Better Living through Coordination Chemistry: A descriptive study of a prolific papermill that combines crystallography and medicine

David Bimler (smut.clyde@gmail.com)

Research Article

Keywords: Papermills, citation abuse, citation networks, crystallography, coordination polymers, metal-organic frameworks

Posted Date: April 15th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1537438/v1

Abstract

The organised, production-line creation of fake papers by “papermills”, injecting fabricated results into the scientific literature, has become a problem in many domains of biomedical research. Here we propose that at least 800 publications in crystallography and exotic-chemistry journals, from the period 2015-2022, are also the work of a prolific papermill specialising in imaginary Metal-Organic Frameworks and their wholly invented therapeutic applications. The mill is characterised by recycled images, and by oddities of wording in Methods sections, but its most obvious hallmarks appear in Reference sections, with citations to irrelevant research from remote fields of science. One purpose of these irregularities is presumably to manipulate the researchers’ performance indicators. We argue that a paper’s References deserve closer scrutiny from journal editors and publishers than they routinely receive.

1. Background

It is a truism that the rungs of the career ladder for anyone in a research or higher-education profession are publications in academic literature. The expectations of a paper-spangled CV are applied especially broadly in China, where government policy requires even medical practitioners to conduct research, or at least to publish the results of research (Hvistendahl, 2013).

A predictable outcome of the pressure to publish was the emergence of ‘papermills’: production-line academic ghost-writers who assemble manuscripts for customers who are not in a position to conduct research, due to insufficient time, resources, training or ability (Else & Van Noorden, 2021). A mill might be a single individual producing manuscripts on an artisanal basis (Kalliokoski & Heathers, 2020), or a larger studio of collaborating specialists working on an industrial scale, comparable to a medieval scriptorium, or to the “gold-farming” phenomenon.

In practical terms, “assemble” means “fabricate”, and the livelihood of papermills rests on a steady flow of falsehoods into journals, diluting the results of genuine research and wasting the time of anyone who tries to replicate or build on them. The corrosive impact of papermills on research integrity makes them worthy of attention from a practical perspective. Even bogus papers are cited in review papers, literature reviews and grant proposals when authors read only the titles or abstracts (or references in earlier literature reviews), and find support for their own projects. There is a race between retraction and laundering into general acceptance. Of course papermills are also of interest as an academic topic within the sociology of science, as a phenomenon generated by the interaction between funding systems, market forces, and regrettable central-government policy initiatives.

It is the nature of papermills to be agile, synthesising papers in whichever fields are most active and most highly-prioritised by funding bodies at a given time. Like genuine researchers, they follow the funds. If, for instance, a government prioritises herbal traditions as a medical modality, papermills will oblige with reports promising *in-vitro* reports for herbal phytochemicals. If laboratories around the world discover the intra-cellular feedback networks of non-coding RNAs, then journals should brace themselves for an onslaught of ncRNA falsification. But at the same time, the consumer base for papermills is dominated by medical clinicians in China who need a publication in an international journal for each promotional milestone, so the hot topic of the day must receive a therapeutic spin.

Here we set out to delineate a single papermill that piggybacks on the popularity of “metal-organic frameworks” (MOFs) or “coordination polymers” (CPs) (Wikipedia, 2022). The mill’s outputs are a curious hybrid of crystallography and medicine, where these fascinating meta-crystalline structures acquire therapeutic applications, so that they meet the requirements of the mill’s clinician customers. CPs gain anaesthetic properties, or kill cancer cells or bacteria, or stop inflammation.

The present paper is descriptive, in the manner of a case study or a report on a new species. It does not offer specific, selective criteria that distinguish *all* papermill products from authentic science. But despite this narrow scope and modest ambitions, the case study may illustrate some general principles of wider application.

Within the pragmatic field of ‘papermill studies’, one mill was discovered twice. After its repetitive, implausible imitations of Western Blots and flow-cytometric apoptosis assays caught the attention of an *ad-hoc* collaboration of observers, largely anonymous apart from Dr Elisabeth Bik, it was described on the blogs ‘Science Integrity Digest’ (Bik, 2020), and ‘For Better Science’ (Clyde, 2020). Meanwhile a pursuit of malfeasance in gene-function research led Byrne and Christopher (2020) to the mill’s products independently; they memorably described it as “Digital magic, or the dark arts of the 21st century”. This “tadpole papermill” was initially credited with 400 papers (Bik, 2020), but the tally now exceeds 600 after additional journals came to light that the millers had colonised.

Other papermills have been identified. In Editorials to accompany extensive tranches of retractions, journal editors shared their new awareness of papermill activities, and encouraged others in the industry to take similar defensive measures (Behl, 2021; Seifert, 2021).

A broader perspective may be useful. Despite the scale of papermill activity, the niche they occupy is only one part of a larger ecosystem. The industry of ‘publication facilitation’ also contains predatory / parasitical journals, hijacked journals, software-disguised plagiarism (Cabanac, Labbé & Magazinov, 2021), co-authorship markets, citation trading and extortion, data fabrication, etc. These phenomena are all interconnected, blurring the boundaries between them and hindering any rigorous taxonomy, but their common features are also helpful to us. Specifically, Abalkina’s analysis of hijacked journals is also applicable here.

Abalkina (2021) noted that when a website / journal is set up to mimic a long-established journal (which may be still operating, or no longer extant), offering publication opportunities in exchange for a fee – in other words, when a journal has been hijacked or its identity was stolen – its long-term existence is not assured. Legal repercussions might force the website to disappear. This motivates the proprietors of the journal to maximise their *short-term* profiles, and to take all possible shortcuts in constructing and operating the imitation. This is not an industry for long-term planning or high standards of workmanship.

The imitation will have archives of previous publications, to bolster its pretence of a long illustrious history, but these will be copy-pasted from the archives of *other* imitation journals from the same perpetrators (see also Siler et al. 2021).

Papermills operate with the same business model of immediate profit extraction, recycling text, tables and other material to save time and effort. Not to forget the favourable reviews that are sometimes passed to journal editors by faked ‘reviewers’, to suborn the peer-review process and facilitate the acceptance of papermilled manuscripts: these can also be recycled (Day, 2022). Figures are yet another manifestation of this urgency.

Different scientific fields have their individual conventions on what kind of images are appropriate as illustrations. In oncology papers, for instance, one might encounter Western Blots and cell-cycle plots; microphotographs of assays of cell migration and invasion; etc. Reading about nanotechnology or material science or the specific application of drug delivery, one expects to see electron microscopy and X-ray diffractograms. All these display examples of an early stage of data collection, still to be quantified and averaged, whereas the summary statistics and the comparisons among these are the substantive contents of the paper. That is, many of the illustrations in biomedical journals provide the reader with no additional information. As early-stage images, they are primarily a token of good faith, to assure the audience that experiments did occur as described, and that (pictorial) data were collected.

Conversely, if experiments were not conducted, the figures must have come from other sources, and often they will be used repeatedly. They might be fake, with no more realism or attention to detail than is necessary, or acquired from elsewhere. Returning to the ‘tadpole’ papermill mentioned above, it received this label due to the stylised, sinuous Western Blots that dominate its figures, resembling tadpoles or sportive sardines. This mill also favoured a hand-stippled form of illustration that were implausible travesties of flow-cytometry scatterplots. We should note the possibility that papermillers may gain access to the image archives of an active laboratory, and can use genuine though re-purposed “out-takes” from that lab’s research to illustrate their confections.

Another convention of academic writing is the References section of a paper, listing sources to support claims made in the text. These can be laborious to prepare, especially if the authors feel obliged to read each reference to check that their vague recollection or second-hand notion of its contents is correct. So just as with the Figures, these References offer opportunities for papermillers to save time and effort by recycling a pre-existing list, unconcerned about the relevance of specific citations in the context of the text. Ideally reviewers would glance at References to check that they are appropriate to the corresponding citations, but the evidence suggests that does not always occur.

