Individualized prognosis in childhood immune thrombocytopenia

Schmidt, D.E.

Part V

DISCUSSION

12

GENERAL DISCUSSION

In general, using individual patient characteristics to direct the clinical management is the foundation of evidence-based medicine [156, 166]. At the beginning of the OMICs era, the idea to guide individual prognosis and treatment for patients by detailed genetic, transcriptomic, multimodal, and high-dimensional data gave rise to the related personalized medicine movement [166–168]. With the advanced molecular methods that are available today, a deeper understanding of the pathophysiology of diseases and their subtypes can be achieved, such that the classical disease entities are revised by a molecular taxonomy of disease [169, 170]. Some argue that future doctors may work hand-in-hand with computational tools to support clinical decisions [171].

Advancing knowledge in childhood ITP is challenging given the rarity of the disease and the vulnerability of this study population. The central theme of this dissertation is that clinical and molecular characteristics can distinguish the prognosis of childhood ITP. In a translational medicine approach, we integrated the clinical data of patients collected at the diagnosis with molecular analyses and longitudinal follow-up.

Three main factors made this dissertation possible:

1. A very well-defined and complete clinical cohort. The *Treatment with IVIg in Kids with ITP* (TIKI) trial, with a large study pop-

ulation recruited at the diagnosis, sample storage at a central laboratory, and one-year clinical follow-up, allowed a unique opportunity to study the clinical and molecular features of childhood ITP.

2. Focus on computational methods. We considered data from demographic, clinical, genetic, and immune dimensions, and implemented specific methods that were required for integration of these multimodal data sources [158, 172].

3. Collaboration. A multidisciplinary research collaboration was established involving clinicians treating ITP patients, experts on molecular mechanisms and diagnostic testing in immunohematology, and data analysis and modeling experts. Moreover, a European collaboration with the NOPHO ITP study group was initiated to enable the development and validation of ITP childhood prediction models.

Nordic Society of Pediatric Hematology and Oncology (NOPHO)

In the following, the different results of this thesis will be critically discussed, first as they relate to making prognoses for individual patients; second, how they have changed our understanding of the ITP pathophysiology; and third, what the clinical implications are. Finally, a roadmap is provided to advance the clinical care of children with ITP.

12.1 PREDICTION MODELS FOR ITP PROGNOSIS

Making predictions about a patient's prognosis is at the core of clinical medicine [157, 173]. It is the reason we diagnose, and it is

the reason we give or withhold treatments. Expert clinicians make predictions by their intuitive impression of a patient (clinical gestalt), shaped by their extensive experience. Indeed, clinical thinking is fundamentally probabilistic and Bayesian in nature, *"whether [clinicians] realize it or not"* [174]. Clinicians continuously consider pre-test and post-test probabilities, conditional on tests (e.g., a question from history taking, a sign from clinical examination, results of a laboratory test). Depending on the context, experts' decisions can be as good as, or superior to, protocols or checklists [175]. However, "professional intuition is sometimes marvelous and sometimes flawed" [176]. Experts also make mistakes and are subject to biases. Thus, evidence-based medicine suggests to efficiently integrate clinical expertise with the best available evidence [156].

HOW TO DETERMINE THE PROGNOSIS The evidence-based relationship between the patient characteristics and prognosis (risk) can be formalized in a prediction model [177]. The predictions of such a model can then be systematically evaluated. Reflections of the error rate of tests, i.e., sensitivity, specificity, and likelihood ratio, are well-known to clinicians. There are several levels of development of prediction model [178, 179]. First, a model is developed with a set of parameters ("predictors"). In the case of a binary outcome, the model then returns a quantitative probability. This predicted probability makes it possible to examine in a standardized way how well given parameters predict a response, i.e., it allows an accurate estimation of prediction successes and errors. Secondly, a model is evaluated for discrimination and accuracy in a different population (external validation). The clinical utility for assistive or directive

decision support can be assessed scientifically in a model-impact study, i.e., for the guidance of treatments or monitoring (to improve quality of life, health care utilization)[178, 179]. Ideally, a randomized controlled trial is performed, where the decision-supported clinical management is compared to usual care [178, 180]. In practice, the process of model development and evaluation is often not linear, and only a minority of prediction models have been evaluated in clinical practice.

