ORIGINAL RESEARCH
published: 26 July 2021
doi: 10.3389/fdgth.2021.598916

Markup: A Web-Based Annotation Tool Powered by Active Learning

*Samuel Dobbie[1,2], Huw Strafford[1,2], W. Owen Pickrell[2,3], Beata Fonferko-Shadrach[2], Carys Jones[2], Ashley Akbari[1,2], Simon Thompson[1,2] and Arron Lacey[1,2]**

[1] Health Data Research UK, Swansea University Medical School, Swansea University, Swansea, United Kingdom, [2] Swansea University Medical School, Swansea University, Swansea, United Kingdom, [3] Neurology Department, Morriston Hospital, Swansea Bay University Health Board, Swansea, United Kingdom

Across various domains, such as health and social care, law, news, and social media, there are increasing quantities of unstructured texts being produced. These potential data sources often contain rich information that could be used for domain-specific and research purposes. However, the unstructured nature of free-text data poses a significant challenge for its utilisation due to the necessity of substantial manual intervention from domain-experts to label embedded information. Annotation tools can assist with this process by providing functionality that enables the accurate capture and transformation of unstructured texts into structured annotations, which can be used individually, or as part of larger Natural Language Processing (NLP) pipelines. We present Markup (https://www.getmarkup.com/) an open-source, web-based annotation tool that is undergoing continued development for use across all domains. Markup incorporates NLP and Active Learning (AL) technologies to enable rapid and accurate annotation using custom user configurations, predictive annotation suggestions, and automated mapping suggestions to both domain-specific ontologies, such as the Unified Medical Language System (UMLS), and custom, user-defined ontologies. We demonstrate a real-world use case of how Markup has been used in a healthcare setting to annotate structured information from unstructured clinic letters, where captured annotations were used to build and test NLP applications.

Keywords: natural language processing, active learning, unstructured text, annotation, sequence-to-sequence learning

Edited by:
Aurélie Névéol,
Université Paris-Saclay, France

Reviewed by:
Rezarta Islamaj,
United States National Library of Medicine (NLM), United States
Uwe Aickelin,
The University of Melbourne, Australia

****Correspondence:***
Arron Lacey
a.s.lacey@swansea.ac.uk

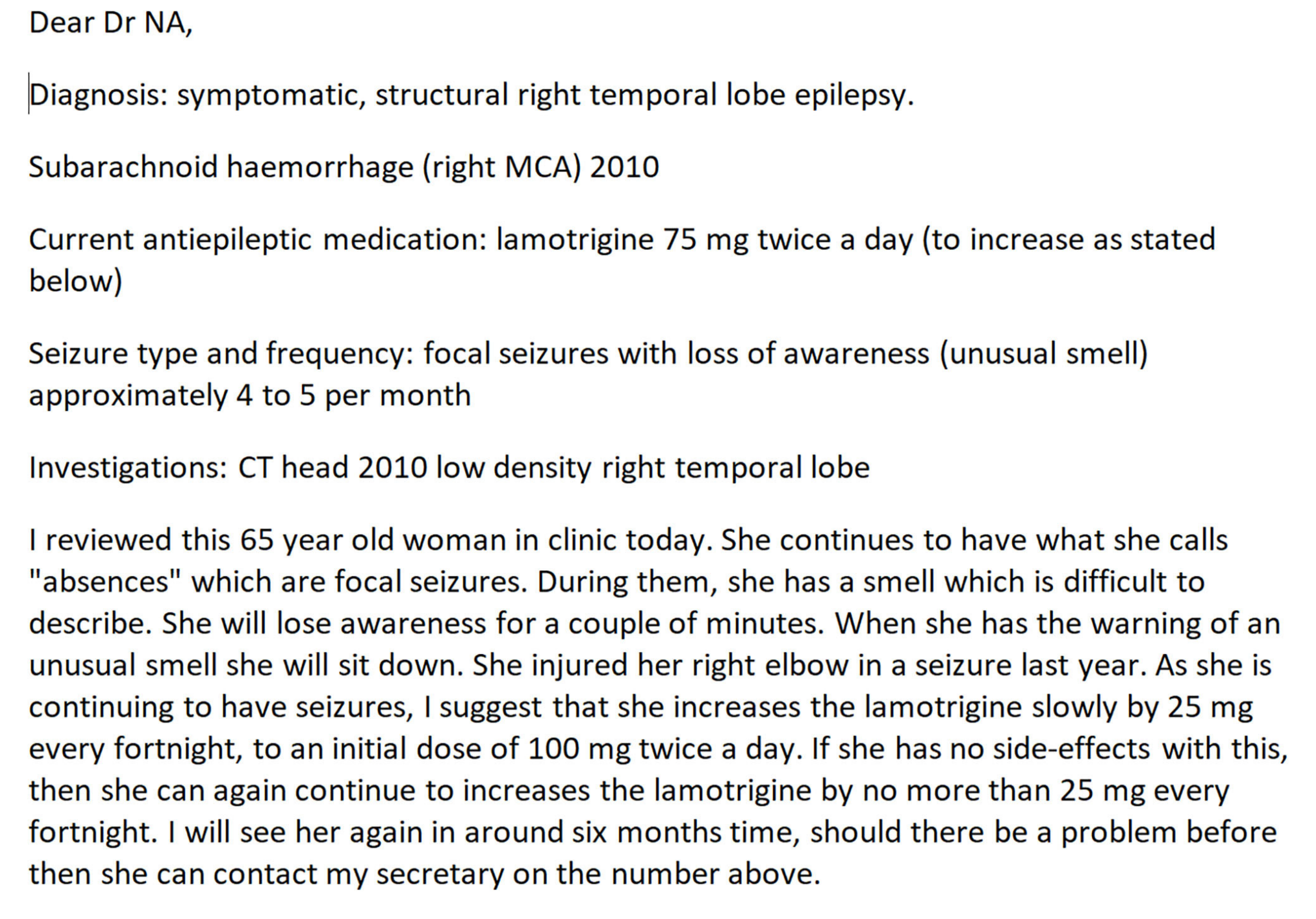
Specialty section:
This article was submitted to Health Informatics, a section of the journal Frontiers in Digital Health