The additional possibility that citations are sometimes *sponsored* should also be considered. We will return to this below.

2. Forms Of Verbal And Visual Coincidence

A single paper can be enough to establish the *existence* of a papermill. “Cervical cancer treatment of Co(II) coordination polymer through miR-9-5p-regulated BRCA1-OCT1-GADD45 pathways” (Zhao et al, 2022) is a chimeric amalgam of laser eye surgery and cervical cancer: readers are assured that *“The cervical cancer rat model was constructed in accordance with the statement of the Association for Visual and Ophthalmic Research on the utilization of animals in ophthalmic and visual research”*. The simplest explanation for its incoherence is that a manuscript was originally written around a promise of faster healing from eye surgery, and then was incompletely refocussed around cervical cancer because the purchasers belong to a Department of Gynaecology and Obstetrics. The millers did not get as far as *describing* the supposed “cervical cancer rat model.”

The cervical cancer cells mice utilized in the study were randomly separated into four diverse groups and injected with miR-9-5p mimic, miR-9-5p inhibitor, mimic control, and inhibitor control. All injections were made directly after laser treatment, and intravitreal injections were manipulated following established protocols...

However, the proposition that a hypothetical papermill generated all the papers considered here (see Section 3) requires more evidence. We draw upon several forms of verbal and visual coincidence. Lengthy text duplications also exist, but this can happen without papermilling.

2.1. Duplicated Crystal-determination Figures

Each of the papers described in Section 3 begins by characterising one or more crystal structures (organic or metal-organic), using X-ray diffraction. Predicted and empirical PXRD plots (powder X-ray diffraction) are the most common form of illustration, even when normal XRD would be possible (i.e. the authors claim to have obtained their material as large high-quality crystals). Frequently these are complemented with thermogravimetric analysis (TGA), infrared spectra (FTIR), and scanning / transmission electron microscopy (SEMs, TEMs). Images repeat, as if the millers’ output exceeded their stockpile. In a tribute to their persuasive power of SEMs, they are doubled, tripled and quadrupled. A doublet of TEMs has illustrated at least five different compounds. A pair of FTIR spectra are equally versatile. TGAs appear repeatedly (Fig. 1).

2.2. Crystallographic datasets

Researchers in this field can make the parameters and results of their crystal-structure determination available, by depositing them in the database at Cambridge Crystallographic Data Center (CCDC) and providing a link as Supporting Data. Many papers in the present corpus follow this trend. 118 structures in the database are credited to a Meng Qinghua, of Hong Zhou University, with many of them providing the foundation for papers, although there is no obvious connection between Meng Qinghua and the authorship of the corresponding papers. 134 structures are credited to a Haitao Yuan of Bohai University, with many providing the foundation for papers from otherwise-unrelated authors. In other cases, the ‘Crystallographer’ and ‘Affiliation’ fields in the CCDC entries have been filled with random keyboardpounding letter-strings like ‘wferf’ or ‘ewfer’, which does not suggest that much pride was taken in the accomplishment. When papers report on more than one compound, they might credit one crystal structure to Meng Qinghua and the other to a random string. Bernès (2022) noted two structures in the CCDC database that had been probed with identical parameters.

The total amount of false information recorded in crystallographic repositories is unknown.

2.3. Duplicated Cell-function Figures

Claims of clinical efficacy can call upon support from *in vitro* and *in vivo* experiments, performed on cells in culture medium and on laboratory animals respectively. With both approaches, raw data can include flow-cytometry assays and Western Blot semi-quantifications of protein expression. The Western Blots in this literature see some recycling. The flow-cytometry scatterplots are frequent (one can speculate that the millers have acquired image archives from a laboratory working in this area). They display a level of similarity that one expects from experimental replicates of research on a single cell-line, rather than independent experiments in multiple laboratories on multiple cell-lines, each with its own metabolic idiosyncrasy. Re-use of a panel has been documented (i.e. identical images), and further duplications may exist, waiting for someone to search for them (Fig. 2).

2.4. Duplicated Cell-viability Figures

In vitro results in this corpus are often accompanied by line graphs that plot cell viability as a function of increasing concentration of the compound in the culture medium: cytotoxicity is sought-after if the cells exemplify a given cancer type or pathogen, and undesirable if they represent normal tissue. In almost all examples, the chart’s horizontal axis (concentration) has a non-intuitive broken-logarithmic scale, in which a small jump in concentration from 8 to 10 units receives the same horizontal increment as factor-of-two increases (2 to 4 units, 4 to 8, 10 to 20, etc) (Fig. 3).

Colony Formation assays are an alternative measurement of the impact of a cytotoxin, by counting the proportion of cells that retain the capacity to proliferate in culture and form colonies, after treatment. Images are repeated across papers (Fig. 4).

2.5. Linguistic oddities

In vivo method descriptions reveal recurring linguistic oddities, similar to the “tortured phrases” hallmarks of thesaurus- or software-assisted plagiarism, although the motivation differs. Notably,

Instead of ‘logarithmic growth phase’, many papers refer to ‘logical’ or ‘logistical growth phase / phage’ – or sometimes a variant, ‘phage of logical growth’.

‘Ultimate / final destiny’ frequently takes the place of ‘final density’.

‘Serious dilution’ substitutes for ‘series dilution’.

Despite the ubiquity of these malapropisms in later papers, they were absent from the papermill's early work. One possible explanation is that increased production required new millers to be recruited, and provided with templates to follow: templates containing errors, which they faithfully copied into the manuscripts they assembled.

In another diagnostic malapropism, *in vivo* as well as *in vitro* Methods sections developed a proclivity for 'preformation' as a substitute for 'procedure'.

The corpus also displays an inordinate fondness for the phrases 'in brief' or 'in short' or 'briefly' or 'in sum', to introduce a Methods paragraph, pre-emptively justifying the absence of crucial details. 'Triumphantly synthesised' is yet another phrase. We note that these are not very selective. In the advanced-chemistry literature, 'triumphantly' has lost its connotations of exultation, becoming accepted as a fancy synonym for 'successfully' – much as the solecism of writing 'facile' to mean 'easy' or 'simple' has entered common usage as a default term, to the despair of prescriptive grammarians.

2.6. Ethics-approval anomalies

Specific to *in vivo* studies, publishing guidelines demand an Ethics disclosure of some kind to reassure readers that any mutilation, vivisection and ultimate dissection of laboratory animals was considered by an independent Ethics Committee and approved (as outweighed by the expected benefits). In the present corpus, the Ethics Approvals never come from a committee at the nominal authors' own institutions (as the millers do not know at the time of writing who will be the customer). They are missing entirely, or meaningless: ascribed to the company that purportedly supplied the supposed laboratory animals, or to committees that do not seem to exist. The journals' editors and reviewers were unaware of this expectation that vivisection requires meaningful ethics oversight, or they suspected that the reports of experiments were false but accepted the manuscripts anyway.

2.7. Corresponding Email anomalies

Papermilled manuscripts in general are often submitted to the journals by the millers directly, on behalf of the signatories, using single-purpose 'burner' email accounts created through social-media companies (e.g. 126.com, 163.com, yeah.com). In the present corpus, a handful of corresponding authors used institutional email accounts, or commercial accounts attested from other contexts. But crucially, 391 of 806 corresponding email identities (48.5%) were generated by simple rules (Table 1), consistent with the working theory that a papermill handled negotiations with journals to remove the signatories from the loop. More recent papers in the corpus (2021–2022) are more likely to carry an e-address that doesn't conform to these simple patterns. Even so, when papers provide e-addresses then the Pubpeer website is designed to message the corresponding author automatically with an invitation to join the discussion, but very few have responded.

