**Case Study: Plagiarism in Medical Dissertations in Germany**

By Debora Weber-Wulff, University of Applied Sciences, HTW Berlin

*weberwu@htw-berlin.de*

1. Introduction

Generally speaking, German medical dissertations are not taken seriously in academic circles. As has now been shown by the collaborative documentation effort of the VroniPlag Wiki academic community, many of these theses are not just short, simple, or shallow, or suffer from fabricated data: There is also heavy and blatant plagiarism. When plagiarism is discovered and publicly discussed, it has been quickly dismissed in the past as a very rare situation, an absolute singularity, as the dean of the medical school in Münster insisted after a public discussion about a plagiarism case in 2011 [1].

Plagiarism in medical doctoral dissertations in Germany is not, as it turns out, a singularity. And it is not just a half-sentence copied here or there or a missing footnote or a mal-formed reference entry. The copying is brazen, often extending over pages, and includes either large swaths of verbatim copying, or slightly rearranged passages from online sources.

Universities are overwhelmed, as in addition to having to deal with the many reported cases of academic misconduct; they must also attempt to staunch the advancing tidal wave of plagiarism. This paper will look more closely at the problem.

2. Medical Dissertations in Germany

Early complaints about medical doctorates can be found in historic documents [2, 3]. More recently, the German Council of Science and Humanities (Wissenschaftsrat), a science policy advisory body, noted back in 2004 that the medical doctorate was not on par with other scientific doctorates and often constituted only “pro forma” research [4, p. 75], although a medical doctorate is fiercely defended by members of the medical academic community [5].

Doctorates in medicine in Germany today are a mass phenomenon. More than 60 % of all graduates will also complete a doctorate, although the percentage has been falling in recent years: Only in physics and chemistry does a higher percentage of graduates do doctoral level work, other fields have much lower rates [6]. The official German statistics show that only about 6 % of the student population at doctorate granting institutions in Germany were studying medicine or dentistry or other health sciences in the winter term 2014/15 (157 166 out of 2 698 910) [7, 4.1, p. 322], while at the same time 7 326 out of 28 147 or 26 % of the doctorates granted were in these fields [7, 4.2, p. 19].

Medical doctoral theses in Germany are frequently very thin volumes that are hurriedly prepared in parallel with the already stressful study of medicine, sometimes even starting in very early semesters, instead of waiting until after graduation. A thesis can only be handed in, however, when the study program has been completed, sometimes years later. The topics are often not chosen by the students, but handed out by the academic advisor.

Some medical schools, for example Münster, accept one or several publications with multiple authors in an international journal as a thesis by publication, often with the advisor as co-author. Although it is understandable that the research groups work
closely together and are eager to publish their results, it is often not made transparent what part each member played in the creation of the publication. And if the advisor is also co-author, he or she is in essence grading their own work. Since there are hundreds of international predatory publishers offering pay-to-publish models that only feign rigorous peer review, it additionally becomes difficult to judge the quality of such publications [8].

The goal for many doctoral students in medicine is to finish the work as quickly as possible, as they are understandably eager to start applying their knowledge in practice. Since the general public in Germany calls a practitioner in medicine the same as an academic who has completed a dissertation, Doktor, the major focus for many students appears to be on obtaining the honorific associated with the degree. However, it is a requirement for the use of that honorific in Germany to have a published dissertation, that is, a contribution to research. Only after publication is the degree conferred and can then be used as part of ones legal name. Thus, students look for efficient ways to complete and publish their theses. Open Access publication channels at a university are seen as a fast track to publication.

The report issued by the German Council of Science and Humanities is a scathing assessment of doctorates in medicine and on the teaching and research conditions in German medical schools [4]. They doubted that there was much to be learned in the preparation of these theses, especially because the supervision situation is so poor that not even basic research methodology can be taught. They found the scientific contribution of this “pro forma” research to be highly questionable, and noted that the focus is more on the occupational and social recognition imparted by the title of “Dr.” than on answering research questions, although there are also many excellent theses to be found.

