Aus der Kinderklinik und Kinderpoliklinik im Dr. von Haunerschen

Kinderspital

Abteilung Pneumologie/Allergologie

Direktor: Prof. Dr. Christoph Klein

Novel statistical approaches for censored immunological data: analysis of cytokine and gene expression data

Dissertation
zum Erwerb des Doktorgrades der Humanbiologie
an der Medizinischen Fakultät der
Ludwig-Maximilians-Universität zu München

Vorgelegt von
Nikolaus Ballenberger

aus München

2012

Mit Genehmigung der Medizinischen Fakultät der Universität München

Berichterstatter: Prof. Dr. med. Erika von Mutius

Mitberichterstatter: Prof. Dr. Hans-Helge Müller
Prof. Dr. Jörg Christian Prinz
Prof. Dr. Matthias Bauer
Prof. Dr. Thorsten Hothorn

Mitbetreuung durch den promovierten Mitarbeiter: Dr. Sabina Illi

Dekan: Prof. Dr. med. Dr. h.c. M.Reiser,FACR,FRCR

Tag der mündlichen Prüfung: 12.11.2012

Content

Figures

Tables

1   Introduction

This chapter gives an overview on environmental determinants of allergic diseases and describes two birth control studies that are aimed to elucidate contributing neonatal immunological mechanisms on atopy in childhood. Furthermore, the complexity of immunological data such as cytokine and gene expression data is illustrated and finally, the aims of the present thesis are formulated.

1.1 Environmental determinants of allergic diseases

1.1.1 General aspects of allergic diseases

Allergy and atopy are two terms that are often used misleadingly as synonyms. Originally, the term 'allergy' was introduced in 1906 by von Pirquet and Bela Schick (from the Greek allos meaning "other" and ergon meaning "reaction") describing a deviation from the original state or normal behaviour of the individual (Turk 1987). With the passage of time, the meaning of the word allergy has changed and is defined nowadays more specifically as "a hypersensitivity reaction initiated by immunological mechanisms". Allergy can be antibody-or cell-mediated. In the majority of cases the antibody typically responsible for an allergic reaction belongs to the IgE isotype and these individuals may be referred to as suffering from an IgE-mediated allergy or IgE-mediated allergic disease" (Johansson, Bieber et al. 2004). In contrast atopy is defined as "a personal and/or familial tendency, usually in childhood or adolescence, to become sensitized and produce IgE antibodies in response to ordinary exposure to allergens, usually proteins" (Johansson, Bieber et al. 2004). As a consequence, such individuals can develop typical symptoms of asthma, rhinoconjunctivitis and/or dermatitis (eczema). However, the association between atopy and manifest clinical illness is loose. Many individuals having IgE antibodies to inhalants or food allergens are not ill and do not have a clinical response upon exposure to disease allergens.

Even though these symptoms are usually not life threatening the burden of allergic diseases is not trivial. Allergic diseases are associated with reduced quality of life and a high financial burden. The costs related to asthma in different western countries ranges from US $300 to US $1300 per patient per year (Accordini 2008).

Taking into account the high disease prevalence and chroniticity of the disease the costs for society are tremendous.

Disease occurrence of allergic diseases is mainly reported as prevalence. There are only few studies investigating incidence of allergic diseases. This is due to the fact that the typical study design applied for measuring disease occurrences of allergic diseases is cross-sectional and therefore, the only valid measure is prevalence. Thus, merely the prevalence is reported in this text. There has been consistent worldwide increase of allergic diseases since the 1960´s especially in western countries until the late 90´s. (von Mutius 2006). The countries with the highest prevalences are England, Scotland and Australia. E.g. among seven year old Australian children a lifetime prevalence of 46% of ever having Asthma or wheeze was assessed.

1.1.2 West-East gradient of allergic diseases

Interestingly, a strong east-west gradient with respect to prevalence of allergic diseases was observed. The worldwide ISAAC study, which was undertaken to identify factors that may explain the rise in atopic diseases, comprised over a half a million children aged 6–7 and 13–14 years from 155 centers in 56 countries (Asher, Anderson et al. 1998; Beasley 1998). By using standardized written and videotaped questionnaires, the highest prevalence of asthma symptoms was mainly found in English speaking countries, and lowest in Eastern Europe, Russia, China, India and Ethiopia. For allergic rhinoconjunctivitis and eczema, the areas of the lowest prevalence were similar to those for asthma symptoms.

Another example of significant evidence of the East–West gradient in the occurrence of atopic diseases comes from studies performed in Germany after the reunification in 1989. Von Mutius et al. (von Mutius, Martinez et al. 1994) showed that in ethnically similar populations, the prevalence of atopic diseases (current asthma, bronchial hyperresponsiveness, atopic sensitization assessed by skin prick tests) were significantly higher among school children ($n = 7445$) living in the former West as compared to the children ($n = 4534$) in former East Germany. These results are in line with those by Nicolai et al. They showed in a study with 5313 adults from the western and 2617 adults from the eastern part of the country that a similar west–east gradient in the prevalence of atopy, which was defined by specific IgE levels, was present among adults aged younger than 40 years (Nicolai, Bellach et al. 1997).

Consequently, the gradient between western and eastern countries has been confirmed in several studies among children and adults performed in Sweden and Poland (BrÅBÄCk, Breborowicz et al. 1994), Sweden and Estonia (JõTgi, Janson et al. 1996), Finland and Russia (Erkki, Tuula et al. 2002), in the Baltic area (BrÃbÃck, Breborowicz et al. 1995) and in a comparison between Sweden, the Baltic countries and Uzbekistan (Bjorksten, Dumitrascu et al. 1998).

Although allergic diseases have a strong hereditary component the West-East gradient is thought to be explained mainly by environmental factors (Von Hertzen and Haahtela 2004) because populations have the same ethnic background. Higher number of siblings, early attendance to day care, exposure to bacterial and viral pathogens, exposure to endotoxin, commensal bacteria and helminth infections and exposure to rural lifestyle such as farming environment are inversely associated with the development of allergic diseases (von Mutius 2000; Karmaus and Botezan 2002; Von Hertzen and Haahtela 2004). All these factors are reflected by a more rural and traditional, and thus less westernized life style. These findings gave rise to the so called hygiene hypothesis. The hygiene hypothesis is based on findings published in an article by Strachan (Strachan 1989) in which he observed that hay fever and eczema were less common in children from larger families with a higher number of siblings leading to more exposure to more infectious agents compared to children from families with only one child. The hygiene hypothesis states that there is a positive association between the lack of early life exposure to infections, symbiotic microorganisms, parasites and the development of allergic diseases by reducing natural development of the immune system. One possible explanation for the underlying biologic mechanism is that exposure to many bacteria and viruses elicits a T_h1-mediated immune response and down regulates T_h2-mediated immune responses, and thus leads to suppressing allergic diseases.

1.1.3 Farm effect on allergic diseases

A large body of studies has emphasized the effect of farming as a strong factor for protection against allergic diseases. Whereas results of studies comparing the prevalence of allergic diseases in urban and rural areas have been inconsistent (Strachan, Anderson et al. 1994) there are large differences in the prevalence of allergic diseases present within rural areas. Farm children from rural areas have a

significantly lower risk of developing these diseases than non-farm children who live in the same rural area. Table 1 that is adapted from a previously published article by von Mutius (von Mutius and Vercelli), shows a summary of studies investigating the farm effect on allergic diseases. Here, the present table was complemented, if available, by "Exposure", the odds ratios of the relevant outcomes and the underlying immunological mechanisms.

The next paragraphs are aimed to describe the most crucial findings of the "farming studies".

Table 1: summary of studies investigating the effect of farm exposure on allergic diseases
This table is adopted from a previously published article by von Mutius (von Mutius and Vercelli). It shows a summary of studies investigating the effect of different exposures of farm environment on allergic diseases and symptoms. The effects are presented in terms of odd´s ratios (OR). Or <1 is considered as "protective", OR>1 is considered as "risk factor".

Country	Age	Exposure	Asthma (OR)	Wheeze (OR)	Atopic dermatitis (OR)	Hay fever symptoms (OR)	Atopic sensitization (OR)	Immunological Finding	References
Germany, Switzerland, Austria, Sweden, Netherland	5 - 13	agriculture, pig farming, barn, shed, silage endotoxin	0.56-0.72 * 0.55*	0,63-0,73 * 0.89	-	-	0.35* 0.38*	expression of CD14 and TLR genes among farm kids is higher	(Ege, Remo et al. 2007) PARSIFAL
Switzerland	6 - 15	farming	1.17	0,77	0,86	0,89	0,31 *	-	(Braun-Fahrländer, Gassner et al. 1999) SCARPOL
Austria	8 - 11	farming	0,24*	0,62*	0,9	0,29*	0,47 *	-	(Riedler, Eder et al. 2000)
Austria, Germany, Switzerland	6 - 13	milk and stable at early life	0,14*	0,17*	-	0,20*	0,12*	-	(Riedler, Braun-Fahrländer et al. 2001)
Bavaria	5 -7	farmer	0,65	0,55*	1,04	0,52*	-	-	(Ehrenstein, Mutius et al. 2000)
New Zealand	5 - 17	farming, animals, milk in utero	0.50*	0.48*	0.46*	0.47*	-	-	(Douwes, Cheng et al. 2008)
Germany, Switzerland, Austria, Sweden, Netherland	5 - 13	farm milk consumption ever, stable in pregnancy	0.76* 0.86	0.77 0.76	-	0.77 0.77	0.76 0.58*	CD 14 , TLR 2 and TLR 4 is significantly higher among exposed	(Ege, Bieli et al. 2006) PARSIFAL
Bavaria	0	farming during pregnancy milk, animals	-	-	-	-	-	higher Treg, foxp3 decreased IL5 increased IL6	(Schaub, Liu et al. 2009) PAULCHEN
Austria, Germany, France, Finland, Switzerland	0	dust samples, farming	-	-	-	-	-	IFN-y and TNF-a higher in E IL5,6 and 11 no effect of dust on cytokines	Pasture (Pfefferle, Gisela et al.)
Austria, Germany, France, Finland, Switzerland	0	toxoplasmose and rubella during pregnancy						inverse associations of cord blood IgE to seasonal allergens with positive maternal records for Toxoplasma gondii and rubella virus	(Ege, Herzum et al. 2008) pasture
England	?	farming, milk	0,67* 0.90	-	0.50 * 0.59*	0.91 0.61*	0.68 0.24*	milk drinkers have lower IGE (total) and higher IFN-y	(Perkin and Strachan 2006)
New Zealand	7- 10	farming, milk, animals	1.00 0.7	0.8 0.6 *	1.1 0.6*	1.8 1.1	1 0.6	-	(Wickens, Lane et al. 2002)
Germany, Switzerland, Austria, Sweden, Netherland	5 - 13	farm milk	0.74*	0.86	0.89	0.56*	0.67*	-	(Waser, Michels et al. 2007)
Austria, Germany, Switzerland	6-14	In mattresses Muramic acid	1.17	0.67*	-	0.87	0.9	-	(van Strien, Engel et al. 2004) ALEX
New Zealand	25-49	farming	0.58*	0.56*	0.78*	0.83*	-	-	(Douwes, Travier et al. 2007)
Austria, Germany, Switzerland	9-10	farming	-	-	-	-	-	TLR2 and CD14 higher in farming kids	(Lauener, Birchler et al. 2002) ALEX

13

Germany, Switzerland, Austria, Sweden, Netherland		stable during pregnancy	0.86	0.76	-	0.74	0.58*	TLR2, TLR4 und CD14 higher, dose response with animal species	(Ege, Bieli et al. 2006) PARSIFAL
Gotland	7-8	farming	0.26*	-	0,39*	0,84	0.93	-	(Klintberg, Berglund et al. 2001)
Austria, Germany, France, Finland,Switzerland	0	farming	-	-	-	-	-	IGE higher in non-farming kids, inverse correlation with IFN-y and TNF-a → higher in farm kids	(Pfefferle, Serdar et al. 2008)
Mongolia	10 - 60	villages, rural towns, capital	1 2.19 1.9	-	-	1 1.9* 2.4*	1 2.0* 1.4*	-	(Viinanen, Munhbayarlah et al. 2005)
Mongolia	10 - 60	town residents, from town to village residents, village residents	1 1.05 0.66	-	-	1 0.68* 0.43*	1 0.62* 0.26*	-	(Viinanen, Munhbayarlah et al. 2007)
Finland	18 - 24	farming	0.81	0.71*	-	0.63*		-	(Kilpeläinen, Terho et al. 2000)
Germany, Switzerland, Austria, Sweden, Netherland	5-13	farming and Anthroposophic lifestyle vs. reference children	0.74* 0.85	1.10 0.78*	0.88 0.83*	0.72* 0.39*	0.73* 0.53*	-	(Alfvén, Braun-Fahrländer et al. 2006) PARSIFAL
Gotland	7-8	farming	0.26*	-	0,39*	0,84	0.93	-	(Klintberg, Berglund et al. 2001)
Austria	7.8	farming	-	-	-	-	0.29*	-	(Horak, Studnicka et al. 2002)
Netherland	45	farming	1.0	0.4*	-	-	-	-	(Smit, Zuurbier et al. 2007)
Germany	18-44	farming	-	-	0.54*	-	0.73*	-	(Schulze, van Strien et al. 2007)
Finland	40	farming women	-	-	-	-	0.18*	-	(Koskela, Happonen et al. 2005)
Germany	33	visits to animal building between 4 and 6 ys.	0.43*	-	-	0.36*	-	-	(Radon, Ehrenstein et al. 2004)
Finland	9	farming	-	-	-	-	0.56*	-	(Remes, livanainen et al. 2003)
Austria	6-10	farming	0.22*	-	-	-	-	-	(Kiechl-Kohlendorfer, Horak et al. 2007)
Sweden	16-75	farming	0.3	-	-	-	-	-	(Wennergren, Ekerljung et al.)
Sweden	16-75	farming	0.78*	-	-	-	-	-	(Eriksson, Ekerljung et al.)
Finland	0-6	farming	0.22*	-	-	-	-	-	(Kilpeläinen, Terho et al. 2002)
Australia	7-12	farming	0.65*	0.55*	0.64*	0.76	-	-	(Downs, Marks et al. 2001)
Canada	0-11	farming	0.22*	--	-	-	--	-	(Midodzi, Rowe et al. 2007)
British Columbia	8-20	farming	0.66*	0.85*	1.0	0.46*	-	-	(Dimich-Ward, Chow et al. 2006)
USA	20-88	farming women	0.55*	-	-	-	-	-	(Hoppin, Umbach et al. 2008)
Canada	12-19	farming	0.70*	0.59*	-	-	0.58*	-	(Ernst and Cormier 2000)
USA	4-17	farming	0.78*	0.78*	0.77*	-	-	-	(Adler, Tager et al. 2005)
USA	6-14	farming	0.88	0.87*	-	-	-	-	(Chrischilles, Ahrens et al. 2004)

1.1.4 Sources of independent protective farming exposures

In several studies three independent protective exposures to the farming environment for developing allergic diseases were identified: contact with livestock, mostly cattle, pigs and poultry, contact with animal feed such as hay, grain, straw and silage and the consumption of unprocessed cow's milk (Ehrenstein, Mutius et al. 2000; Riedler, Eder et al. 2000; Riedler, Braun-Fahrländer et al. 2001; Ege, Remo et al. 2007; Douwes, Cheng et al. 2008).

International studies demonstrated that exposure to livestock contributes to a large extent to the protective 'farm effect' (Riedler, Braun-Fahrländer et al. 2001; Ege, Remo et al. 2007). Even children not living on a farm but being exposed regularly to farm animals also had a lower prevalence of allergic diseases compared to non-exposed non-farm children. The effect reflects an inverse dose response relationship. With an increasing number of animals the protective effect deceases. Furthermore it was shown (Ege, Bieli et al. 2006; Schaub, Liu et al. 2009; Pfefferle, Gisela et al. 2010)that there is no protective effect of farming among children living in crop-farming regions suggesting that the farm effect is at least partly due to contact to animals and livestock (Wickens, Lane et al. 2002).

Studies have shown a protective effect of consumption of unprocessed cow´s milk on developing allergic diseases. (Riedler, Braun-Fahrländer et al. 2001; Perkin 2007; Waser, Michels et al. 2007; Schaub, Liu et al. 2009; Pfefferle, Gisela et al. 2010). However, it is crucial that the milk has not been pasteurized and homogenized yet. It is believed that with the ongoing processing of the milk the protective effect disappears (von Mutius and Vercelli 2010). Also here, analogue to livestock exposure, the protective effect from the consumption of raw milk was also seen among non-farming people consuming unpasteurized and unhomogenized cow's milk (Perkin and Strachan 2006).

1.1.5 Microbial exposure

Children living on farms, especially those with animal sheds, are more exposed to allergens, bacteria, viruses and fungi than children not living on farms. Endotoxin (a substance from Gram-negative bacteria), muramic acids (a component of peptidoglycan from the cell wall of all types of bacteria), extracellular polysaccharides (specific carbohydrates that are secreted or shed during growth of these fungi)

derived from fungi, such as Penicillium and Aspergillus spp and glucans are more abundant in mattress dust from farm children and farming households than to mattresses from non-farm children and non-farming households (Perkin 2007; von Mutius and Vercelli 2010). Children bring their microbial outdoor exposures into the indoor environment. Thus, mattress dust can be considered as a reservoir that represents a microbial exposure to indoor and outdoor environments (von Mutius and Vercelli 2010).

Exposure to endotoxin levels correlate inversely with the prevalence of hay fever, atopic asthma and atopic sensitization. However, high levels of endotoxin were found to be a risk factor for non-atopic wheeze (Vogel, Blümer et al. 2008). On the other hand high levels of muramic acid and extracellular polysaccharides in mattress dust were inversely correlated with wheezing and asthma among rural children (van Strien, Engel et al. 2004; Douwes, Travier et al. 2007; Ege, Remo et al. 2007). It seems that endotoxin may have beneficial effects on atopic diseases and at the same time be a risk factor for non-atopic asthma and wheeze, while muramic acid and extracellular polysaccharides have protective effects for developing wheeze and asthma.

1.1.6 Timing of exposure

Not only is the duration of the exposure to farming environment but also the timing of exposure likely to play a critical role. Studies in which the subject of the timing of exposures has also been addressed demonstrated that the protective effect of farming environments on atopic diseases was strongest when farm contact started early in childhood and was maintained until adulthood (Radon, Schulze et al. 2006; Douwes, Travier et al. 2007; Smit, Zuurbier et al. 2007). An even greater reduction in risk has been demonstrated for those children already exposed prenatally during the mother's pregnancy (Ege, Bieli et al. 2006). Moreover, it could be demonstrated that maternal exposure to animal sheds and unpasteurized cow's milk influences immunological mechanisms, such as the production of specific IgE antibodies, interferon-γ or regulatory T-cell activity in the cord blood of the neonate (Schaub, Liu et al. 2009; Pfefferle, Gisela et al. 2010). It can be concluded that important immunological modulations are already activated during pregnancy. However, many of the underlying mechanisms still are unclear.

1.1.7 Immunological mechanisms

TLRs belong to the Toll-like receptor (TLR) family which plays a fundamental role in pathogen recognition and activation of innate immunity. They recognize pathogen-associated molecular patterns (PAMPs) that are expressed on infectious agents and mediate the production of cytokines necessary for the development of effective immunity. CD14 acts as a co-receptor along with the Toll-like receptor TLR 4 for the detection of bacterial lipopolysaccharide. Peripheral blood leucocytes from farm children in the ALEX study were found to show increased expression of the genes for CD14, TLR2 and TLR4 compared to non-farm children (Lauener, Birchler et al. 2002). In the PARSIFAL study these results were not only confirmed with respect to microbial exposure during childhood but, furthermore, it was demonstrated that exposure of pregnant mothers to stables was associated with this enhanced PRR expression in leukocytes of the cord blood (Ege, Bieli et al. 2006). Even, a dose–response relationship was seen. Expression of TLR2, TLR4 and CD14 increased with the number of different farm animal species with which the mother had had contact during her pregnancy (Ege, Bieli et al. 2006). It can be concluded that microbial exposure during childhood and pregnancy of the mother affects the expression of genes encoding PRRs.

The immunoregulatory effects of farming are not restricted to innate immunity. Also the humoral immune system whose principal function is the production of immunoglobins (IG) is affected by exposure to farming environment. Immunoglobins (also known as antibodies) are gamma globulin proteins that are found in the blood or other bodily fluids of vertebrates, and are used by the immune system to identify and neutralize foreign objects, such as bacteria and viruses. The ALEX study explored the effect of farm exposure on class-switch recombination of IGs depending on allergens. IgE and IgG responses to inhalant allergens (grass, cat hair and house dust mites) were evaluated in school children. Interestingly, farm living did not affect the prevalence of IgG2 and IgG3 isotypes, as expected, but inhibited the development of IgG1, IgG4 and IgE antibodies (Stern, Riedler et al. 2007). These immunoglobulin isotypes are T helper 2 (T_H2)-dependent. The effect was seen for both grass and cats. In contrast, the prevalence of IgE specific for house dust mites was increased among farm children.

A striking finding was that the immunological mechanisms in school-age farm children were also present in newborn babies. The PASTURE birth cohort study (von

Mutius, Schmid et al. 2006) evaluated the effects of maternal farming exposures during pregnancy on IgE responses in the offspring. Seasonal allergen-specific IgE responses were significantly more prevalent in cord blood from newborn babies whose mothers had not been exposed to animal sheds and grass and were strongly inversely related with reduced production of the TH1 cell-associated cytokines. This increased production of IFNγ and TNF was seen when mothers were exposed to multiple animal species and barns during pregnancy and even when butter made from unprocessed milk was consumed by pregnant mothers. In contrast, the TH2 cell-associated cytokine interleukin-5 (IL-5), the regulatory cytokine IL-10 and the TH1-inducing cytokine IL-12 remained unaffected (Pfefferle, Serdar et al. 2008).

Another cell type that is influenced by farming exposure of pregnant mothers is represented by the regulatory T (TReg) cells. Cord blood CD4+CD25hi TReg cells from children born to mothers exposed to stables were both more abundant and more efficient in suppressing T cell proliferation. (Bianca). Additionally, maternal exposure to higher numbers of farm animal species increased the expression of the TReg cell marker glucocorticoid-induced TNF receptor (GITR) and the secretion of IFNγ by cord blood cells in response to allergen (Schaub, Liu et al. 2009).

In summary, it can be concluded that all these effects result in inhibiting a TH2-mediated allergic inflammation. Figure 1 shows a working model of the immunobiology of farm exposure described by von Mutius (von Mutius and Vercelli 2010).