Table 1
Most common patterns of corresponding-author email addresses.

Pattern	Number	Percentage
<name>_<name > 66@.<provider>	68	8.4%
<name>_<name > 666@<provider>	178	22.1%
<name>_<name > 11@<provider>	97	12.0%
<name>_<name > 12@<provider>	32	4.0%
<name>_<name > 22@<provider>	16	2.0%

2.8 Duplicated meaningless References

Reference sections provide the final line of evidence. Many are conspicuous by unexpected overlaps: lists of references reappearing in paper after paper like strips of wallpaper that have been cut to the desired length. Moreover, the duplicated references are often discordant with the corresponding citations in the text, ranging from merely incorrect (papers on inorganic ceramics cited as examples of the diversity of CPs and applications), to glaringly incongruous in their irrelevance (bovine dyspepsia; anti-corrosion coatings; remediation of contaminated soil and wastewater; pollutant degradation; composting corn-husks with chicken manure to maximise their value as a source of biofuel). The absurdity extends to RFID tags, software bug signalling and IoT data structure abstractions. Table 2 lists some of these clusters of co-occurring References.

Table 2
Five clusters of co-occurring irrelevant references.

Cluster 1	**Comparison of hydrogen peroxide and ammonia pretreatment of corn stover: solid recovery, composition changes, and enzymatic hydrolysis**	**Zhao et al (2014)**
	Physical and chemical characterizations of corn stalk resulting from hydrogen peroxide presoaking prior to ammonia fiber expansion pretreatment	Zhao et al (2016)
	Enzymatic hydrolysis and physiochemical characterization of corn leaf after H-AFEX pretreatment	Zhao et al (2016)
	Optimization of liquid ammonia pretreatment conditions for maximizing sugar release from giant reed (*Arundo donax* L.)	Zhao et al (2017)
	Application of hydrogen peroxide presoaking prior to ammonia fiber expansion pretreatment of energy crops	Zhao et al (2017)
	Methane enhancement through co-digestion of chicken manure and oxidative cleaved wheat straw: stability performance and kinetic modeling perspectives	Hassan et al (2017)
	Structural characterization of corn stover lignin after hydrogen peroxide presoaking prior to ammonia fiber expansion pretreatment	Qiao et al (2018)
Cluster 2	β-Hydroxybutyrate induces bovine hepatocyte apoptosis via an ROS-p38 signaling pathway	Song et al (2016)
	Histamine induces bovine rumen epithelial cell inflammatory response via NF-κB pathway	Sun et al (2017)
	Acetoacetate induces hepatocytes apoptosis by the ROS-mediated MAPKs pathway in ketotic cows	Du et al (2017)
	High concentrations of fatty acids and β-hydroxybutyrate impair the growth hormone-mediated hepatic JAK2-STAT5 pathway in clinically ketotic cows	Du et al (2018)
	Inflammatory mechanism of Rumenitis in dairy cows with subacute ruminal acidosis	Zhao et al (2018)
	The ceramide pathway is involved in the survival, apoptosis and exosome functions of human multiple myeloma cells in vitro	Cheng et al (2018)
Cluster 3	Electrochemical Impedance Spectroscopy Evaluation of Corrosion Protection of X65 Carbon Steel by Halloysite Nanotube-Filled Epoxy Composite Coatings in 3.5% NaCl Solution	Zhang et al (2019)
	Synergistic Effect Between Nano-Sb2O3 and Brominated Epoxy Resin on the Flame Retardancy of Poly(butylene terephthalate)	Niu et al (2019)
	A new architecture of super-hydrophilic β-SiAlON/graphene oxide ceramic membrane for enhanced anti-fouling and separation of water/oil emulsion	Kang et al (2019)
	Ensemble Data Reduction Techniques and Multi-RSMOTE via Fuzzy Integral for Bug Report Classification	Guo et al (2018)

Cluster 1	**Comparison of hydrogen peroxide and ammonia pretreatment of corn stover: solid recovery, composition changes, and enzymatic hydrolysis**	**Zhao et al (2014)**
	Identify Severity Bug Report with Distribution Imbalance by CRSMOTE and ELM	Guo et al (2019)
	The Influence Ranking for Testers in Bug Tracking Systems	Li et al (2019)
	Fast splitting based tag identification algorithm for anti-collision in UHF RFID system	Su et al (2019)
	Energy Efficient Tag Identification Algorithms For RFID: Survey, Motivation And New Design	Su et at (2019)
Cluster 4	Mechanisms of emerging pollutant Dechlorane Plus on the production of short-chain fatty acids from sludge anaerobic fermentation	Zhang et al (2021)
	Constructing Straight Pores and Improving Mechanical Properties of Gangue-Based Porous Ceramics	Xu et al (2021)
	Synthesis and catalytic performance of a new V-doped CeO2-supported Alkali-activated-steel-slag-based photocatalyst	Kang et al (2021)
	Effects of A/B-site co-doping on microstructure and dielectric thermal stability of AgNbO3 ceramics	Ran et al (2021)
Cluster 5	Effect of fluoxetine on enhanced biological phosphorus removal using a sequencing batch reactor	Zhao et al (2021)
	Fabrication of direct Z-scheme $FeIn_2S_4/Bi_2WO_6$ hierarchical heterostructures with enhanced photocatalytic activity for tetracycline hydrochloride photodagradation	Shangguan et al (2021)
	Fe3O4 nanoparticles three-dimensional electro-peroxydisulfate for improving tetracycline degradation	Tang et al (2021)
	Synthesis of core-shell α-AlH3@Al(OH)3 nanocomposite with improved low-temperature dehydriding properties by mechanochemical mixing and ionic liquid treatment	Duan et al (2021)

———— Table 2 at end of file ————

These citational incongruities were central in collating the corpus under discussion. When, for instance, papers already assigned to the corpus were more eager to cite “Research and Application of Acoustic Emission Signal Processing Technology” than one might expect, examination of other papers also citing “Research and Application of Acoustic Emission Signal Processing Technology” led to further candidates for inclusion, in successive waves of a progressively broader search. This alternated with discovery of the verbal oddities described in Section 2.5, and their use in another wave of search targets. Systematic searches of journal Tables of Contents also occurred.

3. The Corpus

This paper rests on discussion threads at PubPeer, a platform allowing post-publication reviews and discussions of any paper that has been assigned a DOI or a Pub-Med ID number (Barbour and Stell, 2020). 648 papers were relevant, and are listed in a spreadsheet, archived as Supplementary File 1. They were flagged at PubPeer for possible flaws in the course of December 2021 / January 2022, each receiving its own discussion thread. Critics included Guillaume Cabanac, Alexander Magazinov, N. H. Wise, Sylvain Bernès, "*Hoya camphorifolia*" and "*Parashorea tomentella*", where the last two contributors are biological-species names assigned randomly by the platform's pseudonym generator.

All 648 papers have crystallography as the central discipline (presumably the millers' original speciality), and have medical or pharmacology clinicians as the nominal authors. MOFs or CPs are the targets of crystallographic characterisation in the great majority of cases. Fifty-eight involved crystals of less exotic organic molecules, unlinked and uncoordinated by metal ions. Most of this minority appeared in *Zeitschrift für Kristallographie - New Crystal Structures* or *Main Group Chemistry*, and the therapeutic content (the reason for the nominal authors' interest) was confined to a few passing citations.

A later development saw CPs associated with two kinds of application: industrial (e.g. catalytic potential, photocatalytic degradation of pollutants, gas separation, fluorescent detection of toxins or the volatile molecular fingerprints of explosives, etc.), but also therapeutic use, to fit the purchasers' specialities. 'Shimmer' comes to mind: "It's a floor-wax and a dessert topping!" (Gould & Murray, 1976).

