
Gerald H. Lushington¹* and Mary I. Zgurzynski²

¹Qnapsyn Biosciences, Inc. 16 Dekalb Pike, Suite 248, Blue Bell, PA 19422; ²Boston College, Communication Dept.,
140 Commonwealth Ave. Chestnut Hill, MA 02467

A picture may be worth a thousand words, but in navigating the morass of modern drug development, the right word could be worth a thousand data points.

Dating back several decades, pharmaceutical innovation has suffered a state of prolonged stagnation, characterized by slow elucidation of meaningful drug targets. This problem spans all medical applications, but the greatest gap between practical need and tangible advancement may lie in areas emerging as crucial societal challenges, such as neurological, autoimmune, endocrine, and metabolic (NAEM) disorders.

Beyond the direct personal and economic impact of NAEM conditions, our lack of progress on medically underserved challenges like diabetes, hypertension, lupus, Alzheimer’s, and related afflictions imposes huge indirect burdens since these pathologies are risk factors for cardiovascular disease, cancer, and for vulnerability to infectious disease.

One conceptual barrier to transformative innovation in NAEM-oriented medicine is the absence of clear adversaries. Cancer and infectious disease present distinguishable enemies (e.g., tumors, bacteria, or viruses) to which drug substances can be honed according to differential capacity to harm an obvious pathogen. But what should your medicine ‘kill’ in cases of Alzheimer’s or diabetes? Amyloid plaques, for example, resemble graveyards more than battlefields; outcomes in targeting plaques for Alzheimer’s treatments have been largely neutral or negative [1]. Similarly, attempts to mitigate diabetes through pancreatic alpha cell antagonism [2] have produced no tangible therapeutic benefit.

In the absence of viable adversarial strategies, artificial intelligence (AI) has been applied to seek root causes of pathology within post-genomic molecular interaction networks. Omics analyses, while potentially insightful, have generally faltered for novel NAEM target perception; however, such etiologies are multifactorial, arising from both environmental and genetic factors [3], such that no single determinant substantially alters disease progression.

Fortunately, non-traditional sources of insight are proliferating. Never before has society had such diverse sources of quantitative and qualitative evidence (analytical databases, electronic publications, clinical reports, internet buzz, and electronic health records) and the algorithms to assimilate them. The bright flame of the Baconian scientific method has sputtered, but these new sources of information might be key to reignition. Text mining of peer-reviewed publications and official public documents has recently accelerated; thus, natural language processing may facilitate the processing of anonymized health records and even derive fuzzy inferences from electronic news and social media [4].

However, there are cultural barriers to this. Text inference has long proven analytically valuable for information-intensive disciplines like security, intelligence, law enforcement, and economics, but applications to medical and pharmaceutical discovery remained somewhat taboo until the pandemic. COVID-19, the most extensively communicated health event in history [5], spurred the rapid adoption of new communication modes and standards. The first few months of 2020 produced a blizzard of information, both technical (viral sequencing, structural biology, virion morphology and transport, cleavage eitology, host immune response, drug repurposing prospects, etc.) and qualitative. As the virus first ravaged China, extensive physician/patient interaction reports rapidly transformed our basis for assessing pathology and epidemiology.

Scientific research is only gradually recognizing the transformative values incumbent in our vigorous pandemic response. However, public information repositories such as those initiated by Google [6] have shown major promise in tracking the variant spread and emergent immunity. At the same time, rigorous profiling of trends across patient-specific symptoms has proven instrumental in assessing opportunities for therapeutic repurposing [7, 8], which has been the dominant mode for practical COVID-19 drug development to date.

*Address correspondence to this author at the Qnapsyn Biosciences, Inc. 16 Dekalb Pike, Suite 248, Blue Bell, PA 19422;
E-mail: glushington@qnapsyn.com
In terms of COVID-19 pharmacoeconomics, one opportunity that will persist long after infections subside relates to persistent post-infection autoimmune and neurological sequelae as seen in COVID ‘long-haulers’. A complete understanding of the NAEM-related consequences of COVID may require extensive clinical transcript analysis via protocols akin to those of von Economo, who determined retrospectively (in 1931) that a prolonged worldwide surge in cases of encephalitis lethargica and postencephalitic parkinsonism corresponded to sequelae of the 1918 H1N1 pandemic [9]. If manual text analysis of old clinical reports fostered such dramatic insight in the absence of sophisticated bioanalytical techniques, our body of modern text resources surely also avails golden new opportunities to characterize and treat post-COVID NAEM pathologies.

Von Economo’s clinical transcript analysis transcends traditional ‘omics’ molecular interrogation by statistically assessing qualitative trends across prior and concurrent patient complaints, family history, lifestyle habits, and various other observations that formal clinical reports may neglect. Such analyses have been proposed for adjunctive diagnosis or recharacterization of the complex, multifactorial pathologies inherent in schizophrenia [10] or lupus [11]. Unlike the limited tools available to von Economo, these modern text mining studies are greatly empowered by access to sophisticated artificial intelligence (AI) analysis.

However, the effective application of AI to text-based pathology characterization is limited by data availability. Fortunately, the oddly unconventional nature of NAEM pathologies, while tangibly hindering conventional pharmaceutical countermeasures, may actually strengthen the predictive value of text mining strategies. The myriad of cross-correlative features spanning different NAEM disorders, for example, may expose valuable analogies between common afflictions like diabetes, arthritis, and autism that also inform the molecular pathology of semi-related orphan indications. In other words, surveying clinical transcripts from patients suffering one set of NAEM disorders may illuminate new insight into other challenges.

The fact remains, however, that research on even the most robust applications has access to only a fraction of all potentially useful information. Modern healthcare produces a motherlode of bioanalytical and clinical transcript insight, but much remains obscured for proprietary or privacy reasons. Barriers to accessing even rigorously de-identified clinical transcripts may render many AI training sets inadequate for exposing important nuances.

By way of preliminary conclusion, it is becoming clear that text mining and natural language processing present major advancement opportunities in hitherto frustrating areas of medicine, but legal and practical access to salient clinical records remains an obstacle. To this end, part 2 of this series will discuss prospective insight buried in various underutilized medical text resources, the practical considerations for eventually extracting this insight, some techniques by which the insight can effectively map to medical and pharmaceutical innovation, and conditions under which a greater volume of such resources may become available for the broader good.

CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

Declared none.

REFERENCES


DISCLAIMER: The above article has been published, as is, ahead-of-print, to provide early visibility but is not the final version. Major publication processes like copyediting, proofing, typesetting and further review are still to be done and may lead to changes in the final published version, if it is eventually published. All legal disclaimers that apply to the final published article also apply to this ahead-of-print version.