Figure 1: a working model of the immunobiology of farm exposure
described by von Mutius (von Mutius and Vercelli 2010).
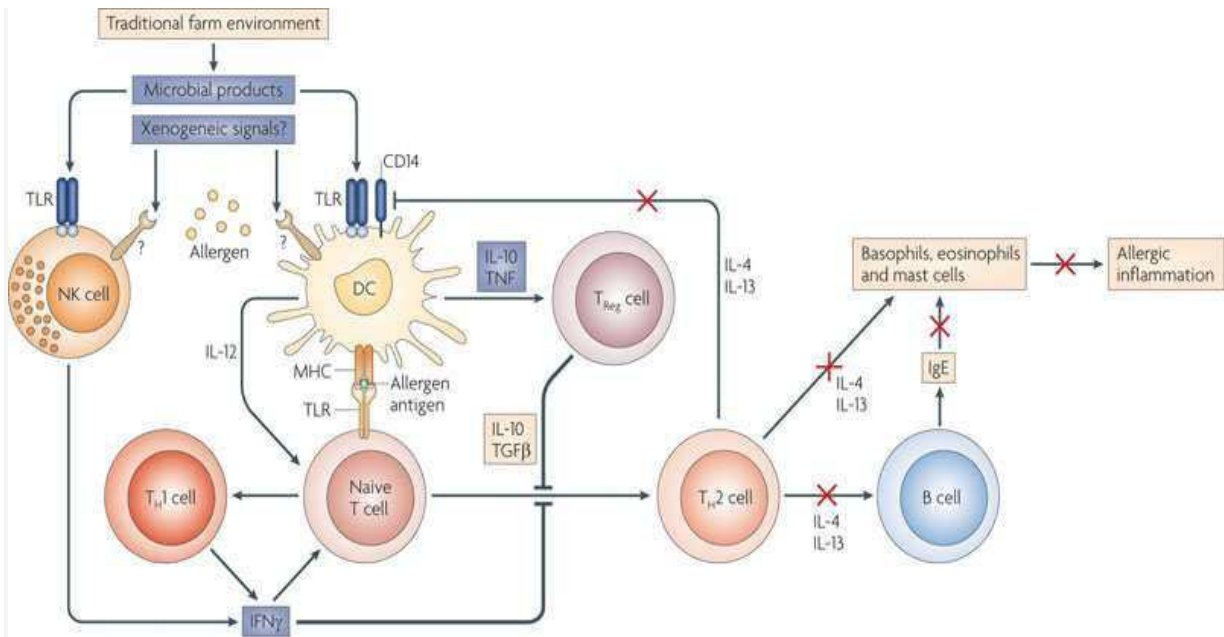
In this model, the biological diversity of a traditional farm environment (in particular, high numbers of farm animal species) results in intense microbial pressure on the innate immune system. This in turn directs vigorous tumour necrosis factor (TNF)- and interleukin-10 (IL-10)-promoted regulatory T (T_{Reg}) cell activation, which balances adaptive immune responses and suppresses key effector mechanisms of allergic inflammation (allergen-induced T helper 2 (T_H2) cell-associated cytokine production and T_H2 cell-dependent IgE synthesis). Moreover, decreased IL-4 and IL-13 production relieves T_H2 cell-associated cytokine-dependent inhibition of CD14 expression, which leads to further enhancement of pattern-recognition receptor expression and amplification of innate immune responsiveness. Upregulation of interferon-γ (IFNγ) in children of mothers exposed to multiple farm animal species depends primarily on enhanced innate immune activation that is induced by high microbial burden through dendritic cells (DCs) and Toll-like receptor (TLR)-expressing natural killer (NK) cells, but may also be related to the constant, robust xenogeneic pressure generated by close contact with multiple animal species. Xenogeneic signals (delivered through a currently undefined mechanism) may stimulate NK cells to secrete IFNγ, which counteracts allergen-induced T_H2 cell-associated cytokine production and accelerates maturation of T_H1-type responses by activating DC-derived IL-12 production. All of these effects synergize in preventing T_H2-mediated allergic inflammation. TGFβ, transforming growth factor-β.

The extreme biological diversity of a traditional farm environment with high numbers of animal species activates the innate immune system and the expansion of Treg cells mediated by the cytokines TNF and IL-10 (Chen, BÃ¤umel et al. 2007). Treg cells, balancing adaptive immune responses, decrease allergen-induced, TH2 cell-associated cytokine production and, therefore, IgE production. Moreover, a decrease in IL-4 and IL-13 expression alleviates IgE class switching and increases the TH2 cytokine-dependent expression of CD14 (Lauener, Goyert et al. 1990). Consequently, pattern-recognition receptor expression is elevated and innate immune responsiveness is improved, which inhibits TH2-type immune responses. The finding of an increase in PRR expression among school-age farm children contributes to the fact that immunological effects of early farm exposure may be maintained into adulthood.

Another key factor in this model is represented by the cytokine IFNγ. Low levels of IFNγ at birth are known to be associated with an increased risk for the later development of allergic symptoms and atopic disease (Wright 2004; Vuillermin, Ponsonby et al. 2009). It is assumed that contact with several animal species may activate dentritic cells and TLR-expressing natural killer (NK) cells which release IFNγ (Muzio, Bosisio et al. 2000). IFNγ at birth can reduce allergen-induced TH2 cell differentiation and stimulate high IL-12 production by dendritic cells (Snijders, Kalinski et al. 1998) . Consequently, accelerated TH1-type immune responses are enhanced.

However, many of the proposed possible underlying mechanisms in the model are based on hypothesis. Therefore, future investigation of the cellular, genetic and epigenetic mechanisms of immunological pathways with respect to farm environment is crucial.

In conclusion, it is evident that farm environment especially during childhood and pregnancy of the mother conveys protection from respiratory allergies with a prolonged effect into adulthood. Studies suggest that contact with farm animals, their fodder and their products, such as unprocessed milk consumed directly from the farm, and the microbial burden in farming environment contribute to the 'farm effect'. Even though a number of steps in innate and adaptive immunity are involved many immunological mechanisms remained unravelled.

1.2 PAULCHEN and PAULINA: two birth cohort studies

In order to contribute to explaining mechanisms on neonatal immune responses PAULCHEN und PAULINA were conducted. The general goal of the two birth cohort studies is the identification of underlying immunological mechanisms and interactions of different cell populations in early life in association with epidemiological data.

1.2.1 PAULCHEN- prospective cord blood study in rural Southern Germany

Pregnant mothers were recruited in an obstetric clinic in rural southern Germany. Study enrolment was conducted from July 2005 to September 2007 by trained midwives in the last trimester of pregnancy. Inclusion criteria comprised healthy neonates and mothers with uncomplicated pregnancies. Exclusion criteria included preterm deliveries, perinatal infections and maternal use of antibiotics in the last trimester and chronic diseases. From 84 mothers enrolled in the study, 82 (97%) cord blood samples were included in the study. Two subjects were excluded because of perinatal infections. Mothers completed a detailed questionnaire regarding rural lifestyle, including detailed farming exposures. Maternal farm exposure was defined as the mother living and regularly working on a farm in the last 1 to 5 years and during pregnancy. Non-farming mothers lived in the same rural area but not on a farm. Specific exposure to stables/barns and animal species and milk intake were documented during pregnancy. Potential covariates, including delivery mode, sex, birth characteristics, siblings, education, maternal atopy, smoking, and miscarriage, were determined by using a questionnaire. Informed consent was obtained from the mothers for participation in the study, including cord blood collection. Approval was obtained from the local human research committee of the Bavarian Ethical Board, LMU Munich, Germany.

The aim of PAULCHEN is to assess the influence of environmental, lifestyle factors and genetic background on neonatal immune responses. A selection of different immune responses in cord blood mononuclear cells is investigated in association with lifestyle factors and atopic history of the parents.

1.2.2 PAULINA - Pediatric Alliance for Unselected Longitudinal Investigation of Neonates for Allergies

Pregnant mothers were recruited in an obstetric clinic in Munich, Germany. Study enrollment was conducted from July 2005 to September 2007 by trained midwives in the last trimester of pregnancy. Inclusion criteria comprised healthy neonates and mothers with uncomplicated pregnancies. Exclusion criteria included preterm deliveries, perinatal infections, maternal use of antibiotics in the last trimester and chronic diseases. From 161 mothers enrolled in the study, 148 (92 %) cord blood samples were included in the study. Mothers completed a detailed questionnaire regarding lifestyle and atopic background. Potential covariates, including delivery mode, sex, birth characteristics, siblings, education, maternal atopy, smoking, and miscarriage, were determined by using a questionnaire. Informed consent was obtained from the mothers for participation in the study, including cord blood collection. Approval was obtained from the local human research committee of the Bavarian Ethical Board, LMU Munich, Germany.

The aim of Paulina is to assess the effect of microbial stimulation of cord blood cells on distinct cellular immune responses in association with a selection of epidemiological data.

1.2.3 Complexity of cytokine and gene expression data

The understanding of immunological mechanisms underlying human disease has increased greatly over the last decades. In this context, different regulatory mechanisms involving gene regulation (of e.g. cytokines) at mRNA level and expression of proteins play an important role. When a foreign antigen is recognized by the immune system a complex cascade of regulatory immune mechanisms become activated subsequently resulting in the transcription of genes, thus expression on mRNA level. This is followed by several complex pathways, finally leading to the release of cytokine secretion (protein expression), which function as mediators of immune and inflammatory reactions (Abbas and Lichtman 2009). In order to take the complex relationships into account appropriate statistical analysis is crucial (Genser, Cooper et al. 2007).

1.2.4 Data distribution

In order to statistically analyse immunological data it is important to assess the structure of the data, such as the underlying distribution. This is crucial because many statistical methods may only be validly applied if the data follow a certain kind of distribution as normality or log-normality. Unfortunately, the data are rarely normally distributed and fail to be transformable into a distribution on which parametric tests may be applied (e.g. a logarithmic transformation to make skewed data approximately normally distributed).

1.2.5 Multivariate structure

Complex and multiple relationships are often present between immunological parameters. Immunologists are generally interested in many different outcomes (e.g. clinical outcomes or cytokine concentrations) depending on various exposures (e.g. genetic variation or environment) in the presence or absence of other intervening immunological parameters (e.g. cytokines). Furthermore, it is frequently aimed to explain the complete causal pathways from a certain exposure (e.g. exposure to an allergen) to a (clinical) outcome (e.g. atopy or concentration of a certain cytokine). Therefore, data sets usually contain a large number of interacting variables that have to be taken into account by appropriate statistical models that allow adjusting for potential confounding and interaction effects.

1.2.6 Censoring

Another common characteristic is that datasets including cytokines may contain non-detectable values and thus have to address the issue of so called "censored data". When protein expression is measured by Luminex technology in early life, where several immune parameters are expressed at low concentrations, left censoring at a single censoring level can occur as the measured concentrations fall below the detection threshold and are, consequently, not quantifiable any more. Often single values are erroneously imputed (e.g. 0.01 or half of the detection limit) in order to include the data for the analysis.

On the other hand, "right censored" data may occur when the measurement exceeds a certain threshold on the top measurement scale and thus, is also not exactly quantifiable. Right censoring at multiple censoring levels may occur when genes or

cytokines are measured by real-time RT-PCR at mRNA level and are expressed in relation to a given housekeeping gene by the Δct formula (Heid, Stevens et al. 1996). However, in contrast to measurement by Luminex technology, right censoring in the context of measurement by real-time RT-PCR is less obvious.

The literature recommends different methods of dealing with censored or "non-detectable" data. These suggestions comprise substitution of the values above or below the detection level (Buckley, Liddle et al. 1997), Tobit regression (Tobin 1958), multiple Imputation (Lubin, Colt et al. 2004; Uh, Hartgers et al. 2008) and deletion (Hobbs, Muir et al. 2003) among others (Helsel 2005). However, simple substitution is not advisable as it may lead to strongly biased results (Helsel 2005). Other methods like Tobit regression require strong parametric assumptions which can rarely be fulfilled by cytokine data (Arabmazar and Schmidt 1982; Austin, Escobar et al. 2000). Multiple Imputation has been shown to be valid (Lubin, Colt et al. 2004; Uh, Hartgers et al. 2008) but is quite time consuming, especially when data sets are large. Furthermore, multiple imputation is not supported by all statistical packages. Therefore, it is important to use statistical tools that take censoring into account, are not prone to violating parametric assumptions, allow adjusting for covariates, potential confounding and interaction effects and are available in common statistical packages.

1.3   Aims of Thesis

Based on the complexity of cytokine data as described in the introduction the aims of the thesis are:

- To describe the characteristics of the PAULCHEN and PAULINA data sets especially with respect to censoring and data distribution

- To present distribution-free methods on how to compute summary statistics for censored immunological data

- To illustrate non-parametric statistical testing procedures on differences between two or more groups when censoring is present

- To introduce the Tobit regression model on ranks: a novel regression approach for non-normal censored data in order to allow adjusting for potential confounding and interaction effects

The methods and procedures described in this thesis are accompanied by numerous illustrative examples both taken from the PAULCHEN and PAULINA studies and artificially created. Additionally, SAS codes are given at the end of each chapter for all statistical methods and tests so that they can be reproduced by the reader.

2 Characteristics of the PAULCHEN and PAULINA data sets

The aim of this chapter is to define the concept of censoring, describe the kind and dimension of censoring in PAULCHEN and PAULINA and illustrate by means of two examples the typical distribution of the immunological variables in both data sets.

2.1 Definition of censoring

2.1.1 Left censoring

Measurements whose values are known to be below a certain threshold are non-detectable and thus left-censored because the exact value of the data is not known. It is merely known that the observation is below a certain threshold. e.g. If a laboratory kit is not capable of measuring exactly a concentration below, e.g. 1, the measured values that are below 1 are left censored. If the same kit is used in one laboratory there is only one detection limit. If different laboratories use different kits the detection limits may vary. Consequently, different detection levels may be present. Figure 2: shows an example of left censoring with more than one detection level. The concentration levels of case 3, 4 and 5 can be directly determined. However, the concentration levels of case 1 and 2 cannot be directly determined because they fall below the laboratory specific detection level. The only information available for these cases is that the measured concentration lies between 0 and 1 for case 1 and between 0 and 10 for case 2 respectively.

Figure 2: Example of left censoring with more than one detection level from different laboratories

The concentration levels of case 3, 4 and 5 can be directly determined. However, the concentration levels of case 1 and 2 fall below the detection levels and thus, cannot be directly determined. The only information available for these cases is that the measured concentration lies between 0 and 1 for case 1 and between 0 and 10 for case 2 respectively. The two different detection levels may arise from the fact that the measurements were performed in two different laboratories using different laboratory equipments. The first line at concentration 1 represents the detection threshold of laboratory 1, the second line at concentration 10 represents the detection threshold of laboratory 2.
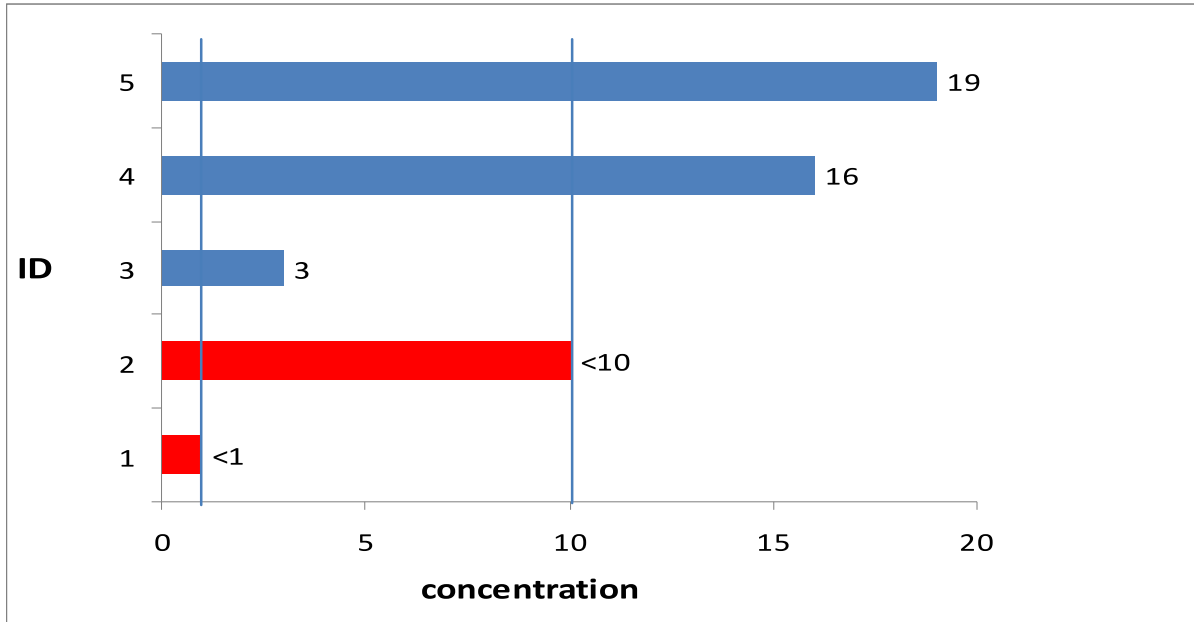
Luminex data in PAULCHEN and PAULINA are left censored

A typical example for "left censoring" in the PAULCHEN and PAULINA data is the measurement of cytokine concentrations at protein level by Luminex technology. To be detectable the concentration has to exceed a certain threshold determined by the respective lab technique, e.g. 1.3 pg/ml for IFN-y. Measurements by luminex may also occur as "right censoring" when the measurement exceeds a certain threshold on the top measurement scale and thus, are not exactly quantifiable. However, this rarely happens in the context of cytokine measurement by luminex. The exact detection threshold may vary according to the cytokine of interest. Left censoring at one or more detection levels is highly obvious and thus can easily be dealt with because every measurement below a certain value (detection level) can be directly marked as "censored". Table 2 shows three observations of the variable IFN-y. Case A and B are left censored because they fall below the minimal detectable concentration of 1.3 pg/ml. The information that can be drawn from these two cases is merely that the concentration is less than 1.3 pg/ml. Case C is not censored

because the measured concentration exceeds the minimal detectable concentration of 1.3 pg/ml.

Table 2: Example with two left censored observations from PAULCHEN data
Case A and B are left censored because they fall below the minimal detectable concentration of 1.3 pg/ml. Case C is not censored because the measured concentration exceeds the minimal detectable concentration of 1.3 pg/ml.

Case	Concentration in pg/ml	Censored
A	<1.3	yes
B	<1.3	yes
C	2.5	no

2.1.2 Right censoring

Data in medical and industrial studies are most often right censored, merely known as being greater than a certain threshold. In other words if a measurement exceeds a certain threshold it is not quantifiable any more. Typically, this happens with survival studies whose goal is to assess the length of time until an event occurs e.g. peoples´ death after receiving a certain drug. Figure 3 shows such a scenario. When patients survive the complete study time or drop out of the study at a certain time point without dying the conclusion is that patients survive until this time or longer. As a consequence, the observations are right censored at different thresholds (here time points). In Figure 3 subjects 2, 3, 5 die after 6, 10 and 14 months (no censoring occurs). Subjects 1 and 4 survive at least 15 and 20 months respectively (right censoring occurs). They either drop out or reach the study end without dying.

Figure 3: Example of right censoring taken from classical survival analysis
The figure shows the survival time of five patients. If patients survive the complete study time or drop out of the study at a certain time point without dying it can be concluded that these patients survive until this time or longer (time point of death is unknown). As a consequence, the observations are right censored at different thresholds (here time points). Here, subjects 2, 3, 5 die after 6, 10 and 14 months (no censoring occurs). Subjects 1 and 4 survive at least 15 and 20 months respectively (right censoring occurs). Subject 4 drops out, subject 1 reaches the study end without dying.
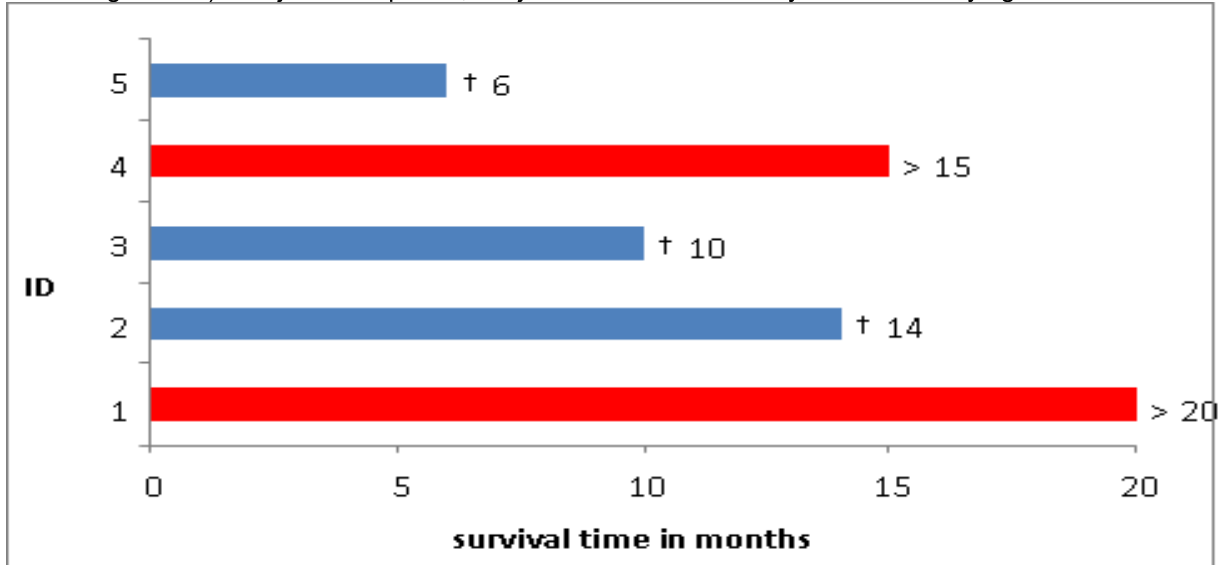
real-time RT-PCR at mRNA level data in PAULCHEN and PAULINA are right censored

Right censoring in cytokine and other gene expression measurements may be present when they are assessed at mRNA level by real-time RT-PCR and expressed as Δct (cycle threshold) in relation to a housekeeping gene according to the formula $\Delta ct = ct$ (gene of interest) - ct (housekeeping gene) (Livak and Schmittgen 2001). Although censoring is less obvious than for low protein concentrations it is as crucial. Δct values are often right censored at multiple detection levels. Figure 54 explains the underlying mechanism for the existence of right censored data at multiple detection levels. During real-time RT-PCR, mRNA expression of a single gene becomes amplified and can be assessed quantitatively. Amplification of DNA results in an increase of a fluorescence signal. The threshold for detection of fluorescence above background is determined. The cycle at which the fluorescence from a sample crosses the threshold is called the cycle threshold (ct). In Figure 4 the ct-value for gene of interest E2 is 31.1. If the threshold cycle exceeds a certain predetermined maximum cycle threshold (ct) (e.g. the maximum number of cycles run by the real-time RT-PCR cycler) the value is considered to be very low and not-quantifiable (comparable to non-detects using LUMINEX technology). In the figure not-quantifiable gene expressions are represented by gene G12 and gene H12. A high

ct-value refers to low expression of a gene, a low ct-value describes high gene expression. Thus, by showing the highest possible ct-value for the specific PCR-cycler, the respective value is greater than the predetermined maximum cycle threshold (in figure approximately 39) and is thus right-censored. The determined ct from the gene of interest is set relative to the ct value of a housekeeping gene (Δct). In the figue the housekeeping genes are represented as E1, G1 and H1. Housekeeping genes are characterized by stable expression over time and stimulation conditions and per definition their cycle threshold may vary only slightly from sample to sample. Thus, the values are normalized for possible variation in the amount and quality of mRNA between different samples. If the ct from the gene of interest is right censored, the resulting Δct contains a right censored value and has multiple quantification (censoring) levels. A numerical example taken from Figure 4 is illustrated in Table 3 the determined cycle threshold is 39.

Figure 4: Example of right censored mRNA expression data with multiple censoring levels
D2, G12, E2 are genes of interest. These are set relative to the house keeping genes D1, E1 and H12. However, D2 and G12 are not quantifiable and thus, right censored.
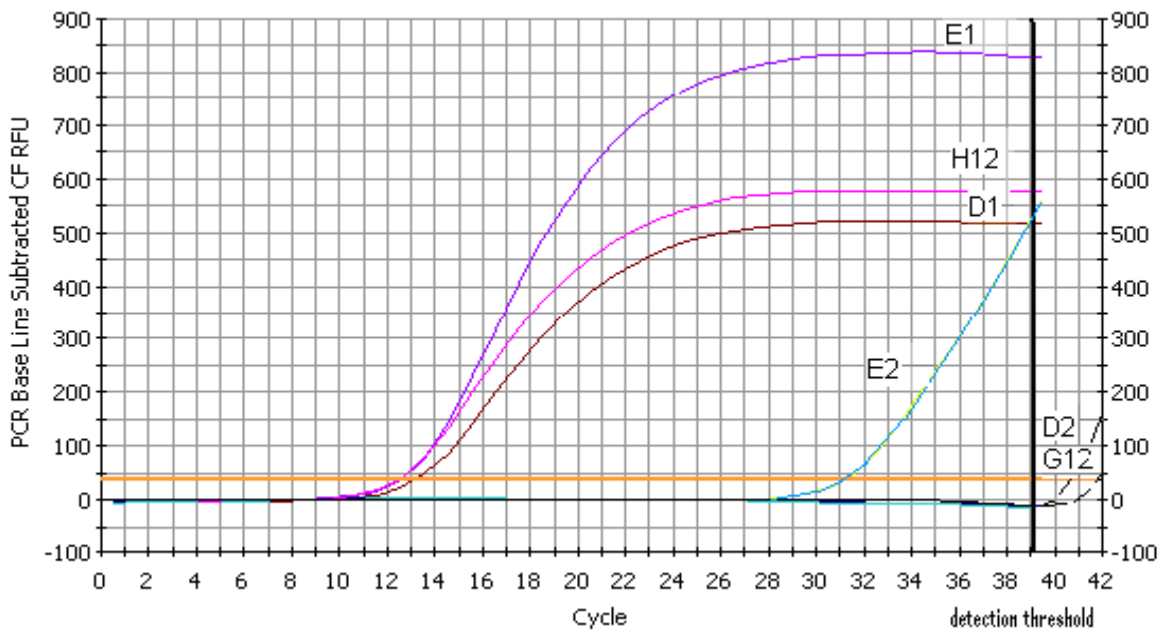
Table 3: Numerical example of right censored mRNA expression data with multiple censoring levels.
The fact that mRNA expression may be right censored with different censoring levels is illustrated by an example with three cases taken from PAULCHEN data.

