Quality Control Measures in Clinical Trials: Risk-Based Monitoring and Central Statistical Monitoring

Von der Naturwissenschaftlichen Fakultät
der Gottfried Wilhelm Leibniz Universität Hannover

zur Erlangung des Grades

Doktor der Naturwissenschaften (Dr. rer. nat.)

genehmigte Dissertation
von

Firas Fneish, M. Sc.

2023

Referent: PD. Dr. rer. hort. Frank Schaarschmidt

Korreferent: Prof. Dr. sc. hum. Gerhard Fortwengel

Korreferent: PD. Dr. med. Philipp Ivanyi

Tag der Promotion: 21.04.2023

Contributing publications:

1. Fneish, F., Limaye, D., Struever, V., Schaarschmidt, F., Fortwengel, G. (2020).
   Comparison of Non-Commercial Risk-Based Monitoring Tools by Their Application on Clinical Trial Protocols.
   American Journal of Biomedical Science & Research, Vol 8, issue 2, Pages 216 - 220. *DOI: 10.34297/AJBSR.2020.08.001276*

2. Fneish, F., Schaarschmidt, F., Fortwengel, G. (2021).
   Improving Risk Assessment in Clinical Trials: Toward a Systematic Risk-Based Monitoring Approach.
   Current Therapeutic Research, Vol 95. *DOI: 10.1016/j.curtheres.2021.100643*

3. Fneish, F., Ellenberger, D. Frahm, N., Stahmann. A., Fortwengel, G., Schaarschidt, F. (2022)
   Application of statistical methods for central statistical monitoring and implementations on the German Multiple Sclerosis Registry
   Therapeutic Innovation & Regulatory Science (Under Review)

Zusammenfassung

Die Regulierungsbehörden haben die Verwendung eines risikobasierten Monitoringsystems (RBM) in klinischen Versuch gefördert. Die risikobasierte Monitoring beinhaltet neben der Identifizierung möglicher Risiken auch deren Bewertung , um eine gezielte Monitoring zu ermöglichen. Risiken sind dabei definiert als Gegebenheiten, die die Sicherheit der Patienten und die Integrität der Studie beeinträchtigen könnten. Verschiedene Studien haben gezeigt, dass RBM in der Praxis zunehmend eingesetzt wird. Die Anwendung der zahlreichen verfügbaren RBM-Instrumente wurde jedoch nicht untersucht. Die zentrale statistische Monitoring (CSM), die unter die FernMonitoring des RBM-Systems fällt, hat ebenfalls an Aufmerksamkeit gewonnen, da ihre Effizienz bei der Monitoring klinischer Studien anerkannt wurde. Diese Dissertation widmet sich der Verbesserung der Qualitätsbewertungen im risikobasierten Monitoring und im zentralen statistischen Monitoring.

Das erste Kapitel der Dissertation gibt einen Überblick über die klinische Forschung und die Arten von klinischen Studien. Darüber hinaus wird speziell auf die klinische Forschung in klinischen Versuch eingegangen. Es werden die verschiedenen Arten von klinischen Versuch dargestellt, gefolgt vom Managementprozess der Versuch und den Monitoringsaktivitäten. In Abschnitt 2.1 werden die Grenzen der derzeitigen RBM-Instrumente aufgezeigt. Es wird gezeigt, wie unterschiedlich eine Risikobewertung der Ergebnisse einer klinischen Versuch ausfallen kann, wenn sie mit verschiedenen RBM-Instrumenten bewertet wird. Darüber hinaus zeigt dieser Abschnitt die verschiedenen Risiken auf, die von den RBM-Instrumenten abgedeckt werden. Es wird deutlich, dass ein Risikobewertungsinstrument benötigt wird, das jedes Risiko in einer klinischen Versuch abdecken kann. Daher wird in Abschnitt 2.3 eine neue Risikomethodenbewertung (RMA) vorgeschlagen, die auf jede klinische Versuch angewendet werden kann und die Möglichkeit bietet, zusätzliche Risiken in die Bewertung aufzunehmen. Es wird eine Bewertungsmethode vorgestellt, die es den Beteiligten ermöglicht, das Ausmaß eines Risikos zu visualisieren und zu quantifizieren. Dies kann die Beteiligten leiten und ihnen bei der Entscheidungsfindung helfen, ein bestimmtes Risiko durch eine wirksame Maßnahme zu mindern und es zu überwachen. Der theoretische RMA-Ansatz wird in einer Web-App mit einer benutzerfreundlichen Schnittstelle präsentiert, um seine Umsetzung in der Praxis zu erleichtern. In Abschnitt 2.4 wird ein neuer Ansatz zum Nutzen von CSM vorgeschlagen. Er stellt Mehrfachvergleiche der Mittelwerte einzelner Zentren mit dem großen Mittelwert aller Zentren vor. Der Ansatz ist bereits verfügbar und wurde in verschiedenen Kontexten angewandt. Hier wird seine Anwendung vorgeschlagen, um ein auffälliges Zentrum zu erkennen. Da der Ansatz für verschiedene Datentypen verfügbar ist, wird speziell der Vergleich für kontinuierliche, binomiale und ordinale Datentypen gezeigt. In einer Monte-Carlo-Simulationsstudie werden verschiedene Modelltypen, die GM-Vergleiche schätzen, auf die Kontrolle des Typ-I-Fehlers und die höchste Power für ausgeglichene und unausgewogene Szenarien getestet, die in klinischen Studien und Beobachtungsstudien beobachtet werden. Außerdem wird die Validierung des Ansatzes anhand von Real-World-Daten (RWD) aus dem Deutschen Multiple-Sklerose-Register (GMSR) gezeigt. Schließlich wird der Ansatz in Form von Web-Apps vorgestellt, um eine gemeinsame grafisch dargestellte Schlussfolgerungl für unterschiedliche Endpunkte zu ermöglichen.

Schalgworte:

Risikobewertung, Mehrfachvergleiche, ausgelöste Monitoring, Einhaltung des Protokolls, gute klinische Praxis

Abstract

Regulatory authorities have encouraged the usage of a risk-based monitoring (RBM) system in clinical trials. In addition to the identification of possible risks, risk-based monitoring also includes their evaluation to enable targeted monitoring. Risks are defined as conditions that could affect patient safety and the integrity of the study. Various studies demonstrated the increasing usage of RBM in practice. The application of the many RBM tools available has not been investigated. Central statistical monitoring (CSM) which falls under the remote monitoring of the RBM system has also been gaining more attention due to the recognition of its efficiency in monitoring clinical trials. This dissertation is dedicated to improving the quality assessments in risk-based monitoring and central statistical monitoring.

The first chapter of the thesis provides an overview of clinical research and the types of clinical studies. Furthermore, it specifically focuses on clinical research structure, management, and activities in clinical trials. The different types of clinical trials are illustrated, followed by the management process of the trial and monitoring activities. Section 2.1 highlights the limitations of the current RBM tools. It shows how different an outcome risk assessment of a clinical trial can be when assessed with different RBM tools. Furthermore, this section shows the different risks covered within RBM tools. It shows the need for a risk assessment tool that can cover any risk in a clinical trial. Hence section 2.3 proposes a new risk methodology assessment (RMA) that can be applied to any clinical trial with the ability to add additional risks to the assessment. It presents a scoring method that allows stakeholders to visualize and quantify a risk size. This would guide stakeholders and assist them in the decision plan for mitigating a certain risk by an effective measure and monitoring degree in the monitoring plan. The theoretical RMA approach is presented in a shiny web app with a user-friendly interface to ease its implementation in practice. Section 2.4 proposes a new approach for the benefit of CSM. It presents multiple comparisons of individual center means to the Grand Mean of all centers. The approach is available and has been applied in different contexts. Here its implementation to detect a deviating center is recommended. As it is available for different data types, it shows specifically the comparison for continuous, binomial, and ordinal data types. In a Monte-Carlo simulation study, different model types estimating GM comparisons were tested for the control of Type I error and the highest power for balanced scenarios and unbalanced scenarios observed in clinical trials and observational studies. It also shows the validation of the approach on Real-world data (RWD) from the German Multiple Sclerosis Registry (GMSR). Finally, the approach is presented in shiny web apps to facilitate a common graphical conclusion style for different endpoints.

Keywords:

Risk assessment, multiple comparisons, triggered monitoring, protocol compliance, good clinical practice

Abbreviations

Adverse Event	AE
Analysis of means	ANOM
Bayesian Generalized Linear Model	BayesGLM
Bias-reduced Generalized Linear Model	BrGLM
Case Report Form	CRF
Central Statistical Monitoring	CSM
Clinical Research Associate	CRA
Electronic Data Capture	EDC
European Medicines Agency	EMA
False Discovery Rate	FDR
Generalized Linear Model	GLM
German Multiple Sclerosis Registry	GMSR
Good Clinical Practice	GCP
Good Clinical Practice - Inspectors Working Group	GCP-IWG
Grand Mean	GM
International Council for Harmonisation	ICH
Investigational medicinal product	IMP
Multiple sclerosis	MS
National Institute of Health	NIH
Non-parametric GM comparison - Multiple Comparisons and Simultaneous Confidence Intervals	Nparcomp
Principal Component Analysis	PCA
Real-World Data	RWD
Risk Indicator Taxonomy	RIT
Risk Methodology Assessment	RMA

List of Figures:

Contents

Chapter 1

Introduction

1.1 Clinical Studies

1.1.1 Types of clinical research

Clinical studies involve research carried out on human participants to enrich medical knowledge (Food and Drug Administration, 2019). Various types of clinical studies exist (Collier, 2009; Food and Drug Administration, 2018; Fortwengel, 2011). Based on the objective of the study, the type is determined. The National Institute of Health (NIH) classifies clinical studies into two types clinical trials and observational studies (National Institute of Health, 2019).

Research carried out on human beings investigating an intervention is referred to as Clinical trials (M. Friedman et al., 2015). They present the primary way in which researchers determine if a new drug or a medical device is both safe and effective (National Institute on Aging, 2020). They are the golden standard for answering a specific research question (Shamley & Wright, 2017), specifically, they answer questions about health and illness (UK Clinical Research Collaboration, 2014). Clinical trials are the best way to determine whether a treatment works for a specific disease. Furthermore, they inform researchers whether the investigated treatment can be an alternative option to a standard treatment. (e.g., not being effective in some people or causing side effects) (American Cancer Society, 2020). The type of intervention could be an investigational medicinal product (IMP), a new medical device, or a new approach to surgery/therapy (Food and Drug Administration, 2018). While observational studies do not involve any sort of intervention (Song & Chung, 2010). Similar to clinical trials, observational studies do involve human participants, however, there is no direct intervention (Gilmartin-Thomas et al., 2018). A distinguishing characteristic is that the intervention is determined by clinical practice and not through a study protocol (Thiese, 2014; Yang et al., 2010).

1.1.2 Overview of Clinical Trials

Clinical trials are regulated through specific guidelines to ensure the rights, safety, and well-being of trial participants are protected and the reported results are reliable (European Medicines Agency, 2018). Good Clinical Practice (GCP) is an international ethical and scientific guideline used as the primary standard for the conduct of clinical trials developed by the International Council for Harmonization of Technical Requirements for Registration of Pharmaceuticals for Humans (ICH) in 1996 (Imperial College Clinical Research Governance Office, 2007). Since its establishment, it has been modified to tackle unclear, inconsistent, and contradictory definitions (Landray et al., 2017). These guidelines are under continuous improvement regarding their ethical and regulatory challenges (Bhatt, 2010).

The new intervention examined in a clinical trial must pass through several phases. Each phase answers a specific research question regarding the safety and efficacy of the intervention. Figure 1 shows the goals and the study participants required in each phase the IMP must pass through before receiving market approval from regulatory authorities such as the EMA or the FDA (National Institute of Health, 2015; Umscheid et al., 2011). Clinical trials involve an intervention according to a protocol established by clinical investigators. Accordingly, each trial has a specific protocol detailing the study objectives, design, and organization of the trial phase(Chan et al., 2013; Jones & Abbasi, 2004; Rivera et al., 2020).

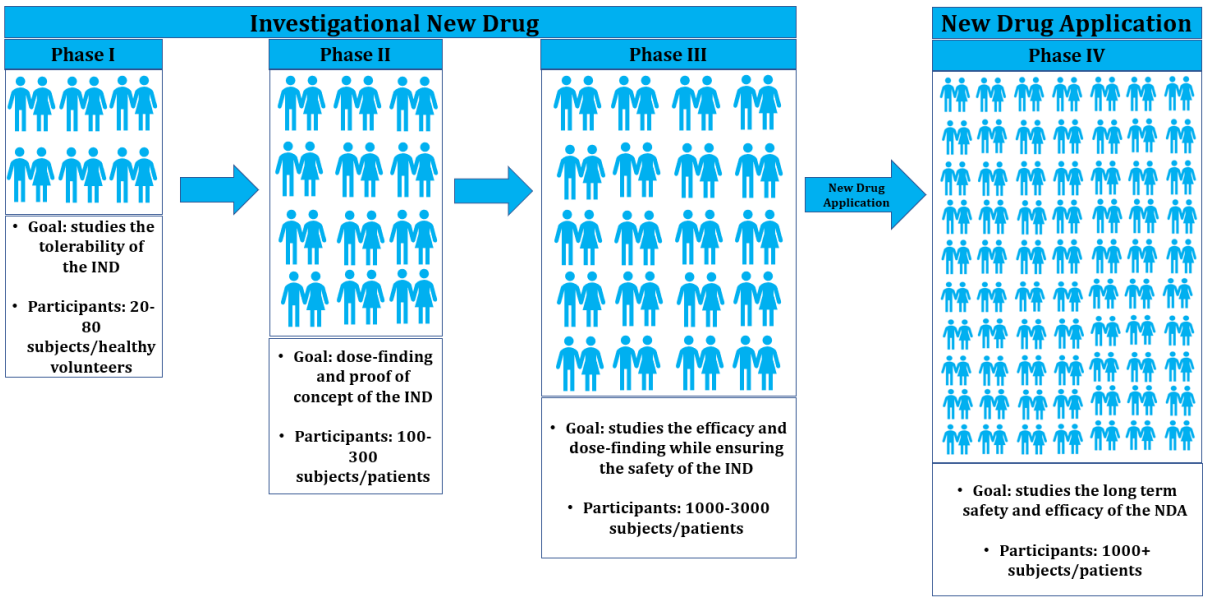


Figure 1: Structural differences and goals of clinical trial phases.

1.2 Clinical trial management

1.2.1 Clinical trial team

The clinical trial team is one of the most essential factors for the conducted trial. All clinical trials require a team of people with specific qualifications to perform their corresponding tasks. Members can come from a wide range of skills and expertise. Throughout the complete trial period, the team must follow the protocol of the trial and the GCP standards (World Health Organization, 2005). A clinical trial team consists of sponsors, principal investigators, study coordinators, biostatisticians, and data managers. Additional individuals such as nurses and pharmacists could be part of the trial team depending on the trial requirements (Baer et al., 2011; Canter & Lewis, 2014). The responsibilities and duties of individuals vary based on their role in the trial. Figure 2 shows the major tasks performed by each research team member. Ensuring the well-being and safety of participants is a common task for all team members from the moment the trial starts till the end, in some cases also the post-trial period. (e.g., follow-up visits) (Fortwengel, 2004).

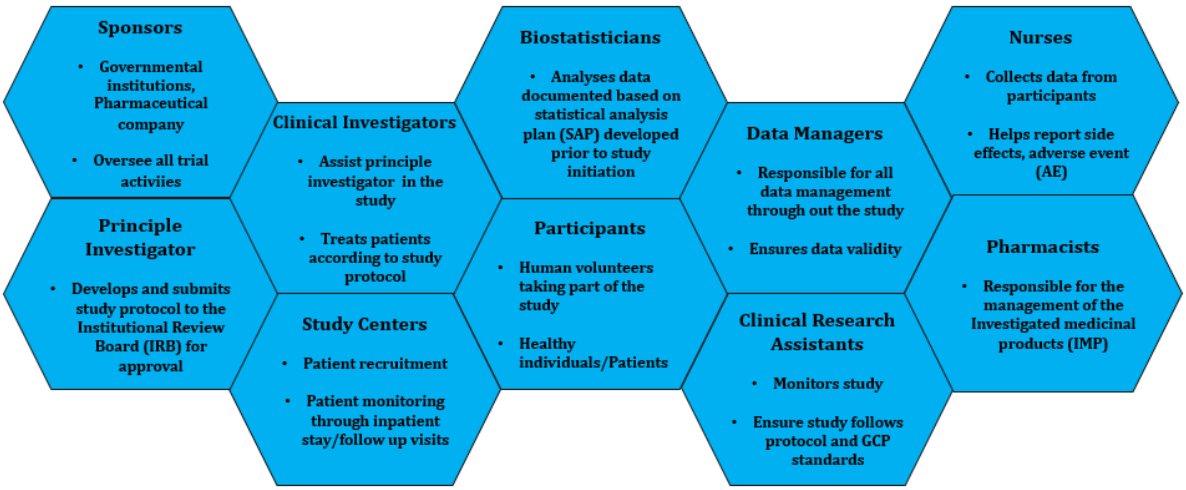


Figure 2: Main responsibilities of clinical trial team members.

1.2.2 General activities in a clinical trial

Trial activities start from the first-day stakeholders decide on running a clinical trial. The sponsor develops the trial protocol detailing the specifics of how the study will be conducted (ICH, 2016). All team members of the trial participate in the protocol development in which they contribute to the section related to their expertise e.g., a

biostatistician calculates the power and the sample size of the trial as well as determines and implements the statistical methods to detect a true difference between treatment groups (Ellen van Bavel, 2016). Following protocol development, a feasibility report is established for recruiting study center(s) in the trial. The clinical research associate (CRA) conducts interviews with the potential clinical investigator(s) from the study center(s). Once the foundation of the trial is established, the sponsor is required to register the trial and request authorization from competent authorities (Dupin-Spriet, 2005; Poolman et al., 2007).

1.2.3 Traditional Monitoring and ICH recommendation

The ICH-GCP requires the sponsor to simultaneously develop a monitoring plan before the trial initiation which the CRA follows (ICH, 2016). Monitoring procedures should be as detailed as possible. (National Institute for Health Research (NIHR) Clinical Trials Toolkit, 2012). This would enhance the participant's safety throughout the trial as well as improve data integrity. Traditional monitoring involves intensive on-site monitoring visits at clinical trial centers and exhaustive SDV of clinical trial data (FDA, 2020).

Onsite monitoring entails visits conducted to study sites before and during the trial. The monitoring team verifies whether a center is following the study protocol. Specifically, a monitor e.g., a CRA checks whether case report forms (CRF) filled comply with the study protocol. Specifically, the monitor verifies the consistency of data collected with participants' medical files (Dupin-Spriet, 2005). Additionally, the monitor inspects for missing data, faulty completion of forms, questionable values, or other protocol violations.

In recent years, clinical researchers have questioned the validity and necessity of traditional monitoring methods (Uren et al., 2013). Many consider it to be an expensive, time-consuming, and resource-demanding activity that does not improve the quality of clinical trial data or the protection of trial participants (Olsen et al., 2016). Onsite monitoring might be realistic or easily implemented in a single-center trial phase I or II, however, it becomes complicated for a large trial or even multi-center trial. Additionally, Clinical trials have developed in different aspects such as complexity, globalization, and technological means. Intensive onsite monitoring showed to be of high cost and does not always identify errors (Baigent et al., 2008a). To tackle these difficulties and the advancement in clinical trial setups, regulatory authorities recommended the

implementation of a novel risk monitoring approach. The ICH recognized this need and released a new GCP guideline requiring sponsors to implement a risk-based approach to improve the effectiveness and efficiency of monitoring. (International Council for Harmonisation, 2018). This facilitates the option for sponsors to divide monitoring activities between onsite and centralized monitoring or a combination of both. The ICH defines centralized monitoring as a remote evaluation conducted at a location other than the study centers conducting the trial. Due to the current advancement in electronic data capture (EDC), centralized monitoring can offer a greater advantage (International Council for Harmonisation, 2018). It incorporates the utilization of statistical methods to detect deviating or erroneous data. Following the recommendation several risk-based monitoring approaches became available. Hurley et al. identified and summarized RBM tools available for clinical trial monitoring (Hurley et al., 2016). However, the state-of-the-art RBM tool was not identified. Similarly, since centralized monitoring became available, the approaches focused on presenting summary statistics, principal component analysis (PCA), and detecting inlier/outlier data.

This thesis focuses on the procedures that fall under the RBM umbrella. In detail, it presents a risk assessment methodology of a trial and a statistical approach to monitoring centers in a clinical trial. First, differences are distinguished between current non-commercial RBM monitoring tools in practice by their application to real clinical trial protocols. Secondly, a thorough risk methodology assessment (RMA) of clinical trials is developed. The proposed method is adaptive, it can include any potential risk. The risk assessment within the existing RBM tools is predefined to fixed risks, lacking the option of tallying new ones. Finally, for the benefit of CSM, the use multiple comparisons of single centers to the Grand mean (GM) of all centers is proposed. It can be used to detect centers that are significantly deviating from GM. In a Monte Carlo simulation study, the approach is assessed for the ability to control type I error (α) and achieve the highest possible power ($1-\beta$) for specific data types in different parameter settings. Further, the approach is demonstrated to real-world data (RWD) from the German Multiple Sclerosis Registry (GMSR).

The following sections describe background information and the basis for research conducted in this thesis. Thus, the findings of the three manuscripts are summarized and set into context.

1.3 Risk-Based Monitoring

1.3.1 State of the art

A risk-based monitoring approach entails the identification of any risk that might influence areas routinely subject during monitoring activities. These risks should be identified at the system and clinical trial level, followed by a systematic evaluation of these risks and their likelihood of occurring and the extent of detecting these errors, and their impact on human subject protection, trial data reliability, and GCP- and protocol compliance (International Council for Harmonisation, 2018). To date, various tools for risk identification have been developed as either paper-based or electronic RBM. These tools have been compared on their characteristics, and their respective strategy to minimize the risk. Jungen et al. formulated a risk indicator taxonomy (RIT) to serve the RBM purpose (Jongen et al., 2016). Hurley et al. compared RBM tools to RIT and found only 12 RBM tools cover risk indicators listed in RIT. However, Hurley et al. indicate that the risks covered within most RBM tools do cover ICH-GCP demands for RBM (Hurley et al., 2016). This research aimed to study the effectiveness of RBM tools and their differences in practice. Specifically, a special interest was focused on the implementation and outcome processes when applied to real clinical trial protocols. .

1.3.2 Application of non-commercial RBM tools

Non-commercial RBM tools were applied (ADAMON (Brosteanu et al., 2009), OPTIMON (Journot et al., 2011), Transcelerate (TransCelerate, 2012), SWISS (Swiss Clinical Trial Organisation, 2014), NORM (NORM, 2015), YEE (Yee, 2017), and MHRA (MRC/DH/MHRA, 2011)) to real clinical trial protocols covering all phases (I-IV) to compare the overall risk assessment of each. Real clinical trial protocols were retrieved from the registry of clinical trials "ClinicalTrials.gov". Moreover, a direct comparison was performed between risks covered between the Transcelerate tool and other RBM tools. Almost all investigated RBM tools showed different overall risk assessments. Furthermore, the risks detected and their impact within each RBM tool were different, hence the outcome mitigation plans of each RBM tool varied as well. The broad differences between the RBM tools indicate that an ideal risk assessment tool is currently missing. Each RBM tool focuses on predefined risk areas lacking the option of adding additional ones. Given the different risks each clinical trial can have, a main setback for the risk assessment is to be fixed on specified

risks. A vital characteristic feature of an RBM tool is the identification and classification of potential risks associated with a planned study. As risks are different from one trial to another, current risk assessments within investigated RBM tools partially fulfill this requirement. Potential risks in a clinical trial can be identified in a well-structured study protocol. Further, risks must be recognized in advance to enable a systematic screening by RBM tools. The potential risks should be weighted to allow for an internal assessment of a risk characteristic and thus enable the generation of a study risk score. Preparatory work on this is still missing and was, therefore, the next step in the further development of RBM systems whose developmental approach then no longer decides which risk level a clinical study has.

1.4 Risk Methodology assessment

1.4.1 ICH-GCP requirement for risk identification and evaluation

A key quality feature of RBM entails a robust risk assessment. The ICH-GCP requires sponsors to identify potential risks critical to the trial process and data (ICH, 2016). It is left to the sponsor to decide on an appropriate system to identify risks in a clinical trial. The ICH specified the criteria each risk should be evaluated on:

- The likelihood of errors occurring. (Probability)
- The extent to which such errors would be detectable. (Detectability)
- The impact of such errors on human subject protection and reliability of trial results. (Impact)

However, the ICH did not specify the standards each of the above criteria should be assessed on. Most of the risk assessment methods within current RBM tools provide an overall assessment score of the trial. Although the ICH did not specify this need, it does play a helpful role in indicating the critical level of the overall risks present in a specific trial. Still, a general limitation of current risk assessment methods within the RBM tools is the lack of transparency of the final decision rule for the determined monitoring plan. In this research, faults detected by the Good Clinical Practice - Inspectors Working Group (EMA GCP-IWG) are considered as the main risks that should be covered by a risk assessment tool (EMA-IWG, 2018). Since clinical trials are diverse in complexity and

structure, hints are provided on how the identification process should be and a rationale for a risk assessment process is established. Specifically, standards for the Impact and detectability criteria of the evaluation process are provided.

1.4.2 Individual assessment of each risk

The ICH indicates that the monitoring team must verify the rights and well-being of human subjects, the reliability of data, and compliance to study protocol and GCP guidelines must be followed. Since these points are the main tasks the monitoring team must observe, they are used as the main aspects for impact criteria of risk evaluation. Additionally, different weights for the individual points are provided. As for the detectability criteria, the detection technique required to point out the risk outcome is suggested, hence either as onsite or remote monitoring. Table 1 shows the assessment category and weights for each criterion of potential risk.

Criteria	Assessment category	Score
Impact	1. Well-being/safety of subjects	3
	2. Reliability of data	2
	3. Compliance with GCP/protocol guidelines	1
Probability	1. Very likely	5
	2. Likely	4
	3. Even chance	3
	4. Unlikely	2
	5. Very unlikely	1
Detectability	1. Onsite Monitoring	2
	2. Remote Monitoring	1

Table 1: Risk assessment criteria. Following risk identification, each risk is evaluated based on the category it impacts, the probability of a risk occurring, and the monitoring technique required for detection (Fneish et al., 2021b).

The main challenge in risk assessment is understanding the consequence of a certain risk on the trial. In the presented RMA, radar charts are utilized to aid the stakeholders in the assessment and decision process for individual risk. This would support them in having a better understanding of the risk effect. The proposed assessment scale shows the importance of the risk in the trial. Together with the radar chart, it gives indications to the assessor how the monitoring plan should tackle the assessed risk. The area under the

radar chart is computed by the assessment criteria of the risk. This means the larger the area under the chart, the more monitoring is required to control the assessed risk. Figure 3 shows an example of a specific risk assessment. Assuming a risk would be having a discrepancy between source data documents and case report forms, this would impact the reliability of data and GCP/protocol compliance, with the assumption it would have a probability score of 3 and a detection technique by performing an onsite visit. Then with the proposed algorithm, this risk has a score of 40% of 100. This score aids the sponsor in deciding the frequent visits to be taken to check this risk. The risk assessment must be an ongoing process performed before, during, and after the trial. Any amendments to the study protocol require a new risk assessment procedure. Additionally, any new faults detected during onsite visits require a reassessment as well. Whereas any faults detected during remote procedures might indicate a trigger for an onsite visit.