It is quite possible that the papermill provides manuscripts to other (smaller) customer bases, merely characterising imaginary CPs or claiming applications of this industrial nature. Apart from spurious entries in the CCDC database, it is not clear how these could be identified (though see Xu et al, 2019).

Publication dates were distributed as in Table 3:

Table 3. Number of observed papermill products per year.

Year	2015	2016	2017	2018	2019	2020	2021	2022
Papers	2	9	13	55	73	226	245	25

The dates paint a picture of accelerating publication, as if a small-scale cottage industry had been scaled up to a production line with a larger staff. One can imagine crystallographers initially ghostwriting manuscripts as a favour for friends, moonlighting from their day job, and becoming progressively more professional, though this must remain speculation.

The most well-represented publishers in the corpus are Taylor & Francis (204 papers), Springer (154, plus 54 from the 'Pleiades' imprint), Elsevier (88), De Gruyter (48), IOS (41) and Wiley (25). Table 1 singles out some noteworthy examples. The Table includes some scholarly-society journals (masked with asterisks) even if they are not well-represented.

12	**Wiley**	_**Journal of the Chinese Chemical Society ***_
4		_Chemistry Select *_
40	Springer	_Journal of Cluster Science_
27		_Journal of Inorganic and Organometallic Polymers and Materials_
26		_Journal of the Iranian Chemical Society *_
18		_Journal of Fluorescence_
18		_Journal of Polymer Research_
34	Elsevier	_Journal of Molecular Structure_
16		_Journal of Solid State Chemistry_
15		_Arabian Journal of Chemistry *_
8		_Journal of the Indian Chemical Society *_
148	Taylor & Francis	_Inorganic and Nano-Metal Chemistry_
37		_Journal of Coordination Chemistry_
46	Pleiades (Springer)	_Journal of Structural Chemistry_
39	De Gruyter	_Zeitschrift für Kristallographie - New Crystal Structures_
41	IOS	_Main Group Chemistry_
7	Japan Oil Chemists' Society	_Journal of Oleo Science *_
5	CSIRO	_Australian Journal of Chemistry *_
2	Chemical Society of Japan	_Bulletin of the Chemical Society of Japan *_
1	Associação Brasileira de Divulgação Científica	_Brazilian Journal of Medical and Biological Research *_

Individually these were not high-profile papers, but their impacts add up. Between them, the 648 papers were cited 518 times, i.e. an average of 0.82 times each. The distribution of citations was skewed, with most (410 papers) receiving zero citations and 125 being cited once, ranging up to three papers being cited 10 times and one paper 11 times (from Dimension.ai, as of 7 April 2022).

For the sake of completeness, three journals that do not assign DOIs are included as worksheets in the spreadsheet. For the _Indian Journal of Chemistry A_, the three problematic papers are recent (all December 2021) and it may be that the hypothesised papermill only recently began submitting papers there. Conversely, four papers in _Biomedical Research_ are from 2016–2017, with nothing more recent coming to light. That journal's publication fee may have become a deterrent. *Ind.J.Chem.A* and *Biomed Res.* are published by, respectively, Council for Scientific and Industrial Research (an Indian government body), and a branch of the predatory publisher OMICS ('Allied Academies'). One can safely describe the former as more prestigious.

The third DOI-free journal is *Latin American Journal of Pharmacology*, from Buenos Aires. Its archives have yielded 155 entries, and that is without looking at the years 2020, 2019, 2017 or half of 2016 – a thorough survey of the archives might double the number. Library subscriptions do not seem to be available for *Lat.Am.J.Pharmacol.*; instead, papers must be bought individually (US$50 per paper for access).

All papers came from Chinese medical institutions. Most authors in the table appear only once. Most of the exceptions come from a group at Tongliao City Hospital, and spread over the medical schools and hospitals attached to Tongliao's Inner Mongolia University for Nationalities. Across that group's 27 papers, Bao-Hua Song, Fu-Long Bi and Wen-Li Tang each signed five; Chen Li, Dong-Song Bai and Yang Liu each signed six; while Lei Pan, Gui-Feng An and Zhi-Fang Zhang signed seven, eight and 16.

4. Discussion

The postulated papermill has convinced journals to accept nearly 800 manuscripts with medical clinicians as the nominal authors, in which crystallographic inventions are adorned with fabricated claims of therapeutic efficacy. This is a tribute to the broadness of the curricula in Chinese medical schools, and to the ubiquity of crystallographic apparatus.

The millers' accomplishment is all the greater when one considers that despite the allure of their exotic chemistry, MOFs and CPs do not lend themselves to therapeutic goals (except from their potential as drug-delivery vehicles, sequestering a cytotoxin and releasing it in a targeted environment). Administration is the paramount problem. Many MOFs do not dissolve in solvents fit for internal use, while if an MOF can be dissolved, it is no longer a MOF: metal ions and ligands are no longer locked in the specific alignments that they took while crystallising under tightly-controlled conditions of heat, pressure, timing, pH, etc. Papers typically finessed this issue by referring vaguely to "administration" without specifying details, leaving readers to guess whether the MOF was injected as a suspension of finely-pulverised particles, or swallowed, or inhaled, or held against laboratory animals' skin by strips of adhesive tape. "In brief" does a lot of work in these Methods paragraphs.

The millers' success was possible because journal editors and reviewers were blithely unaware of biological blunders and absurdities in the texts. In defence of the editors and reviewers, they are chemists and physicists by training, and no wiser in the ways of biologists than the millers are. Some absurdities have been mentioned, but many others could be adduced. On the *in vitro* side, neurons (Huang et al, 2020), cochlear hair cells (Sun et al, 2021) and corneal endothelial cells (Bian et al, 2021, 2022) are described as multiplying in culture media, which does not happen in reality. Tuberculosis bacteria multiply in 'Rain Heart Infusion' rather than requiring a special medium (Yan et al. 2020). Non-malignant cell-lines are rebranded as “cancer” (Ni et al, 2021). *Candida albicans* is demoted from fungus to bacterium (Wang, Song & Jiang, 2021).

In vivo errors are equally intriguing. Rodents switch between ‘rats’ and ‘mice’ in successive sentences (Yang et al, 2021), or even from ‘mice’ to ‘rabbits’ (Bao et al, 2021). We learn of 220-g BALB/c mice, and 220-g “Sprague-Dawley mice” (Wang et al, 2021), and of rats that were presumably female and pregnant (for models of pre-eclampsia or puerperal infection: Sun et al, 2022; Yang et al), although too young to be sexually mature. Mice “were seeded in 5% CO_2 and 37°C standard condition”, or “placed in the 5% CO_2 and 37 C standard conditions” (Liu et al, 2021). Rodent strains were specified that do not react to toxic diets in the way required from the disease model. Sometimes an animal model of disease is named but not accompanied by any details of what was done to induce disease (e.g. Zhang et al, 2020; Zhao et al, 2022).

In silico docking studies feature in a number of papers. Here, the strength of the bond is computed between the ion/ligand spatial cell of the MOF lattice on one hand, and an arbitrarily-chosen protein on the other hand, for a range of possible alignments, leading up to a choice of one alignment as lower-energy than the others. These calculations establish only that the millers have access to the necessary software. The objective of all this is not obvious, as the unit cell never exists in isolation: in a solution, it is non-existent: in a crystal, it subsumed within a much larger structure.

Individual papers had little direct impact, with most remaining uncited and ignored in altmetric indicators. Collectively and indirectly, though, they normalise implausible claims. Rresearchers with misguided CP / MOS projects can cited them to present their own results as part of the research mainstream.

