Supplementary Material for

Biomedical Text Mining for Research Rigor and Integrity: Tasks, Challenges, Directions

Halil Kilicoglu

April 10, 2017

Contents

S1. CONSORT checklist

Table 1 shows the CONSORT (Schulz *et al.*, 2010) checklist items that apply to Methods sections of randomized clinical trial (RCT) papers.

Topic	**Checklist item**
Trial design	Description of trial design (such as parallel, factorial) including allocation ratio
	Important changes to methods after trial commencement (such as eligibility criteria), with reasons
Participants	Eligibility criteria for participants
	Settings and locations where the data were collected
Interventions	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered
Outcomes	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed
	Any changes to trial outcomes after the trial commenced, with reasons
Sample size	How sample size was determined
	When applicable, explanation of any interim analyses and stopping guidelines
Randomisation: Sequence generation	Method used to generate the random allocation sequence
	Type of randomisation; details of any restriction (such as blocking and block size)
Randomisation: Allocation concealment mechanism	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned
Randomisation: Implementation	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions
Blinding	If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how
	If relevant, description of the similarity of interventions
Statistical methods	Statistical methods used to compare groups for primary and secondary outcomes
	Methods for additional analyses, such as subgroup analyses and adjusted analyses

Table 1: CONSORT (Schulz *et al.*, 2010) checklist items that apply to Methods sections of clinical trials papers

S2. Illustration of conceptual, relational, and contextual semantic levels

(1)	*The EP2 receptor for prostaglandin E2 (PGE2) is a membrane receptor that mediates at least part of the action of PGE2. It has been shown that EP2* ***plays a critical role***$_{r_1}$ *in tumorigenesis* ***in***$_{r_2}$ *mouse mammary gland and colon. However, the possibility that the EP2 receptor is* ***involved***$_{r_3}$ *in the development of skin tumors was unknown. The purpose of this study was to investigate the* ***role***$_{r_4}$ *of the EP2 receptor in mouse skin carcinogenesis. . . .* *Tumors from WT mice produced more blood vessels and fewer apoptotic cells than those of EP2 knockout mice as determined by immunohistochemical staining. Our data suggest that the EP2 receptor* ***plays a significant role***$_{r_5}$ *in the protumorigenic action of PGE2 in skin tumor development.*		

(2)	*EP2 receptor*
	EP2
	prostaglandin E2
	PGE2
	tumorigenesis
	mouse
	mouse mammary gland
	colon
	skin carcinogenesis
	skin tumors
		`NCBI Gene:19217:Ptger2:Gene`
		`NCBI Gene:19217:Ptger2:Gene`
		`MESH:D015232:Dinoprostone:Chemical`
		`MESH:D015232:Dinoprostone:Chemical`
		`MEDIC:D002471:Cell Transformation, Neoplastic:Disease`
		`NCBI Taxonomy:10090:Mus musculus:Species`
		`UMLS:C1512980:Mouse Mammary Gland:Tissue`
		`UMLS:C1522281:Mouse Colon:Tissue`
		`UMLS:C1519346:Skin Carcinogenesis:Neoplastic Process`
		`MEDIC:D012878:Skin Neoplasms:Disease`	

(3)	***plays a critical role***
	in
	involved
	role
	plays a significant role
		r_1: `19217-AFFECTS-D002471`
		r_2: `C1512980-LOCATION_OF-D002471`
		r_3: `19217-ASSOCIATED_WITH-D012878`
		r_4: `19217-ASSOCIATED_WITH-C1519346`
		r_5: `19217-AFFECTS-D012878`
			`Fact, Prior`
			`Uncommitted, Current`
			`Fact, Current`
			`Uncommitted, Current`
			`Probable, Current`

Table 2: Abstract of a PubMed article (PMID 16230392), with examples of annotations corresponding various levels of semantic information.

For illustration of conceptual, relational, and contextual semantic levels, consider the fragment of the abstract of a PubMed article (PMID 16230392), shown in row (1) of Table 2. Some of the biomedical entities (the conceptual level) are shown in row (2) and are underlined in text. In row (2), textual mentions of these entities are in the second column, and the corresponding concepts are shown as `Nomenclature:Identifier:Official Symbol:Semantic Type` tuples in the third column. The mentions *EP2* and *EP2 receptor* are normalized to the same concept (with the unique identifier `19217` in the NCBI Gene database). Such concept normalization accounts for acronyms (*prostaglandin E2* vs. *PGE2*) as well as inflectional/derivational forms (e.g., *skin tumors* vs. *skin tumor*) and synonymy.

