

Enhanced Signal Detection and Management Enables More Effective Pharmacovigilance

Introduction: Early Detection of Adverse Drug Reactions Essential

Thanks to growing government and public concern, intense media scrutiny, and increased access to public registries, drug safety has become a zero-tolerance issue. With each introduction of a potent new drug, pharmaceutical firms must detect adverse reactions quickly in order to ensure patient safety, minimize costs, protect their brands and ease regulatory compliance.

Early-detection efforts are hindered, however, by a lack of integration among diverse information sources, such as spontaneous reporting databases, prescription event monitoring databases, linked administrative databases and electronic medical records. Manual processes are another roadblock. Researchers can no longer afford to take days or weeks to analyze data; they need to automate the data analysis process, with tools commonly used during post-market surveillance, such as data mining and signal detection. Particularly during the target discovery phase, researchers would greatly benefit from expanding beyond pure genome driven target discovery.

Clearly, pharmaceutical firms can no longer rely on traditional pharmacovigilance methods. In today's fiercely competitive and risk-averse industry, it has become essential for pharmaceutical firms to proactively identify and expeditiously manage emerging safety risks. To do this, they

need to establish more sophisticated programs that capitalize on the best available information from multiple data sources, using the most advanced set of tools available for understanding potential signals in earlier stages of the process.

The Need for Enhanced Signal Detection

For pharmacovigilance to be effective, a radical change must take place in the signal detection and management process. The use of data mining and signal detection during post-market surveillance is not unheard of. Using these technologies in clinical trials is a relatively new concept and one worth exploring. In the past, analysis of clinical data required a mutual dialog between researchers, clinicians and statisticians, and could take anywhere from a few days to a few weeks to complete. Using advanced signal detection technologies throughout the clinical trials process allows researchers to perform ad hoc and close-to-real-time data analysis, reducing time in the interchange between researchers, clinicians and statisticians.

The 21 CFR Part 822.31 guidelines recommend the use of data mining tools to aid in signal detection. They also specify which analyses a company should undertake and suggest comparisons with external databases such as the U.S. Federal Drug Administration (FDA) AERS and World Health Organization (WHO) data. However, data mining



tools alone are not sufficient to perform signal detection. While these tools have an important role to play, all the processes involved in signal detection should be considered. These include the collection and delivery of adverse event data for analysis, the triaging of potential signals, and the writing and updating of pharmacovigilance plans. The systems supporting these processes should be validated, and the processes themselves should be governed by written SOPs.

Early detection of safety issues could also lead to termination of future clinical trials, which protects patients from possible adverse events, and saves companies money that could be spent on clinical trials, litigation and regulatory fines.

The WHO defines a signal in pharmacovigilance as, “any reported information on a possible causal relationship between an adverse event and a drug, the relationship being unknown or incompletely documented previously.”

Traditionally, analysis is carried out by a systematic manual review of every report sent by physicians to pharmacovigilance experts. These reports are registered in pharmacovigilance database systems (see Figure 1).

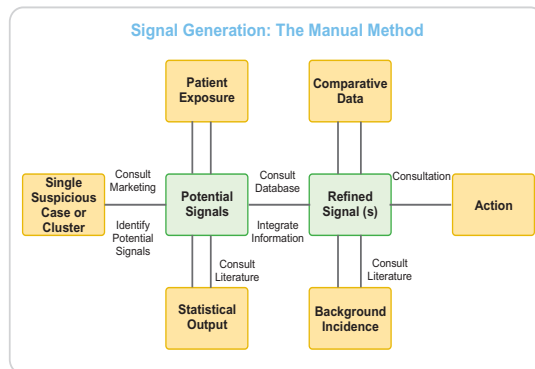


Figure 1: *The manual method for signal detection*

During the last five years, automated signal detection methods have been developed to supplement qualitative clinical methods. While these automated methods cannot replace expert clinical reviewers, they can assist with the difficult task of screening huge numbers of drug-event combinations in databases for potential signals. Through commonly used methods that are based on an underlying model of statistical association, databases are scrutinized for a significant occurrence of disproportionalities or dependencies between drug-event pairs.

