Bridging from drug registration trials to meaningful clinical evidence: The case of schizophrenia

Mattila, T.K.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 6
Reliable data on efficacy and safety of medicinal products is essential for decision making by regulatory authorities, for composing treatment guidelines by professionals and ultimately for prescribing medications by clinicians to their patients. Randomized, controlled trials constitute the basis of this evidence, including drug registration trials, performed to obtain marketing authorisation for a new medicinal product. In an ideal world of transparency, data from these trials would be available to everyone. However, in the current situation, we rely mainly on publications of clinical trials in scientific journals and unfortunately, not all of these trials are published, and in some cases the published trials may lead to distorted conclusions due to selective reporting. This became evident to me when I examined publication bias in insomnia medication trials.

The benefit of access to raw study data in contrast to relying only on published data is that raw study data enables not only analyses on efficacy of medicinal products, but also creates the opportunity to test additional hypotheses relevant for regulatory and clinical practice. This is particularly interesting in the field of psychiatry, where drug development has been declining in recent years (1,2). One of the factors related to the fact that this field is becoming less appealing to pharmaceutical industry is the unpredictable behaviour of a compound in Phase III clinical trials despite promising results in earlier phases of development (1). This is partly due to the complex nature of psychiatric disorders, which are far from homogenous, as currently acknowledged by the experts in the field and also reflected in the most recent (5th) edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) (3). This is certainly true for schizophrenia, which is nowadays acknowledged as a complex disorder presenting in different signs and symptoms which vary from patient to patient and over time. This creates challenges for drug research in schizophrenia: how to disentangle the true effects of a drug in a jungle of symptoms which are not always strongly interconnected to each other? There is also an array of patient characteristics which may modify response to treatment. For example, there are differences between recent onset and chronic schizophrenia which may result in different outcomes between patients with different disease durations (4). Another commonly mentioned problem in clinical trials in schizophrenia is the lack of external validity: the included patients do not reflect the patients in clinical practice with comorbidities, such as substance abuse (5,6). Furthermore, there are uncertainties with regard to certain trial design aspects, such as sufficient trial duration, lack of hard outcomes and the lack of consensus about the definition of response on the commonly used assessment scales, which may complicate performing and interpreting schizophrenia trials (4,7-9). Studying the impact of these factors on compliance, retention and outcomes is often difficult or, impossible in the case of aggregate data, because they do not have sufficient details. Access to raw study data and analyses on this individual patient data (IPD) may help to study these factors and may facilitate regulatory and clinical decision making. I had the opportunity to investigate three important current
challenges in schizophrenia research using an IPD database, including 22 double-blind, randomized, placebo-controlled trials of five different second generation antipsychotics (SGAs) for the acute treatment of schizophrenia. Access to this database enabled me to investigate geographic differences in efficacy of these compounds, the implications of changing diagnostic criteria over time and the effect of these medications on insight in schizophrenia, an important but hardly studied aspect in the disease course.

Summary of findings

Publication bias in the context of insomnia medications
In chapter 2, I examined the extent of publication bias in insomnia medications and found that only half (55%) of placebo controlled clinical trials performed with three insomnia medications recently submitted for registration in the EU were published in peer reviewed scientific journals. Positive trials were more likely to be published, leading to a more biased view of efficacy of these products compared to the true effect size based on all performed studies: the effect on sleep latency was 1.6 times larger in the published data as compared to the complete data. In addition, the published trials selectively reported positive study endpoints and additional post hoc analyses. Furthermore, important safety data concerning withdrawal effects, rebound insomnia and next day performance were lacking in majority of publications.

Geographic variation in efficacy of second generation antipsychotics
In chapter 3, I moved on to examine whether there are differences across geographic regions (North America, EU and ‘other’) in the efficacy of SGAs for the acute treatment of schizophrenia. SGAs were more effective than placebo in all geographic regions, but I found that the magnitude of the effect of the five compounds was smaller in studies conducted in North America (Hedge’s g 0.37) as compared to studies conducted in the EU (Hedges’s g 0.56) and the rest of the World (Hedge’s g 0.53). This difference was not explained by an unequal distribution of studied compounds across regions.