***Received:** 25 August 2020*
***Accepted:** 16 June 2021*
***Published:** 26 July 2021*

Citation:
Dobbie S, Strafford H, Pickrell WO, Fonferko-Shadrach B, Jones C, Akbari A, Thompson S and Lacey A (2021) Markup: A Web-Based Annotation Tool Powered by Active Learning. Front. Digit. Health 3:598916. doi: 10.3389/fdgth.2021.598916

INTRODUCTION

Across various domains, there are increasing quantities of unstructured free-text data being produced. For example, in healthcare, clinical letters are created to document details of patient consultations. These letters often contain valuable information relating to patient symptoms, history, risks, outcomes, prescriptions, and diagnoses, as shown in **Figure 1**. However, the unstructured nature of these letters poses a challenge for the retrieval and utilisation of embedded data in novel healthcare research. Whilst Natural Language Processing (NLP) techniques can be employed to perform information extraction, gold standard annotation datasets are still necessary for training and validation.



Dear Dr NA,

Diagnosis: symptomatic, structural right temporal lobe epilepsy.

Subarachnoid haemorrhage (right MCA) 2010

Current antiepileptic medication: lamotrigine 75 mg twice a day (to increase as stated below)

Seizure type and frequency: focal seizures with loss of awareness (unusual smell) approximately 4 to 5 per month

Investigations: CT head 2010 low density right temporal lobe

I reviewed this 65 year old woman in clinic today. She continues to have what she calls "absences" which are focal seizures. During them, she has a smell which is difficult to describe. She will lose awareness for a couple of minutes. When she has the warning of an unusual smell she will sit down. She injured her right elbow in a seizure last year. As she is continuing to have seizures, I suggest that she increases the lamotrigine slowly by 25 mg every fortnight, to an initial dose of 100 mg twice a day. If she has no side-effects with this, then she can again continue to increases the lamotrigine by no more than 25 mg every fortnight. I will see her again in around six months time, should there be a problem before then she can contact my secretary on the number above.

FIGURE 1 | A de-identified and anonymised unstructured free-text clinical letter.

It is therefore important to provide domain experts with the tools necessary to annotate unstructured text rapidly and accurately. Support from such tools could consist of highlighting key phrases, capturing detailed attributes within a phrase, or suggesting domain-specific annotations (1).

Several tools have been developed with the aim of assisting annotators throughout the annotation processes. The brat rapid annotation tool (brat) is widely used and allows users to annotate both entities and attributes within entities (e.g., negation status), and define linkage, or relationships between entities (2). Some tools, such as the extensible Human Oracle Suite of Tools and Knowtator, include the ability to annotate against pre-existing ontologies such as UMLS. Like brat, these tools are run locally on user machines (3, 4).

More recent annotation tools, such as the brat based WebAnno, TeatTat, and Marky have introduced useful features that emphasise distributed annotation tasks (5–7). Users can set up projects with annotation schemas and are able to compute inter-annotator agreement across sessions.

Machine learning has been used in some tools to speed up the annotation process by providing annotation suggestions to the user. These approaches fall into two categories: pre-annotation and active learning. The Rapid Text Annotation Tool generates pre-annotations from a gold standard annotated corpus. INCEpTION and ezTAG use active learning, adapting to user annotations and improving suggestions over time (8, 9). There is also a growing emphasis on web-based software that does not rely on local installations, as shown by tools such as Anafora (10).

The structured annotations that result from annotating a document with an annotation tool can be used as the building blocks of datasets for training and developing NLP tools and Artificial Intelligence (AI) systems. However, it takes an experienced annotator an average of 15–30 min to annotate a document that has 41 data elements embedded (11). The time-consuming nature of this process, combined with the value of annotator time, can make it infeasible for individual groups to annotate the quantities of documents necessary to train large-scale, accurate AI systems. As such, opportunities to utilise the information embedded in unstructured documents for NLP and AI may be limited.

Markup aims to incorporate desirable features from existing tools, whilst introducing novel features to assist annotators and streamline the annotation process. The primary novel feature offering of Markup is an extension to traditional active learning based Named Entity Recognition (NER) suggestions. Markup uses Long Short Term Memory (LSTM) networks to perform sequence-to-sequence conversion of sentences into relevant attributes for named entities. The networks are trained on synthetically generated templates *via* a bespoke data generation interface offered by Markup. Markup is fully integrated with UMLS and uses phrase approximation to map annotations to UMLS Concepts. The tool is available as a central resource *via* a website, https://www.getmarkup.com/, and as a local application (no internet required) which can be installed from https://github.com/samueldobbie/markup/.

MATERIALS AND METHODS

Display

Markup provides a compartmentalised display consisting of a configuration panel (left), a document panel (centre), and an annotation panel (right), with the option of a dark mode (**Figure 2**).

Configuration Panel

The configuration panel contains selectable entities and attributes parsed from the user-defined configuration file. The panel also lists suggested ontology mappings and acts as the central hub for adding and exporting annotations.

Document Panel

The document panel contains the text of the currently opened document. Spans of text the user has annotated will be coloured based on the selected entity, as described in section Annotation.

Annotation Panel

The annotation panel displays a categorised and ordered list of annotations added by the user along with predicted annotations, as discussed in section Predictive Annotation Suggestions.

