Asia Pacific Journal of Corpus Research
Vol. 1, No. 1, 95-125

http://www.icr.or.kr/apjcr
e-ISSN 2733-8096
https://doi.org/10.22925/apjcr.2020.1.1.95

A Novel Theory of Support in Social Media Discourse

Bazil Stanley Solomon
(Oxford Brookes University)

Solomon, B. S. (2020). A novel theory of support in social media discourse. *Asia Pacific Journal of Corpus Research*, 1(1), 95-125.

This paper aims to inform people how to support each other on social media. It alludes to an architecture for social media discourse and proposes a novel theory of support in social media discourse. It makes a methodological contribution. It combines predominately artificial intelligence with corpus linguistics analysis. It is on a large-scale dataset of anonymised diabetes-related user's posts from the Facebook platform. Log-likelihood and precision measures help with validation. A multi-method approach with Discourse Analysis helps in understanding any potential patterns. People living with Diabetes are found to employ sophisticated high-frequency patterns of device-enabled categories of purpose and content. It is with, for example, linguistic forms of Advice with stance-taking and targets such as Diabetes amongst other interactional ways. There can be uncertainty and variation of effect displayed when sharing information for support. The implications of the new theory aim at healthcare communicators, corpus linguists and with preliminary work for AI support-bots. These bots may be programmed to utilise the language patterns to support people who need them automatically.

Keywords: AI, Facebook, Linguistics, Purpose, Support

1. Introduction

The paper explores the problem of support patterns in social media discourse for diabetes support. The research is predominately from a computing and artificial intelligence perspective with guidance from linguistics. AI can help reduce the big data and large-scale text corpus from many potential linguistic dimensions to a few high-frequency ones. AI and topic modelling alone can produce too broad base categorisations of text data. Searching for an 'architecture' of these types of discourses is analogous to the search for an architecture of sentences. People use sentences daily, and take it for granted, producing them in a certain way. For instance, a sentence may consist of verbs' doing words' and nouns 'the entities' Lehmann (1987). People use ubiquitous posts and discourses in social media, and they can benefit from a comprehensive and subtle renaissance of study.

In 2020, the Corona Virus Pandemic caused the United Kingdom government like many countries across the world to make their citizens go into a state of lockdown. People formed many more large and significant data interactions in online social media groups. These are to communicate and support each other throughout the national and global health crisis emergency. The frequency of the use of COVID-19 terms and topics increased significantly on social media. Understanding online technologies and human behaviour are at the forefront of a national and global response.

Extensive computing analysis in Linguistics is well-established in corpus linguistics. It can be each word versus its keyness, concordance lines or collocations. This paper is on a Social Media: Facebook Diabetes UK support group and with the use of computational analyses (Blei et al., 2003). It offers a way to use AI to analyse ngram high frequencies in large-scale linguistic data. The subtly of insights from the field of linguistics guides with questions such as can online discourses organise into different categories analogous to the architecture of sentences? The research is about a predominate utilisation of computing and AI analyses with guidance from the field of linguistics.

Human languages are intricate and complex. Blei et al. (2003)

describe ‘what people converse about’ in the sense that a social media post contains many topics. For them, each social media post can have different rankings of topics by comparing the topics in the Post to all other posts, be it consecutive-linear or non-consecutive-linear postings for that particular social media discourse. This AI machine learning understanding of topics in the text, for example, a Facebook post, may not be equivalent to Halliday’s (1967) understanding of a ‘topic’. Halliday considers a topic to be the first expression of any sentence or its theme. What exactly is the topic produced with AI text analysis? Latent Dirichlet allocations (LDA) ‘topics’ of the ‘sentences’ in posts are, however, about the interactions of many individuals in many discourses with many different interactions in overall communications. They may indicate what is remarkable about a corpus of conversations. LDA can cluster similar posts under a set of for, example, top trigrams. So, in effect, the highest frequency top trigram maybe what all those posts are inherently about, even though the human eye would not see the posts as immediately similar. So, both the first idea expressed in a sentence or theme or Post or the highest frequency theme in many posts are significant.

The following sections will elaborate and focus on particularly Advice with stance-taking and domain-specific targets. These concepts help to place the research proposed novel device-enabled discourse categories in context.

Some other terms used and developed in this paper will be helpful to bear in mind. The capitalised ‘TOPIC’ or ‘nent’ will be used in the research to describe the device-enabled discourse content categories. It involves the removal of stopwords in the LDA analysis. The lower-case ‘topic’ or ‘vose’ will describe the device-enabled discourse purpose categories. It involves keeping the stopwords in the LDA analysis. The ‘Target’ will describe the entities mentioned in the posts. Both Blei’s and Halliday’s notions can help to define ‘topic’, ‘TOPIC’ and ‘target’ as used in the research. They are essential in understanding what people converse about and how they purposefully do this.

The specific questions which drive the research are:

RQ1: What attitudes, opinions, and sentiments are people expressing, about their conditions and issues in Facebook Diabetes UK posts?
RQ2: How do people express their attitudes, opinions, and sentiments about their conditions and issues in Facebook Diabetes UK posts?

2. Background

A seminal study in this area is the work of Harvey and Koteyko (2013, p.185). They suggest that peer support connects to advising, stance-taking and meaning. They provide the example' …instead of briefly outlining her position on the topic and giving direct advice through the use of imperatives...' Participant 15 'chooses to state her views indirectly and takes time to elaborate on her background and experiences…creating common ground between herself and the advice-seeker'. Another excellent example of a linguistic approach is in an earlier study by Davison and Pennebaker (1997). They looked at the use of particular words in similar types of online discourse.

Harvey and Koteyko (2013, pp. 165–187) consider some CMC theories to be a negative view of communication. They can also be mechanistic in not allowing for the fluidity and dynamism of participant identities. People do have norms of online behaviour.

2.1. Online Peer Support and Advice and Stance-taking Framework

In Locher and Limberg's (2012) discussion on Advice, they criticise how some authors have understood support practices. For example, they argue for the firmly embedded linguistic form of advice-giving in the study of pragmatics - a study of language in use.

Martin and White's (2005) comprehensive and extensively used framework was developed within SFL to account for how evaluative language functions within situational and cultural contexts. Other prominent social media scholars, such as Zappavigna (2012) excellently use SFL as an approach. She uses a methodology for Twitter social media analysis for analysing micro-post that is a form of corpus-based DA applying SFL. It posits language as a meaning-making resource. It is a theory tailored to answering questions about how meanings work within some contexts, are made, and in this sense, are 'functional'. It is distinct among linguistic theories, as SFL can help to both develop a theory about the social process and a description of language patterns.

