European Respiratory Society
Annual Congress 2012

Abstract Group: 12.1. Medical Education, Web and Internet
Keyword 1: E-health **Keyword 2:** Asthma - management **Keyword 3:** COPD - management

Title: Semantic pattern analysis of patient perceptions using automated co-occurrence information mapping

Dr. Mark 30414 Rolfe mark.rolfe@pharmagenesis.com [1], Dr. Jo 30415 Powell-Bright
Jo.Powell-Bright@pharmagenesis.com [1] and Mr. Daniel 30416 Dodd drd@thoughtwaregroup.com [2]. [1]
Research Evaluation Unit, Oxford PharmaGenesis Ltd, Oxford, United Kingdom and [2] The Thoughtware
Group, Byron Bay, NSW, Australia .

Body: Background: The majority of patients with chronic diseases seek additional information from the internet following medical consultations. Text analytics is a widely recognized, validated system for modelling the structure and information content of text. Here we describe a unique method for identifying and analysing linguistic information from the internet to provide quantitative, unprompted insights into patients' sentiments about their conditions. Methods: Boolean- and thesaurus-based, machine-learning software is used to conduct an iterative nonlinear search of the web for all relevant texts containing broad keywords related to a given chronic disease. Texts are then analysed by Leximancer (v4.0), a text-mining tool that identifies themes and concepts from large bodies of text using a statistics-based algorithm. Results: This innovative approach ensures that a comprehensive disease-specific dataset is captured from the web. Leximancer automatically identifies commonly occurring concepts (weighted combinations of words that co-occur within the text). These are presented visually as maps, showing the strength of the relationship between different concepts (relative frequency and inter-connectedness) to facilitate semantic classification. Positive and negative sentiments about specific aspects of the disease and its management can be identified and selected for statistical analysis, demonstrating the validity of this technique. Conclusion: The internet is a data-rich source of patient-to-patient and patient-to-healthcare professional communications. The sentiment analysis method described can facilitate broader understanding of patient perceptions of their disease and its management.