Setup

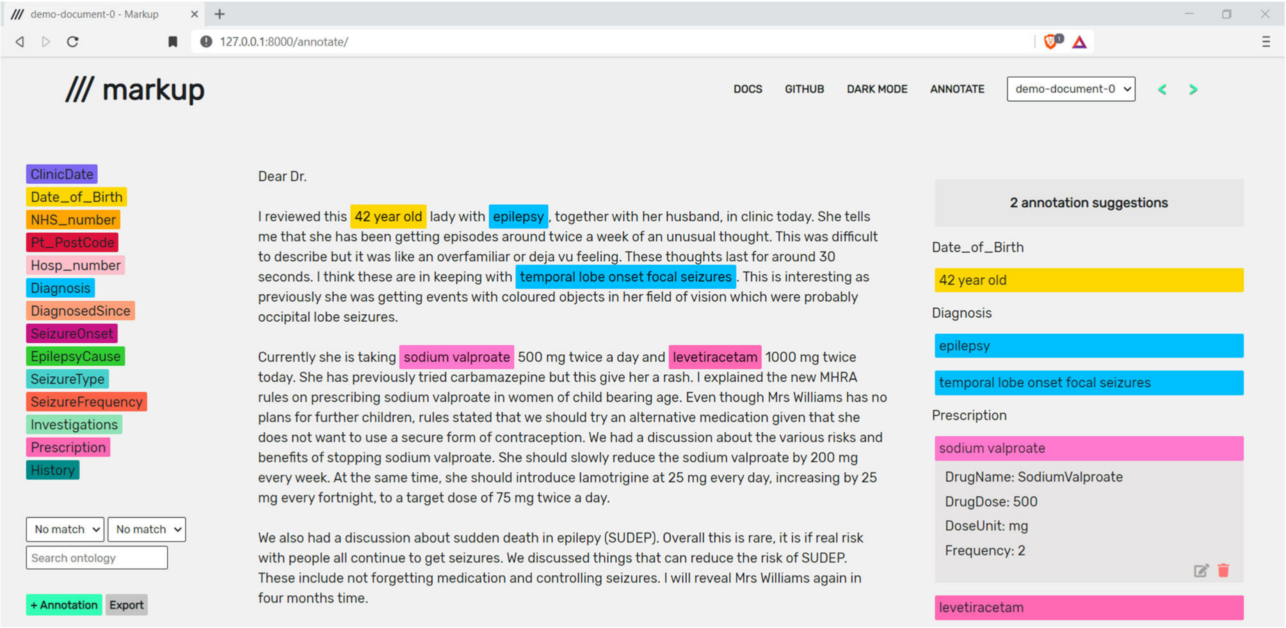
Markup has a quick and intuitive setup process, which enables users to open, navigate, and annotate any number of plaintext documents during a single session. To define the entities and attributes that will be available throughout the annotation task, a configuration file in brat standoff format must be provided. To assist users, Markup offers an in-built configuration file creator that outputs in brat standoff format (**Figure 3**).

Existing annotation files may be specified during setup, with the existing annotations being displayed for each related document. These annotation files must have the .ann file extension and be in brat standoff format.

Users have the option of uploading a custom ontology to be used for automated ontology mappings and direct querying, as discussed in section Automated Ontology Mapping. The custom ontology must be a plain text file, with each line being of the form [TERM][TAB][CODE]. If no custom ontology is provided, pre-loaded ontologies, such as UMLS, may be used. Certain pre-loaded ontologies require additional user permissions, thus users may need to authenticate *via* an external account, such as a UMLS account, prior to access. Markup allows users to authenticate via the UMLS authentication API.

Annotation

Markup annotations can be added by selecting a span of text within a document along with the desired entity and any number



Dear Dr.

I reviewed this 42 year old lady with epilepsy, together with her husband, in clinic today. She tells me that she has been getting episodes around twice a week of an unusual thought. This was difficult to describe but it was like an overfamiliar or deja vu feeling. These thoughts last for around 30 seconds. I think these are in keeping with temporal lobe onset focal seizures. This is interesting as previously she was getting events with coloured objects in her field of vision which were probably occipital lobe seizures.

Currently she is taking sodium valproate 500 mg twice a day and levetiracetam 1000 mg twice today. She has previously tried carbamazepine but this give her a rash. I explained the new MHRA rules on prescribing sodium valproate in women of child bearing age. Even though Mrs Williams has no plans for further children, rules stated that we should try an alternative medication given that she does not want to use a secure form of contraception. We had a discussion about the various risks and benefits of stopping sodium valproate. She should slowly reduce the sodium valproate by 200 mg every week. At the same time, she should introduce lamotrigine at 25 mg every day, increasing by 25 mg every fortnight, to a target dose of 75 mg twice a day.

We also had a discussion about sudden death in epilepy (SUDEP). Overall this is rare, it is if real risk with people all continue to get seizures. We discussed things that can reduce the risk of SUDEP. These include not forgetting medication and controlling seizures. I will reveal Mrs Williams again in four months time.