Consider the following three cases in which a maximum cycle threshold (CT) of 39 is thought to be the detection limit.

Case	gene of interest	CT of gene of interest	Gene detection	housekeeping gene	CT of housekeeping gene	Δct value	Censored Δct value	Interpretation of Δct value
A	E2	31.1	yes	E1	12.6	18.5	no	18.5
B	G12	>39	no	H12	12.4	26.6	yes	>26.6
C	D2	>39	no	D1	13.1	25.9	yes	>25.9

A. The threshold cycle of gene of interest E2 for case A is 31. The threshold cycle of the housekeeping gene E1 is 12.6. The Δct value is 18.5 (31.1 -12.6 =18.5). In this case no censoring is present.

B. The threshold cycle of gene of interest G12 for case B exceeds 39 and thus, is not quantifiable any more. The value has to be considered as "greater than 39" and, consequently, is right censored. The threshold cycle of the housekeeping gene H12 is 12.4. The Δct value is 26.6 (39 – 12.4 = 26.6). However, as this Δct value contains the right censored value of 39 the resulting Δct value is also right censored. Thus, the Δct value has to be interpreted as greater than 26.6.

C. The threshold cycle of gene of interest D2 for case C exceeds 39 and thus, is also not quantifiable anymore and, consequently, as in example B, also right censored. The threshold cycle of the housekeeping gene D1 is 13.1. It follows that the Δct value is 25.9 (39 – 13.1 = 25.9). Again, as this Δct value contains the right censored value of 39 the resulting Δct value is also right censored. The resulting Δct value has to be interpreted as greater than 25.9. However, as the threshold cycle of the housekeeping gene is different to the one in case B (26.9 vs. 25.9) the resulting right censored Δct value has a different censoring level than in case B.

It can be concluded that Δct values may be right censored data at different censoring levels.

2.2 Dimension of censored data in PAULCHEN and PAULINA

2.2.1 PAULINA

Out of 148 variables assessing immunological parameters 62 variables contain censored observations. This corresponds to a proportion of 42%. From these 62 variables 35 (56%) are derived by real-time RT-PCR at mRNA level and 27 (44%) by luminex technology.

Table 4 illustrates the dimension of censoring in the PAULINA data set. It depicts the number of variables with censored observations and descriptive statistics for the proportion of censored observations per variable. All variables that contain at least one censored observation are taken into account. The median proportion of censored observations per variable is 30.27%. The upper quartile is represented with 71.79% and the lower quartile with 17.28%. The minimum and maximum proportions lie between 1.51% and 98.63 % resp.

Table 4: descriptive statistics for the proportion of censored observations per variable in PAULINA data
Number of variables with censored observations and descriptive statistics for the proportion of censored observations per variable for the PAULINA data set. Variables with at least one censored observation are taken into account

N° of variables with censored observations	Median (IQR) proportion of censored observations per variable	Minimum proportion of censored observations per variable	Maximum proportion of censored observations per variable
62/148 (42%)	30.27% (17.28% -71.79%)	1.5%	98.6%

2.2.2 Paulchen

Out of 162 variables assessing immunological parameters 92 variables contain censored observations. This corresponds to a proportion of 57%. From these 92 variables 56 (61%) are derived by real-time RT-PCR at mRNA level and 36 (39%) by luminex technology.

Table 4 illustrates the dimension of censoring in the PAULCHEN data set. It depicts the number of variables with censored observations and descriptive statistics for the proportion of censored observations per variable. All variables that contain at least one censored observation are taken into account. The mean proportion of censored observations per variable is 41.61%, the median proportion 33.68%. The upper quartile is represented with 64.88% and the lower quartile with 12.50%. The minimum and maximum proportions lie between 1.30% and 100.00 % resp.

Table 5: descriptive statistics for the proportion of censored observations per variable in PAULCHEN data
Number of variables with censored observations and descriptive statistics for the proportion of censored observations per variable for the PAULINA data set. Variables with at least one censored observation are taken into account

N° of variables with censored observations	Median (IQR) proportion of censored observations per variable	Minimum proportion of censored observations per variable	Maximum proportion of censored observations per variable
92/161 (57%)	33.68% (12.50% -64.88%)	1.30%	100%

2.3 Non-normal distribution of the variables in PAULCHEN and PAULINA

The majority of the data in both PAULCHEN and PAULINA are not normally distributed and cannot be transformed into a distribution in order to meet parametric assumptions. This is illustrated by two examples: The distributions of a variable derived by both luminex and real-time RT-PCR at mRNA level are presented in Figure 5 -Figure 8. The probability plots show how exactly the values of the cytokine the gene IL17F_PHA_dct (real-time RT-PCR at mRNA level) and IFN_g_M (luminex) fit with the percentiles of an underlying distribution like normality or lognormality. The underlying distribution with its mean and variance is represented by the reference line with the corresponding intercept and slope. The reference line of the underlying distribution is computed by maximum likelihood estimation taking into account the censored observations. If the data distribution of the variable of interest matches the underlying distribution the data points lie as close as possible on the reference line. The non-censored determined observations are represented by circles along the reference line and the censored observations by small vertical lines on the x-axis. Figure 5 and Figure 6 compare the distributions of the variables IL17F_PHA_dct and IFN_g_M with a normal distribution. It can clearly be shown that neither of the variables follow a normal distribution as most of the data points are pretty distant from the reference line.

Figure 5: non-normal distribution of variable with right censored observations from PAULCHEN data

The probability plot compares the distribution of the variable IL17F_PHA_dct to a normal distribution. It can clearly be noted that the variable does not follow a normal distribution as most of the data points are pretty distant from the reference line. The reference line represents normal distribution. Censored observations are marked along the x-axis.
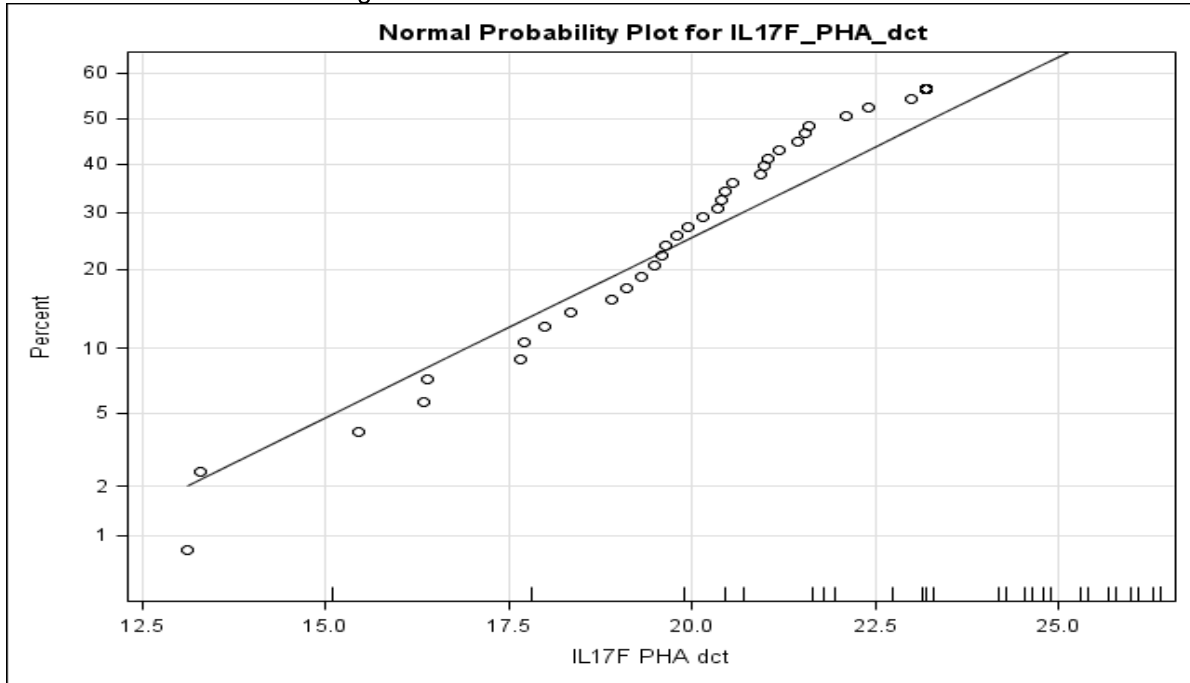
Figure 6: non-normal distribution of variable with left censored observations from PAULCHEN data

The probability plot compares the distribution of the variable IL17F_PHA_dct to a normal distribution. It can clearly be noted that the variable does not follow a normal distribution as most of the data points are pretty distant from the reference line. The reference line represents normal distribution. Censored observations are marked along he x-axis.
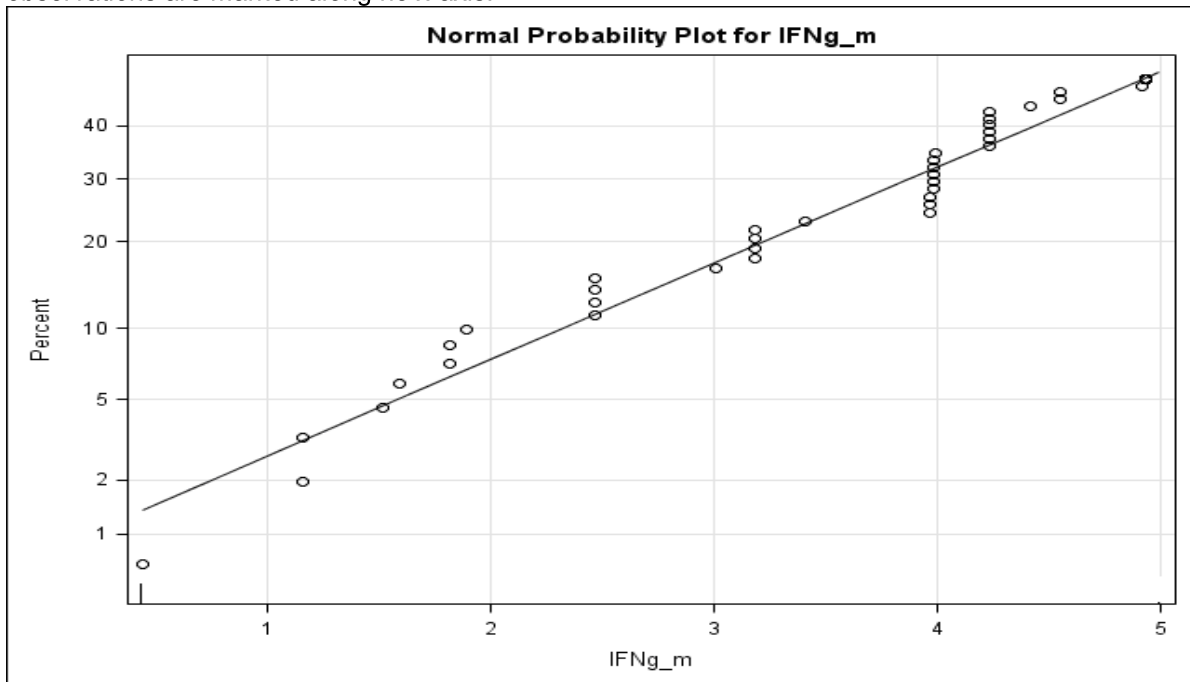
A similar pattern can be seen when the percentiles of a lognormal distribution are fit with the variables IL17F_PHA_dct and IFN_g_M (see Figure 7). Again the points do not fall on the straight line that represents the lognormal distribution. Thus, neither normality nor lognormality can be assumed.

Figure 7: non-lognormal distribution of variable with right censored observations from PAULCHEN data

The probability plot compares the distribution of the variable IL17F_PHA_dct to a lognormal distribution. It can clearly be noted that the variable does not follow a normal distribution as most of the data points are pretty distant from the reference line. The reference line represents lognormal distribution. Censored observations are marked along the x-axis.
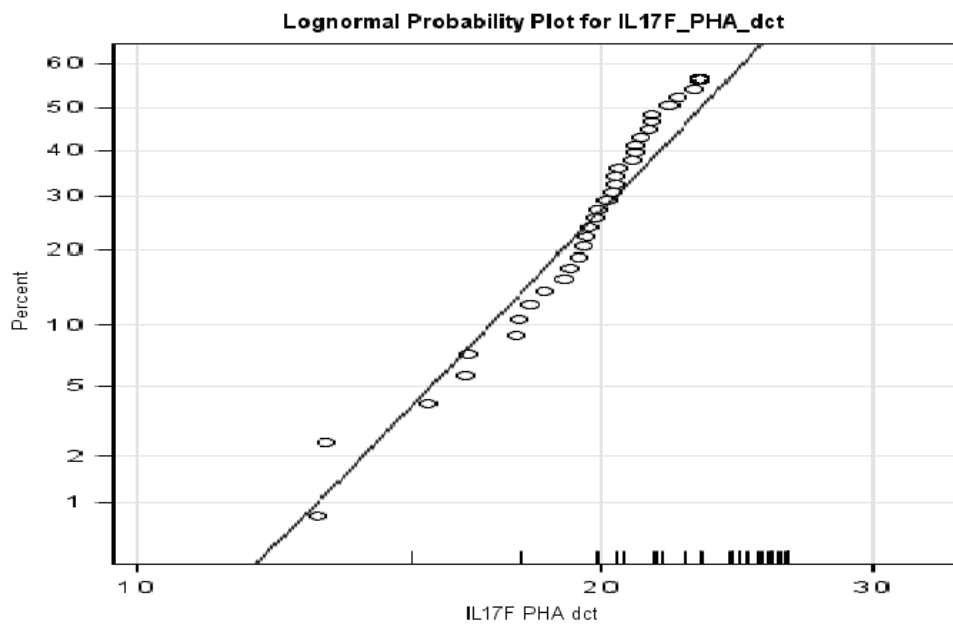
Figure 8: non-lognormal distribution of variable with left censored observations from PAULCHEN data

The probability plot compares the distribution of the variable IL17F_PHA_dct to a lognormal distribution. It can clearly be noted that the variable does not follow a lognormal distribution as most of the data points are pretty distant from the reference line. The reference line represents lognormal distribution. Censored observations are marked along the x-axis
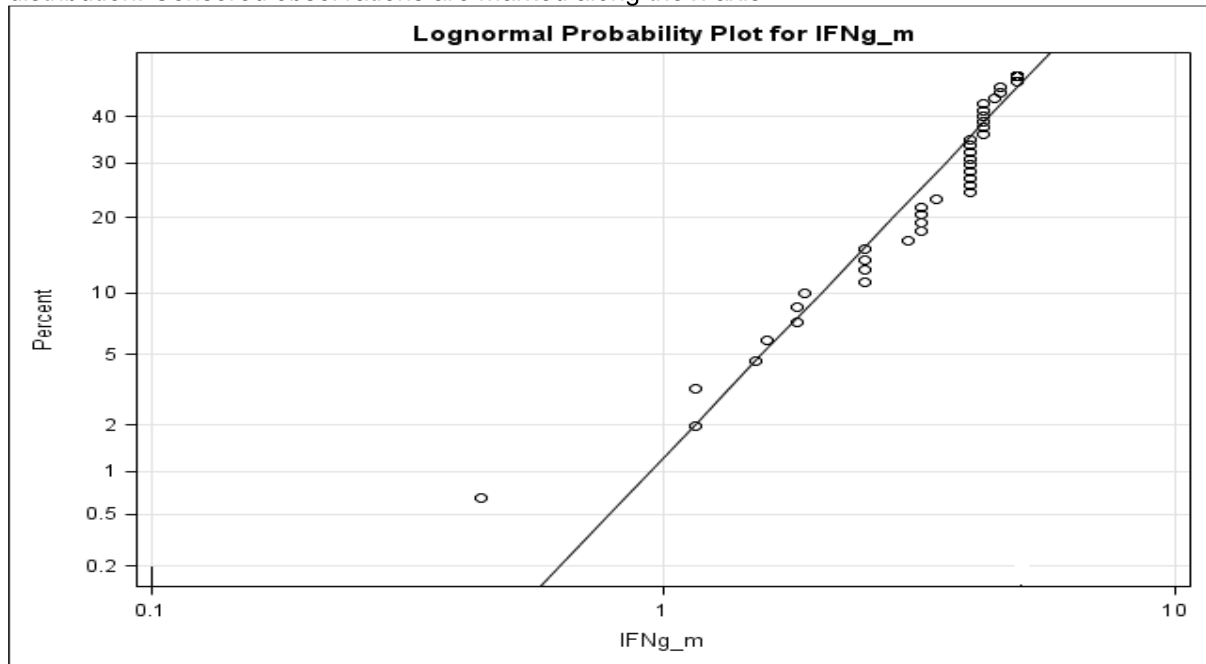
2.4 SAS code for distribution fit and summary of censoring

The SAS code in order to check the distribution fit and summary of censoring is as follows:

```
ods graphics on;
   proc lifereg data=paulchen;
       model IL17_m_dct*Censor(1) = / d = normal;
       probplot;
   run;
ods graphics off;
```

2.5 Conclusion of chapter

This chapter demonstrated two typical characteristics of the PAULCHEN and PAULINA data:

Data often contain left and right censored observations with both single and multiple detection levels.

Furthermore, the assumptions of an underlying distribution applying parametric tests are mostly violated in the setting of immunological measurements. Consequently, statistical instruments are required that take both characteristics of the data into account.

3   Summary statistics for non-normal censored data

In this chapter it is described how to compute summary statistics such as the median and its confidence intervals for immunological data (Luminex and Real-time RT-PCR measurements) when these are censored and do not follow a distribution upon which parametric tests can be applied.

3.1 Summary statistics for Luminex data can be computed by standard methods

The median is used as a statistical measure of location when a distribution is skewed or when one requires reduced importance to be attached to outliers. The median of a sample is defined to be the middle value when the observations are ordered from lowest to highest. Due to the fact that immunological data rarely follow a certain distribution (see 2.3) the median should be used as the measure of choice for these data in order to produce summary statistics.

When computing summary statistics such as the median and confidence intervals for protein expression data measured by Luminex technology, standard methods (as implemented in all statistical packages, e.g. Proc univariate in SAS) are applicable (Helsel 2005). The measured values are assigned ranks from lowest to highest. All values below the detection level are assigned tied ranks. As there is generally only one detection level per variable there is no loss of information. Based on ranks the median can be computed and its corresponding confidence intervals can be derived from a table with a continuous distribution based on order statistics (Helsel 2005) or calculated by Campell and Gardners´ formula (Campbell and Gardner 1988) as described below. In contrast to the interquartile range, which represents the dispersion in the study sample, the confidence interval gives an estimated range of values which is likely to include the parameter of interest in the population.

3.1.1 Median for an odd number of observations

When the sample contains an odd number of observations the median is calculated as follows:

$$Median = observation\ at\ R(\frac{n+1}{2}) \qquad (1)$$

Where

R=Rank
n=number of observations

Consider the following example with five observations that are already ranked from lowest to highest:

2 4 6 9 10
n=5

$$Median = observation\ at\ R(\frac{5+1}{2}) =\ R(3)$$

The 3rd rank corresponds to the observation with value 6. Consequently the median= 6.

3.1.2 Median for an even number of observations

When the sample contains an even number of observations then the median is calculated as follows:

$$Median = \frac{observation\ at\ R\ (\frac{n}{2})\ +\ observation\ at\ R\ (\frac{n}{2}+1)}{2} \qquad (2)$$

Consider the following example with six observations that are already ranked from lowest to highest:

2 4 6 9 10 12

n=6

$$Median = \frac{observation\ at\ R\ (\frac{6}{2})\ +\ observation\ at\ R\ (\frac{6}{2}+1)}{2} = 7.5$$

The median is the mean of the values at the 3rd and the 4th rank. Consequently the median= (6+9)/2= 7.5.

3.1.3 Computation of the median: Application to data measured by Luminex

Computation of the median by standard methods as described may easily be applied to variables measured by Luminex technology as these measurements generally contain censoring at one single detection level. The cytokine IFN_y_m is taken as an example. Values below the detection level are assigned tied ranks which are marked as 0.01 in the data set and ranked from lowest to highest. The median of the variable IFN_y_m is derived as follows:

The variable contains 76 observations. Consequently, formula (2) is applied.

$$Median = \frac{observation\ at\ R\ (\frac{76}{2})\ +\ observation\ at\ R\ (\frac{76}{2}+1)}{2} = 0.04$$

The median is the mean of the values at the 38th and the 39th rank. According to data set (not shown here) the 38th rank corresponds to value 0.03 and the 39th rank to value 0.05. Consequently the median= (0.3+0.5)/2=0.04.

3.1.4 Nonparametric confidence interval for the median

The confidence interval is the interval that contains with a probability of, in general, 95%, the true population value. Here, it is of interest to compute the confidence intervals of the median. The confidence intervals may be derived by methods that both assume and do not assume a certain distribution. Due to the fact that immunological data rarely follow a certain distribution the confidence interval of the median should be derived by a nonparametric method. The formula is given by (Campbell and Gardner 1988):

$$r = \frac{n}{2} - (N_{1-\alpha/2} \times \frac{\sqrt{n}}{2})\ \text{and}\ s = 1 + \frac{n}{2} + (N_{1-\alpha/2} \times \frac{\sqrt{n}}{2}) \tag{3}$$

Where

    r= lower bound of the confidence interval

    s= upper bound of the confidence interval

    n= sample size

    $N_{1-\alpha/2}$= is the corresponding value from the standard Normal distribution for the 100(1-α/2) percentile

3.1.5 Computation of the confidence interval: Application to data measured by Luminex

The formula (3) for the confidence interval of the median is applied to the variable IFN_y_m:

$$r = \frac{76}{2} - (1.96 \times \frac{\sqrt{76}}{2}) = 29^{th} \text{ rank}$$

$$s = 1 + \frac{76}{2} - (1.96 \times \frac{\sqrt{76}}{2}) = 48^{th} \text{ rank}$$

Rank 48 and rank 29 correspond to the values 0.77 and 0.01 respectively in the data set.

The confidence intervals (CI) may also be looked up in a binomial table (Table 6) with nonparametric two sided confidence intervals for the median: (http://www.math.unb.ca/~knight/utility/MedInt95.htm). With a known sample size the corresponding ranks of the upper and lower endpoints may be derived directly from the table.

Example (IFN-y): N=76

       Lower endpoint (L in table)= rank 29

       Upper bound (U in table)= rank 48

Again, rank 48 and rank 29 correspond to the values 0.77 and 0.01 respectively in the data set.

Conclusion: the median 0.04 has the corresponding CI (0.01; 0.77).

Table 6: Nonparametric two sided 95% (or better) confidence intervals for the median
With a known sample size N the corresponding ranks of the upper (U) and lower (L) endpoints may directly be derived from the table. The p-value gives the probability that the median lies between the lower and upper endpoints which is 95% or better

N	L	U	P<0.05	N	L	U	P<0.05	N	L	U	P<0.05
1	.	.	.	41	14	28	0.02753	81	32	50	0.04483
2	.	.	.	42	15	28	0.04356	82	32	51	0.03524
3	.	.	.	43	15	29	0.03154	83	33	51	0.04752
4	.	.	.	44	16	29	0.04877	84	33	52	0.03753
5	.	.	.	45	16	30	0.03570	85	33	53	0.02946
6	1	6	0.03125	46	16	31	0.02590	86	34	53	0.03985
7	1	7	0.01563	47	17	31	0.03999	87	34	54	0.03142
8	1	8	0.00781	48	17	32	0.02930	88	35	54	0.04221
9	2	8	0.03906	49	18	32	0.04438	89	35	55	0.03342
	...				...				...		
27	8	20	0.01916	65	25	41	0.04635	105	42	64	0.03130
28	9	20	0.03570	66	25	42	0.03558	106	43	64	0.04087
29	9	21	0.02412	67	26	42	0.04980	107	43	65	0.03295
30	10	21	0.04277	68	26	43	0.03846	108	44	65	0.04281
31	10	22	0.02945	69	26	44	0.02949	109	44	66	0.03462
32	10	23	0.02006	70	27	44	0.04139	110	45	66	0.04476
33	11	23	0.03508	71	27	45	0.03193	111	45	67	0.03631
34	11	24	0.02431	72	28	45	0.04437	112	46	67	0.04674
35	12	24	0.04096	73	28	46	0.03442	113	46	68	0.03802
36	12	25	0.02882	74	29	46	0.04739	114	47	68	0.04872
37	13	25	0.04703	75	29	47	0.03695	115	47	69	0.03975
38	13	26	0.03355	76	29	48	0.02863	116	47	70	0.03227
39	13	27	0.02370	77	30	48	0.03954	117	48	70	0.04150
40	14	27	0.03848	78	30	49	0.03079	118	48	71	0.03379

3.1.6 Summary statistics computed in SAS for data measured by Luminex

The most convenient form in order to compute summary statistics is to use a statistical program such as SAS. Unsurprisingly, the results computed by SAS and by hand are identical (Table 7). Furthermore, the table compares the median and its corresponding confidence intervals derived by the standard methods to the median and its corresponding confidence intervals derived by the Kaplan-Meier method. The Kaplan-Meier method is the method of choice that takes censoring into account (see chapter below).