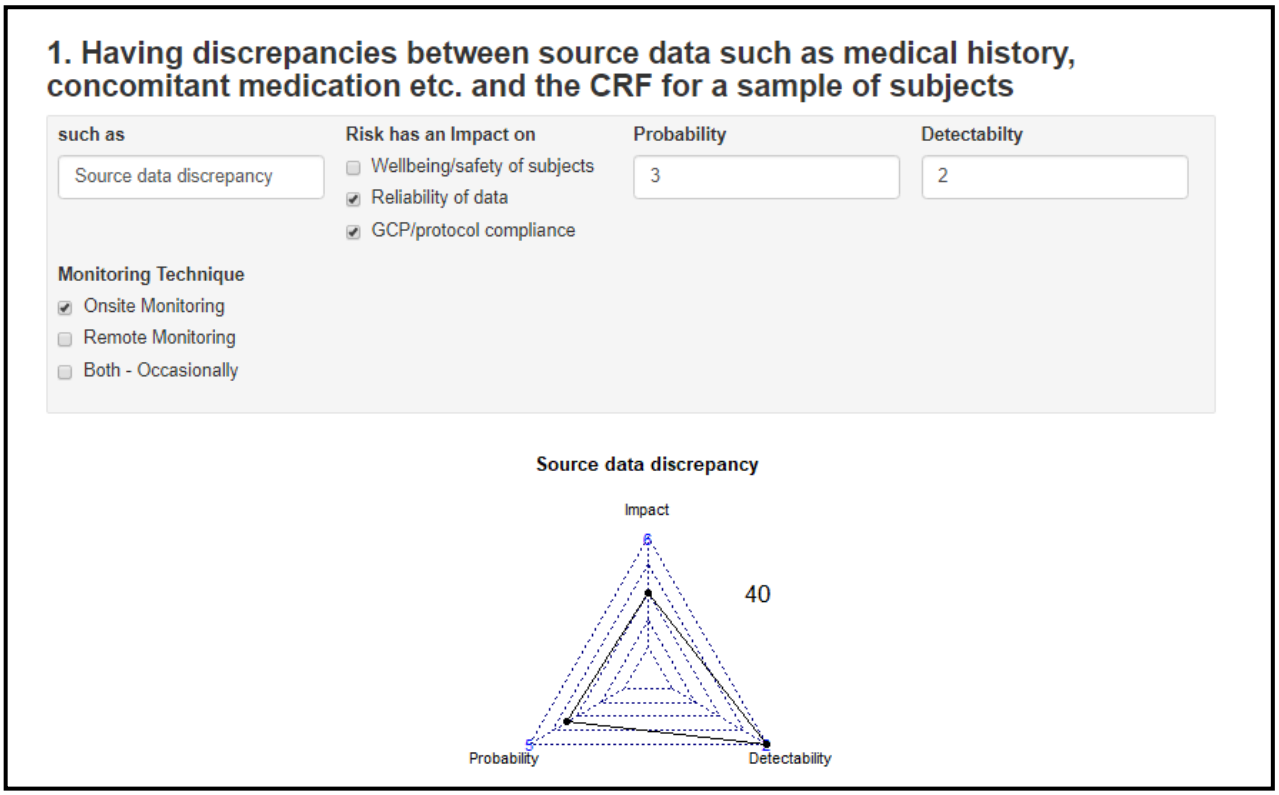


Figure 3: Individual risk assessment. Each risk must be evaluated on Impact, Probability, and Detectability.

1.4.3 Overall score for risk assessment

The overall score of the trial is represented by the average score of the risks assessed. Although it should not play a major role in any decisive measure, it still conveys an idea of how critical a trial is. Thus, an established monitoring plan would be reflected in the overall score. For example, a monitoring plan consisting mainly of remote monitoring techniques where an overall score is 80% means the established monitoring plan does not reflect the assessment procedure. Accordingly, an overall score of 40% would be reflected in the monitoring plan by having more remote procedures and some onsite visits.

Remote monitoring also known as Centralized statistical monitoring (CSM) has been proposed as an effective way and less costly than on-site monitoring (Bakobaki et al., 2012). In the next section, current methods of CSM implementations are presented and a new approach to improve CSM was investigated and validated on real-world data.

1.5 Centralized statistical monitoring

1.5.1 State of the art

The EDC has revolutionized the monitoring strategies in clinical trials enabling the remote monitoring approach. As regulatory authorities indicated CSM was introduced to improve data reliability. It ensures the quality and validity of data collected in multicenter clinical trials (Desmet et al., 2014a). Different approaches have been implemented for CSM (Baigent et al., 2008a; Oba, 2016a; Venet et al., 2012b). These approaches mainly focused on implementing visualization techniques, outlier/inlier detection, data distributions, and principal component analysis (PCA). They have proven to be effective in detecting unusual or fraudulent data. Furthermore, they may indicate some issues in a certain center, but they do not elucidate whether a center deviation is due to chance. The main objective of CSM is improving data reliability (International Council for Harmonisation, 2018), this means CSM should be able to detect single centers that might deviate from a study protocol or detect a center that has misunderstandings concerning data reporting such as adverse events (AEs). These types of deviations will not produce single extreme values in the data, they would rather lead to deviating summary statistics, AE rates, or class frequencies of categorical data. The literature related to CSM does not include a clear

overview in detecting a deviating center that would eventually alarm the monitoring team to initiate an onsite visit to a center due to data variation likely due to chance or not.

1.5.2 Comparisons to Grand Mean

All centers participating in a clinical trial are expected to follow the study protocol. Accordingly, if a center is violating any of the protocol requirements it should be reflected by the data collected and thus detected by comparing the data coming from an individual center to the data of all centers. In this research, the utilization of multiple comparisons of single centers to the Grand Mean (GM) of all centers is proposed. This approach has been applied in fields comparing treatments in laboratory experiments (Silverstein, 1974). It is also available in the analysis of means (ANOM) context for quality control (Pallmann & Hothorn, 2016). Hothorn et al. (2008) provide a general framework software for simultaneous inference procedures in general parametric models, which includes multiple comparisons to the GM procedure (Hothorn et al., 2008a). Konietschke et al. (2015) provide a software for non-parametric multiple comparisons which could perform GM comparisons (Konietschke et al., 2015a). For several data types, this approach allows the detection of centers that are deviating from the GM. A simulation study was carried to investigate whether this contrast can be implemented by different statistical methods for continuous, binary, and ordinal endpoints while controlling type I error and achieving the highest power for balanced and unbalanced designs common in clinical registries and clinical trials.

1.5.3 Model Types and Endpoints

In this research, the GM comparisons is considered for three data types (Continuous, Binomial, and Ordinal). Different statistical approaches were investigated for the performance in the multiple comparisons. Specifically, the control of type I error and the power for performing the contrasts of individual center mean to GM of data for different statistical approaches was investigated.

For a continuous non-normally distributed endpoint, the non-parametric approach available in Nparcomp (Multiple Comparisons and Simultaneous Confidence Intervals) R package was implemented (Konietschke et al., 2012a). Konietschke et al. indicate that the non-parametric method is based on asymptotic results on the distribution of rank statistics. Thus, this hints at a limitation in the non-parametric approach for small sample sizes. For this reason, it was noteworthy to test whether the non-parametric would

manage to control type I error for small sample sizes that could be found in clinical trial phases 1 or 2 where the number of patients is limited. Nparcomp was also implemented for an ordinal outcome and tested for the same scenarios as the continuous endpoint.

A binary outcome such as a patient suffering from an AE is common in clinical trials. Study protocols might be misunderstood, where for example a vaccination reaction to a vaccine might be reported as an AE in a specific center. Another misconception would a misunderstanding of a relapse definition by the clinical trial staff when they report an observed symptom as an indication of a patient relapse. In both examples, the center would be pointed out by GM comparisons. This would trigger the monitoring team to visit the center for further checkups and training. For such binary endpoints, the classical generalized linear model (GLM) with logit link was implemented. However, GLMs have limitations when 0 excess is present (Fneish & Schaarschmidt, 2019a) and since 0s are commonly seen in clinical trials, two alternative methods to account for the 0 problem were investigated. A Bayesian generalized linear model based on the prior distribution (Gelman et al., 2008a) and Bias-reduced generalized linear model which is based on Jeffreys-prior (Kosmidis & Firth, 2021a) are proposed to handle the 0 problem. In the simulations, methods are compared for their performance and ability to control type I error and achieve the highest power when implementing the comparisons to GM for scenarios with and without 0 problem. Figure 4 summarizes the data types and model types considered in this research.

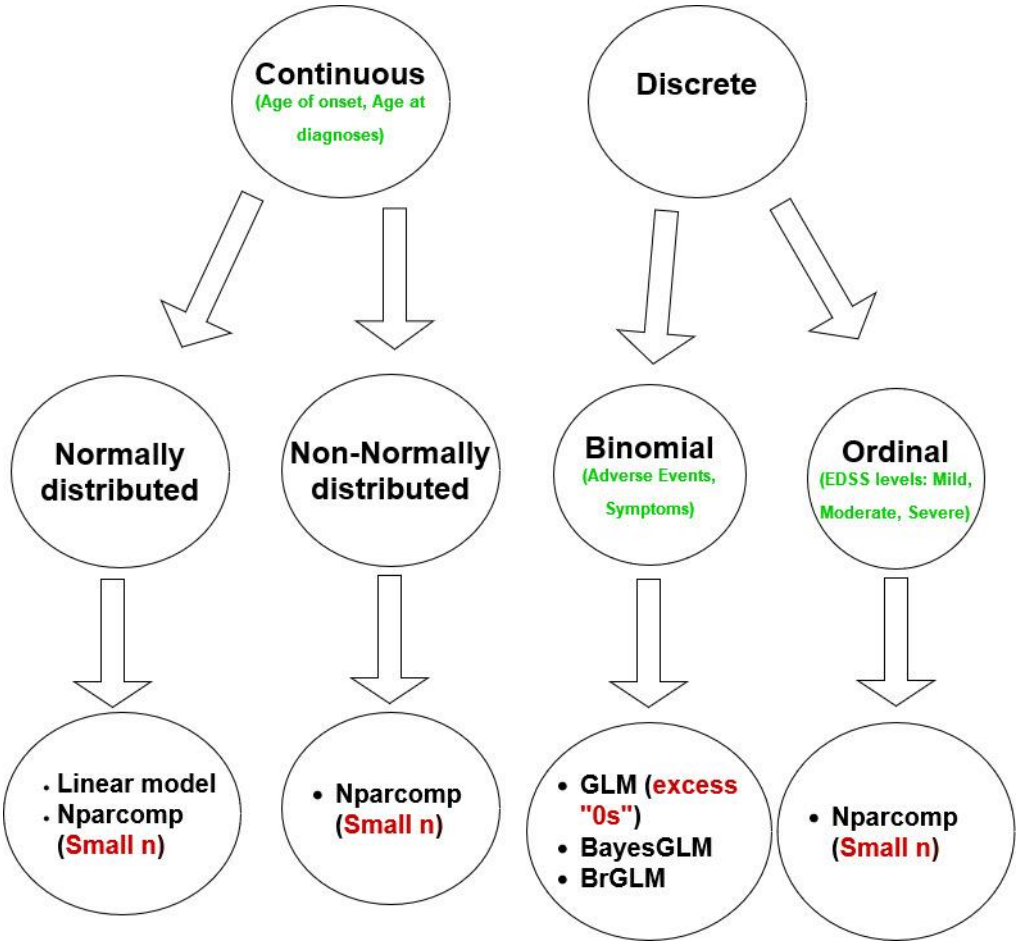


Figure 4: Proposed estimators for each data endpoint. Non-parametric method (Nparcomp), Generalized linear model (GLM), Bayesian Generalized Linear model (BayesGLM), Bias-reduced Generalized Linear Model (BrGLM).

1.5.4 Application on Real-World Data

Current practices of CSM focus on visualization techniques such as the distribution of data in specific centers. RWD found in clinical trials can be very diverse, it might include a few to hundreds of patients within a varying number of centers. The complexity of clinical trials and registries makes it more challenging in identifying a proper methodology for performing CSM. Some visualization methods on data from the GMSR are shown, followed by the demonstration of GM implementation on specific data types. Fifteen largest centers that are part of the GMSR's pharmacovigilance module were included. The data consists of a wide range of variables covering demographic and clinical data such as patient profile, disease status, and medication treatments. Ohle et al. (2021) cover further details on the GMSR (Ohle et al., 2021). In this research, three endpoints are covered (Continuous, Binomial, and Ordinal), the age at diagnoses, data missingness, and diseases severity as continuous, binomial, and ordinal data types respectively. The approach of multiple

comparisons is investigated. Specifically, center means are compared to the Grand Mean of all centers. If a center behaves differently from other centers, then it would be reflected by the reported data. This would indicate to atypical center, specifically when there is a significant mean difference between an individual center mean and the GM. This would eventually trigger an onsite visit if the deviations cannot be justified.

Continuous Endpoint: Age at diagnosis

Figure 5 shows the distribution of patient cohorts for age at diagnoses in each center. This graph clearly illustrates a difference in the distribution among centers. It also shows that some centers include some data outliers. As intuitive as this graph may be, it does not identify a problem in a specific center. It won't directly serve CSM purposes where the focus lies on checking the compliance of centers to study protocol and GCP guidelines.

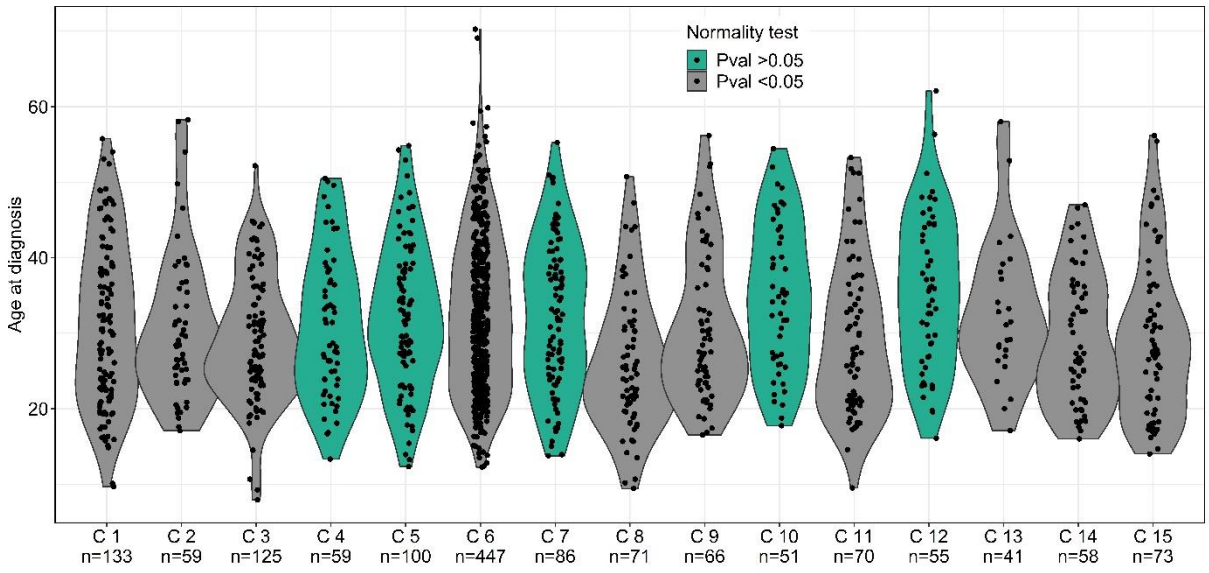


Figure 5: Violin plot including kernel density estimates of age at diagnoses indicating possible violations of normality. *pval*<0.05 indicates a non-normally distributed data.

The implementation to GM comparison clearly illustrates that C 6 has a larger mean compared to the GM whereas C 8 and C 15 have smaller means compared to GM (Figure 6). In a real clinical trial, this would indicate whether a certain center has a cohort significantly different from the inclusion criteria indicated in a study protocol. However, in the GMSR data, this could be justified since some centers might have a specific specialization in which a difference would not be alarming.

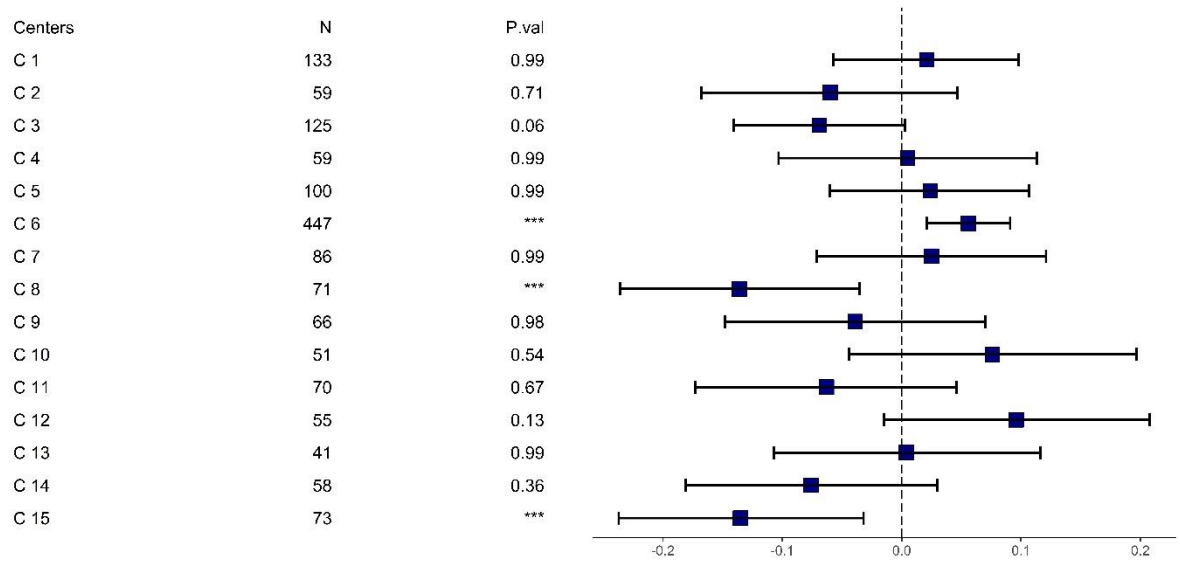


Figure 6: Simultaneous confidence intervals for contrasts of center means with GM for continuous variable age at diagnosis using the non-parametric method (Nparcomp).

Binary Endpoint: Missing data

Poor data quality affects any research and hence decreases its respective power, specifically if it includes a lot of missing data. Although it's best to have complete data, it is not uncommon to have missing's in a data set. For this reason, the GM comparison for such scenarios would serve the detection of a deviating center having more missingness compared to the GM of centers. In this way, it would be pointing out a center that would require some attention for better documentation. Figure 7 shows a binary endpoint for a patient with missingness for age at diagnoses or age at onset variables in each center. It shows the performance of the centers in documenting these variables. Most of the centers have the required documentation needed for almost all their patients, however, some centers have higher proportions of missingness. Statistically, this would indicate whether a center has more deviations of non-documented data than the GM.

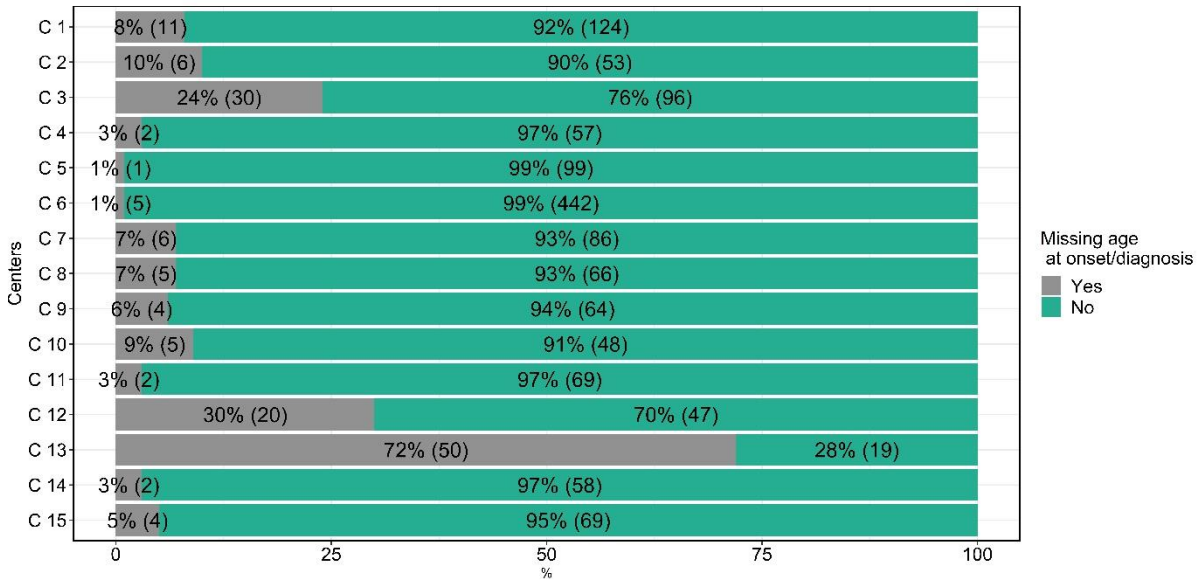


Figure 7: Centers documentation of age at diagnosis and age at onset for each patient (%).

The utilization of GM comparisons to center means indicates that centers C 3, C 12, and C 13 have higher means than the GM of the data. In other words, these centers have more missingness for age at onset or diagnoses than the average (Figure 8). C 6 has a smaller mean compared to the GM, in this context it would indicate a higher documentation performance for this center than the average. The BayesGLM model type for binary endpoint is preferred in these scenarios since 0s could be present in the data.

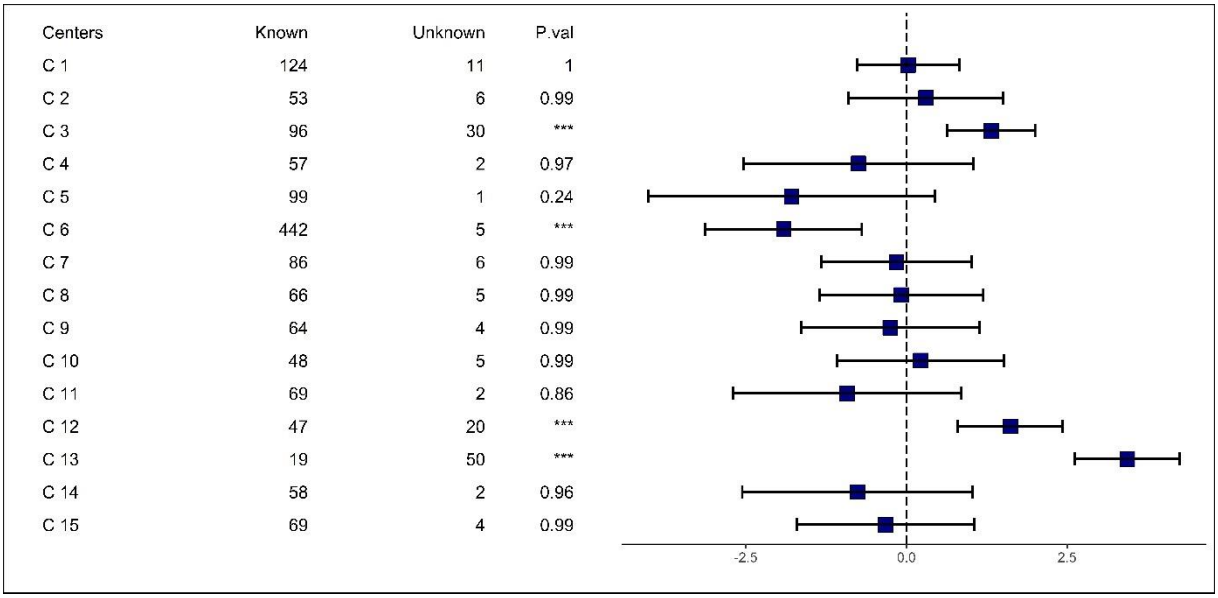


Figure 8: Simultaneous confidence intervals for contrasts of center means with GM for a binary variable missing age at diagnosis or age at onset reported using Bayesian Generalized Linear model (BayesGLM).

Ordinal Endpoint: Disease severity

Another endpoint commonly observed would be an ordinal variable. Figure 9 presents the disease severity level of patients documented at the latest visit in each center. Here the comparison would be whether a certain center would have a different mean compared to the GM of all centers. In other words, it would indicate whether a specific center has a cohort that has a higher or a lower disease severity compared to the GM. (since the disease severity is in increasing order). Unlike clinical trials, the GMSR does not have a strict inclusion criterion for patient recruitment regarding the level of disability. Centers would include any patients diagnosed with multiple sclerosis. Hence the difference between center cohorts for disease severity is not unforeseen in this specific case, specifically if a center is specialized such as a rehabilitation center.

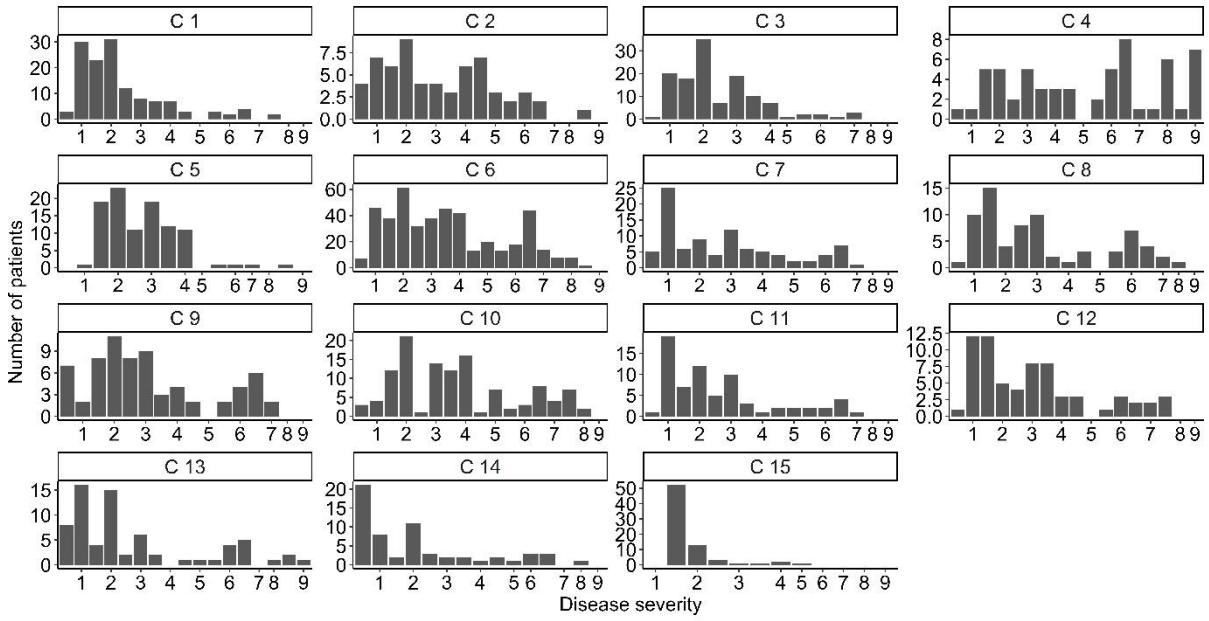


Figure 9: Histograms of disease severity (EDSS) for patients at the latest visit.

Figure 10 shows the comparisons of GM to the mean of individual centers for disease severity. C 4, C 6, and C 10 show a mean larger than the GM, in other words, it indicates that the cohorts belonging to these centers have a higher disease severity than the GM of the whole cohort. As for C 1, C 3, C 14, and C 15 they have a smaller mean than the GM. These differences could be alarming and could be a natural deviation due to the cohort

itself. Stakeholders at the registry would indicate whether this difference is expected or not. However, in a clinical trial where cohorts are likely to be comparable, it would indicate whether a center tends to give smaller or higher measurements, then in this case an onsite visit should be conducted to have a justification for the deviation.