2 annotation suggestions

Date_of_Birth

42 year old

Diagnosis

epilepsy

temporal lobe onset focal seizures

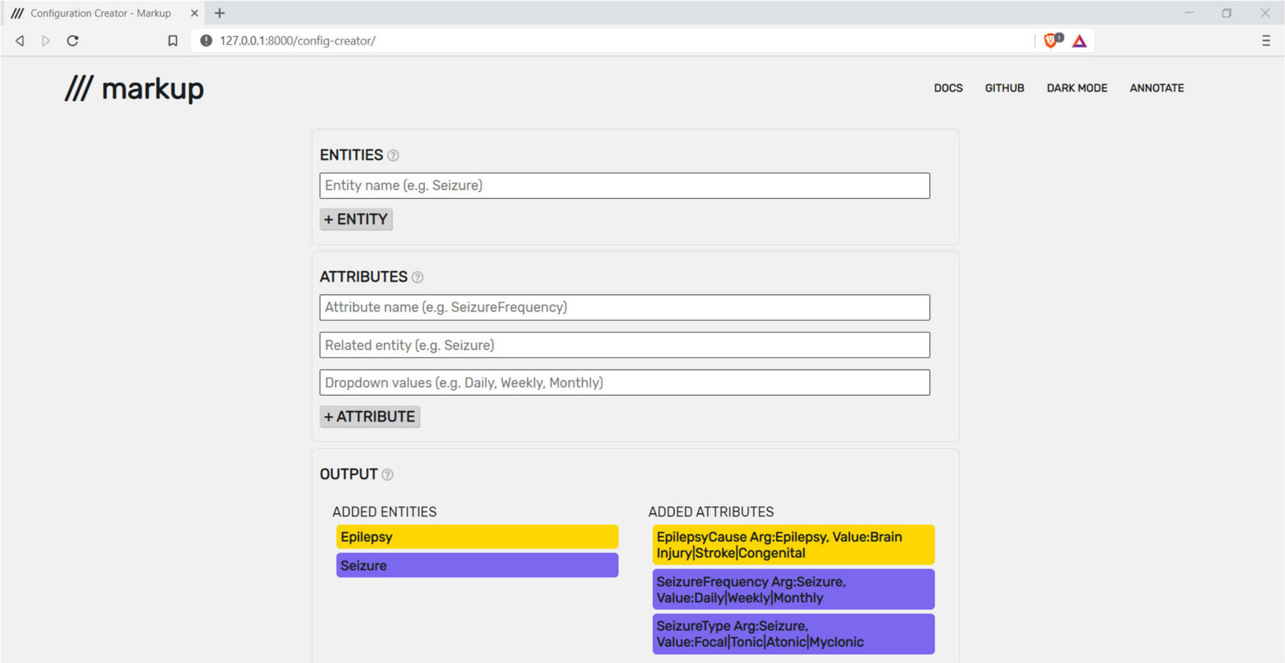
Prescription

sodium valproate

DrugName: SodiumValproate
DrugDose: 500
DoseUnit: mg
Frequency: 2

levetiracetam

FIGURE 2 | A live annotation session in Markup. Note that the document shown is a synthetic letter that has been created by a clinician to reflect real-world clinic text.


Configuration Creator - Markup
127.0.0.1:8000/config-creator/
/// markup
DOCS
GITHUB
DARK MODE
ANNOTATE
ENTITIES
Entity name (e.g. Seizure)
+ ENTITY
ATTRIBUTES
Attribute name (e.g. SeizureFrequency)
Related entity (e.g. Seizure)
Dropdown values (e.g. Daily, Weekly, Monthly)
+ ATTRIBUTE
OUTPUT
ADDED ENTITIES
ADDED ATTRIBUTES

Epilepsy
Seizure

EpilepsyCause Arg:Epilepsy, Value:Brain Injury|Stroke|Congenital
SeizureFrequency Arg:Seizure, Value:Daily|Weekly|Monthly
SeizureType Arg:Seizure, Value:Focal|Tonic|Atonic|Myclonic

FIGURE 3 | Markup's built-in configuration file creator being used to construct a file with epilepsy and seizure entities and attributes.

of attributes. Annotation placement is unaffected by existing annotations, and thus complex data can be captured *via* repeated annotation of the same region.

Given the intricate details required for capture within documents, configuration files can be complex and lengthy, making them difficult to navigate throughout the annotation process. Markup provides an intuitive alternative by dynamically displaying attributes based on the selected entity, thus preventing irrelevant options from appearing in the users' view. Each attribute also allows for free-text input, enabling a user to capture attributes that have not been previously defined in the configuration file.

Adding an annotation will colour the annotated text based on the colour of the corresponding entity and the annotation will be added to the categorised display in the annotation panel, as shown in **Figure 2**. Information related to each annotation (e.g., attribute values) can be viewed by selecting the desired annotation in the annotation panel, or by hovering over an annotation with the cursor. Additional options (e.g. delete or edit an annotation) can also be seen upon selection of an annotation within the annotation panel.

Upon completion of the annotation process, the user can export the annotations into the brat standoff format, thus enabling external NLP software, such as the General Architecture for Text Engineering (12), to make use of annotations produced by Markup.

Predictive Annotation Suggestions

Manually adding an annotation is a time-consuming task across all domains, despite annotations often following predictable formats. For example, when annotating a sentence that contains a prescription, it is likely that the sentence will contain a drug name, dose, unit, and frequency.

During the annotation process, Markup learns to automate the process of identifying, suggesting, and adding complex annotations, with appropriate entities and attributes. Markup will continuously suggest annotations for the document within the users' view (**Figure 4**).