Table 7: median with confidence intervals derived by the standard method and Kaplan-Meier method
The results from the standard methods are identical to the ones from the Kaplan-Meier method. Consequently, in the setting with censoring at one single detection level it is not necessary to use the more complicated Kaplan-Meier method which is considered to be the classical nonparametric method for censored data.

Standard methods (Calculated by hand and by SAS)		Kaplan-Meier method	
Median	*CI (lower; upper)*	*Median*	*CI (lower; upper)*
0.04	0.01; 0.77	0.04	0.01; 0.77

The results are also identical here. Consequently, it may be concluded that when computing summary statistics for protein data measured by Luminex (censoring at one single level) standard methods may be applied without distorting the result. In the setting with censoring at one single detection level it is not necessary to use the more complicated Kaplan-Meier method which is considered to be the classical nonparametric method for censored data.

3.1.7 SAS code for summary statistics of Luminex data

Standard method:

```
proc univariate data=paulchen CIQUANTDF ;
var IFN_g_M;
run;
```

3.2 Summary statistics for Real-time PCR data should be computed by the Kaplan-Meier method

In contrast to Luminex data with left censoring, summary statistics for right censored Real-time PCR data have to be computed by the Kaplan-Meier method (Kaplan and Meier 1958). The non-parametric Kaplan-Meier (K-M) method is considered the standard method for producing summary statistics for right censored survival data at multiple detection levels by calculating the survival probability (Kaplan and Meier 1958). However, for immunological analysis, it is not relevant to calculate the probability of surviving but rather the probability of exceeding a certain gene expression and thus Δct value. Nevertheless, the underlying idea is similar. In the following, the outcome is called exceedance probability, instead of survival probability. How to produce a lifetable according the Kaplan-Meier method is illustrated and described in detail in order to compute summary statistics for Δct values. Therefore, an example (see Table 8) of immunological data is taken to illustrate the application of the Kaplan-Meier method. The Real-time PCR data that are used are from the Paulchen Study (Schaub, Liu et al. 2009). It is aimed to assess whether the gene expression (Δct value) of a certain cytokine (IL-17F-PHA) is dependent on a certain exposure. The data were not normally distributed and could not be transformed to a valid distribution.

Table 8: Characteristics of an example with right censored gene expression (IL-17F-PHA) expressed as Δct and assessed by real time RT-PCR

Exposed	Sample Size	Number right Censored	Percent right Censored
no	45	25	55.55
yes	16	4	25.00
total	61	29	47.54

3.2.1 Creating a lifetable according to the Kaplan-Meier method

Before the median and the corresponding confidence interval can be derived a lifetable and a survival plot of the survivor function have to be created. Table 9

shows how the survivor function according to the K-M method is applied to the Δct data for the unexposed group.

Table 9: The Kaplan-meier method is applied in order to produce summary statistics for right censored gene expression data

The survival function according to the K-M method is applied to example with right censoring (IL-17F-PHA) at multiple detection levels for the unexposed group (see Table 8).

ΔCT-value	Number Censored (NC)	Number Detected (ND)	Number Left (NL) = NL(before) – (NC(before)+ND(before))	Incremental probability P=(NL-ND)/NL	Probability of exceeding Δct value S=P*S(before)
0	0	0	40	(40-0)/40=1	1
14.45	1	0	40	-	-
15.70	0	1	39	(39-1)/39=0.97	0.974*1.00=0.97
16.80	0	1	38	(38-1)/38=0.97	0.97*0.97=0.95
...					
20.15	0	1	15	(15-1)/15=0.93	0.933*0.516=0.48
...					
23.30	1	0	3	-	-
23.70	0	1	2	(2-1)/2=0.50	0.50*0.177=0.08
25.05	1	0	1	-	-

Δct value: measured cytokine concentration of the observation.
Number censored: number of censored measurements with given Δct value
Number detected: number of detected (uncensored) measurements with given Δct value
Number Left: number of observations left that exceed given Δct value
Incremental probability: probability of the NL to exceed the cytokine concentration of the given observation. In case of censoring the incremental probability is not calculated, but NL is reduced by NC
Probability of exceeding Δct value: the product of incremental probabilities up to that point

The first column represents the measured cytokine concentration of the observation expressed as Δct. The next column shows whether censoring occurred (0=no 1=yes). In the third column the number of observations that are detected is presented. The column (Number Left) depicts the number of observations left that exceed the last measured concentration. The incremental probability is the probability of the observations left (Number Left) to exceed the cytokine concentration of the given observation. The column "survival probability" presents the product of the incremental probabilities up to that point. Lifetables for the exposed group and both groups together are computed analogue (tables not shown).

3.2.2 Creating a survival plot

From the life table based on the Kaplan-Meier method a survival plot which plots the survival functions can be created. Censored values are represented by crosses. In the setting of immunological data the plot should not be interpreted as "survival" but as the probability of exceeding a certain gene expression level depending on the exposure. Consequently, the x-axis presents the measured gene expression level expressed as Δct. The y-axis represents the exceedance probability. The medians of each group may be easily visualized in the survival plot (see Figure 9) by drawing a horizontal line at 0.5 on the vertical axis and drawing a perpendicular line to the x-axis at the points where the horizontal line intersects the survival function. This procedure can be performed for any percentile of interest. For each measured gene expression level an exceedance probability can be achieved depending on the exposure.

Figure 9: Kaplan-Meier curve of a cytokine expressed as Δct and assessed by real time RT-PCR
The Kaplan-Meier curve represents the survival functions achieved by the life table (table 9). Right censoring with multiple detection levels are present. The plot should not be interpreted as "survival" but as the probabilities of exceeding a certain gene expression level depending on the exposure.
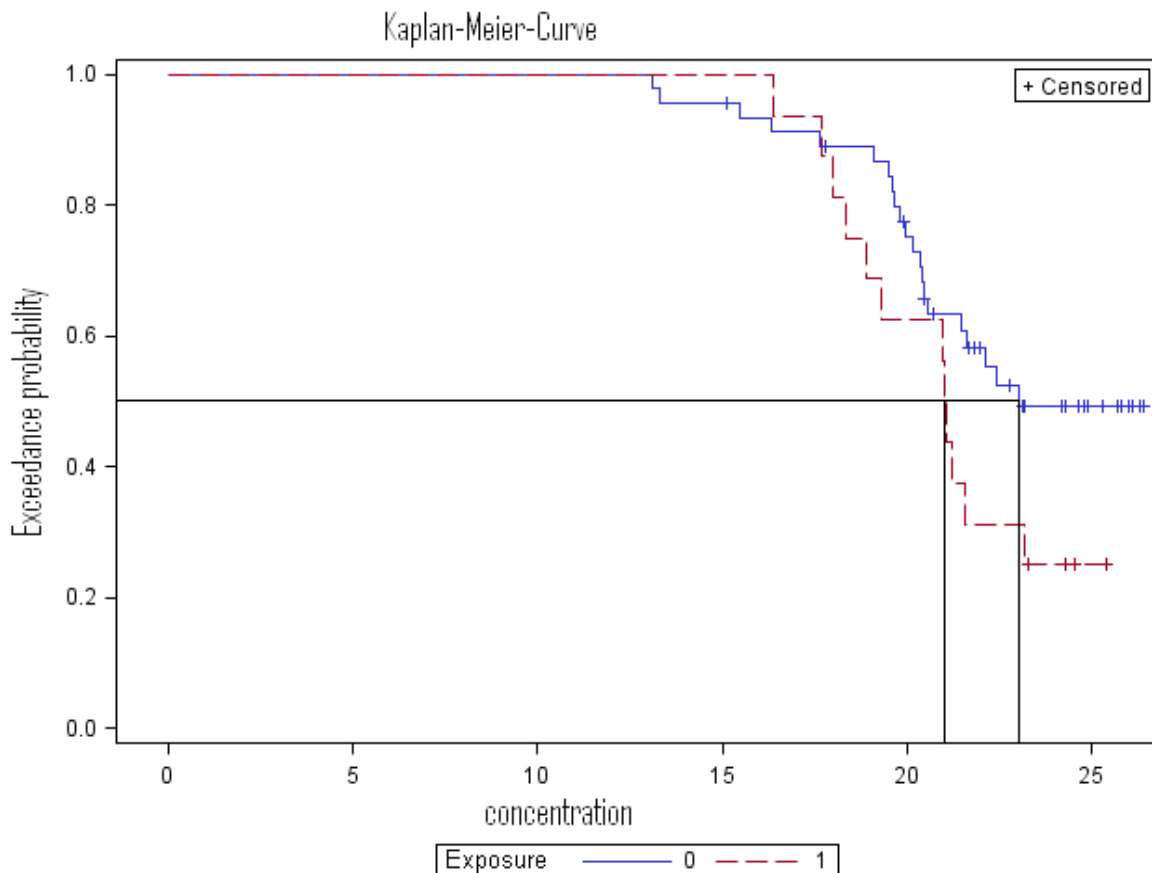
Example:

In the exposed group the probability of exceeding the Δct-value of 20 is approximately 60%. In the unexposed group the probability of exceeding the Δct-value of 20 is approximately 75%. Consequently, the gene expression level in the unexposed group is higher.

3.2.3 Median for right censored data

The median of the DCT-value according to the Kaplan-Meier method for right censored data is defined as the smallest DCT-value for which the survival probability (here: probability of exceeding the DCT-value) is less than or equal to 0.5. According to the lifetables the medians for the exposed are 21.02, for the unexposed 23.00 and both groups together 22.10 respectively. This way the Δct-value for any percentile of interest can be achieved. Additionally, the medians or any percentile of interest of each group may be easily visualized in the survival plot as described in the previous paragraph.

3.2.4 Confidence interval of the median for right censored data

In order to produce confidence intervals for the median nonparametric procedures are required as no underlying distribution can be assumed. Klein and Moeschberger (Klein and Moeschberger 2003) discourage the use of the confidence interval according to Kaplan-Meier as it becomes unreliable with small sample sizes. They rather recommend producing the confidence interval for the median by the B-C method after Brookmeyer and Crowley (Brookmeyer and Crowley 1982) which is based on Greenwood´s (Greenwood 1926) formula for the standard error of the survival function:

$$\text{Standard error of the survival function= s.e.(S)=}$$

$$S_j \times \sqrt{\sum_{j=1} \frac{ND_j}{NL_j(ND_j - NL_j)}} \qquad (4)$$

Where

S_j = survival function of observation at time *j*

ND_j = number of observations detected at time *j*

NL_j = number of observations left at time *j*

However, applying Greenwood´s formula is time consuming and can be approximated by Peto´s (Peto, Pike et al. 1976) formula in order to be easily calculated by hand:

$$\textit{Standard error of the survival function by Peto = s.e.(S)}= \sqrt{S_j{}^2 \times \frac{1-S_j}{NL_j}} \qquad (5)$$

Where

S_j = survival function of observation at time j

NL_j = number of observations left at time j

For illustration the formula is applied to the median of the variable IL-17F-PHA = 22.10 with its corresponding survival function S= 0.489 which are obtained from a lifetable as described in paragraph 3.2.1.:

$$\text{S.e. (0.489)} = \sqrt{0.489^2 \times \frac{1-0.489}{26}} = 0.07$$

Consequently, the survival function S=0.489 has the standard error S.e. =0.07.

The B-C sign method can be applied based on the standard errors of the survival functions of each observation. For each observation in the lifetable a ratio is created whose variation is approximately normal (see Table 10).

The ratio depends on the survival function of each observation, its standard error and the percentile of interest (here: the median):

$$\textit{B-C sign ratio:} \quad \frac{S_j - p}{s.e.(S_j)} \qquad (6)$$

Where

S_j = is the survival function of the observation at time j

p = target percentile (here: median)

S.e.= standard error of the survival function of the observation at time j

Example for computing the B-C ratio:

The B-C ratio for the median: As calculated before the median= 22.10 has the survival function S=0.489 with the corresponding standard error S.e. =0.07:

$$B\text{-}C \text{ sign ratio}: \frac{0.489 - 0.5}{0.07} = -0.22$$

Table 10: In order to compute confidence interval of the median for right censored data the B-C ratio method is applied.
The B-C ratio is based on the Survival function and its standard error. Its variation is approximately normally distributed. The survival times (here ΔCT-values) of detected observations that lie closest to the boundaries of the B-C ratio values >= + 1.96 or <=-1.96 represent the lower and upper bounds of the confidence intervals

ΔCT-value	Censoring Indicator	Survival function	Standard Error	B-C ratio
13.10	no	0.98	0.02	30.24
13.30	no	0.97	0.02	20.85
15.10	yes			
15.45	no	0.95	0.03	16.47

...

20.40	no	0.67	0.06	2.74
20.45	no	0.65	0.06	2.43
20.45	yes			
20.55	no	0.63	0.06	2.02
20.70	yes			
20.95	no	0.61	0.06	1.80
21.00	no	0.60	0.06	1.50

…

22.40	no	0.47	0.07	-0.52
22.75	yes			
23.00	no	0.44	0.07	-0.84
23.15	yes			
23.20	no	0.42	0.07	-1.16
23.20	yes			
23.30	yes			

...

26.40	yes			
26.40	yes			

Consequently, The survival times (here DCT-values) of detected observations that lie closest to the boundaries of the B-C ratio values >= + 1.96 or <=-1.96 represent the lower and upper bounds of the confidence intervals.

Example: The closest B-C ratio >= +1.96 is 2.02. The corresponding DCT-value is 20.55. It follows that the lower bound of the confidence interval is 20.55. However, a B-C ratio <= -1.96 cannot be computed because all observations >23.20 are censored. The lowest possible B-C ratio to be calculated is -1.16. Consequently, the

upper confidence interval cannot be computed. It is merely known that the upper bound is greater than 23.20.

3.2.5 Summary statistics computed in SAS for data measured by Real time PCR

Table 11 demonstrates that manual calculation of the B-C ratio method and the Kaplan-Meier method implemented in SAS yield identical results.

Table 11: Results for the median and its corresponding confidence interval produced by SAS and manual calculating of the B-C ratio method

Median	CI(lower,upper)
22.10	20.55, n.a.

3.2.6 SAS code for summary statistics of RT-PCR data

Kaplan-meier method:

```
proc lifetest method =pl data=N.paulchen;
       time IL17F_PHA_dct*il17fphact_c(1);
       strata pregbarn
run;
```

3.3 Consequences when censored data at multiple detection levels are calculated by inappropriate methods

In order to demonstrate the consequences when censored data at multiple detection levels are calculated by inappropriate methods an example of right censored data with multiple detection levels is analyzed by both appropriate and inappropriate methods. The appropriate method is the Kaplan-Meier method for the median and B-C ratio test for the confidence interval. In contrast, the inappropriate methods are the standard methods as described in paragraph 3.1 which are only applicable for left censored data with one detection level. Data from Table 8 are used as an example. In order to demonstrate the great impact of censoring on computing summary statistics four situations are used to illustrate this:

1. The median and CI are calculated on the original data of example. Censoring is properly taken into account in the Kaplan-Meier method.

2. Censoring is ignored in the exposed group. Censoring is not taken into account in the exposed group

3. Censoring is ignored in the unexposed group. Censoring is not taken into account in the exposed group

4. Censoring s ignored in both groups. Censoring is not taken into account in both the exposed and the unexposed group

In all four situations the original values of the data are maintained. Merely the proportion of censored observations is ignored. Ignoring means that the censored observations are not marked as censored for the analysis in the corresponding group.

The results are shown in table 12 and Figure 10 to 13.

Table 12: Consequences are shown when summary statistics are computed on censored data at multiple detection levels by inappropriate methods
This table compares the results from the Kaplan-Meier method (appropriate method) to the standard method (inappropriate method) when applied to right censored data with multiple detection levels. The median and the corresponding CIs according to the K-M method and the B-C ratio test depend on the censored observations in the data. In contrast, summary statistics by the standard method do not capture the information of censoring. Both median and confidence intervals remain unaffected.

N° per group	standard method		Kaplan-Meier method	
	median	CI	median	CI
Censoring is properly taken into account by the Kaplan-Meier method				
40 (55.55 % censored)	21.62	20.40-23.15	23.00	20.45- NA
13 (25 % censored)	21.02	18.90-23.30	21.02	18.35-23.20
Censoring is not taken into account in the exposed group by the Kaplan-Meier method				
40 (55.55 % censored)	21.62	20.40-23.15	23.00	20.45- NA
13 (0 % censored)	21.02	18.90-23.30	21.02	18.35-23.20
Censoring is not taken into account in the unexposed group by the Kaplan-Meier method				
40 (0 % censored)	21.62	20.40-23.15	21.62	20.35-23.00
13 (25 % censored)	21.02	18.90-23.30	21.02	18.35-23.20
Censoring is not taken into account in both groups by the Kaplan-Meier method				
40 (0 % censored)	21.62	20.40-23.15	21.62	20.35-23.00
13 (0 % censored)	21.02	18.90-23.30	21.02	18.35-23.20

The median and the corresponding CIs according to the K-M method and the B-C ratio test depend on the censored observations in the data. In situation 1 (Censoring is properly taken into account) and situation 2 (censoring is not taken into account in the exposed) the median and confidence interval in the unexposed group is 23.00 (CI: 20.45 - NA) and 21.03 (CI: 18.35 – 23.20), respectively. "NA" describes that the

horizontal line at 0.5 does not intersect the upper limit of the confidence interval. In situation 3 (censoring is not taken into account in the unexposed group) and situation 4 (censoring is not taken into account in both groups) the median and CI in the non-exposed group is 21.625 (CI: 20.35 – 23.00) and 21.025 (CI: 18.35 – 23.20), respectively. The information whether an observation is censored or not may affect summary statistics calculated by appropriate methods even though the values of the observations are not altered. In contrast, the median and the CI according to the standard methods stay unaffected by the amount of censoring. This fact underlines the importance of calculating summary statistics of right censored data with different detection levels by appropriate descriptive methods.

The survival plots in Figure 10 present graphically all four situations. It can be clearly seen that the shapes of the survival functions strongly differ according to the information of censoring.