The relational (or *propositional*) level is illustrated in row (3). In this case, relationships are represented as subject-predicate-object triples in the third column. Predicate types (`AFFECTS`, `ASSOCIATED_WITH`, etc.) are defined in the UMLS Semantic Network and the subject and object arguments refer to normalized entities in row (2). The phrases indicating the predicates are in bold (in row (1) and in row (3) second column) and have the relation identifiers as subscripts in row (1). Here, we focus on relationships from second and third sentences as well as the last sentence. The first relation shown (r_1) corresponds to the proposition that *EP2 affects tumorigenesis*, expressed in the second sentence. The main knowledge claim of the abstract under discussion (sometimes referred to as *claimed knowledge updates* (Sándor and de Waard, 2012)) is expressed in the last sentence and captured by the relation r_5 in row (3). Other representational formalisms, such as simple binary relations (no predicate) and nested event representations, are sometimes also used to represent this semantic level.

The next level in semantic interpretation is the contextual level (or *extra-propositional*), illustrated in the last column of row (3). Two dimensions are considered for illustration: *factuality* and *source of evidence*. The first relation (r_1: proposition that *EP2 affects tumorigenesis*) is represented as `Fact` and attributed to `Prior` knowledge (inferred from the clause *It has been shown that*). The relation corresponding to the research question

(r_4: the proposition that *EP2 has a role in skin carcinogenesis*) is represented as `Uncommitted` (due to *The purpose of this study was to investigate*). Interpretation at this level helps us represent the unfolding of the argumentation within the abstract, culminating with the main knowledge claim represented with the relation r_5 (the proposition that *EP2 plays a role in skin tumors*), assigned the factuality value `Probable` and attributed to `Current` evidence (based on the clause *Our data suggest that*). For brevity, we have not shown the sentences in the middle section of the abstract or their interpretation. Extracting entities/relations in this section and resolving coreference indicated by the phrase *Our data*, it could be possible to link the relations expressing the experimental results to the main knowledge claim (r_5) as evidence (rather than the coarse-grained `Current` shown in Table 2); in effect, creating a document-level argumentation graph.

S3. Named entity recognition and normalization

Named entity recognition and normalization tasks focus on the conceptual level. Named entity recognition (NER) is the task of identifying mentions of a specific semantic class in free text. Named entity normalization (NEN) refers to mapping these mentions to specific entries (concepts) defined in an ontology or knowledge-base. It is a foundational task that has seen much activity since the early days of bioNLP. One of the earliest and best-known systems is MetaMap (Aronson and Lang, 2010), a knowledge-based, broad-coverage system that maps free biomedical text to concepts in the UMLS Metathesaurus (Lindberg *et al.*, 1993). MetaMap pre-processing consists of tokenization of the input text, sentence segmentation, acronym/abbreviation detection, and part-of-speech tagging with MedPost tagger (Smith *et al.*, 2004). Then, a minimal commitment parser breaks text into noun phrases and identifies their heads. The UMLS SPECIALIST Lexicon (McCray *et al.*, 1994) is used to generate lexical variants which are then mapped to candidate strings in Metathesaurus. The best mapping is identified by combining four criteria: centrality, variation, coverage, and cohesiveness. MetaMap can also perform word sense disambiguation, determining, for example, which UMLS concept the textual mention *cold* refers to: `Cold Temperature`, `Common Cold`, or `Cold Sensation`. MetaMap is highly configurable, with numerous data, processing, and output options.

While MetaMap aims to maps all types of biomedical text to UMLS concepts, many systems have aimed to recognize specific semantic classes (e.g., diseases, chemicals/drugs, gene/proteins) and map them to more focused vocabularies. Unlike MetaMap, such systems are often trained on corpora annotated for specific classes using machine learning techniques. For example, DNorm (Leaman *et al.*, 2013) was based on the NCBI disease corpus (Doğan *et al.*, 2014) and targeted the MEDIC disease vocabulary (Davis *et al.*, 2012). They performed NER using the BANNER system (Leaman and Gonzalez, 2008), based on Conditional Random Fields (CRF). They retrained the system on the NCBI disease corpus and enhanced it with abbreviation detection. The normalization step was based on the pairwise learning-to-rank method, which found the best match between a mention and the disease names in MEDIC, by converting both into vectors and searching for the disease name that maximizes a scoring function learned from the training data. Their method achieved an F_1 score of 0.78, compared to 0.57 of MetaMap. In a similar vein, tmChem (Leaman *et al.*, 2015) focused on chemical/drug name recognition. It was trained on the CHEMDNER corpus (Krallinger *et al.*, 2015) using an ensemble of CRF models and mapped recognized entities to MeSH (Coletti and Bleich, 2001) and ChEBI (de Matos *et al.*, 2010) vocabularies, reaching 0.87 F_1 score in abstract-level evaluation. PubTator (Wei *et al.*, 2013a) is a biocuration tool that provides a common user interface and API for NER/NEN on PubMed abstracts. It incorporates DNorm and tmChem as well as tools that consider other entity types (species (Wei *et al.*, 2012), gene/proteins (Wei *et al.*, 2015), mutations (Wei *et al.*, 2013b)). Unlike these systems, TaggerOne (Leaman and Lu, 2016) is not limited to a specific entity type, although it requires an annotated corpus for training. It employs a semi-Markov structured linear classifier, which combines a feature-based approach for NER and supervised semantic indexing for normalization. They achieved results superior to DNorm and tmChem in disease and chemical/drugs recognition (0.81 vs. DNorm's 0.78 for diseases, 0.90 vs. tmChem's 0.87 for chemical/drugs).