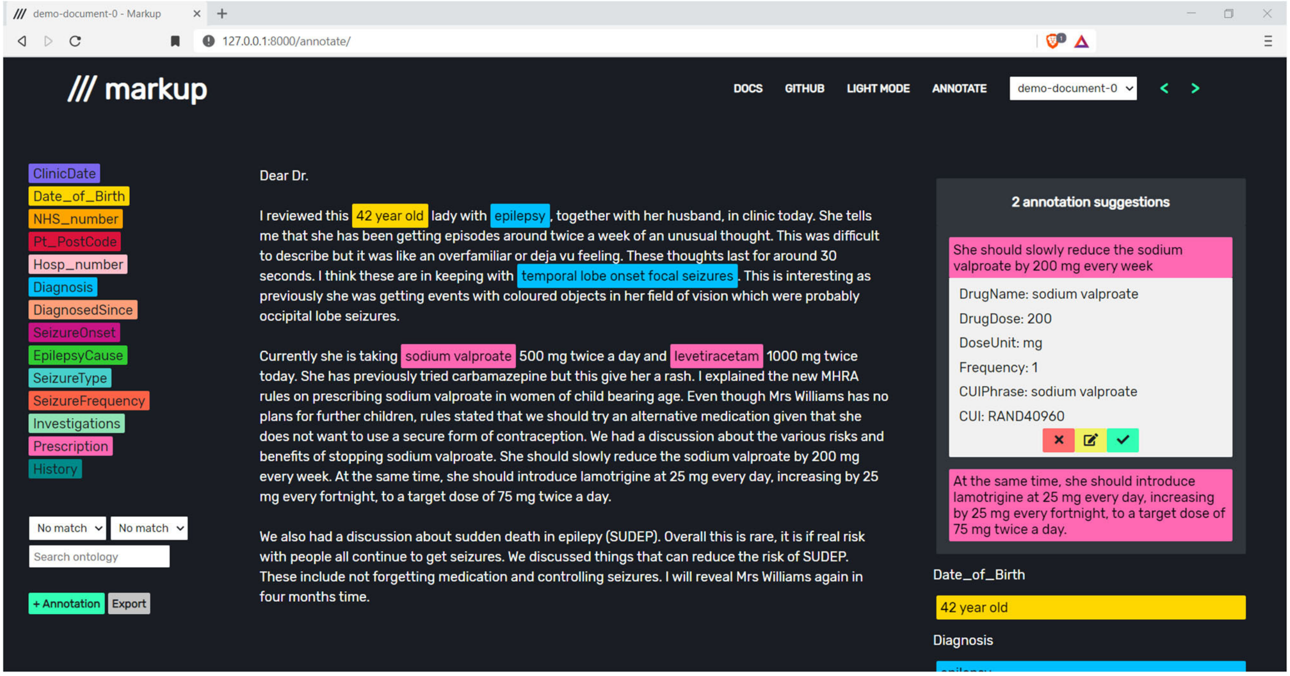
Markup achieves this using active learning (AL) and sequence-to-sequence (Seq2Seq) models. The AL model uses an underlying Random Forest classifier to identify target sentences, such as sentences that contain prescriptions (13). The word-level Long Short-Term Memory (LSTM) based Seq2Seq model converts each target sentence into relevant attributes (14–16). The models are continuously adjusted based on feedback from user interactions with the predicted annotation suggestions (**Figure 5**).

We have demonstrated prediction of prescription-based annotations. Markup is however unrestricted in its annotation predictions and offers functionality that enables users to generate their own synthetic data and train models with no technical expertise.

Automated Ontology Mapping

Misspellings and unconventional wordings often contained within documents make it beneficial to map annotated terms to standardised and recognised versions of a similar term in external ontologies, as shown in **Table 1**.

Markup assists annotators in the production of normalised annotation datasets by providing functionality for automatically



Dear Dr.

I reviewed this 42 year old lady with epilepsy, together with her husband, in clinic today. She tells me that she has been getting episodes around twice a week of an unusual thought. This was difficult to describe but it was like an overfamiliar or deja vu feeling. These thoughts last for around 30 seconds. I think these are in keeping with temporal lobe onset focal seizures. This is interesting as previously she was getting events with coloured objects in her field of vision which were probably occipital lobe seizures.

Currently she is taking sodium valproate 500 mg twice a day and levetiracetam 1000 mg twice today. She has previously tried carbamazepine but this give her a rash. I explained the new MHRA rules on prescribing sodium valproate in women of child bearing age. Even though Mrs Williams has no plans for further children, rules stated that we should try an alternative medication given that she does not want to use a secure form of contraception. We had a discussion about the various risks and benefits of stopping sodium valproate. She should slowly reduce the sodium valproate by 200 mg every week. At the same time, she should introduce lamotrigine at 25 mg every day, increasing by 25 mg every fortnight, to a target dose of 75 mg twice a day.

We also had a discussion about sudden death in epilepy (SUDEP). Overall this is rare, it is if real risk with people all continue to get seizures. We discussed things that can reduce the risk of SUDEP. These include not forgetting medication and controlling seizures. I will reveal Mrs Williams again in four months time.

She should slowly reduce the sodium valproate by 200 mg every week

DrugName: sodium valproate

DrugDose: 200

DoseUnit: mg

Frequency: 1

CUIPhrase: sodium valproate

CUI: RAND40960

At the same time, she should introduce lamotrigine at 25 mg every day, increasing by 25 mg every fortnight, to a target dose of 75 mg twice a day.

Date_of_Birth

42 year old

Diagnosis

FIGURE 4 | Accepting, editing, or rejecting prescription suggestions to store annotations and improve future annotation suggestions. Note that the document shown is a synthetic letter that has been created by a clinician to reflect real-world clinic text.

mapping annotations to standardised terms in both existing ontologies, such as UMLS, and user-defined ontologies. Upon selecting a span of text within Markup, a dropdown list is automatically populated with all relevant matches from the chosen ontology.

To achieve this, the lexical similarity is computed between the span of text being annotated and the terms within the ontology, using Cosine Similarity. To account for misspelt words, terms are broken down into n-chars before computing the similarity. To avoid the computational expense of repeatedly converting all terms to n-chars and computing the similarity between the span of text and all terms within the ontology, SimString has been used (17). With SimString, a database is constructed containing the n-char representations of each term, and an initial filtering is performed upon inputting the selected span of text, with Cosine Similarity then being computed against the remaining terms. The terms that surpass the similarity threshold are then ordered by relevance within the dropdown list by computing the Levenshtein Distance between the terms and the selected span of text.

Whilst the above process occurs automatically each time a user selects a span of text, users also have the option of manually querying the ontology *via* Markup, which will populate the dropdown list based on the input search term. If a mapping is accepted by the user, all data relevant to that item within the ontology, such as its unique identifier, will be associated with the annotation.

Safety and Security

Markup makes use of the Transport Layer Protocol (via HTTP over TLS) to encrypt data that is communicated between the client and Markup server(s). Given the broad range of documents that Markup may be used to annotate, it is recommended that a local installation and server is used when annotating sensitive documents, or where organisational policy dictates that data may not leave a local environment or organisation, to help minimize security risks.