As a theoretical concept, 'stance' has been described as Appraisal (Martin, 2000; Martin and White, 2005) or attitude (Halliday, 1994). Online healthcare support writing is a highly social practice. It is one that needs the careful positioning of claims within communities of intelligent people against a background of prior views and voices Tannen and Trester (2013). In consideration of the work of the above researchers and the online healthcare researchers in this paper, this paper suggests the view that people may express their online healthcare needs through epistemic and affective stances. It is following the general view that epistemic stance can be a level of disposition that people could share. In contrast, the affective stance can be a feeling, attitude, mood or degree of emotional intensity that people could also share.

Goldsmith's (2000) typology of advice asking can make the research study more accurate and comprehensive. He shows that there are different themes that people use to call for Advice. He considers the asking for Advice to be also related to asking questions for Advice. Sillence (2013) uses and adds a fifth advice-guidance pattern (same boat no.5, e.g., 'is anyone in the same boat as me?') to Goldsmith's typology as shown in Table 1. There are social relations based on power or solidarity.

Sillence (2013) suggests that the employment of advice strategy while aiming to get a response is to its degree of honesty and clearness. He argues that the level of directness needs examination because of the tension that exists between showing support and appearing to force help on people. These strategies can include questions.

Table 1. Types of Advice Solicitation: Goldsmith's (2000) Typology

No.	Advice pattern	Description frequency	Advice-asking labelling keys in this paper
1	Request for advice	Explicit solicitation of advice using the following phrases: (a) '*I need your advice*'; (b) '*What should I do?*'; and (c) '*Should I do X?*'	AA_R
2	Request for opinion or information	Questions such as '*What do you think?*' or '*What do you think of X?*' that can often generate advice responses even though they may be ambiguous about whether the poster wants to solve a problem or obtain emotional support.	AA_{OI}
3	Problem disclosure	Also, potentially ambiguous, as it can be interpreted as a *request for advice, sympathy, or solidarity*.	AA_P
4	The announcement of a plan of action	The poster may receive advice after announcing their intentions.	AA_A
5	Anyone in the same boat? Sillence (2013)	The poster asks specifically to hear from anyone in the same boat as themselves or those who are going through the same experience.	AA_B

Kouper (2010) focuses on showing the patterns that can help to identify advice-giving in Table 2. He argues that people searching for or giving advice must make choices about how they exchange advice and how this affects the beliefs of the community. He suggests the development of social relations through the giving and taking of advice.

Table 2. Levels of the Directness of Advice: Kouper's (2010) Categories

No.	Type of advice	Description	Advice giving labelling keys
1	Direct advice	Any comment that included imperatives or the modal verb '*should.*'	AG_D
2	Hedged advice	Any comment that contained explicit hedges or hedging devices, e.g., '*I think*', '*It seems*', or '*Why don't you?*'	AG_H
3	Indirect advice	Any comment that lacked explicit or hedged advice but had enough information to act upon it, for example, '*Here's one possibility*' *or 'There are some options.*'	AG_I
4	Description of personal experience	An account of how the person dealt with the situation the advice-seeker had described.	AG_E

2.2. Artificial Intelligence topic-modelling, Computing Analysis and Corpus Linguistics

AI researchers have demonstrated the unsupervised classification of dialogue acts using a Dirichlet process mixture model. Blei (2012) for instance suggests that in LDA, the number of topics can be between 50 and 150; however, the optimal amount usually depends on the size of the dataset or the researcher's knowledge of the domain.

Davison and Pennebaker (1997), in their seminal study, used LIWC to study the chronic illness of Diabetes, capturing emotion words and cognitive words in people's online postings. LIWC text analysis is a linguistic analysis tool that can help to reveal thoughts, feelings, personality, and motivations in a corpus.

Abdallah et al. (2016) have highlighted the difficulties with extracting structured information from unstructured text. Many approaches and systems help with the named entity recognition (NER) task. Abdallah et al. (2016) have investigated MeaningCloud for advanced opinion mining functionality, the globally aggregated polarity of the text and more in-depth analysis and sentence-level breakdown. It gives the polarity, extracting entities and concepts and the sentiment associated with each of them.

2.3. Position on Support in Social Media Discourse and AI/Linguistic Analysis

The conjecture in the research is that support in social media discourse conversations is the result of language-based high-frequency sophisticated patterns. These sophisticated large-scale patterns can be made discoverable by AI machine learning with linguistic analysis. Of primary importance are device enabled discourse categories of purpose and content. It can include linguistic forms such as Advice and stance-taking. There are many other interactional activities amongst peers. It is carried out by people in their meaningful and shared interactions. Together they can influence real outcomes about people's issues and concerns. There is remarkable a high-frequency stance-taking with diverse affective stance and low epistemic stance, not always necessarily in consecutive posts but rather indirectly across many non-consecutive posts concerning a similar target. People can post support at any time and not always in consecutive linear order. The theory may be falsified if counterexamples of such patterns in support become available. Support-bots may be developed to produce meaning in the broader context of human living. A support-bot can be related to how a universal support machine can best help support any other human or any other support machine at any time and anywhere.

Figure 1. Conjecture on Support in Social Media Discourse and AI/Linguistic Analysis

Support is hard to define. Hunt and Koteyko (2015) offer a critique of Social Media Diabetes Support Pages like those on Facebook. The notion of support develops throughout this research. It can be a single detailed knowledge of a domain of social practice. Analysis of advice with stance-taking can help to discover and place any device enabled discourse categories in the corpus. This paper's conjecture in Figure 1 is after much consideration of against what corpus linguistics, for example, Krishnamurthy (1996) suggests that analysts need to keep guard. It is looking at the corpus compared to looking for the things expected to be found or compared to looking for something not supposedly discoverable in the corpus.

This paper suggests that meaning in the broader context of human living can be related to how a universal support machine can best help support any other human or any other computer, at any time, and at anywhere.

This research uses appraisal and SFL of prominent linguistic researchers, for example, Martin and White (2005). They suggest that writing can be emotive, evaluative, and meaningful. It can have a particular stance.