This essay grew out of an interest in Reference Sections as a source of absurdities and repetitions. Particular constellations of co-occurring references found their way into scores of papers, often gloriously mismatched to the corresponding citations in the text. These recycled, nonsensical constellations provided one method of searching for potentially papermilled papers. The process of finding further constellations might lend itself to scaling up and automation by publishers, much as they apply plagiarism screening, to identify other papermills that cut corners in the same way.

The situation is complicated by the importance of citations as a source of career benefits to the authors of the cited papers (Biagioli & Lippman, 2020). This importance needs no introduction; it sometimes motivates journal reviewers and the Guest Editors of Special Issues to abuse their gate-keeper roles by demanding that the contributors of prospective papers add citations to their own work, as a condition of acceptance (Van Noorden, 2020). One anonymous PubPeer contributor testified to the existence of ‘citation circles’ of otherwise-unconnected authors who cite one another’s papers, for mutual benefit, organised by brokers who establish these circles and invite authors to join them (‘*Oriensubulitermes inanis*’, 2022). These phenomena are further niches within the ecosystem of publication.

The evidence is consistent with some level of collaboration between papermills and citations manipulators. This remains speculative, but an alliance would also account for the rapid growth in the output of the present papermill.

On their own, recycled References in a paper do not prove that a papermill was involved, for authors might have joined a citation-manipulation circle independently. *Oriensubulitermes inanis*' comment came within the context of the perovskite-themed oeuvre of Jun Zhao and Zhiqin Zheng (Supplementary File 2). The constellations of irrelevant references found in that group's 14 papers overlap with the constellations encountered in the crystallography corpus. The citations generally appear in the last few sentences of each paper, and in at least one case they were added as corrections at the final stage of proof-reading, and not present in the version accepted by reviewers.

In another small corpus of (so far) 40 potentially papermilled products, authors reported the synthesis of various ceramic nanomaterials with photocatalytic or drug-delivery applications. These papers are dated 2019 to 2021, and were largely published in *Ceramics International* from Elsevier (Supplementary File 3). This is not the place to describe the problematic electron microscopy within them, or the other shared features that are consistent with a common source. Those papers are relevant here because they also display lists of irrelevant references, which overlap substantially with the lists addressed in Section 2.8. Many cited a 'bovine dyspepsia' quartet of papers, and others that cluster around them (Cluster 2 from Table 2). Others were apparently inspired by the work of Zhao et al on corn-husk processing for biofuel production (Cluster 1), including the 'chicken manure' reference mentioned above.

It is puzzling that all the invocations of bovine ketosis, chicken manure, RFID tags, sound processing etc., as examples or applications of CPs, went unnoticed by so many reviewers. It is possible, though, that these were also added as final-stage corrections to the papers, and not even seen by the reviewers. Here the two versions of "Assembly of Two New Coordination Polymers: Luminescent Properties and Anti-inflammatory Activity on Postoperative Infectious Endophthalmitis" (Bian, Zhang & Qu 2022) are relevant. As well as the published version of record, a preprint is archived at Research Square (Bian, Zhang & Qu, 2021). A comparison shows that the block of meaningless but heavily-promoted References was a late addition to the published version, replacing a block of perfectly relevant citations in the preprint.

The point of this digression is that the reuse of a given reference in a papermill's output (whether or not it is relevant) can have at least two possible explanations. The organised manipulation of citation statistics may be a factor. However, we return to our starting point, that saving effort in the production line is a priority in any papermill's operation, and recycled strips of 'references wallpaper' may be prominent in the present corpus of papers for that reason alone.

In either case, References sections are a tempting target for both organised manipulation and corner-cutting by papermills. Though ignored by casual readers, they can be a source of promotion and income for researchers. They deserve more scrutiny from journals and reviewers than they seem to have received so far.

That is one lesson of this study. A second lesson is that if the scope of a journal had previously been limited to a specific domain of physics or chemistry, and manuscripts about medical applications of that speciality suddenly dominate the submissions, the editors need to temper their pleased surprise with caution. They need to find a pool of reviewers who are wise in the ways of biologists and can judge whether the therapeutic results of those submissions are plausible, or at least coherent. Reviewers suggested by the authors are not a good idea. The failure of so many journal to reject so many papers rife with biological absurdities speaks to some broader malaise.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets supporting the conclusions of this article are available in the Dryad repository, https://doi.org/10.5061/dryad.2280gb5v6.

Competing interests

The authors declare that they have no competing interests

Funding

None

Authors' contributions

The corresponding author collected the data and interpreted them as described.

Acknowledgements

The material analysed here was also the basis for a blogpost at “For Better Science” (Clyde, 2022). This manuscript could not have reached fruition without productive discussions with colleagues, including Guillaume Cabanac; PubPeer contributors; and others who prefer to remain anonymous. G. Cabanac compiled the scientometric values. We thank Digital Science for making Dimensions data (https://dimensions.ai) and Altmetric data (https://altmetric.com) available for research.

References

1. Anna Abalkina (2021). “Detecting a network of hijacked journals by its archive”. Scientometrics 126 7123–7148. doi: 10.1007/s11192-021-04056-0

2. Cai-Ying Bao, Yan-Ya Zheng, Huan Zhang, Qian Zhuo (2021). “Zn(II) and Co(II)-based coordination polymers: anesthesia effect by regulating animal blood pressure in hip replacement operation”. Inorganic and Nano-Metal Chemistry 41(4), doi: 10.1080/24701556.2020.1799398 Pubpeer thread https://pubpeer.com/publications/F6578BC62F38277FA06FBC993510CB
3. Christian Behl (2021). “Science integrity has been never more important: It's all about trust”. J Cellular Biochemistry 122:694–695. doi: 10.1002/jcb.29906
4. Mario Biagioli & Alexandra Lippman (2020). “Gaming the Metrics: Misconduct and Manipulation in Academic Research”. MIT Press, doi: 10.7551/mitpress/11087.001.0001
5. Guang-Qing Bian, Shi-Xi Zhang, Xiao-Li Qu (2021). “Assembly of Two New Coordination Polymers: Luminescent Properties and Anti-inflammatory Activity on Postoperative Infectious Endophthalmitis”, Research Square. doi: 10.21203/rs.3.rs-635113/v1
6. Guang-Qing Bian, Shi-Xi Zhang, Xiao-Li Qu (2022). “Assembly of Two New Coordination Polymers: Luminescent Properties and Application Values on Postoperative Infectious Endophthalmitis”. J. of Inorganic and Organometallic Polymers & Materials, 32 75–84. doi: 10.1007/s10904-021-02107-w
7. Elisabeth Bik (2020). “The Tadpole Paper Mill”. https://scienceintegritydigest.com/2020/02/21/the-tadpole-paper-mill/
8. Jennifer A. Byrne and Jana Christopher (2020), “Digital magic, or the dark arts of the 21st century—how can journals and peer reviewers detect manuscripts and publications from paper mills?” FEBS Letters 594 583–589 doi: 10.1002/1873-3468.13747
9. Guillaume Cabanac, Cyril Labbé & Alexander Magazinov (2021). “Tortured phrases: A dubious writing style emerging in science”. ArXiv https://arxiv.org/pdf/2107.06751.pdf
10. “Smut Clyde” (2022). “Polymer Papermillers Taking the Piss”. https://forbetterscience.com/2022/01/26/polymer-papermillers-taking-the-piss/
11. Adam Day (2022). “Exploratory analysis of text duplication in peer-review reveals peer-review fraud and paper mills”. ArXiv doi: 10.48550/arXiv.2202.03310
12. Holly Else & Richard Van Noorden (2021). “The fight against fake-paper factories that churn out sham science”. Nature 591, 516–519 1doi: 0.1038/d41586-021-00733-5
13. Elliot Gould & Anne Murray (1976). https://snltranscripts.jt.org/75/75ishimmer.phtml
14. Jia-Peng Huang, Ping Zhang, Jing-Gui Song, Jian Zhao (2020). “Two Cu(II)-based coordination polymers: Photocatalytic dye degradation and treatment activity combined with BDNF modified bone marrow mesenchymal stem cells on craniocerebral trauma via increasing complement C3 expression”. Arabian Journal of Chemistry 13(9), doi: 10.1016/j.arabjc.2020.07.010 PubPeer https://pubpeer.com/publications/3DE3DAC190EFDE6ED1EA6DB5EFAF3B
15. Mara Hvistendahl (2013). “China’s Publication Bazaar”, 342 (6162) 1035. doi: 10.1126/science.342.6162.1035
16. Otto Kalliokoski & James Heathers (2020). “The Mills have Ayes”. Archived at https://osf.io/ds6hk/