BENEFITS OF PREDICTION MODELS A key benefit to the design of prediction models is that a model's performance characteristics can be compared, such as for simple vs. complex models (such as clinical characteristics alone vs clinical *and* genetic predictors). This ability to compare is particularly relevant when expensive or invasive laboratory tests are proposed, as their added benefit can now be quantified. A prediction model can be designed to consider evidence-based multivariate effects and even interactions, given observed data. A prediction model may allow non-experts, such as pediatricians evaluating a newly diagnosed child in an emergency department, to make similar good or bad predictions as experts. The limitations of prediction models are that they often do not capture all available data, and new information may become available after a model has been adopted. For example, Wells' decision rule for pulmonary embolism does not include estrogen use or pregnancy. A further issue is the generalizability ("could my patient have been included in this study?"). Prediction models may need to be adapted to the local context of patients (e.g., ethnic background, environmental factors). The necessary simplification of a model due to constraints

in the available sample size may imply that the model insufficiently explains complex predictor relationships (e.g., interactions). However, such relationships are thought to be rare in practice [181].

BENEFIT OF MODELS FOR THE INDIVIDUAL PATIENT To understand how prediction models benefit the individual patient, it is essential to formalize the clinical decision process. For an individual patient, a binary prediction is either correct or false, irrespective of whether a model is used or not. When no individual data are used for decision support, the best guess for the expected outcome is the baseline risk of the group, e.g., 75% chance for transient ITP (pre-test probability). A prediction model predicts the expectation of an outcome after taking into account individual patient characteristics and gives an individual post-test probability (e.g., for two separate individuals, 35% and 95% probability for transient ITP). The rate of correct predictions and errors can now be evaluated in a group of patients, measuring discrimination and calibration. If it is a good model, it can be applied to future patients to make better predictions than without it, and this is how the individual patient benefits. How helpful a prediction model is on patient-centered outcomes and impact on clinical care can only be quantified in a separate study.

12.1.1 *Individualized prediction of ITP spontaneous recovery and IVIg treatment responses*

This dissertation aimed to distinguish the prognosis of newly diagnosed childhood ITP by clinical and molecular factors. Two new tools (multivariate prediction models) are proposed that integrate

multiple predictors and allow the determination of an individual patient's risk, the Childhood ITP Recovery Score (chapter 9) and a biological model to stratify ITP disease courses (chapter 10). Both use information available at the diagnosis or obtained within a reasonable timeframe and with limited costs. These features are of significant benefit for timely clinical adaptation.

Different model strategies are available to predict outcomes from data [164]. The model strategies essentially provide varying tradeoffs between model flexibility and interpretability, ranging from approaches like logistic regression (less flexible, very interpretable) to random forests or artificial neural networks (very flexible, hard to interpret). The two models we present are intentionally chosen to be interpretable and straightforward, such that clinicians do not have to believe in a magical 'black-box'. Our models can be readily interpreted and can theoretically be calculated by hand. An added benefit is that simple models may generalize better than complex ones.

CLINICAL PREDICTION SCORE In the Childhood ITP Recovery Score, developed in 377 and externally validated in 200 patients with newly diagnosed ITP, we show that clinical characteristics alone can discriminate transient ITP with an AUC ROC of 0.71 (95% CI, 0.62 – 0.80; external validation; chapter 9) . Assessment of calibration showed good agreement between predicted and observed recovery probabilities. The discrimination may be considered low but is in line with what can be expected from a clinical prediction model. Perhaps most importantly, the actual predictive value of clinical characteristics to differentiate disease courses are now known. The large effect sizes

of clinical predictors identified in previous studies [141] were much more limited in our model, which may hint towards some patient selection, overfitting, and absence of multivariate effects in previous studies (although the effect of shrinkage of the regression coefficients in our case needs to be considered). Significantly, the Childhood ITP Recovery Score is developed and externally validated in two independent multicenter cohorts. It may be seen as a limitation that patients with platelet counts above $20x10^9$/L were omitted; however, this score is relevant for the patients with low platelet counts and active bleeding symptoms where there is a potential indication for treatment. From this point forward, a candidate prediction marker's additional value should be compared to the Childhood ITP Recovery Score [179].