Biber and Finegan (1989, p. 124) define stance as 'the lexical and grammatical expression of attitudes, feelings, judgements, or commitment concerning the propositional content of a message'. It can include adverbs, verbs, and adjectives, which mark affect, certainty, doubt, hedges, emphasis, possibility, necessity, and prediction.

Fundamental research from Du Bois (2007) has established at least three things needed to know a given occasion of stance-taking. It is beyond what may be overtly present in the words and structures of the stance sentence itself: (1) Who is the stance-taker? (2) What is the object of stance? (3) What stance is the stance-taker responding to most?

An explanation for any main patterns utilises Suler's (2004) idea. People connect to communities based on their shared interests and that they express them on social media. Suler's (2004) disinhibition or boldness effect can help to explain why people trust each other enough to share their direct experiences in the first place.

Some key researchers, for example, Harvey and Koteyko (2013) amongst others, suggest potential topics or linguistic devices concerning online healthcare support, as shown in (Table 3).

Table 3. Examples of Support Categorisations

Topic Sub-Categories	Description
Organisational Events	Duk holds events for a person with Diabetes
Charity	Duk holds events to raise charity and also for research
Advice	Duk owners and users offer advice
Knowledge Topics on Diabetes	Posters posts about medical conditions, e.g., blood glucose levels
Emotional Support issues	Posters look (for) and gain a desire to do something and may feel better

Blei et al. (2003) offer an expansion of the above notions by suggesting that people can share, seek to understand, express themselves and explain the meaning of, for example, a single word or n-gram. The use of AI machine learning is not without its challenges, and it helps to understand AI chatbots and their usage of patterns of human conversations. AI researchers, for example, Smith et al. (2011), deal with natural language processing done by machines and the challenges of having conversations with humans in natural ways. They have demonstrated that dialogue systems can benefit from using the language-based abstractions of human dialogue acts and speech acts.

This research adopts amongst others, Oxford Brookes University Ethics Committee and Facebook's policies and Ethics specialist's researchers, for example, Townsend and Wallace's (2017). The data in this research is anonymised.

3. Methodology and Results

The Facebook Diabetes UK corpus is selected, and specialist AI LDA computer programs used for analysis. It follows well-established corpus linguists, for example, Adophis and Knight (2020). They show that it can allow human researchers to perform essential statistical analysis. A broad aim of this paper is that AI, together with human researchers, can discover high-frequency and salient patterns in the data. Adophis and Knight (2020) show of scrutinising any discovered patterns from a more qualitative, theory-informed perspective with Discourse analysis.

The critical work of Partington (2008) on how to conduct corpus linguistics research gives the research ways to do contextualisation of the research questions.

The research uses insights from the vital researchers mentioned in this paper to weave together quantitative and qualitative approaches. The data was collected from the Facebook Diabetes UK page for this investigation. The data collected ranges from February 2008 to July 2015. A total of 226,385 posts is obtained for analysis with a total of 16,137 users/peers who posted 218,068 posts, 96% of the total. It has 6,960,998 tokens and 64,904-word types. Well-known procedures for machine learning and widely used software such as Spyder, with Anaconda and Python, MS Excel and R programming are used to collect, store, clean, anonymise and analyse the data. The anonymised data in this paper, for example, have all individual names, addresses, gender and identification, removed. For example, ‘he’, ‘she’ was replaced with ‘partner’. Also, spelling mistakes, for example, ‘dieing of kidney failier’ becomes ‘dying of kidney failure’, so they were corrected in the parts of the posts finally presented.

A summary of the research mixed-methods steps is given in Figure 2.

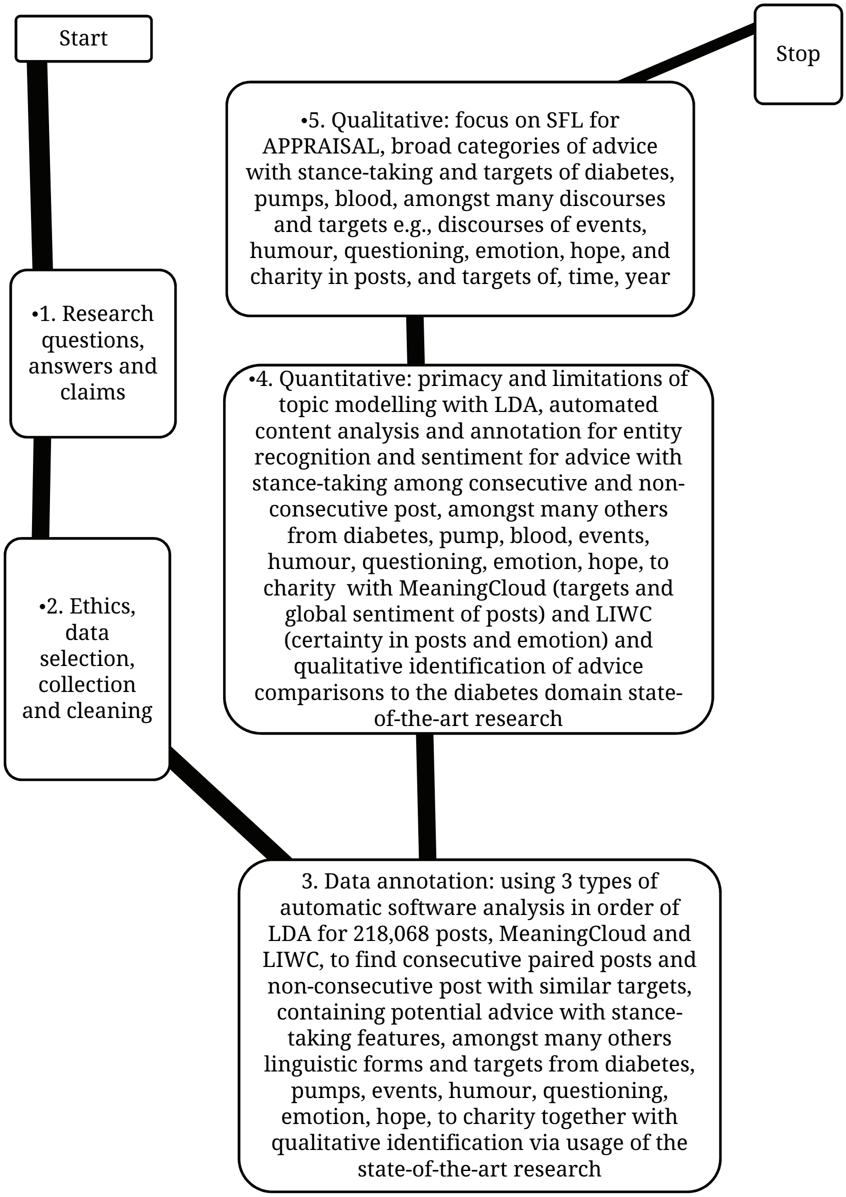

Figure 2. A Diagrammatic Overall Summary of the Methods.

