

Aus dem Institut für Sozialmedizin, Epidemiologie und
Gesundheitsökonomie der Medizinischen
Fakultät Charité – Universitätsmedizin Berlin

DISSERTATION

Innovational Approach to Simultaneously Analyze Numerous
Subgroups and to Identify Subgroup Effects in Clinical and
Epidemiological Studies

Innovativer Ansatz zur simultanen Analyse zahlreicher
Subgruppen und Identifikation von Subgruppeneffekten in
klinischen und epidemiologischen Studien

zur Erlangung des akademischen Grades
Doctor rerum medicinalium (Dr. rer. medic.)

vorgelegt der Medizinischen Fakultät
Charité – Universitätsmedizin Berlin

von

Christoph Muysers
aus Kamp-Lintfort

Datum der Promotion: 30.11.2023

Inhaltsverzeichnis

Abbildungsverzeichnis	iii
Abkürzungsverzeichnis	iv
Zusammenfassung	1
1 Einleitung	3
1.1 Subgruppen in Klinik und Epidemiologie	3
1.2 Verwendete Studien	5
1.3 Subgruppenanalysen	7
1.4 Kontroversen zu Subgruppenanalysen	8
1.5 Motivation zur Exploration	10
1.6 Innovativer Ansatz für ein Subgruppenscreening	10
1.7 Fragestellungen	11
1.7.1 Methodische Fragestellungen	11
1.7.2 Inhaltliche Fragestellungen	12
2 Methodik	13
2.1 Daten und Studien	13
2.1.1 Klinische Studien	13
2.1.2 Simulationsstudie	13
2.1.3 Epidemiologische KiGGS Studie	14
2.2 Datenaufbereitung	16
2.2.1 Skalenniveau	16
2.2.2 Fehlende Daten	16
2.3 Subgruppenscreening	16
2.3.1 Durchführung	16
2.3.2 Arbeitsoberfläche des ‚Subgroup Explorers‘ und Interpretation	17
2.3.3 Zusätzliche Funktionen des ‚Subgroup Explorers‘	19
3 Ergebnisse	22

3.1	Onkologische Studie.....	22
3.2	Fertilitätskontrollstudie.....	26
3.3	Simulationsstudie	27
3.4	Epidemiologische KiGGS Studie	27
4	Diskussion.....	29
4.1	Zusammenfassung der Ergebnisse.....	29
4.2	Interpretation der Ergebnisse	29
4.3	Einbettung der Ergebnisse in den bisherigen Forschungsstand.....	30
4.4	Stärken und Schwächen.....	31
4.4.1	Ambivalenz und Schwächen des Subgruppenscreenings	31
4.4.2	Stärken des ‚Subgroup Explorers‘.....	32
4.5	Implikationen für die Praxis und zukünftige Forschung	32
4.5.1	Implikationen für die Praxis.....	32
4.5.2	Zukünftige Forschung	33
5	Schlussfolgerungen.....	35
	Literaturverzeichnis	36
	Eidesstattliche Versicherung	39
	Anteilerklärung an den erfolgten Publikationen.....	40
	Auszug aus der Journal Summary List.....	42
	Druckexemplare der Publikationen.....	44
	Lebenslauf.....	66
	Komplette Publikationsliste.....	67
	Danksagung	70

Abbildungsverzeichnis

ABBILDUNG 1: EXEMPLARISCHE GRAPHIK DES ‚SUBGROUP EXPLORERS‘	18
ABBILDUNG 2: ‚SUBSCREEN EXPLORER‘ - SUBGRUPPENERGEBNISSE DER ONKOLOGIESTUDIE	23
ABBILDUNG 3: MOSAIKPLOT DER ONKOLOGIESTUDIE (SUBGRUPPEN ‚ALTER‘ UND ‚VORBEHANDLUNG‘)	24
ABBILDUNG 4: ‚SUBSCREEN COMPARER‘ DER ONKOLOGIESTUDIE	25
ABBILDUNG 5: ‚SUBSCREEN EXPLORER‘ DER FERTILITÄTSKONTROLLSTUDIE	26

Abkürzungsverzeichnis

BMI	Body Mass Index
CRAN	Comprehensive R Archive Network
EMA	European Medicines Agency
FDA	Food and Drug Administration
HR	Hazard Ratio
KI	Künstliche Intelligenz
KiGGS	Kinder- und Jugendgesundheitssurvey
OR	Odds Ratio
OS	Overall Survival
PFS	Progression-Free Survival
RKI	Robert Koch-Institut
SES	Sozioökonomischer Status (engl. socio-economic status)

Zusammenfassung

Hintergrund: Subgruppenanalysen sind Bestandteil klinischer und epidemiologischer Studien. Meist werden jedoch nur wenige, präspezifizierte Subgruppen betrachtet. Durch gewachsenes Interesse an Subgruppeneffekten, Intensivierung von Forschungsfeldern in diesem Bereich und gesteigener Rechenkapazität bei komplexen Analysemodellen, werden umfangreichere Subgruppenanalysen vorgenommen. Die Homogenität wichtiger Subgruppen im Verhältnis zum Gesamtergebnis kann dabei verallgemeinernde Aussagen rechtfertigen. Erwartbare Subgruppenergebnisse stützen bisherige Annahmen, während unerwartete Ergebnisse weitere Forschung anstoßen können.

Methode: Die Daten zweier klinischer Studien (Onkologie, Fertilitätskontrolle), einer Simulationsstudie sowie einer bevölkerungsbasierten (epidemiologischen) Studie wurden mit einem innovativen ‚Subgruppenscreening‘ untersucht. Die explorativen Analysen umfassten graphische Verfahren, Künstliche Intelligenz gestützte Random Forests und deskriptive Statistiken zu möglichst allen mutmaßlich relevanten Subgruppen. Analysen wurden mit dem entwickelten R-Paket ‚subscreen‘ durchgeführt.

Ergebnisse: In den klinischen Studien wurden in unerwarteten und kombinierten Subgruppen positive Effekte entdeckt. Bei anderen Faktoren konnten Trends in den ordinalen Faktorstufen aufgezeigt werden und wissenschaftliche Erkenntnisse stützen. Die Simulationsstudie hat Aufschlüsse zur Sensitivität der Methode mittels der eingefügten Effekte liefern können, wobei die ‚Variable Importance‘ aus dem Bereich Machine Learning als Referenz dienen konnte. Weitere Szenarien zeigten positive Robustheitseigenschaften des R-Pakets. In der epidemiologischen Studie konnten Effekte für bestimmte Subgruppen bestätigt, sowie neue unbekannte Auffälligkeiten beobachtet werden.

Diskussion: Für alle drei gegebenen Studien wurde ein besseres Datenverständnis erlangt. Dem Bedarf einer umfassenden Recherche wurde Rechnung getragen und neue Erkenntnisse bezüglich der Zusammenhänge mit Subgruppen wurden offenkundig. Des Weiteren konnten mögliche Zusammenhänge zwischen potenziellen Einflussfaktoren und Erkrankungen aufgezeigt werden, wobei zusätzliche Funktionen des R-Pakets hilfreich bei den umfassenden Explorationen waren. Die unterstützenden Analysen sind sowohl für klinische als auch epidemiologische Studien routinemäßig zu empfehlen. Nachteile, insbesondere Multiplizitätsprobleme, sollten durch inhaltliche Diskussionen und Sensitivitätsanalysen ausreichend berücksichtigt werden.

Abstract

Background: Subgroup analyses are part of clinical and epidemiological studies. Usually, only few, pre-specified subgroups are considered. Lately, extensive subgroup analyses are performed due to growing interest in subgroup effects, intensification of research in this area, and increased computational capacities for complex analysis models. The homogeneity of important subgroups in relation to the overall result can justify generalized statements. Expected subgroup results support previous assumptions, while unexpected results may prompt further research.

Methods: The data from two clinical studies (oncology, fertility control), a simulation study, and a population-based (epidemiological) study were analyzed using an innovative 'subgroup screening'. The explorative analyses included graphical methods, artificial-intelligence supported random forests, and descriptive statistics for all presumably relevant subgroups. Analyses were carried out with the developed R-package 'subscreen'.

Results: In the clinical studies, positive effects were discovered in unexpected and combined subgroups. For other factors, trends in the ordinal factor levels could be shown and support scientific findings. The simulation study was able to provide information on the sensitivity of the methods based on the added effects, with the 'Variable Importance' from the field of machine learning serving as a reference. Other scenarios showed positive robustness properties of the R-package. In the epidemiological study, effects for certain subgroups were confirmed and new unknown effects were observed.

Discussion: A better understanding of the data was obtained for all three given studies. The need for comprehensive research was taken into account and new findings regarding the relationships with subgroups became apparent. Furthermore, possible relationships between potential influencing factors and diseases could be shown, with additional functions of the R-package being helpful in the comprehensive explorations. The supporting analyses are recommended for both clinical and epidemiological studies. Disadvantages, in particular multiplicity problems, should be sufficiently taken into account through substantive discussions and sensitivity analyses.

1 Einleitung

1.1 Subgruppen in Klinik und Epidemiologie

Als Subgruppen werden Teilmengen von Studienpopulationen bezeichnet, die durch die Faktorstufen eines Faktors definiert sind. Als Faktoren kommen zu Studienbeginn (Baseline) erfasste Variablen in Frage, wie z.B. demographische Variablen (Familienstatus, Geschlecht, Geburtsort, etc.) deren Faktorstufen (,ledig‘, ,divers‘, ,Berlin‘, etc.) die Einteilung in nicht überlappende Subgruppen erlauben. Subgruppenanalysen dienen dazu, zu untersuchen, ob es unterschiedliche Resultate, im Hinblick auf eine Zielgröße, zwischen den interessierenden Faktorstufen gibt. Die Unterschiede können dabei hinsichtlich der Haupt- oder Nebenzielgröße (primäre / sekundäre Endpunkte) mittels Wiederholung einer geplanten Hauptanalyse in den verschiedenen Subgruppen durchgeführt werden. Komplexere Verfahren berücksichtigen Interaktionen (Wechselwirkungen bei der Kombination verschiedener Faktorstufen) oder statistische Multiplizitätsprobleme (z.B. Adjustierung von p-Werten) mittels erweiterter statistischer Analyseverfahren.

Subgruppenanalysen sind mittlerweile etablierter Bestandteil klinischer, aber auch epidemiologischer Studien. Bisher wurden jedoch meist nur wenige und möglichst präspezifizierte Subgruppen betrachtet. Allerdings sind durch jüngere Entwicklungen in den Bereichen personalisierter Medizin, epidemiologischer Beobachtungsstudien, Anforderungen der Gesundheitsbehörden [1] mit neuen Fragestellungen und, mithin durch Forschungsarbeiten in der Statistik, relevante Fortschritte in der Methodik zu erweiterten Subgruppenanalysen erzielt worden [z.B. 2, 3, 4]. Nicht zuletzt sind auch durch verbesserte Rechnerleistungen Möglichkeiten geschaffen worden, die zuvor nicht oder nur unzureichend für komplexe Verfahren oder große Datenmengen umgesetzt werden konnten.

Im Bereich der klinischen Studien sind Subgruppenanalysen bei der Beurteilung einer Intervention nützlich, u.a. um die Ergebnisse der Subgruppen mit dem Ergebnis der gesamten Studienpopulation zu vergleichen. Dadurch kann bzgl. dieser Subgruppen Homogenität bestätigt bzw. Heterogenität aufgedeckt und in diesem Fall relevante Subgruppen für die Einschätzung der Wirksamkeit und Sicherheit einer Intervention identifiziert werden. Grundsätzlich kann nicht a priori von Homogenität einer Studienpopulationen ausgegangen werden [1]. Eine solche nicht geprüfte Homogenitätsannahme impliziert etwa die Feststellung „die statistisch signifikante Überlegenheit eines Testpräparats in

der Studienpopulation gilt gleichermaßen für alle Subgruppen“. Eher sollte eine Verschiedenheit unterstellt werden. Solange es keine ausreichende Überprüfung der Homogenität gibt, sind verallgemeinernde Aussagen über die Studienpopulation nicht gerechtfertigt. Rob Hemmings [2] beschreibt einen fließenden Übergang von ‚groben‘ Analysen einzelner Gruppen mit wenigen einfachen Faktorstufen bis hin zur personalisierten Medizin, was einer sehr feingliedrigen Gruppierung unter Beachtung vieler Faktoren entspricht. Im ersteren Fall würden beispielsweise zwei verschiedene Altersgruppen verglichen; die Hinzunahme von zunehmend mehr Faktoren führt dann zu detaillierteren und kleineren Subgruppen, bis hin zur Individualebene.