Comparing Active Learning and Manual Annotation

We used Markup's built-in data generator to synthetically generate 50 training samples (medication prescriptions) as a seed for the AL and 50,000 for the Seq2Seq model. The data generator allows users to define multiple terms of interest such as subsets of UMLS concepts, arbitrary strings and number range fields that can be used to define attribute variables within template sentences. These template sentences are then resampled, where attribute values are permuted to generate a set number of training samples. All 50,000 training samples contained target outputs for the Seq2Seq model, whereas the samples for the AL contained no target output, and were simply split into positive and negative samples, with 25 being positive samples (sentences containing prescriptions) and 25 being negative samples (sentences not containing prescriptions). No real-world training data was used to train the AL or Seq2Seq models.

Using the local version of Markup on standalone machines we compared annotation speed within Markup using manual annotation and active learning. Four experienced annotators each annotated the prescriptions (including drug name, dose strength, frequency, and further directions) found within two sets of 25 pseudonymised clinic letters.

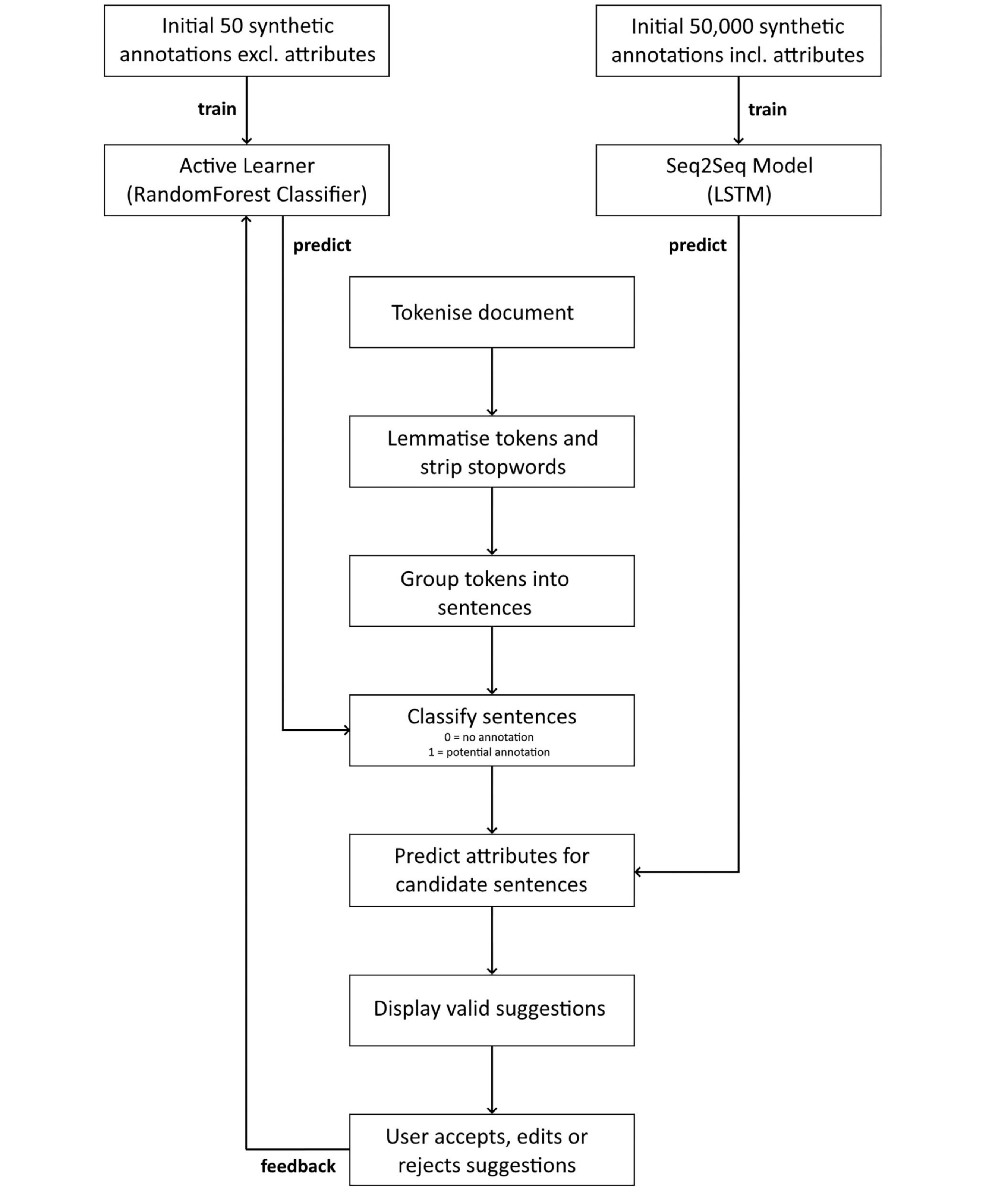

Initial 50 synthetic annotations excl. attributes
train
Active Learner (RandomForest Classifier)
predict
Initial 50,000 synthetic annotations incl. attributes
train
Seq2Seq Model (LSTM)
predict
Tokenise document
Lemmatise tokens and strip stopwords
Group tokens into sentences
Classify sentences
0 = no annotation
1 = potential annotation
Predict attributes for candidate sentences
Display valid suggestions
User accepts, edits or rejects suggestions
feedback

FIGURE 5 | A flow diagram of the annotation suggestion process. Named Entity Recognition is performed using an Active Learner with an underlying Random Forest classifier, and sentence to attribute conversion is performed using an LSTM-based Sequence-to-Sequence model.

One set was manually annotated and the other was annotated using active learning. Annotators 1 and 2 used manual annotation for set 1, and annotators 3 and 4 used active learning suggestions for set 1. The annotators swapped methods for set 2 with annotators 1 and 2 using active learning and annotators 3 and 4 using manual annotation. This method reduced potential bias from differences in annotation difficulty between the sets of letters. Manual annotation is generally quicker when documents have been previously viewed and so annotators used different methods for the different sets of letters.