The WHO and the FDA are currently using automated detection algorithms based upon Bayesian analysis to achieve signal generation. However, to date, these methods have not been evaluated, and there is no gold standard for signal detection. A recent evaluation of the WHO Bayesian approach showed good overall sensitivity but rather low specificity. One limitation is the small number of occurrences of each drug-event association in the database. Moreover, the performance of these quantitative methods are limited by the fact that they do not take into account the semantic information existing in the controlled vocabularies used to code adverse events in case reports. The Medical Dictionary for Regulatory Activities (MedDRA) is the unified standard terminology currently used for recording and reporting adverse drug event data in most countries. Other similar dictionaries used for coding are WHO-ART, COSTART and ICD.

Traditional Approaches and Tools/Algorithms for Signal Detection

Automated signal generation is a growing field in pharmacovigilance that relies on data mining of huge spontaneous reporting systems for detecting unknown adverse drug reactions (ADR). Previous implementations of quantitative techniques did not take into account issues related to MedDRA terminology used for coding ADRs. MedDRA is a first-generation terminology lacking formal definitions; the grouping of similar medical conditions is not accurate due to taxonomic limitations.

The FDA uses a data mining algorithm called the Multi-item Gamma Poisson Shrinker (MGPS) to interrogate its spontaneous reporting database, as it has the ability to look at drug-drug interactions. The MGPS algorithm examines the ratio of an observed adverse drug event to the total number of adverse drug events (over 56 million) in order to detect a signal using Bayesian statistical analysis.

Information systems available to pharmaceutical or biotech companies must employ a range of different data sources, each with their own data structure and mode of presentation. This diversity among the data sources hinders their integration, and thus hampers the complete understanding of the information they contain.

Several academic and industrial research groups active in biology-related fields are growing ever more convinced that their research would advance at a faster pace and that they would gain better insight into the subjects of their research if only different information sources could be integrated together.

Above all, research conducted during the target discovery phase of the drug discovery process would profit from higher levels of integration. It would allow research scientists to expand beyond pure genome-driven target discovery to a situation where other types of relevant information are also made available.

Need for Proactive Tools to Enable Strategic Pharmacovigilance

The growing government, media, and public concern about drug safety has both increased the importance of pharmacovigilance activities and highlighted the limitations of relying entirely on traditional methods such as individual case review. Today's zero tolerance drug safety environment calls for new strategies to proactively identify and expeditiously manage emerging safety risks.

In this highly competitive and risk-averse environment, it is essential for pharmaceutical companies to establish pharmacovigilance programs that capitalize on the best available information from multiple data sources, using the most advanced set of tools available for developing a thorough and well-understood safety profile.

A balance must be found between the sensitivity, specificity and false positive and false negative signals in order to establish an optimal signal-to-noise ratio. False negative signals are potentially harmful, since important adverse drug reactions may be missed. A large number of false positive signals, on the other hand, may indirectly cause harm in that they may prevent patients from receiving effective treatment. For this reason, the threshold in quantitative signal detection must be established very precisely. Similar to the detection of adverse drug reactions, drug-drug interactions may only be detected after drugs have been marketed, which may have serious implications for the drug(s) involved.

Due to the growing availability of more potent

drugs acting on fundamental biochemical pathways or receptors, the number of drug-drug interactions is likely to increase. The elderly are at particular risk of adverse drug-drug interactions, and furthermore, the risk increases with the number of drugs prescribed and taken concurrently.

In the event that a possible drug-drug interaction causes a decreased effect of one of the drugs, the adverse drug reaction used to identify the drug-drug interaction should be indicative of a lack of efficacy of one of the drugs involved.

Although the first initiatives of spontaneous reporting originated in the middle of the previous century, pharmacovigilance is continuously being refined. In addition to the classical approach -- where for each case an assessment is made as to whether the reported association represents a new adverse drug reaction -- quantitative techniques are being developed to facilitate the signal detection process.

These new techniques enable a more detailed analysis of patient characteristics, drugs and the ADRs involved, thus enhancing the quality and widening the scope of pharmacovigilance. Given the rapid advances in information technology, it is expected that quantitative signal detection will establish itself as a standard source of information in spontaneous reporting, which will clearly benefit the study and monitoring of the safety of drugs after marketing.

The good news is that every stakeholder benefits. Manufacturers can increase compliance with regulatory timelines (100% is achievable), deliver higher quality data, reduce time spent in physician follow-up, increase brand protection with earlier safety warnings, provide competitor product intelligence and potentially open new revenue streams through product line extensions. Healthcare providers gain standard tools and processes for reporting adverse events that save time and offer a higher level of protection for their patients. Healthcare payers can access standardized data to make more informed coverage decisions. With standardization, regulators' safety reviews will be faster and more accurate. Ultimately, patients win with an efficient system that protects them, and, as a result, their perception of the industry will improve.