Impact of DSM-5 changes in schizophrenia
In chapter 4, I tested the impact of the changes in diagnosis of schizophrenia from the 4th edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) (10) to the 5th edition (DSM-5) (3). I found that the change in diagnostic criteria does not drastically change the diagnosed patient population, indicating that results of clinical trials in patients diagnosed according to pre-DSM-5 criteria are also valid for patients diagnosed according to current (DSM-5) criteria. In addition, I found that the decision to delete schizophrenia subtypes from the DSM-5 is justified from a pharmacotherapy perspective, because the effects of SGAs do not differ between the DSM-IV subtypes.
also found that individualizing treatment to target certain symptom dimensions may not be realistic or desirable, as patients who respond to treatment improve in all symptom dimensions.

**Impact of second generation antipsychotics on insight**

Finally, in chapter 5, I examined the effect of SGAs on insight in schizophrenia and found that SGA treatment resulted in a significantly larger improvement in insight than placebo. However, this improvement was significantly associated with improvements in other symptoms. This is in contrast to earlier studies which have indicated that the effect on insight cannot be attributed to the improvement in other symptoms. I further found that the effect of SGAs on insight was not associated with geographic regions, illness duration or drop-out status.

**General discussion and future directions**

In this thesis I examined three important issues in modern schizophrenia research that are relevant for the pharmaceutical industry, regulatory authorities and clinicians and their patients. Optimizing the design of Phase III clinical trials for psychopharmacological agents is highly desirable and identifying and including factors that may moderate the (side) effects of drugs may result in (much) better effect sizes in patient subgroups. In addition, adaptations of the scientific regulatory guidelines regarding the performance and analysis of clinical trials may help to improve the current situation in drug development in psychiatry and ultimate the treatment of patients.

In this era of general globalization, one issue that which regulatory authorities and investigators struggle with is extrapolation of study results from one geographic region to another. Decisions are made on a case-by-case basis by the Food and Drug Administration (FDA) in the US and the European Medicines Agency (EMA) in the EU, and mainly depend on an analysis of (dis)similarities between geographic regions in intrinsic and extrinsic factors which may contribute to differences in efficacy and safety of a product between regions. My study on geographic variation in efficacy of SGAs showed that differences between geographic regions (e.g. North America, EU and other countries) do exist. However, since the regional differences in effect size were not statistically significant, one might conclude that extrapolation of study results in this case is possible. It should be noted that a rather small number of patients from the EU was included in the studies which might explain the statistical insignificance. On the other hand, the effects of the different SGAs were significant and clinically relevant in all geographic regions, which seem to justify extrapolation. However, in case of smaller effect sizes with different medication in different conditions (e.g. depression) overall,
Chapter 6

regional differences could lead to different conclusions in different geographic regions with significant effects in one region and non-significant in another region. Unfortunately it was outside the scope of the current study to examine possible explanations for the observed regional differences. However, a recent study in acute mania demonstrated that the only variable that contributed to the observed, statistically significant difference in the efficacy of antipsychotics in the treatment of acute mania in bipolar disorder was baseline severity, but the amount of explained variance was very small (11). This indicates that adjusting the study population and study design across regions may not overcome the problem of uncertainty with regard to the extrapolation of trial results from one geographic region to another. Identifying the reasons behind geographic variation in efficacy of medicinal products is not easy, because there may be many factors that each play a small role. Moreover, it is not only simple factors such as patient characteristics or trial methodology, but probably also more complex factors such as regionally different therapeutic approaches, drug compliance and the entire treatment context, including a range of socio-economic factors. Therefore it seems that there cannot be universal guidance on how to design a study in one geographic region with unproblematic extrapolation of results to other regions. In addition, extrapolation might be disease and medication specific and thus a previously mentioned case-by-case approach seems the only possible solution at the moment. A decision to extrapolate or not should at least take into account the observed effect sizes across regions, known heterogeneity in patient populations regarding genetic, physiological and psychopathological factors. A certain level of uncertainty also in this respect, as always with drug development and authorisation, has to be accepted. Studying geographic differences and examining the reasons behind it is nevertheless highly recommended, in an attempt to better understand the phenomenon and possibly amend regulatory guidance in line with the findings.