Subgruppenanalysen haben bei diesem betrachteten breiten Spektrum unterschiedlichste methodische Ansätze und medizinisch-biologische Fragestellungen. Von Interesse ist es, das Gesamtergebnis einer Untersuchung auf alle Subgruppen entweder übertragen zu können oder aber Subgruppen mit (un-)erwarteten Ergebnissen zu identifizieren. Bei der Interpretation von Subgruppenergebnissen sind neben inhaltlich-medizinischen auch statistische Eigenschaften zu berücksichtigen. Insbesondere sind statistische Multiplizitätsprobleme [3] zu nennen. Hierbei steht die Anzahl der analysierten Subgruppen in einem direkten Zusammenhang mit der zuverlässigen Interpretierbarkeit der Ergebnisse. Je mehr Analysen, die multiples Testen nicht berücksichtigen, umso größer ist die Wahrscheinlichkeit für das Vorkommen zufälliger falscher Schlussfolgerungen. Das heißt, es ist mit zunehmender Sicherheit davon auszugehen, dass mindestens ein falsch-positives oder falsch-negatives Ergebnis vorliegt. Die Aussagekraft ist daher für solche Subgruppenergebnisse beschränkt; i.d.R. können sie nur rein explorativ interpretiert werden.

Darüber hinaus beschäftigt sich die Epidemiologie *„mit der Verteilung von Krankheiten und mit Faktoren, welche diese Verteilung beeinflussen“* [4]. Weiterhin wird *„große Aufmerksamkeit auch der Kombination von Faktoren gewidmet ... Faktoren, denen man sich gleichzeitig aussetzt, sich in ihrem Effekt nicht einfach addieren“* lassen. Solche Interaktionen mit z.B. synergistischen Effekten sind bekannt, werden aber oft bei der Modellierung und Ursachenforschung vernachlässigt. Auch hierbei können Subgruppenanalysen einen wichtigen unterstützenden Beitrag liefern.

1.2 Verwendete Studien

Die Daten aus zwei klinischen Studien, die für diese Arbeit herangezogen wurden, sind aus der pharmazeutischen Entwicklung der Phase III gewählt, da für diese späten Studienphasen meist genügend Probanden zur Verfügung stehen, um auf möglichst vielen Faktorstufen der subgruppendifinierenden Faktoren aussagekräftige Fallzahlen gewährleisten zu können.

Die onkologische ‚Mission‘ Studie war eine randomisierte, doppelblinde, zweiarmige, placebokontrollierte Phase-III-Studie, die in mehr als 30 Ländern Europas, Nord- und Südamerikas sowie im asiatischen Raum durchgeführt und 2015 [5] publiziert wurde. Die Patienten erhielten nach Randomisierung entweder das Prüfpräparat mit dem pharmazeutischen Wirkstoff Sorafenib oder Placebo, jeweils inklusive begleitender state-of-the-art Behandlung. Dieses entweder bis zum Fortschreiten der Krankheit (progression-free survival; PFS), bis zum Tod (overall survival; OS) oder zu einem Zeitpunkt, der wegen vorzeitigem Studienende als zensiert evaluiert wird, wie etwa unverträglicher Toxizität, Rücknahme der Einwilligung des Patienten oder nach Ermessen des Prüfarztes. Es wurden über 700 Patienten zu nahezu gleichen Anteilen in der Verum- und der Placebogruppe untersucht. Die hier interessierenden klinischen Endpunkte PFS, OS und deren Zusammenhang untereinander aber auch ihrer Einflussfaktoren sind Teil vieler medizinischer und biostatistischer Diskussionen.

Ein möglichst andersartiges Studiendesign wurde für die zweite klinische Studie gewählt, um das Anwendungsspektrum der Subgruppenanalyse der vorliegenden Arbeit zu erweitern. Im Vergleich zur obigen zweiarmigen Onkologiestudie wird mit dieser zweiten Studie demonstriert, welchen Beitrag Subgruppenanalysen für unkontrollierte Studien leisten können. In einer Fertilitätsstudie wurden ca. 1500 Frauen eingeschlossen, die sich für eine hormonelle Empfängnisverhütung mit niedrigdosierten Levonorgestrel entschieden. In der offenen (nicht randomisierten, nicht verblindeten) einarmigen Studie, die 2016 beendet und 2020 publiziert wurde [6], war das Ziel, während einer möglichst einjährigen Dauer, unbeabsichtigt auftretende Schwangerschaften zu evaluieren. Dieses unkontrollierte Studiendesign entspricht den Richtlinien der amerikanischen Food and Drug Administration (FDA) [7] und European Medicines Agency (EMA) [8] zur Überprüfung der Wirksamkeit von hormonellen Kontrazeptiva, wobei die benötigte Fallzahl von der angenommenen Wirksamkeitsrate, hier der Pearl Index, abhängt. Die Studie wurde wegen mut-

maßlicher unzureichender Wirksamkeit vorzeitig beendet, da die beobachtete Schwangerschaftsrate nicht mit der angestrebten Rate, auch unter Berücksichtigung von zufälligen statistischen Schwankungen, in Einklang zu bringen war.

Weiterhin wurde vom Promovenden, in Anlehnung an eine Arzneimittelstudie, eine Simulationsstudie programmiert, um die Eigenschaften der verwendeten Methodik gezielt zu untersuchen. In dieser Studie werden Zufallszahlen als fiktive Zufallsergebnisse generiert, die virtuellen Probanden zugeordnet werden. Hierbei sind unterschiedliche Subgruppeneffekte implementiert worden. In einem Vergleich zweier Gruppen mit jeweils 600 Subjekten waren, zusätzlich zu einer prägnanten gesamten Hazard Rate, unterschiedliche simulierte Effekte zur Beeinflussung der Hazard Rate bestimmten Subgruppen zugeordnet. Weitere Effekte wurden in zusätzlichen Untersuchungen implementiert und befassten sich mit Ungleichgewichten der Subgruppengrößen und mit höheren Anzahlen an Faktorstufen, die die Einschätzung von Subgruppeneffekten erschweren könnten. Diese zusätzlichen Effekte und die zugehörigen Untersuchungen sind aufschlussreich für ein besseres Verständnis der angewandten Methodik dieser Arbeit, sind jedoch aus Platzgründen in der ausgewählten Publikation [9] nicht weiter beschrieben. Ziel einer solchen Simulationsstudie ist es besondere Umstände, die sowohl (un-)beabsichtigt als auch (un-)beobachtbar in tatsächlichen Studien auftreten könnten, in der Simulation kontrolliert einzusetzen, um deren Auswirkungen zu evaluieren. So ist etwa die Untersuchung, ob die o.g. Ungleichgewichte einen relevanten Einfluss auf die Sensitivität der Analyse haben, mit dem Wissen über die implementierten Effekte einfacher zu evaluieren. Im Gegensatz zu tatsächlichen Studien, bei denen eine Variation der Ungleichgewichte nicht einfach zu realisieren und das Wissen um die tatsächlichen Effekte nicht vorhanden ist, sondern nur statistisch geschätzt werden kann.

Um die Anwendung der Subgruppenanalysen zu erweitern, wurde neben den klinischen Studien die Methodik zusätzlich in einer epidemiologischen Studie untersucht [10]. Im Vergleich zu typischen klinischen Studien, die mit einem kontrollierten Studiendesign (z.B. Anzahl Untersuchungsgruppen, festgelegte zu untersuchende Faktoren, interventionelle Behandlung) viele Planungsaspekte berücksichtigen können, ist in der epidemiologischen Studie mit deutlich mehr, aber unkontrollierten und nicht-interventioneller Behandlung zu rechnen. Bei dieser epidemiologischen Studie handelt es sich um die sogenannte zweite Welle des Kinder- und Jugendgesundheitsurvey (KiGGS) des Robert Koch-Instituts (RKI). Die Intention der KiGGS-Studie ist mittels einer „repräsentativen bundesweiten Untersuchung und Befragung von Kindern und Jugendlichen im Alter von

0-17 Jahren die Datenlage zur Gesundheit der heranwachsenden Generation in Deutschland zu verbessern und Informationslücken zu schließen“ [11]. In der zweiten Welle der KiGGS-Studie nahmen in drei Jahren über 15.000 Jungen und Mädchen der Bundesrepublik an der Studie teil. Dabei wurden Indikatoren zur Prävalenz von Risikofaktoren, gesundheitsrelevanten Verhaltensweisen und Lebensbedingungen und Krankheiten erfasst. Die Daten der zweiten Welle zur KiGGS-Studie basieren auf altersspezifischen Fragebögen sowie medizinischen Untersuchungen und Laborergebnissen. Der Fokus der Subgruppenmethodik war für die vorliegende Arbeit auf die 12-Monats-Prävalenzen von allergischen Formen von Asthma bronchiale (im Folgenden kurz Asthma) und Rhinitis - mit Heuschnupfen (Pollinose) als einer speziellen Unterart – (im Folgenden kurz Rhinitis) gelegt. In den Subgruppenanalysen wurde insbesondere der Sozioökonomische Status (SES) nach Winkler [12] untersucht.

1.3 Subgruppenanalysen

Fortschritte der konfirmatorischen Statistik im Bereich des multiplen Testens (Multiple Comparison Procedures) umfassen z.B. Gatekeeping oder Adaptive/Group Sequential Designs [13]. Diese liefern, mit zum Teil komplexen statistischen Verfahren, Möglichkeiten für konfirmatorische Ergebnisse der Subgruppen und erlauben somit verallgemeinernde Aussagen über größere Populationen. Aktuelle Methoden und Überlegungen zu Subgruppenanalysen bzw. dem angrenzenden Thema der multiplen Vergleiche sind, mitunter auch unter Beteiligung des Promovenden, in neueren Fachbüchern und Artikeln einsehbar [z.B. 3, 14, 15].

Im Kontrast zu den klassischen, frequentistischen Methoden der Statistik mit dem Ziel konfirmatorische Aussagen zu treffen, steht beim Data Mining die Erkenntnisgewinnung und die Deskription im Vordergrund, ohne notwendigerweise eine konfirmatorische Aussage zu treffen. Dank der gesteigerten Rechnerleistung sind als nützliche Methoden für den Bereich des Data Mining Verfahren die Künstliche Intelligenz (KI) und Maschine Learning zu nennen. Neuronale Netze, KI-gestützte Clusteranalysen und Random Forests sind somit dank verbesserter Rechnerleistungen in den letzten Jahren mit geringerem Ressourcenaufwand auch für große Datensätze anwendbar.

Weiterhin sind zahlreiche graphische Verfahren zu Subgruppenanalysen, mit unterschiedlichen Zielsetzungen, verfügbar. Die Darstellungen basieren auf einem breiten Repertoire vom einfachen deskriptiven Überblick der Resultate von Subgruppenanalysen,

wie z.B. auch Forest Plots, bis hin zu methodisch-analytisch basierten Ansätzen, die eine Adjustierung beinhalten, etwa im R-Paket ‚gMCP‘ [16]. Einen Übersichtsartikel zur Thematik graphischer Verfahren von Subgruppenanalysen liefern Ballarini et al. [17]. Hierin ist auch die erste ausgewählte Publikation des Promovenden [9] und der dort vorgestellten Applikation ‚subscreen‘ [18] genannt, wobei die Mächtigkeit, d.h. die große Anzahl durchführbarer Subgruppenanalysen, und die Interaktivität der Applikation hervorgehoben wird.

1.4 Kontroversen zu Subgruppenanalysen

Extreme Beispiele haben schon frühzeitig zu kritischen Diskussionen bezüglich unkontrollierter Subgruppenanalysen (etwa unlimitiert, nicht präspezifiziert oder nicht adjustiert) geführt. Ende der 1980‘er Jahre wurde in der randomisiert-kontrollierten ISIS-2-Studie [19] Aspirin bei Patienten mit Verdacht auf einen akuten Myokardinfarkt untersucht. In einer ungeplanten post-hoc Analyse wurde berichtet, dass Aspirin bei fast 3000 Patienten nicht wirksam ist, wenn sie das Sternzeichen Waage oder Zwilling haben.