To add insult to injury, the European Research Council decided in May 2010 that the German “Dr. med.” is not acceptable as an equivalent to a Ph.D.—this prevents German medical graduates from obtaining research funding and post-doc positions in other countries. They must do additional work, such as a habilitation, to be considered equivalent. [9].

Even though there has been a sacrifice of quality in order to gain efficiency in the process of obtaining a doctorate, the perception in the medical community has been that as long as the data is not falsified or fabricated, a thesis is considered to be acceptable. Made-up mice, creative picture manipulations, or dropped outliers are quickly seen as academic misconduct, but plagiarism was not considered to be an issue in medicine.

But as the chair of the Committee on Doctoral Dissertations at the Charité in Berlin, Jörg-Wilhelm Oestmann, noted in a recent article in the Deutsches Ärzteblatt: “Up until recently, I was of the opinion that with my almost 20 years of experience in the process of granting doctorates, plagiarism is not a problem in medicine. ‘We fake data, but plagiarism is not an issue.’ The developments of the past year have disabused me of that notion. There are currently 27 cases of plagiarism under investigation at the Charité for theses submitted in years past. Other medical schools are confronted with the same problem. Quite a number of doctors will probably be losing their academic degree.” (Translation by the author) [10].

What led up to the discovery of these cases of plagiarism that caused Prof. Oestmann to change his mind?
3. Finding Plagiarism in Theses

It is necessary to take a step back and look at what has led to the phenomenon of public documentation of plagiarism in dissertations in Germany.

A group of German-language researchers set up an online public platform called *VroniPlag Wiki* [11] in March 2011 for collaboratively documenting plagiarism in published academic works. This online academic community has been documenting text parallels in dissertations and habilitations that could be considered to be plagiarism continuously since then. They have published documentations of 152 cases in various disciplines and all over Germany, as well as a number of other European countries, as of October 1, 2015.

The group was formed in response to widespread public discussions about plagiarism that arose in Germany in early 2011. The popular Minister of Defense, Karl-Theodor zu Guttenberg, had submitted a dissertation in law to the University of Bayreuth in 2007 that turned out to be an extreme plagiarism on more than 90% of the pages, starting from the first line of the thesis. In an intensive collaborative effort, this plagiarism was publicly documented at a site called *GuttenPlag Wiki* [12]. The media echo was extensive, and zu Guttenberg stepped down as minister on March 1, 2011.

The VroniPlag Wiki group has been documenting plagiarism in medical dissertations and habilitations from its beginnings. To date, the group has published documentations on 53 medical doctorates, 6 medical habilitations, 25 doctorates in dentistry, and three in veterinary medicine. These 87 theses were documented by the group on the basis of tips from well-informed whistleblowers, or from serendipitous finds, or were identified by either cross-checking a small cluster of theses or by a more systematic search for potential plagiarisms through tens of thousands of theses available online. Five of these theses documented as being plagiarisms were not published digitally, the rest were available from Open Access offerings of the various universities affected. Two of the 87 documentations are for theses accepted at Spanish universities, one from Italy, the rest are from medical schools throughout Germany.

The first medical thesis documented was a curious habilitation from the Charité [11: Ut] that contained large portions verbatim from three dissertations for which the author had been the advisor. Two of the dissertations were published before the habilitation, one was published a year after the habilitation was submitted. However, one detail about the numbering of the illustrations showed that this dissertation must have been available at the time the habilitation was submitted. The Charité reported that they found no academic misconduct in this case, as the authors were all members of the same working group. More detail about this case, and the work of the VroniPlag Wiki group, can be found in the author’s book, *False Feathers* [13].

Contrary to popular notions, the VroniPlag Wiki group does not have specialized software for automatically finding and documenting plagiarism. It also does not use a software system such as *Turnitin* or *PlagScan* or any of the many other software systems that purport to determine plagiarism. Instead, the sources for plagiarism are discovered using simple tools such as a search machine, and good old-fashioned, time-consuming research methods that also include reading the thesis and looking for signs of copying.