3.1. AI LDA Machine Modelling

The analyses involve the crucial data cleaning and modelling of the large-scale 226,385 posts with keeping the stopwords for topics and removing them for TOPICS and example, keeping the (?) by calling them questionmark. There was a lower log-likelihood of -1316169 for 50 topics as compared to 25 or 100 topics in the model. LDA 50 topics and 500 features are moderately selected after trying parameters for dimension reduction. Trigrams were of significance as it could show patterns for voses and nents.

Five topics (i.e., 0, 24, 13, 31, and 34) were randomly selected from the total fifty analysed topics. The posts within them were then selected for consecutive linear posts. So, there are linear consecutive and non-consecutive posts. Together they can cover the entire corpus. Any post may be related to similar targets and conversations across the corpus. These analyses resulted in moderate seventy-three posts and their substantial quantity of text for further analysis. The assumption is that these posts will tend to be about peers asking for and advising with stance-taking. It is on the similar support issues and therefore may also contain similar stance-taking features. Large-scale text data can be examined by bringing down the scale to basic high-frequency patterns. In this context, it then becomes sensible to use AI and topic modelling for analysis.

The posts tend to be about some aspect of the 'voses' 'device-enabled discourse purpose categories' and the 'nents' 'device-enabled discourse content categories'.

'*Device-enabled discourse purpose categories*' – 'topic' patterns:

topic 0: *do you think*

topic 13: *questionmark questionmark questionmark*

topic 24: *you so much*

topic 31: *diagnosed with type*

topic 34: *you will get*

'*Device-enabled discourse content categories*' – TOPIC' patterns:

The seventy-three posts and their substantial quantity of text were then assigned their respective '*device-enabled discourse* *content categories*' or top TOPIC from a total of 50 TOPICs. They each contain one of the following TOPICS.

TOPIC 27: feel better soon
TOPIC 43: good morning hope
TOPIC 28: questionmark good luck
TOPIC 38: high blood sugars
TOPIC 19: happy new year
TOPIC 12: message add friend
TOPIC 47: feel free add
TOPIC 28: questionmark good luck
TOPIC 4: fast acting insulin
TOPIC 11: questionmark questionmark questionmark
TOPIC 21: hope comes soon
TOPIC 40: ha ha ha
TOPIC 8: Monday Friday pm
TOPIC 31: diagnosed type years

There is a striking observation to emerge from the data analysis of the device-enabled discourse categories. These include a variety of interactional activities amongst peers. They compare to the findings in well-established research.

3.2. MeaningCloud: Findings of Targets and Global Sentiment

MeaningCloud was used to analyse the random posts and their moderate quantity of text. It is to find the target elements in a post, such as the concepts, time expression and quantity. The targets found in the data include, for example, Diabetes, pump, and blood. The research focuses on these domain-specific targets from many others ranging from usernames, places, time, to years.

MeaningCloud is also used to assign global sentiment values to the text of the seventy-three Posts. All the words used together in a post are related to the sentiment of the entire Post and not only just one single word or sentence within the Post. The global sentiment is given for each Post as positive, neutral or negative. Figure 3 is an example of targets and the global sentiment for Post 86.

Targets and global sentiment pattern

MeaningCloud gives targets of blood amongst other targets, for example, place, guy, kind, blood, for the following original Post: 86. It is given a positive global sentiment.

1, 86, '...just came out of hospital…developed keto-acidosis and kidney infection…my blood glucose level was approximately fifty AG_E …. great…happy…' , High positive emotion and positive global sentiment and affective stance; and zero certainty and high tentative and zero insight and epistemic stance for high healthcare, TOPIC 27: feel better soon, topic 24: you so much, targets: place, guy, kind, blood

2, 87, ...living with type 1 diabetes….about ten years……inject twice a day with humalin i - a long lasting insulin….fun!! …blood sugars are generally pretty good AG_E …. if anybody on here wants to chat about type 1 diabetes just drop me a message AG_D …, High positive emotion and global neutral sentiment affective stance; and high certainty and high tentative and zero insight and epistemic stance for high healthcare, TOPIC 38: high blood sugars, topic 24: you so much, Targets: username, long, Humalog, blood

Figure 3. Post 86 and Post 87

3.3. LIWC: Findings of Emotion, Certainty and Healthcare

LIWC is used to analyse the posts for values of either positive, neutral or negative for healthcare, emotion and certainty. Figure 3 Post 86 shows an example.

Emotion and certainty and healthcare pattern

Post 86 LIWC gives high positive emotion and zeroes certainty and high healthcare values for the original Post.

3.4. Manual Identification of Advice Features Amongst the Many Targets and other Potential Linguistics Features

The posts were manually examined against, for example, Tables 1, 2 for advice patterns. With the critical literature on similar work, a broad approach is taken as social media text is not straightforward for a corresponding one-to-one match. There can be many other interactional activities amongst peers. For example, for advice asking pattern $AA_{OI;}$ the text may suggest a question without a question mark when a person who posts asks for information advice. A closer inspection of all the text in the entire posts can help to make a broad fit without losing the gist of the established literature on the matter. Other targets and linguistic features are also investigated, but with a focus on Advice as a potentially crucial element for support patterns.

Post 189527 Figure 4 and post 216361 Figure 5 are examples showing that they can be made up of different sentences, and different targets, discourses categories of content and purposes. These can be tackled in a single post by the person or across other posts by many different people. For example, targets may be for novomix, Levemir, lantus, blood, school, Diabetes, and fourteen years. They cover the diversity of posts by peers. The Post can contain many discourse contents. For example, TOPIC 43: good morning hope potentially about Hope and Greetings, and TOPIC 28: questionmark good luck potentially about Questioning, Good Luck. The discourse can have many purposes, for example, topic 24: you so much potentially focusing lots on the other person; and topic 13: questionmark questionmark questionmark potentially about Questioning.