Ein statistisches Spannungsfeld bezüglich des Umgangs mit Subgruppenanalysen wird bereits auch in früheren Publikationen beschrieben: *„Nearly everything that has been learned in epidemiology has been derived from the analysis of subgroups. In a randomized clinical trial, the entire purpose is the comparison of the test subjects and the controls, and when there is particular interest in the results of treatment in a certain section of trial participants, a subgroup analysis is performed. ... Nevertheless, the reliability of subgroup analysis can often be poor because of problems of multiplicity and limitations in the numbers of patients studied. The naive interpretation of the results of such examinations is a cause of great confusion in the therapeutic literature”* [20]. Während also Subgruppenanalysen zu wesentlichen Erkenntnissteigerungen beigetragen haben, sind die Ergebnisse, die statistisch nicht adjustiert wurden, nicht oder nur sehr eingeschränkt zu verwenden, um z.B. Verallgemeinerung von der Studienpopulation auf eine Gesamtpopulation zur übertragen.

Interessante oder auffällige Resultate aus explorativen Subgruppenanalysen sollten lediglich als Hinweise zu verstehen sein, die z.B. in konfirmatorischen Untersuchungen weiterer Studiendaten bestätigt werden. Eine Auswahl an nicht reproduzierbaren Ergebnissen aus Subgruppenanalysen, sowie weiteren grundsätzlichen Problematiken zur Subgruppenanalyse wird von Rothwell beschrieben [21]. In den 1980er Jahren ist ein

systematischer Fehlgebrauch in der Epidemiologie angemerkt worden: „*Problems can arise through failure to specify the subgroups of interest a priori and through examining large numbers of subgroups after the fact, i.e., through multiple testing*” [22]. Diese Betrachtung wird unverändert vertreten, wie z.B. in aktuelleren Richtlinien des CONSORT Statements [23] beschrieben „*Post hoc subgroup comparisons (analyses done after looking at the data) are especially likely not to be confirmed by further studies*“. Statt einer rein explorativen Subgruppenanalyse sind eher zunehmend mehr Verfahren zur Adjustierung entwickelt worden. Außerdem hat eine methodische Effizienzsteigerung es ermöglicht (z.B. mittels ‚ α -propagation‘) eine Vielzahl an Hypothesen in einem Studien-setting auch mit konfirmatorischen Verfahren zu prüfen.

Trotz dieser stringenten Vorgehensweise muss die Frage gestellt werden, ob ein möglicher Erkenntnisgewinn aus vorliegenden Daten nicht verloren geht, wenn aus formellen Gründen eine Deskription der Daten bewusst unterlassen wird, um keine falsch-positiven Signale zu riskieren. Dabei sollte ein exploratives Screening (siehe Abschnitt 1.6) als Ergänzung verstanden werden und weitere Einblicke in die Daten geben können. Diese müssen zwar in weiteren Analysen und Studien bestätigt werden – es hätte sie aber wohlmöglich ohne Exploration erst gar nicht gegeben.

Auch wenn, wie etwa durch das o.g. CONSORT, korrekterweise von einer hohen Wahrscheinlichkeit für falsch-positive Signale ausgegangen wird, sollte einer anderen akzeptierten Sichtweise Rechnung getragen werden, nämlich dass heterogene Studienpopulationen erwartbar sind, die auffällige Signale in explorativen Subgruppenanalysen liefern können.

Weiterhin ist es nicht realistisch, dass alle relevanten subgruppenspezifischen Faktoren prospektiv bekannt sind. Ein Ansatz post-hoc Subgruppenanalysen zu unterlassen, würde offensichtlich Erkenntnisgewinne zu wichtigen Subgruppen ausschließen, die nicht a priori bekannt waren. Die explorative Subgruppenanalyse basiert dabei auf dem Ansatz, dass nichts übersehen werden sollte. Hierzu bedenke man etwa die Bewertung von Nebenwirkungen im Bereich der klinischen Studien. Auffälligkeiten in zahlreichen deskriptiven Häufigkeitstabellen können ein (unerwartetes) Signal liefern und dürfen nicht ignoriert werden. Genauso sollte auch gelten dürfen, dass Potenziale von Patientengruppen, die besonders von einer Therapie profitieren, explorativ detektiert werden können.

1.5 Motivation zur Exploration

Bei epidemiologischen Studien sind auch die Kombinationen von Faktoren, die einen möglichen Einfluss auf Erkrankungen haben, von Interesse. Nicht für jeden Faktor, und umso weniger für jegliche Faktorkombinationen, ist genügend Expertenwissen vorhanden, um eine präspezifizierte Selektion von Faktoren vornehmen zu können. Ein möglichst unvoreingenommener breiter Ansatz bei der Faktorauswahl ist somit wünschenswert.

Ein etablierter Ansatz, um die Wahrscheinlichkeit für falsche Entscheidungen (Hypothesen verwerfen oder nicht zu verwerfen) gering zu halten, ist der Fokus auf wenige Hypothesen (folglich unter Berücksichtigung weniger Faktoren). Im Gegensatz dazu, war die Idee der explorativen Subgruppenanalyse für die vorliegende Arbeit nicht eine kleine Übersicht bestimmter Subgruppen zu erstellen, sondern einen allumfassenden Blick auf möglichst alle Subgruppen zu erhalten. Mit diesem Vorgehen sollte keine Konkurrenz, sondern eine Ergänzung zur Methodik der adjustierten und präspezifizierten Ansätze etabliert werden.

1.6 Innovativer Ansatz für ein Subgruppenscreening

Für klinische und epidemiologische Studien ist ein Ansatz gewünscht, mit dessen Hilfe ein umfassender Überblick von möglichst allen erhobenen und relevanten Faktoren, als auch deren Kombinationen von Faktorstufen geschaffen werden kann. Das Subgruppenscreening mit dem ‚Subgroup Explorer‘ (R-Paket ‚subscreen‘ [18]) wurde zu diesem Zweck entwickelt.

Mit dem Subgruppenscreening lässt sich die Analyse zahlreicher Subgruppen vornehmen, um zu erwartende Subgruppeneffekte zu bestätigen oder unerwartete Subgruppeneffekte zu entdecken, die es in weiteren Untersuchungen noch zu verifizieren gilt. Die Detektion von Effekten geschieht etwa, indem der Fokus auf extreme Punkte, solche die weit vom Gesamtergebnis sind, gelegt wird.

Als ursprünglich inspirierendes Grundgerüst des ‚Subgroup Explorers‘ ließe sich ein Forest Plot anführen. Anstelle einer inhaltlich sortierten Darstellung der Ergebnisse zu den ausgewählten Faktoren mit Konfidenzintervallen in einem Forest Plot, werden beim ‚Subgroup Explorer‘ alle Ergebnisse der Faktoren und aber auch aller Faktorkombinationen in einem Scatterplot, geordnet nach deren Gruppengröße, dargestellt. Hierin bedeutet

jede Faktorkombination als einzelner Punkt im Scatterplot die Darstellung einer Subgruppe, wobei dessen Position auf der horizontalen Achsenrichtung die Gruppengröße und in der Vertikalen den Effekt relativ zu Gesamteffekt der Studienpopulation darstellt. Typischerweise erfolgt die Nutzung des ‚Subgroup Explorer‘ interaktiv an einem großen Bildschirm. Gegenüber den exemplarischen Graphiken in [9] und [10], sowie den folgenden Abbildungen 1 - 5, können am gemeinsamen Bildschirm die interdisziplinären Expert:innen der Studie zusammen Auffälligkeiten diskutieren und die Darstellungsformen nach Bedarf unmittelbar wechseln.

1.7 Fragestellungen

Da es sich beim Subgruppenscreening um einen innovativen Ansatz handelt, der erstmals für klinische und epidemiologische Studien publiziert wurde, sind auch Fragestellungen zu diesem Ansatz selbst ein Teil dieser Arbeit und im Abschnitt 1.7.1 beschrieben. Zusätzlich sind dort typische Fragen aufgeführt, die sich häufig im Anschluss einer Studiendurchführung stellen; z.B. im Zuge der Ursachenforschung einer nicht erfolgreichen klinischen Studie. Spezifische Fragen zu den in dieser Arbeit herangezogenen Studien sind im Abschnitt 1.7.2 aufgeführt.

Auf Grund des explorativen Charakters der Analysen mit ergebnisoffenen Untersuchungen ergibt sich eine umfangreiche Liste an Fragestellungen im Vergleich zu einer Untersuchung mit einer expliziten Hypothese.

1.7.1 Methodische Fragestellungen

Um konkrete Aussagen zu Effekten und Auswirkungen machen zu können, bietet sich für die methodischen Fragestellungen eine Simulationsstudie an.

- Ist die Methode des Subgruppenscreenings sensitiv genug ist, um Effekte zu detektieren? Mittels der Simulationsstudie sollte überprüft werden, ob sich die eingefügten simulierten Subgruppeneffekte mit der verwendeten Methodik aufdecken lassen, um Aussagen über die Sensitivität des Subgruppenscreenings treffen zu können.
- Welchen Einfluss haben stark unterschiedlich große Subgruppen, sowie große Anzahlen an Faktorstufen?
- Lässt sich eine unerwartet große Streuung in den Daten einer klinischen Studie durch eine inhomogene Studienpopulation erklären?

1.7.2 Inhaltliche Fragestellungen

Während bei prospektiv geplanten Studien eine primäre Fragestellung (Hypothese) im Vordergrund steht, ergeben sich für das retrospektive Subgruppenscreening eine Vielzahl an Fragen, die typischerweise auch in anderen klinischen und epidemiologischen Studien auftreten (siehe auch [9]):

- Im Ergebnis einer klinischen Studie, die das erhofft signifikante Ergebnis liefern konnte: Gibt es eine oder mehrere Subgruppen, die das Gesamtergebnis allein tragen oder zusätzlich noch bessere Ergebnisse gezeigt haben? Daraus resultieren weitere Fragen: Können Subgruppen in frühen Studienphasen identifiziert werden, um die Studienpopulationen in späteren Studienphasen und confirmatorischen Studien enger zu fassen? Oder aber sollten unterschiedliche Dosierungen oder Behandlungsdauern für unterschiedliche Gruppen angestrebt werden?
- Onkologische Studie [5]: Gibt es Subgruppen, bei denen der Zusammenhang der klinischen Endpunkte PFS und OS besonders stark oder besonders schwach ausgeprägt ist? Gibt es Subgruppen, die bei einem Endpunkt erfolgversprechend, bei dem anderen aber nicht erfolgreich sind?
- Fertilitätskontrollstudie [6]: Gibt es Auffälligkeiten, die Aufschluss zu der nicht ausreichenden Effektivität der Studienpopulation gibt? Gibt es z.B. eine oder mehrere Subgruppen, die das Gesamtergebnis nivellieren? (Gruppen, die nicht von der Behandlung profitieren, wohingegen die restliche Population das erhoffte Resultat erbringen)
- Epidemiologische KiGGS-Studie [11]: Gibt es, möglicherweise auch unerwartete, subgruppendifinierende Faktoren, die einen positiven oder negativen Zusammenhang mit Asthma und Rhinitis haben? Ist weiterhin bei spezieller Berücksichtigung des SES das Risiko für Asthma oder Rhinitis, über sämtliche Subgruppen betrachtet, homogen oder gibt es Subgruppen, die Hinweise auf relevante Interaktionen mit dem SES geben? Alle Fragen auch insbesondere im Hinblick auf einer sehr großen Anzahl von kombinierbaren Faktorstufen.

2 Methodik

2.1 Daten und Studien

2.1.1 Klinische Studien

Für die onkologische Studie [5] mit dem primären Endpunkt Overall Survival (OS) und als weiteren Endpunkt Progression Free Survival (PFS), basierte die Auswertung durch das damalige Studienteam auf einem Cox Model zur Berechnung des Hazard Ratio (HR) für beide Endpunkte separat. Allerdings sind Aussagen über den Zusammenhang von OS und PFS im Hinblick auf die Verwendung von PFS als Surrogatparameter für OS von Interesse, wobei die Validität nicht ausreichend geklärt ist [24]. Auffällige Subgruppen können jedenfalls in diesem Zusammenhang einen Beitrag zur Diskussion liefern.