BIOLOGICAL MODEL As the next step beyond the clinical prediction of disease courses, when extensive molecular and clinical data are available, what biological markers best predict disease outcomes? Although many prediction markers have been proposed for ITP disease courses (chapter 11)[141], this question has never been assessed before. One strategy is to develop prediction models based on candidate markers identified by experts and the literature [157]. Alternatively, completely unbiased analyses can be performed to discover predictors (e.g., whole-genome sequencing; plasma proteomics). This dissertation chose a balance between these two approaches, with a targeted albeit still broad assessment of predictors from genetic and immune domains. In chapter 10, to build a biological model, we used the concept of systems medicine, where detailed biological data is linked with clinical characteristics and outcomes [182]. We used

supervised statistical learning to identify a profile that best predicts disease outcomes. To deal with missing data that are inevitable in clinical studies, we developed an analysis strategy that integrated supervised statistical learning with a multiple imputation strategy by chained equations [183]. We show that the biological model detects the *absence of a sustained response* (ASR) to IVIg with a sensitivity of 0.91 (95% CI, 0.80 – 1.00; chapter 10). With high sensitivity, patients who "test negative" for ASR (thus, would have a complete sustained response) can be expected to respond to IVIg. The biological model alone has a ROC AUC of 0.84 (fitted predictions; 95% CI, 0.75 – 0.93) and ROC AUC of 0.70 in cross-validation. We also show that the biological model is associated with bleeding symptoms during the full one-year follow-up, which is generally more important than the surrogate marker of platelet counts. For the individual patient, this means that, although spontaneous recovery within three months is likely, administration of IVIg would improve platelet counts and diminish the bleeding risk immediately (if applied).

GENERAL LIMITATIONS The discriminative ability of the scores may be seen as low, yet we interpreted the models as applicable with *room for improvement*. Importantly, with the Childhood ITP Recovery Score, most prediction errors with the clinical prediction model are observed in patients with low and intermediate probability of spontaneous recovery (chapter 9). We, therefore, set out to improve prediction in a biological model. In this biological model, we show that molecular predictors improve the discrimination in particular among these patients (chapter 10). Thus, our work provides the

current best model to predict disease courses going beyond clinical characteristics alone.

A separate biological model (or interaction term) for patients above seven years would be wishful, but limited data were available for patients in this age group. An approach that would have fit within our overall strategy is the fitting of interaction effects by hierarchical group-lasso regularization [184], to include, e.g., differences by patient age. To yield a valid prediction model, a study would likely require international collaboration to achieve sufficient inclusions and is challenging because of logistical issues arising from biological sample collection and preparation at different institutions. External validation of the biological model is desirable. However, the evaluation of the prediction markers in patients not included in the model development (randomized to observation) already provides independent validation and indication of the model's usefulness.

Notably, by our study design, heterogeneous treatment effects are not formally demonstrated [185, 186]. With any treatment episode, a patient's response may be affected by random and non-random variation. A 70% response rate could reflect that a treatment works for 100% of patients 70% of the time, or 70% of the patients 100% of the time [185]. Formal proof of heterogeneous treatment effects would require observing variation in individual responses to IVIg treatment in a repeated cross-over study [187]. Alternatively, we may allow the demonstration of an interaction effect between treatment and clinical or molecular characteristics. At this point, to believe in individual responses to ITP therapy, we would have to assume consistent non-random treatment responses [185] - which may be quite likely *a priori*, given the broad ITP clinical disease definition.

12.2 INSIGHTS INTO ITP PATHOPHYSIOLOGY

Childhood ITP is thought of as an acquired autoimmune disease, with humoral and cellular immune reaction against platelet self-antigens [18, 188]. The best predictors of prognosis are key causal factors in the ITP pathogenesis. Thus, we assessed immune characteristics that may be implicated in the pathogenesis of ITP.

12.2.1 *Genetic heterogeneity*

The IgG-Fcγ receptors (FcγR) allow immune cells to sense and attack antibody-opsonized cells, and genetic variants in the FcγR may thus be related to immune (dys)function. Here we show that the previously identified *FCGR2C*-ORF variant, associated with predisposition to Kawasaki disease[189], is also associated with susceptibility to childhood ITP (chapter 3). This variant can also be found in chronic ITP. Still, the transient form of childhood ITP is firmly associated with enrichment of this variant, with an odds ratio for transient vs chronic ITP of 4.7 (95% confidence interval, 1.9-14.3), arguing for a pro-inflammatory mechanism predisposing to transient ITP. Importantly, this finding suggests the potential use of genetic testing early in the disease to determine expected disease courses and possible responses to therapy. Since the variant is in linkage disequilibrium with the 2B.4 promoter polymorphism and the *FCGR2A*-27Q/W variant, we could not conclusively rule out the effects of these variants, although biologically, this is unlikely. The *FCGR2B*-p.232I allele is associated with early spontaneous recovery from ITP [90] but has no significant impact on disease courses beyond one week (chapter 3).