Figure 10: The plot shows the Kaplan-Meier curve based on the situation in which censoring is properly taken into account in both groups
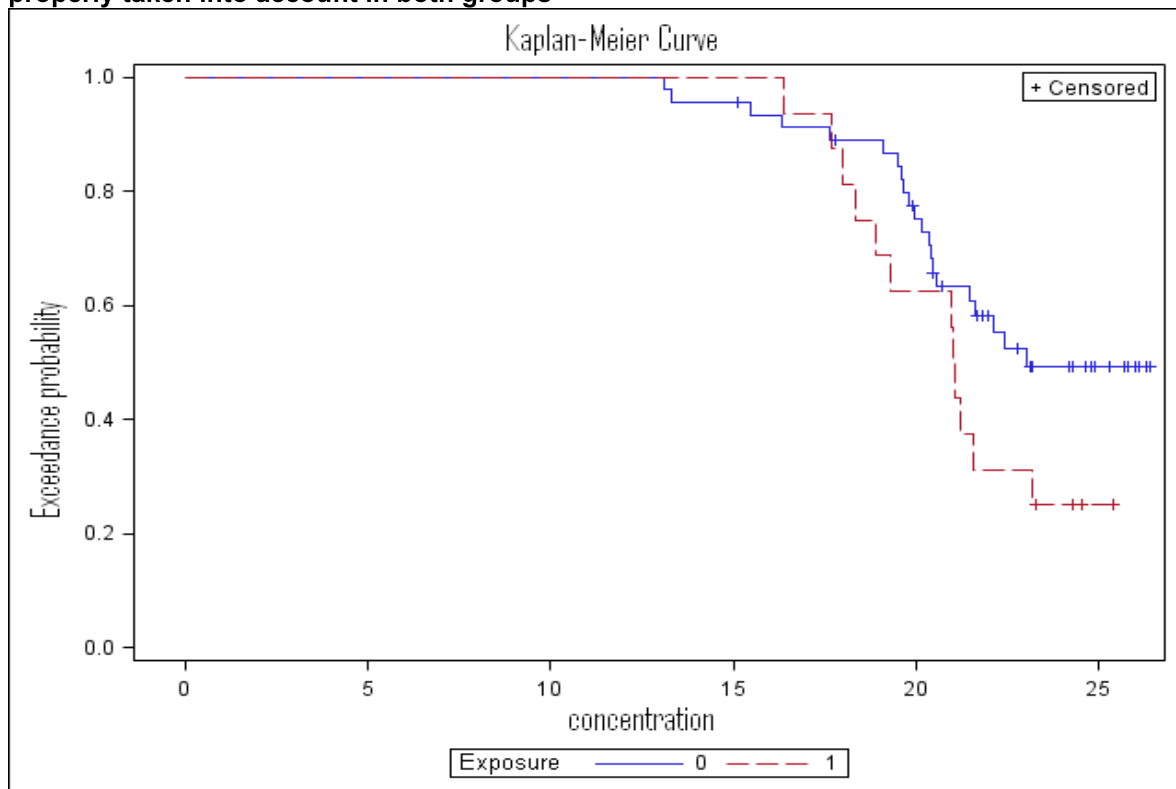
Figure 11: The plot shows the Kaplan-Meier based on the situation in which censoring is taken into account in the unexposed group and ignored in the exposed group
It can be clearly seen that the function of the unexposed group has a sharper drop than in figure 10. Consequently, the two functions differ more strongly.
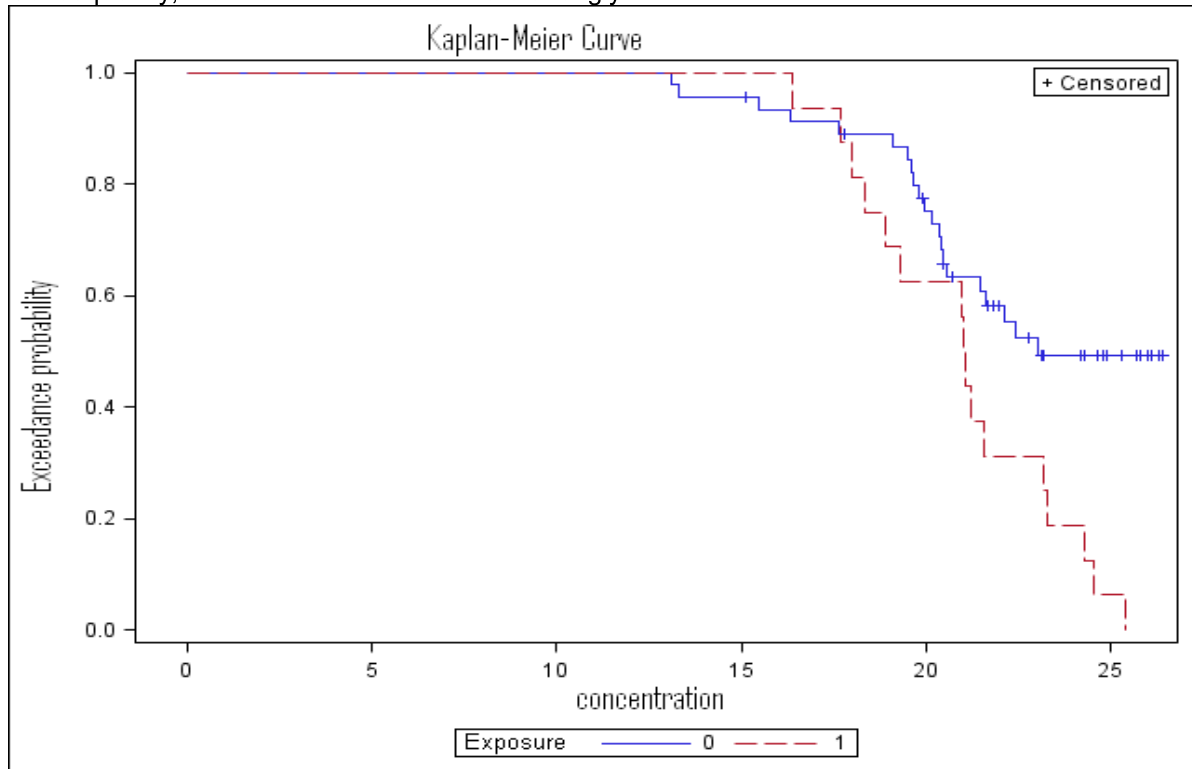
Figure 12: The plot shows the Kaplan-Meier based on the situation in which censoring is taken into account in the unexposed group and ignored in the exposed group
It can be clearly seen that the function of the exposed group has a sharper drop than in figure 10. Consequently, the two functions are more similar to each other
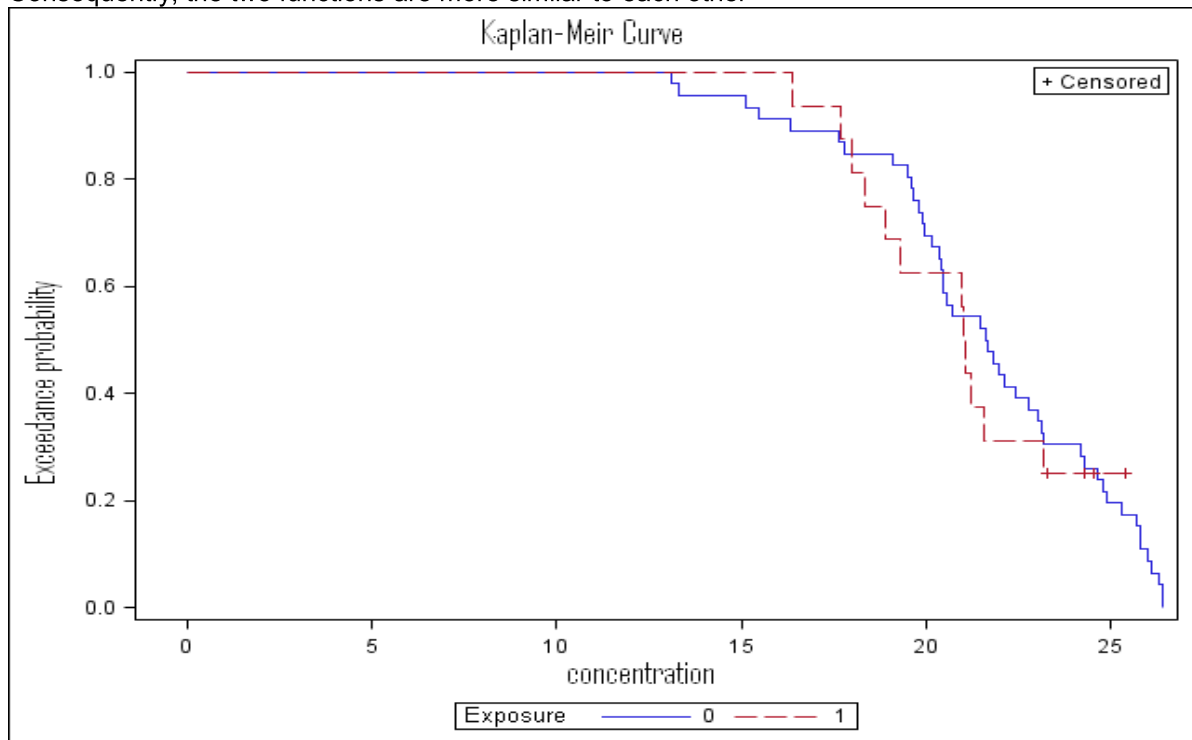
Figure 13: Censoring is not taken into account in both groups
Here in this scenario both functions have a sharper drop. Their shape is different to the one in figure 10 in which censoring is properly taken into account.
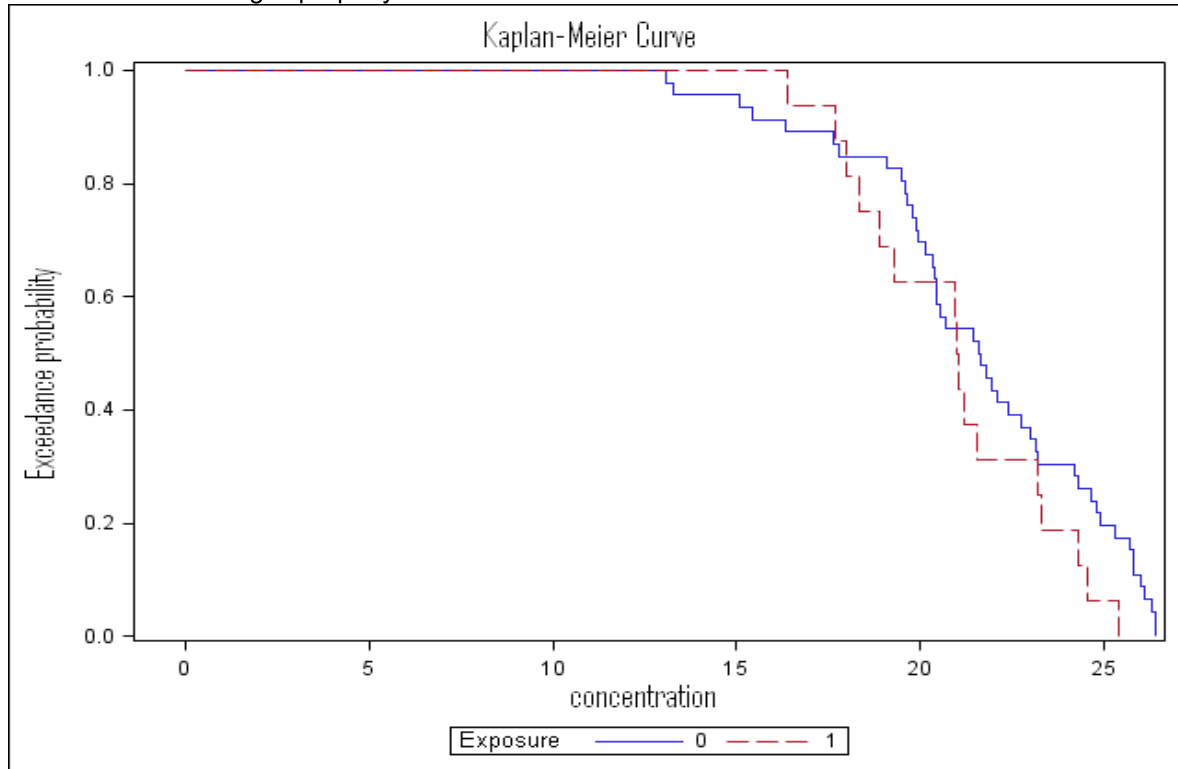
3.4 Conclusion of chapter

When computing summary statistics (e.g. median with confidence intervals) for left censored data with one detection level such as Luminex data, standard methods may be applied without loss of information or biasing the results. In contrast summary statistics for right censored data with multiple detection levels should be computed by methods that particularly take censoring into account such as the Kaplan-Meier method in order to avoid biased results.

4 Non-parametric statistical testing for group comparisons when censoring is present

In this chapter it is described how to perform statistical testing on differences between two or more groups in the context of immunological data derived by both Luminex and Real-time RT-PCR measurements when these are censored and do not follow a distribution upon which parametric tests can be applied.

4.1 Luminex data may be analyzed by the Wilcoxon rank-sum test

When statistically analyzing non-normally distributed left censored data at one single detection level, such as measurements by Luminex, standard non-parametric tests like the Wilcoxon rank-sum-test or the Kruskal-Wallis-Test may be applied without a loss of information (Helsel 2005).

In order to apply the Wilcoxon rank-sum test the measured values are assigned ranks from lowest to highest as depicted in Table 13. In case of ties (e.g. 0.01) the average rank is assigned (e.g. rank 19.5 to the value 0.01).

Table 13: Before applying the Wilcoxon ranks sum test the data are ranked
the measured values of the concentration are assigned ranks from lowest to highest. Furthermore exposure (yes/no) is indicated. Based on this information the Wilcoxon rank-sum-test or the Kruskal-Wallis-Test may be applied

Concentration in pg/ml	exposure	rank
0.01	no	19.5
0.01	no	19.5
....		
0.07	no	39
0.08	no	40
0.45	no	41.5
0.45	no	41.5
0.58	no	45.5
0.77	yes	45.5
....		
1.82	yes	62
1.82	yes	62
1.99	no	64
2.53	no	66.5
....		
3.84	no	74.5
4.56	no	76

Due to the fact that there is generally only one detection level per variable in Luminex data there is no loss of information by assigning tied ranks to all censored values below (or above) the detection (here: non-detectable values are represented by 0.01). Based on the ranks the Wilcoxon rank-sum test and the Kruskal-Wallis-Test respectively can be applied by the PROC NPAR1WAY procedure in SAS or calculated manually. To illustrate this the variable IFN_y_m is used as an example (Table 13). The aim is to check whether the cytokine concentration is significantly different depending on a certain exposure. The test static U from the Wilcoxon rank sum test is achieved by the following formula:

$$\textit{Test statistics } U = n_1 \times n_2 + \frac{n_1 \times (n_1 + 1)}{2} - R_1 \ (7)$$

Where
n_1 = number of observations in unexposed group
n_2 = number of observations in exposed group
R_1 = sum of ranks in unexposed group

The values achieved from Table 13 are entered in the formula:

$$U = 59 \times 17 + \frac{59 \times (59 + 1)}{2} - 2362 = 411$$

The test statistic U is approximately normally distributed. Consequently, Z can be computed from U by

$$Z = \frac{U - \frac{n_1 \times n_2}{2}}{\sqrt{\frac{n_1 \times n_2(n_1 + n_2 + 1)}{12}}} \approx N(0;1) \qquad (8)$$

The values achieved from Table 13 and formula 7 are entered in the formula 8:

$$Z = \frac{411 - \frac{59 \times 17}{2}}{\sqrt{\frac{59 \times 17(59 + 17 + 1)}{12}}} = 1.13$$

The corresponding p-value derived from a table with standard normal distribution is: 0.26.

4.1.1 Results computed by SAS

Table 14 shows that the results computed by SAS and by hand are identical. Furthermore, the table compares the z-score with the corresponding p-value derived by the standard methods to the z-score with the corresponding p-value derived by the generalized Wilcoxon test. The generalized Wilcoxon test is the method of choice that takes censoring into account at multiple detection levels (further description below).

Table 14: Comparison of z-scores calculated by the Wilcoxon-Test and the generalized Wilcoxon-Test

The z-scores and corresponding p-values by either method are identical. Consequently, in the setting of left censoring at one single detection level the Wilcoxon test can be applied without loss of information

Wilcoxon (calculated by hand and SAS)		Generalized Wilcoxon (SAS)	
z-score	p-value	z-score	p-value
1.13	0.26	1.13	0.26

The results are also identical here. Consequently, it may be concluded that when computing summary statistics for protein data measured by Luminex (censoring at one single level) standard methods may be applied without distorting the result.

4.1.2 SAS code for standard method

```
proc npar1way data=N.paulchen wilcoxon ;
class pregfmilk;
var IFN_y_m;
run;
```

4.2 Real-time PCR data should be analyzed by the generalized Wilcoxon-Test

In the case of analyzing censored gene expression data (shown as Δct) when testing for differences between two or more groups, statistical tests need to be applied that capture the information in the censored proportion of the Δct data. One possibility is the generalized Wilcoxon test (Peto and Peto 1972; Prentice 1978). This test is a modification of the rank-sum test and takes right censored data - leading to multiple detection limits- into account by estimating the U score of each censored value

according to the survival function of the previous uncensored observation. Basically, the test investigates whether survival distributions in two or more groups differ significantly. In SAS the Proc lifetest procedure covers the generalized Wilcoxon test. The generalized Wilcoxon test is applied to the example from Table 8 with right censoring at multiple detection levels (Δct assessed by real time RT-PCR). In order to conduct the generalized Wilcoxon-Test a table (Table 15) according to the Kaplan-Meier method has to be created.

Table 15: The generalized Wilcoxon test is based on the survival function according to the Kaplan-Meier method
Based on the survival function by the Kaplan-Meier method the U score is computed. Based on the U score the test statistic W can be derived.

ΔCT-value	Exposure	Censoring Indicator	Number Detected (ND)	Number Left (NL)	Survival (S_t)	U	U²
13.1	no	no	1	61	0,984	0.984	0,968
13.3	no	no	1	60	0.968	0,952	0.906
15.1	no	yes	0	59		-0,032	0.001
15.45	no	no	1	58	0.951	0.919	0.845
.....							
20.45	no	no	1	38	0.650	0.317	0.100
20.45	no	yes	0	37		-0.350	0.123
20.55	no	no	1	36	0.632	0.282	0.080
20.7	no	yes	0	35		-0.368	0.135
20.95	yes	no	1	34	0.614	0.246	0.061
21	yes	no	1	33	0.596	0.210	0.044
21.05	yes	no	1	32	0.578	0.174	0.030
....							
26.4	no	yes	0	1		-0.578	0.334
26.4	no	yes	0	0		-0.578	0.334

The U score based on the survival function at that observation is achieved as follows:

$$U \text{ score for uncensored observations: } U = S_t + S_{t-1} - 1 \quad (9)$$

$$U \text{ score for censored observations: } U = S_{t-1} - 1 \quad (10)$$

where

S_t is the value of the survival function at corresponding observation

S_{t-1} is the value at the survival function for the previous uncensored observation

Example:

U score for DCT-value (15.45) = 0.951 + 0.968 – 1 = 0.919

The test statistic W is obtained by summing the U scores for one group

$$W = \sum U_t \quad (11)$$

The variance of W is as follows:

$$Var(W) = m \times n \frac{\sum U^2}{(m+n) \times (m+n+1)} \quad (12)$$

Where

    m = number of observations in unexposed group

    n = number of observations in exposed group

The values achieved from Table 13 are entered in the formula 11 and 12

$$W = \sum U_t = -2.79$$

$$Var(W) = 3.52$$

The Z statistic is produced by dividing W by the square root of the variance of W:

$$Z = \frac{W}{\sqrt{Var(W)}}$$

$$Z = \frac{-2.79}{\sqrt{3.52}} = -1.49$$

Comparing z=-1.49 to table of the standard normal distribution yields a p-value of 0.14.

4.2.1 Results computed by SAS

The table compares the z-score with the corresponding p-value calculated by hand to the z score calculated by SAS with the LIFETEST procedure. The results are identical

Table 16: Results computed by SAS
The results calculated by hand and SAS are identical

Generalized Wilcoxon (calculated by hand and SAS)	
z-score	*p-value*
-1.49	0.14

4.2.2 SAS procedure for the generalized Wilcoxon test

proc lifetest method =pl data=N.paulchen;
time IL17F_PHA_dct*il17fphact_c(**1**);
strata pregbarn/test=wilcoxon;

run;

4.3 Consequences when censored data are analyzed by inappropriate methods

In order to demonstrate the consequences when censored data at multiple detection levels are calculated by inappropriate methods an example of right censored data with multiple detection levels is again analyzed by both appropriate and inappropriate methods. The appropriate method is the generalized Wilcoxon test. In contrast, the inappropriate method is the standard Wilcoxon Test as described in paragraph 4.1 which is only applicable for left censored data with one detection level. Data from Table 8 are used as an example. The same four situations are again used to illustrate in order to demonstrate the great impact of censoring on computing summary statistics:

1. The p-value is calculated on the original data of example. Censoring is properly taken into account

2. Censoring is ignored in the exposed group. Censoring is not taken into account in the exposed group

3. Censoring is ignored in the unexposed group. Censoring is not taken into account in the exposed group

4. Censoring is ignored in both groups. Censoring is not taken into account in both the exposed and the unexposed group

In all four situations the original values of the data are maintained. However, when censoring is ignored it means that the censored observations are not marked as censored for the analysis. Consequently, censoring is not taken into account for the analysis.

The results are shown in **Fehler! Verweisquelle konnte nicht gefunden werden.**. It can be seen that the p-value given by the generalized Wilcoxon test is strongly dependent on the amount of censoring in each group. In situation 1 the p-value amounted to 0.15, in situation 2 to 0.06, in situation 3 to 0.62, and in situation 4 to 0.39. In contrast, when data are analyzed by the standard Wilcoxon test the results

do not differ according to the amount of censoring in each group. The results are identical in each situation. Obviously, only the generalized Wilcoxon test captures the information whether an observation is censored which may affect the p-value even though the values of the observations are not altered. This fact underlines the importance of analyzing right censored data with different detection levels by appropriate statistical methods.

Table 17: Consequences are shown when inappropriate methods are applied to censored data at multiple detection levels
This table compares the results from the generalized Wilcoxon test (appropriate method) to the standard Wilcoxon test (inappropriate method) when applied to right censored data with multiple detection levels. The p-values according to the generalized Wilcoxon test depend on the censored observations in the data. In contrast, the p-value by the standard Wilcoxon test does not capture the information of censoring. The p-values remain unaffected.

N° per group	Wilcoxon Test	Generalized Wilcoxon Test
	p-value	p-value
Censoring is properly taken into account by the Kaplan-Meier method		
40 (55.55 % censored)	0.39	0.14
13 (25 % censored)		
Censoring is not taken into account in the exposed group by the Kaplan-Meier method		
40 (55.55 % censored)	0.39	0.06
13 (0 % censored)		
Censoring is not taken into account in the unexposed group by the Kaplan-Meier method		
40 (0 % censored)	0.39	0.63
13 (25 % censored)		
*Censoring is **not** taken into account in both groups by the Kaplan-Meier method*		
40 (0 % censored)	0.39	0.39
13 0 % censored)		

4.4 Conclusion of chapter

When performing statistical testing on differences between two or more groups in the presence of left censored data with one detection level such as Luminex data, standard methods as the Wilcoxon rank sum test may be applied without loss of information or biasing the results. In contrast, testing on differences between two or more groups in the presence of right censored data with multiple detection levels should be computed by methods that particularly take censoring into account such as the generalized Wilcoxon Test in order to capture the information of censoring in the data. Otherwise, biased results may be the consequence.

5 Tobit regression on ranks: a model for censored data without parametric assumptions

Multivariate structures are often present in immunological settings. Consequently, the statistical methods described in paragraph 3 and 4 are not applicable any more. More sophisticated methods such as regression techniques that may take possible confounding and interaction effects into account are crucial. However, in the context of cytokine and gene expression data these methods need to be robust against violating parametric statistical assumptions due to the fact that underlying distributions may not be assumed. Furthermore, the methods should capture the information of possible censored observations in the data. Therefore, the Tobit regression on rank transformed data is introduced in the following chapter. This model makes use of both the classical Tobit regression and the regression on rank transformed data. Additionally, its application will be shown by means of two examples taken from the PAULCHEN study. For the purpose of easiness simple regression is used as an illustration. Self evidently Tobit regression may be extended as any other regression to multivariate comparisons in order to adjust for possible confounding and interaction effects.

5.1 Description of the model

As mentioned before the generalized Wilcoxon does not allow adjustment for covariables and potential confounding. Hence, a regression technique such as the Tobit model is required to assess the simultaneous effect of different variables on the outcome in the presence of censoring. However, the disadvantage of the Tobit regression is that it is vulnerable to violation of the statistical assumptions like normality of the error distribution and equal variances of the residuals (homoscedasticity). E.g. In the presence of heteroscedasticity the Tobit estimates may become inconsistent (Arabmazar and Schmidt 1982; Austin, Escobar et al. 2000) and lead to biased results.

Based on the concept of Iman and Conover (Iman and Conover 1979; Conover 1980; Conover and Iman 1981) the data is ranked from lowest to highest and subsequently, the Tobit regression can be performed on the rank transformed data. Rank transformation procedures lead to distribution free tests.

The classical Tobit model according to the economist James Tobin (Tobin 1958) uses three types of information: 1) the observed values above detection limits b) the proportion of data below each detection limit and c) the assumed underlying distribution which may be e.g. normal or log-normal. The Tobit model estimates a regression model for the data above the detection level, and assumes the same distribution of errors for the censored data (below DL) as for the observed data. The Tobit regression is applicable for both left and right censoring and both single or multiple detection levels.

The Tobit model can be described in terms of a latent variable $y*$. Suppose

$$y* = \alpha + \beta x + \varepsilon \text{ where } \varepsilon \sim N(0, \delta^2) \qquad (13)$$

and the observed variable y satisfies :

$$y = y* \quad if \quad y > a \qquad (14)$$

$$y = a \quad if \, y* \, <= a \qquad (15)$$

where a=detection limit.

In contrast to linear regression whose parameters are derived by the least square method the parameters in the Tobit regression are derived by applying maximum likelihood estimation (MLE). MLE methods are computed by maximising a likelihood function L. For detailed information on computation of MLE refer to pertinent literature.

Tobit regression model on ranks is not calculated on the real values of the observations but on their corresponding ranks. Conover and Iman stated that calculating parametric tests on rank transformed data produce equivalent results as the corresponding non-parametric test (Iman and Conover 1979; Conover 1980; Conover and Iman 1981). Thus, according to Iman and conover the corresponding Tobit regression on ranks is given by:

$$R(y*) = \frac{n+1}{2} + \beta(R(x) - \frac{n+1}{2}) \quad (16)$$

$$R(y) = R(y*) \text{ if } R(y*) > R(a) \qquad (17)$$

$$R(y) = R(a) \text{ if } R(y*) <= R(a) \qquad (18)$$

where

$R(y)$ =is the rank of the corresponding variable.

$R\,(a)$ =detection limit of rank transformed variable.

The estimates of the Tobit model on ranks are estimates of the rank transformed variables R(y) which may be retransformed to their original values according to iman. This procedure is described in paragraph 5.7.

As Tobit regression in general is applicable to both left and right censored data with both single or multiple detection levels the Tobit regression on ranks can be used for both Luminex and Real time PCR data when expressed as Δct.

5.1.1 Ranking censored data

Ranking is performed according to the following method: the entire set of observations is ranked from smallest to largest. The smallest observation obtains the rank 1 and the largest observation the highest possible rank. Average ranks are assigned in case of ties. In Table 18 an example is presented on how to transform a variable y with right censoring into its corresponding rank. The variable y turns into the rank transformed variable R(y). In case of censoring the censored variable >y turns into the censored rank transformed variable >R(y).

Table 18: Example of rankig a variable with right censoring
The entire set of observations is ranked from smallest to largest. Consequently, the variable y turns into the rank transformed variable R(y). In case of censoring the censored variable >y turns into the censored rank transformed variable >R(y).

y	R(y)	censoring
1	1	no
1.5	2	no
>2	>3	yes
2.5	4	no

5.1.2 Regression on ranked data: parametric assumptions hold

When data are not normally distributed it can be transformed into normal distribution by e.g. taking the logs. However, transformation into normality does not always work. In scenario a) in Figure 14 log transformation of y results in the normally distributed

variable log(y). In contrast this transformation does not work in scenario b). Here instead, the variable y can be rank transformed into rank(y) resulting in a uniform symmetric distribution.

The reason that regression on ranks is considered to be a non-parametric procedure is based on the fact that linear regression or Tobit regression can be performed on uniform distributed (rank transformed) variables while parametric assumptions still hold. Parametric assumptions of linear regression are, among others things, equal variances of the residuals (homoscedasticity) and normality of the residuals.

The fact that these assumptions are not violated is demonstrated by figure 15. In a) residual diagnostics of a linear regression model applied to a heavily skewed variable reveal that the residuals are not normally distributed. Additionally, heteroscedasticity is present and outliers have a potential impact on the model. Consequently, the assumptions of linear regression are violated. In contrast, when linear regression is applied to the same variable after rank transformation the parametric assumptions hold. In b) the residuals become approximately normally distributed and the variances become homoscedastic. The influence of potential outliers is reduced.