Für die zweite klinische Studie zur Fertilitätskontrolle [6] wurde die Rate an Schwangerschaften in Subgruppen betrachtet, und es wurden zielgerichtet Faktoren untersucht, bei denen ein konkreter Einfluss auf die Schwangerschaftsrate dem aktuellen Wissensstand entsprach. Eine Bestätigung oder Abschwächung dieses Wissensstands kann dabei zum Verständnis für die nicht erfolgreiche Studie hilfreich sein, wie auch die Detektion von weiteren auffälligen Subgruppen, die in diesem Zusammenhang nicht erwartbar sind. Insbesondere wird bei der Analyse mit dem Subgruppenscreening der Fokus auf Subgruppen gelegt, die das Gesamtstudienresultat erklären könnte. Dazu wurden unter anderem, entsprechend der Publikation von Yamazakia et al. [25], gezielt Interaktionen zu Body Mass Index (BMI) und Alter, sowie BMI und Ethnie untersucht. Im Gegensatz zur oben genannten onkologischen Studie wurde die Untersuchung auf einen einzelnen Endpunkt beschränkt. Allerdings gab es keine vergleichende Kontrollgruppe, was Interpretationen, neben dem ohnehin deskriptiven Charakter der hier angewandten Methodik, erschweren.

2.1.2 Simulationsstudie

Für die virtuelle Studie wurden insgesamt 1200 Subjekte mit einer zufälligen Überlebenszeit basierend auf einer Hazardfunktion in der Programmiersprache R simuliert. Exakt aufgeteilt in zwei Behandlungsgruppen (Verum und Placebo) mit jeweils 600 virtuellen Subjekten, die einer Subgruppe (mit Faktorstufen ‚A‘ oder ‚B‘) zugeordnet wurden, und unterschiedlich stark simulierte Hazard Rates haben, wurde untersucht, ob diese Unterschiede mit der angewandten Methode entdeckt werden konnten. In dieser Simulation

wurde die Faktorstufe ‚B‘ mit jeweils 100 virtuellen Subjekten deutlich kleiner gewählt, als die jeweils 500 Subjekte auf der Faktorstufe ‚A‘, um auch unbalancierte Situationen zu erfassen (z.B. mehr Frauen als Männer in einer Studie).

Zusätzlich zur visuellen Überprüfung wurde aus dem Bereich Machine Learning die so genannte ‚Variable Importance‘ [26] als Untersuchungsmethode herangezogen, um die visuelle Untersuchung zu unterstützen. Diese Möglichkeit ist grundsätzlich mit dem Subgroup Explorer gegeben, wurde aber für diese Simulationsstudie exemplarisch mit weiteren Graphiken und Kenngrößen zur ‚Variable Importance‘, außerhalb des üblichen Subgruppenscreenings, ausgeweitet. Als weiterer Bestandteil der Simulation, wurde in mehreren Durchläufen der Simulation (zu betrachten als weitere Simulationsstudien) unterschiedliche Faktoren programmiert, die wiederum unterschiedliche HRs für Subgruppen beinhalteten. Hierbei allerdings mit deutlich unterschiedlichen Subgruppengrößen, d.h. starken Ungleichgewichten in den Besetzungszahlen der Behandlungsgruppen; in einem anderen Fall mit mehr als zwei Faktorstufen, denen wiederum unterschiedliche HRs zugeordnet wurden. Beide Erweiterungen der Simulation, d.h. Ungleichgewichte und multiple Faktorstufen, entsprechen typischen Situationen in realen Studien. Die angewandte Methodik bei der Untersuchung war dabei die gleiche, wie im einfacheren oben genannten Aufbau der Simulationsstudie.

2.1.3 Epidemiologische KiGGS Studie

In der zweiten ausgewählten Publikation [10] standen in der epidemiologischen Untersuchung die beiden Erkrankungen Asthma und Rhinitis im Fokus. Insgesamt wurden drei Endpunkte berücksichtigt, die 12-Monats-Prävalenz von Asthma (ohne Rhinitis), die 12-Monats-Prävalenz der allergischen Rhinitis (ohne Asthma) und das Auftreten beider Erkrankungen innerhalb eines 12-Monats-Zeitraums. Mit dieser Einteilung sollte auch eine allergische Multimorbidität dieser Atemwegserkrankungen separat betrachtet werden können. Alle Kinder und Jugendliche ohne diese Erkrankungen dienten hierbei als Kontrollgruppe. Mithin wurde der Altersbereich für die Promotionsarbeit auf 4-17-Jährige beschränkt. Kinder der Altersgruppe 0-3 Jahre wurden nicht für Analysen herangezogen. In dieser Altersgruppe lässt sich eine allergische Rhinitis nicht klar von einem Erkältungsschnupfen unterscheiden; zudem ist das bei älteren Kindern eher asthmatypische Wheezing (d.h. eine giemende, pfeifende Atmung) in den ersten Lebensjahren häufig viral bedingt und somit eine Fehlklassifikation nicht auszuschließen.

In den Auswertungen wurden ausschließlich Kinder und Jugendliche mit einem sozioökonomischen Status (SES) nach Winkler [12], bei dem Bildung, Einkommen und Beruf der Eltern die zentralen Dimensionen des Sozialen Status darstellen, der Kategorien ‚niedrig‘ und ‚hoch‘ in Betracht gezogen. Die dritte Kategorie ‚mittel‘ wurde aus methodischen Gründen nicht verwendet. Diese Reduzierung der Studienpopulation um die Kategorie ‚mittel‘ geht mit einem Informationsverlust einher. Um jedoch Auffälligkeiten zu entdecken, können die am stärksten unterscheidenden Extreme vorteilhafter sein und deutlichere Ergebnisse liefern. Alternativen zum Auslassen dieser Kategorie, wie z.B. die Dichotomisierung an einem medianen Cut-Point der sozioökonomischen Score-Daten, würden eine geringere Vergleichbarkeit mit anderen Publikationen, die sich auf den spezifischen SES nach Winkler [12] beziehen, bedeuten. Sie könnten möglicherweise auch in weniger diskriminierende Ergebnisse resultieren. Danach verblieben für die hier vorgestellte Untersuchung der Subgruppenanalyse über 4.600 Kinder und Jugendliche der KiGGS-Studie.

Für das Subgruppenscreening dieser Studie wurden 56 Faktoren berücksichtigt [10]. Die Auswahl der Faktoren basierte auf Diskussionen mit Statistiker:innen und Epidemiolog:innen unter Hinzunahme eigens hierzu erstellter deskriptiver Statistiken, wobei aus der Vielzahl der zur Verfügung stehenden Faktoren alle berücksichtigt wurden, bei denen ein Einfluss nicht ausgeschlossen werden konnte. Die hieraus definierten Subgruppen wurden hinsichtlich ihres potenziellen Einflusses auf die einfachen Erkrankungsraten für die drei Endpunkte Asthma, Rhinitis und ‚beide Erkrankungen zugleich‘ untersucht, um zunächst einen Überblick zu schaffen, welche Faktoren die Inzidenzen beeinflussen.

Anschließend wurde als zentraler Kern der Publikation [10] mit einem weiteren Subgruppenscreening der SES hinzugenommen und detailliert untersucht. Dazu wurden für alle drei Endpunkte das OR als Vergleich zwischen den Gruppen ‚niedriger‘ SES und ‚hoher‘ SES gebildet. Somit ergaben sich drei ORs für die gesamte Studienpopulation. Bezogen auf diese drei Gesamtergebnisse wurde das Subgruppenscreening angewandt, um subgruppenspezifische Faktoren zu detektieren, die im Vergleich zum Gesamtergebnis auffällig stärker oder schwächer erscheinen. Vergleiche der Erkrankungen, wie in der onkologischen Studie zwischen den Endpunkten, waren ebenso Teil des methodischen Ansatzes. Dadurch sollten Zusammenhänge zwischen den beiden allergischen Erkrankungen deutlicher werden, sofern sie durch weitere subgruppenspezifische Faktoren erklärt werden konnten.

2.2 Datenaufbereitung

2.2.1 Skalenniveau

Zur Beurteilung von Subgruppeneffekten werden ausschließlich subgruppendifinierende Faktoren (wie z.B. Geschlecht) mit ordinaler oder nominaler Skala genutzt, um aus einzelnen Faktorstufen (z.B. ‚weiblich‘, ‚männlich‘, ‚divers‘) oder aus der Kombination mit Faktorstufen eines anderen Faktors Subgruppen zu definieren. Bei stetigen Faktoren muss daher für das Subgruppenscreening zunächst eine Kategorisierung vorgenommen werden, indem geeignete Schwellenwerte zur Klassenbildung definiert werden (z.B. kg-Werte, um Gewichtsklassen zu erhalten). Nachteile der Klassenbildung sind bekannt [9], andererseits ermöglicht die Klasseneinteilung konkrete und leicht kommunizierbare Aussagen für Subgruppen im Vergleich zu einer Änderungsrate in Abhängigkeit von einem stetigen Faktor.

2.2.2 Fehlende Daten

Fehlte bei einem Subjekt die Information für eine Faktorstufe, so wurden der fehlende Wert auf die spezielle Kategorie ‚keine Daten‘ gesetzt, um zu viele leere Untergruppen zu vermeiden, die ansonsten auch durch die Kombination verschiedener Faktorstufen entstehen würden. Da das Ziel aller Untersuchung die Detektion auffälliger Faktoren ist, wird durch dieses Ersetzen mithin informatives Fehlen berücksichtigt. Würde eine Auffälligkeit im Zusammenhang mit dieser Kategorie auftreten, so wäre dies ein Hinweis darauf, dass fehlende Daten nicht zufällig verteilt auftreten, sondern möglicherweise in einem Zusammenhang mit dem Endpunktergebnis stehen.

2.3 Subgruppenscreening

2.3.1 Durchführung

Als zentrales Software-Element des Subgruppenscreenings diente der ‚Subgroup Explorer‘ (R-Paket ‚subscreen‘, Version 3) und ist auf dem CRAN-Server (Comprehensive R Archive Network) frei verfügbar [18]. Der ‚Subgroup Explorer‘ ist ein visuelles Werkzeug zur Unterstützung von Subgruppenberechnungen und -bewertungen. Es ist als eine auf Shiny [27] basierende interaktive grafische Benutzeroberfläche konzipiert. Während die Ideenentwicklung und der erste Prototyp als Vorstufe zum R-Paket ‚subscreen‘ noch von

einem Biostatistiker allein entstanden war, sind mittlerweile in der Version 3 der Applikation [18] mehrere tausend Zeilen Programmiercode entstanden, die von mehreren Statistikern und Programmieren gepflegt und bearbeitet werden.

Mit der Durchführung ist idealerweise eine interdisziplinäre Diskussion aller beteiligten wissenschaftlichen Bereiche verbunden. Typischerweise hat vorab, unter Beteiligung dieser Bereiche, die Auswahl der Faktoren stattgefunden. Durch den allumfassenden Ansatz sollten alle erhobenen Faktoren in Betracht gezogen werden, die nicht kategorisch ausgeschlossen werden können. Nach Vorliegen der Daten soll in dieser interdisziplinären und interaktiven Sitzung eine Diskussion geführt werden, in der alle subgruppenbezogenen Fragen sofort beantwortet werden können.

Eine Beschränkung auf die Kombination von drei oder maximal vier Subgruppenfaktoren erweist sich als sinnvoll, da Kombinationen von noch mehr Faktorstufen dann Subgruppen definieren, die möglicherweise leer oder sehr klein, aber in jedem Fall schwer zu interpretieren sind. Beispielsweise wären für die Faktoren Geschlecht, Altersgruppe, Gewichtsgruppe und Herkunft eine mögliche Subgruppe ‚weiblich, jung, mittlerer BMI und Nordamerika‘, die, je nach Studiengröße, evtl. schon klein ist, aber in der Kombination mit einer weiteren Faktorstufe dann schon leer oder zu klein für eine relevante Aussage ist.

Erwähnenswert ist die Effizienz der Durchführung. In der praktischen Vorgehensweise wird zu Beginn des Subgruppenscreenings in einem ersten zeitintensiveren Rechenschritt der Datensatz so im ‚Subgroup Explorer‘ vorbereitet, dass die Ergebnisse der Subgruppen in einer Liste erstellt werden. Damit sind im zweiten Schritt diese Ergebnisse sofort für die verschiedenen Darstellungen abrufbar. Auf diese Weise wird ein reibungsloser Wechsel zwischen verschiedenen Darstellungen für den interaktiven Teil des Subgruppenscreenings in der Teamarbeit ermöglicht.

2.3.2 Arbeitsoberfläche des ‚Subgroup Explorers‘ und Interpretation

Ein wesentliches Ergebnis des ‚Subgroup Explorers‘ ist ein Diagramm, in dem jeder einzelne Punkt für eine Subgruppe steht (Abbildung 1). Das Diagramm kann Tausende von ihnen zeigen. Rechnerisch und praktisch auch Millionen von Subgruppen, wie etwa für die KiGGS Studie durchgeführt [10], wobei aus Gründen der unübersichtlichen Darstellung auf eine Abbildung verzichtet wurde. Die Ergebnisse dieser Millionen Datenpunkte waren jedoch jederzeit abrufbar.