The *FCGR2B*-p.232T allele that leads to expression of a non-functional FcγRIIb receptor [190] might preclude homozygous p.232T/T patients from responding to IVIg. Still, only minimal data are available given the rarity of this homozygous non-functiuonal variant [90]. A recent systematic review showed the *FCGR2A*-p.H131R variant to be associated with susceptibility to childhood ITP [64], but we could not confirm this in our study population. Ethnic variation needs to be taken into account when comparing different study populations [189]. Overall, these data identify *FCGR2C*-ORF as a promising target to determine ITP disease courses early.

12.2.2 *Platelet autoantibodies*

Our data show that circulating platelet glycoprotein-specific antibodies are present in 62% and 10% of patients for IgM and IgG (chapter 5), respectively. Although we show that such antibodies can also be observed in healthy individuals (anti-GPIIb/IIIa IgM, 10/65; IgG, 3/65) or other patient groups, the likelihood that a positive result indicates an ITP patient is high. No difference in TPO levels between antibody-positive and -negative patients were observed. Patients enrolled in the TIKI study were already diagnosed with ITP, and alternative diagnoses were already discarded. Thus, the comparison of TIKI patients to other patient groups or healthy individuals may overestimate the diagnostic accuracy by comparing extremes. A key question is if the findings of our study will be consistent in a yet undiagnosed population with thrombocytopenia suspected of ITP (diagnostic uncertainty).

intention to diagnose

Mechanistically, the (temporary) loss of tolerance to platelet antigens may be due to epitope spreading[30]. We observed that platelet-reactive antibodies against multiple platelet antigens typically occurred together, which may support this hypothesis. Our data are supported by the findings of our systematic review that exclusively focused on children with ITP (chapter 4). More diagnostic yield can likely be achieved by even more sensitive detection techniques like surface-plasmon resonance with coated FcγR or flow-cytometry-based FcγR-tetramer testing (Szittner et al.; manuscript in preparation). This may allow the direct testing on patient platelets with the limited platelet numbers that can be obtained from children.

12.2.3 *Cellular immune response*

The self-limiting character of childhood ITP, often linked to a preceding infectious episode, inspires the long-standing thought that a state of immune dysbalance may contribute to or cause the disease. Children's immune system is developing, and significant changes in the lymphocyte subsets can be observed during the first ten years of age [191]. Compared to healthy children, critical differences in the immune subsets of children with ITP have been proposed [18, 51, 52, 188, 192]; in addition, some studies in children suggest that specific lymphocyte subsets are related to the ITP prognosis [27, 52, 192]. However, none of these markers have shown usefulness in daily clinical practice (yet). The prerequisites for clinical implementation of a marker include the adjustment for physiological age differences, assessment in a homogenous patient population (at diagnosis, without

treatments), and a complete analysis of related immune markers (e.g., all relevant lymphocyte subsets instead of specific subsets only).

To our knowledge, we performed the largest analysis of immune subsets and their relevance for ITP patient prognosis to date, using centralized analysis in a diagnostic reference laboratory. First, we observed that the overall leukocyte and lymphocyte counts and CD4, CD8, CD19, and NK cell counts were all within the age-appropriate expected reference ranges (chapter 2), indicating no difference to healthy children. This is a major difference to previously reported data [18, 51, 52, 188, 192].

By analyzing the association of immune markers with the observed disease course, adjusted for age, we observed that increased CD8 levels were associated with reduced recovery rates (chapter 6). In addition, the increased frequency of CD4 effector T cells was associated with reduced recovery rates. Using single-cell RNA sequencing in 8000 immune cells across six age-matched individuals by 10X Genomics, paired with single cell T and B cell receptor V(D)J sequencing, we observed that these activated effector CD4 cells were present at a higher frequency in chronic ITP cases. These cells and showed a transcriptomic signature with the co-expression of *IL7R*, *ITGB1* (CD29), desmoyokin (*AHNAK*), annexin A2 (*ANXA2*), *KLF6*, CD161 (*KLRB1*), and interleukin 32 (*IL32*). No other T, B, NK or monocyte cluster frequency identified by single-cell RNA sequencing was differentially changed between transient and chronic ITP. Both CD8 levels and CD4 effector cell frequencies remained significant in a combined multivariate analysis. They were not associated with a (self-reported) preceding infection, leukocyte or total lymphocyte counts, or presence of anti-platelet autoantibodies. Thus, we

T regulatory cells

conclude that T lymphocyte subsets may have a role in predicting spontaneous recovery and IVIg treatment responses. In contrast to the suggestions of previous studies, an unexpected finding was that CD_4 T regulatory cells were not decreased compared to expected values from healthy controls (chapter 6). Moreover, our data showed that Treg frequencies were not associated with recovery from ITP.