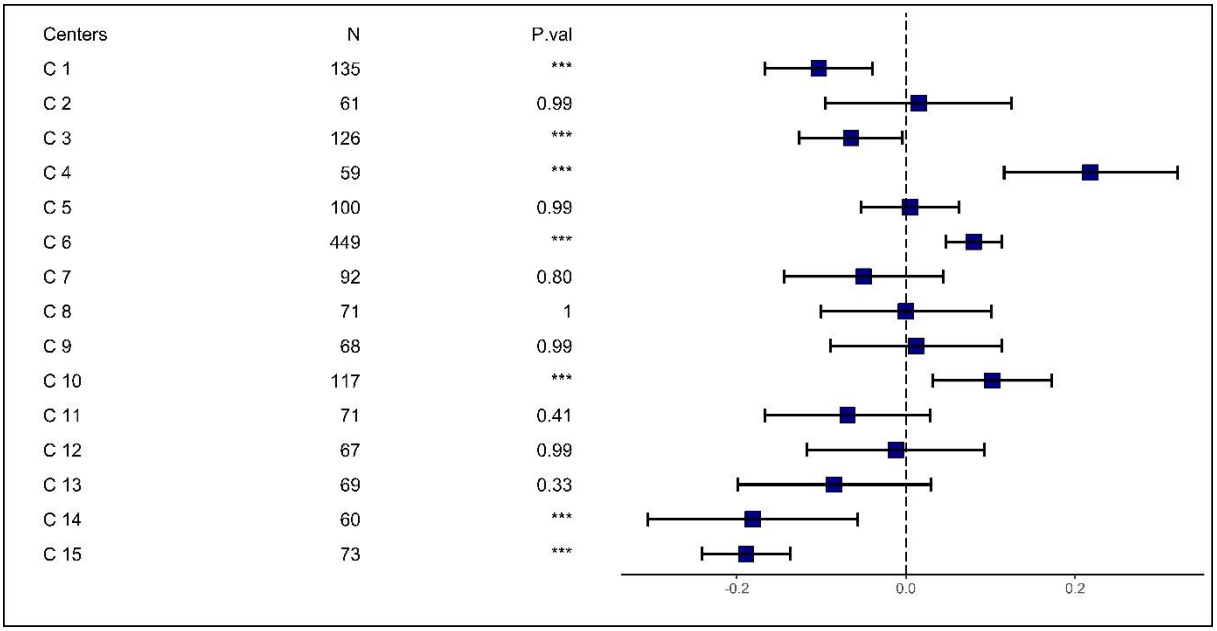


Figure 10: Ordinal variable of disease severity (EDSS) comparison for each center to the GM using the non-parametric method (Nparcomp) .

1.6 Contributions to the field

In the framework of this thesis, the differences between current RBM tools were recognized through their implementation in clinical trial protocols. This revealed their heterogeneity and thus their output when implemented for the same clinical trial. Furthermore, to tackle the wide differences clinical trials have, a rationale for a robust risk assessment methodology was established. The approach presented has an adaptive ability to include any risk in the assessment regardless of clinical trial phase or complexity. It gives the assessor the chance to include or exclude risks related or not related to the trial. To ease the implementation of RMA, a user-friendly shiny web application is formed. The user can run the web app locally, run the analysis, and document the risk assessment performed to further develop a monitoring plan. Each assessed risk would be summarized and assigned a mitigation technique. This would also help the assessor in reviewing the risk assessment previously done when needed. The

syntax of RMA is available on GitHub under the following path: https://github.com/firasfneish/Risk-Methodology-Assessment

For remote monitoring techniques, this thesis proposed the utilization of Grand Mean comparisons to individual centers. This comparison would direct to a deviating center that is unlikely due to chance. The comparison to GM was implemented for binary, ordinal, and continuous endpoints using different statistical methods. The deviating settings were identified when performing the comparisons to GM. Nparcomp fails to control type I error for $n_i < 20$ and for extremely unbalanced designs. BayesGLM outperforms GLM and BrGLM for small sample sizes and can deal with the 0 problem. Since the available models for GM comparisons are scattered between different R packages, a unified interface for their implementation is shown and the same graphical output displaying simultaneous confidence intervals of GM comparisons for different variable types is presented. Their implementation is also verified on RWD from the German Multiple Sclerosis Registry. Two shiny apps for BayesGLM and Nparcomp methods are provided as an interactive and visual platform for users to easily implement the comparisons by providing their datasets. Both shiny apps run the comparisons of each center to the GM and plot simultaneous confidence intervals for contrasts of center means with the GM of the given dataset. Syntax of both apps can be found on GitHub under the following path: https://github.com/firasfneish/CSM.

1.7 Conclusions and future research

1.7.1 Risk Methodology Assessment

The established RMA currently provides stakeholders the ability to assess any risk in a clinical trial that can be covered by RBM. It also gives stakeholders the chance to incorporate additional risks into the assessment process. Future work should include real problems found in clinical trials in the assessment list. It would allow stakeholders to further include risks they did not previously consider. A database that combines historic faults and errors in clinical trials is currently missing. Thus, a database that includes risks found in clinical trials would be a major step that would benefit clinical research. This would give RMA the ability to forecast a certain risk. Such predictions require a database consisting of numerous factors. This requirement is currently lacking and hence would be the focus for future work. The COVID-19 pandemic has pushed pharmaceutical companies further to implement RBM and adopt remote monitoring strategies (Barnes et al., 2021).

Therefore, the risk assessment within the RBM system should be under continuous development to tackle the changes and shifts in the possible monitoring activities.

1.7.2 Centralized Statistical Monitoring

The multiple comparisons of center mean to GM means of all centers were investigated and validated for the identification of a deviating center. This research focused on performing the comparisons for specific data endpoints (Binomial, ordinal, Continuous). This approach can be further implemented for more data types such as Time to Event(time2event) endpoint, Poisson/count data, and Nominal data. Monte Carlo simulations are needed to identify appropriate estimators and detect limitations in each type. Future work would focus on considering Weibull models and Cox regression for time2event data. GLM and BayesGLM and negative binomial model for count data. Desmet et al. (2017) investigated the detection of a deviating center by employing a beta-binomial (Desmet et al., 2017a), it would be noteworthy to compare beta-binomial, GLM, and BayesGLM for GM comparisons. Ordered categorical regression and cumulative link models should also be investigated for ordinal data since the Nparcomp method failed to control type I error for small sample sizes and heavily unbalanced designs. Some centers could naturally differ from other centers due to their patient cohort selection. This is more relevant to registry data rather than clinical trials as registries could include data coming from centers with different specializations whereas clinical trials include centers with very specific inclusion criteria. This natural difference could be accounted for by considering the effect of additional covariates (e.g. center type) when performing the comparisons by each estimator. In the current approach, many tests are performed, thus false positive conclusions might be concluded. To counteract the multiple testing problem, α correction methods such as Bonferroni adjustment, and False Discovery Rate (FDR) should be examined, however, clinicians and statisticians should decide whether this correction is needed. Equivalence intervals are a further option that could serve CSM. Nevertheless, this would require clinicians to determine an interval in which a certain center deviation could still be accepted.

1.8 References

Afroz, M. A., Schwarber, G., & Bhuiyan, M. A. N. (2021). Risk-based centralized data monitoring of clinical trials at the time of COVID-19 pandemic. *Contemporary Clinical Trials*, *104*, 106368. https://doi.org/10.1016/j.cct.2021.106368

Agresti, A. (2013). *Categorical Data Analysis, 3rd Edition | Wiley*. Wiley.Com. https://www.wiley.com/en-us/Categorical+Data+Analysis%2C+3rd+Edition-p-9780470463635

American Cancer Society. (2020, August). *Clinical Trials: What You Need to Know*. https://www.cancer.org/content/dam/CRC/PDF/Public/6800.00.pdf

Arm.pdf. (n.d.). Retrieved May 11, 2022, from https://cran.r-project.org/web/packages/arm/arm.pdf

Baer, A. R., Zon, R., Devine, S., & Lyss, A. P. (2011). The Clinical Research Team. *Journal of Oncology Practice*, *7*(3), 188–192. https://doi.org/10.1200/JOP.2011.000276

Baigent, C., Harrell, F. E., Buyse, M., Emberson, J. R., & Altman, D. G. (2008a). Ensuring trial validity by data quality assurance and diversification of monitoring methods. *Clinical Trials (London, England)*, *5*(1), 49–55. https://doi.org/10.1177/1740774507087554

Baigent, C., Harrell, F. E., Buyse, M., Emberson, J. R., & Altman, D. G. (2008b). Ensuring trial validity by data quality assurance and diversification of monitoring methods. *Clinical Trials (London, England)*, *5*(1), 49–55. https://doi.org/10.1177/1740774507087554

Bakobaki, J. M., Rauchenberger, M., Joffe, N., McCormack, S., Stenning, S., & Meredith, S. (2012). The potential for central monitoring techniques to replace on-site monitoring: Findings from an international multi-centre clinical trial. *Clinical Trials (London, England)*, *9*(2), 257–264. https://doi.org/10.1177/1740774511427325

Barchard, K. A., & Pace, L. A. (2011). Preventing human error: The impact of data entry methods on data accuracy and statistical results. *Computers in Human Behavior*, *27*(5), 1834–1839. https://doi.org/10.1016/j.chb.2011.04.004

Barnes, B., Stansbury, N., Brown, D., Garson, L., Gerard, G., Piccoli, N., Jendrasek, D., May, N., Castillo, V., Adelfio, A., Ramirez, N., McSweeney, A., Berlien, R., & Butler, P. J. (2021). Risk-Based Monitoring in Clinical Trials: Past, Present, and Future. *Therapeutic Innovation & Regulatory Science*, *55*(4), 899–906. https://doi.org/10.1007/s43441-021-00295-8

Bhatt, A. (2010). Evolution of Clinical Research: A History Before and Beyond James Lind. *Perspectives in Clinical Research*, *1*(1), 6–10.

brglm function—RDocumentation. (n.d.). Retrieved March 31, 2022, from https://www.rdocumentation.org/packages/brglm/versions/0.7.2/topics/brglm

Brosteanu, O., Houben, P., Ihrig, K., Ohmann, C., Paulus, U., Pfistner, B., Schwarz, G., Strenge-Hesse, A., & Zettelmeyer, U. (2009). Risk analysis and risk adapted on-site monitoring in noncommercial clinical trials. *Clinical Trials*, *6*(6), 585–596. https://doi.org/10.1177/1740774509347398

Buyse, M., George, S. L., Evans, S., Geller, N. L., Ranstam, J., Scherrer, B., Lesaffre, E., Murray, G., Edler, L., Hutton, J., Colton, T., Lachenbruch, P., Verma, B. L., & ISCB SUBCOMMITTEE ON FRAUD. (1999). The role of biostatistics in the prevention, detection and treatment of fraud in clinical trials. *Statistics in Medicine*, *18*(24), 3435–3451. https://doi.org/10.1002/(SICI)1097-0258(19991230)18:24<3435::AID-SIM365>3.0.CO;2-O

Buyse, M., Trotta, L., Saad, E. D., & Sakamoto, J. (2020). Central statistical monitoring of investigator-led clinical trials in oncology. *International Journal of Clinical Oncology*, *25*(7), 1207–1214. https://doi.org/10.1007/s10147-020-01726-6

Canter, C., & Lewis, K. (2014). Training Clinical Trial Teams Current Training Practices and Educational Needs of Clinical Research Personnel. *Monitor*, 25–30.

Chan, A.-W., M. Tetzlaff, J., G. Altman, D., Laupacis, A., C. Gøtzsche, P., Krleža-Jerić, K., Hróbjartsson, A., Mann, H., Dickersin, K., A. Berlin, J., J. Doré, C., R. Parulekar, W., S.M. Summerskill, W., Groves, T., F. Schulz, K., C. Sox, H., W. Rockhold, F., Rennie, D., & Moher, D. (2013). SPIRIT

2013 Statement: Defining Standard Protocol Items for Clinical Trials. *Annals of Internal Medicine*. https://www.acpjournals.org/doi/10.7326/0003-4819-158-3-201302050-00583

Chung, K. C., & Song, J. W. (2010). A Guide on Organizing a Multicenter Clinical Trial: The WRIST study group. *Plastic and Reconstructive Surgery*, *126*(2), 515–523. https://doi.org/10.1097/PRS.0b013e3181df64fa

Collier, R. (2009). Legumes, lemons and streptomycin: A short history of the clinical trial. *CMAJ : Canadian Medical Association Journal*, *180*(1), 23–24. https://doi.org/10.1503/cmaj.081879

Desmet, L., Venet, D., Doffagne, E., Timmermans, C., Burzykowski, T., Legrand, C., & Buyse, M. (2014a). Linear mixed-effects models for central statistical monitoring of multicenter clinical trials. *Statistics in Medicine*, *33*(30), 5265–5279. https://doi.org/10.1002/sim.6294

Desmet, L., Venet, D., Doffagne, E., Timmermans, C., Burzykowski, T., Legrand, C., & Buyse, M. (2014b). Linear mixed-effects models for central statistical monitoring of multicenter clinical trials. *Statistics in Medicine*, *33*(30), 5265–5279. https://doi.org/10.1002/sim.6294

Desmet, L., Venet, D., Doffagne, E., Timmermans, C., Legrand, C., Burzykowski, T., & Buyse, M. (2017a). Use of the Beta-Binomial Model for Central Statistical Monitoring of Multicenter Clinical Trials. *Statistics in Biopharmaceutical Research*, *9*(1), 1–11. https://doi.org/10.1080/19466315.2016.1164751

Desmet, L., Venet, D., Doffagne, E., Timmermans, C., Legrand, C., Burzykowski, T., & Buyse, M. (2017b). Use of the Beta-Binomial Model for Central Statistical Monitoring of Multicenter Clinical Trials. *Statistics in Biopharmaceutical Research*, *9*(1), 1–11. https://doi.org/10.1080/19466315.2016.1164751

Dupin-Spriet, A. S., T. (2005). *Good Practice of Clinical Drug Trials | Karger Book*. https://www.karger.com/Book/Home/230349

Ellen van Bavel. (2016). *Multidisciplinary Approach to Protocol Writing*. Applied Clinical Trials Online. https://www.appliedclinicaltrialsonline.com/view/multidisciplinary-approach-protocol-writing

EMA-IWG. (2018). *Annual Report of the Good Clinical Practice Inspectors Working Group*. 20.

European Medicines Agency. (2018, September 17). *Clinical trials in human medicines* [Text]. European Medicines Agency. https://www.ema.europa.eu/en/human-regulatory/research-development/clinical-trials-human-medicines

FDA. (2013). *Guidance for Industry, Oversight of Clinical Investigations—A Risk-Based Approach to Monitoring*. 22.

FDA. (2020, June 11). *Oversight of Clinical Investigations—A Risk-Based Approach to Monitoring*. U.S. Food and Drug Administration; FDA. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/oversight-clinical-investigations-risk-based-approach-monitoring

Fneish, F., Limaye, D., Struever, V., Schaarschmidt, F., & Fortwengel, G. (2020). Comparison of Non-Commercial Risk Based Monitoring Tools by Their Application on Clinical Trial Protocols. *American Journal of Biomedical Science & Research*, *8*(3), 216. https://doi.org/10.34297/AJBSR.2020.08.001276

Fneish, F., & Schaarschmidt, F. (2019a). *Multiple Treatment Comparisons Of Binomial Data With Only 0s In One Treatment Group*. https://doi.org/10.13140/RG.2.2.24658.20162

Fneish, F., & Schaarschmidt, F. (2019b). *Multiple Treatment Comparisons Of Binomial Data With Only 0s In One Treatment Group*. https://doi.org/10.13140/RG.2.2.24658.20162

Fneish, F., Schaarschmidt, F., & Fortwengel, G. (2021a). Improving Risk Assessment in Clinical Trials: Toward a Systematic Risk-Based Monitoring Approach. *Current Therapeutic Research, Clinical and Experimental*, *95*, 100643. https://doi.org/10.1016/j.curtheres.2021.100643

Fneish, F., Schaarschmidt, F., & Fortwengel, G. (2021b). Improving Risk Assessment in Clinical Trials: Toward a Systematic Risk-Based Monitoring Approach. *Current Therapeutic Research*, *95*, 100643. https://doi.org/10.1016/j.curtheres.2021.100643

Food and Drug Administration. (2018, November 3). *What Are the Different Types of Clinical Research?* FDA; FDA. https://www.fda.gov/patients/clinical-trials-what-patients-need-know/what-are-different-types-clinical-research

Food and Drug Administration. (2019, May 28). *Clinical Research Versus Medical Treatment*. FDA; FDA. https://www.fda.gov/patients/clinical-trials-what-patients-need-know/clinical-research-versus-medical-treatment

Fortwengel, G. (2004). *Guide for Clinical Trial Staff | Karger Book*. https://www.karger.com/Book/Home/229514

Fortwengel, G. (2011). *Guide for Investigator Initiated Trials*. Karger Medical and Scientific Publishers.

Gelman, A., Jakulin, A., Pittau, M. G., & Su, Y.-S. (2008a). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, *2*(4). https://doi.org/10.1214/08-AOAS191

Gelman, A., Jakulin, A., Pittau, M. G., & Su, Y.-S. (2008b). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, *2*(4). https://doi.org/10.1214/08-AOAS191

Gelman, A., Su, Y.-S., Yajima, M., Hill, J., Pittau, M. G., Kerman, J., Zheng, T., & Dorie, V. (2021). *arm: Data Analysis Using Regression and Multilevel/Hierarchical Models* (1.12-2) [Computer software]. https://CRAN.R-project.org/package=arm

Genz, A., & Bretz, F. (2009). Special Cases. In A. Genz & F. Bretz (Eds.), *Computation of Multivariate Normal and t Probabilities* (pp. 7–20). Springer. https://doi.org/10.1007/978-3-642-01689-9_2

Gilmartin-Thomas, J. F., Liew, D., & Hopper, I. (2018). Observational studies and their utility for practice. *Australian Prescriber*, *41*(3), 82–85. https://doi.org/10.18773/austprescr.2018.017

Gliklich, R. (2009). *Clinical Trials vs Registries*. *0*(0). https://www.appliedclinicaltrialsonline.com/view/clinical-trials-vs-registries

Goldberg, S. I., Niemierko, A., & Turchin, A. (2008). Analysis of Data Errors in Clinical Research Databases. *AMIA Annual Symposium Proceedings*, *2008*, 242–246.

Hothorn, T., Bretz, F., & Westfall, P. (2008a). Simultaneous inference in general parametric models. *Biometrical Journal. Biometrische Zeitschrift*, *50*(3), 346–363. https://doi.org/10.1002/bimj.200810425

Hothorn, T., Bretz, F., & Westfall, P. (2008b). Simultaneous inference in general parametric models. *Biometrical Journal. Biometrische Zeitschrift*, *50*(3), 346–363. https://doi.org/10.1002/bimj.200810425

Hurley, C., Shiely, F., Power, J., Clarke, M., Eustace, J. A., Flanagan, E., & Kearney, P. M. (2016). Risk based monitoring (RBM) tools for clinical trials: A systematic review. *Contemporary Clinical Trials*, *51*, 15–27. https://doi.org/10.1016/j.cct.2016.09.003

ICH. (2016). *GUIDELINE FOR GOOD CLINICAL PRACTICE*. 66.

Imperial College Clinical Research Governance Office. (2007). *Good Clinical Practice*. Imperial College London. https://www.imperial.ac.uk/research-and-innovation/support-for-staff/research-office/research-governance-and-integrity/what-is-research-governance/regulatory-frameworks/good-clinical-practice/

International Council for Harmonisation. (2018). *E6(R2) Good Clinical Practice: Integrated Addendum to ICH E6(R1)*. U.S. Food and Drug Administration; FDA. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/e6r2-good-clinical-practice-integrated-addendum-ich-e6r1

Interventional Clinical Trial or Registry Study? Writing Protocols that Demonstrate the Difference. (2018, July 24). WCG IRB. https://www.wcgirb.com/insights/interventional-clinical-trial-or-registry-study-writing-protocols-that-demonstrate-the-difference/

Jones, G., & Abbasi, K. (2004). Trial protocols at the BMJ. *BMJ (Clinical Research Ed.)*, *329*(7479), 1360. https://doi.org/10.1136/bmj.329.7479.1360

Jongen, P. M. J. M., van den Bogert, C. A., van de Laar, C. W. E., Notenboom, K., Hille, E. T. M., & Hegger, I. (2016). Risk indicator taxonomy for supervision of clinical trials on medicinal products. *Current Medical Research and Opinion*, *32*(7), 1269–1276. https://doi.org/10.1185/03007995.2016.1170671

Journot, V., Pignon, J.-P., Gaultier, C., Daurat, V., Bouxin-Métro, A., Giraudeau, B., Preux, P.-M., Tréluyer, J.-M., Chevret, S., Plättner, V., Thalamas, C., Clisant, S., Ravaud, P., & Chêne, G.

(2011). Validation of a risk-assessment scale and a risk-adapted monitoring plan for academic clinical research studies—The Pre-Optimon study. *Contemporary Clinical Trials*, *32*(1), 16–24. https://doi.org/10.1016/j.cct.2010.10.001

Kahan, B. C. (2014). Accounting for centre-effects in multicentre trials with a binary outcome – when, why, and how? *BMC Medical Research Methodology*, *14*(1), 20. https://doi.org/10.1186/1471-2288-14-20

Kirkwood, A. A., Cox, T., & Hackshaw, A. (2013). Application of methods for central statistical monitoring in clinical trials. *Clinical Trials (London, England)*, *10*(5), 783–806. https://doi.org/10.1177/1740774513494504

Konietschke, F., Hothorn, L. A., & Brunner, E. (2012a). Rank-based multiple test procedures and simultaneous confidence intervals. *Electronic Journal of Statistics*, *6*(none), 738–759. https://doi.org/10.1214/12-EJS691

Konietschke, F., Hothorn, L. A., & Brunner, E. (2012b). Rank-based multiple test procedures and simultaneous confidence intervals. *Electronic Journal of Statistics*, *6*(none), 738–759. https://doi.org/10.1214/12-EJS691

Konietschke, F., Placzek, M., Schaarschmidt, F., & Hothorn, L. A. (2015a). nparcomp: An R Software Package for Nonparametric Multiple Comparisons and Simultaneous Confidence Intervals. *Journal of Statistical Software*, *64*, 1–17. https://doi.org/10.18637/jss.v064.i09

Konietschke, F., Placzek, M., Schaarschmidt, F., & Hothorn, L. A. (2015b). nparcomp: An R Software Package for Nonparametric Multiple Comparisons and Simultaneous Confidence Intervals. *Journal of Statistical Software*, *64*, 1–17. https://doi.org/10.18637/jss.v064.i09

Kosmidis, I., & Firth, D. (2009). Bias reduction in exponential family nonlinear models. *Biometrika*, *96*(4), 793–804. https://doi.org/10.1093/biomet/asp055

Kosmidis, I., & Firth, D. (2021a). Jeffreys-prior penalty, finiteness and shrinkage in binomial-response generalized linear models. *Biometrika*, *108*(1), 71–82. https://doi.org/10.1093/biomet/asaa052

Kosmidis, I., & Firth, D. (2021b). Jeffreys-prior penalty, finiteness and shrinkage in binomial-response generalized linear models. *Biometrika*, *108*(1), 71–82. https://doi.org/10.1093/biomet/asaa052

Landray, M. J., Bax, J. J., Alliot, L., Buyse, M., Cohen, A., Collins, R., Hindricks, G., James, S. K., Lane, S., Maggioni, A. P., Meeker-O'Connell, A., Olsson, G., Pocock, S. J., Rawlins, M., Sellors, J., Shinagawa, K., Sipido, K. R., Smeeth, L., Stephens, R., … Casadei, B. (2017). Improving public health by improving clinical trial guidelines and their application. *European Heart Journal*, *38*(21), 1632–1637. https://doi.org/10.1093/eurheartj/ehx086

M. Friedman, L., D. Furberg, C., L. DeMets, D., M. Reboussin, D., & B. Granger, C. (2015). *Fundamentals of Clinical Trials* (fifth). https://www.audible.de/pd/Fundamentals-of-Clinical-Trials-Hoerbuch/B09TD5X1LS

McCullagh, P., & Nelder, J. A. (1983). *Generalized Linear Models* (2nd ed.). Chapman & Hall. https://doi.org/10.1201/9780203753736

MRC/DH/MHRA. (2011). *Risk-adapted Approaches to the management of Clinical Trials of Investigational Medicinal Products*. 31.

Multcomp.pdf. (n.d.). Retrieved May 10, 2022, from https://cran.r-project.org/web/packages/multcomp/multcomp.pdf

National Institute for Health Research (NIHR) Clinical Trials Toolkit. (2012). *Trial Management & Monitoring*. https://www.ct-toolkit.ac.uk/routemap/trial-management-and-monitoring/

National Institute of Health. (2015, May 14). *The Basics*. National Institute of Health (NIH). https://www.nih.gov/health-information/nih-clinical-research-trials-you/basics

National Institute of Health. (2019, March). *Learn About Clinical Studies—ClinicalTrials.gov*. https://clinicaltrials.gov/ct2/about-studies/learn

National Institute on Aging. (2020). *What Are Clinical Trials and Studies?* National Institute on Aging. https://www.nia.nih.gov/health/what-are-clinical-trials-and-studies

NORM. (2015). *Guideline for a coordinated GCP-monitoring of clinical trials in the Nordic countries*.

Oba, K. (2016a). Statistical challenges for central monitoring in clinical trials: A review. *International Journal of Clinical Oncology*, *21*(1), 28–37. https://doi.org/10.1007/s10147-015-0914-4

Oba, K. (2016b). Statistical challenges for central monitoring in clinical trials: A review. *International Journal of Clinical Oncology*, *21*(1), 28–37. https://doi.org/10.1007/s10147-015-0914-4

Ohle, L.-M., Ellenberger, D., Flachenecker, P., Friede, T., Haas, J., Hellwig, K., Parciak, T., Warnke, C., Paul, F., Zettl, U. K., & Stahmann, A. (2021). Chances and challenges of a long-term data repository in multiple sclerosis: 20th birthday of the German MS registry. *Scientific Reports*, *11*(1), 13340. https://doi.org/10.1038/s41598-021-92722-x

Olsen, R., Bihlet, A. R., Kalakou, F., & Andersen, J. R. (2016). The impact of clinical trial monitoring approaches on data integrity and cost—A review of current literature. *European Journal of Clinical Pharmacology*, *72*(4), 399–412. https://doi.org/10.1007/s00228-015-2004-y

Pallmann, P., & Hothorn, L. A. (2016). Analysis of means: A generalized approach using R. *Journal of Applied Statistics*, *43*(8), 1541–1560. https://doi.org/10.1080/02664763.2015.1117584

Poolman, R. W., Hanson, B., Marti, R. K., & Bhandari, M. (2007). Conducting a clinical study: A guide for good research practice. *Indian Journal of Orthopaedics*, *41*(1), 27–31. https://doi.org/10.4103/0019-5413.30522

R: The R Stats Package. (n.d.). Retrieved March 31, 2022, from https://stat.ethz.ch/R-manual/R-devel/library/stats/html/00Index.html

Reflection paper risk based quality management clinical trials. (n.d.). Retrieved March 31, 2022, from https://www.ema.europa.eu/en/documents/scientific-guideline/reflection-paper-risk-based-quality-management-clinical-trials_en.pdf

Rivera, S. C., Liu, X., Chan, A.-W., Denniston, A. K., Calvert, M. J., Ashrafian, H., Beam, A. L., Collins, G. S., Darzi, A., Deeks, J. J., ElZarrad, M. K., Espinoza, C., Esteva, A., Faes, L., Ruffano, L. F. di, Fletcher, J., Golub, R., Harvey, H., Haug, C., … Yau, C. (2020). Guidelines for clinical trial protocols for interventions involving artificial intelligence: The SPIRIT-AI extension. *The Lancet Digital Health*, *2*(10), e549–e560. https://doi.org/10.1016/S2589-7500(20)30219-3

Shamley, D., & Wright, B. (2017). *A Comprehensive and Practical Guide to Clinical Trials—1st Edition* (1st ed.). https://www.elsevier.com/books/a-comprehensive-and-practical-guide-to-clinical-trials/shamley/978-0-12-804729-3

Silverstein, A. B. (1974). Comparing All Treatment Means with the Grand Mean. *Psychological Reports*, *34*(1), 221–222. https://doi.org/10.2466/pr0.1974.34.1.221

Smith, C. T., Stocken, D. D., Dunn, J., Cox, T., Ghaneh, P., Cunningham, D., & Neoptolemos, J. P. (2012). The Value of Source Data Verification in a Cancer Clinical Trial. *PLOS ONE*, *7*(12), e51623. https://doi.org/10.1371/journal.pone.0051623

Song, J. W., & Chung, K. C. (2010). Observational Studies: Cohort and Case-Control Studies: *Plastic and Reconstructive Surgery*, *126*(6), 2234–2242. https://doi.org/10.1097/PRS.0b013e3181f44abc

Swiss Clinical Trial Organisation. (2014). *Guidelines for Risk-Adapted Monitoring*.