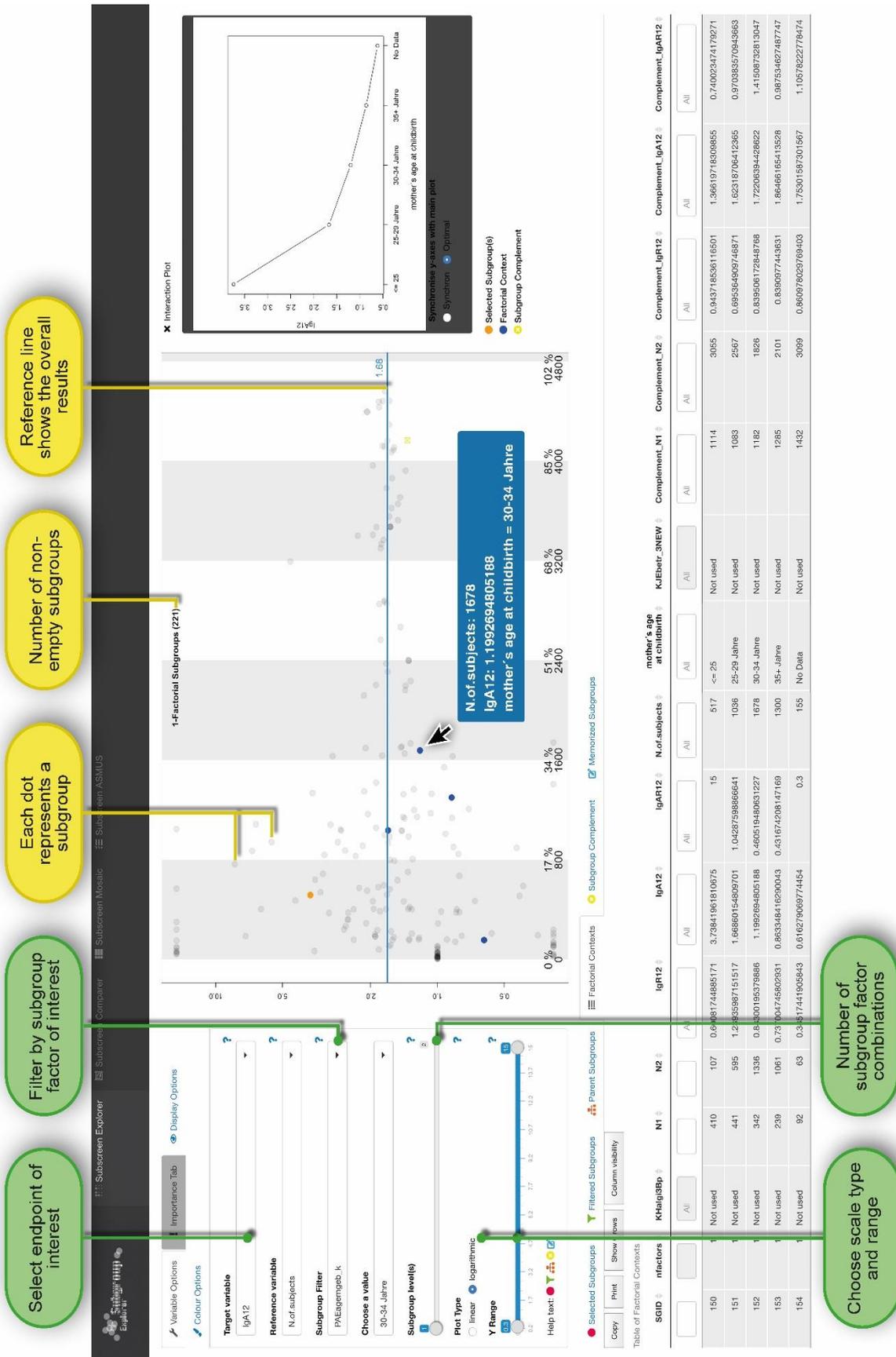


Abbildung 1: Exemplarische Graphik des ‚Subgroup Explorers‘

Quellenangabe: aus Muysers et al. [10, Figure 1]

Jede Subgruppe wird durch einen Punkt im Diagramm dargestellt, wobei die Position durch die Größe der Subgruppe (horizontale Achse) und das statistische Maß des Effekts (vertikale Achse) bestimmt wird. Die blauen Punkte im Diagramm und die tabellarischen Ergebnisse unterhalb des Diagramms für den Faktor „mother's age at childbirth“ erscheinen nach Anklicken des Reiters „Factorial Context“ für eine zuvor aktiv ausgewählte Subgruppe (orange). Die gelbe Komplementärgruppe hebt die Subgruppe aller Personen mit Ausnahme der Personen aus der ausgewählten (orange) Subgruppe hervor. Der „Interaction Plot“ (Liniendiagramm) rechts kann bei Bedarf um einen weiteren Faktor erweitert werden.

Für kleine Subgruppen (linker Bereich des Diagramms) sind größere statistisch zufällige Abweichungen vom mittleren Studieneffekt zu erwarten. Je größer eine Subgruppe ist (in Richtung des rechten Diagrammbereichs), umso kleiner werden die Abweichungen vom Studienmittel sein, da große Subgruppen einen großen Teil des Studienmittels beeinflussen. Typischerweise ist zu beobachten, dass die Punkte in dem Diagramm einen charakteristischen Funnel Plot bilden. Für Studien ohne Subgruppeneffekte würden die Abgrenzungen des Funnel Plots klar zu erkennen sein und nur wenige statistisch zufällige Ausreißer aufweisen. Markante Abweichungen von dieser Trichterform weisen auf auffällige Subgruppen hin. Die relevante Unterscheidung zwischen ‚zufällige Ausreißer‘ und ‚auffällige Subgruppen‘ ist zwar ein typisches Problem in der Statistik. In der Entwicklung des Formats eines Subgruppenscreenings und in der Entwicklung der ‚Subgroup Explorers‘ hat es sich aber als sinnvoll erwiesen auf p-Werte, Konfidenzbänder oder andere Maßzahlen zur Einteilung in (Nicht-)Signifikanz, z.B. mittels Resampling Methoden, zu verzichten. Die rein explorative Darstellung trägt zu einer wesentlich stärkeren inhaltlichen Diskussion bei, die ein wesentliches Element bei der Frage der Reliabilität und Hypothesengenerierung ist. Wohingegen p-Werte und Konfidenzbänder in diesem Kontext eine Trennschärfe suggerieren, die retrospektiv ohnehin nur von geringer Aussagekraft wären.

2.3.3 Zusätzliche Funktionen des ‚Subgroup Explorers‘

‚Subscreen Comparer‘: Für die Untersuchung von Studien mit zwei oder mehr Endpunkten bietet neben dem Diagramm des Screenings (Abbildung 1 im Abschnitt 2.2) die weitere Darstellung mittels des ‚Subscreen Comparer‘ [9] eine Möglichkeit zur Exploration. Hierbei werden die Subgruppenergebnisse gleichzeitig für zwei Endpunkte untereinander in einem Bild verglichen. Interaktive Selektionen und Betrachtungen einzelner Subgruppen sind dabei für den Vergleich unterstützend. Die Diskrepanzen bzw. Übereinstimmungen zwischen zwei Endpunkten bezüglich bestimmter Subgruppenergebnisse können dabei auch über die alternative Darstellung des ‚Bubble-Plot‘ visualisiert werden [9].

‚Bubble-Plot‘: In dieser Darstellung repräsentiert jeder ‚Bubble‘ eine Subgruppe. Zusätzlich lässt sich die Größe einer Subgruppe durch die relative Größe des ‚Bubbles‘ erkennen. Hierbei befinden sich perfekt übereinstimmende (hinsichtlich der beiden Endpunkte) Subgruppen auf einer Winkelhalbierenden und große Differenzen (hinsichtlich der beiden Endpunkte) weiter entfernt von dieser. Die Koordinaten eines ‚Bubbles‘ auf horizontaler und vertikaler Achse sind durch die Ergebnisse der beiden Endpunkte definiert.

„Variable Importance“: Als Kompromiss zu den genannten p-Werten und Konfidenzbändern einerseits und einer rein deskriptiven Darstellung der Subgruppenergebnisse andererseits, ist die Bereitstellung einer Liste mit ‚wichtigsten‘ Faktoren im ‚Subgroup Explorer‘ eingeführt. Mittels Random Forests aus dem Bereich Machine Learning wird die ‚Variable Importance‘ (VIMP) berechnet [28]. Hierbei wird keine Klassifizierung in (nicht-)signifikant vorgenommen, welche die erwähnte ergebnisoffene Diskussion behindern würde. Andererseits wird aber ein objektives Werkzeug bereitgestellt, das helfen kann, bei der Fülle an Faktoren ein Kriterium zu liefern, herausragende Faktoren zu detektieren. Der Random Forest gibt zwar keinen Hinweis auf die Richtung eines Effekts, hat aber den Vorzug auch sehr große Datensätze effizient zu klassifizieren.

Tabellenzeilen: In Abbildung 1 sind neben den bereits erklärenden Textboxen noch die Tabellenzeilen unterhalb des Diagramms zu erwähnen. Die Tabellen liefern, je nach Auswahl, einen schnellen Überblick zu den interessierenden Subgruppen. Die Auswahl kann sich dabei auf eine spezielle auffällige Subgruppe, die aus dem Diagramm direkt gewählt wurde beziehen. Oder es wird eine Auswahl für die Tabelle erstellt, indem gezielt Subgruppen über den einfach zu bedienenden Filter benannt werden. Umfangreiche Informationen ergeben sich dann zu den Ergebnissen solcher Subgruppen in Kombination mit anderen Faktoren, zu den ggf. übergeordneten Subgruppen (weniger Faktorkombinationen) und zu ‚Komplementärgruppen‘ (die Faktorstufen, die gerade nicht zu den auffälligen oder gewählten Subgruppen gehören). Auf diesem Weg soll ein möglichst aufschlussreiches Bild bestimmter subgruppendifinierender Faktoren bereitgestellt werden.

Interaktionenplot: Die kleinere Graphik neben dem Diagramm, der ‚Interaction Plot‘, zeigt bei ordinal skalierten Faktoren Monotonieeigenschaften, wie etwa in Abbildung 1, wo aufsteigende Altersklassen dargestellt sind. Bei nicht notwendigerweise ordinal skalierten Faktorstufen sind bei zwei oder mehr Faktoren Interaktionen gut darstellbar.

„Subscreen Mosaic“: Als weitere Option visualisiert das ‚Subscreen Mosaic‘ (Figure 6 in [9]) den interessierenden Effekt für alle Faktorenkombinationen von bis zu drei Faktoren und ermöglicht so eine genauere Untersuchung der komplementären Gruppeneffekte jedes Faktors und der Faktorkombinationen. In diesem Zusammenhang bedeutet eine komplementäre Gruppe, dass beispielsweise bei der Untersuchung einer weiblichen Subgruppe auch die männliche Subgruppe dargestellt wird. Die Berücksichtigung von Komplementärgruppen wird u.a. in den Richtlinien der EMA [1] gefordert. Da es bei mehr als zwei Faktorstufen und/oder der Kombination von Faktorstufen graphisch einfacher zu er-

fassen ist als in einer Tabelle, können viele Faktoren untersucht werden. Die nebenstehende Farbskala im Mosaik Plot kennzeichnet die Ergebnisse der jeweiligen Subgruppen, während deren Rechteckgrößen die Subgruppengrößen repräsentieren.

Allen Darstellungen gemeinsam sind die intuitiv einstellbaren Auswahlmöglichkeiten von Faktoren und die Mouseover Funktionen oder Möglichkeiten Subgruppenpunkte anzuklicken, um weitere Informationen zu Subgruppen zu erhalten. Vielfältige Farbgestaltungsmöglichkeiten und Skalentypen (linear, logarithmisch) vereinfachen die Interpretation der verschiedenen Darstellungen.

Zusammengefasst enthält der ‚Subgroup Explorer‘ umfangreiche Darstellungsmöglichkeiten, um einen umfassenden Überblick aller Subgruppen zu erhalten, Subgruppenvergleiche von Endpunkten in verschiedenen Varianten anzustellen, und komplementäre Gruppen in die Betrachtungen einzubeziehen.