After interpreting these data, we believe that adherence to previously published immune predictors of childhood ITP disease courses, like T regulatory cells, which is often considered dogma in the field, should be interpreted cautiously.

12.2.4 *Age heterogeneity*

It has been long recognized that age is a critical factor in the prognosis of ITP. Adults have more often chronic disease courses than children [6], and there are also differences within childhood ITP (e.g., a child of 3 vs 15 years). We now defined these differences in much more detail using data of 577 patients from the NOPHO and TIKI study (chapter 2) and show that between one and 16 years of age, the clinical characteristics and the prognosis of ITP gradually change. As opposed to the previously recognized threshold at eight or ten years of age, where a favorable disease course was believed to be less likely, we observed that a shift towards more persistent and chronic ITP occurs already before five years of age. Surprisingly, responses to IVIg (with sustained platelet count $\geq 100x10^9$/L) above ten years of age were largely absent in our cohort, although only a few cases were available. These data suggest that there is a severe potential for heterogeneous treatment effects in this age group.

12.2.5 *Further thoughts on mechanisms*

The type of glycosylation of IgG-Fc parts in serum antibodies is associated with multiple autoimmune diseases, particularly abundant Fc-galactosylation and sialylation [193]. Genetic variants in the genes regulating IgG glycosylation are also associated with autoimmunity [194]. Therefore, a change in IgG-Fc glycosylation may contribute to antibody-mediated autoimmune pathology by changing the threshold of antigen-antibody immune complexes binding to FcγR [193]. Crucially, we observed no difference in the IgG-Fc glycosylation between 108 refractory ITP patients and 120 healthy control (chapter 7). Overall, these data suggest that the immune pathophysiology of ITP may be different from other autoimmune diseases.

THROMBOPOIETIN (TPO) is the main hormone driving platelet production. Paradoxically patients with ITP show normal or only slightly elevated TPO levels, despite thrombocytopenia. One explanation of the 'relative' TPO deficiency is that in ITP, there is a sufficient mass of megakaryocytes and platelets to bind and adsorb TPO via their TPO-binding receptor c-MPL. A recently discovered mechanism of skewed TPO production due to anti-GP Ib antibodies (in mice) has suggested that TPO production may be impaired in ITP [195]. This may potentially explain why ITP patients do not have increased TPO levels despite a shortage of platelets. In an extensive study of patients with ITP at our institute, where paired antigen-specific antibody data and TPO levels were available, we found this effect is not present in human ITP (chapter 8). Apart from these data, we discovered that TPO levels were increased, particularly among young

children aged below five years (chapter 2). Future studies should investigate if the increased TPO in these patients reflected a different ITP pathophysiology.

12.3 CLINICAL IMPLICATIONS

12.3.1 *Diagnosis of ITP*

The purpose of a medical diagnosis is to guide the clinical management and inform prognosis [196]. A diagnostic error may lead to inappropriate management (i.e., immunomodulatory treatment in a patient with a hereditary thrombocytopathy mistakenly diagnosed as ITP) or misaligned expectations (i.e., fear of long-standing bleeding symptoms, systemic autoimmunity, or leukemia, when the actual disorder is self-limiting, transient ITP). The clinical diagnosis of ITP entails that multiple diseases with a typical clinical presentation may be combined under one entity, as was previously suggested by the name idiopathic in the definition of ITP. The clinical diagnosis may question the current focus on a common, shared immune pathogenesis, also because 30% of patients do not respond to immune treatment with IVIg [90]. The "low sensitivity" of platelet antibody tests in clinical populations may not be a shortcoming or limitation but be a feature. When multiple pathogeneses are combined under the umbrella of one clinical diagnosis, it is expected that only a subgroup would test positive for a laboratory assay (chapter 4, chapter 5). Our data support such a paradigm. Similar ideas have been advanced in rheumatologic diseases; here, composite clinical and laboratory criteria are used for the diagnosis [197]. In Alzheimer's disease, a clinical diagnosis is combined with imaging findings to distinguish vascular and non-Alzheimer neurodegeneration [198]. This allows stratifying clinical trials and the accurate definition of treatment effects [198].