S4. Relation extraction

Relation extraction is the task of extracting specific types of biomedical relationships from free text. It generally builds on NER/NEN and can be viewed as the basis of identifying knowledge claims.

Representation formalisms

Different formalisms have been used to represent biomedical relations. In the simplest case, relationships are represented as binary relations, where the type of the relationship is implicit or underspecified. This representation has been used in extraction of protein-protein interactions (Tikk *et al.*, 2010) and gene-disease associations (Özgür *et al.*, 2008), for example. More informative is the subject-predicate-object triple representation, where the predicate indicates a named relationship and the subject/object pair (arguments) indicates the entities involved in the relationship. The relationship can be unidirectional (e.g., causal) or bidirectional (e.g., interaction)[1].

A more generalized representation is the predicate-argument structure representation, used particularly in representing biological events expressed in biomedical literature. In this representation, an event, triggered by a predicate, can have one or more arguments (participants), with specific roles (`Theme`, `Cause`, etc.) and it can be nested (i.e., it can have other events as participants). The example below, taken from the GENIA event corpus (Kim *et al.*, 2008), shows that the sentence contains two events (triggered by *stimulates* and *production*), and the theme of the simulation event (shown as e_1) is the *interleukin-10 production* event (shown as e_2), taking place in *human monocytes*.

(1) (a) *Our previous results show that recombinant gp41 ...* ***stimulates***$_{e_1}$ *interleukin-10* ***production***$_{e_2}$ *in human monocytes.*

(b) *stimulates*: `Positive_Regulation`(e_1, `Cause`:*gp41*: `Protein`, `Theme`:e_2)
production: `Gene_Expression`(e_2, `Theme`:*interleukin-10*: `Protein`, `Site`:*human monocytes*:`Cell`)

Nanopublications (Mons and Velterop, 2009), a life sciences Semantic Web community effort, represent relations using a richly annotated RDF triple format, incorporating relation context (e.g., under what conditions) and provenance (e.g., which article it is attributed to). Nanopublications have been proposed as a general solution for storing and exchanging scientific knowledge, so their applicability goes beyond relation extraction tasks. An example nanopublication is shown in Figure 1. It represents the assertion that *inhibition of mTOR by rapamycin can slow or block AD progression in a transgenic mouse model of the disease.*

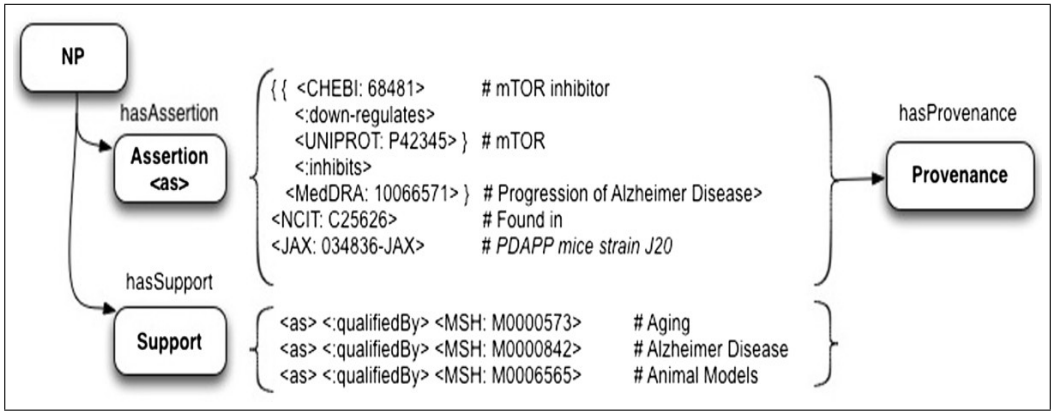

Figure 1: An example nanopublication (Image source: Clark *et al.* (2014); use permitted under the Creative Commons Attribution License CC BY 4.0).