Thiese, M. S. (2014). Observational and interventional study design types; an overview. *Biochemia Medica*, *24*(2), 199–210. https://doi.org/10.11613/BM.2014.022

Timmermans, C., Venet, D., & Burzykowski, T. (2016). Data-driven risk identification in phase III clinical trials using central statistical monitoring. *International Journal of Clinical Oncology*, *21*(1), 38–45. https://doi.org/10.1007/s10147-015-0877-5

TransCelerate. (2012). TransCelerate. https://www.transceleratebiopharmainc.com/rbminteractiveguide/what-is-risk-based-monitoring-rbm/introduction/

Trotta, L., Kabeya, Y., Buyse, M., Doffagne, E., Venet, D., Desmet, L., Burzykowski, T., Tsuburaya, A., Yoshida, K., Miyashita, Y., Morita, S., Sakamoto, J., Praveen, P., & Oba, K. (2019). Detection of atypical data in multicenter clinical trials using unsupervised statistical monitoring. *Clinical Trials (London, England)*, *16*(5), 512–522. https://doi.org/10.1177/1740774519862564

UK Clinical Research Collaboration. (2014). *Understanding Clinical Trials*. https://www.ukcrc.org/wp-content/uploads/2014/03/iCT_Booklet.pdf

Umscheid, C. A., Margolis, D. J., & Grossman, C. E. (2011). Key Concepts of Clinical Trials: A Narrative Review. *Postgraduate Medicine*, *123*(5), 194–204. https://doi.org/10.3810/pgm.2011.09.2475

Uren, S. C., Kirkman, M. B., Dalton, B. S., & Zalcberg, J. R. (2013). Reducing clinical trial monitoring resource allocation and costs through remote access to electronic medical records. *Journal of Oncology Practice*, *9*(1), e13-16. https://doi.org/10.1200/JOP.2012.000666

Venet, D., Doffagne, E., Burzykowski, T., Beckers, F., Tellier, Y., Genevois-Marlin, E., Becker, U., Bee, V., Wilson, V., Legrand, C., & Buyse, M. (2012a). A statistical approach to central monitoring of data quality in clinical trials. *Clinical Trials (London, England)*, *9*(6), 705–713. https://doi.org/10.1177/1740774512447898

Venet, D., Doffagne, E., Burzykowski, T., Beckers, F., Tellier, Y., Genevois-Marlin, E., Becker, U., Bee, V., Wilson, V., Legrand, C., & Buyse, M. (2012b). A statistical approach to central monitoring of data quality in clinical trials. *Clinical Trials*, *9*(6), 705–713. https://doi.org/10.1177/1740774512447898

World Health Organization. (2005). *Handbook for good clinical research practice (GCP): Guidance for implementation.* World Health Organization.

Yang, W., Zilov, A., Soewondo, P., Bech, O. M., Sekkal, F., & Home, P. D. (2010). Observational studies: Going beyond the boundaries of randomized controlled trials. *Diabetes Research and Clinical Practice*, *88*, S3–S9. https://doi.org/10.1016/S0168-8227(10)70002-4

Yee, D. (2017). *DATA AND SAFETY MONITORING PLAN*. 50.

Chapter 2

Publications and Manuscripts

2.1 Comparison of Non-Commercial Risk Based Monitoring Tools by Their Application on Clinical Trial Protocols.

Firas Fneish[1], Dnyanesh Limaye[2], Vanessa Strüver[2], Frank Schaarschmidt[1], Gerhard Fortwengel[2]

1. Leibniz Universität Hannover, Herrenhäuser Strasse 2, 30419, Germany
2. Hochschule Hannover, Expo plaza 12, 30539, Germany

Type of authorship: First author

Journal: American Journal of Biomedical Science & Research

Impact factor: 1.042

Number of citations: 2 (Google Scholar)

DOI: 10.34297/AJBSR.2020.08.001276

Contributions:

1. Conceptualized and designed the study **(Firas Fneish, Dnyanesh Limaye, Vanessa Strüver, Frank Schaarschmidt, Gerhard Fortwengel)**
2. Protocol assessment **(Firas Fneish)**
3. Data analysis and interpretation **(Firas Fneish)**
4. Writing the manuscript **(Firas Fneish)**
5. Reviewed the manuscript for intellectual content **(Dnyanesh Limaye, Vanessa Strüver, Frank Schaarschmidt, Gerhard Fortwengel)**

American Journal of
Biomedical Science & Research

www.biomedgrid.com
ISSN: 2642-1747

Research Article

Comparison of Non-Commercial Risk Based Monitoring Tools by Their Application on Clinical Trial Protocols

Firas Fneish[1]*, Dnyanesh Limaye[2], Vanessa Struever[2], Frank Schaarschmidt[1] and Gerhard Fortwengel[2]

[1]*Leibniz Universität Hannover, Herrenhäuser Straße 2, 30419 Hannover, Germany*

[2]*Hochschule Hannover, Expo Plaza 12, 30539 Hannover , Germany*

***Corresponding author:** Firas Fneish, Herrenhäuser Straße 2, Hannover 30419, Lower Saxony, Germany

To Cite This Article: *Firas F, Dnyanesh L, Vanessa S, Frank S, Gerhard F. Comparison of Non-Commercial Risk Based Monitoring Tools by Their Application on Clinical Trial Protocols. 2020 - 8(3). AJBSR.MS.ID.001276.* DOI: 10.34297/AJBSR.2020.08.001276.

Received: March 31, 2020; **Published:** April 07, 2020

Abstract

Clinical trial monitoring involves intensive on-site monitoring visits at clinical trial sites and exhaustive source data verification (monitoring) of clinical trial data [1]. Clinical researchers have questioned the validity and necessity for traditional monitoring methods [2], which have been under investigation due to their ineffectiveness in improving the quality of clinical trial data or in protecting trial participants [3]. Implementing a risk based monitoring (RBM) system is suggested by the ICH's newly adapted guidelines to improve overall quality management [3]. The RBM involves the identification of any risk that might have an effect on areas routinely subject during monitoring activities. Risks should be identified by a RBM system followed by an evaluation of their likelihood of occurring and the extent to detect these errors and their impact on human subject protection, trial data reliability, and GCP- and protocol compliance [4]. To date various tools for risk identification have been developed with both in paper based or electronic RBM [5,6]. These tools have been compared on their characteristics and the strategy to decrease risk. However the application and subsequent effectiveness of RBM tools is yet to be examined [6]. The aim of this research is to apply each non-commercial RBM tool to clinical protocols and compare the potential risks detected in each, additionally the overall risk assessment of the protocols. Here we show that RBM tools result in different overall risk assessment when applied to the same clinical trial protocols, interestingly, each RBM tool detected distinct risks which thus resulted in a variation in the outcome mitigation.

Keywords: Clinical Trial Protocols, Risk Based Monitoring Tools, Risk Adapted Monitoring

Background

Source data verification (Monitoring) is an essential requirement for all clinical trials in phases I-IV as stated by World Health Organization (WHO) guidelines for good clinical practice (GCP) of clinical trials on pharmaceutical products, the Food and Drug Administration (FDA) code of federal regulations, and by the International Council for Harmonization (ICH) [1]. However, regulatory agencies have stressed the need for oversight approaches to identify different risk levels in each specific trial prior its commencement [1]. Moreover, it has been reported that onsite monitoring is costly with a limited outcome to clinical trial data quality onsite monitoring/SDV [2].

Clinical trial monitoring often involves intensive on-site monitoring visits at clinical trial centers and extensive SDV of clinical trial data [3]. Clinical researchers have questioned the validity and necessity of traditional monitoring methods [4]. It has been considered to be an expensive, time-consuming and resource demanding activity that does not necessarily improve the quality of clinical trial data or the protection of trial participants [5].

Over the years, clinical trials have developed complex designs, became more globalized, and used advanced technological means at various stages, which resulted in more recommendations to the guidelines for GCP. ICH has given the sponsors flexibility to utilize innovative approaches to plan, conduct and evaluate clinical trials. Nevertheless, greater emphasis has been placed on data completeness and accuracy than on critical aspects such as risk management of outcome data. For this reason, an integrated addendum to the ICH (GCP) Guideline was released in order to

request improved and more effective methods to protect the rights of clinical trial participants, and to ensure data reliability as well as GCP and trial protocol compliance. The existing ICH guideline has been modified with respect to points such as principles of GCP, investigator responsibilities, sponsor responsibilities, and the essential documents [6]. The amended ICH (GCP) guideline suggests different recommendations to the sponsor to improve overall quality management in a clinical trial. One of the recommendations is to implement a risk based approach monitoring system.

A risk based monitoring (RBM) approach involves the identification of any risk that might have an effect on areas routinely subjected to monitoring activities. These risks should be identified based on protocol mandated requirements and procedures, protocol related logistics, clinical trial phase and country of conduct. Identification should be followed by risk evaluation instead, risk likelihood, the extent to detect these errors and their impact on human subject protection, trial data reliability, GCP, and protocol compliance [7].

To date various tools for risk identification have been developed as either paper based or in electronic format [8,9]. These tools have been compared regarding their characteristics and their employed strategies in identification and classification of potential risks. Additionally it has been stated that the lack of evidence to show superiority of RBM over traditional onsite monitoring has held back their utilization [10]. Recent research using ADAMON negates the inferiority of risk adapted monitoring to extensive monitoring [10]. We aim to evaluate the effectiveness of RBM as a tool for onsite risk based monitoring, given the lack of investigation into such a method so far [9].

Methods

Search Strategy on RBM tools and Clinical Trial Protocols

For RBM tools:

Google Scholar was used in October 2018 to search the following key terms: risk based monitoring tools, risks assessment of clinical trials and risk analysis of clinical trials. The search resulted in 16 pages and after page 10 there were no suitable publications. Additional restrictions for the advance search option in Google Scholar were not used. Additionally, PubMed search engine was used with the terms: risk based monitoring tools, risks assessment of clinical trials where it resulted in 4 pages.

For Protocols:

Google Scholar was used with following terms: clinical trial protocols and summary protocols of clinical trials phases. The search resulted in 13 pages of results of which 10 were suitable. We did not use any additional restrictions for the advance search option in Google Scholar. Additionally "clinicaltrial.gov" has been used with the terms clinical trial protocols. An advanced search was used to specify available study protocols.

Assessment of Clinical Trial Protocols:

The first objective was to find out whether non-commercial RBM tools give similar overall risk assessment for the selected protocols. Noncommercial RBM tools (ADAMON [10], NORM [11], MHRA [12], Yee [13], Transcelerate [14], OPTIMON [15] and SWISS [16]) were applied to perform risk assessment of 18 clinical trial protocols from different phases with different indications. Based on the outcome the risk was categorized into high, medium, low for the respective clinical trial.

Comparison of RBM Tools Risk Covered

The second objective was to investigate whether the tools cover different risk aspects. Transcelerate RBM tool has been used as a standard by six commercial RBM tools [9]. For this reason it was used for the second investigation as a base for risk category to be compared to each RBM tool by its risk category structure: safety, study phase, complexity, technology, subject population, data collection, endpoints, staff experience, Investigational medicinal drug (IMP), logistics, blinding, operation complexity, geography in order to investigate the different risks covered between the RBM tools. Evaluation of the different risk statements was done by the following rating scale that we developed to identify whether the risk is also investigated by the other RBM tools and to which level as shown in Table 1.

Table 1: Rating scale for risks covered by Risk Based Monitoring tool.

Rating scale of risk			
	Addressed	Partially	Not addressed
Description	Risk is investigated by several questions relating to its importance	Risk is investigated as a minor risk of limited importance	Risk is not addressed by the tool at all

Statistical methods

Fisher test was used to detect differences between the risks investigated by the RBM tools if any using R statistical software version (3.6.0). The flow chart (Figure 1) was developed in R as well using packages "grid" and "Gmisc".

Ethical consideration

Neither human subjects were involved nor were personal subjects data were collected and/or processed in this research, hence no ethical permission needed for this study.

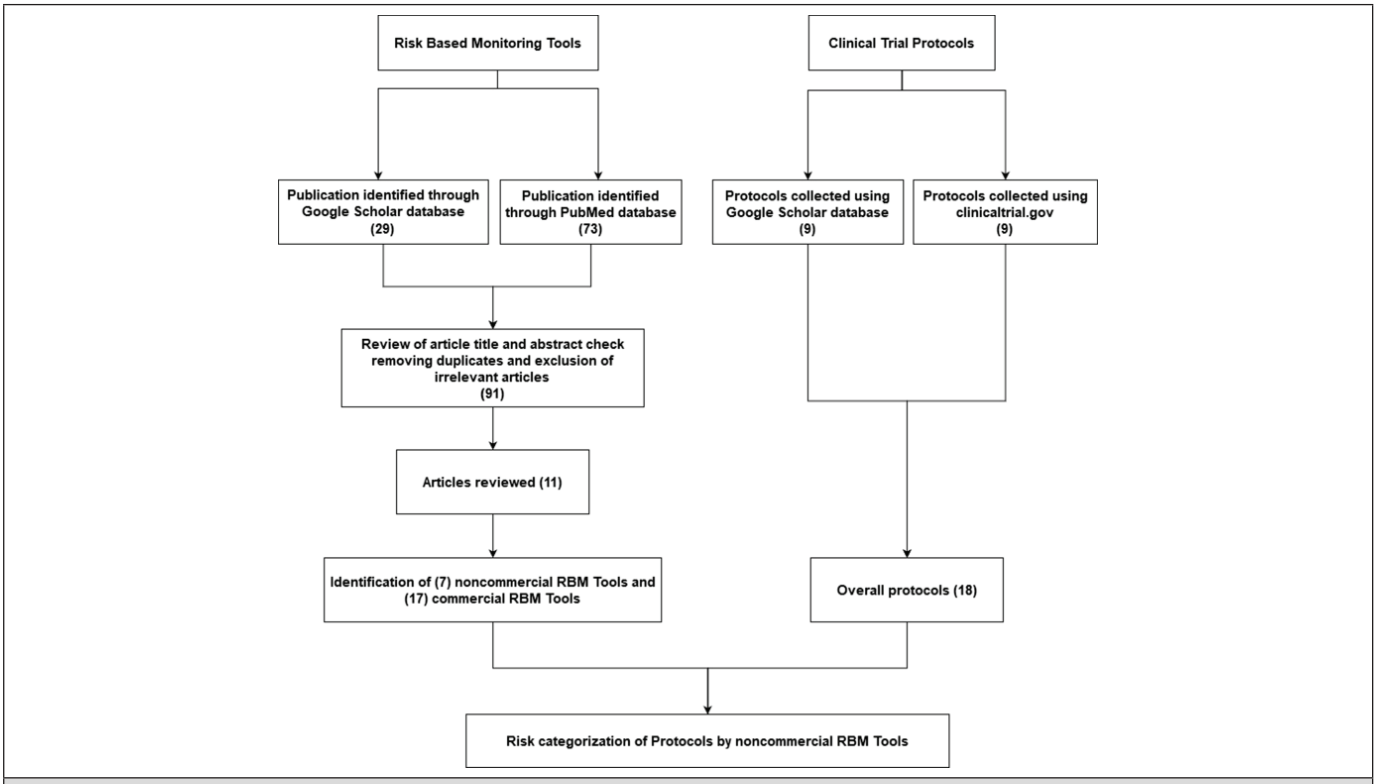


Figure 1: Selection of 18 Clinical Trial Protocols to be assessed by 7 noncommercial RBM tools.

Results & Discussion

Search Strategy

In total 24 RBM tools were identified based on a systematic review article [9], of which 7 were publicly available (Figure 1).

Assessment of Clinical Trial Protocols (Figure 2)

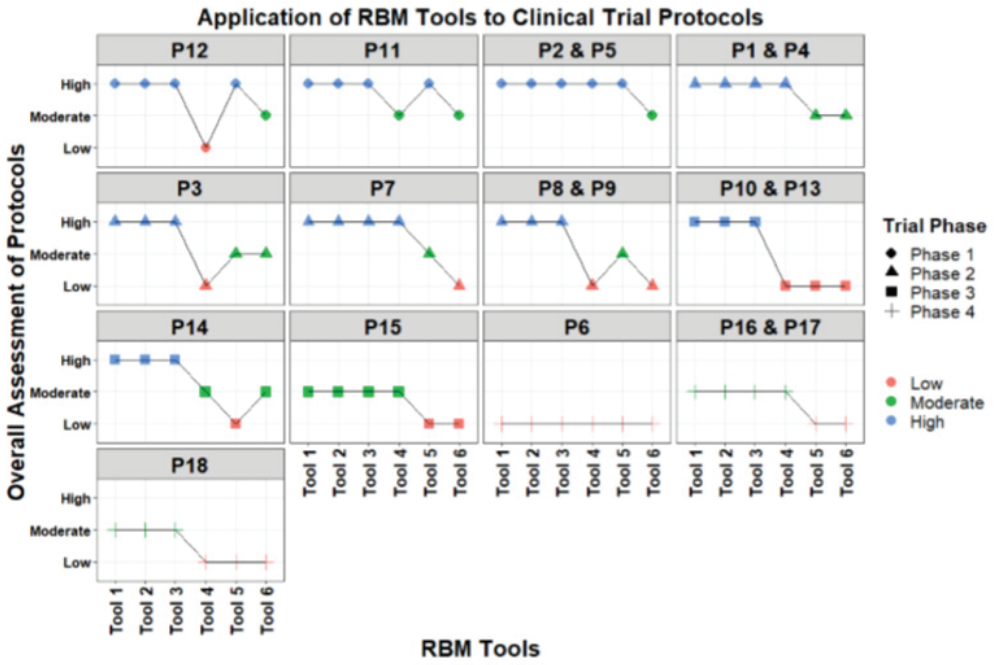


Figure 2: Overall risks assessment of Clinical Trial Protocols by each RBM tool.

For the overall risk assessment of each protocol by different RBM tools, results are reported anonymously. Out of the 18 protocols, 4 protocols belonged to phase 1(P12, P11, P2, P5), six protocols to Phase 2 (P1, P3, P4, P7, P8, P9), four protocol to Phase 3 (P10, P13, P14, P15), and four protocols to Phase 4 (P6, P16, P17, P18). Out of the 7 RBM tools, one tool did not provide an overall outcome assessment of the whole trial.

For phase 1, four (T1, T2, T3, T5) out of six tools classified the above mentioned protocols as high risk level. While tool T6 categorized these protocols into moderate risk level. Remaining tool (T4) categorized the protocol P12 as low, P11 as moderate and two protocols P2, P5 as high risk level. While for phase 2 trials, three (T1, T2, T3) out of 6 tools classified 4 protocols into high risk level, while T4 classified them as high and low risk levels. The other two tools (T4, T5) assessed 3 protocols as Medium risk while 1 protocol was assessed as Medium and Low risk levels. For phase 3 protocols, 3 tools (T1, T2, T3) categorized P10, P13, P14 as High risks and P15 as Moderate risks while T4 categorized P10, P13 as Low risks while P14 and P15 as Moderate risks, nevertheless T5 categorized all phase 3 protocols as Low risks and remaining tool T7 categorized P10, P13 and P15 as Low risks and P14 as Moderate risks. For phase 4 protocols, all tools categorized P6 as Low risks while 3 tools (T1, T2, T3) categorized P16, P17 and P18 as Moderate risks but T4 categorized P16 and P17 as Moderate risks and P18 as Low risk. Remaining Tools (T5, T6) categorized all Phase 4 Protocols as Low risks.

Risk category covered by each RBM Tool (Figure 3)

Tool 6 is the Transcelerate RBM tool being compared to the other non-commercial RBM tools. Risk category "blinding in the study design" is fully addressed by T7, while it is addressed partially by T1, T2, T3 and is not addressed by T4 and T5. Complexity risk category is fully addressed in 4 tools (T1, T2, T3, T7), partially addressed by T5 and not addressed in T4. Data collection is only addressed in T7. Endpoints are partially covered in 4 tools (T1, T2, T3, T7), while not addressed in 2 tools (T4, T5). Geography risk is not covered by any of the tools. Risk category related to investigational medicinal product (IMP) is addressed in T7 while partially addressed by 5 tools (T1, T2, T3, T4, T5). Logistics risk category is addressed in T7 and partially addressed in 5 tools (T1, T2, T3, T4, T5). Operation Complexity risk category is only addressed in T7 and not addressed in other tools as well (T1, T2, T3, T4, T5). As for safety risk category, it is addressed by 4 tools (T1, T2, T3, T7) while partially addressed and not addressed by 2 tools, T4 and T5 respectively. Staff experience risk category is addressed in 4 tools (T1, T2, T3, T7) and not addressed by 2 tools (T4, T5). Study phase is addressed by all tools except 1 tool (T4) where it is partially addressed. Risks related to subject population are addressed by 3 tools (T1, T2, T3) while partially addressed by 2 tool (T4, T5) moreover not addressed by 1 tool (T7). Risks related to Technology are only addressed by T7. Significant differences ($p < 0.05$) were observed between risks covered by each RBM tool.

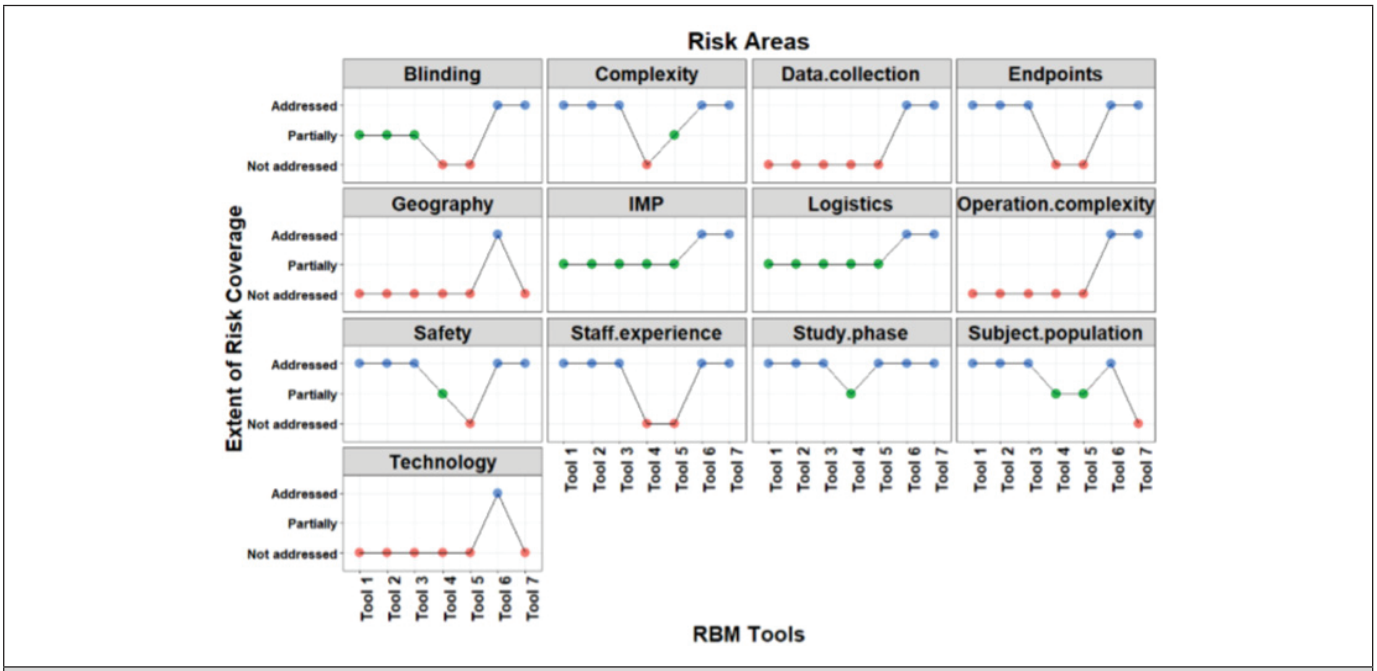


Figure 3: Risk area covered by each RBM Tool.

The assessment of protocols by the non-commercial RBM tools has shown that they result in different risk outcomes regardless of the clinical trial phase. Hence the mitigation plan to manage these risks will differ as well. The mitigation plan of an assessed risk should be implemented with either onsite monitoring or centralized monitoring [3]. The observed differences in the assessment clearly show that there is not yet an ideal non-commercial RBM. Each RBM tool focuses on specific risk aspects. Our findings highlight

differential risk considerations between RBM tools. Of the latter that fail to cover risk categories, their comparison revealed a significant difference. Moreover the weight age of a certain risk and its importance is usually assessed by the individual risk assessor implementing the RBM tool. The risks covered by each RBM tool should guarantee the safety and rights of the human subjects nevertheless the accuracy and reliability of data [3]. Our research points to apparent heterogeneity in the different risks being covered by each RBM tool.

Conclusion

An ideal RBM tool should cover risks related to a clinical trial. Further classification and scoring system should be included for the RBM tool user. Furthermore, a detailed monitoring strategy equipped with a proper plan to prevent detected risk should be readily available for the user.

A key quality feature of an RBM tool is the identification and classification of potential risks associated with a planned study. As described, the requirements are very different and sometimes only partially fulfilled by selected software tools as investigated. The development of such software requires a well-structured illustration of a clinical trial as it should be in the study protocol. In addition to this structural mapping, potential risks have to be defined in advance in order to enable a systematic screening by the software. Ideally, the potential risks should be weighted to allow for an internal assessment of a risk characteristic and thus to enable the generation of a study risk score. A corresponding preparatory work on this is still missing and is therefore, in the opinion of the authors, the next step in the further development of RBM systems whose developmental approach then no longer decides which risk level a clinical study has.

Limitations

The quality of the protocols was not taken into consideration, as our aim was to assess the protocols that follow ICH GCP guidelines and have already been reviewed, approved and accepted by the institutional review board (IRB). The author has solely done the assessment of the protocols with RBM tools, critical questions have been discussed within the author's group before the decision making process.

Conflict of Interest

None to declare

References

1. C Smith (2012) The Value of Source Data Verification in a Cancer Clinical Trial. Plos One 7(12): e51623.
2. C Baigent, FE Harrell, M Buyse, JR Emberson, DG Altman (2008) Ensuring trial validity by data quality assurance and diversification of monitoring methods. Clinical Trials 5(1): 49-55.
3. (2013) Food and Drug Administration (FDA). Guidance for Industry. Oversight of Clinical Investigations-A Risk- Based Approach to Monitoring.
4. Uren SC, Kirkman MB, Dalton BS, Zalcberg JR (2013) Reducing clinical trial monitoring resource allocation and costs through remote access to electronic medical records. Journal of oncology practice 9(1): e13-6.
5. Olsen R, Bihlet AR, Kalakou F, Andersen JR (2016) The impact of clinical trial monitoring approaches on data integrity and cost--a review of current literature. European journal of clinical pharmacology 72(4): 399-412.
6. E6 (R2) Good Clinical Practice: Integrated Addendum to ICH E6 (R1)-Guidance for Industry - Guideline for Good Clinical Practice - E6 (R2), 2018.
7. V Journot, JP Pignon, C Gaultier, V Daurat, A Bouxin-Métro, et al. (2011) Validation of a risk-assessment scale and a risk-adapted monitoring plan for academic clinical research studies-the pre-option study, Control Clin Trials 32(1): 16-24.
8. MRC/DH/MHRA Joint Project, Risk-Adapted Approaches to the Management of Clinical Trials of Investigational Medicinal Products, 2011 cited 2016 March 24.
9. C Hurley, F Shiely, J Power, M Clarke, JA Eustace, et al. (2016) Risk based monitoring (RBM) tools for clinical trials: A systematic review. Contemporary Clinical Trials.
10. O Brosteanu, G Schwarz, P Houben, U Paulus, A Strenge-Hesse, et al. (2017) Risk adapted monitoring is not inferior to extensive on-site monitoring: Results of the ADAMON cluster-randomized study. Clinical Trials 14(6): 584-596.
11. M Colban, A Sandsta, A Nikander, L Nybond, L Astrand, et al. (2015) Guideline for a coordinated GCP- monitoring of clinical trials in the Nodric contries.
12. MRC/DH/MHRA Joint Project (2011) Risk-adapted Approaches to the Management of Clinical Trials of Investigational Medicinal Products.
13. D Yee (2017) Data and safety monitoring plan.
14. (2013) Trans Celerate Bio Pharma Inc. Risk-Based Monitoring Methodology.
15. A Bouxin-metro, P Bertoye, G Chene, V Daurat, C Gaultier, et al. (2001) Risk assessment scale.
16. Quality Assurance Working Group of the Swiss Clinical Trial Organisation (SCTO) (2015) Guidelines for Good Operational Practice.