3 Ergebnisse

3.1 Onkologische Studie

Im ‚Subgroup Explorer‘ zeigt sich der charakteristische Funnel Plot in Abbildung 2; ähnlich zu Abbildung 1, durch die größere Anzahl dargestellter Subgruppen jedoch mit einer klareren Abgrenzung der Punktwolke. Dabei wird eine Subgruppe im linken unteren Bereich ($N = 53$, $HR = 0,34$) sofort auffällig. Nach Markierung (interaktiv per Mausklick in der Shiny-Graphik des ‚Subgroup Explorers‘) werden die Details unterhalb der Graphik aufgeführt und zeigen, dass es sich um i) die chinesische Population, die ii) keine Vorbehandlung mit einem Epidermal Growth Factor Receptor (EGFR) hatten. Eine Subgruppe, die nicht präspezifiziert war und mutmaßlich ohne das Subgruppenscreening in dieser Kombination auch nicht aufgefallen wäre.

Die Subgruppe ‚East Asia‘ war präspezifiziert und zeigte ein HR nahe des Gesamtstudienresultates (d.h. nahezu kein Behandlungseffekt). Jedoch wurde diese Subgruppe in einer weiteren Analyse nach japanischer und chinesischer Population differenziert, die gegenteilige Effekte zeigten (verdeutlicht durch Markierung der japanischen Subgruppe)[9]. Dieses unerwartete Resultat hat nach inhaltlichen Diskussionen und Recherchen weitere Analysen angeregt (nicht Bestandteil dieser Arbeit), die dann auch eine valide Rationale für den positiven Effekt in der chinesischen Subgruppe erbringen konnten.

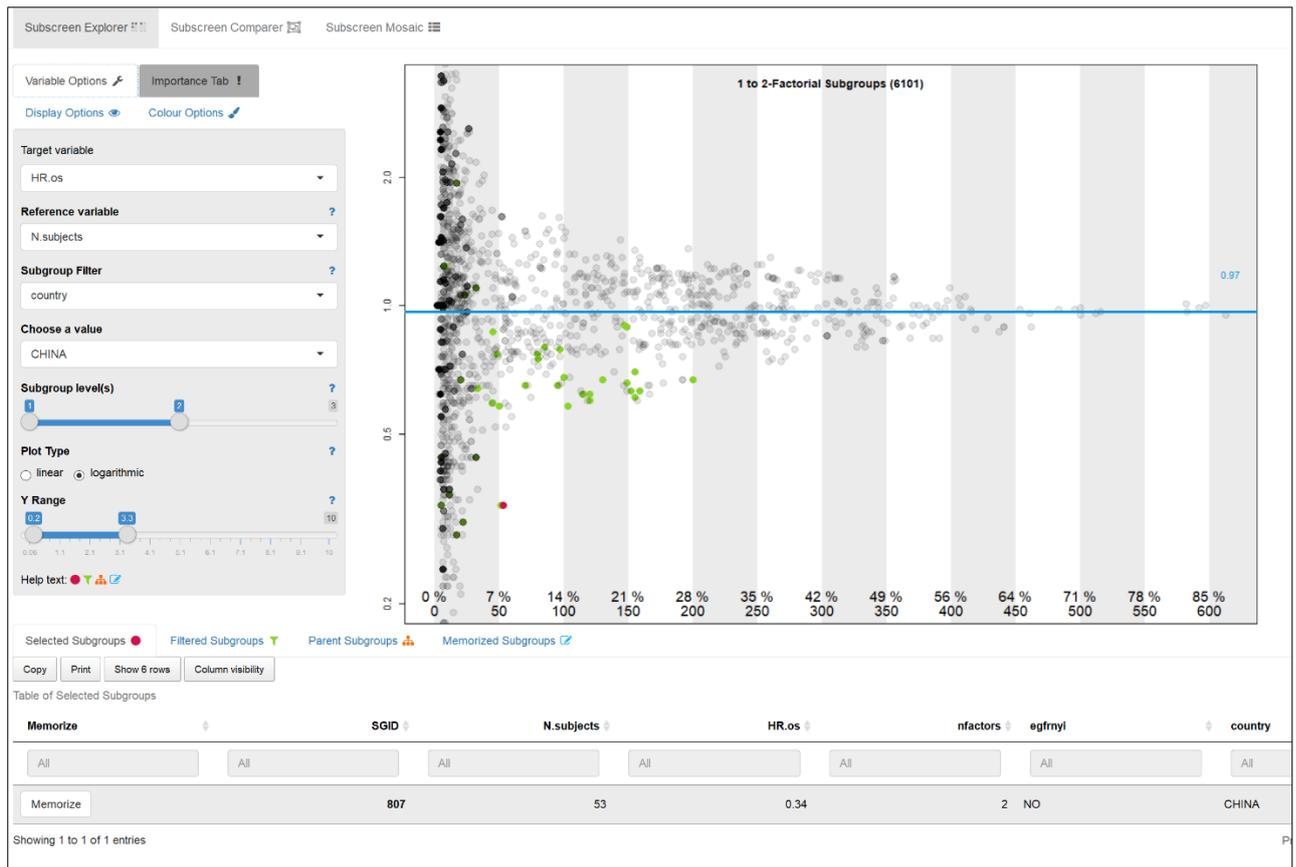


Abbildung 2: ‚Subscreen Explorer‘ - Subgruppenergebnisse der Onkologiestudie

Die blaue horizontale Linie dient als Referenzwert des Gesamtstudienresultates ($HR = 0,97$) und zeigt, dass die Intervention für den primären Endpunkt OS nahezu keinen Effekt hat. Die auffällige Subgruppe wird in der interaktiven Shiny Graphik per Mausklick als roter Punkt hervorgehoben. Wie im unteren Tabellenteil dargestellt, repräsentiert dieser Punkt die Subgruppe der „chinesischen Patienten ohne Vorbehandlung mit einem Epidermal Growth Factor Receptor (EGFR)“ mit einem Ergebnis zugunsten von Sorafenib ($HR = 0,34$). Motiviert durch dieses Ergebnis werden alle Subgruppen angesehen in denen chinesische Patienten enthalten sind. Nach Anwendung des entsprechenden Filters sind diese Subgruppen folglich grün markiert und zeigen konsistent günstige HRs. Quellenangabe: aus Muysers et al. [9, Figure 3]

Eine weitere auffällige Subgruppe wird in Abbildung 3 ersichtlich. Die Darstellungsform des ‚Subgroup Explorers‘ ist in dieser Graphik ein Mosaikplot.

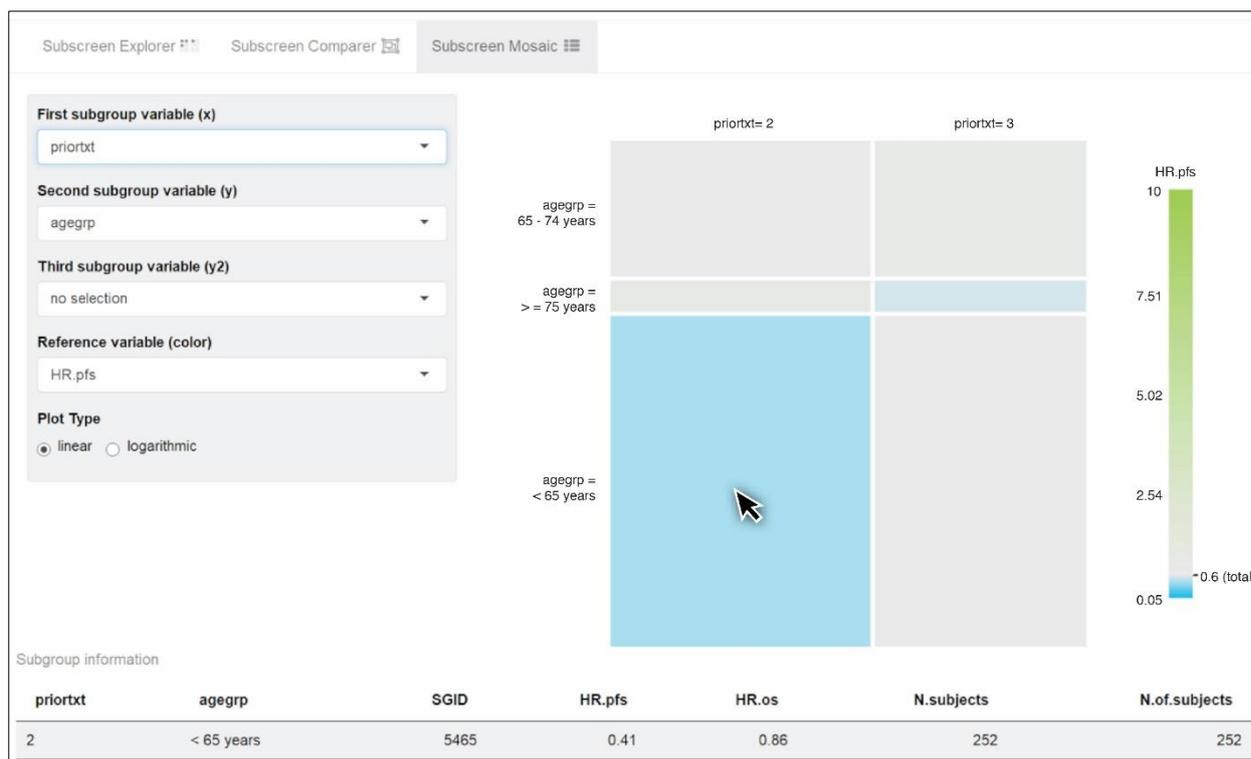


Abbildung 3: Mosaikplot der Onkologiestudie (Subgruppen ‚Alter‘ und ‚Vorbehandlung‘)

Im Mosaikplot sind komplementäre Subgruppen - auch als Kombination verschiedener Faktoren – dargestellt; die Rechteckgröße ist proportional zur Subgruppengröße. Der Punktschätzer der Subgruppe ist erkennbar durch den Farbcode entsprechend der Skala rechts neben dem Mosaikplot. Detailliertere Informationen sind interaktiv ablesbar in der Shiny Graphik durch die Mouse-Over Funktion. In obiger Abbildung wird hierdurch für das linke untere Rechteck ablesbar, dass für jüngere Patienten ohne Vorbehandlung ein deutlich besseres Behandlungsergebnis erzielt wurde (bezüglich PFS (HR = 0,41) im Vergleich zum Ergebnis der gesamten Studienpopulation (HR = 0,6)). Quellenangabe: aus Muysers et al. [9, Figure 6]

In diesem Fall zeigte sich, dass die Subgruppe der „unter 65-jährigen“, die „keine spezifische Vorbehandlung haben“ im direkten Vergleich zu allen anderen (komplementären Gruppen) am stärksten von der Therapie profitieren (HR = 0,41 für PFS und HR = 0,86 für OS). Die Größe des Rechtecks zeigte zugleich, dass es sich nicht um eine vernachlässigbar kleine Gruppe handelte, sondern mit über 250 Patienten (Statuszeile unter dem Mosaikplot mit Detailinformationen) eine relevante Aussagekraft besitzt.

Die Gesamtstudienresultate für die Endpunkte OS (HR = 0,97, nahezu kein messbarer Behandlungseffekt) und PFS (HR = 0,5, also zu Gunsten der untersuchten Intervention im Vergleich zu Placebo) waren deutlich verschieden. Dennoch ist der direkte Vergleich beider Endpunkte in einer gemeinsamen Grafik zur Untersuchung möglichst aller Subgruppen informativ. So wird durch eine gezielte Analyse interessierender Subgruppen,

so wie exemplarisch in Abbildung 4 für den subgruppendifinierenden Faktor ‚Region‘, die Homogenität der Subgruppen bzgl. beider Endpunkte untersucht.

