersonides Website - `http://www.gersonides.com/r/`

Harrell, F. E., (2005), "Exploratory Analysis of Clinical Safety Data to Detect Safety Signals," http://biostat.mc.vanderbilt.edu/wiki/pub/Main/FHHandouts/gsksafety.pdf.

Huster, W. J. (1991), "Clinical Trial Adverse Events: The Case for Descriptive Techniques," *Drug Information Journal*, 25, 447–456.

Jones, H. E., Ohlssen, D. I., Neuenschwander, B., Racine, A., and Branson, M., (2011), "Bayesian models for subgroup analysis in clinical trials", *Clinical Trials*, 8, 129–143.

Kerman, J., Neuenschwander, B., and Branson, M., (2007), "Bayesian Monitoring of Drug Safety Data", ROeS meeting., http://www.meduniwien.ac.at/ROeS/ROeS_Seminar_Bern_2007/talks/ROeS2007_Kerman.pdf

Mehrotra, D. V. and Adewale, A. J., (2012), "Flagging clinical adverse experiences: reducing false discoveries without materially compromising power for detecting true signals,", *Statististics in Medicine*, 31 1918-1930.

Mehrotra, D. V. and Heyse, J.F. (2004), "Use of the false discovery rate for evaluating clinical safety data," *Statistical Methods in Medical Research*, 13, 227–238.

Ohlssen, D., et al (2009), "A case-study examining the use of Bayesian methods for subgroup analysis in Clinical Trials," http://www.mdanderson.org/education-and-research/departments-programs-and-labs/departments-and-divisions/division-of-quantitative-sciences/pdf/davidohlssen.pdf.

Ohlssen, D. (2013), "Bayesian approaches to subgroup analysis, selection problems and signal detection in drug development."

Pocock, S. J., Assmann, S. E., Laura E. Enos, L. E., and Kasten, L. E. (), "Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems," *Statististics in Medicine*, 21, 2917–2930

Rothwell, P. M., (2005), "Subgroup analysis in randomised controlled trials: importance, indications, and interpretation," *Lancet*, 365, 176–86

Sargent, D., (2006), "Subgroup analyses: can We smooth out the Rough Edges?", Mayo Clinic.

Stein-Cheung, C., et al (2015), "Detecting safety signals in subgroups," Chapter 16 in *Quantitative Evaluation of Safety in Drug Development Design, Analysis, and Reporting*, by Q. Jiang and H. A. Xia (eds), Dekker.

Su, X., Tsai, C.L., Wang, H., Nickerson, D.M., and Li, B. (2009), "Subgroup analysis via recursive partitioning," *Journal of Machine Learning Research*, 10, 141–158.

Wang, R. et al, (2007), "Statistics in Medicine Reporting of Subgroup Analyses in Clinical Trials," *New England Journal of Medicine*, 357, 2189–2194.

Zink, R., et al (2010), "Considerations for Subgroup Identification of Patients With Enhanced Treatment Response in Clinical Trials."