2.2 Improving Risk Assessment in Clinical Trials: Toward a Systematic Risk-Based Monitoring Approach.

Firas Fneish[1], Frank Schaarschmidt[1], Gerhard Fortwengel[2]

1. Leibniz Universität Hannover, Herrenhäuser strasse 2, 30419, Germany
2. Hochschule Hannover, Expo plaza 12, 30539, Germany

Type of authorship: First author

Journal: Current Therapeutic Research

Impact factor: 2.31

Number of citations: 2 (Google Scholar)

DOI: 10.1016/j.curtheres.2021qs.100643

Contributions:

1. Conceptualized and designed the study **(Firas Fneish, Frank Schaarschmidt, Gerhard Fortwengel)**
2. Programmed the shiny web app **(Firas Fneish)**
3. Writing the manuscript **(Firas Fneish)**
4. Critically reviewed the manuscript for intellectual content **(Frank Schaarschmidt, Gerhard Fortwengel)**

Current Therapeutic Research 95 (2021) 100643

Contents lists available at ScienceDirect

Current Therapeutic Research

journal homepage: www.elsevier.com/locate/curtheres

Improving Risk Assessment in Clinical Trials: Toward a Systematic Risk-Based Monitoring Approach

Firas Fneish[1,*], Frank Schaarschmidt[1], Gerhard Fortwengel[2]

[1] Institute of Cell Biology and Biophysics, Department of Biostatistics, Leibniz University Hannover, 30419 Hannover, Germany
[2] Faculty III Media Information and Design, Hochschule Hannover, 30539 Hannover, Germany

ARTICLE INFO

Article history:
Received 15 March 2021
Revised 26 July 2021
Accepted 4 August 2021

Key words:
Clinical trials
ICH-GCP
Mitigation plan
Risk-based monitoring
Risk methodology assessment

ABSTRACT

Regulatory authorities have encouraged the usage of a risk-based monitoring (RBM) system in clinical trials before trial initiation for detection of potential risks and inclusion of a mitigation plan in the monitoring strategy. Several RBM tools were developed after the International Council for Harmonization gave sponsors the flexibility to initiate an approach to enhance quality management in a clinical trial. However, various studies have demonstrated the need for improvement of the available RBM tools as each does not provide a comprehensive overview of the characteristics, focus, and application.

This research lays out a rationale for a risk methodology assessment (RMA) within the RBM system. The core purpose of RMA is to deliver a scientifically based evaluation and decision of any potential risk in a clinical trial. Thereby, a monitoring plan can be developed to elude prior identified risk outcome.

To demonstrate RMA's theoretical approach in practice, a Shiny web application (R Foundation for Statistical Computing) was designed to describe the assessment process of risk analysis and visualization tools that eventually aid in focusing monitoring activities.

RMA focuses on the identification of an individual risk and visualizes its weight on the trial. The scoring algorithm of the presented approach computes the assessment of the individual risk in a radar plot and computes the overall score of the trial. Moreover, RMA's novelty lies in its ability to decrease biased decision making during risk assessment by categorizing risk influence and detectability; a characteristic pivotal to serve RBM in assessing risks, and in contributing to a better understanding in the monitoring technique necessary for developing a functional monitoring plan.

Future research should focus on validating the power of RMAs to demonstrate its efficiency. This would facilitate the process of characterizing the strengths and weaknesses of RMA in practice.

Introduction

Clinical trials are conventionally monitored by source data verification that is costly, requires ample resources, and exhibits several limitations.[1,2] The International Council for Harmonization (ICH) has provided sponsors with the flexibility to initiate a novel approach called risk-based monitoring (RBM) to enhance quality management in a clinical trial.[3] Regulatory authorities such as European Medicines Agency (EMA) define RBM as a systematic process that involves identification, assessment, controlling, communicating, and reviewing the risks in a clinical trial before its initiation.[4] With this methodology, not only would the occurrence of the assessed risk be prevented, but it would also minimize onsite monitoring duties to some extent. Following the ICH recommendation for approach utilization, several RBM tools were developed. The available RBM tools have been identified and summarized based on their structural approaches, similarities, and differences.[5] Additionally, noncommercial RBM tools were compared in their application on real clinical trial protocols to assess the overall risk level of each protocol by each tool; furthermore, each noncommercial RBM tool was compared directly with the Transcelerate RBM tool (commonly accepted as the standard in pharmaceutical industry) to investigate the risk category and risk coverage in each.[6]

These studies reveal distinct approaches employed by the available RBM tools to assess a certain risk, demonstrate the unique assessment of each RBM Tool to the same clinical trial protocol, and

* Address correspondence to: Firas Fneish, Institute of Cell Biology and Biophysics, Department of Biostatistics, Leibniz University Hannover, Hannover, Germany.
E-mail address: fneish@cell.uni-hannover.de (F. Fneish).

https://doi.org/10.1016/j.curtheres.2021.100643

exhibit the different risks investigated within each RBM tool. The Food and Drug Administration (FDA) encourages all clinical trials regardless of the phase to implement RBM. Currently a standardized RBM approach for clinical trials is lacking,[7] which presents a challenge to implement RBM by the industry.[8] Our objective is to fill the gap by presenting a systematic risk analysis in clinical trials to standardize RBM. To have an efficient RBM tool, a potent risk assessment has to be performed first. For this reason, we propose a novel methodology and a robust algorithm to assess any risk in a clinical trial. The methodology can be implemented on any clinical trial regardless of the phase and complexity. Moreover, the algorithm aids the assessor in the decision-making process of monitoring technique needed and monitoring level required during the development of the monitoring plan.

Risk Identification Process

The quality feature of an RBM system entails risk assessment of a study and a mitigation plan that details a monitoring strategy for the concerned trial. However, the crucial question arising is how to define a certain risk.

The presence of varying risk criteria covered and examined within the risk assessment by each RBM tool suggests the need to restructure the definition of a certain risk. A risk is defined as the unsolicited outcome of a certain process. Any event that is likely to have a negative influence on the trial should be counted as a risk. The identified risk must be assessed through its influence on the safety of the human participant, trial integrity, the chance of its occurrence, and the ease by which it can be detected. Several systems such as Delphi[9] or SWOT analysis[10] can be oriented toward identifying risks in clinical trials. The Delphi method is a process that utilizes a questionnaire circulated among experts such as clinical research associates, statisticians, clinical investigators, sponsors, and any member involved in a clinical trial stage.[9] SWOT analysis is yet another strategy that aids organizations to pinpoint strengths, weaknesses, opportunities, and threats to a business or a project planning, in this case a clinical trial.[10] The application of both methods is simple, and their outcome is highly dependent on the diverse groups involved.[11] Another approach is utilizing risk summaries from monitoring reports of completed clinical trials; however, it is unlikely to access those reports as they are only accessible by the sponsors.[11]

An Ideal RBM System

Clinical trial sponsors along with the involved clinical trial members are responsible for guaranteeing the safety and well-being of the human participants, their rights, and the data quality.[12] The regulatory authorities require sponsors to ensure proper monitoring during the initiation and progress of a clinical trial.[13] RBM is expected to be an imperative tool in guiding the sponsor to identify and mitigate risks.[14] Similarly, EMA's reflection article concerning risk-based management demonstrates that a risk-based approach is needed to enhance quality management of clinical trials.[15] To date, FDA's guidance on RBM approach is divided into 3 parts, the detection of critical data and processes, the risk assessment categorization tool, and developing an appropriate monitoring plan following the risk-based approach.[16] Such a revolutionized technology played a huge role in achieving RBM in the field of mitigation monitoring techniques developed as remote monitoring.[17] The focus of any mitigation plan is shaped by the outcome of a risk assessment. Although 100% source data verification can certainly be reduced by the available mitigation plans, it does not reflect the focus of the personnel carrying out onsite monitoring activities, as the FDA entailed.[16]

Proposed Risk Methodology Assessment in Clinical Trials

An RBM tool that covers risks in any clinical trial including a monitoring plan of appropriate technique is still missing.[18] Additionally, there still exists ambiguity in the assessment methodology behind a certain risk. In this study we propose a novel risk methodology assessment (RMA) that enables the user to visualize the assessment of individual or overall risks present in a specific trial. RMA follows the concept of failure mode and effect analysis, specifically a systematic failure mode and effect analysis.[19] The focus is on system-related deficiencies in which hazards are identified, studied, and prevented.

The fundamental process is to initially focus on the most common faults detected in previous trials. For this reason, the RMA approach includes the frequent findings detected by Good Clinical Practice- Inspectors Working Group (EMA GCP-IWG) report.[20] The EMA GCP-IWG objective is to harmonize and coordinate GCP activities in the European Union. The annual report, which emphasizes GCP practice in the European Union, can be used as a reference for risk identification. The report sheds light on the number of inspections done routinely and non-routinely to active clinical trial sites and reports deficiencies detected in the trials.

Our article follows the recommendation of the ICH to favor risk based monitoring by providing a methodology of risk assessment that evaluates the occurrence likelihood of a risk, summarizes the extent of monitoring required with the help of a radar plot-based visualization of said risk and hence aids in the decision making of the mitigation step to be put forth. RMA does not suggest a prevention strategy due to the miscellaneous outcome of a certain risk in an individual trial. For instance, a risk associated with investigational medicinal products in a Phase I trial might have a higher impact than a Phase III trial. The anticipation step and the overall mitigation plan should be developed by the stakeholders responsible for the planning procedure. The FDA specifically highlights the sponsors' responsibility to have a mitigation approach for defined risks irrespective of the implemented risk assessment technique.[14] Figure 1 shows RMA's approach to identify, assess, and form a mitigation plan.

Theoretical Implementation of RMA Methodology

Each clinical trial is based on an explicit study protocol outlining the study end point(s), study procedures, medical investigations, and so on, which necessitate appropriate consideration during risk identification. The results presented by the GCP-IWG annual report signify the definite complications that a monitoring team can detect during a routine site visit. For this reason, the identification process of potential risks could be derived from GCP-IWG report as a starting point. Accordingly, a risk assessment should reflect the detected faults as risks that must be assessed before trial initiation.

A risk assessment system should consist of components in which a risk is identified, assessed, visualized for its monitoring level, and classified into the type of monitoring required. The assessment process is classified based on the FDA's recommendation of impact, probability, and detectability.[7] Nonetheless it does not indicate standards each category should be assessed on. It is left to the stakeholders to decide the appropriate decision process. In the presented methodology we propose defined standards required for impact and detectability measurements.

According to the ICH-GCP guidelines,[21] monitoring is conducted to ensure the well-being/safety of participants, the reliability of data and compliance with GCP/protocol guidelines. A risk that does not affect at least 1 of these criteria must not be deliberated as a risk that can be covered by RBM monitoring. The individual criteria should be differentially weighted based on the critical aspect re-

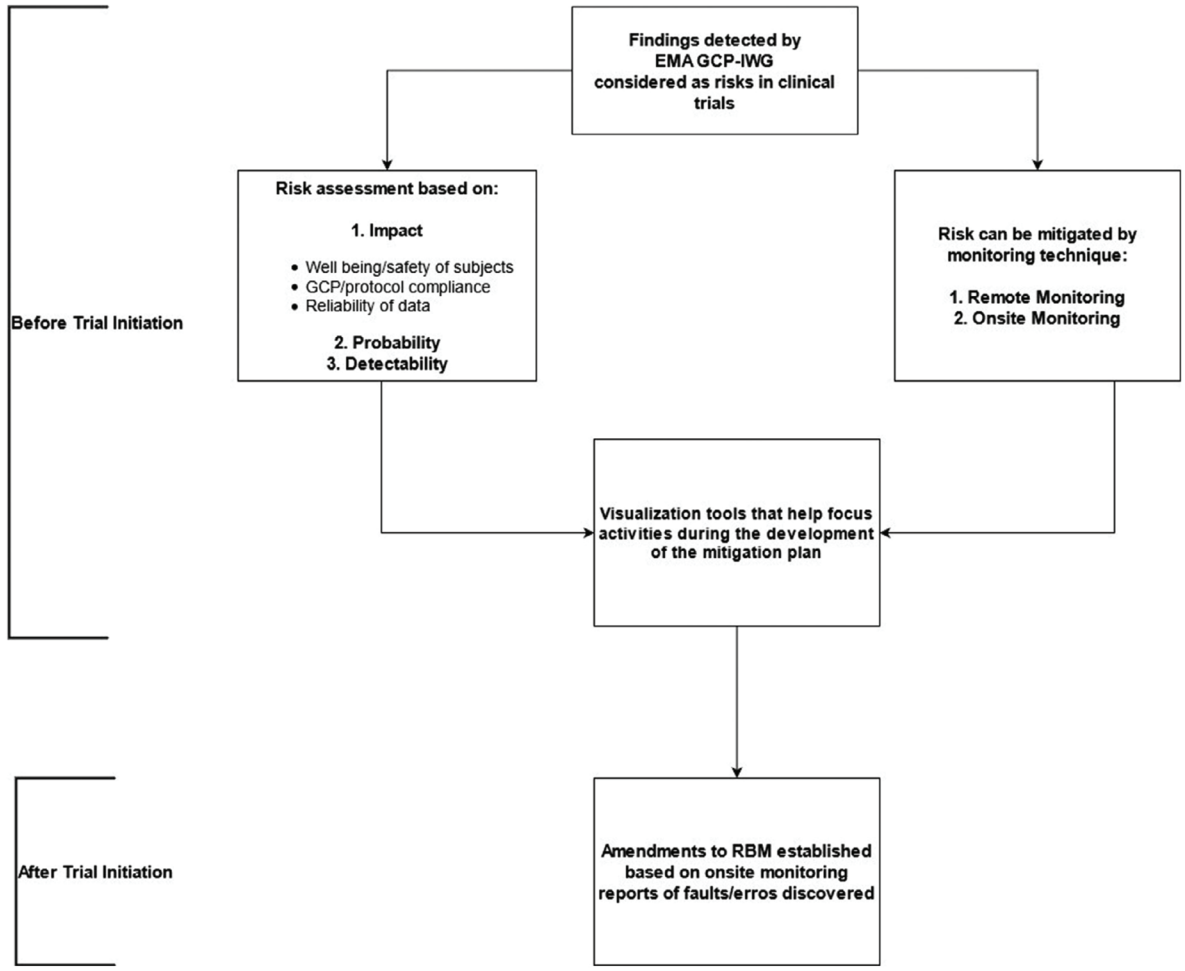


Figure 1. Flowchart of risk methodology assessment (RMA) risk assessment process before and after trial initiation. This flowchart shows the methodological approach of risk-based monitoring (RBM). Following risk identification, each risk is evaluated and assigned a mitigation technique. Following the assessment, stakeholders develop the monitoring plan based on the assessment. The assessment must be repeated if any amendments were established to the protocol or when unidentified faults are discovered.

Table 1
Risk assessment criteria. Following risk identification, each risk is evaluated based on the category it impacts, the probability of risk occurring, and the monitoring technique required for detection.

Criteria	Assessment category	Score
Impact	1. Well-being/safety of subjects	3
	2. Reliability of data	2
	3. Compliance with GCP/protocol guidelines	1
Probability	1. Very likely	5
	2. Likely	4
	3. Even chance	3
	4. Unlikely	2
	5. Very unlikely	1
Detectability	1. Onsite monitoring	2
	2. Remote monitoring	1

GCP = Good clinical practice.

sulting from each separately. For instance, a risk affecting the well-being/safety of participants alone will have a higher impact than a risk affecting GCP/protocol compliance. The detectability and probability should be assessed by the stakeholders based on their decision process. However, probability is weighed based on the likelihood of a risk occurrence and detectability is evaluated based on the monitoring detection technique either as remote monitoring or onsite monitoring. We propose a score measure for the category of each criterion (Table 1).

Scoring Method

The scoring algorithm of RMA allows the stakeholders a unique prospect to visualize the risk size and quantify it. The goal of risk communication is to guide the stakeholders in the risk assessment in a transparent manner and to assist them in the decision plan to mitigate its occurrence by an effective measure.[22] Visual representation can help stakeholders observe the assessment of the risk and understand its needed monitoring level. The visualization process can be achieved by radar charts as they enhance comparisons of quality measurements.[23]

With the defined scaling system, the area would reflect the extent of how critical a risk is, which subsequently hints to the extent of monitoring required. The larger the area, the more monitoring is required; however, it does not reflect the type of monitoring technique needed as this must be decided by the stakeholders themselves (Figure 2). Following the assessment, a monitoring technique should be assigned. According to regulatory agencies, the main techniques can either be traditional onsite monitoring, remote monitoring, or a combination of both.

Area Under the Radar Chart

The aim of radar chart is to present multivariate data, the main advantage is to translate the data to a meaningful sense. The area under the radar is equivalent to the cumulative area of the sep-

Figure 2. The area under the radar chart. This figure shows the total area of the radar chart. Each area of the subtriangles is calculated based on the conventional formula.

arate triangles (Figure 2). The area under the radar chart is then reported as a percentage of the maximum score possible.

Each area is detected by the sides of the respective triangle input

$$Area\ A1 = 0.5\ \times input\ (Impact) \times input\ (Detectability) \times \sin(120)$$

$$Area\ A2 = 0.5\ \times input\ (Impact) \times input\ (Probability) \times \sin(120)$$

$$Area\ A3 = 0.5\ \times\ input\ (Probability) \times input\ (Detectability) \times \sin(120)$$

$$Total\ Radar\ Area\ = Area\ A1 + Area\ A2 + Area\ A3$$

Practical Implementation of RMA

A shiny web application was formed to illustrate the theoretical approach of RMA. The application includes risks that could be assessed and visualized under the radar plot (Figure 3).

Following the assessment of the individual risks, the input scores provided by the assessor and the subsequent score areas are documented. The following process can aid stakeholders in comparing the assessment report with monitoring reports after trial initiation to get a better understanding of the faults/weaknesses and strengths of the performed assessment (Figure 4).

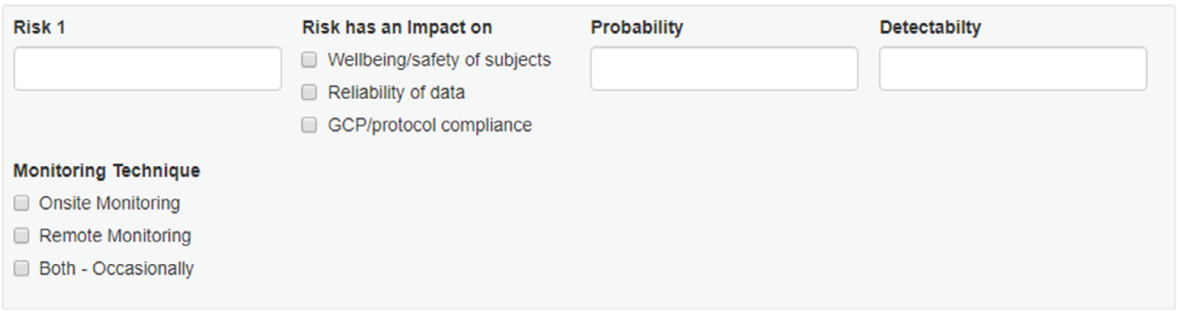
The score of the distinct risks assessed allows stakeholders to distinguish high score risks that necessitate more extensive monitoring in the monitoring plan (Figure 5a). Consequently, based on the profile input of each risk (Figure 5b) and its relation to the threshold for maximum score, represented by dashed lines, stakeholders can decide on the extent of monitoring visits/checks required in the monitoring plan. Finally, an overview of the sum of risks to be monitored by each technique (Figure 5c) imparts a clearer understanding of the type of monitoring plan needed, which is highly essential in the application of RBM.

The assessment process should repeated as soon as amendments are made to the trial protocol or when identifying new risks during monitoring process after trial initiation. This would require the stakeholders to conduct a new risk assessment to engage a proper mitigation action in the monitoring plan. It is essential to act on a new identified risk to understand its direct effect on the overall score of the risk assessment as a whole and on the monitoring technique required to prevent its occurrence.

Generally, the monitoring activities of the clinical research coordinator/monitoring team should focus on the requirements, responsibilities, and hazards that can carry potential liabilities to the trial assurances. The final assessment report will stipulate the potential risks to be monitored and frequency of monitoring needed.

Because RBM is becoming a principle stage in clinical trials,[24] both RMA's strategy and approach have the potential to improve data quality and reduce clinical costs. Undoubtedly, the risk assessment within other RBM systems can also identify certain risks; however, the assessment methodology of the individual risk criteria is either not reported or vague. As for their systems, they are fixed on prespecified risks lacking the ability of tallying new ones. For this reason, RMA's scoring system provides a means to facilitate confirmation of a certain risk and assess its outcome measure. Additionally, it incorporates flexibility in directly, including an additional risk area in the assessment report. Finally, once the entire risk assessment is completed, risks could be grouped based on the monitoring technique to assist the stakeholders in the trial monitoring plan development. The established method can be considered a primary step toward a practical monitoring guidance in which a monitoring plan form will be based on different risks in a trial, individual process, and required monitoring.[25]

The innovative approach of RBM will facilitate establishment of adequate and focused monitoring activities, reduce 100% source data verification activities, and enhance the quality of the trial and patient safety.[26] This goal should be clearly communicated to stakeholders and clinical trials to prevent misconceptions among clinical research coordinators regarding RBM's outcome in increasing workload, a concern that has been previously reported, despite



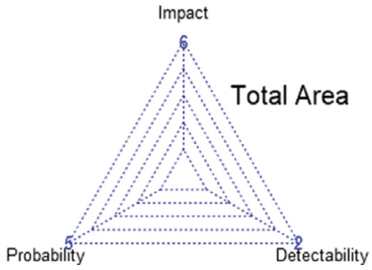


Figure 3. The individual risk assessment presented by the radar chart. This figure shows the criteria of risk assessment that should be completed by the assessor. The individual risk is assessed by its impact, probability, and detectability. Accordingly, the total area of the risk is presented by the radar chart.

	Risk	Impact	Probability	Detectability	Monitoring	Score
1	Missing/Lack of essential document(s)					
2	Receipt of IMP shipment to site (Delay, ect.)					
3	Records of blood samples shipment to the central laboratories(Delay, etc.)					
4	Having incomplete documentation					
5	Incomplete screening list (Not following screening appointments)					
6	lack of contemporaneous independent copy of the CRF filed on site					
7	SOPs won't be followed/used					
8	SOPs won't be updated as required					
9	The implementation of an efficient quality management system by the Sponsor					
10	Risk of having discrepancies between source data and data reported in the CSR					

Figure 4. Assessment score of each individual risk with corresponding input. This figure shows the documentation of the individual risks assessed with its input criteria score and the computed overall score.

its capacity to do the opposite. RBM is a continued improvement process that requires all stakeholders and clinical trial staff to initiate the risk assessment before and during the trial period. An effective monitoring plan can only be achieved after a successful implementation of RBM.[27] We believe the RMA approach can aid stakeholders in distinguishing and evaluating any potential risk. Future investigation should focus on validating the power of RMAs to demonstrate efficiency in practice.

RMA could be further developed to software that utilizes existing data to forecast a certain risk outcome and provide a mitigation plan based on the risk score. Further work is required to achieve the desired prediction. Classification models may be employed to predict the existence of a specific risk and measure its individual score; however, numerous factors such as data quality and model fit variability require consideration during the utilization of such models.[28] Artificial intelligence algorithms should be the next phase of any risk assessment. Transparent risk methodologies such as RMA should be made available to both regulatory authorities and the public. The prospect of being able to estimate a risk outcome and potential mitigation serves as a continuous incentive for future research. We believe the efficiency of RBM has been well established and proven; yet the ultimate design of RBM development will be a challenge for us for years.

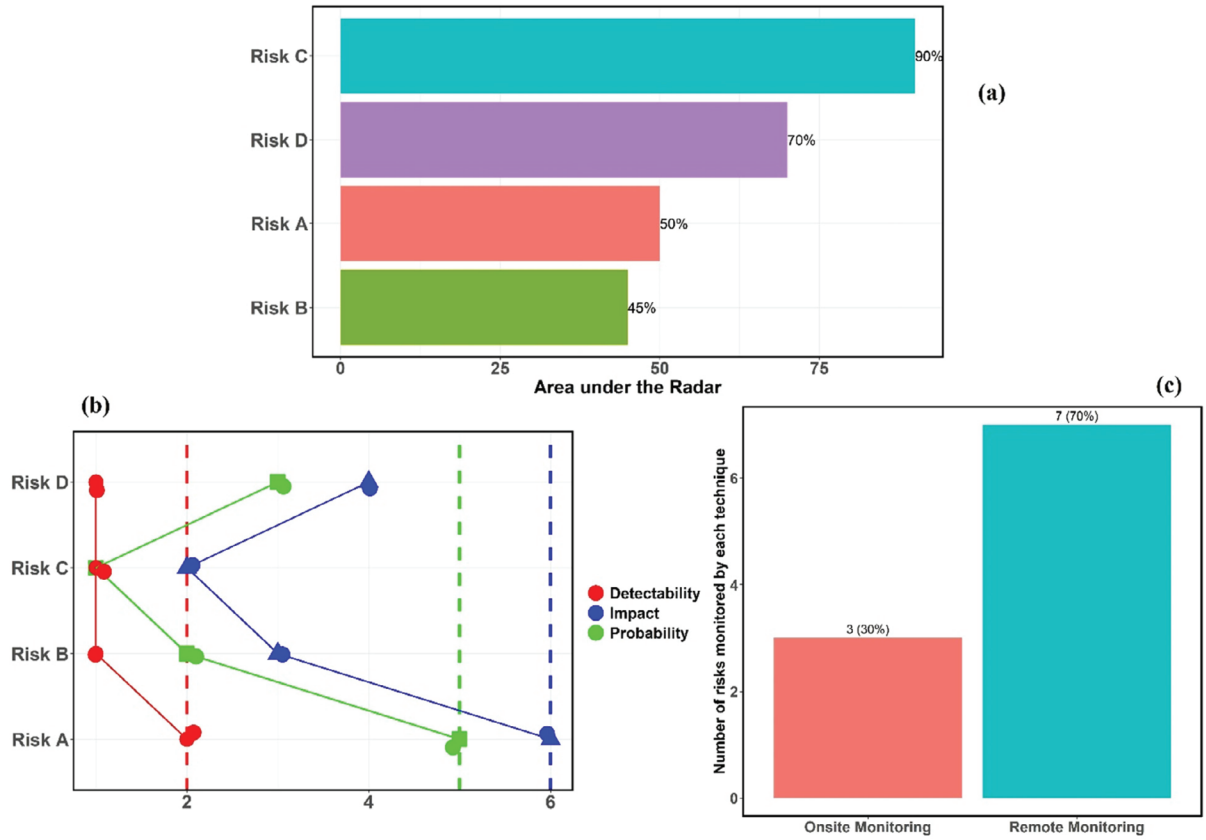


Figure 5. Risk methodology assessment. (A) Overall scores (area under the radar) of each risk. (B) Risks based on the input of the assessment; red, green, and blue points are compared with their respectively colored dashed lines representing the maximum score. (C) Overall counts of risks covered by each monitoring technique.

Acknowledgments

The authors thank the journal for its editorial support and the anonymous referees for their numerous constructive comments toward improving our manuscript.

Shinyapp & Code

The shinyapp was coded using R software (R Foundation for Statistical Computing) and is platform independent; specifically, an interactive hypertext markup language document is produced using Rmarkdown runtime shiny. The Rmarkdown shiny syntax is deployed to shinyapps server in which it preserves the functionality of the code. The syntax is available on Github at https://github.com/firasfneish/Risk-Methodology-Assessment. Project home page: https://firasfneish.shinyapps.io/Risk_Based_Monitoring_Methodology/

Author Contributions

F. Fneish designed and conceptualized the methodology of the tool, programmed the shinyapp, and drafted the manuscript for intellectual content. F. Schaarschmidt, critically reviewed the manuscript. G. Fortwengel designed and conceptualized the methodology of the tool and critically reviewed the manuscript. The authors approved the final version of the manuscript.