Figure 14: transforming a heavily skewed variable into ranks results in uniform distribution.
Scenario a) shows that log-transformation of a log-normally distributed variable y result in a normally distributed variable log(y). However, some distributions like in scenario b) are not transformable into normality. Here, rank transformation can be conducted resulting in a uniform symmetricly distributed variable rank(y). Data parametric tests can be applied on rank transformed.
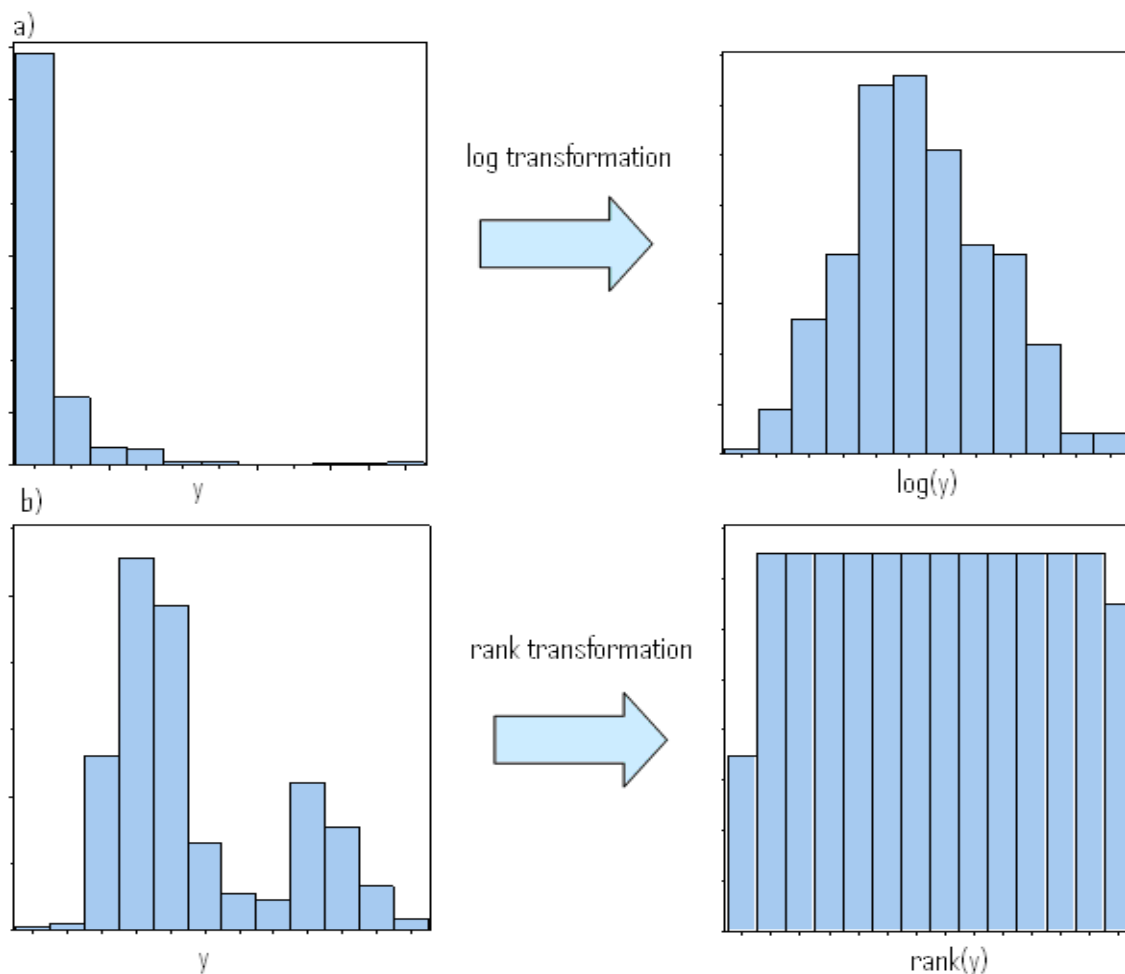
Figure 15: Parametric assumptions hold when regression is performed on rank transformed data

In a) residual diagnostics of a linear regression model applied to a heavily skewed variable reveal that the residuals are not normally distributed and heteroscedasticity is present. Consequently, the assumptions of linear regression are violated. In contrast, when linear regression is applied to the same variable after rank transformation the assumptions hold. In b) the residuals become approximately normally distributed and the variances become homoscedastic
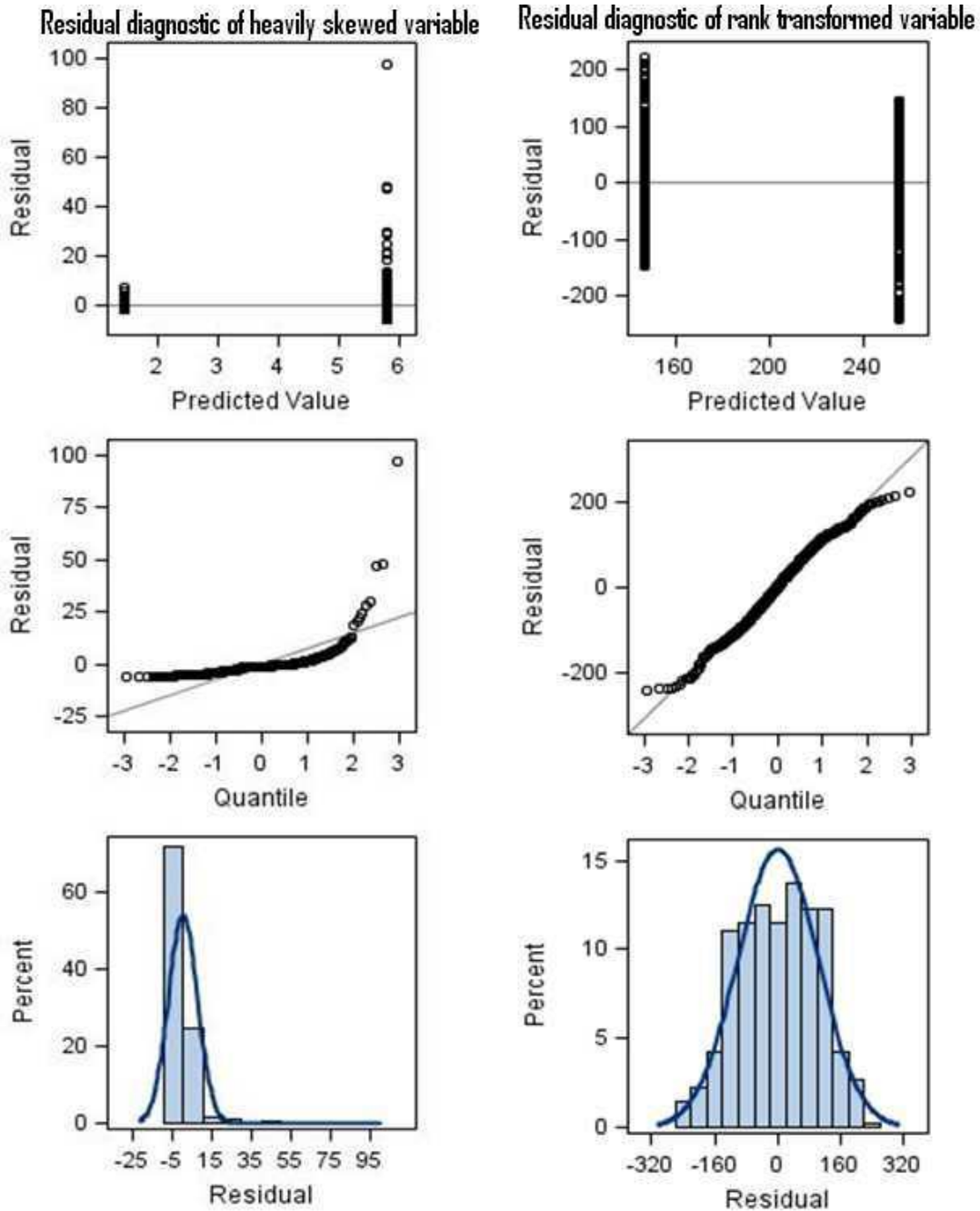
5.2 Performance check of the Tobit model on ranks by simulation study

Simulation studies are generally used to compare different statistical methods and models under realistic data conditions. These realistic conditions are artificially created according to certain criteria which are of interest. In this way, the performance or bias of different competing models can be assessed when "reality" is known. E.g. the validity of a model can be evaluated in the presence of heteroscadasticity. A simulation study was conducted here in order to check the performance of the Tobit regression on ranks in comparison to other competing statistical tests and methods when an increasing amount of censored observations are added to the data. Performance was defined to be the Type-I error rate and power of a statistical method or test.

5.2.1 Type I error

Type I error, also known as „type-I", α error, or "false positive" rate, is the error of rejecting a null hypothesis when it is actually true (Table 19). e.g. a diagnostic test wrongly identifies somebody as having a disease when in fact he/she has no disease. The false positive rate is the proportion of absent events e.g. no diagnosis that yield positive test outcomes, i.e., the probability of a positive test result given an absent event. The false positive rate is also known as the specificity of a test and is related to the level of significance of a statistical test. The significance level is the probability of making the wrong decision when the null hypothesis is true

Table 19: concept of false positive (α error) and false negative (β error) test results

		Reality	
		H0 is true (absent)	*H1 is true (present)*
Test result	*H0 is true*	correct	False negative (β-error)
	H1 is true	False positive (α error)	correct

5.2.2 Type-II error and power

Type II error, also known as a "type-II", a β error, or „false negative" rate, is the error of failing to reject a null hypothesis when it is actually not true (Table 19). It may occur when a diagnostic test wrongly identifies somebody as not having a disease when in fact he/she has the disease. The false negative rate is the proportion of present events that yield negative test outcomes, i.e., the probability of a negative test result given present event.

The power of a test is defined as: $1\text{-}\ type-II\ error$ .It is also known as the sensitivity of a test.

5.2.3 Procedure of the simulation study

10000 data sets were simulated by drawing samples from a population similar to the PAULCHEN and PAULINA studies with respect to sample size and proportion of censored observations.

A two group comparison was conducted with an interval scaled dependent variable. Normal error distribution was assumed as it was desired to include the Tobit regression model as the golden standard in the simulation study. Two scenarios were considered: In the first scenario both groups had the same mean of 20 and a variance of 2 N(20, 2). Scenario 1 was meant to assess the Type-I error rate. In the second scenario there was a mean difference of 1 (mean= 20 and 21 resp.) and a variance of 2 (group 1: N(20,2), group 2: N(21,2)). Scenario 2 was meant to assess the power. In each scenario three different sample sizes per group were considered (25, 50 and 100 resp.). Additionally to different sample sizes, decreasing detection levels were added to the simulation settings starting from DL > 26. The next simulation setting had DL > 25 and so on until DL > 18. Consequently, the proportion of censored observation increased each simulation setting and the proportion of censoring was always higher in the group with the higher mean which was to be expected in the context of gene expression data expressed as Δct.

In the simulation study four statistical models and the standard Wilcoxon rank-sum were used for comparison. As described in previous chapters and according the literature (Helsel 2005) both the models and the standard Wilcoxon rank sum test are

considered to be appropriate methods in this context of right censored data at one single detection level. :

1. Quantile regression (Koenker 2005):   In contrast to the method of least squares as used in linear regression that approximates the conditional mean of the response variable given certain values of the predictor variables, Quantile regression approximates either the median or other quantiles of the response variable. This model was recommended in literature to be a possible option for application to censored data. However, its performance has not been assessed in the context of censored data

2. Logistic regression: logistic regression is a model from the family of the generalized linear models. It is used for the prediction of the probability of an event occurring by fitting data to a logit function. From the probability of occurring an event the odds ratio can be computed. Logistic regression describes the relation between one or more independent variables (e.g., age, sex, etc.) and a binary response variable (e.g. yes/no). The cutoff value for dichotomizing the data in the simulation study was the median. When the proportion of censoring was >= 50 % the detection level was used as the cut-off value.

3. The Wilcoxon rank-sum test: As described in paragraph 4.2 the test is a non-parametric statistical hypothesis test and thus assumes no certain distribution of the data. It can be applied here as the detection occurs at one single level.

4. The Tobit regression: the tobit model as described in paragraph 5.1 was used in the simulation study as the golden standard.

5. The Tobit regression on ranks as described in paragraph 5.1 was the model of main interest in the simulation study.

5.3 Results of performance check

Figure 16 to Figure 18 show the performance of the Tobit regression model on ranks in comparison to four statistical models and tests as described in paragraph 5.2.3 The x-axis indicates the proportion of censored observations, the y-axis on the right-hand side the power and the y-axis on the left-hand side the Type-I error rate.

5.3.1 Performance of Tobit model on ranks with sample size N=100 per group

Figure 16 shows the performance of the Tobit regression on ranks with a sample size N=100 per group. With respect to power the Wilcoxon test, the Tobit model and the Tobit model on ranks perform equally well until app. 70% of censoring. However, the power decreases slightly with increasing an amount of censoring from app. 95% (0% censoring) until app 75% (70% censoring). With a higher amount of censoring (>70% censoring) the Wilcoxon test produces the least false negative rate and performs slightly better than both the Tobit model and the Tobit model on ranks. Between 0% and 50% of censoring logistic regression performs with a constant power of approximately 80%. With more than 50% of censored observations its power is reduced. However, when the proportion of censoring reaches approximately 80% logistic regression has as much power as Tobit regression and Tobit regression on ranks. In contrast to the other tests and models, quantile regression shows very poor performance with respect to power across all proportions of censoring. With more than 50% of censoring quantile regression, when the estimator of interest is based on the 5th quantile (median), is no longer applicable.

The type-I error rate for all statistical tests and models does not depend on the proportion of censoring and lies at around 5%. An exception is the Type-I error rate of quantile regression which increases strongly with more than 30% of censored observations.

Figure 16: performance of Tobit model on rank transformed data in comparison to four statistical tests and models with a sample size of N=100 per group
The x-axis indicates the proportion of censored observations, the y-axis on the right- hand side the power and the y-axis on the left-hand side the Type-I error rate. With respect to power the Wilcoxon test, the Tobit model and the Tobit model on ranks perform equally well until approximately 70% of censoring. The type-I error rate for all statistical tests and models does not depend on the proportion of censoring and lies at around 5% except for quantile regression.

5.3.2 Performance of Tobit model on ranks with sample size N=50 per group

Figure 17 shows the performance of the Tobit regression on ranks with a sample size N=50 per group. With respect to power the Wilcoxon test, the Tobit model and the Tobit model on ranks perform equally well until app. 70% of censoring. However, the power decreases slightly with increasing amount of censoring from app. 90% (0% censoring) until approximately 60% (70% censoring). With a higher amount of censoring (>70% censoring) the Wilcoxon test produces the least false negative rate and performs slightly better than both the Tobit model and the Tobit model on ranks. Between 0% and 50% of censoring logistic regression performs with a constant power of approximately 65%. With more than 50% of censored observations its

power is reduced. However, when the proportion of censoring reaches approximately 80% logistic regression has as much power as Tobit regression and Tobit regression on ranks. In contrast to the other tests and models, quantile regression shows very poor performance with respect to power across all proportions of censoring. With more than 50% of censoring quantile regression, when the estimator of interest is based on the 5th quantile (median), is no longer applicable.

The type-I error rate for all statistical test and models does not depend on the proportion of censoring and lies at around 5%. An exception is the Type-I error rate of quantile regression which increases strongly with more than 30% of censored observations.

Also with a sample size of N=50 the type-I error rate for all statistical tests and models does not depend on the proportion of censoring and lies at around 5%. An exception is again the Type-I error rate of quantile regression which increases strongly with more than 30% of censored observations.

Figure 17: performance of Tobit model on rank transformed data in comparison to four statistical tests and models with a sample size of N=50 per group
The x-axis indicates the proportion of censored observations, the y-axis on the right hand side the power and the y-axis o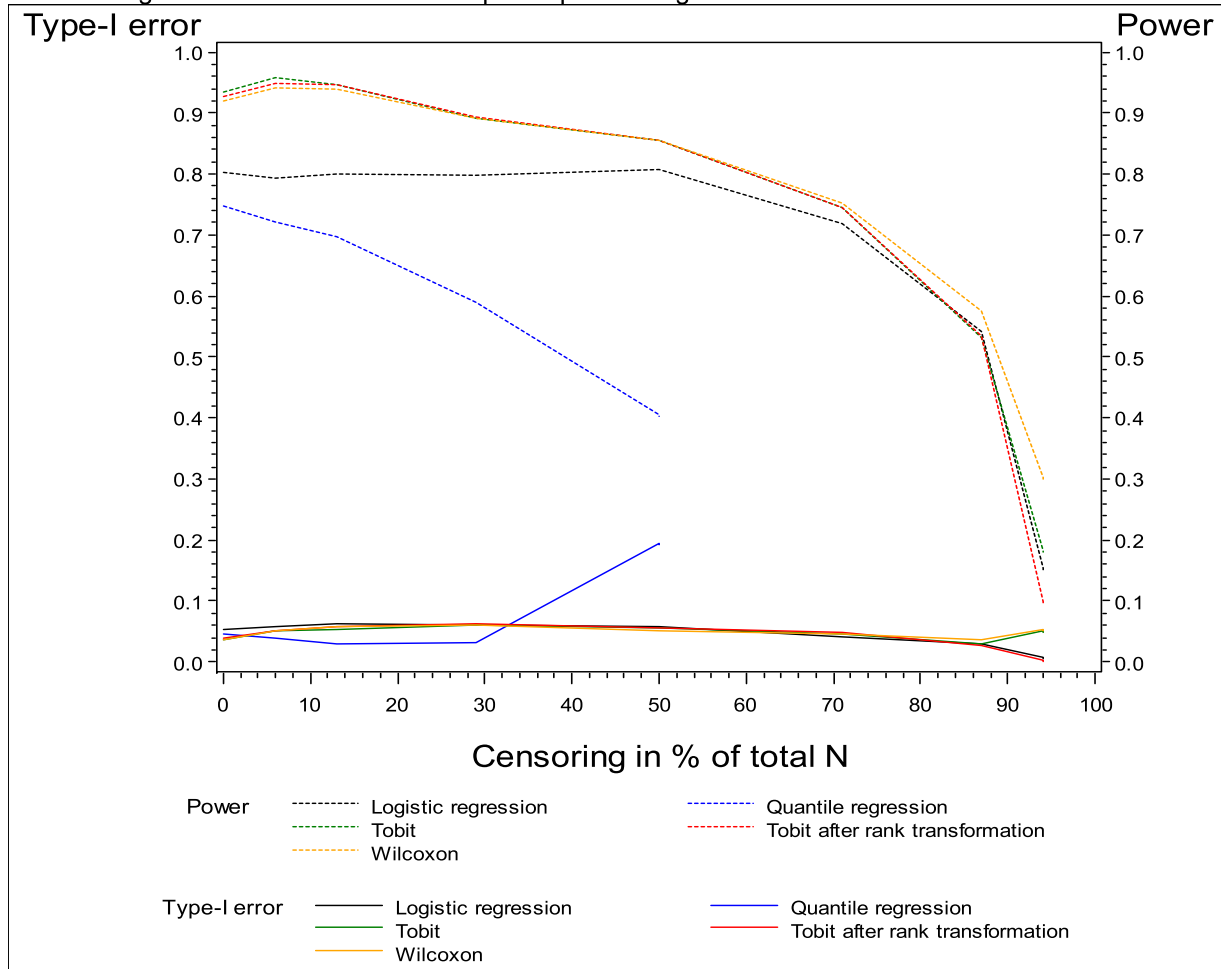n the left hand side the Type-I error rate. With respect to power the Wilcoxon test, the Tobit model and the Tobit model on ranks perform equally well until approximately 70% of censoring. The type-I error rate for all statistical test and models does not depend on the proportion of censoring and lies at around 5% except for quantile regression
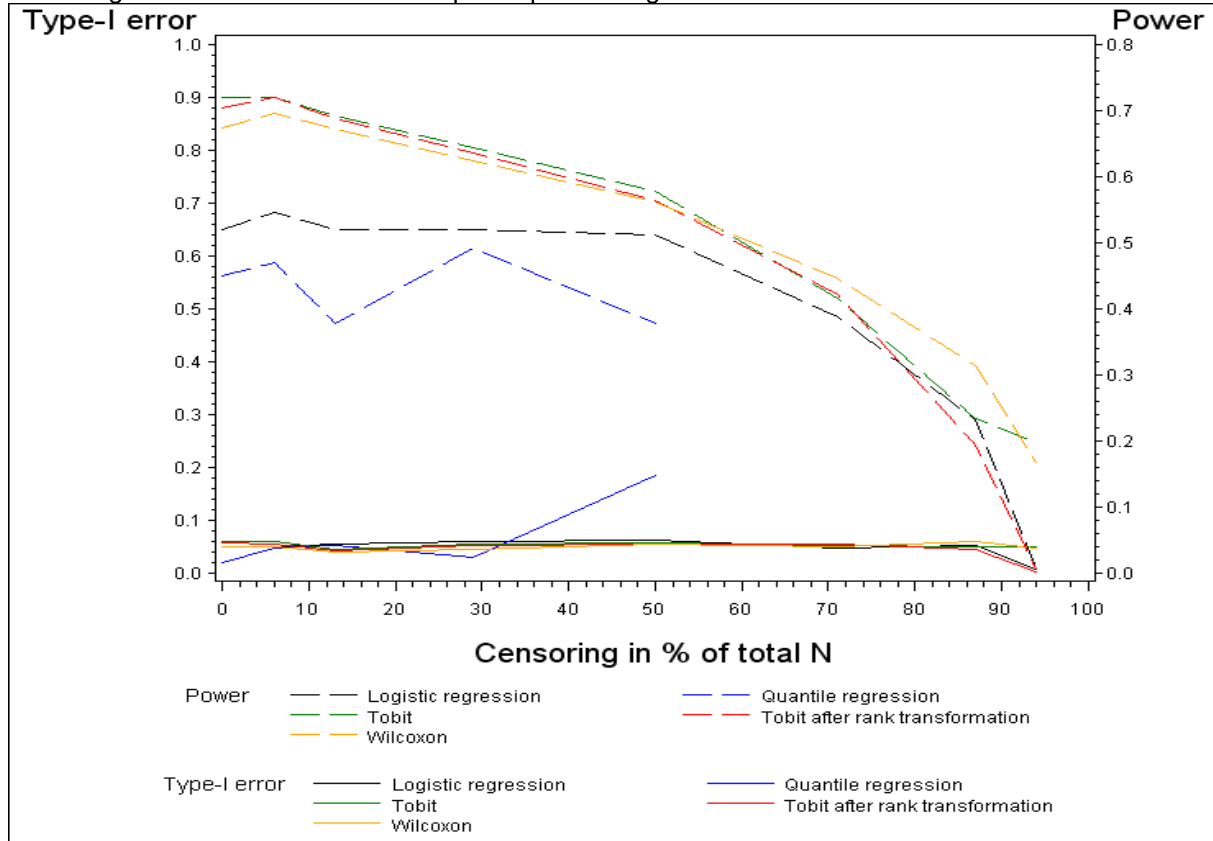
5.3.3 Performance of Tobit model on ranks with sample size N=25 per group

Figure 18 shows the performance of the Tobit regression on ranks with a sample size N=25 per group. Here, again the power of the Tobit model on ranks, the Tobit model and the Wilcoxon test is the greatest, even though strongly reduced compared to the power in simulation settings with a greater sample size. The power slightly decreased with increasing amount of censoring from approximately 42% (0% censoring) until approximately 28% (70% censoring). With increasing proportion of censoring the Tobit regression produces the least false negative rate. However, at the same time point its type-I error rate strongly increases. Logistic regression has constant power of approximately 30% until 50 % of censoring. The power of the quantile regression becomes inconsistent and varies strongly with increasing proportion of censoring.

Also with a sample size of N=25 the type-I error rate for all statistical tests and models does not depend on the proportion of censoring and lies at around 5%. Exceptions are the Type-I error rate of quantile regression which increases with more than 30% of censored observations and the Type-I error rate of Tobit regression which increases with more than 85% of censored observations.

