Paper IC09

Codeless Queries Harnessing AI to Translate Clinical Trial Questions into Custom Listings Through Intuitive UI Interactions

Aditya Gadiko, Saama, Raleigh, US

ABSTRACT

The traditional methods of analyzing clinical trials data, often involving complex coding processes, have long been a bottleneck in deriving insights efficiently. This paper introduces a groundbreaking AI-driven solution that revolutionizes the approach to clinical trials data interpretation. By leveraging a human-in-the-loop approach, our innovative UI permits 80% code completion without direct coding. Subject matter experts can articulate questions in English, which are then translated into UI blocks. These blocks capture the essence of the query, and if machine comprehension falters, experts can iteratively refine the machine's understanding. The final outcome: tailored custom listings generated with precision. This method accelerates data analysis and fosters a more intuitive interaction with data, particularly pertinent for those with a clinical trial statistical programming background. The case studies spotlight the efficacy and the transformative potential of this methodology in real-world scenarios. Dive in to witness the future of clinical trials data analysis.

INTRODUCTION

Current methods in clinical trial data analysis are cumbersome, requiring in-depth knowledge of programming languages (SAS, SQL, R, Python) and an understanding of clinical trial data. This necessitates coordination across multiple functions such as Biostatistics, Programming, Pharmacovigilance, and Data Management, increasing the dependency on specialized resources. The process is resource-intensive and time-consuming, often requiring project management to handle logistics. Clinicians and non-technical staff find it challenging to access and interpret data without relying on technical experts, leading to inefficiencies and potential delays in decision-making, impacting patient safety.

Necessity for Innovation in Data Interpretation

The need for innovation in clinical data interpretation is multifaceted:

1. Complexity of Traditional Methods: Traditional approaches, involving manual coding, slow down analysis and increase error potential (Zhang, Chen & Zhang, 2019).
2. Volume of Data: The exponential growth of diverse data types in clinical trials demands more efficient management and interpretation methods (Chitungo, 2021).
3. Need for Precision and Speed: Accurate and rapid data interpretation directly impacts patient safety and treatment efficacy (Patel, Shah & Deshpande, 2017).
4. Advancements in AI and Machine Learning: AI and machine learning offer solutions to overcome traditional limitations, enabling faster, more accurate data processing(Liu, Wang & Wang, 2020) .
5. Interactive and Intuitive Data Analysis: AI-driven solutions allow SMEs to interact with data more intuitively.
6. Iterative Refinement and SME Involvement: Ensuring AI interpretations align with expert understanding enhances output validity.
7. Real-world Application and Validation: Case studies demonstrate the practical value of these innovative approaches.

The escalating complexity and diversity of data in clinical trials necessitate innovation in data interpretation methods. Traditional approaches, often reliant on manual coding, struggle to manage and interpret this vast, varied data efficiently. AI and machine learning technologies present an opportunity to surmount these challenges, offering advanced tools for efficient data management and interpretation. (Gadiko & Shroff, 2022)

AI-driven solutions, like 'Smart Filters', offer transformative potential. Traditional methods, hindered by manual coding and complexity, are inadequate for the growing volume of clinical data. Smart Filters enable the translation of natural language queries into actionable data commands. The capabilities of AI in enhancing data processing efficiency and accuracy are well-articulated (Gadiko & Shroff, 2022), underscoring the substantial impact of these technologies in clinical research. The iterative refinement process with SMEs enhances the reliability of analysis, democratizing data analysis and indicating significant potential in drug development and medical research.

Traditional methods in clinical trials rely heavily on manual processes and statistical techniques (e.g., t-tests, chi-squared tests, ANOVA) for data interpretation. Tools like Electronic Data Capture (EDC) systems face challenges in handling complex datasets. Conventional methods, while critical for determining statistical robustness, are time-consuming, resource-intensive, and error-prone, struggling with data volume and diversity (Fischbach, Kohli & Beam,

2010). A gradual shift towards digital and automated solutions is evident, though aligning with regulatory standards remains a challenge.

Current data analysis methods in clinical trials are challenged by their time-intensive nature, complexity in handling diverse data types, and susceptibility to human error (Bates, 2014). Resource intensity, data standardization issues, and scalability problems are prominent. Regulatory compliance, limited predictive analytics, accessibility barriers, and adaptability to rapid changes are significant concerns, underscoring the need for more adaptable methods like AI-driven solutions.