17. Chang-Liang Li, Pei-Zi Xia, Yu-Xiang He, Hua Zhou, Bo Zhao & Da-Wei Zhang (2021). "Two mixed-ligand coordination polymers: selective detection of acac and treatment activity on peripheral artery disease by reducing the peptidyl arginine deiminase activity". Journal of Molecular Structure 1229 129613. doi: 10.1016/j.molstruc.2020.129613 Pubpeer https://pubpeer.com/publications/EF4B05E9887E35A650AC5453BE2311
18. Yan Liu, Wei Zhang, Jiang-Bo Nan, Kai-Xiong Lei & Hui-Li Zou (2021). "Two Cu(II)-Based Coordination Polymers: Magnetic Properties and Treatment Activity on the Restless Leg Syndrome". Journal of Inorganic and Organometallic Polymers and Materials, 32, 134–142. doi: 10.1007/s10904-021-02092-0 Pubpeer thread https://pubpeer.com/publications/E34202CFDD277598F5F3FD454118F7
19. Li Ni, Shu-Li Wang, Jun-Qing Zhao, Gang Wu & Lu-Yan Lang (2021). "Two coordination polymers: selective detection of TNP and treatment activity on uterine fibroids combined with ultrasound treatment via inducing cancer cell apoptosis". Journal of Molecular Structure 1227 129364. doi.org/10.1016/j.molstruc.2020.129364 Pubpeer thread https://pubpeer.com/publications/B9F904876903F8E78F3CCC22920196
20. "*Oriensubulitermes inanis*" (2022). Comment https://pubpeer.com/publications/940C291607CF03969C6A936F8BA5B9#2
21. Roland Seifert (2021). "How *Naunyn-Schmiedeberg's Archives of Pharmacology* deals with fraudulent papers from paper mills". Naunyn-Schmiedeberg's Archives of Pharmacology volume 394, 431–436. doi: 10.1007/s00210-021-02056-8
22. *Kyle Siler*, *Philippe Vincent-Lamarre*, *Cassidy R. Sugimoto* & *Vincent Larivière (2021). "Predatory publishers' latest scam: bootlegged and rebranded papers". Nature 598 563–565*
23. *Xiao-Yan Sun*, *Jun-Nan Li*, *Gui-Mei Dong*, *Yan-Li Tao, Yao Chen & Shu-Qin Yang* (*2021). "A porous Er(III)–organic framework as a highly efficient catalyst for chemical fixation of CO*<biusub>2</biusub> *and treatment activity on senile deafness via inducing the cochlear hair cell apoptosis". Journal of the Iranian Chemical Society, 18, 1303–1309. doi*: 10.1007/s13738-020-02111-w PubPeer *thread* https://pubpeer.com/publications/775EECDE1DD88F2FC45279FA95D112
24. *Yan-Lan Sun*, *Fen-Mei Zhou*, *Zhen-Jing Chi*, *Yu-Lei Zhang, Rui-Rui Zhang (2022*). "*Treatment activity of a new luminescent thermostable Sm(III) complex on preeclampsia"*. *Inorganica Chimica Acta*, *530*, *120662. doi*:
10.1016
/j.ica.2021.120662 PubPeer thread
https://pubpeer.com/publications/13502822879ECF3A*8C6DB9E16EC2FC*
25. *Richard Van Noorden* (*2020*). "*Highly cited researcher banned from journal board for citation abuse"*. *Nature 578*, *200–201. doi*: 10.1038/d41586-020-00335-7
26. *Dan Wang*, *Yue Song* & *Di Jiang (2021). "Co(II) coordination polymer: application values on vulvovaginal candidiasis through reducing Candida albicans growth". Inorganic and Nano-Metal*

Chemistry doi: 10.1080/24701556.2021.1966446 PubPeer thread
https://pubpeer.com/publications/B3D5AE*8C8EF236167DA504A506E278*

27. *Zhi Wang*, *Ming Ye*, *Ya-Yi Huang*, *Fang Hu*, *Feng Qi*, *Dong Li* (*2021*). *"A new 3D pillar-layered Cd(II) compound: Luminescence sensing property for detection of Fe 3 + ions and treatment activity on focal cerebral ischemia/reperfusion"*. *Journal of the Chinese Chemical Society*, *68*(*10) pages 2013–2020. doi: 10.1002/jccs.202100108 PubPeer thread*
https://pubpeer.com/publications/B0298EB565E
5349E8A14E627EE57A4
28. *Wikipedia (2022). "Coordination Polymer".* https://en.wikipedia.org/w/index.php?title=Coordination_polymer&oldid=1081316335
29. Qi-Wei Xu, Qiu-Shuang Wang, Shan-Shan Li & Xia Li (2019). "Cu(ii)/Ni(ii)−organic frameworks constructed from the homometallic clusters by 5-(2-carboxyphenoxy)isophthalic acid and N-ligand: synthesis, structures and visible light-driven photocatalytic properties". RSC Advances 9, 16305. doi: 10.1039/c9ra01496a PubPeer thread
https://pubpeer.com/publications/DF7786414E46430E05AFFAFB87A74B
30. Li-Ping Yan, Qin Sun, Zhi-Bin Liu, Min Wu, Yan-Ping Ge & Qing Zhang (2020). "A new porous Co(II)-coordination polymer for the chemical fixation of CO_2 and treatment effect against tuberculosis by reducing the pckA expression in Mycobacterium tuberculosis". Applied Organometallic Chemistry, 34 (11). doi: 10.1002/aoc.5908 PubPeer thread
https://pubpeer.com/publications/42520B45C588B92FE2E5E0D3A70920
31. Juan Yang, Wei Song, Guoping Xiong & Ting Shen (2021). "Two new mixed ligand coordination polymers: Photo-catalytic properties, treatment and nursing application values on puerperal infection". Journal of Molecular Structure 144 130849. doi: 10.1016/j.molstruc.2021.130849 Pubpeer https://pubpeer.com/publications/E11529960A74EA472AB2BE544CBA70
32. Xia Zhao, Wei-Lei Dong, Gui-Fang Luo, Jing Xie, Jie Liu & Fu-Rong Yu (2022). "Cervical cancer treatment of Co(II) coordination polymer through miR-9-5p-regulated BRCA1-OCT1-GADD45 pathways". Arabian Journal of Chemistry 15(5) 103713. doi: 10.1016/j.arabjc.2022.103713 PubPeer thread https://pubpeer.com/publications/6F3E1D61D5FE9E8414CEEE2AEC4877
33. Jin-Mei Zhang, Mei-Ling Zeng, Xiao-Xin Liang, Lin Kong & Zhuo Zhang (2021). "Two transition metal coordination polymers: Selective gas sorption and treatment activity in combination with Latanoprost on primary open-angle glaucoma". Journal of Solid State Chemistry 301 122292. doi: 10.1016
/j.jssc.2021.122292 PubPeer thread
https://pubpeer.com/publications/936DC3E9F06FCBE0F2B62E24835CF2