During the active learning session, annotators could accept, edit, or reject suggestions. Where Markup did not suggest a valid prescription for a prescription phase the annotators were instructed to annotate it manually.

RESULTS

Markup was developed using React and TypeScript (front-end) and Python (back-end), and makes use of the Flask web framework. De-identified and anonymised clinic letters were used throughout Markup's development for validation of functional outputs.

A multi-disciplinary team of annotators, including data analysts and clinicians, provided extensive feedback whilst testing Markup on epilepsy and plastic surgery clinic letters. The captured annotations are being used to build and test real-world NLP applications, such as an updated and expanded version of ExECT (18). Re-annotation, following iterations of annotation definitions, was incorporated for flexibility. UMLS codes, certainty context, and multiple components from complex phrases were captured and exported in a structured format.

TABLE 1 | Mappings of terms, as found in a clinical letter, to standardised versions of the terms in an external ontology.

Annotated term	Mapped ontology term
Focal eeppilepsy	Focal epilepsy
Patient broke their arm	Broken arm
ADR	Adverse drug reaction

Evaluation of Active Learning Annotation Suggestions

The use of active learning annotation suggestions reduced annotation time by 14.2% (11.01 min, 95% CI 0.62–0.87, $p < 0.001$) and increased the mean pairwise inter-annotator F-measure by 0.02 (95% CI 0.05–1, $p = 1$) (**Table 2**).

One annotator tested whether the Active Learner (AL) improved suggestions over time by comparing baseline suggestions for all 25 letters compared to suggestions when given input by the annotator. The AL model reduced the number of incorrect suggestions by 52% ($N = 50$) when compared to baseline suggestions without human intervention ($N = 106$) and marginally increased the number of correct suggestions by 3% ($N = 68$) compared to the baseline ($N = 66$).

An initial system usability survey was run within the annotation team, resulting in an average score of 91. Surveys found at https://github.com/samueldobbie/markup-sus/.

DISCUSSION

Markup is a general-purpose, open-source, web-based annotation tool that offers a range of features to enhance the annotation experience, tackle limitations from existing tools, and reduce annotation time.

Markup has four primary strengths. The first is the inclusion of integrated ontology access and automated ontology mappings. The second is the inclusion of predictive annotation suggestions *via* an Active Learner that improves with user feedback, and a Sequence-to-Sequence (Seq2Seq) model which helps to automate the capture of annotations attributes. The third is that Markup's models can be trained without a pre-annotated corpus, as Markup's data generator enables users to define synthetic templates to produce training data. The fourth is Markup's availability as a web application, enabling annotation without the installation of software components, and as a standalone application *via* a local server.

Markup's active learning reduced annotation time by 14.2% and decreased pairwise inter-annotator F-measure by only 0.02.

TABLE 2 | Comparing manual and predictive annotation.

Method	Letter set	Annotation time (minutes)		Mean time (minutes)	Annotated items N	Pairwise F-measure
		First annotator	Second annotator			
Manual	1	41.24	51.22	46.23	85	0.96
	2	35.36	63	49.18	96	0.89
Predictive	1	32.33	40.46	36.40395	91	0.93
	2	33.00	41	37.00	96	0.88
				Mean decrease in time	14.2% (11.01 min) CI 0.62–0.87, $p < 0.001$	
				Mean decrease in F	0.02 (95% CI 0.05–1, $p = 1$)	

Each letter set contained 25 pseudo-anonymised letters. Each of 4 annotators manually annotated 1 set and used annotation predictions to annotate the other set (see Method section Comparing Active Learning and Manual Annotation). Pairwise F-measure is defined as the average of F-measures between multiple annotators where each annotator is sequentially selected as the gold standard annotator.

The time taken for the Seq2Seq to convert target sentences into annotation attributes is the largest delay in the suggestion process; taking a mean of 20.2 s across 5 random documents, which contain a mean of 55 sentences. Reducing this time will be a priority for ongoing development.

The reduced annotation time, with a negligible decrease in the pairwise inter-annotator F-measure, point to good quality suggestions provided to the user. The Active Learner is, however, sensitive to human input. Incorrect suggestions were reduced by 52%, with user input to the Active Learner, when compared to baseline pre-annotations.

The Neves systematic review identifies accurate pre-annotation as one of the main areas of focus to improve annotation tools (19). To our knowledge, most existing annotation tools that use pre-annotation and active learning based suggestions are limited to Named Entity Recognition (NER) tasks. Users then have the task of manually entering attributes for the suggested entity. Markup's annotation suggestion process automatically suggests attributes using a Seq2Seq model that has been trained on synthetically generated data.

The INCEpTION annotation platform uses Active Learning to predict and assign a label category to an annotation span that has been suggested to the user following NER in the active learning phase (20, 21). INCEpTION achieves this using uncertainty sampling when trained from a gold standard pre-annotated corpus. As Markup only requires synthetically generated sentence templates to begin suggesting complex annotations, it removes the need to pre-annotate a large corpus for training purposes, as users can simply define a set of base templates. ezTag provides an interactive learning mode that does not require training data. It allows the use of pre-trained taggers, or a string-matcher, to suggest NER to users. ezTag is similar to Markup in this respect. However, Markup goes beyond standard NER problems with its use of LSTM. This expands NER recognition into more challenging areas such as sequence mapping, and automated tagging of sub-annotations within the main annotation span.

Whilst Markup has been used successfully on complicated, real-world clinic letters, there are still several limitations that will need to be addressed during future development. The most notable limitation is the lack of user presence within the application, as no option to create a user account currently exists. As such, there is no option to share configurations, documents, annotations, and ontologies within an annotation group as is possible with tools such as Webanno (5).