Early attempts at integrating AI in clinical trial data analysis involved basic machine learning models for pattern recognition and predictive analytics. NLP technologies processed unstructured data (Jiang, Chen & Zhang, 2022) and AI was used for image analysis in certain medical fields (Ker, Chen & Qi, 2017). AI algorithms improved patient recruitment and real-time data monitoring. Challenges included data privacy concerns, model interpretability, and regulatory hurdles. Collaborative efforts between tech companies and research institutions have fostered more sophisticated AI solutions in clinical trials.

Addressing the challenges in traditional data analysis, this paper presents an AI-driven solution leveraging semantic parsing technology. Users can input queries in natural language, which the AI model translates into logical, machine-comprehensible 'Smart Filters'. The process involves autocomplete based on current data and medical dictionaries, entity recognition, categorization of data types, and mapping to data models like CDISC SDTM standards. These processes are in line with the methodologies described in (Gadiko & Shroff, 2022), demonstrating the practical application of AI in enhancing clinical trial data analysis.

Defining the Role of SME in the AI Validation Process

SMEs, including Medical Monitors and Safety Physicians, play a crucial role in refining AI interpretations. They interact with Smart Filters through a user-friendly interface, allowing them to modify queries, add conditions, switch functions, and create nested logic without needing programming knowledge. This process ensures that the AI's outputs are accurate and contextually relevant, facilitating efficient decision-making and enhancing patient safety.

Proposed Solution with Sample UI

Subject Matter Experts (SMEs), encompassing roles such as Medical Monitors, Safety Physicians, Clinical Scientists, and Medical Reviewers, are integral to the AI validation process. Typically, these experts lack programming expertise and depend on programmers for data analysis. This study utilizes an AI model to bridge this gap. For instance, an SME's query, 'medical history schizophrenia,' is transformed by the AI into actionable 'Smart Filters.' This transformation involves mapping the query to specific data model elements, such as 'MHTERM' (a column in the Medical History table) and 'SCHIZOPHRENIA' (a categorical value within MHTERM). The AI model facilitates multiple avenues for SMEs to engage in iterative refinement, significantly enhancing the query's precision and relevance.

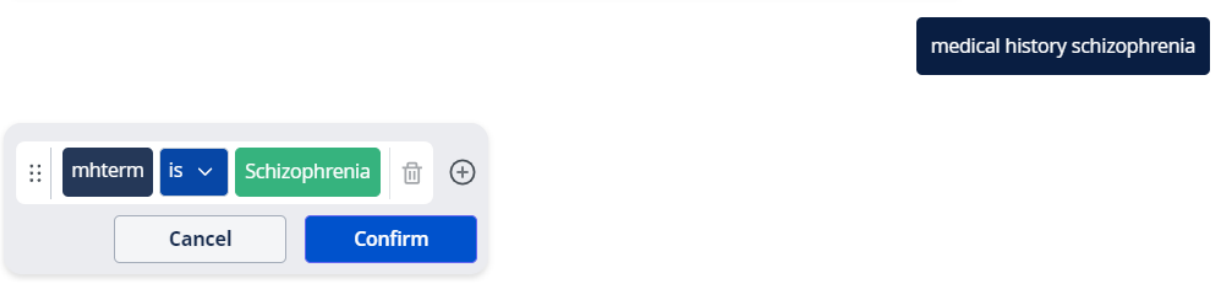


FIGURE 1

Should an SME wish to include an additional condition, such as 'Paraphrenia,' the process is streamlined and user-friendly. By simply clicking the edit icon on the 'SCHIZOPHRENIA' Smart Filter, a pop-up window displays the most current data associated with the 'MHTERM' column. Here, if 'Paraphrenia' is listed in the database, the SME can select it directly. In instances where 'Paraphrenia' is not pre-listed, the SME has the option to add it as a new category. Consequently, the Smart Filters are updated to reflect 'MHTERM IS SCHIZOPHRENIA PARAPHRENIA.' This update triggers the generation of a table or listing, compiling all patient records matching the criteria 'MHTERM IS SCHIZOPHRENIA PARAPHRENIA.