Supplementary

Supplementary Files 1-3 are not available with this version

Figures

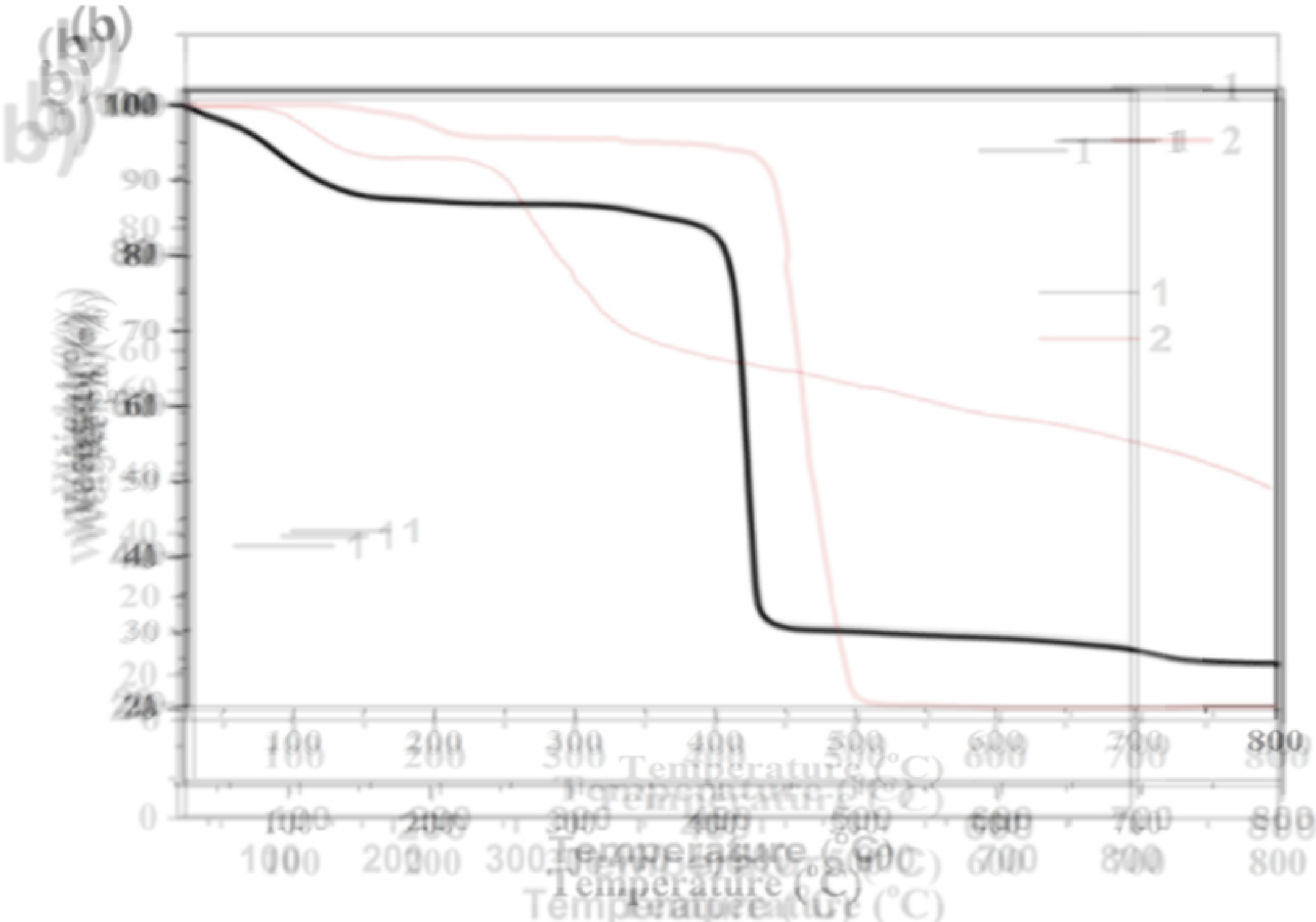

Figure 1

Thermogravimetric Analysis plots from eight papers, superimposed in LibreOffice Impress to dramatise the lack of differences apart from displacements along horizontal and vertical axes (temperature and mass).