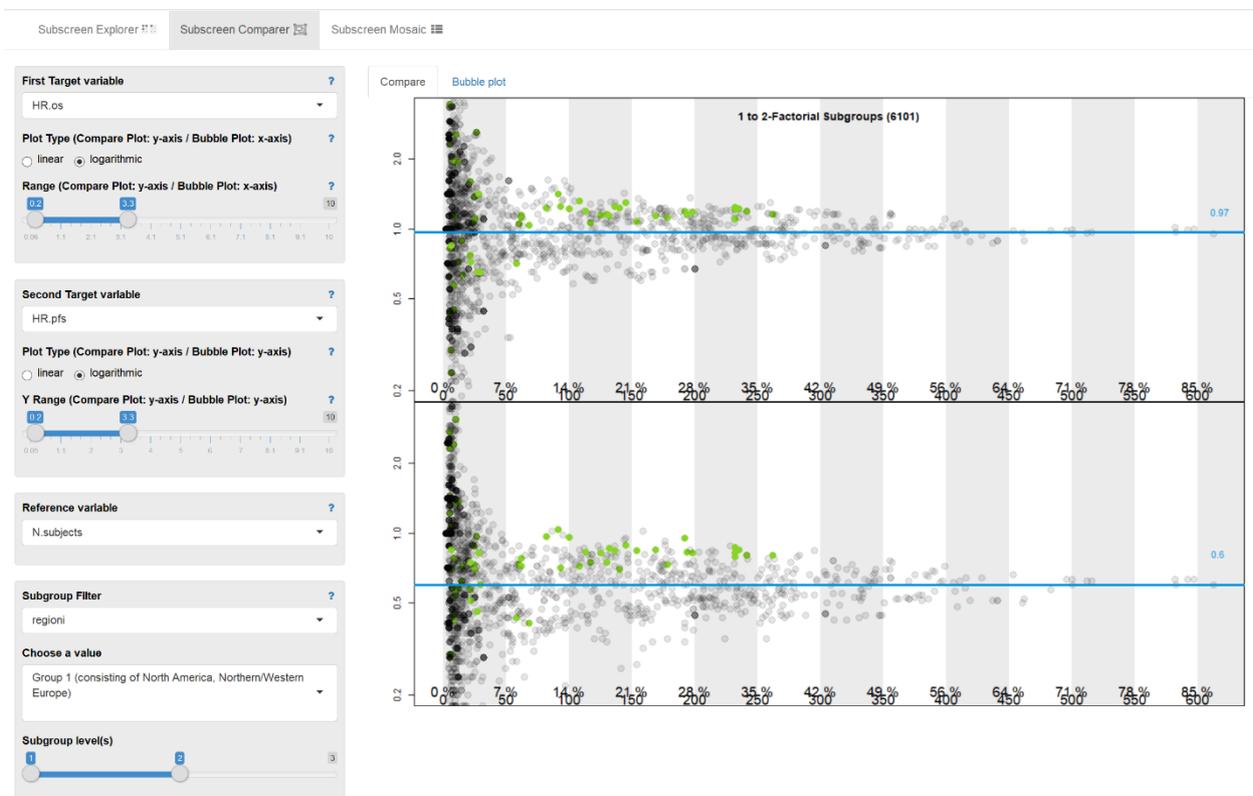


Abbildung 4: ‚Subscreen Comparer‘ der Onkologiestudie

In dieser Abbildung zeigt sich der Vergleich des Overall Survivals (OS, oberer Bildteil) und der Progression-Free Survival (PFS, unterer Bildteil). Durch Auswahl einer Faktorstufe im Filterbereich (linke Seite), können die Subgruppenergebnisse für beide Endpunkte gleichzeitig verglichen werden. In diesem Fall wurde eine bestimmte Region ausgewählt, die dadurch die entsprechenden Subgruppen in der Graphik grün hervorhebt. Quellenangabe: aus Muysers et al. [9, Figure 5]

Im Beispiel zeigte sich, dass alle Subgruppen, in denen die Faktorstufe ‚Gruppe 1‘ enthalten ist, kontinuierlich oberhalb des Studiengesamtergebnisses lagen. Eine ähnlich konsistente Verteilung der hervorgehobenen Punktwolken ist auf diesem Wege für alle relevanten Subgruppen erfolgt und kann somit als wichtiges Indiz für einen ähnlichen Trend der beiden Endpunkte gewertet werden.

3.2 Fertilitätskontrollstudie

Die hormonelle Fertilitätskontrollstudie zeigt in Abbildung 5 einen erwartbaren Trend, welcher bei zunehmenden BMI auf eine geringere kontrazeptive Wirksamkeit hinweist. Die dortige Gruppe '5' mit den größten BMIs fügt sich nicht voll in den beschriebenen Trend ein, kann jedoch als zufällige Abweichung in Betracht gezogen werden.

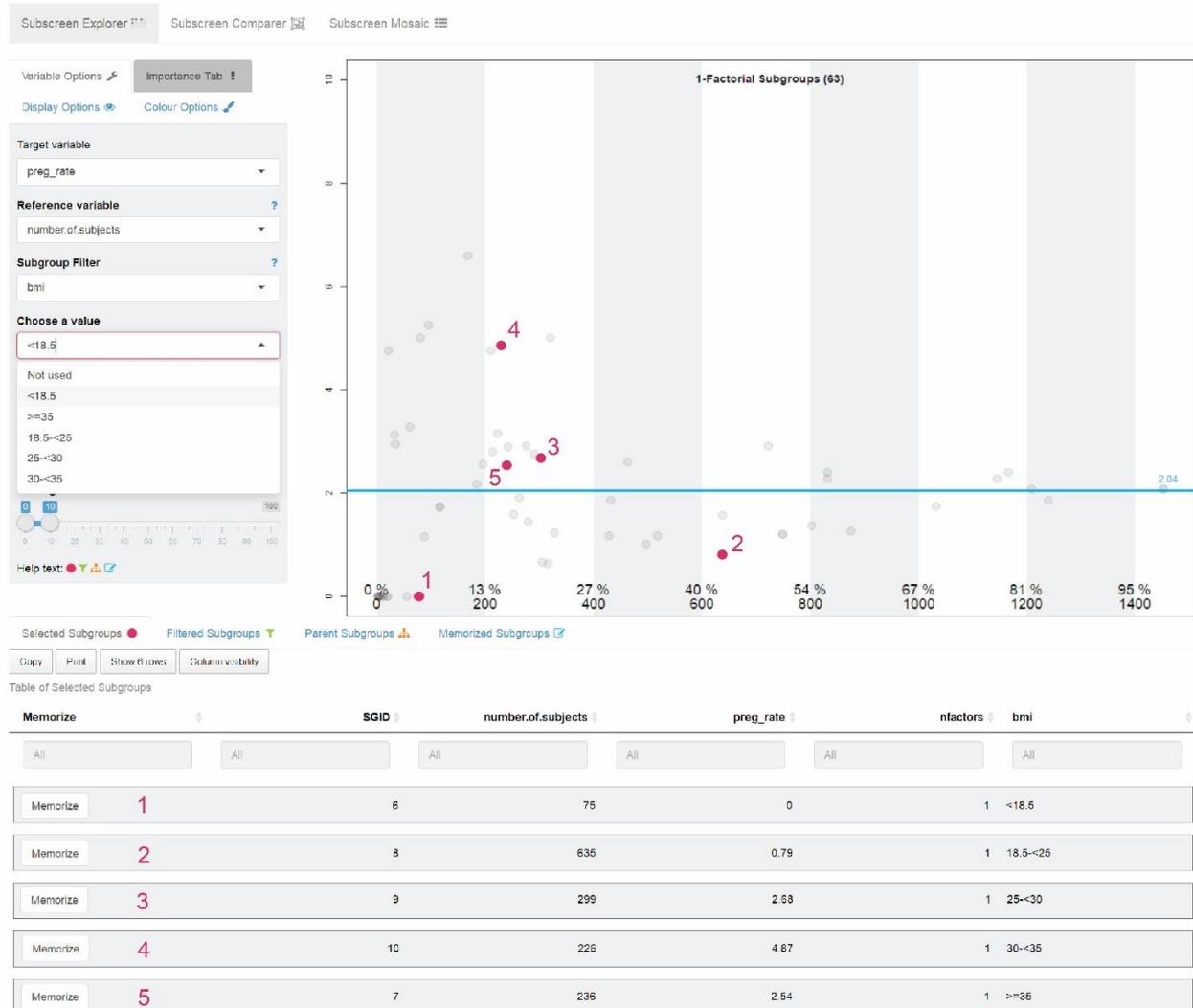


Abbildung 5: ‚Subscreen Explorer‘ der Fertilitätskontrollstudie

Die Abbildung zeigt die Ergebnisse mit Hervorhebung der BMI-Klassen. Jeder Punkt stellt eine Punktschätzung der Schwangerschaftsrate einer bestimmten Untergruppe dar. Die fünf rot markierten Punkte zeigen beispielhaft die Ergebnisse für die BMI-Untergruppen in fünf BMI-Kategorien (Stufen). Quellenangabe: aus Muysers et al. [9, Figure 7]

Bei der weiteren Analyse lag der Fokus auf Subgruppen, die mutmaßlich das Gesamtstudienresultat der nicht ausreichenden Wirksamkeit erklären könnten. Bei der gezielten Untersuchung von Interaktionen zu BMI und Alter, sowie BMI und Ethnie waren diese

Faktoren in der ausgewählten Publikation [9] wegen früherer Publikationen im Fokus [25]. Allerdings waren beim Subgruppenscreening keine klaren Hinweise auf Zusammenhänge auffällig. In sämtlichen BMI-Gruppen waren die Subgruppen zu den genannten Faktoren scheinbar zufällig um die horizontale Linie des Gesamtstudienresultates verteilt.

Darüber hinaus haben auch sämtliche weitere Untersuchungen im Subgruppenscreening der zwei- und drei-faktoriellen Kombination von Faktorstufen in den verschiedenen Methoden und Darstellungsmöglichkeiten des ‚Subscreen Explorers‘ keine weiteren Auffälligkeiten gezeigt.

3.3 Simulationsstudie

In der Simulation war der eingefügte Effekt einer Subgruppe auf den zwei Faktorstufen ‚A‘ und ‚B‘ sofort auffällig [9]. Das Ergebnis der Subgruppe mit Ausprägung ‚B‘ zeigte eine deutlich höhere HR (roter ausgewählter Punkt im Diagramm) und war oberhalb des Funnel; mithin waren alle Subgruppen mit den Faktorstufen ‚A‘ konsistent mit einer kleineren HR. Die ‚Variable Importance‘ aus dem Bereich Machine Learning hatte diesen Effekt ebenso als relevant an erster Stelle der Variablenpriorisierung gelistet. Weiterhin waren keine Auffälligkeiten beim Subgruppenscreening ersichtlich. Diese Ergebnisse zusammengefasst waren kongruent mit den simulierten Daten.

Der Einfluss von einer größeren Zahl von Faktoren und Faktorstufen oder eine starke Imbalance der Fallzahlen auf den Faktorstufen war gering und führte kaum zu auffälligen falsch-positiven Befunden. Parallel dazu durchgeführte Analysen in dem vorgesehenen Cox Modell zeigten häufiger falsch-positive, d.h. nominal signifikante p-Werte für Subgruppen bei denen tatsächlich kein Effekt simuliert wurde.

3.4 Epidemiologische KiGGS Studie

Für das Subgruppenscreening dieser Studie waren bei den einfachen Erkrankungsraten in der Subgruppe der Kinder und Jugendlichen, die im Alter von 0-6 Jahren keinen Kontakt zu Nutztieren hatten, mit über 10% besonders viele Rhinitis-Erkrankungen zu beobachten im Vergleich zur relativen Gesamthäufigkeit von unter 8%. Ebenso hatte die Tabakrauchexposition einen starken Einfluss auf eine Rhinitis-Erkrankung (etwa 9,5% bei jeglicher Tabakrauchexposition und über 12% bei aktivem Raucherstatus oder wenn

die Mutter während des Stillens geraucht hatte). Hingegen dazu waren bei den hier genannten Faktoren keine relevanten Auffälligkeiten in Bezug auf die Asthma-Erkrankung zu beobachten. Nach Durchführung des Random Forests zur Bestimmung der ‚Variable Importance‘ waren als wichtigste diskriminierende Faktoren für die Rhinitis Erkrankung der ‚Allgemeine Gesundheitszustand‘, ‚Aktive Rauchbelastung‘, ‚Atopische Dermatitis‘ und ‚Altersgruppe‘ berechnet worden. Allerdings gibt der Random Forest prinzipiell keine Richtung des Effekts an.

Für das OR (bezogen auf SES) ist in Abbildung 1 eine Übersicht der 1-faktoriellen Subgruppen für Asthma gegeben. ORs größer als 1 weisen auf eine stärkere Assoziation der Kinder mit niedrigem SES mit Asthma hin. Hierbei war der Faktor "Alter der Mutter bei der Geburt des Kindes" auffällig. Die weitere Untersuchung mit dem ‚Interaction Plot‘ (oben rechts in der Abbildung 1) zeigte deutlich einen abnehmenden Trend für das OR. Das bedeutet mit zunehmendem Alter der Mütter bei Geburt des Kindes sank das Risiko für die Kinder bei niedrigem SES an Asthma zu erkranken.