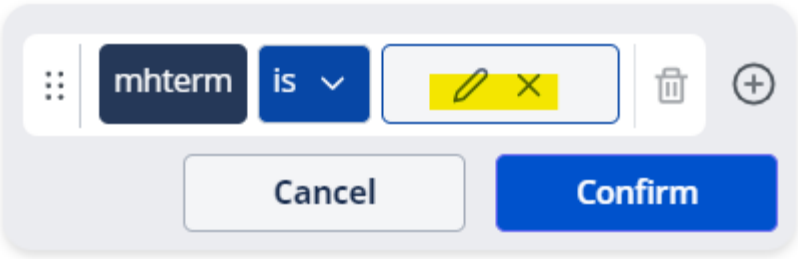
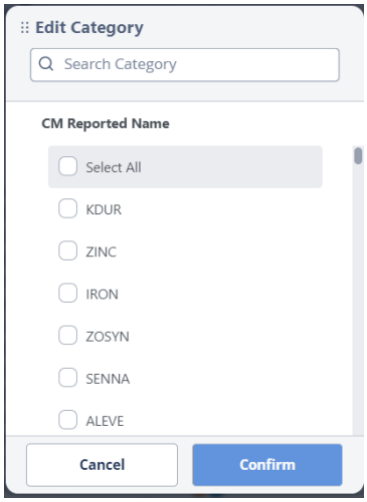


FIGURE 2

SMEs can enhance the Smart Filters by adding supplementary logic, even without any programming knowledge. For instance, to append further conditions to 'MHTERM IS SCHIZOPHRENIA' SMEs have the option to either phrase a follow-up question in natural language or utilize the plus icon to access a pop-up menu displaying all available data tables. Upon selecting a table, such as 'Concomitant Medication,' SMEs can then choose a specific column – say, 'Concomitant Medication Term' – and select or add terms within that column. If the terms 'ZYPREXA' and 'BACLOFEN' are chosen, the Smart Filters would update to 'MHTERM IS SCHIZOPHRENIA PARAPHRENIA AND CMTERM IS ZYPREXA BACLOFEN' resulting in a comprehensive listing of patients meeting these combined criteria.



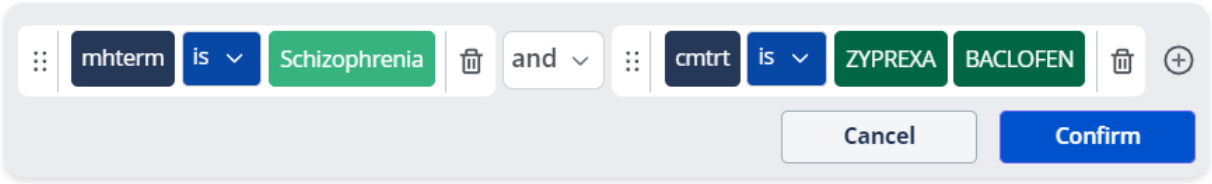


FIGURE 3

Furthermore, SMEs can modify the logical operators within the Smart Filters for broader data retrieval. By altering the 'AND' operator to 'OR' in the Smart Filter, the system will include patients meeting either of the specified conditions. Consequently, the updated Smart Filter 'MHTERM IS SCHIZOPHRENIA PARAPHRENIA OR CMTERM IS ZYPREXA BACLOFEN' broadens the scope of the query to encompass a larger patient set from the clinical trial database.

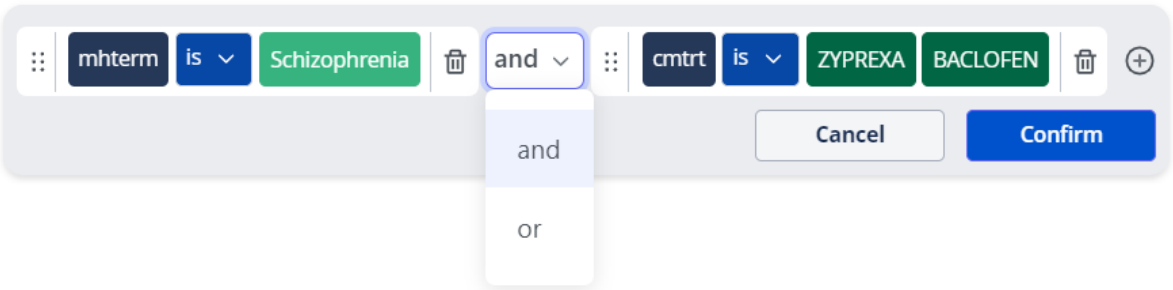


FIGURE 4

Similarly to the flexibility in changing logical operators from 'AND' to 'OR', SMEs possess the capability to toggle between 'IS' and 'IS NOT' within the Smart Filters. For example, modifying the conditions to 'MHTERM IS SCHIZOPHRENIA PARAPHRENIA OR CMTERM IS NOT ZYPREXA BACLOFEN' enables the system to filter out patients who either have a diagnosis of schizophrenia or are not on the specified concomitant medications, Zyprexa or Baclofen.