SYSTEMS MEDICINE Overall, a good understanding of the underlying pathophysiologies of ITP can only come from a systematic, deep analysis of phenotypes and their relationship with disease outcomes (systems medicine)[158, 182]. The use of composite diagnostic criteria for ITP could allow a better definition of prognosis already at the time of diagnosis [199]. A potential pathway for clinical adaptation is to diagnose separate entities for transient ITP and a further "unclear" category, e.g., "non-transient ITP with unclear prognosis". This dissertation shows that the positive inclusion of transient ITP could be possible by a combination of clinical features (Childhood ITP Recovery Score, chapter 9), potentially supported with selected molecular markers (e.g., *FCGR2C*-ORF; chapter 3, chapter 10). A high probability of transient ITP can be given by the Childhood ITP Recovery Score alone for 64% of patients (external validation in TIKI trial). There is a very high chance of recovery and a much-improved response to IVIg. Although we don't have any data, we postulate that other treatments also show differential effectiveness between these disease categories.

12.3.2 *Prognosis versus diagnosis*

Prognosis is central

The purpose of a medical diagnosis is to guide the clinical management and inform prognosis [196]. "Prognosis is central to medicine. All diagnostic and therapeutic actions aim to improve prognosis." [157] Importantly, the traditional dichotomization of diagnosis as present and absent may not always be of benefit to patients [173, 200], a concept that has gained ground in, e.g., type 1 Von Willebrand disease [201]. In contrast to the traditional way of diagnosis, disease

detection should aim to identify patients that may experience benefit from clinical intervention [180, 202], using information about their prognosis [200, 202]. In the case of ITP, we focus on patients with severe ITP (platelet count $\leq 20x10^9$/L), and for this group, this may mean detecting patients that can benefit from treatment, close monitoring, or additional diagnostic tests, or otherwise from 'hands-off' observation. Thus, focus on informed and evidence-based prognosis is a preferred clinical decision-making strategy [173], but in ITP, this was only possible to a rudimentary degree.

Now, in this dissertation, we stepped beyond the concept of "diagnosis" for childhood ITP by effectively using longitudinal follow-up data. We aimed to identify better which patients may require treatment, may benefit from such treatment, and which patients could be monitored "hands-off" in a relatively safe way (prediction of prognosis; chapter 9, chapter 10, chapter 11). Stratified decision making based on a child's prognosis, using data available at the diagnosis or during follow-up (e.g., age, platelet count, longitudinal platelet trajectory), is likely more suitable to yield clinical benefit than the diagnostic term "ITP".

12.3.3 *Additional diagnostic tests*

A remaining open question is if and when additional diagnostic tests should be performed to distinguish different underlying etiologies (case-mix) of the *thrombocytopenia syndrome* diagnosed as ITP. Over time, as children with transient ITP eventually recover (30%, 60%, and 90%, respectively one, six, and twelve months after the diagnosis), the remaining patients with possibly different underlying

disorders are enriched in the group that is left over. Screening for thrombocytopathies or autoimmunity could be performed at such timepoints (genetic analysis, autoimmune markers, immunoglobulin measurements, direct antiglobulin test). Still, at present, it is not clear what the best timepoint would be.

Stratifying diagnostic testing

The tools we propose in this dissertation allow enrichment of patients with persistent and chronic ITP disease courses already earlier, at the diagnosis. Identifying such patients may enable clinicians to target additional diagnostic tests early in the disease and prevent unnecessary treatments and side effects. To illustrate this, data presented at ASH 2019 annual meeting[150] indicated that 10% of patients with chronic ITP show IgG levels below the reference range, potentially indicating an immune disorder, as compared to an expected 2.5% in a healthy population (empirical 95% reference interval). The odds ratio for chronic ITP was 5.7, which seems high. However, if IgG levels were measured at the diagnosis in 200 newly diagnosed ITP patients, because chronic ITP occurs in only 10% of patients, 2/20 chronic patients would be identified with low IgG (10%), versus 5 patients (empirical 2.5%) with a false-positive result. The predictive value for chronic ITP of a positive test would thus be 2/7, or 28%. If cases with a high probability for transient ITP were excluded by the Childhood ITP Recovery Score, leaving 73/200 patients (36.5%) and assuming all 20 chronic cases are in this group, there would be only 2 false-positive cases, and the positive predictive value would be 2/4, or 50%. Thus, *the positive predictive value is almost doubled*. Of course, this example also shows that a marker present in just 10% of chronic patients is likely not valuable at the diagnosis.