Another issue in clinical trials, also in schizophrenia, is the diagnostic criteria which are regularly updated. The process of updating the criteria generally takes a long time with extensive discussions between experts, and reaching consensus may take years, as was the case with the most recent (5th) update of the DSM, which was published in 2013 after 13 long years of preparation. As standard practise, patients are included in drug registration trials according to the diagnostic criteria in place at that time the study protocol is submitted. If considerable changes in these criteria are made, the target population to be treated with the study drug may differ from the studied patient population. The changes implemented in the diagnosis of schizophrenia in the DSM-5 did not raise major concerns over the implications to clinical practise (12), although a concern was raised that the simplification of the criteria which took place might increase inclusion of false positive cases (13). My analysis on the drug registration trials dataset revealed
that >99.5% of patients diagnosed with the DSM-IV criteria would also be diagnosed with schizophrenia according to the DSM-5 criteria, suggesting that the results of pre-DSM-5 trials can be extrapolated to DSM-5 diagnosed patients. My dataset included patients diagnosed with DSM-III (14), DSM-III-R (15) and DSM-IV (10) criteria, thus the numerous changes across the years in diagnostic criteria have had very little impact on the diagnosed patient population. This is reassuring information for the field – results from old studies are relevant for patients in present day practise.

Examining the impact of changes in diagnostic criteria is relevant in disorders where commonly used medicinal products have been tested decades ago and diagnostic criteria are amended with advances in knowledge over the disorders, such as in rheumatologic conditions. For example, the ACR/EULAR classification criteria for rheumatoid arthritis (RA) published by the American College of Rheumatology and the European League Against Rheumatism was updated in 2010, and in comparison to the previous criteria from 1987, a larger (and probably more heterogeneous) patient population would be diagnosed with RA, as patients with early disease would fall under the new diagnostic criteria (16). The prognosis and perhaps response to treatment may differ between early and established disease, and therefore it may be questioned whether results from early studies are applicable to the current patient population. Therefore, the current European Medicines Agency (EMA) draft guideline on clinical investigation of medicinal products other than NSAIDs for treatment of rheumatoid arthritis (17) calls for separate trials in early versus advanced RA patients.

The issue of changing diagnostic criteria is also acknowledged when assessing the efficacy of a new medicinal product against the available treatments when direct comparison in a clinical trial is missing. Analyses of large raw datasets can clarify the differences in patient populations and therefore either support or advice against using historical controls. In this respect, I strongly support examining the impact of changing diagnostic criteria to facilitate the assessment of the position of a new medicinal product in the treatment guidelines.

To some extent, raw study data can also be used to validate the implemented changes in diagnostic criteria. In my study on changing diagnostic criteria, I confirmed that the response to SGAs does not differ between schizophrenia subtypes, supporting the decision to omit these subtypes from the DSM-5. It should be noted though, that response to treatment is only one aspect of subgroup differences.

I was also able to evaluate whether the newly implemented psychosis assessment scale in DSM-5, Clinician-Rated Dimensions of Psychosis Symptom Severity (C-RDPSS), is a valuable addition to the assessment battery in clinical practise. This scale was included in the DMS-5 to reflect the current view of schizophrenia as a disorder with symptom dimensions, each of which requiring their own characterisation and assessment. At first sight, it seemed that the response to SGAs was clearly different for the various symptom
domains with only minimal improvements in abnormal psychomotor behaviour, negative symptoms and impaired cognition, and substantial improvements in hallucinations, delusions and disorganized speech. However, when examining patients defined as responders (>30% reduction in total BPRS score), all symptom domains improved in these patients. This highlights the complexity of schizophrenia: symptom domains are only weakly correlated, but in the response to SGAs they seem to change together, and this may complicate the analysis and interpretation of the effect of a medication on individual symptoms or symptom domains.