Beyond categorical or character-based data, the system also adeptly handles numeric logic. SMEs can introduce numeric criteria either directly through a natural language query, such as asking for 'medical history schizophrenia and age greater than 70 years,' or by selecting a specific numeric column through the user interface. The system offers a variety of numeric operations, including 'Age is greater than 70', 'Age is less than 70', and 'Age is between 70 and 80', among others. As a result, Smart Filters could be configured as 'MHTERM IS SCHIZOPHRENIA PARAPHRENIA OR CMTERM IS ZYPREXA BACLOFEN AND AGE IS GREATER THAN 70'

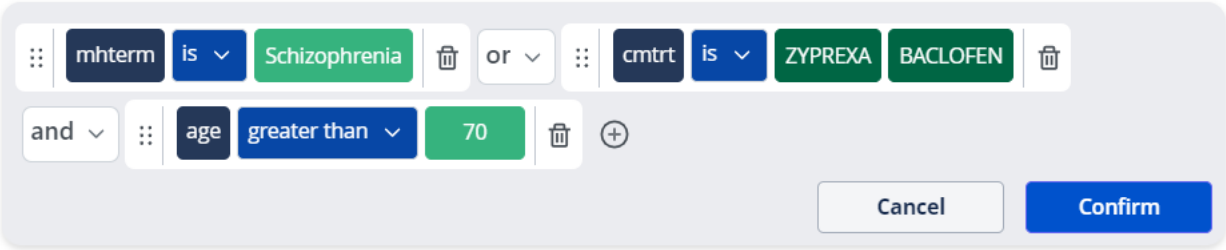


FIGURE 5

Furthermore, the system's versatility extends to handling date formats. An SME can incorporate dates into their query, for instance, 'medical history schizophrenia age is more than 70 years with study drug exposure start date after July 6th 2023.' Options for date logic include a range of operators like 'EXSTDC IS greater than July 6th 2023' or 'EXSTDC IS between July 6th 2023 and October 10th 2023.' This functionality enables complex query formulations such as 'MHTERM IS SCHIZOPHRENIA PARAPHRENIA OR CMTERM IS ZYPREXA BACLOFEN AND AGE IS GREATER THAN 70 AND EXSTDC IS GREATER THAN July 6th 2023.'

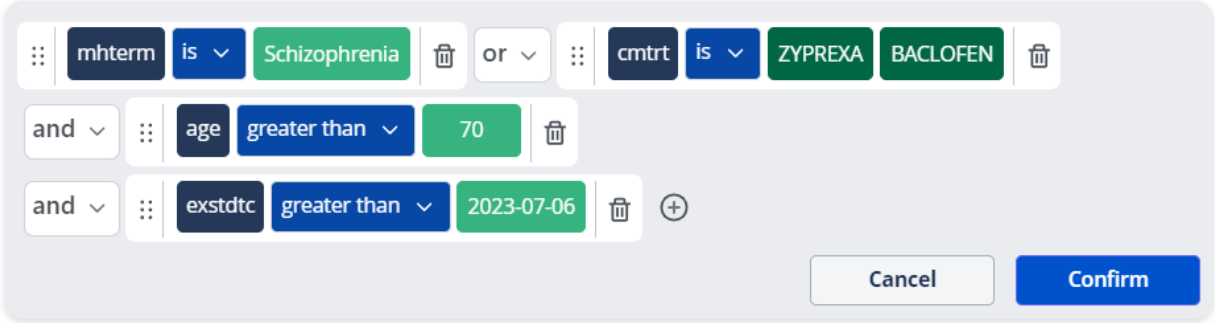


FIGURE 6

A particularly innovative feature for SMEs is the capability to construct grouped or nested logic within Smart Filters, accomplished through a simple drag-and-drop action. This user-friendly functionality obviates the need for complex coding. For instance, SMEs can group 'AGE IS GREATER THAN 70' with 'MHTERM IS SCHIZOPHRENIA PARAPHRENIA,' leading to a logical grouping like '(MHTERM IS SCHIZOPHRENIA AND AGE IS GREATER THAN 70) OR (CMTERM IS ZYPREXA BACLOFEN AND EXSTDC IS GREATER THAN July 6th 2023).