Figure 18: performance of Tobit model on rank transformed data in comparison to four statistical test and models with a sample size of N=25 per group
Tthe power of the Tobit model on ranks, the Tobit model and the Wilcoxon test is the greatest, even though strongly reduced compared to the power in simulation settings with a greater sample size. The type-I error rate for all statistical test and models is not depending on the proportion of censoring. Exceptions are the Type-I error rate of quantile regression which increases with more than 30% of censored observations and the Type-I error rate of Tobit regression which increases with more than 85% of censored observations
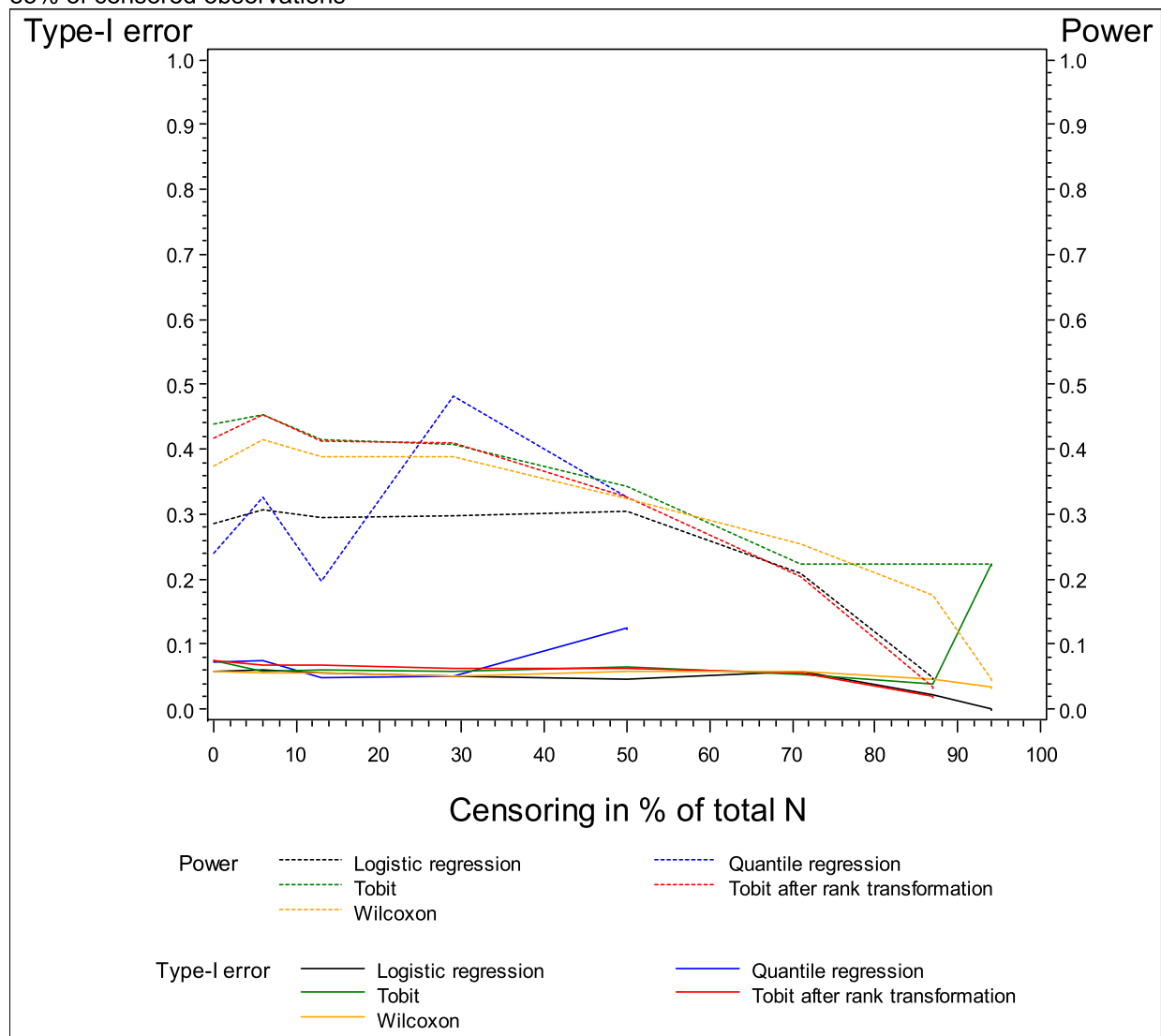
5.4 SAS code for simulation study

Step 1: 10000 datasets are computed

```
data test;
do numsim = 1 to 100;
      do group = 0,1;
              do i = 1 to 10000;
              if group = 0 then y = rand('normal',21,2);
              else y = rand('normal',20,2);
              if y>20 then y=21;
              if y>20 then cutoff=1;
              else cutoff=0;
              output;
              end;
      end;
end;run;
```

Step 2: the data sets are ranked

```
data test;set test;
proc rank data=test out=testrank ties=low ;
      var y;
      by numsim;
      ranks yrank ;
run;
```

Step 3: The statistical models and tests (As an example the Wilcoxon test) are applied to the 10000 datasets

```
ods select none;
ods output WilcoxonTest=wil;
proc npar1way data = test wilcoxon;
by numsim;    class group;
var y;
run;
ods select all;
```

Step 4: the proportion of significant results is assessed and stored

```
data propswil;set wil;
where (name1 = "P2_WIL");
reject = (nvalue1 lt .05);
run;

proc freq data = propswil;
tables reject/binomial(level='1');
output out=wiltype1 binomial;
run;

data wiltype1;set wiltype1;
keep _bin_;
rename _bin_=wilcoxon;
run;
```

Step 5: Repeating this procedure for all tests and models with varying sample sizes

5.5 Application of Tobit regression on ranks to left censored variable

In order to illustrate the application of the Tobit regression on ranks the variable IFN_y with left censored observations from the PAULCHEN study is taken as an example . Again, it is of interest whether IFN_y is dependent on a certain exposure. Before applying the tobit model on ranks the variable IFN_y is ranked from lowest to highest (see Table 20). The variable IFN_y is left censored at one single detection level which is < 0.07. Consequently, all values below 0.07 (here marked as 0.01) represent censored observations and are replaced by average tied ranks. However, it is arbitrary which rank the censored observations are assigned to because the model in SAS receives the information that all rank transformed observations below the detection level of 0.07 are censored. Thus, all ranks below rank 39 are censored.

Table 20: Before applying the tobit model on ranks to the variable IFN_y ranking is performed from lowest to highest
The variable IFN_y is left censored at one single detection level which is < 0.07. 0.07 represents rank 39. Consequently, all ranks below rank 39 are censored.

Concentration in pg/ml (y)	censoring	Exposure (x)	Rank (y)
0.01	yes	no	19.5
0.01	yes	no	19.5
....			
0.07	no	no	39
0.08	no	no	40
0.45	no	no	41,5
0.45	no	no	41,5
0.58	no	no	45,5
0.77	no	yes	45,5
....			
1.82	no	yes	62
1.82	no	yes	62
1.99	no	no	64
2.53	no	no	66.5
....			
3.84	no	no	74.5
4.56	no	no	76

Subsequently, the tobit model is performed on the rank transformed data where the exposure x represents the independent variable and Rank(y) represents the dependent variable. In order to take censoring into account all tied ranks below rank 39 are indicated as censored in the model.

The results are presented in Table 21. The estimate of rank(y) for both the intercept and the exposure and the corresponding confidence intervals are shown. For the purpose of comparison the p-value achieved by the Wilcoxon rank-sum test is also shown. As described before, in situations with left censored data at one single detection level the p-value by the Wilcoxon test is considered to yield the correct result and can thus serve as a golden standard to be compared with. Both p-values depicted in the table are of similar magnitude. This finding confirms the results from the simulation study that demonstrates similar performance of the Tobit regression on ranks and the Wilcoxon test. How to retransform the estimated ranks into the original values is shown in paragraph 5.7.

Table 21: The results of the Tobit model on ranks for the variable IFN_y_m are presented
For the purpose of comparison the p-value achieved by the Wilcoxon rank-sum test is also shown. Both p-values are of same magnitude confirming the results from the simulation study that demonstrates similar performance of the Tobit regression on ranks and the Wilcoxon test.

	Estimate Rank(y)	Confidence interval (lower bound; upper bound)	p-value	
			Tobit on ranks	Wilcoxon test
intercept	41.10	36.75;45.41	0.24	0.25
Exposure(x)	-7.89	-14.33;1.44		

5.6 Application of Tobit regression on ranks to right censored variable

In order to illustrate the application of the Tobit regression on ranks the variable IL17F_PHA_dct with right censored observations from the PAULCHEN study is taken as an example. Again, it is of interest whether IL17F_PHA_dct is dependent on a certain exposure.

Before applying the tobit model on ranks to the variable IL17F_PHA_dct the variable is ranked from lowest to highest (Table 22). The variable IL17F_PHA_dct is right censored at multiple detection levels. All values with the symbol ">" represent the right censored observations and have to be interpreted as values "greater than". Consequently, their corresponding ranks are similarly treated as ranks "greater than".

Table 22: Before applying the tobit model on ranks to the variable IL17F_PHA_dct ranking is performed from lowest to highest

Right censoring at multiple detection levels is present. All values with the symbol ">" represent the right censored observations and have to be interpreted as values "greater than". Consequently, their corresponding ranks are similarly treated as ranks "greater than".

DCT-value (y)	Censoring Indicator	Exposure (x)	Rank (y)
13.10	no	0	1
13.30	no	0	2
>15.10	yes	0	>3
15.45	no	0	4
16.35	no	0	5
16.40	no	1	6
17.65	no	0	7
17.70	no	1	8
>17.80	yes	0	>9
18.00	no	1	10
...			
>21.65	yes	0	>35
>21.80	yes	0	>36
>21.95	yes	0	>37
22.10	no	0	38
22.40	no	0	39
>22.75	yes	0	>40
23.00	no	0	41
>23.15	yes	0	>42
>23.20	yes	0	43
...			
>26.10	yes	0	>59
>26.30	yes	0	>60
>26.40	yes	0	>61
>26.40	yes	0	>62

Subsequently, the tobit model on ranks is performed on the rank transformed data where the exposure x represents the independent variable and Rank (y) represents the dependent variable. In order to take censoring into account all censored ranks that represent the censored observations are indicated as censored by the model.

The results are presented in Table 23. For the purpose of comparison the p-values achieved by the generalized Wilcoxon are also shown. As described before in situations with right censored data at multiple detection levels the p-value by the generalized Wilcoxon test is considered to yield the correct results and can thus serve as a golden standard to be compared with. Again, both p-values depicted in the table are of similar magnitude. This finding underlines the validity of the Tobit regression on ranks. How to retransform the estimated ranks into the original values is shown in paragraph 5.7

Table 23: application of Tobit regression on ranks to right censored variable
For the purpose of comparison the p-value achieved by the generalized Wilcoxon is also shown. Both p-values are of same magnitude underlining the validity of the Tobit model on ranks

	Estimate Rank(y)	Confidence interval	p-value	
			Tobit on ranks	Generalized Wilcoxon
intercept	45.48	35.76; 54.94	0.13	0.14
Exposure (x)	-13.13	-29.64; 3.39		

5.7   Retransformation of the estimated rank into originally measured values

In contrast to the classical linear regression model whose estimates result in estimates of the originally measured values y regression on ranks results in estimates of the ranks $R(y)$ which are not directly interpretable. In order to obtain an interpretable value expressed in originally measured units the estimated parameters of the Tobit regression on ranks can be retransformed.

The interpretable value y' out of the estimated rank $R(y)$ is achieved according to the following formula (Iman):

$$y' = y_1 + (y_2 - y_1) \times \frac{R(y) - R_1}{R_2 - R_1} \qquad (19)$$

Where

$R(y)$ = the estimated regression parameter by the Tobit regression on ranks

R_1 = the rank before $R(y)$

R_2 = the rank after $R(y)$

y_1 = the value of the observation belonging to R_1

y_2 = the value of the observation belonging to R_2

The following example demonstrates that the retransformed estimators of the Tobit regression on ranks validly estimate the true effect.

A simple hypothetical, randomly created group comparison with a mean difference of approximately 2 (mean group 1 = approximately 20, mean group 2 = approximately 22) is used to illustrate (data not shown). The variable is normally distributed. Additionally, the variable contains 25% right censored observations. This example is chosen to allow the classical Tobit regression to be used for comparison with the

Tobit regression on ranks when the "real situation" is known (group difference = 2). For this example the parametric assumptions for the Tobit model hold. Consequently, it can be assumed that the classical Tobit regression estimates are valid.

The estimates of Tobit regression on ranks as obtained by SAS are presented in Table 24. The estimates are not interpretable yet.

Table 24: The Tobit model on ranks is applied to a hypothetical, randomly created example
A group comparison with a mean difference of approximately 2 (mean group 1 = approximately 20, mean group 2 = approximately 22) is used to illustrate this(data not shown). The variable is normally distributed. Additionally, the variable contains 25% right censored observations. Results from the Tobit regression on ranks are not interpretable in original values yet.

	Estimate R(y)	Confidence interval	p-value
intercept	37.85	30.01; 45.69	0.0001
Group (x)	28.07	16.66;39.48	

Based on Table 24 the mean rank with corresponding confidence intervals for group 1 is:

$$R(y) = 37.85 \ \ CI(30.01;45.69)$$

the mean rank with corresponding confidence intervals for group 2:

$$R(y) = 65.92 \ \ CI(54.51;77.33)$$

Applying formula 19 y' for group 1 can be derived:

$$y' = 20.23 + (20.23 - 20.01) \times \frac{37.85 - 37}{38 - 37} = 20.18$$

Analogue, the estimated ranks of the estimate in group 2 and the confidence intervals of the estimates of the ranks in both groups are retransformed. The retransformed estimates are depicted in Table 25: the tableFor comparison the results from the classical Tobit model applied on the same data are presented. The achieved values are of similar magnitude which underlines that the retransformed estimates of the ranks yield valid results.

Table 25: the table compares the retransformed values and the p-values from the Tobit model on ranks to the values achieved from the classical Tobit regression
The achieved values are of similar magnitude which underlines that the retransformed estimates of the ranks yield valid results.

Tobit regression on ranks			Classical Tobit regression		
Estimate group 1	**Estimate group 2**	**p-value**	**Estimate group 1**	**Estimate group 2**	**p-value**
20.18 (19.70;20.65)	21.96 (21.22;23.15)	<0.0001	19.88 (19.23;20.55)	22.21 (21.30;23.22)	<0.0001

5.8 SAS Code for Tobit model on ranks

Step 1: the variable is ranked

```
data test;set test;
proc rank data=test out=testrank ties=low ;
     var y;
     ranks yrank ;
run;
```

Step 2: Tobit regression is applied to the rank transformed variable

```
ods select none;
ods output ParameterEstimates=tobit;
     proc qlim data=testrank;
     model yrank=group;
   endogenous yrank~ censored(ub=20);
run;
ods select all;
```

5.9 Tobit regression on ranks estimates the median of the original measured values

In contrast to the classical regression model whose estimates represent the mean estimates from regression models on rank transformed data represent the mean rank which corresponds to the median of the originally measured values. To illustrate this fact a simple hypothetical example with highly skewed data is considered: Table 26 presents hypothetical data in originally measured value and their corresponding ranks. The mean and the median of the originally measured values are 20276.25 and 700 respectively. The mean of the rank transformed values is 4.5. Retransforming the mean rank according to formula 19 into the originally measured values:

$$y' = 400 + (100 - 400) \times \frac{4.5 - 4}{5 - 4} = 700$$

Table 26: Tobit regression on ranks estimates the median of the original measured values

Originally measured values	Ranks of originally measured values
1	1
10	2
100	3
400	4
1000	5
10000	6
50000	7
100000	8
Mean: 20276.25	**Mean rank:** 4.5
Median: 700	**Retransformed in original values**: 700

5.10 Application of the retransformation procedure to examples from PAULCHEN

In this paragraph it is demonstrated how the procedure of retransformation is applied to two examples from the PAULCHEN study. Both an example with right and left censored observations is presented. Therefore, the example with left censored data is variable IFN_g_m from paragraph 5.5 and the example with right censored data is variable IL-17F_pha from paragraph 5.6.

Table 27 presents the results of the Tobit regression on ranks for the left censored variable IFN_g_m. The estimates from the Tobit regression on ranks are already retransformed into the original measured values according to formula 19. The table additionally depicts the results of the standard method as described in chapter 3.1 and 4.1 which are considered to be appropriate here. The magnitude of both the estimates (median and confidence interval) and the p-values of both methods are similar. This finding suggests that the results of the Tobit regression on ranks are unbiased when applied on left censored data.

Table 27: results of the Tobit regression on ranks (already retransformed) for the left censored variable IFN_g_m are shown
The results are unbiased as they are of same magnitude as the results from the standard method.

Tobit regression on ranks			standard method		
Estimate group 1	**Estimate group 2**	**p-value**	**Estimate group 1**	**Estimate group 2**	**p-value**
0.05 (0.01;0.70)	0.09 (0.03;0.93)	0.24	0.01 (0.01;0.77)	0.08 (0.01;1.01)	0.25

Table 28 presents the results of the Tobit regression on ranks for the variable IL-17F_pha. The estimates from the Tobit regression on ranks are again already retransformed according to formula 19. The table additionally depicts the results from the Kaplan-Meier-method as described in chapter 3.2 and 4.2. Even though the

results differ slightly the magnitude of both the estimates (median and confidence interval) and the p-values of both methods are similar. This finding suggests that the results of the Tobit regression on ranks are unbiased when applied on right censored data.

Table 28: results from the Tobit regression on ranks (already retransformed) for the right censored variable IL-17F-PHA are shown
The results are unbiased as they are of same magnitude as the results from the Kaplan-Meier method.

Tobit regression on ranks			Kaplan-Meier method		
Estimate group 1	**Estimate group 2**	**pvalue**	**Estimate group 1**	**Estimate group 2**	**p-value**
23.23 (21.20; n.a)	21.28 (19.22;24.10)	0.13	23.00 (20.75; n.a.)	21.02 (18.75-23.80)	0.14

5.11   Conclusion of chapter

This chapter introduced the Tobit model on rank transformed data which may be applied to censored data that violate parametric assumptions. This model makes use of both the classical Tobit regression and the idea of regression on rank transformed data and may be used to adjust for possible confounding and interaction effects. Additionally, its application was shown by means of two examples with both left and right censored observations taken from the PAULCHEN study. Its validity could be demonstrated by a simulation study. Furthermore, it was explained how to retransform the estimates from the Tobit regression on ranks into original values. The retransformed estimates from the Tobit regression on ranks always result in the median of the original measured values.

6 Discussion

Motivated by the PAULCHEN and PAULINA data sets the main goal of the dissertation was to highlight the characteristics and complexity of immunological data and present novel statistical approaches that take these into account. The complexity arises from skewed data distributions which are not transformable into appropriate distributions, the fact that censoring may hamper the data analysis and the presence of complex and multiple relationships between immunological and epidemiological parameters.

The majority of the data in both PAULCHEN and PAULINA is not normally distributed and cannot be transformed into an appropriate distribution. This holds both for data measured by Luminex technology at protein level and data measured by real-time RT-PCR at mRNA level. This is depicted in Figure 1 which compares the distribution of two variables measured by both Luminex and real-time RT-PCR to a normal and lognormal distribution. Consequently, in order to apply parametric tests and models the corresponding assumptions such as an appropriate underlying distribution may be violated.

The concept of censoring is a great issue in both the PAULINA and the PAULCHEN data sets. A large number of variables assessing immunological parameters contain a non-negligible proportion of censored observations. These variables are both derived by real-time RT-PCR at mRNA level and luminex technology. When protein expression is measured by Luminex technology left censoring at a single censoring level can occur as the measured concentrations fall below the detection threshold. Here, censoring is obvious. In contrast, right censoring at multiple censoring levels may occur when genes or cytokines are measured by real-time RT-PCR at mRNA level and are expressed in relation to a given housekeeping gene by the Δct formula. Here, censoring is much less obvious.

Furthermore, the complex and multiple relationships that are often present between immunological and epidemiological parameters must not be neglected. Immunologists are generally interested in many different outcomes depending on various exposures when other intervening immunological parameters may be present. Thus, data sets usually contain a large number of interacting variables that have to be taken into account by appropriate statistical models that allow adjusting

for potential confounding and interaction effects. However, such models need to be both non-parametric and take censoring into account.

Having highlighted the characteristics of both PAULINA and PAULCHEN data, statistical methods were presented that take these into consideration.

In order to compute summary statistics like the median and its corresponding confidence intervals for censored data at multiple detection levels it was shown how to apply the Kaplan-Meier-method and B-C ratio test using an example with gene expression data (shown as Δct). Furthermore, it was illustrated that this method is preferable to the standard method since the Kaplan-Meier method takes censoring into account. It could be demonstrated that summary statistics differed according to varying proportions of censored observations when analyzed by the Kaplan-Meier method. In contrast, summary statistics remained unaffected according to varying proportions of censored observations when analyzed by standard methods. This fact underlines the sensitivity of the Kaplan-Meier method to capture the information of censoring in the data and its appropriateness as a descriptive method. The Kaplan-Meier method may be applied regardless of the underlying distribution. The major drawback of this method is that it is only applicable for less than 50% of censoring. For data with at least 50% censoring literature recommends other methods and ways of action. These are based on maximum likelihood estimation, multiple imputation or regression on order statistics (Helsel 2005). However, applying these methods require assumptions for an underlying distribution and a minimum sample size. Ignoring these may result in biased results. Consequently, it remains questionable whether cytokine and gene expression data with more than 50% of censoring and small sample sizes should be interpreted at all.

Not only summary statistics of censored data at multiple detection levels need to be adapted but also testing on group differences should be computed by appropriate methods in the setting of cytokine and gene expression data. In order to correctly calculate p-values it is also essential that a statistical test takes the amount of censoring in each group into account. For statistical testing on differences between two or more groups the use of the generalized Wilcoxon test is preferable to the classical Wilcoxon test. The achieved p-value by the generalized Wilcoxon test in the example with right censored expression data highly depended on the proportion of censoring in each group. In contrast the p-values from the standard Wilcoxon test did

not capture the information of the censored proportion and were highly biased compared to the correct analysis by the generalized Wilcoxon test. Consequently, right censored gene expression data with multiple detection levels should be calculated by the generalized Wilcoxon test.

In contrast, when censored cytokine data measured by Luminex are censored at one detection level standard methods for both summary statistics and statistical testing on differences between two or more groups may be applied without loss of information.

In order to take the often complex and multiple relationships among immunological parameters and between immunological and other characteristics (e.g. environment) into consideration appropriate statistical models are of high importance (Genser, Cooper et al. 2007). The recommendations given in literature are often not applicable: substitution of the values above or below the detection level (Buckley, Liddle et al. 1997) may lead to strongly biased results. Classical Tobit regression is prone to violating parametric assumptions (Arabmazar and Schmidt 1982; Austin, Escobar et al. 2000) and multiple Imputation (Lubin, Colt et al. 2004; Uh, Hartgers et al. 2008), which is a valid alternative, may be time consuming and not supported by all statistical packages. Furthermore, multiple imputation requires the assumption of a certain underlying distribution.

Another method mentioned in literature (Helsel 2005) as a suitable regression model for censored data is the semiparametric cox regression which is mainly applied in classical survival analysis. The outcome of cox regression is the hazard (the probability of an event occurring) given a certain predictor. In classical survival analysis the predictor is time, transferred into immunology the predictor would be e.g. concentration of a certain cytokine. Even though cox regression has no assumption of the underlying distribution of the data, it has not been translated well into the setting of immunological studies. This is potentially due to two reasons: Firstly, the concentration of cytokine may not be estimated by the model as in Tobit regression and secondly, the cox regression model is based on the assumption of proportionality of hazard ratios. This means e.g. that the survival functions of two groups of interest may not cross each other. In the context of cytokine measurements this condition is hardly to be fulfilled. An example for crossing survival functions can be seen in Figure 9. Consequently, cox regression is not a recommendable alternative for cytokine measurements.

As a possible solution for the analysis of non-normal censored cytokine data the Tobit regression on rank transformed data was presented. The model takes censoring into account, does not require parametric assumptions, allows adjusting for covariates and potential confounders and is available in common statistical packages. The performance of the Tobit model on ranks with respect to power and type I error rate was assessed in a simulation study. The performance was comparable to both the classical Tobit regression and the Wilcoxon test over different sample sizes and varying amounts of censoring. In comparison to the logistic regression and the quantile regression the Tobit model on ranks performed much better. The fact that logistic regression has less power than other models is due to dichotomizing, resulting in loss of information (Uh, Hartgers et al. 2008). Quantile regression has even less power and can only be applied to up to 50% of censoring. Thus, both logistic regression and quantile regression cannot be recommended. Additionally, the good performance of the Tobit model on ranks was confirmed by applying it to immunological examples. The resulting p-values were of similar magnitude to those obtained from the generalized Wilcoxon test.

The simulation study was not conducted with skewed data, different error distributions or great outliers due to two reasons. It was aimed to include the classical Tobit model as the golden standard. Applying this model requires strong parametric assumptions as (log)normality and homoscedasticity of the data (Arabmazar and Schmidt 1982; Austin, Escobar et al. 2000). Additionally, rank regression is considered to be a distribution-free analysis (Conover 1980; Conover and Iman 1981). As shown before when skewed data with heteroscedastic residuals are rank transformed parametric assumptions still hold. Therefore, there is no need to check the performance of regression models on rank transformed data when the parametric assumptions of the non-transformed variable do not hold.

The obtained estimate from the Tobit on ranks model is the estimate of the rank transformed variable instead of the estimate of the originally observed variable. However, the estimate of the rank transformed variable can be easily retransformed into the original observed values (Iman and Conover 1979) resulting in the median of the originally measured variable.