The posts also tend to be about Advice. Different types of 'Advice' features are identified which tend to match those in the established literature. The posts may contain more than one type of Advice feature because a post can contain more than one sentence; it may seek Advice in one sentence and advise in another. Thus, it was not easy to assign advice types to an entire post but rather to the different parts of the Post. However, the Post may tend towards an overall broad advice pattern from its inherent advice related features. This systematic way of identification can help with the possible bias of trying to find only what one is trying to find in a corpus, and thus that expectation may be counteracted. Targets include school, Diabetes, fourteen years. There tends to be, for example, Figure 3 Post 86 (AG_E) advice-giving with a description of personal experience (see Table 2). It is an account of how the person dealt with the situation that the advice-seeker had described. There tends to be, for example, Figure 4 Post 189527 (AA_P) advice-asking by problem disclosure. It can be potentially ambiguous, as it can be interpreted as a request for Advice, sympathy or solidarity.

62, 189527,…i am on novomix approximately three times a day…most people taking this seems to inject approximately twice a day… i find it ok for routine days…. but when i do anything different during holidays…doing more activity…eating out…find that my blood sugars go all over the place…. am considering changing to levemir and lantus AA_P … would appreciate any comments on the pros and cons AA_R , TOPIC 43: good morning hope, topic 24: you so much, Targets: novomix, Levemir, lantus, blood

Figure 4. Post 189527

4. Research Findings

LDA and DA are used together. LDA is of primary importance to the analysis. It is to identify relationships and concepts for threads of posts and posts across the corpus. They are studied with examples and shown in section 4.1.

For non-consecutive posts, the individual patterns are combined. They can, for example, give evidence for Advice with stance-taking in posts concerning the target word, ‘blood’ and many other interactional activities amongst peers. There is remarkably a high-frequency stance-taking with diverse affective stance and low epistemic stance but not necessarily in consecutive posts but rather indirectly in usage across many posts about the same target. The evidence from this study suggests that the stance of the users may have some alignment on ‘certainty’. They can tend to take up occasionally similar emotion positions or diverging global sentiment positions. These targets are part of the components of the stances, and the participants are responding to a stance on ‘blood’ about blood glucose levels.

The consecutive posts are like the non-consecutive Post for the way support is offered in the non-consecutive posts. They are about similar targets and similar stance-taking. They are not always direct consecutive support posts to persons in the thread. The salience and importance of assessment and advice moves are explained as being contributed to by peers who share personal experiences. The text in the posts contains discursive properties of emotive language. There is general uncertainty and peers tend to construct posts in this manner to seek and give advice. They emphasise descriptions of their targets and do take a stance towards those targets.

4.1. Combining Quantitative Results

Non-consecutive posts

Generally, across posts, Post 86 Figure 2, Post 189527 Figure 3, Post 987 Figure 4 and Post 216361 Figure 6, for example, they can share particular ‘*device-enabled discourse purpose categories.*’ topic 24: you so much, topic 34: you will get, and topic 13: questionmark questionmark questionmark.

They also tend to share different ‘*device-enabled discourse content categories*’. TOPIC 27: feel better soon, TOPIC 43: good morning hope, TOPIC 31: diagnosed type years and TOPIC 28: questionmark good luck.

They could use Advice giving features of AG_E (potential advice-giving in the social group via problem disclosure), AA_P (problem disclosure), AA_R (explicit solicitation of advice), and AG_I (indirect advice, any comment that lacked explicit or hedged advice but had enough information to act upon it, for example, ‘Here’s one possibility’ or ‘There are some options.’)

They tend to share Targets of:

Targets: place, guy, kind, blood

Targets: novomix, Levemir, lantus, blood

Targets: user, type, Diabetes, pill

Targets: blood, injections, people, Diabetes

These meanings tend to be imbued in the quantitative patterns. They are essential in representing online chronic illness support in a certain way by peers.

Consecutive posts

Post 86 and 87 shown in Figure 3

The peers tend to utilise of primary importance device-enabled discourse categories of purpose and content to frame the disease. It is framed as relentless for example TOPIC 8: Monday Friday pm, with long descriptions of how they are coping, yet somehow managing for example topic 24: you so much and topic 13: questionmark questionmark questionmark. It is even though their health is also in an inevitable decline for example with topic 34: you will get, and TOPIC 27: feel better soon, but these devices can be about solidarity and having power over the problematic disease.

4.2. Qualitative Analysis

The most striking observations to emerge from the data analysis is that the high-frequency *device-enabled discourse categories* tend to relate to well-established linguistic patterns. These can be Advice with stance-taking about several and diverse targets of the domain. The FDP postings have examples of high levels of self-positioning that are expressed through both affective and epistemic stances. There are high levels of uncertainty for epistemic stances. There tend to be some elements of a depersonalised stance. They can position some peers as passive recipients of the targets that they converse about in the text. Peers can frequently express zero epistemic stance about how certain they are about their Advice for targets such as 'diabetes' or 'medication'. Peers tend to have support, and they are shown to be supported via the use of trigrams such as topic 0: '*do you think*'. They can utilise a passive voice, but their epistemic stance can include topic 1: '*should be able*', which can mitigate their fluency in the support.

Their usage of positivity, for example, Post 86: '*great…happy*', is also a positive form of appreciation. However, they may not always ascribe it to themselves in an agentive manner. Others presented their language in a matter of fact, epistemic stances. Peers who had similar targets, for example' blood', expressed clear uncertainties about the target as a controllable entity in their daily lives.

The sentiment and emotion that are expressed can vary.

Therefore, there are broader viewpoints available on dealing with the concerns and issues in a community of support. It means that part of the attitudes towards the targets mentioned above shows variation. It is not only one type of sentiment. Opinions which may be considered in the literature to be based on people's sentiments and emotions still differ without a need to have a complete agreement on any target. People are shown to express their own diverse personal experiences about their conditions and issues in the Facebook Diabetes UK posts rather than stating facts.

This sharing of 'information' is not done with high certainty. Support discourses of for example advice with stance-taking, events, humour/sarcasm, questioning, emotion, hope, and charity is not only about sharing 'information' as a fact-based system of knowledge but about finding information and support. There is more going on in 'support', and the research considers it to be about topics and TOPICS with, for example, advising with stance-taking. A possible suggestion is that it allows the objective medical facts and the subjective experience to come under the scrutiny of the public gaze. Participants are seen to be agents or passive participants when dealing with illness and wellness.