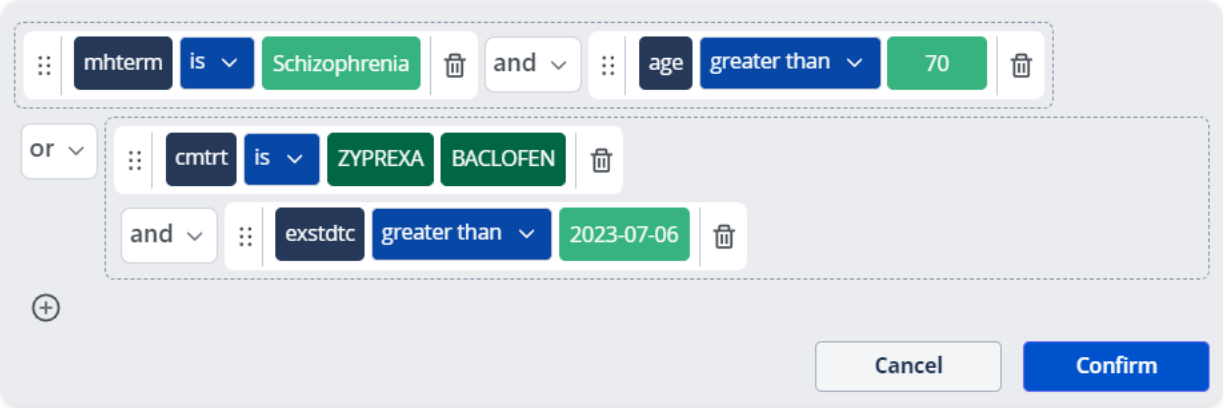


FIGURE 7

The selected case study focuses on real-world scenarios where Medical Monitors, Safety Physicians, and Clinical Scientists require quick access to custom data listings for ongoing clinical trial monitoring. These listings are crucial for identifying trends, reviewing protocol deviations, and addressing patient safety concerns. Traditional methods for generating these listings involve complex data extraction and manipulation, often requiring weeks of a coordinated effort between clinical SMEs and programmers.

Let us use the same example as a case study, the AI-driven system was employed to address a specific query: "(MHTERM IS SCHIZOPHRENIA AND AGE IS GREATER THAN 70) OR (CMTERM IS ZYPREXA BACLOFEN AND EXSTDC IS GREATER THAN July 6th, 2023)" This scenario exemplifies the need for rapid data synthesis across various domains, including medical history, medication records, demographics, and study timelines.

Using the AI-powered system, the query was processed in natural language, and the required listing was generated within seconds, a significant reduction from the traditional three-week timeframe. This efficiency was not only in data retrieval but also in validation, as the AI system provided SQL code for further verification in addition to the Clinical SME to validate the logic via Smart Filters.

The proposed solution promotes equity and inclusion by enabling non-technical experts in clinical trials to directly query data using natural language, thus democratizing data analysis, fostering cross-disciplinary collaboration, and ensuring diverse, rapid, and informed decision-making that enhances patient safety and trial outcomes.

DISCUSSION

The outcomes from the case studies underscore the transformative potential of AI in clinical trial data analysis. The ability of the AI system to quickly process complex queries and generate validated listings marks a significant departure from traditional, time-intensive data analysis methods.

This paradigm shift towards AI-driven data analysis hints at a future where clinical trials are more adaptive and responsive (Wickramasinghe & Wishart, 2020). The speed and accuracy offered by AI can lead to more timely interventions, enhanced patient safety, and overall more efficient Clinical trial management (Topol, 2019). This approach can be particularly beneficial in large-scale or multi-site trials where data complexity and volume are substantial.