Yet, it remains unclear as to what extent the Tobit on ranks estimates the true population parameter when retransformed into its original values. Therefore, a simulation study evaluating the root mean square error (RMSE) of the Tobit on ranks

estimator might be performed. RMSE is an estimate of both bias and consistency of a parameter of interest. However, as the Type-I error rate and power of the Tobit model on ranks is comparable to classical Tobit regression, likewise the estimated RMSE of Tobit on ranks is expected to be comparable to the RMSE of classical Tobit model. In other words, a biased estimator would have resulted in reduced performance revealed by the simulation study.

A remaining aspect open for discussion is about the course of action when the explaining variable is also censored. This scenario is imaginable when the concentration of a measured cytokine that contains censored measurements is aimed to be predicted by the concentration of another cytokine that is also constrained by censored observations. Literature gives recommendations about the course of action when the independent variable is censored (Austin and Hoch 2004; Schisterman, Vexler et al. 2006). However, no substantial recommendations are given when both explaining and outcome variable are hampered by censored observations.

One approach to adjust one censored variable by another censored variable is to take a ratio or the difference out of both variables as is done by the formula for the fold difference. The fold difference is defined as $fd = 2^{\Delta ct(unstim)-\Delta ct(stim)}$ (Livak and Schmittgen 2001). In this formula the Δct of the stimulated cytokine is subtracted from the Δct of the unstimulated cytokine. In other words the Δct of the stimulated cytokine is adjusted for the Δct of the unstimulated cytokine. In the case of right censoring of the unstimulated cytokine, the resulting fold difference also becomes right censored. In the case of right censoring of the stimulated cytokine the resulting fold difference becomes left censored. In case of right censoring of both Δcts (unstimulated and stimulated) the resulting fold difference can take any value between zero and positive infinity resulting in a right censored observation with a detection level of zero. However, whether this way of dealing with the data leads to unbiased results requires further research.

In summary, it is highly important to treat gene expression data (expressed as Δct values) as censored when the cycle threshold of a real-time RT-PCR measurement exceeds a certain value and thus, is not quantifiable anymore. Applying appropriate statistical methods is crucial. For summary statistics and comparison of simple group differences the Kaplan-Meier-method and the generalized Wilcoxon test respectively

are to be recommended. For multivariate comparison that allows adjusting for potential confounding and interaction effects it is suggested to apply the Tobit regression on ranks. Further advantages in addition to adjusting for potential confounding and interaction effects are that Tobit regression on ranks assumes no certain distribution of the data and is available in standard statistical packages like SAS.

Key points of thesis

- Data often contain left and right censored observations with both single and multiple detection levels.

- The assumptions of an underlying distribution applying parametric tests are mostly violated in the setting of immunological measurements

- Summary statistics for left censored data with one detection level such as Luminex data may be computed by standard methods without loss of information

- Summary statistics for right censored data with multiple detection levels should be computed by methods that particularly take censoring into account such as the Kaplan-Meier method in order to avoid biased results

- Statistical testing on differences between two or more groups in the presence of left censored data with one detection level such as Luminex data standard methods such as the Wilcoxon rank sum test may be applied

- Statistical testing on differences between two or more groups in the presence of right censored data with multiple detection levels should be computed by methods that particularly take censoring into account such as the generalized Wilcoxon test

- Tobit regression on ranks has been shown to be a valid procedure for the application to non-normal censored data

- Tobit regression on ranks can be used to adjust for possible confounding and interaction effects

7 Summary

For several immune-mediated diseases immunological analysis becomes more detailed and complex in the future with large datasets in which cytokine and gene expression data play a major role. These data have certain characteristics that require sophisticated statistical analysis such as strategies for skewed distributions and censoring. Additionally, complex and multiple immunological relationships need to be adjusted for potential confounding and interaction effects.

Consequently, the main goal of the dissertation was to highlight the characteristics and complexity of immunological data by means of two birth cohort studies (PAULCHEN und PAULINA) and present novel statistical approaches in order to take these into account. PAULCHEN und PAULINA were conducted in order to contribute to explaining mechanisms on neonatal immune responses in association with epidemiological data such as life style factors and atopic history of the parents.

It could be shown that the majority of both the PAULCHEN and PAULINA data is neither normally nor log-normally distributed and cannot be transformed into an appropriate distribution. This holds for data both measured by Luminex technology at protein level and data measured by real-time RT-PCR at mRNA level.

Furthermore, a high number of variables from both the PAULCHEN and PAULINA data assessing immunological parameters contain a non-negligible proportion of censored observations. Out of 148 variables in the PAULINA data 62 variables contain censored observations corresponding to a proportion of 42%. Out of 162 variables 92 in the PAULCHEN data contain censored observations corresponding to a proportion of 57%. Censoring occurs in the context of measurements derived by both real-time RT-PCR at mRNA level and Luminex technology at protein level.

Summary statistics without parametric assumptions for left censored data with one detection level such as Luminex data can be computed by standard methods without loss of information. In contrast, summary statistics without parametric assumptions for right censored data with multiple detection levels should be computed by methods that particularly capture the information of the censored proportion in the data such as the Kaplan-Meier method in order to avoid biased results.

In analogy, non-parametric statistical testing on differences between two or more groups standard methods such as the Wilcoxon rank sum test can be applied without loss of information when left censored data with one detection level (e.g. in the

setting of Luminex measurements) are present. However, non-parametric statistical testing of differences between two or more groups should be computed by methods that particularly capture the information of the censored proportion in the data such as the generalized Wilcoxon test when right censored data with multiple detection levels (e.g. in the context of real time RT-PCR measurements) are present.

In order to take the complex and multiple relationships between immunological and epidemiological parameters into account the Tobit regression on ranks was introduced. The non-parametric Tobit regression on ranks can be used for left and right censored data with multiple detection thresholds aiming to adjust for potential confounding and interaction effects. Its performance was evaluated in a simulation study: Both type-I error rate and power are comparable to the classical Tobit regression and the Wilcoxon test over different sample sizes and varying amounts of censoring. In contrast to classical linear or Tobit regression that estimate the mean of the dependent variable Tobit regression on ranks estimates the median of the dependent variable after retransforming into originally measured values.

As a conclusion, it is crucial to apply appropriate statistical methods to complex immunological data that are skewed and contain censored measurements. Otherwise, biased results may be the consequence.

8 Literature

Abbas, A. K. and A. H. Lichtman (2009). Basic Immunology: Functions and Disorders of the Immune System, Saunders.

Accordini, S. (2008). "The socio-economic burden of asthma is substantial in Europe." Allergy **63**: 116-124.

Adler, A., I. Tager, et al. (2005). "Decreased prevalence of asthma among farm-reared children compared with those who are rural but not farm-reared." Journal of Allergy and Clinical Immunology **115**(1): 67-73.

Alfvén, T., C. Braun-Fahrländer, et al. (2006). "Allergic diseases and atopic sensitization in children related to farming and anthroposophic lifestyle – the PARSIFAL study." Allergy **61**(4): 414-421.

Arabmazar, A. and P. Schmidt (1982). "An Investigation of the Robustness of the Tobit Estimator to Non-Normality." Econometrica **50**(4): 1055-1063.

Asher, M. I., H. R. Anderson, et al. (1998). "Worldwide variations in the prevalence of asthma symptoms: the International Study of Asthma and Allergies in Childhood (ISAAC)." European Respiratory Journal **12**(2): 315-335.

Austin, P., M. Escobar, et al. (2000). "The use of the Tobit model for analyzing measures of health status." Quality of Life Research **9**(8): 901-910.

Austin, P. C. and J. S. Hoch (2004). "Estimating linear regression models in the presence of a censored independent variable." Statistics in Medicine **23**(3): 411-429.

Beasley, R. (1998). "Worldwide variation in prevalence of symptoms of asthma, allergic rhinoconjunctivitis, and atopic eczema: ISAAC." The Lancet **351**(9111): 1225-1232.

Bjorksten, B., D. Dumitrascu, et al. (1998). "Prevalence of childhood asthma, rhinitis and eczema in Scandinavia and Eastern Europe." European Respiratory Journal **12**(2): 432-437.

BrÅBÄCk, L., A. Breborowicz, et al. (1994). "Atopic sensitization and respiratory symptoms among Polish and Swedish school children." Clinical & Experimental Allergy **24**(9): 826-835.

BrÅbÃck, L., A. Breborowicz, et al. (1995). "Risk factors for respiratory symptoms and atopic sensitisation in the Baltic area." Archives of Disease in Childhood **72**(6): 487-493.

Braun-Fahrländer, Gassner, et al. (1999). "Prevalence of hay fever and allergic sensitization in farmer's children and their peers living in the same rural community." Clinical & Experimental Allergy **29**(1): 28-34.

Brookmeyer, R. and J. Crowley (1982). "A Confidence Interval for the Median Survival Time." Biometrics **38**(1): 29-41.

Buckley, T. J., J. Liddle, et al. (1997). "Environmental and biomarker measurements in nine homes in the Lower Rio Grande Valley: Multimedia results for Pesticides, metals, PAHs, and VOCs." Environment International **23**(5): 705-732.

Campbell, M. J. and M. J. Gardner (1988). "Calculating Confidence Intervals For Some Non-Parametric Analyses." British Medical Journal (Clinical Research Edition) **296**(6634): 1454-1456.

Chen, X., M. BÃ¤umel, et al. (2007). "Interaction of TNF with TNF Receptor Type 2 Promotes Expansion and Function of Mouse CD4+CD25+ T Regulatory Cells." The Journal of Immunology **179**(1): 154-161.

Chrischilles, E., R. Ahrens, et al. (2004). "Asthma prevalence and morbidity among rural Iowa schoolchildren." Journal of Allergy and Clinical Immunology **113**(1): 66-71.

Conover, W. J. (1980). Practical Nonparametric Statistics New York, John Wiley and Sons.

Conover, W. J. and R. L. Iman (1981). "Rank Transformations as a Bridge Between Parametric and Nonparametric Statistics." The American Statistician 35(3): 124-129.

Dimich-Ward, H., Y. Chow, et al. (2006). "Contact with livestock – a protective effect against allergies and asthma?" Clinical & Experimental Allergy 36(9): 1122-1129.

Douwes, J., S. Cheng, et al. (2008). "Farm exposure in utero may protect against asthma, hay fever and eczema." European Respiratory Journal 32(3): 603-611.

Douwes, J., N. Travier, et al. (2007). "Lifelong farm exposure may strongly reduce the risk of asthma in adults." Allergy 62(10): 1158-1165.

Downs, S. H., G. B. Marks, et al. (2001). "Having lived on a farm and protection against allergic diseases in Australia." Clinical & Experimental Allergy 31(4): 570-575.

Ege, M., F. Remo, et al. (2007). "Not all farming environments protect against the development of asthma and wheeze in children." The Journal of allergy and clinical immunology 119(5): 1140-1147.

Ege, M. J., C. Bieli, et al. (2006). "Prenatal farm exposure is related to the expression of receptors of the innate immunity and to atopic sensitization in school-age children." Journal of Allergy and Clinical Immunology 117(4): 817-823.

Ege, M. J., I. Herzum, et al. (2008). "Specific IgE to allergens in cord blood is associated with maternal immunity to Toxoplasma gondii and rubella virus." Allergy 63(11): 1505-1511.

Ehrenstein, V., V. Mutius, et al. (2000). "Reduced risk of hay fever and asthma among children of farmers." Clinical & Experimental Allergy 30(2): 187-193.

Eriksson, J., L. Ekerljung, et al. "Growing up on a farm leads to lifelong protection against allergic rhinitis." Allergy 65(11): 1397-1403.

Erkki, V., P. y. Tuula, et al. (2002). "Allergic diseases, skin prick test responses, and IgE levels in North Karelia, Finland, and the Republic of Karelia, Russia." The Journal of allergy and clinical immunology 109(4): 643-648.

Ernst, P. and Y. Cormier (2000). Relative Scarcity of Asthma and Atopy among Rural Adolescents Raised on a Farm. 161: 1563-1566.

Genser, B., P. Cooper, et al. (2007). A guide to modern statistical analysis of immunological data. 8: 27.

Greenwood, M. (1926). The Natural Duration of Cancer. London, His Majesty's Stationary Office.

Heid, C. A., J. Stevens, et al. (1996). Real time quantitative PCR. 6: 986-994.

Helsel, D. R. (2005). Nondetects and Data Analysis. Hoboken, John Wiley & Sons.

Hobbs, K. E., D. C. G. Muir, et al. (2003). "Levels and patterns of persistent organochlorines in minke whale (Balaenoptera acutorostrata) stocks from the North Atlantic and European Arctic." Environmental Pollution 121(2): 239-252.

Hoppin, J. A., D. M. Umbach, et al. (2008). Pesticides and Atopic and Nonatopic Asthma among Farm Women in the Agricultural Health Study. 177: 11-18.

Horak, F., M. Studnicka, et al. (2002). "Parental farming protects children against atopy: longitudinal evidence involving skin prick tests." Clinical & Experimental Allergy 32(8): 1155-1159.

Iman, R. L. and W. J. Conover (1979). "The Use of the Rank Transform in Regression." Technometrics 21(4): 499-509.

Johansson, S. G. O., T. Bieber, et al. (2004). "Revised nomenclature for allergy for global use: Report of the Nomenclature Review Committee of the World Allergy Organization, October 2003." Journal of Allergy and Clinical Immunology 113(5): 832-836.

JõTgi, R., C. Janson, et al. (1996). "The prevalence of asthmatic respiratory symptoms among adults in Estonian and Swedish university cities." Allergy 51(5): 331-336.

Kaplan, E. L. and P. Meier (1958). "Nonparametric Estimation from Incomplete Observations." Journal of the American Statistical Association 53(282): 457-481.

Karmaus, W. and C. Botezan (2002). "Does a higher number of siblings protect against the development of allergy and asthma? A review." Journal of Epidemiology and Community Health **56**(3): 209-217.

Kiechl-Kohlendorfer, U., E. Horak, et al. (2007). "Neonatal characteristics and risk of atopic asthma in schoolchildren: results from a large prospective birth-cohort study." Acta Pædiatrica **96**(11): 1606-1610.

Kilpeläinen, Terho, et al. (2000). "Farm environment in childhood prevents the development of allergies." Clinical & Experimental Allergy **30**(2): 201-208.

Kilpeläinen, M., E. O. Terho, et al. (2002). "Childhood farm environment and asthma and sensitization in young adulthood." Allergy **57**(12): 1130-1135.

Klein, J. P. and M. L. Moeschberger (2003). Survivasl Analysis: Techniques for censored and truncated data. London.

Klintberg, B., N. Berglund, et al. (2001). "Fewer allergic respiratory disorders among farmers' children in a closed birth cohort from Sweden." European Respiratory Journal **17**(6): 1151-1157.

Koskela, H. O., K. K. Happonen, et al. (2005). "Effect of farming environment on sensitisation to allergens continues after childhood." Occupational and Environmental Medicine **62**(9): 607-611.

Lauener, R. P., T. Birchler, et al. (2002). "Expression of CD14 and Toll-like receptor 2 in farmers' and nonfarmers' children." The Lancet **360**(9331): 465-466.

Lauener, R. P., S. M. Goyert, et al. (1990). "Interleukin 4 down-regulates the expression of CD14 in normal human monocytes." European Journal of Immunology **20**(11): 2375-2381.

Livak, K. J. and T. D. Schmittgen (2001). "Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2-[Delta][Delta]CT Method." Methods **25**(4): 402-408.

Livak, K. J. and T. D. Schmittgen (2001). "Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2â˘ˆÎ”Î”CT Method." Methods **25**(4): 402-408.

Lubin, J. H., J. S. Colt, et al. (2004). "Epidemiologic Evaluation of Measurement Data in the Presence of Detection Limits." Environ Health Perspect **112**(17).

Midodzi, W. K., B. H. Rowe, et al. (2007). "Reduced risk of physician-diagnosed asthma among children dwelling in a farming environment." Respirology **12**(5): 692-699.

Muzio, M., D. Bosisio, et al. (2000). "Differential Expression and Regulation of Toll-Like Receptors (TLR) in Human Leukocytes: Selective Expression of TLR3 in Dendritic Cells." The Journal of Immunology **164**(11): 5998-6004.

Nicolai, T., B. Bellach, et al. (1997). "Increased prevalence of sensitization against aeroallergens in adults in West compared with East Germany." Clinical & Experimental Allergy **27**(8): 886-892.

Perkin, M. R. (2007). "Unpasteurized milk: health or hazard?" Clinical & Experimental Allergy **37**(5): 627-630.

Perkin, M. R. and D. P. Strachan (2006). "Which aspects of the farming lifestyle explain the inverse association with childhood allergy?" Journal of Allergy and Clinical Immunology **117**(6): 1374-1381.

Peto, R. and J. Peto (1972). "Asymptotically Efficient Rank Invariant Test Procedures." Journal of the Royal Statistical Society. Series A (General) **135**(2): 185-207.

Peto, R., M. C. Pike, et al. (1976). "Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. Introduction and design." British journal of cancer **34**(6): 585-612.

Pfefferle, B. c. Gisela, et al. "Cord blood cytokines are modulated by maternal farming activities and consumption of farm dairy products during pregnancy: The PASTURE Study." The Journal of allergy and clinical immunology **125**(1): 108-115.e3.

Pfefferle, B. c. Gisela, et al. (2010). "Cord blood cytokines are modulated by maternal farming activities and consumption of farm dairy products during pregnancy: The PASTURE Study." The Journal of allergy and clinical immunology **125**(1): 108-115.e3.

Pfefferle, P., Ina, , S. Serdar, et al. (2008). "Cord blood allergen-specific IgE is associated with reduced IFN-Î³ production by cord blood cells: The Protection against Allergyâ€"Study in Rural Environments (PASTURE) study." The Journal of allergy and clinical immunology **122**(4): 711-716.

Prentice, R. L. (1978). Linear rank tests with right censored data. **65:** 167-179.

Radon, K., V. Ehrenstein, et al. (2004). "Childhood visits to animal buildings and atopic diseases in adulthood: An age-dependent relationship." American Journal of Industrial Medicine **46**(4): 349-356.

Radon, K., A. Schulze, et al. (2006). "Inverse association between farm animal contact and respiratory allergies in adulthood: protection, underreporting or selection?" Allergy **61**(4): 443-446.

Remes, S. T., K. Iivanainen, et al. (2003). "Which factors explain the lower prevalence of atopy amongst farmers' children?" Clinical & Experimental Allergy **33**(4): 427-434.

Riedler, Eder, et al. (2000). "Austrian children living on a farm have less hay fever, asthma and allergic sensitization." Clinical & Experimental Allergy **30**(2): 194-200.

Riedler, J., C. Braun-Fahrländer, et al. (2001). "Exposure to farming in early life and development of asthma and allergy: a cross-sectional survey." The Lancet **358**(9288): 1129-1133.

Schaub, B., J. Liu, et al. (2009). "Maternal farm exposure modulates neonatal immune mechanisms through regulatory T cells." The Journal of allergy and clinical immunology **123**(4): 774-782.e5.

Schaub, B., J. Liu, et al. (2009). "Maternal farm exposure modulates neonatal immune mechanisms through regulatory T cells." Journal of Allergy and Clinical Immunology **123**(4): 774-782.e5.

Schisterman, E. F., A. Vexler, et al. (2006). The Limitations due to Exposure Detection Limits for Regression Models. **163:** 374-383.

Schulze, A., R. T. van Strien, et al. (2007). "Characterisation of asthma among adults with and without childhood farm contact." European Respiratory Journal **29**(6): 1169-1173.

Smit, L. A. M., M. Zuurbier, et al. (2007). "Hay fever and asthma symptoms in conventional and organic farmers in The Netherlands." Occupational and Environmental Medicine **64**(2): 101-107.

Snijders, A., P. Kalinski, et al. (1998). "High-level IL-12 production by human dendritic cells requires two signals." International Immunology **10**(11): 1593-1598.

Stern, D. A., J. Riedler, et al. (2007). "Exposure to a farming environment has allergen-specific protective effects on TH2-dependent isotype switching in response to common inhalants." Journal of Allergy and Clinical Immunology **119**(2): 351-358.

Strachan, D. P. (1989). "Hay fever, hygiene, and household size." British Medical Journal **299**(6710): 1259-1260.

Strachan, D. P., H. R. Anderson, et al. (1994). "A national survey of asthma prevalence, severity, and treatment in Great Britain." Archives of Disease in Childhood **70**(3): 174-178.

Tobin, J. (1958). "Estimation of Relationships for Limited Dependent Variables." Econometrica **26**(1): 24-36.

Turk, T. (1987). "Von Pirquet, allergy and infectious diseases: a review." Journal of the Royal Society of Medicine **80**.

Uh, H.-W., F. Hartgers, et al. (2008). Evaluation of regression methods when immunological measurements are constrained by detection limits. **9:** 59.

van Strien, R. T., R. Engel, et al. (2004). "Microbial exposure of rural school children, as assessed by levels of N-acetyl-muramic acid in mattress dust, and its association with respiratory health." Journal of Allergy and Clinical Immunology **113**(5): 860-867.

Viinanen, A., S. Munhbayarlah, et al. (2005). "Prevalence of asthma, allergic rhinoconjunctivitis and allergic sensitization in Mongolia." Allergy **60**(11): 1370-1377.

Viinanen, A., S. Munhbayarlah, et al. (2007). "The protective effect of rural living against atopy in Mongolia." Allergy **62**(3): 272-280.

Vogel, K., N. Blümer, et al. (2008). "Animal shed Bacillus licheniformis spores possess allergy-protective as well as inflammatory properties." Journal of Allergy and Clinical Immunology **122**(2): 307-312.e8.

Von Hertzen, L. C. and T. Haahtela (2004). "Asthma and atopy – the price of affluence?" Allergy **59**(2): 124-137.

von Mutius, E. (2000). "The environmental predictors of allergic disease." Journal of Allergy and Clinical Immunology **105**(1, Part 1): 9-19.

von Mutius, E., F. D. Martinez, et al. (1994). "Prevalence of asthma and atopy in two areas of West and East Germany." Am. J. Respir. Crit. Care Med. **149**(2): 358-364.

von Mutius, E., S. Schmid, et al. (2006). "The PASTURE project: EU support for the improvement of knowledge about risk factors and preventive factors for atopy in Europe." Allergy **61**(4): 407-413.

von Mutius, E. and D. Vercelli "Farm living: effects on childhood asthma and allergy." Nat Rev Immunol **10**(12): 861-868.

von Mutius, E. and D. Vercelli (2010). "Farm living: effects on childhood asthma and allergy." Nat Rev Immunol **10**(12): 861-868.

Vuillermin, P. J., A. L. Ponsonby, et al. (2009). "Microbial exposure, interferon gamma gene demethylation in naïve T-cells, and the risk of allergic disease." Allergy **64**(3): 348-353.

Waser, M., K. B. Michels, et al. (2007). "Inverse association of farm milk consumption with asthma and allergy in rural and suburban populations across Europe." Clinical & Experimental Allergy **37**(5): 661-670.

Wennergren, G., L. Ekerljung, et al. "Asthma in late adolescence – farm childhood is protective and the prevalence increase has levelled off." Pediatric Allergy and Immunology **21**(5): 806-813.

Wickens, K., J. M. Lane, et al. (2002). "Farm residence and exposures and the risk of allergic diseases in New Zealand children." Allergy **57**(12): 1171-1179.

Wright, A. L. (2004). "The epidemiology of the atopic child: Who is at risk for what?" The Journal of allergy and clinical immunology **113**(1): S2-S7.

9 Acknowledgements

Sehr herzlich möchte ich mich bei Frau Prof. Erika von Mutius bedanken für ihre tatkräftige und sehr fachkundige Unterstützung, für die Aufnahme in ihre Arbeitsgruppe und auch dafür, dass sie mir die Freiheit und das Vertrauen gegeben hat meine eigenen Ideen umzusetzen und weiterzuentwickeln.

Mein besonderer Dank gilt Frau Dr. Sabina Illi, die mich durchgehend ausgezeichnet betreut hat. Die vielen konstruktiven Vorschläge und Anmerkungen bei der Ausarbeitung von Vortrag, Paper und Dissertation waren sehr hilfreich und förderlich.

Es freut und ehrt mich gleichermaßen, dass Frau Dr. Bianca Schaub mir die statistischen Analysen ihrer Daten anvertraut und mich in ihre Arbeitsgruppe aufgenommen hat. Außerdem bedanke ich mich bei ihr für die intensive Korrekturlesung des Papermanuskripts und Posters.

Für die sehr gute, kollegiale und angenehme Zusammenarbeit möchte ich mich sowohl bei der Arbeitsgruppe „von Mutius" als auch bei der Arbeitsgruppe „Schaub" bedanken.