My analysis on insight confirmed this finding. Poor insight is considered a factor that negatively impacts treatment motivation, treatment compliance, prognosis, and social functioning in schizophrenia. Interest in studying insight in schizophrenia has grown in the last decades, as some studies indicated that the association between symptoms and insight may not be as strong as previously assumed (18), meaning that insight might be more than just an artefact of symptom severity, automatically improving when symptoms such as delusions and hallucinations subside. Studying the effects of pharmacological treatment on insight is difficult based on aggregate data as individual symptom scores are often not reported in trial publications. In my analysis on individual patient data I was able to overcome this problem.

As stated earlier, schizophrenia is currently perceived as a cluster of symptom dimensions, examining the effects of medicinal products on individual symptoms or symptom dimensions is increasingly interesting. This may also be of interest of pharmaceutical companies, taking into account that a compound might perform poorly in treating the overall disease, but may exhibit efficacy in certain symptoms or symptom dimensions. Analyses on raw study data enable the assessment of efficacy per symptom or symptom dimension/domain, as I demonstrated in my analysis of insight in schizophrenia. However, my study showed that improvement in insight was associated with and probably dependent on improvement in other symptoms, which again suggests that symptoms and symptom dimensions in schizophrenia are connected in a complex way and probably dependent on the overall level severity. Therefore studying the effects of medication on these separate symptoms and/or symptom dimensions is challenging.

It should be also noted that these types of analyses should be based on scientific and clinical rationale, to ensure that for example clustering of symptoms and targeting a treatment to this dimension is relevant in clinical practise. When exercised with caution, individual patient data analyses could be used as an aid for repositioning or re-profiling drugs. However, before accepting a compound for the treatment of specific symptoms or symptom dimensions, initial results need to be (repeatedly) replicated and the clinical relevance of these symptoms or symptoms dimensions should be established in new (FDA or EMA) trial guidelines.
The results of my analysis on the three topics in schizophrenia research highlight the complexity of studying schizophrenia. Geographic variation seems to be dependent on many factors and standardising trials across regions may not solve the issue of extrapolation. Fortunately, changing diagnostics did not have a significant impact on the diagnosed patient population, but implementing new assessment methods may not bring the suggested improvements in clinical practise or provide meaningful data. Furthermore, symptoms and symptom dimensions in schizophrenia are connected in a complex way and probably dependent on the overall level severity. This means that studying individual components, such as impaired cognition which is a hot topic in schizophrenia research and considered an unmet medical need at the moment (19,208), may be complicated. For example with respect to studying impaired cognition, uncertainties remain over appropriate trial design including endpoints, duration and definition of response (19). Many compounds with promising results in Phase II studies have failed to show efficacy in larger studies, which may be at least partly due to the problems of studying this dimension in clinical trials (21,22).

As I demonstrated in my study on publication bias in insomnia trials, the perception of efficacy of treatments can differ considerably between analyses based on only published data and analyses based on all performed trials, even if only aggregate data (and not individual patient data) are available, as was the case in my analysis of the insomnia medications using the European Public Assessment Reports (EPARs). This is something to be aware of also when examining current findings from schizophrenia research. Access to raw study data seems to be the best way to avoid the biases associated with published evidence. At the same time one should be aware of the challenges and requirements involved in working with these type of data and analyses. Performing a study with IPD meta-analysis takes more time and resources, as collaboration between several stakeholders has to be negotiated, which also includes assurance of integrity in data handling and analysis, which in itself is burdensome and requires expertise (23). The challenges in performing IPD meta-analysis are also the main reason why meta-analysts do not even attempt to collect individual patient data, as shown in a survey published by Kovalchik (24). In this respect, regulatory authorities seem to be a very suitable party to obtain, handle and analyse individual patient data from drug registration trials, as they have a comprehensive overview of performed studies and established contacts with the pharmaceutical industry.

It should be also noted that access to individual patient data does not automatically mean high-quality output, as also highlighted by Eichler et al. (25). The quality of IPD meta-analyses varies, due to differences in the quality of data collection, processing, and analysis between researchers, which can complicate interpretation of results and conclusions (26). Recently the Cochrane IPD Methods group published guidance for both reviewers and users of IPD meta-analyses to help identify the key issues in
performing and reporting IPD meta-analyses and to ensure solid conclusions, possibly having implications in clinical practise. This is a point of attention also for regulatory authorities interested in IPD meta-analyses, in-house expertise is crucial to ensure high quality research.