The automated ontology mappings by Markup are currently limited in that suggestions are based solely upon the lexical similarity between terms, thus if the user were to select an acronym within the document, the full-form phrase is unlikely to be detected as a relevant mapping. To tackle this, future work will be done to suggest mappings based on semantic similarity *via* word embeddings.

Further work is required to expand the number of input and output formats Markup can work with, to streamline the process of integrating Markup into new and existing pipelines. For instance, Markup only works with plaintext files, and is generally aimed toward annotation of short documents such as clinic letters. Future support will include the use of Apache Tika to ingest word, pdf and non-standard document formats commonly found in healthcare such as DICOM image files.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

pseudonymised clinical letters routinely collected from the Swansea Epilepsy clinic were used throughout the development of Markup for testing purposes.

AUTHOR CONTRIBUTIONS

AL and SD conceived, designed, and developed Markup and drafted the manuscript. BF-S, HS, CJ, and WP contributed to data collection and/or data analyses. AA and ST provided resources and technical support. All authors revised the manuscript.

FUNDING

This work was funded by Health Data Research UK, which receives its funding from HDR UK Ltd. (HDR-9006) funded by the UK Medical Research Council, Engineering and Physical Sciences Research Council, Economic and Social Research Council, Department of Health and Social Care (England), Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Health and Social Care Research and Development Division (Welsh Government), Public Health Agency (Northern Ireland), British Heart Foundation (BHF), and the Wellcome Trust.

REFERENCES

1. Kim JD, Ohta T, Pyysalo S, Kano Y, Tsujii JI. Overview of BioNLP'09 shared task on event extraction. In: *Proceedings of the BioNLP 2009 Workshop Companion Volume for Shared Task*. Boulder (2009). p. 1–9.
2. Stenetorp P, Pyysalo S, Topić G, Ohta T, Ananiadou S, Tsujii JI. BRAT: a web-based tool for NLP-assisted text annotation. In: *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics*. Avignon (2012). p. 102–7.
3. South B, Shen S, Leng J, Forbush T, DuVall S, Chapman W. A prototype tool set to support machine-assisted annotation. In: *BioNLP: Proceedings of the 2012 Workshop on Biomedical Natural Language Processing*. Montreal, QC (2012). p. 130–9.
4. Ogren Philip VK. A protégé plug-in for annotated corpus construction. In: *Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics (HLT-NAACL)* (2006).
5. Yimam SM, Gurevych I, de Castilho RE, Biemann C. Webanno: a flexible, web-based and visually supported system for distributed annotations. In:

Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics: System Demonstrations. Sofia (2013). p. 1–6.

6. Islamaj R, Kwon D, Kim S, Lu Z. TeamTat: a collaborative text annotation tool. *Nucleic Acids Res.* (2020) 48:W5–11. doi: 10.1093/nar/gkaa333
7. Pérez-Pérez M, Glez-Peña D, Fdez-Riverola F, Lourenço A. Marky: a tool supporting annotation consistency in multi-user and iterative document annotation projects. *Comput Methods Programs Biomed.* (2015) 118:242–51. doi: 10.1016/j.cmpb.2014.11.005
8. Gobbel GT, Garvin J, Reeves R, Cronin RM, Heavirland J, Williams J, et al. Assisted annotation of medical free text using RapTAT. *J Am Med Informatics Assoc.* (2014) 21:833–41. doi: 10.1136/amiajnl-2013-002255
9. Kwon D, Kim S, Wei CH, Leaman R, Lu Z. ezTag: tagging biomedical concepts via interactive learning. *Nucleic acids research.* (2018) 46:W523–9. doi: 10.1093/nar/gky428
10. Chen WT, Styler W. Anafora: a web-based general purpose annotation tool. In: *Proceedings of the conference. Association for Computational Linguistics. North American Chapter. Meeting, Vol. 2013.* Atlanta: NIH Public Access (2013). p. 14.
11. Kuo TT, Huh J, Kim J, El-Kareh R, Singh S, Feupe SF, et al. The impact of automatic pre-annotation in Clinical Note Data Element Extraction-the CLEAN Tool. *arXiv preprint arXiv:*1808.03806 (2018).
12. Cunningham H. GATE, a general architecture for text engineering. *Comput Human.* (2002) 36:223–54. doi: 10.1023/A:1014348124664
13. Breiman L. Random forests. *Machine Learn.* (2001) 45:5–32. doi: 10.1023/A:1010933404324
14. Sutskever I, Vinyals O, Le QV. Sequence to sequence learning with neural networks. *arXiv preprint arXiv:*1409.3215 (2014).
15. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. *Nature.* (1986) 323:533–6. doi: 10.1038/323533a0
16. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput.* (1997) 9:1735–80. doi: 10.1162/neco.1997.9.8.1735
17. Okazaki N, Tsujii JI. Simple and efficient algorithm for approximate dictionary matching. In: *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010).* (2010). p. 851–9.
18. Fonferko-Shadrach B, Lacey AS, Roberts A, Akbari A, Thompson S, Ford DV, et al. Using natural language processing to extract structured epilepsy data from unstructured clinic letters: development and validation of the ExECT (extraction of epilepsy clinical text) system. *BMJ Open.* (2019) 9:e023232. doi: 10.1136/bmjopen-2018-023232
19. Neves M, Leser U. A survey on annotation tools for the biomedical literature. *Briefings Bioinformatics.* (2014) 15:327–40. doi: 10.1093/bib/bbs084
20. Klie JC, Bugert M, Boullosa B, de Castilho RE, Gurevych I. The inception platform: machine-assisted and knowledge-oriented interactive annotation. In: *Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations.* (2018). p. 5–9.
21. de Castilho RE, Ide N, Kim JD, Klie JC, Suderman K. Towards cross-platform interoperability for machine-assisted text annotation. *Genomics Informatics.* (2019) 17:e19. doi: 10.5808/GI.2019.17.2.e19

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.