New Signal Detection Algorithms Provide More Accurate Results

Pharmacovigilance is a natural extension of current data capture and reporting processes. At its core is an advanced analytical engine that allows healthcare stakeholders to monitor product safety data. Based on data warehousing and statistical analysis techniques, this "engine" provides various algorithms to identify trends or clusters of events to help organizations make better decisions over time.

There are two components of this capability: access to the right kind of information in the right format and the use of more advanced statistical methods to see and understand potential signals in earlier stages of the process.

Basic components already exist for leading companies to begin the process of storing and analyzing data for pharmacovigilance. Virtually all pharmaceutical companies use SAS Institute, Inc. tools for analytic processing, particularly to create complex analyses and reports. Complemented by a suite of other SAS solutions, companies can enable proactive pharmacovigilance. The suite includes SAS Data Integration, which allows companies to consistently integrate disparate data sources into industry-standard data structures (e.g., E2B, CDISC and HL7), and SAS Drug Development, which provides a standard, compliant environment to store, retrieve, analyze and report data that unlocks the power of signal detection.

Data mining uses computerized algorithms to identify hidden patterns of unexpected occurrences in large databases. In pharmacovigilance applications, the technique is sometimes used to compensate for the disproportionate reporting patterns in spontaneous adverse event reporting systems. Data mining tools were developed to enhance rather than replace signal detection procedures in large databases. They have been successful as methods for hypothesis generation; analyses of potential signals and detection of complex dependencies (see Figure 2, on next page).

Some of the data mining techniques include the following:

Predictive Modeling

Predictive modeling is a technique used to devel-

op a model to relate a dependent variable with a set of independent variables in a manner similar to multiple regression analysis. There are two types of predictive modeling: classification, for categorical dependent variables, and value prediction, for continuous dependent variables. Classification is appropriate if the goal is to predict group membership of new records based on their characteristics (independent variables). Using classification, the most influential variable is identified and used to split the data into groups. This is then repeated with the next most influential variable until the data are fully characterized.

Clustering or Database Segmentation

Clustering uses an algorithm that segregates a database by evaluating the dissimilarity between records. Pairs of records are compared by the values of the individual fields within them, and clustering into groups provides fast and effective ordering in large datasets. Segmentation could be used to group patients with similar symptoms or diagnoses to determine whether there is a drug association. Thus, clustering is a technique of choice if the goal is to reduce a large sample of records to a smaller set of specific homogeneous subgroups (clusters) without losing much information about the whole sample. Because of the heterogeneity among clusters, this analysis can also be helpful in hypothesis development about the nature of the variation between subgroups.

Link Analysis

Link analysis refers to methods that identify associations or links between records or sets of data. It assesses associations by using an 'if x then y' type rule, by assessing patterns of behavior or by identifying similar time sequences of events.

Deviation Detection

Deviation detection looks for outliers or values that deviate from the norm and can be seen either graphically or statistically. Visualization techniques -- such as scatter plots or histograms, multidimensional graphs for multivariate data, and time series plots -- are used to determine patterns hidden in data. Statistical methods are then employed to measure the significance of deviations once they have been detected. This process could be used to identify patients with idiosyncratic reactions or unusual symptoms that

could be related to medication and may constitute an adverse drug event signal.

Other techniques use Safety Data Mining (SDM)/disproportionality methods to identify adverse events that are reported with greater than expected frequency (statistical independence). Frequency is assessed against the background of all other drugs and events, and results are used for hypothesis generation.

Another statistical tool is the Bayesian methodology to estimate relative reporting rates (risks) of adverse events. It is a tool for further analysis of potential signals and detection of complex dependencies in the data.