Despite its advantages, the current iteration of the AI model has limitations. It does not yet support functions like count operations or complex statistical analyses, which are critical for comprehensive data interpretation. These functions are slated for inclusion in the next version of the system. Additionally, challenges in integrating AI into existing clinical trial infrastructures, ensuring data privacy and security, and maintaining regulatory compliance must be addressed as the technology evolves.

CONCLUSION

This paper presented a groundbreaking AI-driven methodology for clinical trial data analysis, which fundamentally transforms the traditional, code-intensive approach. Key to this innovation is the ability of Subject Matter Experts (SMEs) to directly input queries in natural language. These queries are processed into 'Smart Filters' capturing the core intent, and facilitating rapid and accurate data interpretation. The system's iterative refinement process, involving SME validation, ensures that AI interpretations are aligned with the expert's perspective, enhancing the reliability of outputs. Our case studies highlighted the system's efficiency, demonstrating its capability to generate listings within hours compared to the weeks typically required, significantly accelerating the decision-making process in clinical trials.

The implementation of this AI-driven method marks a significant leap in the field of clinical trial data analysis. By bridging the gap between clinical knowledge and technical data processing, the method enables a more intuitive and interactive engagement with data. This advancement not only speeds up the analysis process but also enhances the precision and dependability of the outcomes, which are crucial for patient safety and effective treatment evaluation in clinical trials. The AI system's capacity to handle complex queries across various data domains exemplifies its potential to transform data analysis in clinical research.

Looking ahead, there are several avenues for further research and development. The current AI model, while robust, is yet to incorporate capabilities for executing count operations or conducting detailed statistical analyses. Enhancing the model to include these functionalities will be a focus of future iterations. Additionally, as AI integration in clinical trials continues to evolve, addressing challenges related to data privacy, regulatory compliance, and the seamless integration of AI systems into existing clinical trial infrastructures will be crucial. This continuous evolution aims to foster a more adaptable, responsive, and efficient approach to clinical trial data analysis, paving the way for advancements in medical research and patient care. We are open to collaboration and happy to share our published research for industry-wide collaboration.

REFERENCES

Q. Zhang, Y. Chen, Y. Zhang, and T. Wang, "An Automated Coding Algorithm for Clinical Trial Data Based on Deep Learning," *IEEE Access*, vol. 7, pp. 18039-18048, 2019

Chitungo, I. (2021). The data explosion in healthcare: What you need to know. *Journal of the American Medical Informatics Association*, 28(2), 351-353

P. R. Patel, P. B. Shah, and P. B. Deshpande, "Precision and Speed in Clinical Data Interpretation: A Review of Automated Methods," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 6, pp. 1468-1476, 2017

H. Liu, Q. Wang, and X. Wang, "A Comparative Study of Manual and Automated Coding for Clinical Trial Data Analysis," *Proceedings of the ACM International Conference on Knowledge Discovery and Data Mining*, pp. 185-194, 2020.

A. Gadiko and N. Shroff, "A Deep Learning-based Virtual Assistant for Better and Faster Clinical Insights," *PhUSE US Connect*, ET01, Year 2022.

M. A. Fischbach, L. H. Kohli, and A. L. Beam, "The Need for Speed and Accuracy in Clinical Decision Support Systems," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 5, pp. 1108-1117, 2010.

Bates, D. W. (2014). The Importance of Accuracy and Speed in Clinical Data Interpretation. IEEE Journal of Biomedical and Health Informatics, 18(2), 395-397.

Jiang, S., Chen, Q., & Zhang, Y. (2022). A review of natural language processing for clinical data analysis. *IEEE Transactions on Biomedical Engineering*, 69(6), 2032-2044.

Ker, J., Chen, L., & Qi, J. (2017). Deep learning for medical image analysis. *IEEE Access*, 6, 7086-7108.

Wickramasinghe, M. V., & Wishart, D. S. (2020). Artificial Intelligence and Machine Learning in Clinical Trial Data Science: Opportunities and Challenges. IEEE Journal of Biomedical and Health Informatics, 24(3), 652-663.

Topol, E. J. (2019). The Future of Clinical Trials: Adaptive, Personalized, and Accelerated Designs Enabled by Big Data and AI. Annual Review of Medicine, 70, 1-13.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Author Name: Aditya Gadiko
Company: Saama
Address: 900 East Hamilton Avenue, CA
Phone: 843 345 4379
Email: Aditya.Gadiko@saama.com
Website: www.Saama.com