Conflicts of Interest

None.

References

1. Tudur SC, Stocken DD, Dunn J, et al. The value of source data verification in a cancer clinical trial. *PloS one*. 2012;7(12):e51623. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0051623.
2. Baigent C, Harrell FE, Buyse M, et al. Ensuring trial validity by data quality assurance and diversification of monitoring methods. *Clinical trials (London, England)*. 2008;5(1):49–55. doi:10.1177/1740774507087554.
3. E6 (R2) Good Clinical Practice: Integrated Addendum to ICH E6 (R1)-Guidance for Industry - Guideline for Good Clinical Practice - E6 (R2). 2018. https://www.fda.gov/files/drugs/published/E6%28R2%29-Good-Clinical-Practice–Integrated-Addendum-to-ICH-E6%28R1%29.pdf:
4. EMA Guidance. Reflection paper on risk based quality management in clinical trials. 2013. https://www.ema.europa.eu/en/documents/scientific-guideline/reflection-paper-risk-based-quality-management-clinical-trials_en.pdf.
5. Hurley C, Shiely F, Power J, et al. RBM Tools for clinical trials: a systematic review. *Contemp Clin Trials*. 2016;51:15–27. https://www.sciencedirect.com/science/article/pii/S1551714416302877.
6. Fneish F, Limaye D, Struever V, et al. Comparison of Non-Commercial Risk Based Monitoring Tools by Their Application on Clinical Trial Protocols. *Am J Biomed Sci & Res*. 2020;8(3). https://biomedgrid.com/pdf/AJBSR.MS.ID.001276.pdf.
7. Miesta E. New FDA Guidance Answers Questions On RBM Use (clinicalleader.com). 2019. https://www.clinicalleader.com/doc/new-fda-guidance-answers-questions-on-rbm-use-0001
8. Limaye N, Jaguste V. Risk-Based Monitoring (RBM) Implementation: Challenges and Potential Solutions. *Ther Innov Regul Sci*. 2019;53:183–189. https://link.springer.com/article/10.1177/2168479018769284.
9. Grime M.M, Wright G. Delphi Method. Wiley StatsRef: Statistics Reference Online. 2016. https://doi.org/10.1002/9781118445112.stat07879
10. Sammut T, Galea B. *book*. Wiley Encyclopedia of Management. Wiley Encyclopedia of Management; 2015 *https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118785317.weom120103*.
11. Olson D, Wu D. *Enterprise Risk Management Models*. Germany: Springer; 2017 *https://link.springer.com/book/10.1007/978-3-662-53785-5*.
12. Beever D, Swaby L. An evaluation of risk-based monitoring in pragmatic trials in UK. *Clinical Trials Units. Trials.*. 2019;20:556 https://doi.org/10.1186/s13063-019-3619-6 21.
13. Food and Drug Authority FDA. Code of federal regulations. 21 CFR part 312, subpart D. 2019. Responsibilities of Sponsors and Investigators.

https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=312&showFR=1&subpartNode=21:5.0.1.1.3.4
14. Food and Drug Administration FDA. Guidance for Industry Oversight of Clinical Investigations — A Risk-Based Approach to Monitoring. 2013. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/oversight-clinical-investigations-risk-based-approach-monitoring
15. Food and Drug Administration FDA. A Risk-Based Approach to Monitoring of Clinical Investigations Questions and Answers Guidance for Industry. 2019 https://www.fda.gov/media/121479/download
16. Buyse M. EMA Guidance Points to Central Statistical Monitoring. *Applied Clinical Trials*. 2014. https://www.appliedclinicaltrialsonline.com/view/ema-guidance-points-central-statistical-monitoring. 23-2.
17. Quanticate. Risk-Based Monitoring. Accessed on 9th December 2020. https://www.quanticate.com/risk-based-monitoring
18. Simović M, Nikolić N. Challenges of risk-based monitoring of clinical trials. *Clinical Research and Regulatory Affairs*. 2015;32(3):83–87. https://www.tandfonline.com/action/showCitFormats?doi=10.3109%2F10601333.2015.1046990.
19. Failure Mode and Effects Analysis (FMEA). Accessed on 9th December 2020. https://quality-one.com/fmea/
20. European Medicines Agency EMA. Annual Report of the Good Clinical Practice Inspectors' Working Group 2018. https://www.ema.europa.eu/en/documents/annual-report/annual-report-good-clinical-practice-inspectors-working-group-2018_en.pdf
21. European Medicines Agency EMA ICH-GCP guideline. *Monitoring section*. 2021. https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-6-r2-guideline-good-clinical-practice-step-5_en.pdf
22. Roth F. Visualizing Risk: The Use of Graphical Elements in Risk Analysis and Communications. CSS Risk and Resilience Reports. 2012. https://www.researchgate.net/publication/255723470_Visualizing_Risk_The_Use_of_Graphical_Elements_in_Risk_Analysis_and_Communications
23. Nowicki H, Merenstein C. Radar chart. 2016 https://skylight.middlebury.edu/~candrews/showcase/infovis_techniques_s16/radar_chart/
24. Cyntegrity. Risk-Based Monitoring, What is It? Accessed on 9th December 2020. https://cyntegrity.com/about-risk-based-monitoring/
25. Love SB, Yorke-Edwards V, Lensen S, et al. Monitoring in practice – How are UK academic clinical trials monitored? A survey. *Trials*. 2020;21(59). doi:10.1186/s13063-019-3976-1.
26. Higa A, Yagi M, Hayashi K, Kosako M, Akiho H. Risk-Based Monitoring Approach to Ensure the Quality of Clinical Study Data and Enable Effective Monitoring. *Therapeutic Innovation & Regulatory Science*. 2020;54(1):139–143. https://link.springer.com/article/10.1007%2Fs43441-019-00037-x.
27. Jeong IS, Jeong JH, Hwang YS, Youn JH. Clinical Research Coordinators' Attitude Toward Risk-Based Monitoring. *Therapeutic Innovation & Regulatory Science*. 2019;53(5):691–695. https://link.springer.com/article/10.1177/2168479018793132.
28. Grant SW, Collins GS, Nashef SM. Statistical Primer: developing and validating a risk prediction model. *European Journal of Cardio-Thoracic Surgery*. 2018;54-2 Pages 203–208. doi:10.1093/ejcts/ezy180.

2.3 Application of statistical methods for central statistical monitoring and implementations on the German Multiple Sclerosis Registry

Firas Fneish[1,2], David Ellenberger[2], Niklas Frahm[2], Alexander Stahman[2], Gerhard Fortwengel[3], Frank Schaarschmidt[1]

1. Leibniz Universität Hannover, Herrenhäuser Strasse 2, 30419, Germany
2. German MS-Register, MS Forschungs- und Projektentwicklungs- gGmbH [MSFP], Hannover, Germany
3. Hochschule Hannover, Expo plaza 12, 30539, Germany

Type of authorship: First author

Journal: Therapeutic Innovation & Regulatory Science (Under Review)

Impact factor: 1.337

Number of citations: Not applicable

DOI: Not applicable

Contributions:

1. Conceptualized and designed the study **(Firas Fneish, David Ellenberger, Alexander Stahmann, Frank Schaarschmidt, Gerhard Fortwengel)**
2. Design of the Monte-Carlo simulations **(Firas Fneish, Frank Schaarschmidt)**
3. Implementation of the Monte-Carlo simulations **(Firas Fneish)**
4. Data analysis of simulation results **(Firas Fneish)**
5. Data analysis of the German Multiple Sclerosis Registry **(Firas Fneish, David Ellenberger)**
6. Programmed the shiny webapps **(Firas Fneish)**
7. Writing the manuscript **(Firas Fneish, Frank Schaarschmidt)**
8. Reviewed the manuscript for intellectual content **(David Ellenberger, Niklas Frahm, Alexander Stahmann, Frank Schaarschmidt, Gerhard Fortwengel)**

Application of statistical methods for central statistical monitoring and implementations on the German Multiple Sclerosis Registry

Authors: Firas Fneish[a,b]*, David Ellenberger[b], Niklas Frahm[b], Alexander Stahmann[b], Gerhard Fortwengel[c], Frank Schaarschmidt[a]

[a] Department of Biostatistics, Institute of Cell Biology and Biophysics, Leibniz University Hannover, Hannover, Germany

[b] German MS-Register, MS Forschungs- und Projektentwicklungs- gGmbH [MSFP], Hannover, Germany

[c] Faculty III – Media, Information, and Design, Hochschule Hannover, Hannover, Germany

* Corresponding author: Firas Fneish

Abstract

Monitoring of clinical trials is a fundamental process required by regulatory agencies. It assures the compliance of a center to the required regulations and the trial protocol. Traditionally, monitoring teams relied on extensive on-site visits and source data verification. However, this is costly, and the outcome is limited. Thus, *central statistical monitoring* (CSM) is an additional approach recently embraced by ICH to detect problematic or erroneous data by using visualizations and statistical control measures. Existing implementations have been primarily focused on detecting inlier and outlier data. Other approaches include principal component analysis and distribution of the data. Here we focus on the utilization of comparisons of centers to the Grand mean for different model types and assumptions for common data types, such as binomial, ordinal, and continuous response variables. We implement the usage of multiple comparisons of single centers to the Grand mean of all centers. This approach is also available for various non-normal data types that are abundant in clinical trials. Further, using confidence intervals, an assessment of equivalence to the Grand mean can be applied. In a Monte Carlo simulation study, the applied statistical approaches have been investigated for their ability to control type I error and the assessment of their respective power for balanced and unbalanced designs which are common in registry data and clinical trials. Data from the German Multiple Sclerosis Registry (GMSR) including proportions of missing data,

adverse events and disease severity scores were used to verify the results on Real-World-Data (RWD).

Keywords: Monitoring, Data quality control, Multicenter clinical trials, Grand mean, Registry data

Introduction

Multicenter clinical trials are imperative to obtain a conclusive assessment concerning the safety and efficacy of medical treatments. They involve diverse clinics or hospitals, and their respective personnel(1). This requires the monitoring team to ensure the compliance of each center to the study protocol and the requirements of good clinical practice. Compliance of a center to the required regulations will make the center's data more reliable. Non-compliance events may lead to errors in patient inclusion criteria, operating procedures and to various types of data entry errors(2,3). Additionally, data tampering or fraud may occur in a single center(4). All these difficulties may result in biased estimates of the investigated treatments efficacy as well as to false positive or false negative detection of safety issues. The monitoring team traditionally performs on-site visits to each study center to ensure compliance of the regulatory requirements; however, these activities have been reported to be costly and of limited outcome with regards to data quality(5,6). In the preceding years, central statistical monitoring (CSM) was proposed as an amendment to a thorough source data verification (SDV) that requires on-site visits(7,8).

CSM utilizes graphical approaches, summary statistics and statistical tests to assess incoming data from all centers in the trial(9–11). The assessment of center compliance can be achieved by statistical models to assess adherence levels. The primary aim is to detect data entry errors, adverse event rates in single centers or safety issues related to individual patients. Moreover, CSM serves to identify centers that could require additional monitoring activities due to deviations or outlier detection. A robust risk assessment of the key risk indicators (KRIs) in clinical trials can target onsite-monitoring activities(12,13). Risk assessment prior to trial initiation can facilitate whether an onsite-monitoring technique or CSM technique is needed to monitor a certain risk. Timmerman et al. (2016) illustrates how CSM can be a means to identify KRIs to target adaptive monitoring(14).

Numerous statistical methods have been applied for monitoring approaches for the implementation of CSM(9,15,16). Based on covariate type, statistical methods were applied to detect atypical/outlier data. For the purpose of risk based centralized monitoring, classical statistical methods have been categorized as unsupervised and supervised monitoring techniques(17). Existing publications on CSM focused on outlier/inlier detection on different levels e.g., , center, country and regional and demonstrated the usage of principle component analysis on the center level.

However, single centers in multicenter clinical trials might deviate from the study protocol or inclusion criteria. They might also deviate in clinical practice, or there might be misunderstandings concerning the definition of adverse events or categorical variables or disease severity scores to be recorded. Such deviations will not produce single extreme values in the data. They will rather lead to deviating summary statistics, adverse event rates, or class frequencies of categorical data. Desmet et al. (2014) proposed the usage of linear mixed effects models to detect location differences between center and other centers for a continuous outcome and a beta binomial model for proportion comparison for a certain event in a center(18,19). In the following paper we propose to use multiple comparisons of single centers to the Grand mean (GM) of all centers. This approach is available for various data types that are abundant in clinical trials. It can be used to detect centers that are significantly deviating from average. Further, confidence intervals are available, such that an assessment of equivalence to the average can be applied. Center comparisons to the GM of the data has been an overlooked aspect. In the following, we will firstly define comparisons to the GM for different model types and assumptions for common data types, such as binomial, ordinal, and continuous response variables. Generalized linear models (GLM), bayesian linear models (BayesGLM), and bias-reduced generalized linear models (BrGLM) were applied for binomial outcomes. For continuous outcomes, a non-parametric and a linear approach are investigated. As for ordinal data, a non-parametric approach is assessed. The correction for multiple testing is accounted for when performing the contrasts. Since approaches are asymptotic and thus depend on the sample size, they were investigated in a Monte Carlo simulation for their ability to control type I error (α) and achieve the highest possible power ($1-\beta$). We demonstrate the implementation of these methods on examples based on data from the German Multiple Sclerosis Registry (GMSR)(20).

Real-world data from GMSR

CSM aids clinical trials and registries in data monitoring for many variables. GMSR collects data directly from participating centers through a certified web-based data capture (EDC) system. The data collected includes a wide range of variables such as patient profile, disease status and medication treatments. We refer to Ohle et al. (2021) for further details on the GMSR(20). We included centers that are participating in the pharmacovigilance module at the GMSR each having at least 50 patients under observation in the database. An overview of the GMSR data is shown in Table 1 for specific variable types considered in this research.

[Table 1 insert here]

Figure 1 shows the dataset of three variables for each center. The dataset covers age at onset, adverse events (AE), expanded disability status scale (EDSS) representing continuous, binomial, and ordinal data types respectively. EDSS and AE are reported for each visit. Figure 1a shows the distribution of patients' age at onset and highlights that data may not be normally distributed. Shapiro test was used to indicate whether the data of individual centers follow the normal distribution. Violations of the normality assumption suggest the need for non-parametric methods to perform center comparisons to the GM. For the same variable Figure 1b shows the missingness found in each center. It illustrates the center performance in terms of data completeness. Although it is common to have missing data, the question arises at what level it is unacceptable? Similarly for adverse events, one center (C3) reports 38% of patients having adverse events while other centers range between 0% and 24% (Figure 1c), this observation again designates a variation in the proportions for a certain event between centers and shows the need for a test to hint for the problematic center(s). As for EDSS measurements (Figure 1d) it exhibits a clear difference for disease severity for patients between centers.

The visualization of these variables provides to the stakeholders an overview of the data at hand. However, it does not directly pinpoint or highlights a problematic center. Although the observed differences between centers could be natural due to patient variation or other factors, it is essential to confirm deviating centers at a given statistical certainty. In some cases, inference of a center being problematic can only be deduced with appropriate statistical testing e.g., complex multicenter clinical trial. This dataset will be

used to demonstrate comparisons of the individual mean center to the GM of all centers for different scales of measurement.

[Figure 1 insert here]

Materials and Methods

We consider a wide spectrum of scenarios relevant to registry and clinical trial data with several centers for the response outcome variable. Let i be the index of the centers in a clinical trial $i = 1, \dots, I$. Within each center i there are n_i subjects, with subject index $j = 1, \dots, n_i$. The GM of all centers within the trial is denoted by $\hat{m}_.$.

Comparisons to $\hat{m}_.$

For a given model with parameters m_i and possibly unbalanced samples sizes n_i the GM $m_.$ can be computed by $m_. = \sum_{i=1}^{I} \frac{n_i}{N} m_i$, where N is the total sample size, $N = \sum_{i=1}^{I} n_i$. Comparisons of each centers parameter m_i to the GM $m_.$ can then be written as a set of $k=1,\dots,K$ linear contrasts, with contrast coefficients $c_k = (c_{k1}, c_{k2}, c_{k3}, \dots, c_{kI})$:

$$c_1 = \left(1 - \frac{n_1}{N}, -\frac{n_2}{N}, -\frac{n_3}{N}, \dots, -\frac{n_I}{N}\right)$$

$$c_2 = \left(-\frac{n_1}{N}, 1 - \frac{n_2}{N}, -\frac{n_3}{N}, \dots, -\frac{n_I}{N}\right)$$

$$\dots$$

$$c_K = \left(-\frac{n_1}{N}, -\frac{n_2}{N}, -\frac{n_3}{N}, \dots, 1 - \frac{n_I}{N}\right)$$

The deviation of the kth center from the GM can then be written as:

$$d_k = m_{i=k} - m_. = \sum_{i=1}^{I} c_{ki} m_i$$

Written in this way, the comparisons to GM are a special case of the framework of testing general linear hypotheses(21). In this framework it is possible to perform hypotheses tests adjusted for multiple comparisons and to compute simultaneous confidence intervals for the parameters defined by the contrasts.

In this application, it can be of interest to test the null hypothesis that no center deviates from the overall mean,

H_0: $d_k = (m_{i=k} - m_{.}) = 0$, for all *k=1,...K,*

versus the alternative hypothesis that at least one center deviates from the overall mean,

H_A: $d_k = (m_{i=k} - m_{.}) \neq 0$, for at least one *k=1,...K.*

In some cases, a test decision concerning a significant deviation might not be of interest. Like in tests on equivalence, objective can be to infer whether single centers do not show a relevant difference from the overall mean. In this case, a prior definition of relevant deviations or equivalence margins,$[-\delta, \delta]$, has to be specified based on subject knowledge. Then, it can be inferred whether the upper and lower confidence limits for each center's deviation d_k are included in this range or not.

For the full details of computing p-values of the above hypothesis tests and simultaneous confidence intervals, we refer to Hothorn et al. (2008)(21). The most important steps from Hothorn et al. (2008) are outlined below. Stacking the *k=1,...,K* vectors of contrast coefficients, c_k, yields a contrast matrix C with K rows and I columns. Fitting linear or generalized linear models yields a vector of estimates of the model parameters with elements $\hat{m}_i$, $\hat{m} = (\hat{m}_1, \hat{m}_2, \hat{m}_3, \dots, \hat{m}_I)^T$ and the corresponding estimated variance-covariance matrix of model parameters, $\hat{V}$. Estimates for the deviations of centers from the GM are then $\hat{d} = C\hat{m}$, the corresponding variance-covariance matrix of these deviations is $\hat{U} = C\hat{V}C^T$, where ^{T} denotes a transposed vector or matrix. The estimated variance of the elements $\hat{d}_k$ in $\hat{d} = (\hat{d}_1, \hat{d}_2, \hat{d}_3 \dots, \hat{d}_K)$ are the diagonal elements of $\hat{U}$, $\hat{u} = diag(\hat{U})$, with elements $\hat{u}_k$. Their square roots are then the estimated standard errors of the $\hat{d}_k$, that is, $\widehat{se}(\hat{d}_k) = \sqrt{\hat{u}_k}$. Finally, the estimated correlation matrix $\hat{R}$ of $(\hat{d}_1, \hat{d}_2, \hat{d}_3 \dots, \hat{d}_K)$ follows from standardizing the matrix $\hat{U}$ with its diagonal elements $\sqrt{\hat{u}_k}$.

Tests of the hypotheses presented above are then based on the test statistics $t_k = \frac{\hat{d}_k}{\widehat{se}(\hat{d}_k)}$, the corresponding adjusted p-values are computed from a multivariate t distribution (or asymptotically from a multivariate normal distribution) with correlation matrix $\hat{R}$, for linear models or generalized linear models, respectively.

Simultaneous confidence intervals for each center's deviation from GM, $\hat{d}_k$, can be computed using the formula

$$\left[\hat{d}_k \pm q_{1-\alpha,two-sided,\hat{R}}\widehat{se}\left(\hat{d}_k\right)\right]=\left[(\hat{m}_{i=k} - \hat{m}_{.}) \pm q_{1-\alpha,two-sided,\hat{R}}\widehat{se}(\hat{m}_{i=k} - \hat{m}_{.})\right]$$

where $q_{1-\alpha,two-sided,\hat{R}}$ is the two-sided equicoordinate (1-α) quantile of multivariate t or multivariate normal distribution, respectively. For further details of computing adjusted p-values and quantiles of multivariate t and normal distributions, we refer to Genz and Bretz (2009)(22).

For a thorough data interpretation, merely relying on rejection/non-rejection at one significance at level, say 0.05, or merely relying on the presented p-values is discouraged (e.g. ASA statement on p-values, Wasserstein & Lazar,2016)(23). Rather, estimated effects (here, deviations from grand mean) and the corresponding confidence limits should be displayed and used for interpretation: Then, the relevance of observed effects can be assessed, or, non-inferiority or equivalence can be assessed based on inclusion of confidence limits in pre-specified equivalence margins for the corresponding parameter.

Response variables

Continuous outcomes

Continuous data may follow the normal distribution, possibly after a suitable data transformation to achieve normality and homogeneous variances. It can then be analyzed by the model used in 1-way analysis of variance

$$Y_{ij} \sim m_i + \varepsilon_{ij} \quad , \quad \varepsilon_{ij} \sim N(0, \sigma^2)$$

Here, m_i is the expected value of center i. In this case, the above multiple comparison procedure is well established and exact. In case of continuous outcomes which are in contradiction to normality before and after transformations, a non-parametric method is described in section 3.3.3 as an alternative.

Binomial outcome

The number of events Y_i is assumed to follow a binomial distribution in which π_i is the event probability in a center i.

$$Y_i \sim Binomial(n_i, \pi_i)$$

For binomial data, we will assume that a generalized linear model (glm) is fitted with the canonical logit link:

$$m_i = \log(\frac{\pi_i}{1-\pi_i})$$

Thus, comparisons to GM will be performed on the logit scale(24).

Excess 0s in binomial data

Fitting a classical generalized linear model for binomial data with zero excess $Y_i = 0$ successes/failures is a common problem in different scientific fields such as clinical trials and toxicological experiments(25). As soon as $Y_i = 0$ in one or several centers, numerically $m_i = \log(\frac{\pi_i}{1-\pi_i})$ becomes very small and $se(m_i)$ will be very large. Several alternatives are available to avoid extreme se. In the next subsection we consider two alternatives, a Bayesian linear model(26) and the Bias-reduced generalized linear model(27).

Estimators and models assumptions

Bayesian generalized linear models (BayesGLM) for binomial endpoint.

The first approach we consider for dealing with zero excess binomial data $Y_i = 0$ in one or several centers is a Bayesian linear model with non-informative priors. Gelman et al. (2008) used scaled Cauchy distributions as priors for each model parameters that estimate effects, e.g. differences on the logit scale. Cauchy priors for a model parameter entail the assumption that extreme center effects on the logit scale are implausible. Prior assumptions for a baseline risk or control group allows a wider range such that $10^{-9} < \log(\pi_i/1-\pi_i) < 1-10^9$ (Gelman et al. 2008)(28). Prior assumptions on parameters impose a restriction on the parameter estimation; this prevents that estimated parameter from becoming extreme and thus prevents the standard error from becoming extreme as well.

Bias-reduced generalized linear models (BrGLM) for binomial endpoint.

A second option to account for binomial data with 0 excess observations $Y_i = 0$ in one or several centers is a bias reduced glm(26). In this approach, the iteratively reweighted least

square algorithm used for fitting generalized linear models is modified by adding pseudo-observations depending on the estimated parameters, such that bias is reduced iteratively(29). This approach always leads to finite estimates of the logits m_i, and of its related variance covariance matrix $\hat{V}$, such that computation can proceed as described in earlier sections. For the computational details we refer to Kosmidis and Firth, 2009(27), and Kosmidis and Firth (2021)(29).

Non-parametric approach for multiple comparisons (Nparcomp) for continuous and ordinal endpoints.

Konietschke et al. (2012) proposed a non-parametric procedure to perform general multiple contrast tests between several samples without relying on assuming any specific distribution for the data Y_{ij}(30). Very briefly, they assume that the data are independent realizations $Y_{ij} \sim F_i$, where the F_i denote, in our context, the distributions in centers $i = 1, \ldots, I$. These distribution functions need not to be explicitly specified, they may differ between centers, including cases like heteroscedastic data, or different levels of skewedness between centers. Their procedure further allows Y_{ij} to be heavily tied data, including ordinal data, such as disease severity scores. The comparisons between centers rely on the generalized relative effects π_i, which are defined as the probability that observations from center i is lower or equal than an observation from the average distribution G resulting from the averaging F_i across all centers. Applying the above contrast matrix C allows to compute adjusted p-values for the deviations of centers from the average as well as simultaneous confidence intervals, again using multivariate-t- (or -normal-) distribution for the test statistics derived from the generalized relative effects. For full computational details we refer to Konietschke et al. (2012).

The method of Konietschke et al. (2012) is an asymptotic one, in other words the control of type-I-error for small samples is unclear. Specifically, Konietschke et al. (2012) state that convergence to normality is slow, especially for many groups (i.e. centers) and small sample sizes. Their simulation study only includes cases with $i = 3, 4, 5$ groups, and only mildly unbalanced sample sizes. Moreover, their simulation study involved only continuous data, while results for highly discrete ordinal data were not shown. In application to real data with ordinal variables, we observed simultaneous confidence intervals indicating quite clear deviations from the null hypotheses, when sample sizes

were extremely small. We therefore ran additional simulation studies specifically tailored for the applications described in this paper.

Simulation Study

A Monte Carlo simulation study was performed to assess the control of type I error (α), the probability to reject H_0 for at least one center if no center deviates from GM in which $H_0: m_i - m_\cdot = 0$, and the power $(1-\beta)$ against an alternative hypothesis $H_A: m_i - m_\cdot \neq 0$ for each method applied on its respective data type, the power represents the probability to reject H_0 for at least one center if $H_A$is true for at least one center. Both GLM and Nparcomp are valid asymptotic methods that require large sample sizes; however, we are interested in their performance under small and unbalanced sample sizes. Since GLM has computational problems when $Y_i = 0$, it is additionally compared to the alternative approaches (BayesGLM and BrGLM) under same settings. Here, power comparisons are of special interest as the three approaches handle the case of $Y_i = 0$ differently. Nparcomp is also assessed when applied to ordinal data with few categories and small sample sizes. Ordinal outcome was simulated from normally distributed data which was then round to 0 digits to create discrete ordinal data.

Simulations were run for balanced and unbalanced designs with varying parameter settings: $I = (5, 10)$ for number of centers in a trial, subjects per center varied between balanced and unbalanced scenarios of $n_{ij} = (2, 3, 4, 5, 6, 10, 20, 40, 50, 80, 100, 150, 200, \ldots., 4000)$. Complete list of parameter settings for all simulations are available in the supplementary material. The n_i in power simulations for unbalanced designs, deviating center constantly had half the number of observations as in other individual centers for covered scenarios, some additional scenarios where run for extreme small n_i in deviating center. For continuous and ordinal power simulations, the true difference between means (δ) were chosen such that for a given sample size a power of 80% is achieved in a two-sample t-test thus one center had a δ deviating from other centers. As for binomial power simulations, the success proportions of centers were chosen such that for a given sample size, a power of 80% is achieved in a two-sample proportion test, consequently, centers had a different success proportion from deviating center. For each parameter setting, a number of 1000 datasets were generated and tested by each method. Note that, with 1000 simulation runs to estimate the type I error, the standard error of an estimated

type I error is ≈ 0.007 and 95% of simulation results are expected in the range $[0.036, 0.063\ ]$ if a method accurately controls type I error at $\alpha = 0.05$.

Software and packages

All simulations were performed in R, version 4.0.5. Implemented methods Linear Model, Generalized Linear Model, Bayesian Generalized Linear Model, Bias Reduction in Binomial response Generalized Linear Models and Non-parametric multiple comparisons are available in R-packages stats v4.0.5(31) (R-core Team), arm v1.11-1(32), brglm v0.6.2(33) and nparcomp v3.0(34) respectively. To compute GM contrasts, "multcomp" package was used(35).