From early 2014, a somewhat more intensive investigation of medical dissertations that have been published online in university library collections was mounted. It seemed
unimaginable that anyone would knowingly publish a plagiarism in an Open Access publication, because then anyone in the world would easily be able to detect the copied portions. However, dozens of theses have been discovered and documented that plagiarize to quite alarming extents, with pages and pages taken verbatim from various sources, and the investigation has only just begun.

It is not difficult to download theses from a university library repository, although each has its own quirks and methods of accessing the files. It is computationally quite simple to compare one file with each of the others in a corpus, noting the similarity factor. It is, however, quite time-consuming, as the number of comparisons grows rapidly with each additional file to be compared with all of the others. The major effort, however, is still to come: the documentation. More technical details of this dissertation-mining project are discussed in two blog articles by the author [14, 15].

When looking for potential plagiarisms, a tremendous number of false positives were found. There were identical copies of theses, as apparently students can upload multiple versions of their theses. However, this is the same thesis submitted by the same person, perhaps with a minor correction on a page or two. Other false positives include uploads that only consist of a title page—the name of the university and the department are then enough to appear to be plagiarism—, or joint work that is stated as such in each thesis.

At first, only the theses from one institution were compared with each other, then the theses from the institutions were compared pairwise with all of the other institutions. Each investigation turned up a shocking number of potential plagiarisms that then had to be carefully documented.

The documentation process, even for a short thesis, is quite laborious, as discussed in [13] and summarized here. A researcher manually creates so-called fragments in order to document online a text parallel between at most one page of the thesis and only one source. If there are multiple sources on the page, multiple fragments must be set up, and each is annotated with metadata about the page and line numbers, in order to assist a reader in locating the portion of text in question. One researcher also documents the bibliographic information for the source, and at least one other person must review each fragment before it is counted. Navigational aids are prepared semi-automatically in order to help the researchers get a good overview of the extent of the plagiarism.

When a documentation is deemed extensive enough, the names of the author and the advisors are published online and a report is generated and sent to the degree-granting institution. There is no hard definition of what "extensive" means: each case must be judged individually. The documentation can still grow, as other researchers find additional sources. It is never possible to state that a thesis is plagiarism-free, as there could still be a source that has not yet been found. The general public often misunderstands this. They expect there to be some sort of magic software that finds all plagiarisms, or gives a text a clean bill of health. The latter is generally impossible. Software can only point to potential plagiarisms, it can never determine originality. And if it does find a text parallel, it cannot determine if it is properly referenced or not. Only a human being is capable of determining plagiarism.

4. Patterns of Plagiarism
There appear to be three major patterns of plagiarism in medical theses, although there are many theses that combine two or even three of these patterns. These are particularly visible in the two major clusters that have been investigated up until now, the University of Münster with 23 documented cases and the Charité in Berlin with 33 cases.

4.1 Intra-institutional Plagiarism

If one suspects plagiarism in a medical thesis in Germany, a good place to start looking for a source is in the local research group. The habilitation of the advisor and the theses of previous doctoral students can be the sources for incredulous amounts of unreferenced text—and data—overlap.

One example is a dissertation with 85 net pages submitted to the University of Heidelberg [11: Nk] in 2002 but not defended until 2006. Almost three quarters of the pages are taken verbatim or with only minor changes, including much of the data, from the habilitation of the advisor, which was published in 1995. Fragments on two pages were taken from another dissertation prepared under the same advisor that was published in 1998. This dissertation, too, copied from the habilitation of the advisor [11: Awb]. In both cases, the university decided that academic misconduct was given, but only a censure was issued. The doctorates were not rescinded. Similar webs of interlocking habilitation and doctoral theses would show up time and again.