There are other efforts concerned with relation representation, such as Biological Expression Language (BEL) (Fluck *et al.*, 2015), focusing on causal and correlative relationships in life sciences. There have been

[1]Subject-predicate-object representation is analogous to RDF triples used in Semantic Web technologies.

attempts to bridge different representation formalisms, as well; for example, Fluck *et al.* (2013) derived BEL statements from event representations. It is also worth noting that in recent years, Abstract Meaning Representation (AMR) (Banarescu *et al.*, 2013) has become increasingly popular as a deep semantic representation formalism in open-domain NLP, partly due to the existence of corpora annotated with it, and it is likely to inform relation representation formalisms in biomedical NLP, as well.

Systems

Let us now look more closely at systems that aim to extract relationships that can be expressed with such representations. Most relation extraction systems focus on binary relationships (e.g., protein-protein interactions (Tikk *et al.*, 2010), drug-drug interactions (Ben Abacha *et al.*, 2015), chemical-induced disease relationships (Xu *et al.*, 2016)). Such systems are trained on corpora annotated specifically for the relationship of interest and use some form of supervised machine learning, often incorporating n-gram, entity, relation trigger features, and syntactic information, generally in the form of dependency parses. Entities involved in relations are sometimes normalized, but not always. Community challenges, such as BioCreative (Hirschman *et al.*, 2005), the BioNLP Shared Task on event extraction (Kim *et al.*, 2009), and the DDIExtraction shared task (Segura-Bedmar *et al.*, 2011), have often provided the stimulus for development of such systems. Below, we describe in more detail several relation extraction systems that have wider coverage. We refer the reader to a recent survey (Luo *et al.*, 2016) for an in-depth discussion of biomedical relation extraction systems.

SemRep (Rindflesch and Fiszman, 2003) incorporates MetaMap as the NER/NEN tool and has a similar philosophy: it is knowledge-based, uses UMLS for domain knowledge, and attempts to extract a wide range of relationships. It uses a subject-predicate-object triple representation. It focuses on various aspects of biomedicine, from clinical medicine (e.g., `TREATS`, `DIAGNOSES`) to molecular interactions (e.g., `STIMULATES`, `INHIBITS`), pharmacogenomics (e.g., `AUGMENTS`, `DISRUPTS`), and genetic basis of disease (e.g., `CAUSES`, `PREDISPOSES`), as well as some types of static relations (e.g., `ISA`, `PART_OF`). Hand-crafted indicator rules are used to map lexical and syntactic triggers to these predicates. Triggers include lexical categories, such as verbs and nominalizations, and syntactic constructions, such as appositives. Syntactic and semantic argument constraints are applied to determine the subject and object arguments of a predicate. Semantic constraints are based on the UMLS Semantic Network, which defines allowable high-level relations between biomedical entity types (e.g., `Pharmacologic Substance-TREATS-Disease or Syndrome`). Evaluations of SemRep have mainly considered precision, which was found to be in 0.58-0.8 range. Recall, in the few cases it was computed, was found to be lower (0.5-0.64).

Another somewhat general system is the Turku Event Extraction System (TEES) (Björne and Salakoski, 2011), a supervised machine learning-based system which adopts the predicate-argument structure representation. First developed for the BioNLP event extraction shared task (Kim *et al.*, 2009) to extract 8 biological event types (e.g., `Gene_Expression`, `Phosphorylation`, `Regulation`), it has since been generalized to other event types (e.g., `Acetylation`, `Catalysis`), and some static relations (e.g., `Subunit-Complex`, `Protein-Component`). Unlike SemRep, TEES uses full syntactic structure, in the form of dependency parses. Their core system consists of several linear SVM classifiers for relation trigger and argument detection and uses a rich set of features, including token and sentence features, trigger features, and features based on syntactic dependency chains between the relation trigger and potential arguments. This core method has been retrained on several corpora, often yielding state-of-the-art performance. Gene name normalization has been incorporated into TEES (Van Landeghem *et al.*, 2013). EventMine (Miwa *et al.*, 2012) is another machine learning-based system first developed for the BioNLP event extraction task, and then generalized by training over several different corpora, yielding good performance.

Relation extraction targeting other representations is relatively rare, though gene-disease associations have been converted to nanopublications (Queralt-Rosinach *et al.*, 2016), and extraction of BEL statements from the literature was studied in a recent BioCreative task (Rinaldi *et al.*, 2016). The top-ranking system in this task yielded an F_1 score of 0.2 (Choi *et al.*, 2016), indicating that systems addressing complex relation formalisms need significant improvements for practical use.