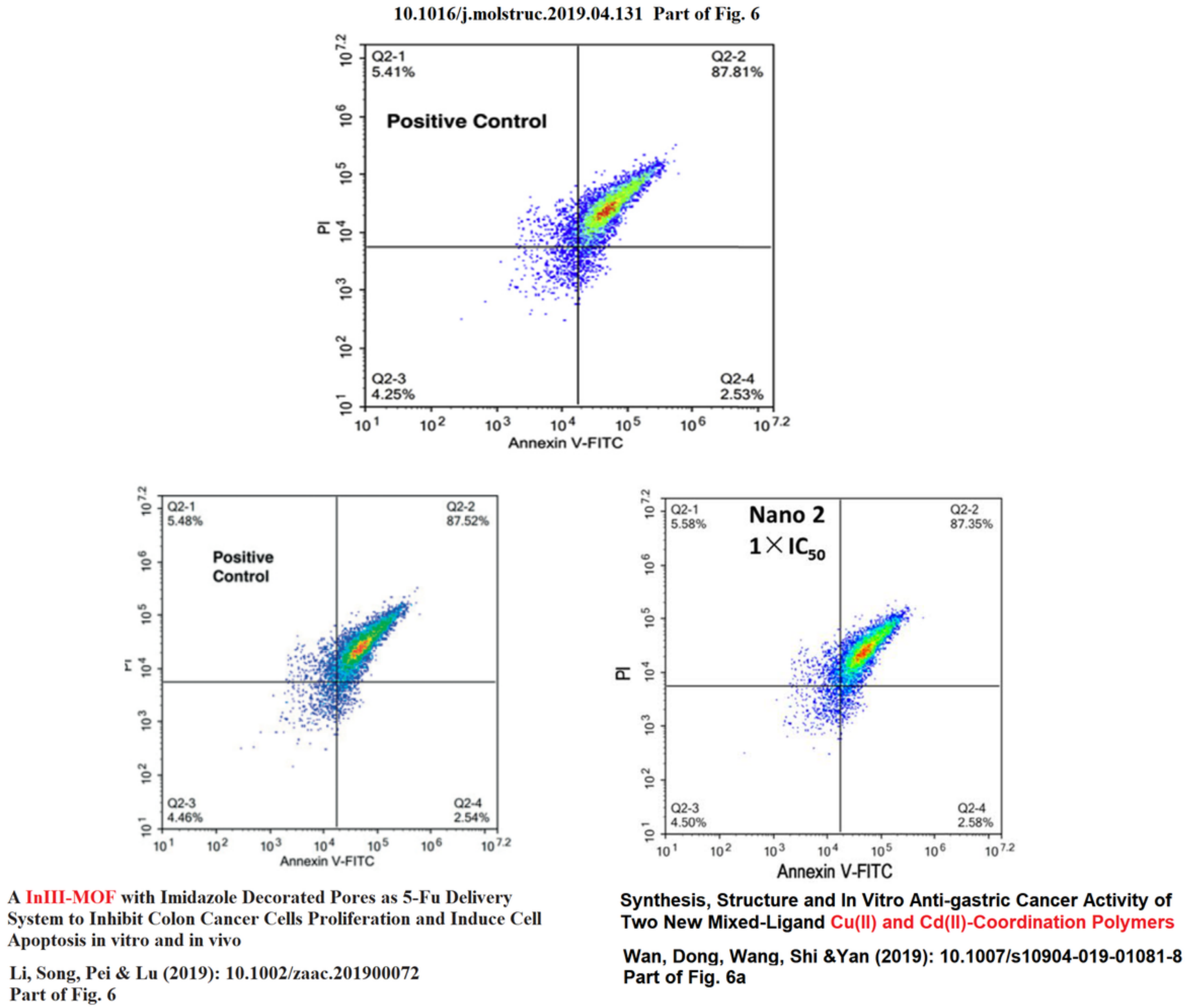

Figure 2

Three copies of a flow-cytometry apoptosis assay: montage created by S. Bernès

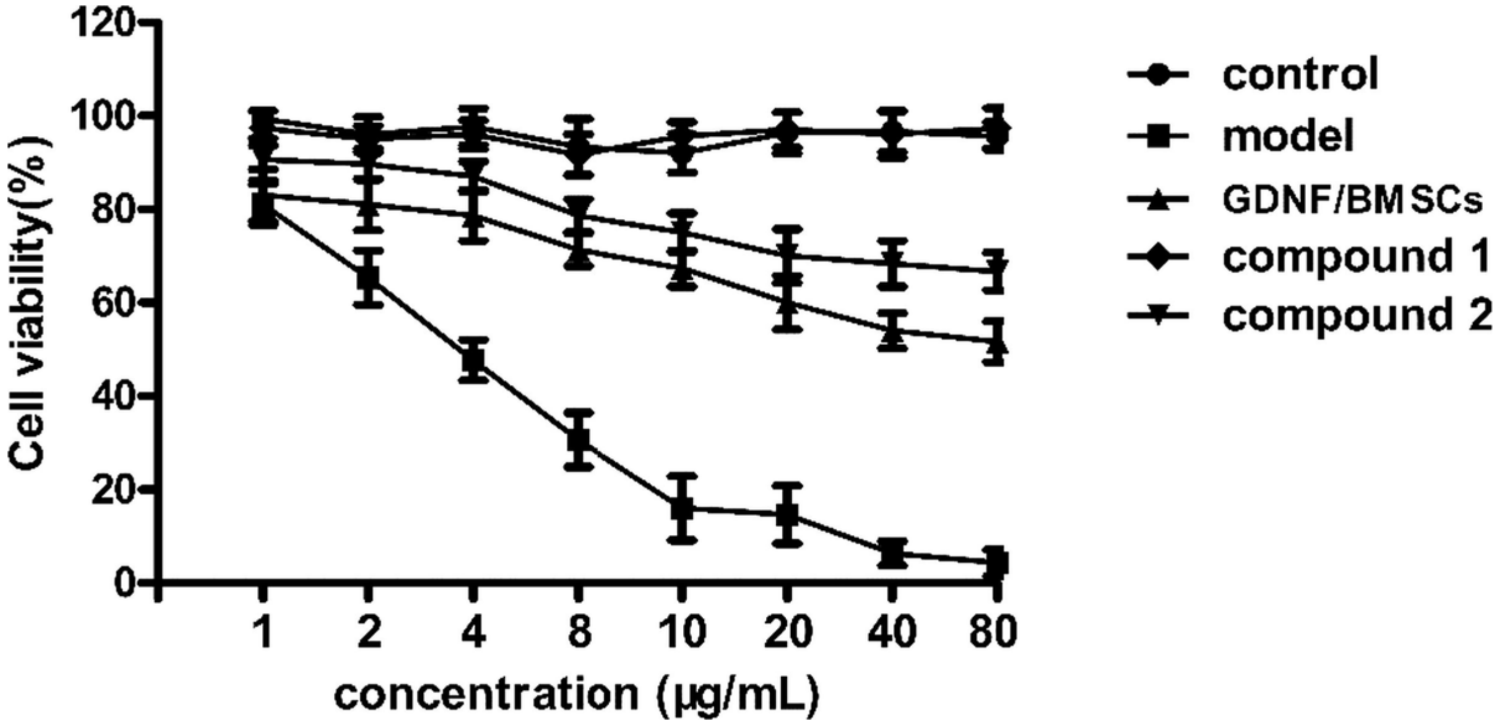

Figure 3

Fig. 6 from Huang et al (2020), “Increased viability of the nerve cells after compound exposure. The nerve cells were planted into 96 well plates and undergo the hypoxia treatment, then compound 1 or 2 was used for treatment”. The text does not mention a ‘hypoxia treatment’, while the figure contradicts the text completely. Under CC BY-NC-ND 4.0 license.

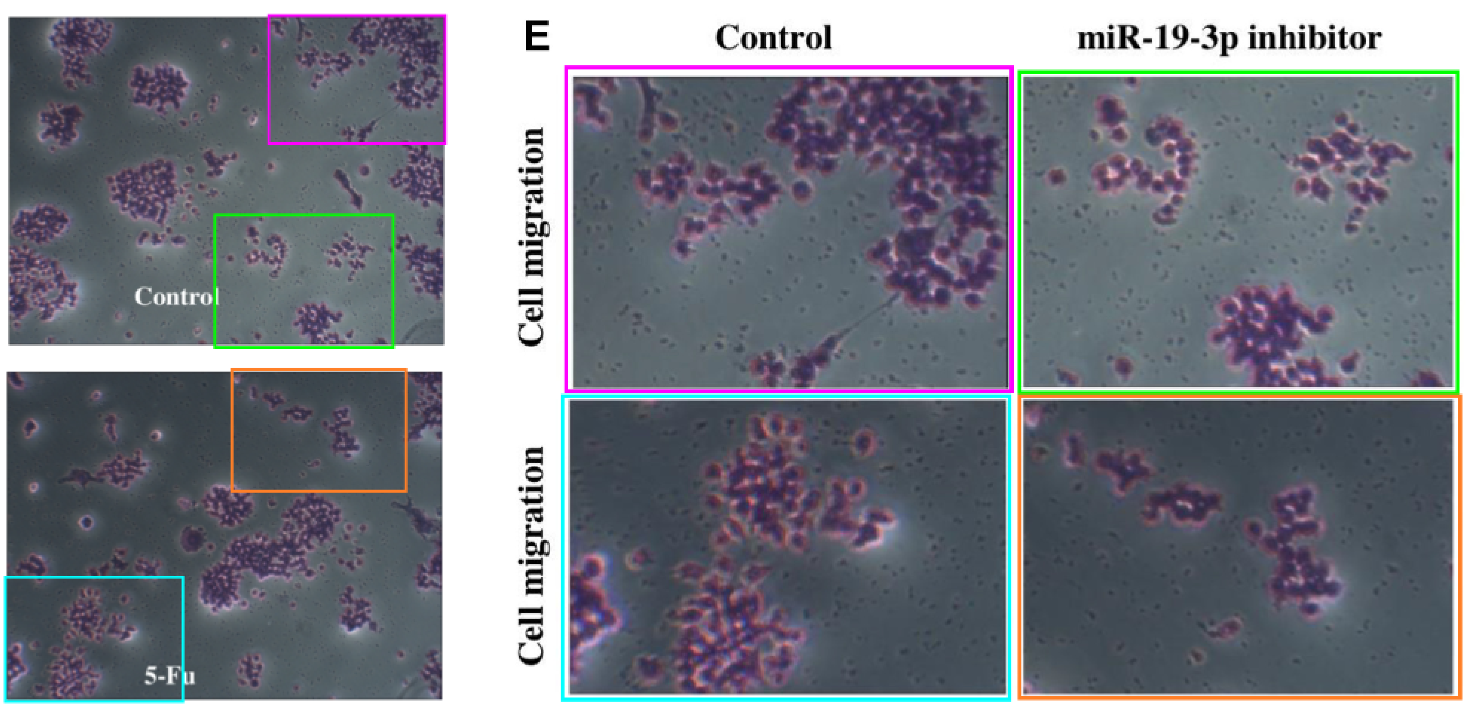

Figure 4

Colony Formation images from two unrelated studies, annotated to show overlaps.