Multiple features are probably required to discriminate the chronic ITP patients in the residual group of 73 patients.

To further enrich patients with persistent or chronic ITP, a possible informative analysis of the TIKI trial is if clinical or molecular characteristics present at one- or three-months follow-up allow the prediction of long-standing disease (dynamic prediction).

12.3.4 *Treatment*

In the TIKI trial, as in previous studies, IVIg treatment effectively resolved the thrombocytopenia and prevented bleeding symptoms. A key question is if the benefit of treatment can be weighed against side effects and costs. Notably, although adverse events are carefully monitored in the setting of a clinical trial, most side-effects of IVIg treatment are of minor significance. In the TIKI trial, treatment did not improve HRQoL[9]. However, we need to consider if a relevant benefit may have been missed as the treatment only worked in 70% of patients at the one-week follow-up[90]. Irrespective of the goal of therapy (bleeding or modification of disease course), if it is possible to target patients that benefit from treatment (prognosis-guided treatment), benefits can be enriched. At the same time, costs and side effects for a larger group can be minimized. Other treatments than IVIg may be suitable, such as corticosteroids or TPO-Ra, with reduced cost or a more beneficial route of administration, but such therapies require formal investigation.

12.3.5 *Questions of generalizability*

Our work focused on children severely affected by ITP, with platelet counts $\leq 20x10^9$/L at the diagnosis. These patients represent a group where potential treatment could be indicated. An open question is how our findings generalize to patients diagnosed with a platelet count above this threshold. We, unfortunately, had no data available to study this. Finally, it is yet unclear if the responses to IVIg would be concordant with the response to other treatments, such as corticosteroids or TPO receptor agonists.

12.3.6 *Follow-up studies*

For clinical development of the predictive models, separate studies are required. For the Childhood ITP Recovery Score, external validation has already shown the validity of the prediction model. The model may be improved by including other clinical markers available at the diagnosis (model extension). We are not aware of currently known clinical characteristics that could be easily added. The model can now be assessed in a clinical impact study, where the guidance of clinical management by the model can be compared against standard-of-care, for outcomes like health care utilization, need for treatment administrations, and quality of life.

For the biological model, external validation of the score would be wishful. Such a study should reserve additional material (DNA, PBMC, plasma) to discover additional markers by satellite studies. This could include evaluation for hereditary thrombocytopenia [87] and immune disorders [203]. At the same time, we are well aware of

the challenges of conducting a large study in childhood ITP. Such a study likely requires international collaboration to include sufficient patients in a feasible timeframe and presents significant logistical issues.

An interesting pathophysiological study would be to compare deep immune phenotyping of T and B cells (e.g., by high-dimensional flow cytometry, or CyTOF) with age-matched healthy controls and children with a recent viral infection. Such a study would allow a definition of the actual immune changes in ITP. Ideally, such a study would include longitudinal follow-up samples. Unfortunately, we did not have access to such material.

All types of studies should consider including patients with a platelet count $20 - 100x10^9$/L in a separate parallel registry and follow these patients as well.

12.4 A ROADMAP TO ADVANCE ITP RESEARCH & CARE

12.4.1 *Dogmalysis*

During this dissertation, it became increasingly clear that we need to be cautious in interpreting ITP datasets. Studies in childhood ITP are challenging. From an ethical point of view, all research should improve care for patients with ITP as the primary aim. Thus, to enhance our understanding of the disease, we require more robust and dependable data (solutions discussed below). We believe now is the time to question what we believe carefully, so we become aware of what can and cannot be supported by the evidence (*dogmalysis*).