7, 987: '…hi…i was diagnosed with type 2 diabetes… approximately eighteen months ago... i am afraid i largely ignore it…just pop the pills and get on with life…i do not like to feel I have an illness… never had any symptoms… thirst…great weight loss or anything... it runs in our family in later life…i am approximately sixty years old… so just accepted that I might get it one day…'AAP; TOPIC 31: diagnosed type years, topic 34: you will get; Targets: username, type, Diabetes, pill.

Figure 5. Post 987

An example is given by Figure 5 post 987 of the discourse. It also contains TOPIC 31: diagnosed type years, which is potentially about discourse content of Diagnosis, Diabetes Type and Years. The discourse also contains topic 34: you will get, which is potentially about the other person and they are getting something. Other Targets of username, type, Diabetes, pill, exist and shows the richness and diversity of posts and discourses.

Thus, we see that a single message of positivity is not the only narrative in the discourses with the daily need of medications, blood glucose tests against an ultimate acceptance of the illness.

This paper suggests that differing sentiments and emotional stances towards the same targets can act as counterbalances and negotiation, and for influencing each other. People can control or manage the exchange of information in social interactions. There may be an influence to become well while facing chronic illness. The posts may be constructed to be about power and solidarity over the illness. Solidarity relations can exist for engaging with the illness and building a network of support. There is an attempt to build friendships with people having similar issues. The concepts of power and solidarity can be useful in future research for a wider elaboration.

4.3. Research Questions

RQ1: What attitudes, opinions, and sentiments are people expressing about their conditions and issues in Facebook Diabetes UK posts?

The results suggest that people take a variety of attitudes. Their stances for the same targets can include a negative or neutral or positive sentiment. The Post can be one of uncertainty. Peers can have a variety of sentiments and zero certainty about the information they are sharing. They can converse about, for instance, many diverse opinions on conditions and issues ranging from their glucose levels (blood), their Diabetes, the use of an insulin pump, school, place, username, events, children, greetings, and insulin

RQ2: How do people express their attitudes, opinions, and sentiments about their conditions and issues in Facebook Diabetes UK posts?

The results indicate that they share information through advice-seeking and advice-giving with stance-taking strategies. It is amongst many other linguistic forms from humour/sarcasm to greetings and many other interactional activities amongst peers.

Listed below are the operationalised research questions and claims about the users:

OQ 1.1: What are the discourse purposes? What are the particular word trigram choices used by people, often together, to express what the discourse purposes are in their posts on social media for supporting people with chronic illness?

What emerges from the results is they use and of primary importance in the research of *device-enabled discourse purpose categories* of for, example:

topic 0: *do you think*
topic 13: *questionmark questionmark questionmark*
topic 24: *you so much*
topic 31: *diagnosed with type*
topic 34: *you will get*

OQ 1.2: About what is the discourse? What frequently used content word trigrams relate to the discourse contents?

This question considers, for example, the Advice in a post that may then contain stance-taking and related content words for support, concerning other posts, or, as the research calls them: for what 'TOPICs'.

What emerges from the results is that they use (and of primary importance in the research) *device-enabled discourse content categories*, for example:

TOPIC 27: feel better soon
TOPIC 43: good morning hope
TOPIC 28: questionmark good luck
TOPIC 38: high blood sugars
TOPIC 19: happy new year
TOPIC 12: message add friend
TOPIC 47: feel free add
TOPIC 28: questionmark good luck
TOPIC 4: fast acting insulin
TOPIC 11: questionmark questionmark questionmark
TOPIC 21: hope comes soon
TOPIC 40: ha ha ha
TOPIC 8: Monday Friday pm

TOPIC 31: diagnosed type years

OQ 1.3: What is its primary target? What nouns/entities are discussed in these posts?

The results that emerge together with the analysis provide essential insights. It is into crucial domain-specific targets such as ‘blood glucose level’, ‘diabetes’, ‘children with diabetes’, ‘insulin pumps’ and ‘medication’. They show the diversity of the corpus ranging from these targets to, for example, school, children, username, events, and greetings. These are available in consecutive and non-consecutive posts.

OQ 2.1: What is the poster’s stance about the certainty of their information? What is the poster’s stance about the certainty of their information? How certain is the person in these posts and do people express these directly to each other in consecutive or in non-consecutive posts?

In summary, these results show that there is mostly an alignment of zero certainties about ‘blood glucose level’, ‘diabetes’, ‘children with diabetes’, ‘insulin pumps’ and ‘medication’.

There is remarkably a high-frequency stance-taking with low epistemic stance but not necessarily in consecutive posts but rather indirectly in usage across many posts about the same target.

OQ 2.2: What is the poster’s stance concerning their feelings about the information and how do they feel in these posts and do people express these directly to each other in consecutive or in non-consecutive posts?

In summary, these results show that there is a varied stance. It is based on the differing sentiment towards the same objects, for example,’ blood glucose level’, ‘diabetes’, ‘children with diabetes’, ‘insulin pumps’ and ‘medication’. There is remarkably a high-frequency stance-taking with diverse affective stance but not necessarily in consecutive posts but rather indirectly in usage across many more non-consecutive posts about the same target.

4.4. Validation and Limitations

Precision quality measures were used (Table 4) for the validation of the LDA Model topic-clustering of posts. The precision measure results suggest a good fit for the overall FDP posts and are a better fit for the users/peers' posts. These precision measure results can be attributed to the smaller number of organisation total posts of 4% when compared to user/peer total posts of 96%. Log-likelihood calculations and the researcher's knowledge of the domain gave 50 topics.