Another example of intra-institutional plagiarism was not documented by VroniPlag Wiki, but was found by a German-language Wikipedia editor. Just a few weeks after zu Guttenberg resigned, the editor was researching a specific type of cancer. Two dissertations from the University of Münster were found on that topic, one from 2006 and one from 2009. The editor was at first excited to have found two theses and not just one, as doctoral dissertations at least tend to have extensive literature surveys, even if they are short. Disillusionment soon set in, as it became apparent that the latter thesis was for all practical purposes a copy of the earlier one, except that the authors were not identical and did not state that they had worked together [16]. The university was informed, and in July 2011 they rescinded the 2009 doctorate. The dean is rather presciently quoted as noting that, “even after years, plagiarists are in danger of being found out” [1].

Twenty-three highly plagiarized medical and dental dissertations have been found to date in Münster. One of the first theses to pop up was [11: Gt]. This 2010 thesis corresponded on 100 % of the 61 pages with a 2009 thesis with the same advisor [11: Ckr]. This prior work is not referenced at all in the thesis. Although the theses were about investigations dealing with the eyes of two different type of ape, the texts were identical, as were many of the figures and illustrations. The numerical data, however, was different although very similar. Looking closely at the numerical results it could be seen (for example on p. 48 of Gt) that the numbers had just been arbitrarily changed. However, the averages and standard deviations reported in Gt did not match the data as presented.

The intra-institutional plagiarism did not stop there. Ckr was just the second link in a chain: A plagiarism of a plagiarism of a plagiarism was found. All these theses were prepared under the same advisor. The University of Münster has just recently announced that it will be sanctioning the professor who advised this chain, as announced shortly after the conference [17].
A second web of major intra-institutional copying was found in Münster that included six dissertations that were again all with the same advisor. These were dissertations for dentists who were writing about the effects of this or that on rat brains. Anywhere from between 30% and 86% of the pages exhibited extensive plagiarism, often from older dissertations at the institution. Another sixteen theses in Münster have been shown to be suspicious, but the volunteer academics at VroniPlag Wiki did not want to continue the documentation, as there was another complete intra-institutional plagiarism found at a different medical school.

This 2010 thesis was submitted to the Charité in Berlin [11: Ali]. The 55-page thesis about prostate cancer was also a copy on 100% of the pages from the text of a thesis prepared two years earlier with the same advisor, without any reference to that thesis. Only five sentences could not be found in that thesis. Here, as with Gt, the numbers were changed in the basic data, but the percentages given did not match the data: they were, however, identical to the source. The audacity of this case, and the prominence of the medical school, led to it receiving at least some press coverage [18].

In addition to three previously documented cases of plagiarism, thirty new ones were found at the Charité with eight more deemed suspicious. Five new ones plagiarized from only one source: another dissertation or habilitation from the same institution. Even if the others had more plagiarized sources, one more thesis [11: Alm] had plagiarized fragments on all of its 24 pages, most of them from the habilitation of the advisor.

There was also an interesting network of six theses in forensic medicine at the Charité about age determination of young people according to properties of teeth. All six had the same advisor. The theses copied extensively from either the habilitation of the advisor, or from each other.

This intra-institutional plagiarism points to a specific form of academic corruption. Here, the results of the research, including text and data, seem to be considered to belong to the group and are reused at will, without making the provenance of the data and text clear. It appears that some sort of tacit approval for this type of collusion, that is, the extensive re-use of text amongst members of the same institution, exists.

4.2 Inter-institutional Plagiarism

Apparently, some dissertation writers also turn to dissertations published online at other universities, and copy bits and pieces from them in order to assemble a text in which to embed their own findings. A typical example of this can be seen in a thesis from the University of Bonn [11: Mak]. This 34-page dissertation that was concerned with calculating sums, averages, standard deviations, and quotients for a given Excel sheet, embedded the tables in 18 pages of material copied out of five other dissertations from three universities.

Another thesis submitted to the University of Münster in 2007 [11: Bm] includes extensive material from 17 doctoral dissertations accepted at other universities, a diploma thesis, and one habilitation that are all available online. It is interesting to note here that although the thesis was submitted in 2007, the entries in the bibliography are all much older. There is one reference from 2002, the rest are from 2001 or earlier. This can be a good indicator of potential plagiarisms.