Limitations

One common limitation of relation extraction systems is that they, predominantly, consider only sentence-bound relationships and are, therefore, unable to extract cross-sentence relationships, expressed through coreference and sometimes implicitly. Two examples from the CDR corpus of chemically-induced disease relationships (Wei *et al.*, 2016) are shown below.

(2) (a) *The current best treatment for HCV infection is combination therapy with pegylated interferon and ribavirin. Although this regimen produces sustained virologic responses (SVRs) in approximately 50% of patients, it can be* ***associated with*** *a potentially dose-limiting hemolytic anemia.*

(b) *We investigated the efficacy and toxicity of a 3-hour paclitaxel infusion in a phase II trial in patients with inoperable stage IIIB or IV NSCLC. . . . Grade 1 or 2 polyneuropathy affected 56% of patients . . .*

Example (2a) shows a cross-sentence causal relationship between the drug combination *pegylated interferon-ribavirin* and *hemolytic anemia*. Recognizing that anaphoric expressions *it* and *this regimen* (double-underlined) co-refer with the drug combination may help in extracting this causal relationship, since the pronoun *it* is a syntactic argument of the causal trigger *associated with*. In Example (2b), the causal relationship between *paclitaxel* and *polyneuropathy* is implicit and can be inferred from the temporal ordering of two events discussed: the administration of *paclitaxel* and the appearance of the adverse effect *polyneuropathy*. Coreference resolution has been explored to aid event extraction (Yoshikawa *et al.*, 2011; Miwa *et al.*, 2012; Kilicoglu and Bergler, 2012; Lavergne *et al.*, 2015; Choi *et al.*, 2016) and its effect has been largely positive, though generally not very significant. Kilicoglu *et al.* (2016) incorporated sortal anaphora resolution (a specific type of anaphora signaled by noun phrases, such as *this regimen* in the example above) into SemRep, enhancing the precision and specificity of its relations. Coreference resolution in PubMed abstracts has also been explored, independently of relation extraction, as one of the subtasks in the BioNLP 2011 Shared Task on event extraction (Kim *et al.*, 2012). Kilicoglu and Demner-Fushman (2016) reported the best performance on the corpus used in this task (F_1 score of 0.68). With regards to implicit relations, Kilicoglu (2016) reported a system that focused exclusively on cross-sentence relationships in the CDR corpus. With a set of lexical, semantic, discourse, and external knowledge features, Kilicoglu achieved an F_1 score of 0.74.

S5. A micropublication example

An example micropublication is shown in Figure 2 below. The micropublication (`MP2`) represents the claim that