**Methodological considerations**

The main strength of my study is obviously the use of a large individual patient dataset of over 5,000 patients with schizophrenia, including 22 studies and five different SGAs. Access to these data allowed me to address questions with regards to some of the important issues in modern schizophrenia research that are difficult if not impossible to examine using aggregate data. In addition, the data originates from randomized, double-blind, placebo-controlled trials which are considered to be the highest level of evidence in medical research. The large sample size also allowed examining subgroups, such as schizophrenia subtypes in my study on changing diagnostic criteria.

The findings of my study need to be interpreted in light of certain limitations. First, the study on publication bias relies on a comparison of publications with the European Public Assessment Reports (EPARs), which are summaries of studies submitted to the EMA for the purpose of marketing authorisation application. It would have been more powerful to perform the comparison against analysis on raw study data, which unfortunately was not available to me. Nevertheless, I found considerable differences between the available journal publications and EPARs, and the observations are in line with previous findings on publication bias, for example in antidepressant trials.

Second, the selection of studies and patients included in the individual patient dataset could be considered limited. Only short-term studies performed with compounds targeted for the treatment of acute psychotic episodes in patients with schizophrenia were included. It can be questioned whether the observations would hold if also long-term studies would have been included or if other outcomes were used. In this case however, the strength of placebo-control and blinding would be lost. In addition, as the data is from RCTs, the included patient population may not be representative of the target patient population, an aspect which is often brought up as criticism towards RCTs for drug registration purposes. This may be especially problematic in patients with an acute psychosis and limited insight, since these symptoms may seriously challenge the requirements for informed consent.

A third limitation is the fact that due to confidentiality reasons I could not specify the compounds included in the analyses, except for the last study on insight. Fortunately the aspiration for transparency is a common interest for both regulatory authorities and pharmaceutical industry, and hopefully in future studies this information can be made public.
Summary and discussion

A fourth limitation is that the data was restricted to a limited number of variables. Other potentially interesting factors, such as treatment history and socio-economic factors were not included in the dataset. This is partly due to the variability in extensiveness of provided data between involved companies. It is therefore recommended that in a request for raw study data it is clearly specified which variables are of interest and should be submitted.

Conclusion

The findings of my thesis clearly show that the clinical development and testing of compounds for the treatment of schizophrenia is complex. Substantial geographic variation in efficacy of medications may hamper extrapolation across regions, but it seems that very little can be done to solve this problem. On the other hand, changes in diagnostic criteria may not challenge extrapolation of past study results to the present day, but attempts to improve assessments to obtain a better correspondence to the current knowledge about schizophrenia as a disorder may not yield meaningful improvement in characterising patients and assessing the efficacy of existing and new medications. Furthermore, the fluctuating connections between symptoms and symptom domains complicate the study of effects of medications on individual symptoms.

Further studies on controversial issues in schizophrenia research are needed. However while gathering evidence it should be kept in mind that data available in the public domain is not necessarily void of biases. While we wait for open access to all clinical trial data, a worthy option is to perform analyses of drug registration trial data to answer these important research questions, which are of interest to both clinicians and regulatory authorities, beyond examining mere efficacy of medicinal products. The results of such analyses can guide not only treatment practises and regulatory decision making, but also the design of future clinical trials. In order to promote the use of drug registration trial data for these purposes and individual patient data analyses in general, awareness of the benefits needs to be raised, while not forgetting the challenges, requirements and investments that are involved. Creating opportunities for researchers to work with drug registration trial data is an esteemed goal of many regulatory agencies, including the MEB, which I fully support.
References

17. EMA. Guideline on clinical investigation of medicinal products other than NSAIDs for treatment of rheumatoid arthritis (CPMP/EWP/556/95 Rev. 2). 2015
Summary and discussion

2013 Aug;23(8):769-78.