Results

This section shows the results of the simulation study. We describe first the results of the type I error control simulations and then the results of power simulations in comparison contrast to the GM.

Simulations of type I error:

The simulations of type I error for all methods are shown in Figure 2. For a binomial outcome, Figure 2a shows the experimental setup used for a balanced design. Simulations show as the sample size per center N increases with increasing success probability of a certain response variable, the more a 5% rejection rate is achieved. For events with a low expected number of events ($n_i\pi_i$) all three methods tend to show α below the nominal level. While for unbalanced designs, Figure 2b shows no difference between the three methods. In extreme settings however, i.e. centers having a smaller number of patients compared to other centers that have a smaller success probability of a certain response variable, the methods appear to be conservative in achieving a 5% rate. For continuous outcomes, Figure 2c and Figure 2d show the simulations of linear model as a comparator to the non-parametric approach for balanced and unbalanced experimental design respectively. As anticipated, the non-parametric method shows increased type I error for small sample sizes (3, 5, 10). A linear model is known to control the familywise type I error rate, the purpose of this comparison is to show the ability of the non-parametric method to control the type I error similarly to the linear model, specifically for extreme settings. For extreme settings such as centers with <10 patients per center, the non-

parametric method rejects up to 10% for balanced designs, however its control is sounder for unbalanced designs for all covered scenarios. For ordinal outcome, Figure 2e and Figure 2f show the simulations of the non-parametric method for balanced and unbalanced experimental designs respectively. Similarly, to the continuous outcome, for centers having <10 patients the control of type error reaches to 18% for balanced designs and is maintained for all scenarios of unbalanced designs. Additional simulations with 5000 runs were performed for the non-parametric method with the same settings, similar type I error control is observed to the 1000 runs (Supplementary figure 1).

[Figure 2 insert here]

Power simulations:

The power simulations for all methods are shown in Figure 3. For a binomial outcome, Figure 3a and Figure 3b show the experimental setups used for balanced and unbalanced designs respectively. Methods show power increase as sample size per center N and success probability increase. Furthermore, BayesGLM is superior in power for small sample sizes relative to GLM and BrGLM, while controlling the type I error. Therefore, we recommend the use of Bayesglm for binomial outcomes that might contain rare events ($Y_i = 0$). For continuous outcome, similarly, to type I error simulations linear model was chosen as comparator to the non-parametric method. Figure 3c and Figure 3d show power simulations of both methods for balanced and unbalanced scenarios respectively. Both methods achieve greater power for balanced designs than unbalanced ones. Additionally, the Non-parametric method has a trivial decrease in power compared to the linear model in all scenarios. Power rather decreases to ~50% for both methods in extreme settings of having small n_i per center. For ordinal outcome, Figure 3e and Figure 3f show the power simulations of the non-parametric method for balanced and unbalanced experimental designs respectively. For balanced designs as the n_i per center increases the power of the non-parametric method increase as well. Power decreases substantially for extreme settings of having small n_i per center in which it reaches a maximum of 35%.

[Figure 3 insert here]

Application to the GMSR dataset:

We illustrate the proposed methods (except GLM and BrGLM as they show inferiority to BayesGLM) by the analysis of GMSR data. Methods were implemented on the corresponding variable type as appropriate. Figure 4 shows simultaneous confidence intervals of the deviations of the 15 centers from GM, for continuous, binary, and ordinal outcomes.

The Non-parametric method was applied on the age at onset variable (Figure 4a), C9 shows a cohort relatively smaller than GM of other centers, whereas C12 shows a cohort larger than the GM. In both cases, it is not a foremost observation for the GMSR data as it does not have a specific inclusion criteria for age onset of patients. However, it could be imperative for clinical trials as they do have a detailed inclusion criterion. A binary variable was derived presenting a missing input of the age at onset variable for each patient (Figure 4b). C3 shows that it has 149 patients, however 41 of them do not have age onset information, although it's not uncommon to have missing data for some variables, C3 shows a higher average than the GM. C12 and C13 have a similar pattern to C3 where both have a higher mean than the GM. While C7 has only five patients with missing information out of 460 patients, it shows a smaller mean for missing information than the GM of other centers. In other words, C7 signifies a superior documentation for age onset than other centers. Another binary variable presenting adverse events (per patient) reported per center is presented in figure 4c. C1, C2, C3 and C5 show higher proportions of adverse events reported than the GM of all centers. Looking in more detail into the AEs documented, C3 reports COVID-19 vaccination reactions as adverse events. This shows a clear example of how centers could perform differently from other centers. The results could indicate the need for stakeholders to approach under reporting centers, in other words it would point the centers that have a significantly smaller average than the GM. Finally Figure 4d shows the contrasts of the centers EDSS measurements to the GM. It shows how center's cohort disease severity for the specific center is different. C4, C7 and C14 show significantly higher EDSS measurements than the GM, while C1 and C15 show a smaller one. These results may alert stakeholders to further investigate the reasons for such differences. Particularly for C7, as it includes a cohort with higher disease severity than average and yet they report fewer adverse events.

[Figure 4 insert here]

Discussion

In this paper we present methods and their implementation to detect center(s) that differ from GM of other centers for a specific variable. The utilization of these procedures serves the aim of CSM in performing data quality checks to improve data integrity. It also minimizes the costs of data monitoring and improves their quality. We were able to show how different statistical methods can be implemented to identify centers in multi-center trials or registry data that might need additional training or is a candidate for on-site monitoring visits. The approach allows the recognition of centers that are significantly deviating from the average. This would eventually enable the monitoring teams to point their attention to problematic sites.

The three methods investigated for Binomial data never strongly exceed type I error. Nevertheless, BayesGLM is superior to GLM and BrGLM in detecting a deviating center when $n_{ij} < 50$. The fact that all three methods tend to be too conservative for small sample sizes and rare events resembles similar problems found for other binomial methods: Due to the discreteness of binomial data, various methods are reported to be either over-conservative or liberal depending on the specific method and parameter configuration (36–38). The non-parametric method has harsh violations of the type I error control; especially for $n_{ij} < 10$, and ordinal data. In other words, the non-parametric method can be applied for clinical trials and registries where centers do not have a relatively small sample size, i.e., centers should have at least 10 patients to identify a true deviation. Our results show that the non-parametric method may result in an increased rate of falsely detecting deviating centers, when sample sizes are small. In some cases, an alternative would be to choose a suitable data transformation followed by application of parametric methods. However, in other cases, like contamination with outliers, bi- or multi-modal distributions, transformation may not settle the problems and non-parametric methods may still be the best choice. Further, it should be noted that the simulation studies in this paper are not suitable for fairly comparing non-parametric with parametric methods, because situations where non-parametric methods may outperform parametric approaches have not been involved.

Desmet et al. (2014, 2017) proposed alternative approaches to detect deviating centers, with differing assumptions(18,19). They assume that some variability between centers

has to be expected and is not of concern, particularly if the number of centers is large. Consequently, they focus on detecting the deviations of a small proportion of contaminated centers from the distribution of the large majority of centers. They cover the important cases of continuous data under the additional assumption of normally distributed center means in a mixed effect model(18), and of binomial data with the assumption of beta distributed variability between centers(19). Conversely, the models underlying the methods in this paper make no assumptions on the distribution of center means and are currently available for a wider range of model types and distributional assumptions for the data, including the non-parametric approach. However, this comes at the price of overfitting and possibly flagging more deviating centers than necessary in cases where variability between centers is allowed, particularly in trials involving a large number of centers. Further research is needed to investigate the approach practically for large multicenter clinical trials covering 20-100 centers with many being very low recruiters.

As Buyse et al. (2020) indicates, the power of a statistical approach lays in performing statistical tests on all variables. This would lead into many numbers of tests conducted and thus the need to combine their conclusions(39). For this reason, a scoring system for an individual center could be further developed for the assessment of the individual data type with appropriate method. Parameters of the scoring system must be individually weighted by stakeholders. Although clinical trials and registries are similar in many aspects, a robust scoring system must be adaptable to consider their differences(40,41). For example, the inclusion criteria of patients differ between both systems. The deviations found in clinical trials are relatively smaller than in registry data as the latter usually have less strict inclusion criteria. Alternatively, it is possible to assess each center for how many variables it has been flagged for and treat it as a binomial measure to finally compare the actual number of how many variables are differing from other centers (see Supplementary Table 1). In other situations, expert knowledge in the CSM team may be used to assess what level of deviations is still acceptable for what variable. The proposed methods are then a statistical tool to assess which centers are within or outside such a range of acceptable deviations for a given variable. In such situations, a method that automatically processes all variables might not be desirable.

Several straightforward extensions of the approach are available. First, in some situations, it might be known that some centers differ from others. For example, centers located at well-known university hospitals might differ in frequencies of disease severity scores from centers at smaller, local hospitals. This again might lead to differences between distributions or summary statistics of several further variables. If it is desired to account for such expected differences between centers, the comparisons to GM can be stratified by the type of center. Alternatively, variables that are known to reflect such expected differences between centers can be included as covariates into (generalized) linear models, such that the comparisons between centers are performed while accounting for the effect of the covariates. Second, there are several variable types for which comparisons to GM can be performed but are not mentioned in detail in this paper. Ordinal data like disease severity scores can be analyzed by cumulative link models(42) with centers (and possibly further covariates) as explanatory variables, such that the tendency to show higher or lower scores can be compared between centers. Additional approaches for ordinal data such as ordered categorical regression and multinomial models for nominal data are available. Time-to-event data or survival times are abundant in clinical trials, and multiple comparisons can be performed for such data, because the cox model as well as Weibull models for survival time are special cases of the framework implemented in the multcomp package(21). Moreover, skewed continuous data can be modelled in generalized linear models assuming exponential, gamma, or inverse Gaussian distribution. Several types of heteroscedasticity can be modelled by generalized least square models. Again, for these model types, comparisons to grand mean can be performed using the multcomp package. Further research is needed to assess the performance of these extensions for limited sample sizes for the investigated approach.

Currently methods are scattered between different packages in R. We provide an easy to use and interactive graphical user interface for the two methods BayesGLM & Nparcomp as two separate shinyapps, https://central-statistical-monitoring.shinyapps.io/BayesGLM-GM/ and https://central-statistical-monitoring.shinyapps.io/Nparcomp-GM/.Users can upload their datasets to compute comparisons to GM, and graphically represent simultaneous confidence intervals for contrasts of center means with GM. We plan to introduce a universal form of the methods demonstrated in a standard R package to tackle different data types easing their implementation and drawing respective decision charts for the benefit of CSM. Central

statistical monitoring serves the core purpose of monitoring goals. It facilitates the detection of deviating centers that are not likely due to chance. This would eventually support monitoring teams to initiate an onsite visit and target their activities.

Contributions

FF designed and conceptualized study, ran simulations, analyzed the data, programmed the shinyapps and wrote the original draft. DE participated in the study design, helped in the analysis of the GMSR data and revised the manuscript. NK revised the manuscript. AS participated in the study design as head of the German MS Register and revised the manuscript. GF participated in the study design, revised the manuscript. FS designed and conceptualized study, helped in drafting the manuscript and revised the manuscript.

Acknowledgments

We would like to thank Zeinab Fneish for proofreading the manuscript.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article

Supplementary

The Syntax of all simulations and respective datasets as well as for the shinyapps are available on Github under https://github.com/firasfneish/CSM.

Competing interests

FF is an employee of Leibniz University Hannover and the German MS Registry. DE had no personal financial interests to disclose other than being an employee of the German MS Registry. NF is an employee of the MSFP. Moreover, he received travel funds for research meetings from Novartis. AS has no personal financial interests to disclose, other than being the leader of the German MS Registry, which receives (project) funding from a range of public and corporate sponsors, recently including G-BA, The German MS Trust, German MS Society, The German Retirement Insurance, Biogen, Celgene (Bristol Myers Squibb), Merck, Novartis, Roche, Sanofi and Viatris. GF has nothing to disclose. FS has nothing to disclose.

References

1. Chung KC, Song JW, WRIST Study Group. A guide to organizing a multicenter clinical trial. Plast Reconstr Surg. 2010 Aug;126(2):515–23.

2. Goldberg SI, Niemierko A, Turchin A. Analysis of Data Errors in Clinical Research Databases. AMIA Annu Symp Proc. 2008;2008:242–6.

3. Barchard KA, Pace LA. Preventing human error: The impact of data entry methods on data accuracy and statistical results. Computers in Human Behavior. 2011 Sep 1;27(5):1834–9.

4. Buyse M, George SL, Evans S, Geller NL, Ranstam J, Scherrer B, et al. The role of biostatistics in the prevention, detection and treatment of fraud in clinical trials. Stat Med. 1999 Dec 30;18(24):3435–51.

5. Baigent C, Harrell FE, Buyse M, Emberson JR, Altman DG. Ensuring trial validity by data quality assurance and diversification of monitoring methods. Clin Trials. 2008;5(1):49–55.

6. Smith CT, Stocken DD, Dunn J, Cox T, Ghaneh P, Cunningham D, et al. The Value of Source Data Verification in a Cancer Clinical Trial. PLOS ONE. 2012 Dec 12;7(12):e51623.

7. EMA. Reflection paper risk based quality management clinical trials [Internet]. 2013 [cited 2023 Apr 27]. Available from: https://www.ema.europa.eu/en/documents/scientific-guideline/reflection-paper-risk-based-quality-management-clinical-trials_en.pdf

8. FDA. Oversight of Clinical Investigations — A Risk-Based Approach to Monitoring [Internet]. U.S. Food and Drug Administration. FDA; 2022 [cited 2023 Apr 27]. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/oversight-clinical-investigations-risk-based-approach-monitoring

9. Venet D, Doffagne E, Burzykowski T, Beckers F, Tellier Y, Genevois-Marlin E, et al. A statistical approach to central monitoring of data quality in clinical trials. Clin Trials. 2012 Dec;9(6):705–13.

10. Oba K. Statistical challenges for central monitoring in clinical trials: a review. Int J Clin Oncol. 2016 Feb;21(1):28–37.

11. Kirkwood AA, Cox T, Hackshaw A. Application of methods for central statistical monitoring in clinical trials. Clin Trials. 2013 Oct;10(5):783–806.

12. Fneish F, Schaarschmidt F, Fortwengel G. Improving Risk Assessment in Clinical Trials: Toward a Systematic Risk-Based Monitoring Approach. Current Therapeutic Research. 2021 Jan 1;95:100643.

13. Fneish F, Limaye D, Strüver V, Fortwengel G. Comparison of Non-Commercial Risk Based Monitoring Tools by Their Application on Clinical Trial Protocols. American Jounal of Biomedical Science & Research. 2020;8(3):216–20.

14. Timmermans C, Venet D, Burzykowski T. Data-driven risk identification in phase III clinical trials using central statistical monitoring. Int J Clin Oncol. 2016 Feb 1;21(1):38–45.

15. Kahan BC. Accounting for centre-effects in multicentre trials with a binary outcome – when, why, and how? BMC Medical Research Methodology. 2014 Feb 10;14(1):20.

16. Trotta L, Kabeya Y, Buyse M, Doffagne E, Venet D, Desmet L, et al. Detection of atypical data in multicenter clinical trials using unsupervised statistical monitoring. Clinical Trials. 2019 Oct 1;16(5):512–22.

17. Afroz MA, Schwarber G, Bhuiyan MAN. Risk-based centralized data monitoring of clinical trials at the time of COVID-19 pandemic. Contemporary Clinical Trials. 2021 May 1;104:106368.

18. Desmet L, Venet D, Doffagne E, Timmermans C, Burzykowski T, Legrand C, et al. Linear mixed-effects models for central statistical monitoring of multicenter clinical trials. Statistics in Medicine. 2014;33(30):5265–79.

19. Desmet L, Venet D, Doffagne E, Timmermans C, Legrand C, Burzykowski T, et al. Use of the Beta-Binomial Model for Central Statistical Monitoring of Multicenter Clinical Trials. Statistics in Biopharmaceutical Research. 2017 Jan 2;9(1):1–11.

20. Ohle LM, Ellenberger D, Flachenecker P, Friede T, Haas J, Hellwig K, et al. Chances and challenges of a long-term data repository in multiple sclerosis: 20th birthday of the German MS registry. Sci Rep. 2021 Jun 25;11(1):13340.

21. Hothorn T, Bretz F, Westfall P. Simultaneous inference in general parametric models. Biom J. 2008 Jun;50(3):346–63.

22. Genz A, Bretz F. Computation of Multivariate Normal and t Probabilities [Internet]. Berlin, Heidelberg: Springer; 2009 [cited 2023 Apr 27]. (Lecture Notes in Statistics; vol. 195). Available from: https://link.springer.com/10.1007/978-3-642-01689-9

23. Wasserstein RL, Lazar NA. The ASA Statement on p-Values: Context, Process, and Purpose. The American Statistician. 2016 Apr 2;70(2):129–33.

24. McCullagh P, Nelder J. Generalized Linear Models [Internet]. 2nd ed. Chapman & Hall; 1983 [cited 2023 Apr 27]. Available from: https://www.utstat.toronto.edu/~brunner/oldclass/2201s11/readings/glmbook.pdf

25. Fneish F, Schaarschmidt F. Multiple Treatment Comparisons Of Binomial Data With Only 0s In One Treatment Group. 2019 Mar 23;

26. Gelman A, Su YS, Yajima M, Hill J, Grazia Pittau M, Kerman J, et al. Data Analysis Using Regression and Multilevel/Hierarchical [Internet]. 2022 [cited 2023 Apr 27]. Available from: https://cran.r-project.org/web/packages/arm/arm.pdf

27. Kosmidis I, Firth D. Bias reduction in exponential family nonlinear models. Biometrika. 2009 Dec 1;96(4):793–804.

28. Gelman A, Jakulin A, Pittau MG, Su YS. A weakly informative default prior distribution for logistic and other regression models. Ann Appl Stat [Internet]. 2008 Dec 1 [cited 2023 Apr 27];2(4). Available from: https://projecteuclid.org/journals/annals-of-applied-statistics/volume-2/issue-4/A-weakly-informative-default-prior-distribution-for-logistic-and-other/10.1214/08-AOAS191.full

29. Kosmidis I, Firth D. Jeffreys-prior penalty, finiteness and shrinkage in binomial-response generalized linear models. Biometrika. 2021 Mar 1;108(1):71–82.

30. Konietschke F, Hothorn LA, Brunner E. Rank-based multiple test procedures and simultaneous confidence intervals. Electronic Journal of Statistics. 2012 Jan;6(none):738–59.

31. R: The R Stats Package [Internet]. [cited 2023 Apr 27]. Available from: https://stat.ethz.ch/R-manual/R-devel/library/stats/html/00Index.html

32. arm.pdf [Internet]. [cited 2023 Apr 27]. Available from: https://cran.r-project.org/web/packages/arm/arm.pdf

33. brglm function - RDocumentation [Internet]. [cited 2023 Apr 27]. Available from: https://www.rdocumentation.org/packages/brglm/versions/0.7.2/topics/brglm

34. Konietschke F, Placzek M, Schaarschmidt F, Hothorn LA. nparcomp: An R Software Package for Nonparametric Multiple Comparisons and Simultaneous Confidence Intervals. Journal of Statistical Software. 2015 Mar 20;64:1–17.

35. Hothorn T, Bretz F, Westfall P, Heiberger RM, Schuetzenmeister A, Scheibe S. multcomp: Simultaneous Inference in General Parametric Models [Internet]. 2023 [cited 2023 Apr 27]. Available from: https://cran.r-project.org/web/packages/multcomp/index.html

36. Brown LD, Cai TT, DasGupta A. Interval Estimation for a Binomial Proportion. Statistical Science. 2001 May;16(2):101–33.

37. Agresti A, Caffo B. Simple and Effective Confidence Intervals for Proportions and Differences of Proportions Result from Adding Two Successes and Two Failures. The American Statistician. 2000 Nov 1;54(4):280–8.

38. Schaarschmidt F, Sill M, Hothorn LA. Approximate Simultaneous Confidence Intervals for Multiple Contrasts of Binomial Proportions. Biometrical Journal. 2008;50(5):782–92.

39. Buyse M, Trotta L, Saad ED, Sakamoto J. Central statistical monitoring of investigator-led clinical trials in oncology. Int J Clin Oncol. 2020;25(7):1207–14.

40. Gliklich R. Clinical Trials vs Registries. 2009 Apr 1 [cited 2023 Apr 27];0(0). Available from: https://www.appliedclinicaltrialsonline.com/view/clinical-trials-vs-registries

41. Is this an Interventional Clinical Trial or Observational Study? How- and Why- it is Important to Write Protocols that Make This Distinction Clear [Internet]. WCG IRB. 2018 [cited 2023 Apr 27]. Available from: https://www.wcgirb.com/insights/is-this-an-

interventional-clinical-trial-or-observational-study-how-and-why-it-is-important-to-write-protocols-that-make-this-distinction-clear/

42. Agresti A. Categorical Data Analysis, 3rd Edition | Wiley [Internet]. Wiley; 2013 [cited 2023 Apr 27]. Available from: https://www.wiley.com/en-gb/Categorical+Data+Analysis%2C+3rd+Edition-p-9780470463635

Age of onset (Median, Quantiles)	30.25 (23.42, 38.42)
Missing Age of onset	153 (8.1%)
Gender	
Females	1356 (72%)
Males	538 (28%)
Adverse Events reported	
Number of Adverse events reported	232
Number of patients experiencing Adverse evenets	183
Disease Course (at last visit)	
RRMS	1623 (86%)
SPMS	237 (13%)
KIS	15 (0.8%)
Unknown	19 (1)

Table 1: Basic and clinical characterization of patient's part of pharmacovigilance module in the GMSR data.

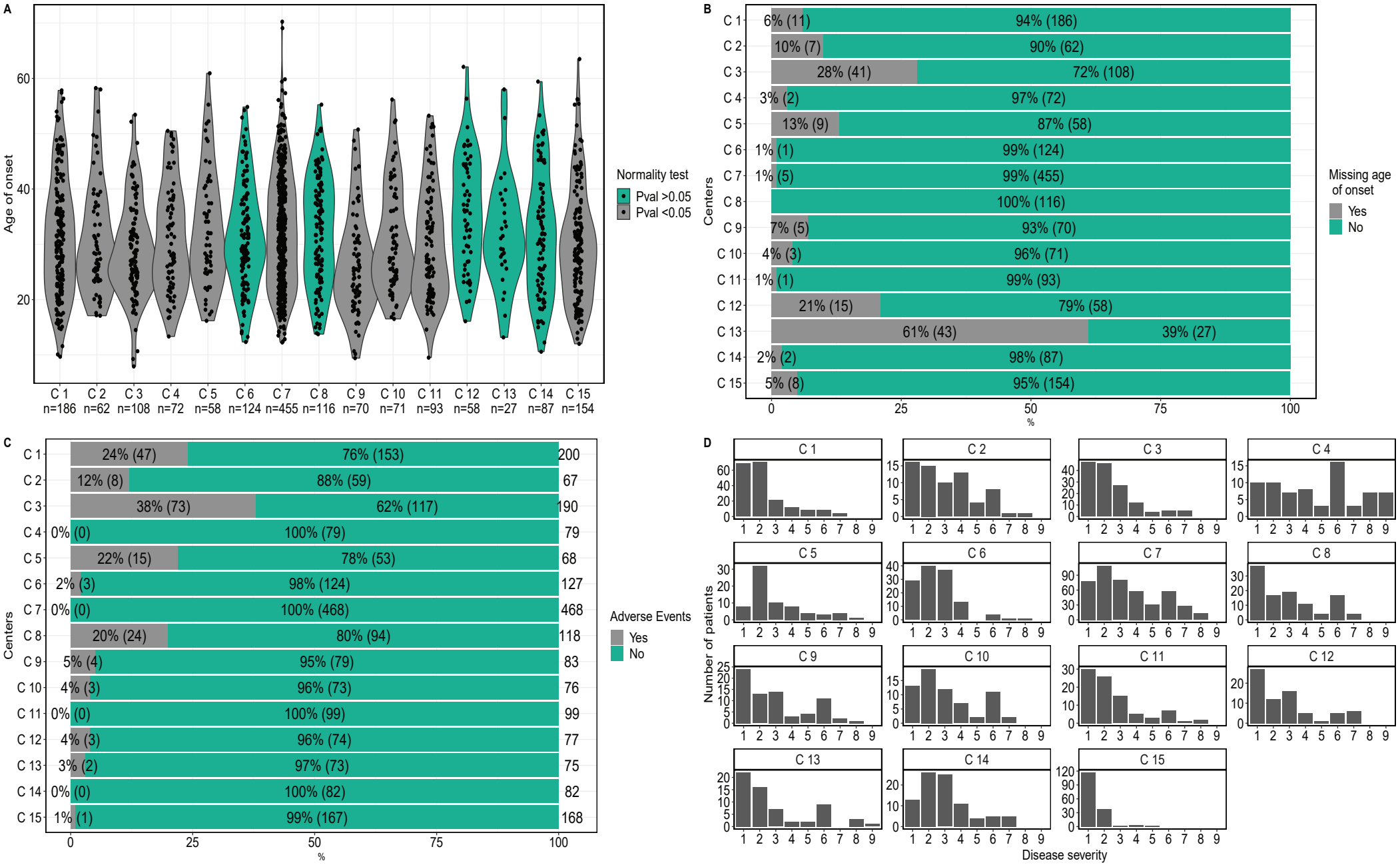


Figure 1: GMSR data stratified by center for four variables A-D. (A) Violin plot including kernel density estimates of age at disease onset indicating possible violations of normality. (B) Patients with missing age onset (%). (C) Reported number of patients having adverse events (%). (D) Histograms of disease severity (EDSS) for patients at latest visit.

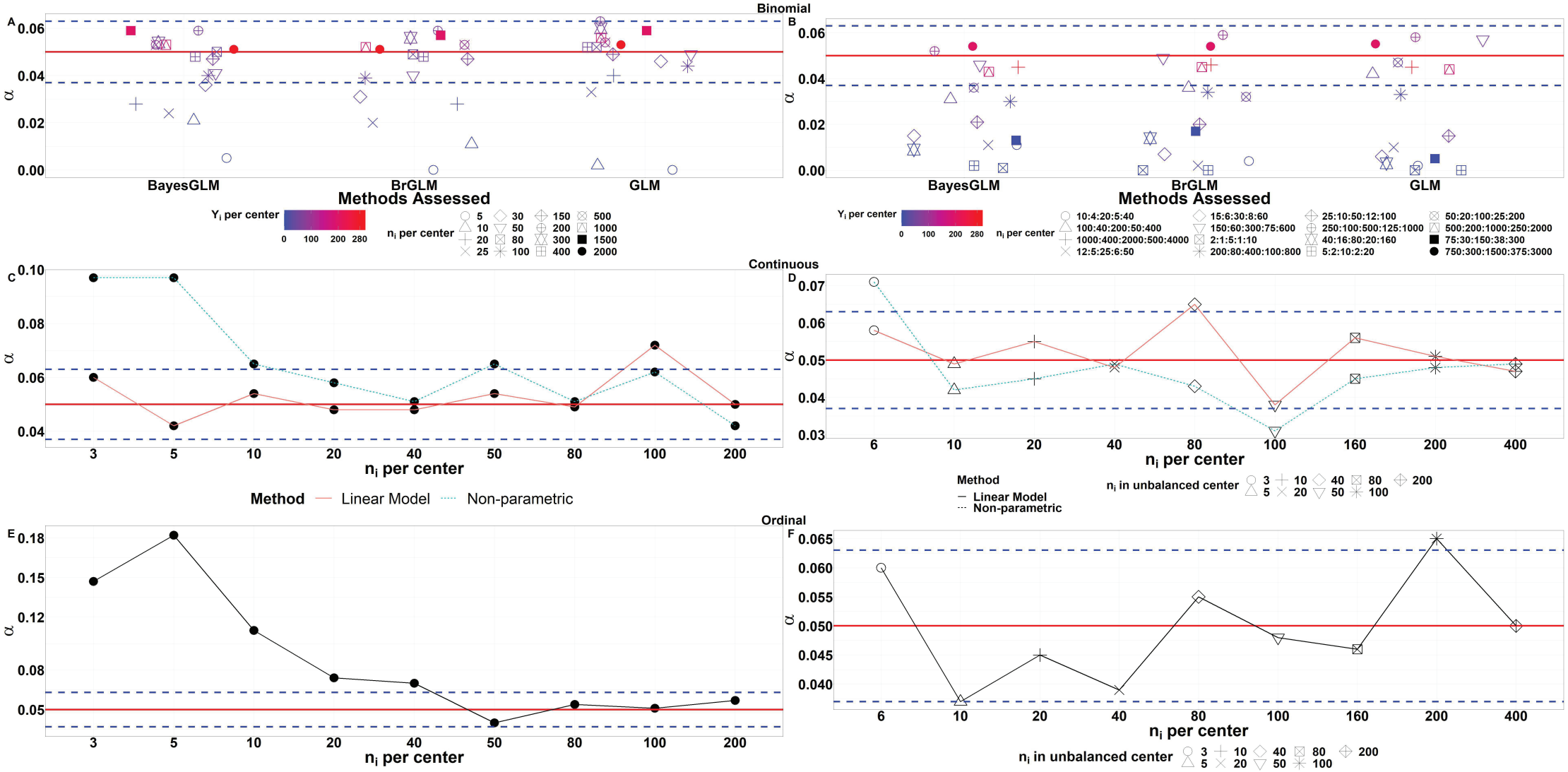


Figure 2: The probability of falsely rejecting the null hypothesis for at least one center as a function of sample size for each method applied on relevant response outcome for balanced (left panel) and unbalanced designs (right panel). The nominal type I error rate (α=0.05) is shown as a horizontal line. BayesGLM Bayesian Generalized Linear model, BrGLM Bias-reduced Generalized Linear Model, GLM Generalized linear Model.