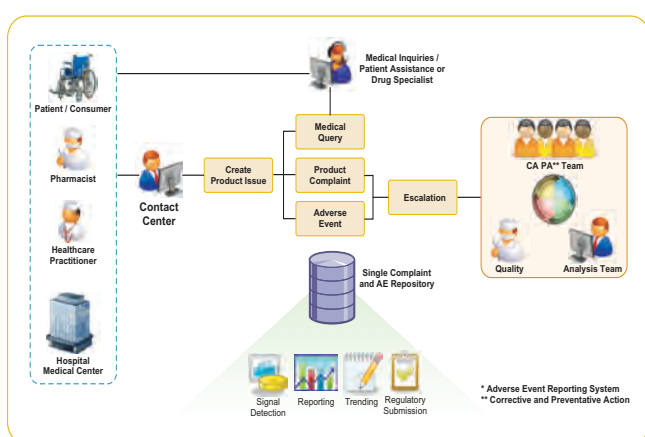


Figure 2: Roadmap to proactive signal detection

Approaches to Improve Source Data Quality

In the domain of molecular biology and related fields, there exists not only a large number of databases containing specialized information, but also a variety of information systems designed to cope with data derived from different disciplines via processes of data integration. The demand for the latter is becoming ever more intense from research groups working on multidisciplinary bioinformatics projects, but such systems are difficult to build because of the heterogeneity of the databases involved.

This heterogeneity is of two sorts:

Syntactic heterogeneity refers to differences in data models (different representation for the same semantic information) and data languages (different data types). These can be easily resolved.

Semantic heterogeneity refers to differences in the underlying meanings of the data represented.

The necessary size of the dataset required is difficult to determine but will depend on the data quality, the background frequency of the event and the strength of the association of the event with the drug.

Some of the external sources of data include the following:

Spontaneous reporting databases

The spontaneous reporting of a possible adverse drug reaction by healthcare providers to governmental agencies or drug companies is an important source of adverse events. However, as spontaneous reporting is a passive system; inconsistent reporting is a limitation, with more frequent reporting for unusual reactions, reactions for new drugs and serious reactions. Furthermore, the accuracy of the data contained within the reports is uncertain. However, spontaneous reporting databases do contain large amounts of data; these databases can therefore be mined to obtain details of adverse drug events.

Prescription event monitoring databases

Prescription event monitoring (PEM) is used to detect adverse drug events by collecting high-quality data from family doctors, on a select group of patients exposed to a specific (new) drug, for a limited period of time. A limitation of PEM database mining is the lack of an adequate control group, as the database contains details of clusters of patients exposed to certain drugs.

Linked administrative databases

Large linked health administrative databases, such as Medicaid in the U.S., contain data on millions of subjects and may also be used as a source for data mining. The data is available at a relatively small additional cost and is not subject to recall or interviewer bias. However, the completeness of details, such as diagnoses, is questionable in many circumstances, and the data tends to apply only to elderly or low-income populations and so may not be representative of the whole population.

Electronic medical records

Electronic medical records (EMR) contain a large number of data fields -- including details such as the use of tobacco products, smoking and non-prescription drugs, symptoms and signs, laboratory data and social circumstances -- on a smaller number of patients. Because of the large number and detail of the variables, which can be combined to generate new diagnoses or adverse events, hypotheses can be explored that are not restricted to existing diagnoses.

Other databases

Clinical trials databases and specialist databases, such as overdose or toxicology databases, may also contain valuable information.

Data retrieved from databases needs to undergo data preprocessing, which involves data sampling and data quality verification to ensure it is clean and well described. Medical data can contain erroneous data, such as ages of 120 years rather than 20 years or a documentation of hysterectomy in males. New variables of interest to be used in the analysis can be generated from the data. For example, it may be possible to estimate socio-economic status from the postal code.

Better data collection applications that employ more rule-based triaging, learning-based automated coding and soft/hard edit checks for data entry will help greatly in reducing time to integrate databases and identifying adverse events.

Cognizant's ASPIRE signal detection and global safety data warehouse framework

Aspire is Cognizant's proprietary drug safety analysis framework. It is a collection of ready-to-use artifacts that can speed up the development process and significantly reduce the cost of building an end-to-end safety integration framework. Aspire provides a framework for global safety data integration from internal and external sources (such as a company's safety database, FDA, WHO safety data, patient exposure data, etc.). It also provides a framework to develop custom analytical reports and incorporate signal detection tools.

Aspire promotes the adoption of industry standards, such as ICH E2B and CDISC SDTM. The features of this framework are outlined below (see Figure 3).

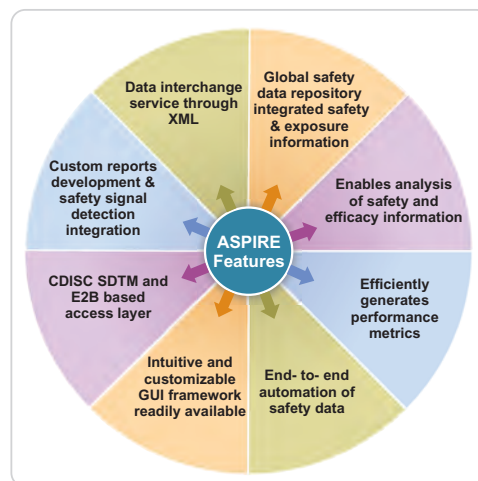


Figure 3. ASPIRE features

The Aspire framework (see Figure 4) can be implemented in the following areas:

- Clinical safety data warehouse.
- Safety signal detection and signal management system.
- Statistical programming platform for creating the safety and efficacy tabulation and listing.