At the University of Mainz, a thesis was submitted in 2009 about wound healing for a doctorate in dentistry [11: Tz]. Curiously, the author already had two doctorates, one in
medicine from Mainz in 2004 and a science doctorate from Frankfurt in 2008 on a medical topic. Over half of the 30 pages that make up the dissertation are taken almost verbatim from a thesis in medicine from Gießen that was published 2003.

Medical theses are also serving as the sources for plagiarism in other, related fields. A thesis in science on a medical topic that was submitted in 2010 to the University of Vienna in Austria [11: Ves] contains material taken verbatim or slightly changed without reference from eighteen theses in medicine or science from eight different universities, as well as fifteen chapters from a handbook. It is interesting to note that in this case, the university states that it has been using plagiarism detection software since the winter term 2007/08 [19]. This demonstrates again that software cannot reliably detect plagiarism. It is only possible if the sources are found in the software databases, which they apparently were not in this case.

In the pattern of inter-institutional plagiarism, the standard excuse of “joint work” that only forgot to properly acknowledge the collaboration, falls away. The pattern also shows that Volker Rieble, a law professor from Munich, was unfortunately correct in his 2010 assessment that Open Access publications would only encourage plagiarism [20, p. 54–55]. However, the solution is not to ban Open Access, but to keep people from plagiarizing. It is because the theses and the sources are publicly available that the plagiarism documentations have been possible.

4.4 Extra-institutional Plagiarism

Textbook material, diverse pages from the Internet and especially the Wikipedia have turned out to be often-used sources, both referenced and not, for medical theses. Comparing theses with the appropriate Wikipedia version has turned up alarming amounts of text copied verbatim. It has been found to be the source for topics in neighboring disciplines such as history, statistics, or learning theories. But one also finds texts describing basic medical knowledge taken 1:1 without reference from the Wikipedia or other web pages. One worries that this is the result of harried medical students not even having the time to look up topics in a textbook, but rather trusting on the wisdom of the crowd, as codified in the ever-changing Wikipedia.

Eighteen medical dissertations to date have been determined to have at least one unreferenced fragment from the Wikipedia. Some have quite extensive sections copied verbatim. For example, a thesis submitted to the University of Bonn in 2010, [11: Go], has nine pages out of the total of 61 thesis pages taken from just the Wikipedia article on epilepsy, as well as two fragments about “Matlab”, the software used in the research.

It appears that those who have copied from the Wikipedia have not stopped there. If there is at least one Wikipedia text found in the thesis, other sources have been found that were copied without reference. Two theses, one from the Charité [11: Anh] and one from Münster [11: Lh] each have unreferenced text taken from six different Wikipedia articles.

5. Dealing with the Plagiarism Problem

The extent of plagiarism found in medical dissertations in just a small sample was unexpected, both for the plagiarism researchers and the affected universities. The investigation is by no means finished. Since the documentation is time-consuming and the researchers with VroniPlag Wiki work as volunteers, the findings thus far can only be interpreted as demonstrating that a problem exists, and it is not just the problem of a
single university or field. Only the tip of the iceberg has been exposed. The universities
must be proactive, or they will be forced into having to react to increased plagiarism
allegations. Not only VroniPlag Wiki can find these plagiarisms, they are published
worldwide. Anyone can—and eventually will—see the problem. The sheer amount of
brazen plagiarism is what has convinced Prof. Oestmann that medicine, too, has a
plagiarism problem.

The reaction of universities to the documentations sent to them has been very varied.
Some do not even bother acknowledging the receipt of the notification. Although the
rules at many universities stipulate that the informer is to be advised of the end result of
the investigations, they often are not. When pressed for information, some universities
respond with allusions to privacy laws, although the object in question is a scientific
publication, not a personal matter. More detail on the reactions of the universities can be
found in [21].