Table 4. LDA Model Precision Measures

LDA	Precision	Recall	F1-score	Support
Diabetesuk	0.666667	0.001414	0.002821	8,489
Peer	0.962382	0.999972	0.980817	216,871
Avg/Total	0.951242	0.962358	0.943977	225,360

5. Discussion and Conclusions

The paper highlights the interdependence of data and theory. The theory can help in explaining 'support in social media discourse'. The device-enabled discourse categories with linguistic forms such as advice with stance-taking patterns point to a super-category of 'Support'. It is shown in Figure 6, the target studied is 'diabetes' and 'blood'. These are shown to be produced in a certain way with TOPIC 28: questionmark good luck; topic 13: questionmark questionmark questionmark; TOPIC 31: diagnosed, type, years; topic 24: you so much

72, 216361, '…people do not see the hard work that goes into trying to manage blood sugar levels AG_I …to try to keep healthy…after approximately twenty-six years since my child was diagnosed at approximately five years old with type 1 diabetes… you would have thought we would have something else to manage the condition… instead of relentless invasive blood testing and injections AG_E …how i wish for a day that my child was free from injections…', Low positive and zero negative emotion and global positive sentiment and affective stance; and zero certainty and high tentative and high insight and epistemic stance; for high healthcare, TOPIC 28: questionmark good luck, topic 13: questionmark questionmark questionmark, Targets: blood, injections, people, diabetes

73, 216362,'…i have type 1…was diagnosed approximately four years ago…after feeling unwell for approximately six months… having hypos and not realising it!...finally getting admitted to hospital as blood glucose levels were really high…after tests…being hooked up to an insulin pump…glucose drip overnight…the consultant came round the following morning… with the obligatory students…just announced this person has type 1 diabetes and will control it with insulin injections…needless to say I was shocked, stunned and devastated…i cannot express my gratitude to the staff…the continuing care I receive from my local hospital..It has been a hard learning curve for me and my family...control was difficult at the start but has been on a course…i can now count carbohydrates…it has stopped being so scary.. at times it is frightening…i have had huge problems with work… at times they just do not understand that I have to rest through illness AG_E , … until I was diagnosed I did not know anything about it... let us get the in-formation out there!.. AG_D .all my friends and family just about understand…just not fully….' , High negative emotion and global negative sentiment and affective stance; and zero certainty and low tentative and high insight and epistemic stance; for high healthcare, TOPIC 31: diagnosed type years, topic 24: you so much; Targets: insulin, family, type, diabetes

Figure 6. Post 216361 and 216362 Support is Illustrated with the Target Words, 'Diabetes' and 'Blood'

5.1. AI and Linguistic Analysis

A novel way is to look for (and of primary importance in the research) the '*device-enabled discourse purpose categories*' or voses and ‘*device-enabled discourse content categories*’ or nents. LDA is used in this instance, as it finds latent word usage patterns across the posts and creates ‘topic’ and ‘TOPIC’ categories. Corpus linguistics offers a way to combine AI and linguistics analysis, so the targets and variable sentiment and similar certainty show that support happens in a particular manner.

5.2. A Proposed Novel Theory of Support

From the analysis of the online chronic illness discourse, it shows high-frequency patterns. They contain *device-enabled discourse purpose categories (voses)* and *device-enabled discourse content categories (nents)*. There is remarkably a high-frequency advice with stance-taking with diverse affective stance and low epistemic stance but not necessarily in consecutive posts but somewhat indirectly in usage across many posts about the same target. The peers or the ‘Discourse’ can act as peers. People may talk to each other sequentially or post at any time or anywhere for support, but it is part of the support pattern.

The representation of the theory in Figure 7 can be utilised and applied to online support discourses by understanding the different levels. The Level 1 high-frequency device enables support enabled discourse purposes and content. There are different levels for the components with an expanding representation in the diagram so that it is more apparent how the components relate to each other in theory. The diagram also shows that the theory is falsifiable. It is if high-frequency patterns of the device-enabled discourse categories of purpose and content are not available in any similar online social media discourse. The limitation of the theory is in its focus on crucial discourse aspects and targets such as Diabetes, blood, pump, medication. The theory may need more exceptional detail from future research on each component. The theory suggests concepts and relationships where people express attitudes, opinions and sentiments about their issues. They may post chronologically or in a non-consecutive non-linear way about topics, TOPICs and targets across the entire online discourse. People on social media platforms tend to support each other during chronic illnesses with a high-frequency sophisticated language pattern.

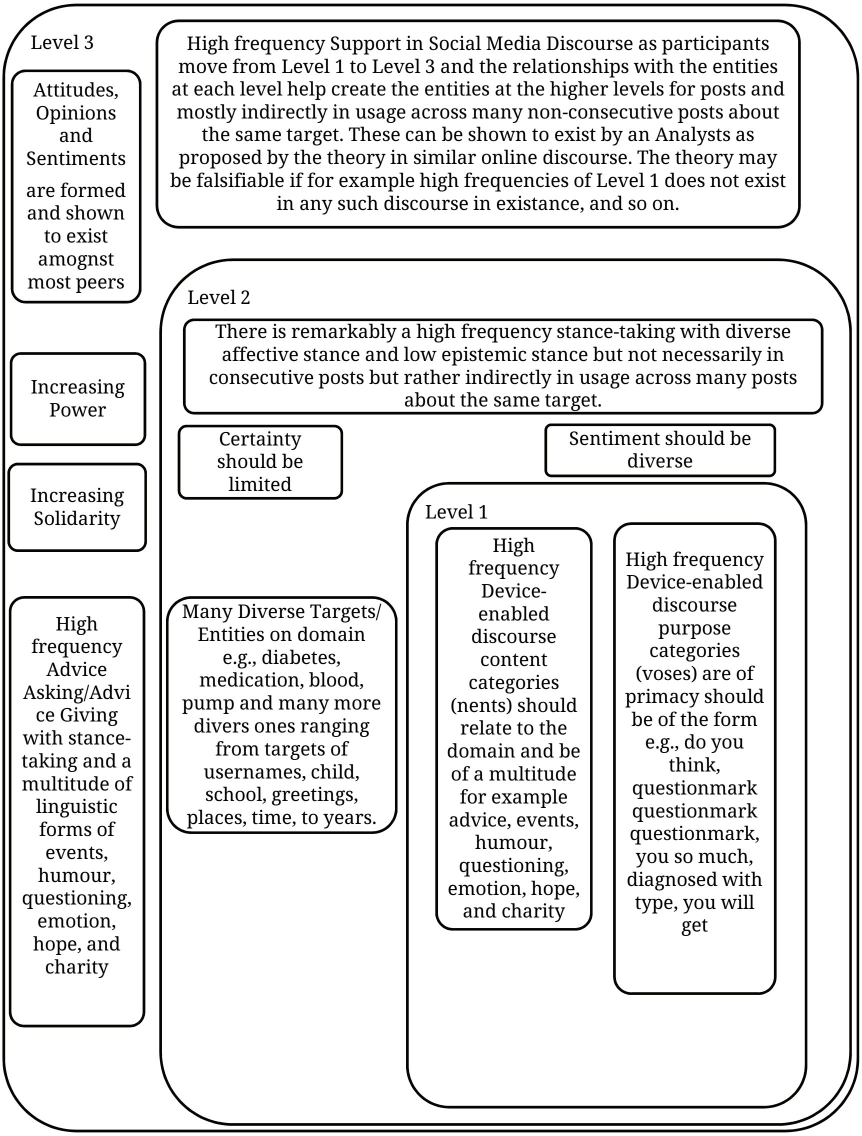

Figure 7. The Concise Points of the Theory of Support for Facebook Diabetes Discourse