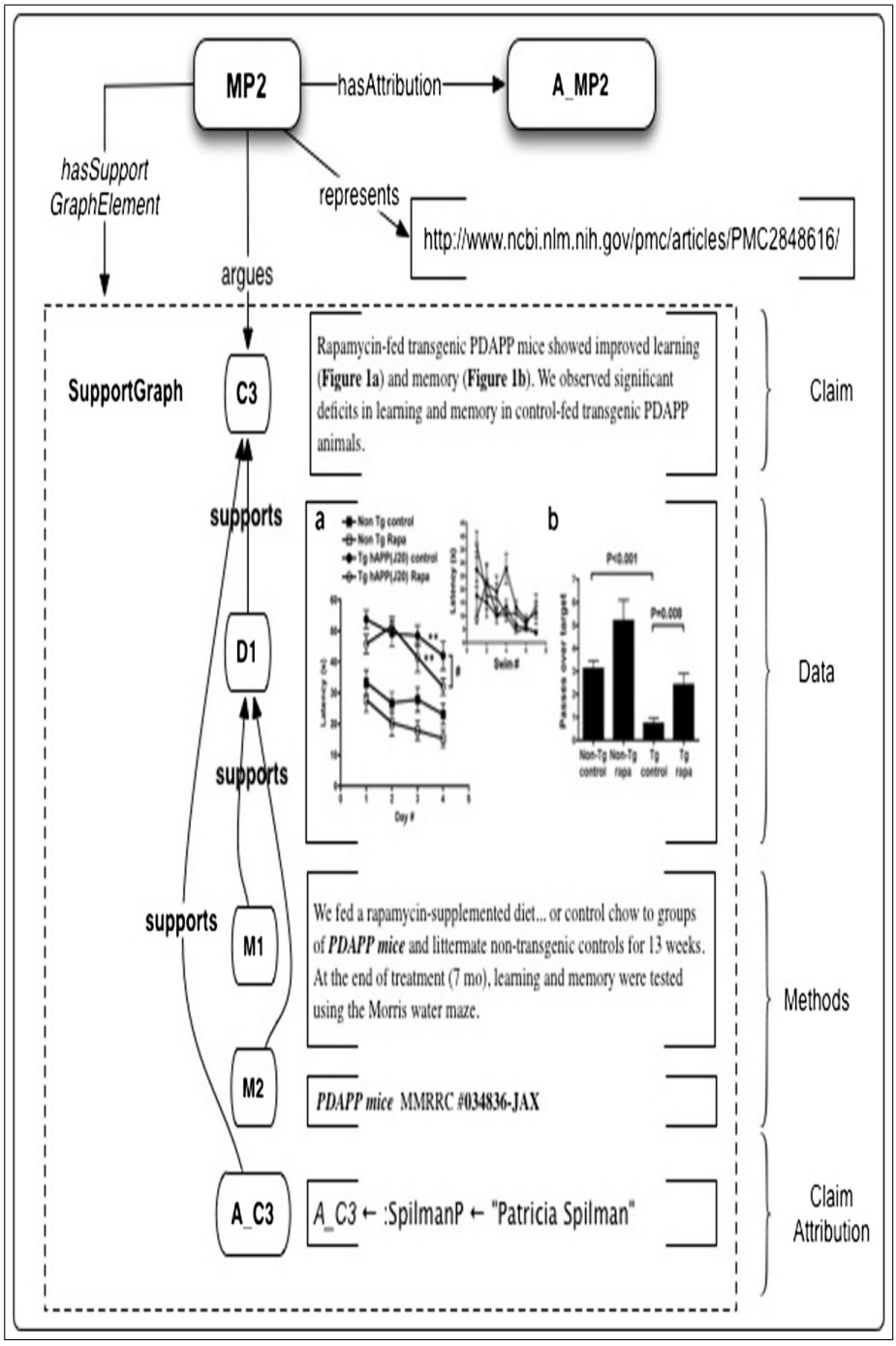

Figure 2: An example micropublication (Image source: Clark *et al.* (2014); use permitted under the Creative Commons Attribution License CC BY 4.0).

Rapamycin-fed transgenic PDAPP mice showed improved learning and memory (`C3`) (`MP2-argues-C3`). The provenance of the micropublication is the PMC article with the identifier PMC2848616 (`represents`). The claim is supported by the two figures (`D1-supports-C3`). The figures are the result of *learning and memory testing using the Morris water maze*, indicated in Methods (`M1-supports-D1`). `M2` indicates a normalized form for the transgenic mouse strain that was used. The claim is attributed to Patricia Spilman, the author of the paper (`A_C3-supports-C3`).

References

Aronson, A. R. and Lang, F.-M. (2010). An overview of MetaMap: historical perspective and recent advances. *Journal of the American Medical Informatics Association (JAMIA)*, **17**(3), 229–236.

Banarescu, L., Bonial, C., Cai, S., Georgescu, M., Griffitt, K., Hermjakob, U., Knight, K., Koehn, P., Palmer, M., and Schneider, N. (2013). Abstract meaning representation for sembanking. In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, pages 178–186.

Ben Abacha, A., Chowdhury, F., Karanasiou, A., Messai, R., Lavelli, A., and Zweigenbaum, P. (2015). Text mining for pharmacovigilance: Using machine learning for drug name recognition and drug-drug interaction extraction and classification. *Journal of Biomedical Informatics*, pages 122–132.

Björne, J. and Salakoski, T. (2011). Generalizing Biomedical Event Extraction. In *Proceedings of BioNLP Shared Task 2011 Workshop*, pages 183–191. Association for Computational Linguistics.

Choi, M., Liu, H., Baumgartner, W., Zobel, J., and Verspoor, K. (2016). Coreference resolution improves extraction of biological expression language statements from texts. *Database*, **2016**, baw076.

Clark, T., Ciccarese, P., and Goble, C. (2014). Micropublications: a semantic model for claims, evidence, arguments and annotations in biomedical communications. *Journal of Biomedical Semantics*, **5**(1), 28.

Coletti, M. H. and Bleich, H. L. (2001). Medical subject headings used to search the biomedical literature. *Journal of the American Medical Informatics Association*, **8**(4), 317–323.

Davis, A. P., Wiegers, T. C., Rosenstein, M. C., and Mattingly, C. J. (2012). MEDIC: a practical disease vocabulary used at the comparative toxicogenomics database. *Database*, **2012**, bar065.

de Matos, P., Dekker, A., Ennis, M., Hastings, J., Haug, K., Turner, S., and Steinbeck, C. (2010). ChEBI: a chemistry ontology and database. *Journal of Cheminformatics*, **2**(Suppl 1), P6.