The ideal roadmap involves basic researchers that collaborate closely with clinicians, targeting patient needs. Many mouse models of "ITP" have been developed, but they almost exclusively study allogeneic platelet antigen responses instead of autoimmunity. It is inevitable to question if results from such studies can genuinely be translated and if such models of "ITP" may well have clouded our view of the clinical syndrome. Moreover, there is a need for clinicians, on the other hand, to better collaborate and focus on performing quality studies that yield robust datasets. Translational studies that aim to unravel the etiology and pathophysiology of ITP need to consider the heterogeneity of ITP. The parable of the blind men with the elephant illustrates this beautifully [204]:

"A group of blind men heard that a strange animal, called an elephant, had been brought to the town, but none of them were aware of its shape and form. Out of curiosity, they said: We must

Figure 12.1: *Blind monks examining an elephant*
Hanabusa Itcho (1652-1724). Japanese painter.

inspect and know it by touch, of which we are capable. So, they sought it out, and when they found it they groped about it. The first person, whose hand landed on the trunk, said, This being is like a thick snake. For another one whose hand reached its ear, it seemed like a kind of fan. As for another person, whose hand was upon its leg, said, the elephant is a pillar like a tree-trunk. The blind man who placed his hand upon its side said the elephant is a wall. Another who felt its tail described it as a rope. The last felt its tusk, stating the elephant is that which is hard, smooth and like a spear." (Modified from Wikipedia)

LESSONS Of course, all of the men are right at the same time. However, some solutions come to mind. The men can work together (collaboration); they can investigate turn-by-turn the different parts of the elephant when they disclose the coordinates that they have examined (complete reporting of patient characteristics), using the same tools (sharing of methods). The men would ideally communicate their findings systematically to assemble a full picture (integrating data across studies, e.g., meta-analysis and meta-regression). We must consider that for ITP, it is even more complex with the case mix, and we do not only investigate an elephant but a frog and a fruit fly at the same time.

12.4.2 *Better clinical study design*

Better data can, in part, be obtained by better study designs and improved dependability of the data. The goal needs to be to gather quality data and sufficient characteristics to be integrated into meta-

analyses and meta-regression. This is particularly relevant for rare childhood disorders [205, 206]. Moreover, there is an overall need to derive quality data on drug development in pediatric disorders [207], and this has been recognized by the recent changes of the European Medicine Agency regulations. The benefit of candidate molecular predictors needs to be assessed using approaches such as cross-validation [164], pre-validation [208], or separate discovery and validation cohorts.

STRATEGIES TO IMPROVE STUDY DESIGNS FOR CLINICAL AND TRANSLATIONAL STUDIES IN ITP

- In a randomized controlled trial, follow patients that don't fit the inclusion criteria and are excluded in a separate parallel arm. For example, children with platelet count $20 - 100x10^9$/L. [209]
- Perform quality multicenter observational prospective studies in settings where it is hard to recruit sufficient patient observations. Key criteria for such studies are stringent study design, patient and outcome definitions. [206, 210]
- Retrospective cohort studies are not inherently wrong, as long as their limitations are recognized and managed. Critically, studies need to report crucial data for integrated analysis (suggestions below)
- Perform MASTER (observational) trials that allow integration of molecular data with disease and treatment outcomes [211]

12.4.3 *Focus on patient-centered outcomes*

As clinicians, we are aware of the catastrophic effects of end-organ bleeding and, in particular intracranial hemorrhage [212]. We naturally care about bleeding, and patients agree with this. Not insignificantly, there is also a stigmatizing effect of bruises in young children. Fatigue is a further important issue for children with ITP [213]. Instead of focusing on (surrogate) platelet or bleeding outcomes, clinical trials should consider improving these patient-centered outcomes by measuring the quality of life. A further important step is the involvement of patients, parents, patient representative groups, and advocates in the trial design.

12.4.4 *Towards a reporting standard*

We recognize as a major challenge for research in pediatric ITP that data collection in the context of a rare disease and a pediatric study population is notoriously difficult. Even small cohorts and case series have a definite benefit in that light. Still, they need to feature sufficient information to make it possible to interpret and integrate results across studies. In our systematic review, we observed a significant lack of reporting of important clinical and metadata of clinical study populations among 40 studies in childhood ITP (chapter 4). It is paramount for a better evidence base in ITP that reliable data are collected and published. Therefore, to improve the clinical interpretation and quality of data in a field, we suggest that future studies in childhood ITP should disclose the following vital data that are readily available from clinical reports.

PROPOSED REPORTING STANDARD FOR CLINICAL AND TRANSLATIONAL ITP STUDIES

- Clinical characteristics, including age, sex, symptom duration, and bleeding symptoms
- Time from diagnosis, in days
- Current and past treatments
- Follow-up time and number of children lost to follow-up