5.3. Limitations

The reduction of large-scale text data into a moderate quantity of high-frequency patterns is not without its problems. Linguistics guides an understanding of the corpus. The context of the corpus is studied with DA. It is utilised on random posts selected from within the LDA' topic' and the 'TOPIC' of the LDA models. This approach also gives further analysis and comparison of high-frequency patterns of the LDA' topic' and 'TOPIC', targets, linguistic forms and interactional types and their posts. It is to gain a context of what people were conversing with each other about and how. However, this is only one possible way of looking at the corpus. The theory needs to be refined against many more online large-scale corpora studies.

5.4. Implications

Linguists can benefit from the language patterns and the theory of support in social media discourse.

Corpus linguists can benefit from the ideas about combined approaches (with predominate AI machine learning, entity recognition, sentiment analysis and DA) for finding patterns in large corpora.

AI practitioners can benefit from the theory of support and language patterns to create what this paper calls support-bots (AI automatic Conversational Support). They may help people profit from the more extensive scales and quicker responses for online support.

References

Abdallah, Z. S., Carman. M., & Haffari, G. (2017). Multi-domain evaluation framework for named entity recognition tools. *Computer Speech and Language, 43*, 34-55.

Adophis, S., & Knight, D. (2020). English language and the digital humanities. In Adophis, S., & Knight, D. (Eds.), *The Routledge*

Handbook of English Language and Digital Humanities (pp. 1-3). Oxford: Taylor & Francis.

Biber, D., & Finegan, E. (1989). Styles of stance in English: Lexical and grammatical marking of evidentiality and affect. *Text, 9*(1), 93-124.

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, *55*(4), 77-84.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, *3*(1), 993-1022.

Du Bois, J. W. (2007). The stance triangle. In Englebretson, R. (Ed.), *Stance-taking in Discourse: Subjectivity, Evaluation, Interaction* (pp. 139-182). Amsterdam: John Benjamins.

Davison, K. P., & Pennebaker, J. W. (1997). Virtual narratives: Illness representations in online support groups. In Petrie, K. J., & Weinman, J. A. (Eds.), *Perceptions of Health and Illness: Current Research and Applications* (pp. 463–486). Amsterdam: Harwood Academic Publishers.

Goldsmith, D. (2009). Soliciting advice: The role of sequential placement in mitigating face threat. *Communication Monograph*, *67*(1), 1-19.

Halliday, M. A. K. (1967). Notes on transitivity and theme in English. *Journal of Linguistics*, *3* (1), 177-274.

Halliday, M. A. K. (1994). *An Introduction to Functional Grammar*. London: Edward Arnold.

Harvey, K., & Koteyko, N. (2013). *Exploring Health Communication: Language in Action*. Abingdon, Oxon: Routledge.

Hunt, D., & Koteyko, N. (2015). What was your blood sugar reading this morning? Representing diabetes self-management on Facebook. *Discourse and Society*, *26*(4), 445-463.

Krishnamurthy, R. (1996). Ethnic, racial and tribal: The language of racism? In Caldas-Coulthard, C. R., & Coulthard, M. (Eds.), *Texts and Practices: Readings in Critical Discourse Analysis* (pp. 129–49). London: Routledge.

Lehmann, W. (1987). *Language*. Chicago: University of Chicago Press.

Locher, M. A., & Limberg. H. (2012). *Advice in Discourse*. Amsterdam: John Benjamins.

Martin, J. R., & White. P. R. R. (2005). *The Language of Evaluation: Appraisal in English*. London: Palgrave Macmillan.

Martin, J. R. (2000). Beyond exchange: APPRAISAL systems in English. In Hunston, S., & Thompson, G. (Eds.), *Evaluation in Text: Authorial Stance and the Construction of Discourse* (pp. 142–75), Oxford: Oxford University Press.

Partington, A. (2008). The armchair and the machine: Corpus-assisted discourse studies. In Taylor Torsello, C., Ackerley, K. & Castello, E. (Eds.), *Corpora for University Language Teachers* (pp. 189-213). Bern: Peter Lang.

Sillence, E. (2013). Giving and receiving peer advice in an online breast cancer support group. *Cyberpsychology, Behaviour, and Social Networking, 16*(6), 480-485.

Smith, C., Crook. N., Dobnik. S., & Charlton, D. (2011). Interaction strategies for an affective conversational agent. *Presence: Teleoperators Virtual Environments*, *20*(5), 395-411.

Suler, J. (2004). The online disinhibition effect. *Cyberpsychology and Behaviour, 7*(3), 321-326.

Tannen, D., & Trester. A.M. (2013). *Discourse 2.0: Language and New Media*. Washington: Georgetown University Press.

Townsend, L., & Wallace, C. (2017). The ethics of using social media data in research: a new framework. In Woodfield, K. (Ed.), *The Ethics of Online Research: Advances in Research Ethics and Integrity* (pp.189-207). Bingley: Emerald Publishing Limited.

Zappavigna, M. (2012). *Discourse of Twitter and Social Media: How We Use Language to Create Affiliation on the Web.* London: Continuum.

THE AUTHOR

Bazil Stanley Solomon is a graduate research student supervised by Prof Nigel Crook at the Oxford Brookes University, Artificial Intelligence Department, Oxford, United Kingdom. He studied and researched the field of Requirements Engineering, Social Media Collaborative Work Systems (SMCW) and AI and corpus linguistics. Bazil works as a politician for the local British Government. He runs his own UK Business: IT Analysis and

Training Ltd. He gives workshops, papers and teaches at Universities, Colleges, Government and Businesses worldwide.

THE AUTHOR'S ADDRESS

First and Corresponding Author
Bazil Stanley Solomon
PhD Student
Department of Artificial Intelligence
Oxford Brookes University
111 Eastmere, Liden, Swindon, Wiltshire, SN3 6LG, UNITED KINGDOM
E-mail: p0011188@brookes.ac.uk

Received: 13 May 2020
Received in Revised Form: 31 July 2020
Accepted: 11 August 2020