Doğan, R. I., Leaman, R., and Lu, Z. (2014). NCBI disease corpus: A resource for disease name recognition and concept normalization. *Journal of Biomedical Informatics*, **47**, 1–10.

Fluck, J., Klenner, A., Madan, S., Ansari, S., Bobic, T., Hoeng, J., Hofmann-Apitius, M., and Peitsch, M. (2013). BEL Networks Derived from Qualitative Translations of BioNLP Shared Task Annotations. In *Proceedings of the 2013 Workshop on Biomedical Natural Language Processing*, pages 80–88.

Fluck, J., Madan, S., Ellendorff, T. R., Mevissen, T., Clematide, S., van der Lek, A., and Rinaldi, F. (2015). Track 4 Overview: Extraction of Causal Network Information in Biological Expression Language (BEL). In *BioCreative V*, Proceedings of the Fifth BioCreative Challenge Evaluation Workshop, pages 333–346.

Hirschman, L., Yeh, A., Blaschke, C., and Valencia, A. (2005). Overview of BioCreAtIvE: critical assessment of information extraction for biology. *BMC Bioinformatics*, **6 Suppl 1**.

Kilicoglu, H. (2016). Inferring implicit causal relationships in biomedical literature. In *Proceedings of the 15th Workshop on Biomedical Natural Language Processing*, pages 46–55.

Kilicoglu, H. and Bergler, S. (2012). Biological Event Composition. *BMC Bioinformatics*, **13 (Suppl 11)**, S7.

Kilicoglu, H. and Demner-Fushman, D. (2016). Bio-SCoRes: A Smorgasbord Architecture for Coreference Resolution in Biomedical Text. *PLOS ONE*, **11**(3), 1–38.

Kilicoglu, H., Rosemblat, G., Fiszman, M., and Rindflesch, T. C. (2016). Sortal anaphora resolution to enhance relation extraction from biomedical literature. *BMC Bioinformatics*, **17**(1), 163.

Kim, J.-D., Ohta, T., and Tsujii, J. (2008). Corpus annotation for mining biomedical events from literature. *BMC Bioinformatics*, **9**, 10.

Kim, J.-D., Ohta, T., Pyysalo, S., Kano, Y., and Tsujii, J. (2009). Overview of BioNLP'09 Shared Task on Event Extraction. In *Proceedings of Natural Language Processing in Biomedicine (BioNLP) NAACL 2009 Workshop*, pages 1–9.

Kim, J.-D., Nguyen, N., Wang, Y., Tsujii, J., Takagi, T., and Yonezawa, A. (2012). The Genia Event and Protein Coreference tasks of the BioNLP Shared Task 2011. *BMC Bioinformatics*, **13(Suppl 11)**, S1.

Krallinger, M., Rabal, O., Leitner, F., Vazquez, M., Akhondi, S., Kors, J., Salgado, D., Lu, Z., Leaman, R., Lu, Y., Ji, D., Lowe, D. M., Batista-Navarro, R., Rak, R., Huber, T., Rocktäschel, T., Campos, D., and et al. (2015). The CHEMDNER corpus of chemicals and drugs and its annotation principles. *Journal of Cheminformatics*, **7**(Suppl 1), S2.

Lavergne, T., Grouin, C., and Zweigenbaum, P. (2015). The contribution of co-reference resolution to supervised relation detection between bacteria and biotopes entities. *BMC Bioinformatics*, **16 (Suppl 10)**, S6.

Leaman, R. and Gonzalez, G. (2008). BANNER: an executable survey of advances in biomedical named entity recognition. *Pacific Symposium on Biocomputing.*, pages 652–663.

Leaman, R. and Lu, Z. (2016). TaggerOne: joint named entity recognition and normalization with semi-Markov Models. *Bioinformatics*, **32**(18), 2839–2846.

Leaman, R., Dogan, R. I., and Lu, Z. (2013). DNorm: disease name normalization with pairwise learning to rank. *Bioinformatics*, **29**(22), 2909–2917.

Leaman, R., Wei, C.-H., and Lu, Z. (2015). tmChem: a high performance approach for chemical named entity recognition and normalization. *Journal of Cheminformatics*, **7**(S-1), S3.

Lindberg, D. A. B., Humphreys, B. L., and McCray, A. T. (1993). The Unified Medical Language System. *Methods of Information in Medicine*, **32**, 281–291.

Luo, Y., Uzuner, Ö., and Szolovits, P. (2016). Bridging semantics and syntax with graph algorithms state-of-the-art of extracting biomedical relations. *Briefings in Bioinformatics*, **18**(1), 160–178.

McCray, A. T., Srinivasan, S., and Browne, A. C. (1994). Lexical methods for managing variation in biomedical terminologies. In *Proceedings of the 18th Annual Symposium on Computer Applications in Medical Care*, pages 235–239.