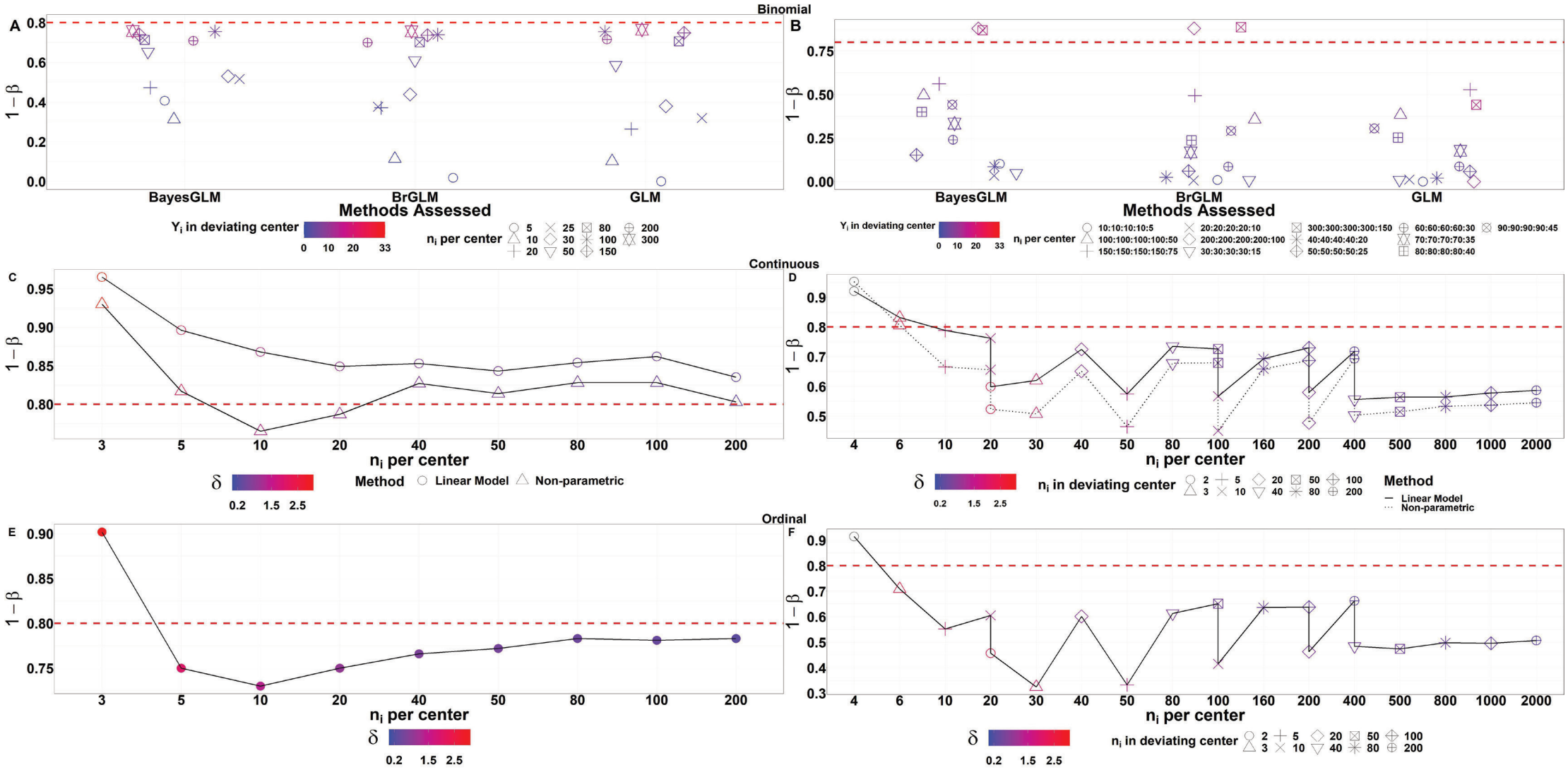


Figure 3: The probability of rejecting at least one null hypothesis (1-β) as a function of sample size for each method applied on relevant response outcome for balanced (left panel) and unbalanced designs (right panel). The group differences (δ) are chosen based on a two-sample test at a power level of 80%, which is shown as a horizontal line. BayesGLM Bayesian Generalized Linear model, BrGLM Bias-reduced Generalized Linear Model, GLM Generalized linear Model.

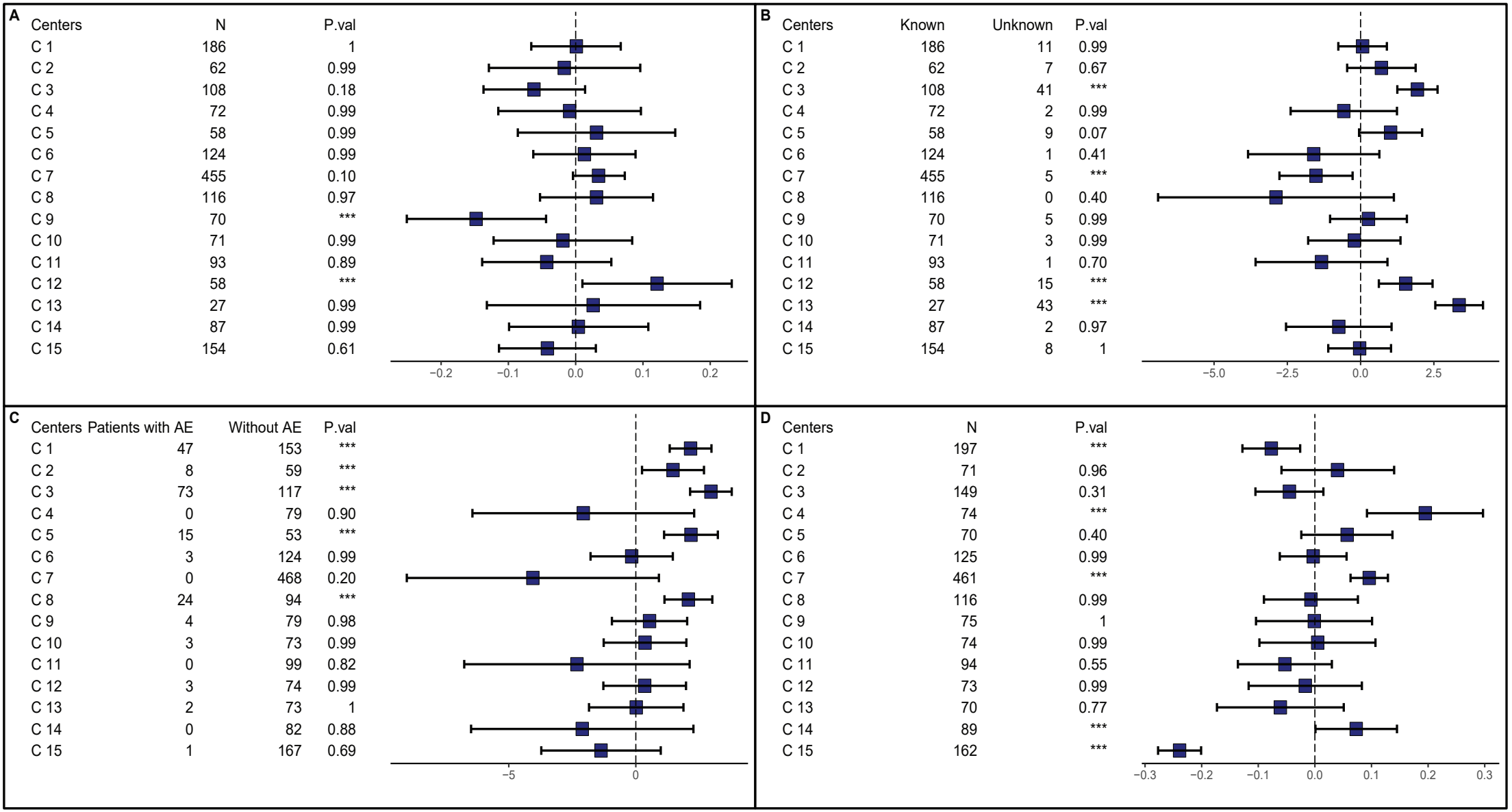
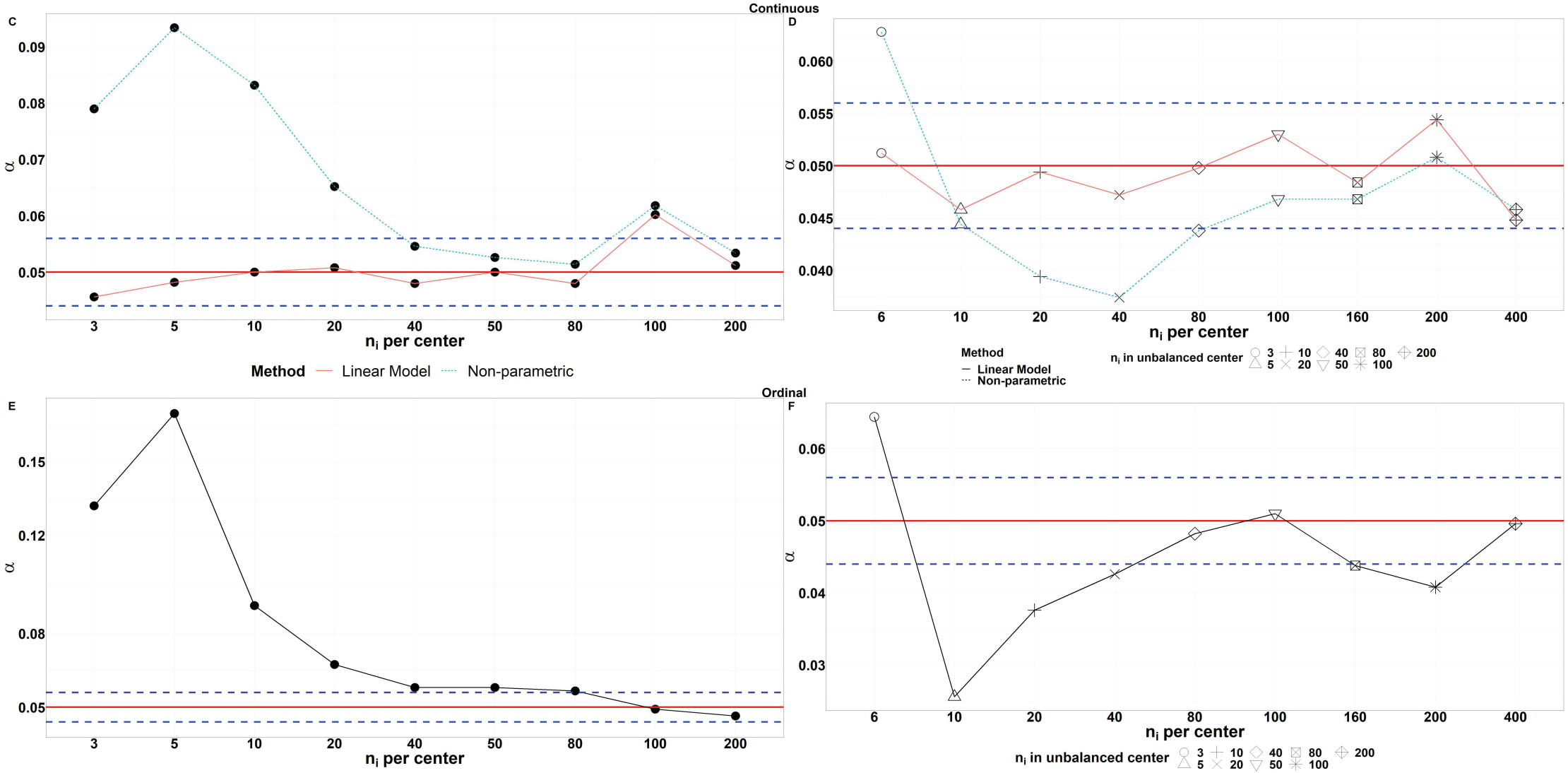


Figure 4: Simultaneous confidence intervals for contrasts of center means with GM for GMSR data set. (A) Continuous variable age onset comparison for each center to GM using non-parametric method. (B) Fitting a BayesGLM for the binary variable of missingness of the age at onset followed by contrasts of center mean towards GM. (C) Fitting a BayesGLM for AEs as binary variable followed by contrasts of center mean and GM. (D) Ordinal variable of disease severity (EDSS) comparison for each center to the GM using the non-parametric method. BayesGLM Bayesian Generalized linear model, AE Adverse events, *** significant $p < 0.05$.



Supplementary figure 1: The probability of falsely rejecting the null hypothesis for at least one center as a function of sample size for each method applied on relevant response outcome for balanced (left panel) and unbalanced designs (right panel). The nominal type I error rate (α=0.05) is shown as a horizontal line. Dotted blue lines indicate error margins for simulations with 5000 runs. Simulated type-I-errors falling outside [0.044; 0.056] indicate a significant deviation from the prespecified level alpha=0.05).

	Age onset	Age onset missing	Adverse events	Disease severity	Total variables flagged for
C1	0	0	1	1	2
C2	0	0	1	0	1
C3	0	1	1	0	2
C4	0	0	0	1	1
C5	0	0	1	0	1
C6	0	0	0	0	0
C7	0	1	0	1	2
C8	0	0	1	0	1
C9	1	0	0	0	1
C10	0	0	0	0	0
C11	0	0	0	0	0
C12	1	1	0	0	2
C13	0	1	0	0	1
C14	0	0	0	1	1
C15	0	0	0	1	1

Supplementary Table S1: Summarized table for flagged variables in each center

Chapter 3

Appendix

3.1 Curriculum Vitae

Work experience:

January 2019 – ongoing, Biostatistician at the German MS-Register, MS Forschungs- und Projektentwicklungs- gGmbH [MSFP], Hannover, Germany

Education:

Leibniz Universität Hannover
Institut für Zellbiologie und Biophysik Abt. Biostatistik
PhD student
(October 2018- April 2023)

Leibniz Universität Hannover
Master of Science
Specialty in Biostatistics
(October 2016- September 2018) (Class of 2018)

American University of Science and Technology, Beirut, Lebanon
Bachelor of Science
Degree of Clinical Laboratory Science
(October 2010- June 2015) (Class of 2015)

3.2 List of publications

1. S. C. Ibeneme, G. Fortwengel, I. J Okoye, A. D. Ezuma, W. O. Okenwa, **F. Fneish**, H. Myezwa, A. T. Ajidahun, A. O. Nwosu, N. Iloanusi, A. Nnamani, G. C. Ibeneme, I. Ulasi, P. Okere. Impact of weight-bearing on bone mineral density and sex-stratified risk factors of bone loss in HIV conditions – findings of the Nigeria HIV-BMD study: An observational study. (Under Review)

2. **F. Fneish**, D. Ellenberger, N. Frahm, A. Stahman, G. Fortwengel, F. Schaarschmidt. Application of statistical methods for central statistical monitoring and implementations on the German Multiple Sclerosis Registry (Under Review)

3. Z. Fneish, L. Becker, F. Mulenge**, F. Fneish**, B. Costa, C. T. Hoffmann, S. Gilles, U. Kalinke. Birch pollen-induced gene signatures are maintained in dendritic cells upon additional exposure to human cytomegalovirus (Under Review)

4. N. Frahm*, **F. Fneish***, M. Peters, D. Ellenberger, A. Stahmann, U. K. Zettl. SARS-CoV-2 vaccination in patients with multiple sclerosis in Germany: First results of a longitudinal observational study. (Under Review)

5. Frahm, N., **Fneish, F**., Ellenberger, D., Haas, J., Loebermann, M., Parciak, T., Peters, M., Pöhlau, D., Rodgers, J., Röper, A.-L., Schilling, S., Stahmann, A., Temmes, H., Zettl, U.K., Middleton, R.M., 2022. SARS-CoV-2 vaccination in patients with multiple sclerosis in Germany and the United Kingdom: Gender-specific results from a longitudinal observational study. Lancet Reg. Health - Eur. 22, 100502. DOI: 10.1016/j.lanepe.2022.100502 (2022)

6. N. Frahm, **F. Fneish**, D. Ellenberger, M. Peters, H. Temmes, A. Stahmann. Rolle von Komorbiditäten bei Multipler Sklerose nicht unterschätzen. NeuroRansmitter. DOI: 10.1007/s15016-022-9410-z (2022)

7. Frahm, N., **Fneish**, F., Ellenberger. D, Flachenecker, P., Friedemann P., Warnke, C., Kleinschnitz,C., Parciak, T., Krefting, D., Hellwig, K., Haas, J., Rommer, P., Stahmann, A., Zettl, U. Therapy Switches in Fingolimod-Treated Patients with Multiple Sclerosis: Long-Term Experience from the German MS Registry. *Neurol Ther* 11, 319–336 (2022). DOI: 10.1007/s40120-021-00320-w (2022)

8. V. Strüver, S. Ali, **F. Fneish**, G. Fortwengel. Patient Benefit of Clinical Research in African Developing Countries. Research in Diversely Advanced African Developing Countries. Current Therapeutic Research, Volume 96, 2022,100656, ISSN 0011-393X, DOI: 10.1016/j.curtheres.2021.100656. (2021)

9. N. Frahm, D. Ellenberger; **F. Fneish**, C. Kleinschnitz, C. Warnke, U. K. Zettl, F. Paul, B. Rauser, A. Stahmann, V. Vogelmann, P. Flachenecker. Characteristics of secondary progressive multiple sclerosis: Disease activity and provision of care in Germany – a registry-based/multicentric cohort study. Multiple Sclerosis & related Disorders. DOI: 10.1016/j.msard.2021.103281 (2021)

10. Setyawati, A., Wahyuningsih, M.S.H., Nugrahaningsih, D.A.A., Effendy, C., **Fneish, F**., Fortwengel, G., 2021. Piper crocatum Ruiz & Pav. ameliorates wound healing through p53, E-cadherin and SOD1 pathways on wounded hyperglycemia fibroblasts. Saudi J. Biol. Sci. 28, 7257–7268. DOI: 10.1016/j.sjbs.2021.08.039 (2021)

11. **F. Fneish**, F. Schaarschmidt, G. Fortwengel. Toward A Systematic Risk Based Monitoring Approach to Improve Assessment of Risks in Clinical Trials. Current Therapeutic Research, 2021. DOI: 10.1016/j.curtheres.2021.100643 (2021)

12. V. Strüver, **F. Fneish**, R. Muche, G. Fortwengel, The Temporal Development of Clinical Research in Emerging Countries. American Journal of Clinical and Experimental Medicine. Vol. 9, No. 1, 2021, pp. 1-7. DOI: 10.11648/j.ajcem.20210901.11 (2021)

13. **F. Fneish**, D. Limaye, V. Strüver, F. Sschaarschmidt, G. Fortwengel. Comparison of Non-Commercial Risk Based Monitoring Tools by Their Application on Clinical Trial Protocols. American Journal of Biomedical Science & Research. 8(3). AJBSR.MS.ID.001276. DOI:10.34297/AJBSR.2020.08.001276 (2020)

14. D. Ellenberger, P. Flachenecker, **F. Fneish**, N. Frahm, K. Hellwig, F. Paul, A. Stahmann, C. Warnke, P. S. Rommer, U. K Zettl, Aggressive Multiple Sclerosis: A Matter of Measurement And Timing. Brain, Volume 143, Issue 11, November 2020, Page e97, DOI:10.1093/brain/awaa306 (2020)

15. P. Rommer, K. Berger, D. Ellenberger, **F. Fneish**, A. Simbrich, A. Stahmann, U. Zettl. Management of treatment aspects in MS patients treated with daclizumab - A case series of 267 patients, Frontiers in Neurology, 11:996, DOI: 10.3389/fneur.2020.00996 (2020)

16. Salter, A. Stahmann, D. Ellenberger, **F. Fneish**, WJ. Rodgers, R. Middleton, R. Nicholas, RA. Marrie. Data harmonization for collaborative research among MS registries: A case study in employment. Multiple Sclerosis Journal. DOI: 10.1177/1352458520910499 (2020)

17. P. Mielzcarek, A. Streichhardt, D. Limaye, V. Limaye, **F. Fneish**, G. Fortwengel. Standard of Care and Transparency in Clinical Trials Conducted in Developing Countries of Africa. Central African Journal of Public Health. Vol. 5, No. 2, 2019, pp. 92-97. DOI:10.11648/j.cajph.20190502.15 (2019)

3.3 Conference Contribution

3.3.1 Oral presentations

1. **F. Fneish**, D. Ellenberger, N. Frahm, A. Stahmann, G. Fortwengel, F. Schaarschmidt. Central Statistical Monitoring approach and implementations on the German Multiple Sclerosis Registry. Deutsche Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie (*GMDS*)- Technology, Methods, and Infrastructure for Networked Medical Research (TMF). (August 2022)

2. J. Haas, D. Ellenberger**, F. Fneish**, A.L. Röpper, U. Zettl, A. Stahmann. Verlauf der Multiplen Sklerose bei Beginn im Kindsalter-Analyse des MS Registers der DMSG, Bundesverband e.V. Deutschland. Deutsche Gesellschaft für Neurologie (DGN). (September 2019)

3. A. Stahmann, D. Ellenberger, P. Flachenecker, **F. Fneish**, J. Haas, C. Kleinschnitz. D. Pöhlau, O. Reinhoff, P. Rommer, L.M. Speikerkötter, U. Zettl. Unterschiede in der Zeit bis zur ersten Krankheitsmodifizierenden Therapie bei MS-Erkrankten mit schubförmigem Verlauf in Deutschland. Deutsche Gesellschaft für Neurologie (DGN). (September 2019)

3.3.2 Poster Presentation

1. M. Peters, N. Frahm, **F. Fneish**, J. Haas, M. Loebermann[e], D. Pöhlau[d], A. L. Röper, S. Schilling, A. Stahmann, H. Temmes, F. Paul, U. K. Zettl[c], D. Ellenberger. Boosterimpfung nach der Grundimmunisierung gegen SARS-CoV-2 bei Menschen mit Multipler Sklerose: Erkenntnisse aus einer prospektiven, longitudinalen Beobachtungsstudie. Deutsche Gesellschaft für Neurologie (DGN). (November 2022)

2. N. Frahm, **F. Fneish**, M. Peters, D. Ellenberger, J. Haas, M. Loebermann, D. Pöhlau, A. L. Röper, S. Schilling, A. Stahmann, H. Temmes[4], F. Paul[6], U. K. Zettl. Relapse activity and disability progression in COVID-vaccinated patients with multiple sclerosis: Homologous vs. heterologous vaccination scheme. European Committee for Treatment and Research in Multiple Sclerosis. (ECTRIMS). (October 2022)

3. **F. Fneish**, N. Frahm, D. Ellenberger, P. Flachenecker, T. Friede, C. Kleinschnitz, D. Krefting, U. Zettl, A. Stahmann. Treatment patterns of patients with aggressive MS in a cohort of early retirees. European Committee for Treatment and Research in Multiple Sclerosis. (ECTRIMS). (October 2021)

4. **F. Fneish**, D. Ellenberger, N. Frahm, A. Stahmann, P. Rommer, U.K. Zettl. Disease modifying therapies in patients with aggressive MS. European Committee for Treatment and Research in Multiple Sclerosis. (ECTRIMS). (October 2021)

5. **F. Fneish**, K. Berger, P. Flachenecker, T. Friede, J. Haas, K. Hellwig, C. Kleinschnitz, F. Paul, D. Pöhlau, O. Rienhoff, P.S. Rommer, C. Warnke, A. Stahmann, U.K. Zettl, D.

Ellenberger. The Story Of Aggressive Multiple Sclerosis. Deutsche Gesellschaft für Neurologie (DGN). (November 2020)

6. **F. Fneish**, F. Schaarschmidt. Treatment comparisons of Binomial data with only 0s in one treatment group at the fifth conference of the Deutsche Arbeitsgemeinschaft Statistik (DAGStat). (March 2019)

7. A. Stahmann*, **F. Fneish***, K. Eichstädt, P. Flachenecker, T. Friede, J. Haas, C. Kleinschnitz, D. Pöhlau, O. Rienhoff, P.S. Rommer, U.K. Zettl. Disease Modifying Treatments (DMTs) in Germany –Changes in treatment patterns. European Committee for Treatment and Research in Multiple Sclerosis. (ECTRIMS). (September 2018)

8. **F. Fneish**. Pharma Ethics Practice in Lebanon at the 12th International Conference on Bioethics, Medical Ethics, & Health Law in Limassol, Cyprus. (March 2017)

9. **F. Fneish**. Adult vaccination status in undergraduate students in Ashrafieh, Lebanon. Society of Clinical Research Associates (SOCRA) 25th Annual Conference, Palais des congres de Montreal, Montreal, Quebec, Canada. (September 2016)

3.4 Acknowledgments

My Colleagues

I would like to thank **Frank Schaarschmidt** for his continuous support during the past years. As a Biostatistician, I have built up my knowledge through his guidance and knowledge. I appreciate the patience he had with me. His door was always open to any question and struggle I had. I have really enjoyed our conversations about everything from politics to cultural differences and religion. I doubt there isn't a topic we haven't discussed. To **Max Menssen** and **Soren Büdig**, thank you for the nice conversations and laughs we shared the past years. To **Clemens Bucilowski** thank you for helping me with all the technical support at the institute.

To **Gerhard Fortwengel**, I honestly do not know how to thank you enough, not only for the guidance through the doctorate research, but since the first day I stepped into Germany. You have helped me in every step along the way. I would have not been here without your continuous support all these years. Thank you for giving me the opportunity to achieve more and believing in me. I hope I can live up to your expectations. To **Dnyanesh Limaye**, thank you for support and help through the years, our friendship has been amazing. I wish you and your family all the best back home.

To my colleagues at the German Multiple Sclerosis Registry, I thank all of you for continuous support through the past years and for your patience with my German speaking skills. I thank **Alexander Stahmann** for his support and giving me the opportunity to gain practical experience in the field. I also thank **David Ellenberger** and **Niklas Frahm** for their support in my research and the nice conversations we have.

My Family

To my lovely wife **Zeinab Fneish**, this work would have not been possible without you being by my side. Your presence gives me the motivation and strength to achieve more and aim higher. I enjoyed every second of writing our theses together. Thank you for everything!

To my parents **Nasser Fneish** and **Iman Fneish**, thank you for your sacrifices you have made to secure us a proper education and a good life. Without your sacrifices this would have not been possible. To my siblings, **Sandra Fneish** and **Hussein Fneish** thank you for the fun moments we always have together.

To my parents in law, **Wajih Kanso** and **Ahlam Kanso**, thank you for your continuous support, I appreciate all the support and the best time we have together.