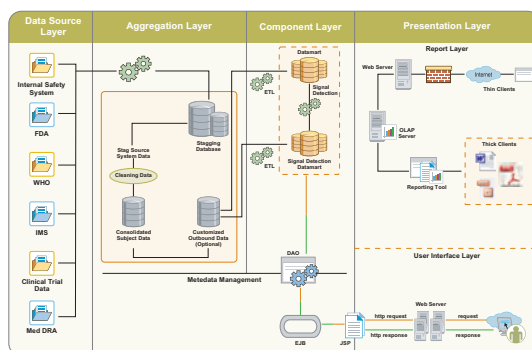


Figure 4. ASPIRE framework

The Aspire signal detection capabilities include but are not limited to the following:

- The ability to identify signals using statistical methods or non-statistical methods: Bayesian methods, PRR, OR, Reporting Rate, Syndrome and Trial Drug - Comparator.
- A collaborative and secure platform to share and delegate the signals.
- Techniques to reduce false positive signals.
- The ability to record analysis of the signals.
- The ability to exclude the cases already in the label.

The ASPIRE model is a graphical tool that helps define a signal and identify parameters that need to be used to detect the signal. It implements a background process that monitors safety data periodically to detect potential signals based on certain detection methods.

Aspire allows pharmaceutical or biotech companies to accomplish the following:

- Create a data mart integrating multiple sources of safety data (both internal and from external regulatory bodies), internal sales data and medical dictionary data.
- Generate safety signals on the data mart using the most appropriate statistical and non-statistical techniques, customized to exactly fit the requirements of the medical safety experts of the company.
- Process and analyze the safety data and generate easy-to-navigate reports (trends, period to period, history, etc.) in any standard tool as per the requirements of the company.
- Substantially reduce the analysis, design and development effort, resulting in a cost savings of 15% to 30%, compared to developing the same solution from scratch. These savings are possible due to reuse of pre-formulated requirements and design documents, data mart design and utilization of re-usable signal generation components and reports.
- Integration with Cognizant's regulatory validation framework to enable compliance of the proposed system with 21 CFR Part 11 requirements. The cost/benefit that can be derived from the solution framework depends on the extent to which it matches the requirements of the pharmaceutical or biotech company.

Cognizant differentiators include the following:

- Ability to pick and chose data sources, detection method, analysis algorithm and visualization tool.
- Support for any program environment.
- Regulatory compliant solution.
- Use of framework reduces the implementation cost by 15% to 30% as opposed to fresh development.

- Already successfully implemented in more than one large pharmaceutical company.
- Very strong life sciences practice and rich experience in the pharmaceutical R&D space.
- A very strong fourth-generation onsite/off shore development and maintenance support infrastructure.

References

1. Nelson RC. "We Need a Post-Marketing Drug Development Process" [commentary]. *Pharmacoepidemiol Drug Saf* 2000; 9:253-5.
2. Benichou C. *Adverse Drug Reactions: A Practical Guide to Diagnosis and Management*. New York; JohnWiley & Sons, 1994.
3. Begaud B, Evreux JC, Jouglard J, et al. "Unexpected or Toxic Drug Reaction Assessment" (imputation). *Actualization of the Method Used in France. Therapie* 1985; 40: 111-8.
4. Benichou C. "Imputability of Unexpected or Toxic Drug Reactions," a chapter in *Adverse Drug Reactions: A Practical Guide to Diagnosis and Management*. New York: Wiley & Sons, 1994: 271-5.
5. Naranjo C, Busto U, Sellers EM. "A Method for Estimating the Probability of Adverse Drug Reactions." *Clin Pharmacol Ther* 1981; 30: 239-45.
6. Bright R, Nelson RC. "Automated Support for Pharmacovigilance: A Proposed System. *Pharmacoepidemiol Drug Safety*."
7. DOH-BFAD-NDP document. "Adverse Drug Reactions: Guide to Participating Hospitals 1996."
8. Edwards IR. "Adverse Drug Reactions," a chapter in *Drug Benefits and Risks, International Textbook of Clinical Pharmacology* 2001. Wiley & Sons.
9. Hartigan-Go K & Wong J. "Inclusion of Therapeutic Failures as Adverse Drug Reactions." *Side Effects of Drugs Annual* 23 2000. Elsevier.
10. Andrew M Wilson, Lehana Thabane, and Anne Holbrook. "Application of Data Mining Techniques in Pharmacovigilance." *British Journal of Clinical Pharmacology* 2004, vol. 57, no. 2, pp. 127-134.