Miwa, M., Thompson, P., and Ananiadou, S. (2012). Boosting automatic event extraction from the literature using domain adaptation and coreference resolution. *Bioinformatics*, **28**(13), 1759–1765.

Mons, B. and Velterop, J. (2009). Nano-Publication in the e-Science era. In T. Clark, J. S. Luciano, M. S. Marshall, E. Prud'hommeaux, and S. Stephens, editors, *Proceedings of the Workshop on Semantic Web Applications in Scientific Discourse (SWASD 2009)*.

Özgür, A., Vu, T., Erkan, G., and Radev, D. R. (2008). Identifying gene-disease associations using centrality on a literature mined gene-interaction network. *Bioinformatics*, **24**(13), i277–i285.

Queralt-Rosinach, N., Kuhn, T., Chichester, C., Dumontier, M., Sanz, F., and Furlong, L. I. (2016). Publishing DisGeNET as nanopublications. *Semantic Web*, **7**(5), 519–528.

Rinaldi, F., Ellendorff, T. R., Madan, S., Clematide, S., van der Lek, A., Mevissen, T., and Fluck, J. (2016). BioCreative V track 4: a shared task for the extraction of causal network information using the Biological Expression Language. *Database*, **2016**.

Rindflesch, T. C. and Fiszman, M. (2003). The interaction of domain knowledge and linguistic structure in natural language processing: interpreting hypernymic propositions in biomedical text. *Journal of Biomedical Informatics*, **36**(6), 462–477.

Sándor, A. and de Waard, A. (2012). Identifying claimed knowledge updates in biomedical research articles. In *Proceedings of the Workshop on Detecting Structure in Scholarly Discourse*, pages 10–17.

Schulz, K. F., Altman, D. G., and Moher, D. (2010). CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMJ*, **340**, c332.

Segura-Bedmar, I., Martinez, P., and Sanchez-Cisneros, D. (2011). The 1st DDIExtraction-2011 Challenge Task: Extraction of Drug-Drug Interactions from Biomedical Texts. In *Proceedings of the 1st Challenge Task on Drug-Drug Interaction Extraction 2011*, pages 1–9.

Smith, L. H., Rindflesch, T. C., and Wilbur, W. J. (2004). MedPost: a part-of-speech tagger for biomedical text. *Bioinformatics*, **20**(14), 2320–2321.

Tikk, D., Thomas, P., Palaga, P., Hakenberg, J., and Leser, U. (2010). A comprehensive benchmark of kernel methods to extract protein-protein interactions from literature. *PLoS Computational Ciology*, **6**(7), e1000837.

Van Landeghem, S., Björne, J., Wei, C.-H., Hakala, K., Pyysalo, S., Ananiadou, S., Kao, H.-Y., Lu, Z., Salakoski, T., Van de Peer, Y., and Ginter, F. (2013). Large-scale event extraction from literature with multi-level gene normalization. *PLOS ONE*, **8**(4), e55814.

Wei, C.-H., Kao, H.-Y., and Lu, Z. (2012). Sr4gn: A species recognition software tool for gene normalization. *PLOS ONE*, **7**(6), 1–5.

Wei, C.-H., Kao, H.-Y., and Lu, Z. (2013a). Pubtator: a web-based text mining tool for assisting biocuration. *Nucleic Acids Research*, **41**(Web Server Issue), W518–22.

Wei, C.-H., Harris, B. R., Kao, H.-Y., and Lu, Z. (2013b). tmVar: a text mining approach for extracting sequence variants in biomedical literature. *Bioinformatics*, **29**(11), 1433–1439.

Wei, C.-H., Kao, H.-Y., and Lu, Z. (2015). GNormPlus: An Integrative Approach for Tagging Genes, Gene Families, and Protein Domains. *BioMed Research International*, **2015**.

Wei, C.-H., Peng, Y., Leaman, R., Davis, A. P., Mattingly, C. J., Li, J., Wiegers, T. C., and Lu, Z. (2016). Assessing the state of the art in biomedical relation extraction: overview of the BioCreative V chemical-disease relation (CDR) task. *Database*, **2016**, baw032.

Xu, J., Wu, Y., Zhang, Y., Wang, J., Lee, H.-J., and Xu, H. (2016). CD-REST: a system for extracting chemical-induced disease relation in literature. *Database*, **2016**, baw036.

Yoshikawa, K., Riedel, S., Hirao, T., Asahara, M., and Matsumoto, Y. (2011). Coreference Based Event-Argument Relation Extraction on Biomedical Text. *Journal of Biomedical Semantics*, **2 (Suppl 5)**, S6.