When universities do get active, they tend to address the symptoms. They purchase so-
called plagiarism detection software in the hopes that it can quickly and reliably
separate the wheat from the chaff. Unfortunately, as the author has repeatedly shown,
all systems are plagued with both false positives—properly quoted material is flagged as
plagiarism—and false negatives, that is, the software finds no plagiarism where there is
much. In addition, each system will generally find different sources and report widely
different “scores” attempting to quantize the amount of plagiarism. See [13, 22, 23] for
more in-depth discussions of the problems associated with plagiarism detection
systems.

Another attempt to deal with the problem is to set up mandatory or elective seminars on
good scientific practice for doctoral students. This is in and of itself always a good idea.
Preventing plagiarism from happening is much better than having to sanction students
after the fact. But it will not solve the problem of plagiarizing students, as some have
learned this behavior in high school and see others, in particular some professors,
plagiarizing material written by, for example, junior researchers, with impunity. It is
essential that the topic of academic misconduct be discussed openly throughout the
university, and that the top-level management of the university be committed to
encouraging good scientific practice in word and deed.

There are communication problems associated with sanctions such as revoking doctoral
degrees. One is concerned with the person, the other with the academic community. On
a personal level, in Germany the honorific conferred with a doctoral degree is often used
socially and professionally. The honorific is documented on the legal identification
documents and prominently displayed on doorbells and shingles. When a doctorate is
revoked, it is up to the person who has had their dissertation revoked to remove all
traces of the title, but there is no agency that oversees this process.

More important to the academic community is the communication about the publication.
This would also be necessary for cases in which a doctorate is not rescinded, but only a
censure given. The text is still a plagiarism, and future researchers will probably be
unaware of the problems with the text. When a doctorate is rescinded, the library
catalogue needs to be amended. Either the notice that the text is a doctorate needs to be
removed, or a statement that the document in question is considered a plagiarism needs
to be placed into the catalogue, such as one for [11: Gt] that is found at [24].
Removing the document from an online repository may be necessary for copyright reasons, but there should be a page that responds to the URL that informs the reader of the situation. It is possible—and there are documented cases of this happening, for example [13, p. 48]—that the thesis is quoted in later years, as authors have no knowledge of the academic misconduct connected with the thesis. Journal papers can be retracted or corrections issued, but this is not possible for a dissertation. Even if a second edition of the thesis is printed (and accepted by the university), it does not replace the first one. The plagiarisms need to be kept by the library for future researchers, perhaps those that are investigating plagiarisms.

There is one solution that is specific for the field of medicine in Germany. All students who successfully complete their medical degrees should receive the title of M.D. together with their license to practice medicine. The study program should include a focus on understanding research, but the practice of accepting substandard theses should be terminated. Students who are interested in doing research can embark on a Ph.D. program, as in other fields.

6. Summary

The work of VroniPlag Wiki has demonstrated that there is a massive plagiarism problem in medicine in Germany. The investigation is neither representative nor anywhere near completion. But it is clear that there are very serious plagiarisms slipping through the quality control processes at many German medical schools.

It is urgent that the medical field gets serious about combatting this problem. Some universities are beginning to understand that they have to act, but purchasing plagiarism detection software or setting up seminars for doctoral students is not enough. This only addresses the problem on a superficial basis.

It is a systemic problem and needs to be addressed at a systemic level.

Acknowledgments

I have been a member of the VroniPlag Wiki academic community since April 2011. The energy and the immense time effort all active members have put into the meticulous collaborative documentation of the plagiarism cases has provided an extensive data basis for observations such as these. Our countless fruitful discussions about the plagiarisms found have been instrumental in helping me to focus on this problem. I am deeply indebted to one member in particular who wishes to remain anonymous. His curiosity and focused determination have contributed immensely to the investigations of medical thesis. This has truly been a collaborative effort, any errors in my reporting are entirely my own.

Bibliography


Note: Individual cases on the VroniPlag Wiki site will be cited as [11: Xxx], using the case abbreviation