Authors

Kanta Chatterjee is a consultant with Cognizant's Life Sciences practice. She is responsible for managing project plans, timelines and resources, and acts as a subject matter expert on LIMS and lab informatics. Kanta has 18 years of experience in consulting for various manufacturing and healthcare clients and has provided support to various implementation and validation teams. She has experience in clinical trials management, with special emphasis on compliance with JCI guidelines and on-site data management and collection. Kanta can be contacted at kanta.chatterjee@cognizant.com

Vimal Kishore Pannala is a solutions manager for Pharmacovigilance in Cognizant's Life Sciences practice and is responsible for evangelizing IT and process solutions for drug & devices safety. He has extensive experience consulting with global top 20 & Indian pharmaceutical, medical devices & biotech

clients, and has executed and led multiple engagements in process improvement & transition, solution frameworks development and systems landscape analysis. Vimal can be contacted at vimalkishore.pannala@cognizant.com

Partha Chakraborty is the head of the Life Sciences Solutions Center of Excellence and has led the transformation and implementation of clinical development and pharmacovigilance across leading pharmaceutical organizations. He has published his point of view at multiple industry forums and publications, and has advocated adoption of industry standards such as CDISC and ICH. Partha led the creation of ASPIRE - Cognizant's solution framework for clinical and safety, which has been successfully implemented at top 10 global pharmaceutical organizations. Partha can be contacted at Partha.chakraborty@cognizant.com

About Cognizant's Life Science Practice

Cognizant's Life Science Practice partners today with 27 of the top 30 global pharmaceutical/biotech organizations in addition to serving the medical devices, CRO and life sciences product companies. Cognizant's passion is in our focus to help life sciences organization navigate today's dynamic healthcare environment by delivering high value business enabled solutions through our business process outsourcing and information technology solution offerings. Cognizant has been delivering solutions to the life sciences industry for over a decade in the areas of Research, Clinical Development, Manufacturing, Sales, Marketing, Corporate functions and Compliance. Our offerings include Application Development, Application Support, Testing Services, Comprehensive Service Management, Business Intelligence & Analytics services, Infrastructure Outsourcing, Package Implementation, Management Consulting and Business Process Outsourcing.

About Cognizant

Cognizant (NASDAQ: CTSI) is a leading provider of information technology, consulting and business process outsourcing services. Cognizant's single-minded passion is to dedicate our global technology and innovation know-how, our industry expertise and worldwide resources to working together with clients to make their businesses stronger. With more than 40 global delivery centers and 59,000 employees as of June 30, 2008, we combine a unique onsite/offshore delivery model infused by a distinct culture of customer satisfaction. A member of the NASDAQ-100 Index and S&P 500 Index, Cognizant is a Forbes Global 2000 company and a member of the Fortune 1000 and is ranked among the top information technology companies in BusinessWeek's Info Tech 100, Hot Growth and Top 50 Performers listings.

Start Today

For more information on how to drive your business results with Cognizant, contact us at inquiry@cognizant.com or visit our website at: www.cognizant.com.



Cognizant
Passion for building stronger businesses

World Headquarters

500 Frank W. Burr Blvd.
Teaneck, NJ 07666 USA
Phone: +1 201 801 0233
Fax: +1 201 801 0243
Toll Free: +1 888 937 3277
Email: inquiry@cognizant.com

European Headquarters

Haymarket House
28-29 Haymarket
London SW1Y 4SP UK
Phone: +44 (0) 20 7321 4888
Fax: +44 (0) 20 7321 4890
Email: infouk@cognizant.com

India Operations Headquarters

#5/535, Old Mahabalipuram Road
Okkiyam Pettai, Thoraipakkam
Chennai, 600 096 India
Phone: +91 (0) 44 4209 6000
Fax: +91 (0) 44 4209 6060
Email: inquiryindia@cognizant.com