Weitere Trendanalysen für auffällige Gruppen zeigten hohe ORs für Asthma in der Gruppe der 11 bis 13-Jährigen (OR = 6,58) und für Rhinitis in der Gruppe der 14 bis 17-Jährigen (OR = 1,13) während alle anderen Altersgruppen für beide Erkrankungen deutlich niedrigere ORs aufweisen.

Andere Faktoren, die Auffälligkeiten aufweisen bei der Beziehung zwischen SES und den drei Krankheitsausprägungen waren die Allergiefaktoren SX1-Screening (d.h. spezifische Antikörper im Serum), Pollen- und Nahrungsmittelallergie. Diese Faktoren zeigten hohe ORs in beide Richtungen für verschiedene Stufen, lassen aber keinen klaren Trend erkennen, was eine plausible Hypothesenbildung erschwert.

Bei nahezu allen untersuchten Subgruppen zeigte sich für den Endpunkt ‚beide Erkrankungen zugleich‘ ein Effekt, der zumeist zwischen den Ergebnissen von Asthma und Rhinitis lag, was durch die häufig gegenläufigen Effekte der beiden Endpunkte erklärbar war.

4 Diskussion

4.1 Zusammenfassung der Ergebnisse

Für beide klinische Studien wurde ein besseres Datenverständnis erlangt. Dieses trug zu gezielten weiteren Analysen und Empfehlungen für zukünftige Studienplanungen bei. Regionale Effekte innerhalb Asiens in der onkologischen Studie wurden offenkundig, die in der regulären Analyse möglicherweise nicht entdeckt worden wären.

Die Fertilitätskontrollstudie zeigte keine unerwarteten Subgruppeneffekte, jedoch erwartbare Schwangerschaftseffekte im Zusammenhang mit Adipositas, die mit der angewandten Methode bestätigt werden können.

Beim Subgruppenscreening der Simulationsstudie wurden genau jene Subgruppenfaktoren auffällig, denen bei der Programmierung der Simulationsdaten Effekte zugefügt wurden. Zusätzliche Funktionen des ‚Subgroup Explorers‘, wie die ‚Variable Importance‘ aus dem Bereich des Machine Learnings, können Ergebnisse zu den wichtigsten Faktoren bestätigen.

Die KiGGS Studie hat Zusammenhänge zwischen potenziellen Einflussfaktoren und den allergischen Erkrankungen Asthma und Rhinitis mittels des Subgruppenscreenings aufzeigen können. Es konnten mehrere erwartbare Risikofaktoren für respiratorische Allergien bestätigt werden (z.B. SES, Rauchen oder Kontakt mit Tieren), aber auch weitere Detailinformationen zu den Zusammenhängen hinzugefügt werden. Mögliche Risikofaktoren wurden im Subgruppenscreening auffällig (Alter des betroffenen Kindes, Alter der Mutter bei Geburt, Bildungsstatus der Eltern und Kontakt mit Tieren kombiniert mit SES), denen weitere Untersuchungen folgen sollten, um diese Zusammenhänge zu bestätigen.

4.2 Interpretation der Ergebnisse

Die regionalen Unterschiede, die mittels Subgruppenscreenings in der onkologischen Studie beobachtet wurden, wurden im Studienteam diskutiert. Dies hatte zur Folge, dass weitere Sensitivitätsanalysen, mit dem Wissen aus dem Subgruppenscreening zielgerichtet ausgelöst werden konnten. Mit weiteren Erkenntnissen und Analysen früherer Studien, ergab sich ein valides Bild, das die unterschiedliche Wirksamkeit bei japanischen und chinesischen Patienten erklärt, mit der Folge einer angepassten Behandlungsempfehlung mit dem untersuchten Präparat für chinesische Patienten. Die weiteren Sensitivitätsanalysen, externe Beiträge und Diskussionen zur Behandlungsempfehlung waren

nicht Teil der Arbeit des Promovenden, sollen aber exemplarisch zeigen, dass das Subgruppenscreening Auslöser für relevante Erkenntnisse sein kann.

Auch wenn die Fertilitätskontrollstudie keine unerwarteten Subgruppeneffekte aufzeigte, wohl aber erwartbare Schwangerschaftseffekte bestätigen konnte, so ist dieses Resultat hilfreich. Einer umfassenden explorativen Untersuchung nach Subgruppeneffekten wurde genüge getan und demnach sind keine subgruppenspezifischen Probleme, die eine erhöhte Schwangerschaftsrate erklären könnten, übersehen worden.

Mittels der Simulationsstudie konnten Eigenschaften des ‚Subgroup Explorers‘, wie etwa das Erkennen der eingefügten simulierten Subgruppeneffekte, gezeigt werden. Der Einfluss einer Vielzahl von Faktorstufen oder eine starke Imbalance der Fallzahlen auf den Faktorstufen ist gering oder moderat, was auf gute Robustheitseigenschaften des Subgruppenscreenings schließen lässt.

Für die epidemiologische Studie zeigten die statistischen Verteilungseigenschaften klarere Zusammenhänge der potenziellen Risikofaktoren mit Asthma, also einen Hinweis auf die Erklärbarkeit der Erkrankung durch die Einflussfaktoren, im Vergleich zur Rhinitis. Somit erscheinen auch bestimmte Subgruppen beim Subgruppenscreening für Asthma auffälliger und können einen Einfluss auf diese Erkrankung, unter Berücksichtigung der ORs von SES, deutlich werden lassen. Dies gilt nicht im gleichen Maße für das Subgruppenscreening von Rhinitis. Hierbei gab es deutlich weniger Auffälligkeiten, so dass eine Erklärung der Erkrankung unter Berücksichtigung der ORs von SES durch subgruppendefinierende Faktoren schwieriger ist. Für die Betrachtung der einfachen Proportionen, d.h. Berücksichtigung der ORs von SES, hingegen lassen sich bei beiden Erkrankungen eine Vielzahl an erklärenden Faktoren als auffällig identifizieren.

Die fachlich-inhaltlichen Resultate aus den Studien stellen nach der Durchführung des Subgruppenscreenings meist erst einen Zwischenstand dar. Eine weitere Bewertung und Überlegungen zu Hypothesen und Analysen, die sich durch auffällige Subgruppen ergeben, unterliegt den entsprechenden Fachdisziplinen in Zusammenarbeit mit biostatistischer Expertise; so geschehen in der o.g. onkologischen Studie.

4.3 Einbettung der Ergebnisse in den bisherigen Forschungsstand

In der Vergangenheit haben die Fragestellungen zu Subgruppeneffekten häufig nur wenig strukturiert und sporadisch Berücksichtigung gefunden. Studententeams haben erst zu einem späteren Zeitpunkt Diskussionen geführt, welche Subgruppen interessant sein

könnten. Wenn die entsprechenden Subgruppenanalysen nicht das erwartete Ergebnis zeigten, wurden weitere Subgruppenanalysen durchgeführt. Dieser nacheinander durchgeführte zeitaufwendige Prozess mit Wiederholungen weiterer Subgruppenanalysen ist nicht effizient und hat schließlich nur einen kleinen Anteil der möglichen Subgruppenanalysen abgedeckt.

Das Subgruppenscreening mit Hilfe des ‚Subgroup Explorers‘ hilft, die Häufigkeit der Durchführungen von post-hoc Analysen zu minimieren, da in einer einzigen, interaktiven und interdisziplinären Sitzung die Analysen und Diskussionen zusammen durchgeführt werden. Dazu werden nahezu alle möglichen Subgruppenanalysen bereitgestellt.

Da keine statistischen Anpassungen und konfirmatorischen Bewertungen der Punktschätzer vorgenommen werden, wie etwa bei Shrinkage-Verfahren und Multiplen Vergleichen, erlauben diese explorativen Subgruppenanalysen zwar keine Verallgemeinerung auf die Gesamtpopulationen, ermöglichen aber begründete weiterführende Planungsmöglichkeiten (z.B. abschließende statistische Modellierung und/oder Hypothesenstellung für prospektive Analysen). Hinzugenommen ist die Betrachtung des nominalen Effekts beim Subgruppenscreening für eine offene Diskussion der beobachteten Resultate intuitiver und leichter. Auffällige Effekte, die nicht zu erwarten waren, sind im wissenschaftlichen Kontext zu diskutieren: „inwieweit erscheinen die Effekte erklärbar?“, „sind es den Ausmaßen nach eher zufällige Ausreißer oder mutmaßlich reproduzierbar?“ und „sind beobachtete Interaktionen von Faktorstufen plausibel?“.

4.4 Stärken und Schwächen

4.4.1 Ambivalenz und Schwächen des Subgruppenscreenings

Die Nachteile von rein explorativen Subgruppenanalysen sind vielfach in der Literatur beschrieben [3, 9, 14, 15]. Unabhängig davon, ob Multiplizitätsprobleme, fehlende Präspezifizierung oder Verzerrungsprobleme im Vordergrund der Publikationen stehen, sind zumeist auch Vorbehalte gegen unadjustierte und nicht a priori definierte Subgruppenanalysen genannt. Diese sind jedoch auf Schlussfolgerungen bezogen, die beim Subgruppenscreening nicht in Anspruch genommen werden, da es hierbei um erste Signale geht, die weiterer Bestätigung bedürfen, wie z.B. klinische und statistische Reliabilität und/oder weiterer Validierung. Diese Validierung kann mittels externer zusätzlicher Daten oder prospektiver neuer Studien stattfinden.

Intensive, ungeplante und unkontrollierte Subgruppenanalysen dienen als warnendes Beispiel vor den Risiken einer Fehlinterpretation, die so auch an das Subgruppenscreening gerichtet wird. Im Vergleich zu vielen Beispielen, wie etwa der ISIS-Studie [19], hat das Subgruppenscreening allerdings bereits bei der Analyse die interdisziplinäre Diskussion zur klinischen und statistischen Reliabilität und empfiehlt ausdrücklich weitere Sensitivitätsanalysen und Validierungen.

Grundsätzlich muss berücksichtigt werden, dass auffällige Subgruppen beim Screening die (extreme) Realisierung einer Stichprobe aus der Grundgesamtheit ohne tatsächlichen Subgruppeneffekt sein können. Somit wäre jede beobachtete Auffälligkeit potenziell auch nur ein Zufallsbefund, der nicht reproduzierbar ist. Umgekehrt lässt sich jedoch nicht ausschließen, dass die beobachtete auffällige Subgruppe durch den tatsächlichen Effekt von subgruppenspezifischen Faktoren bestimmt ist.

4.4.2 Stärken des ‚Subgroup Explorers‘

Das R-Paket ‚subscreen‘, wurde seit der Verfügbarkeit 2017 auf dem CRAN-Server (Comprehensive R Archive Network) deutlich über 20.000-mal heruntergeladen (<https://i-pub.com/dev-corner/apps/r-package-downloads/> Statusabfrage am 8. August 2022). In dieser Zeit hat die kontinuierlich wachsende Anzahl von Nutzern zu einem breiteren Verständnis des Subgruppenscreenings beitragen oder auch die Ideen in andere Plattformen übernommen, wie etwa einem SAS-Makro zur systematischen Evaluierung von Subgruppen [29]. Die vielfache Nutzung und Weiterentwicklung haben zu einer stabilen und nutzerfreundlichen aktuellen Version des R-Pakets beigetragen. Die Anwendung des ‚Subgroup Explorers‘ und Durchführung des Subgruppenscreenings sind intuitiv und einfach umzusetzen. Da es sich um eine zusätzliche explorative Analyse handelt, die zusätzlich zur konfirmatorischen Analyse empfohlen wird, können die detektierten Signale aufschlussreiche Zusatzinformationen liefern.

4.5 Implikationen für die Praxis und zukünftige Forschung

4.5.1 Implikationen für die Praxis

Die Problematik der Multiplizität und falsch positiv auffälliger Subgruppen ist nicht gelöst und hat auch über die Deskription und Hypothesenbildung hinaus eine Relevanz. Falls nach Diskussion einer auffälligen Subgruppe und der Bestätigung der klinischen Relevanz die Entscheidung für weitere Untersuchungen, z.B. in einer weiteren Studie, gefasst

würde, bieten sich zunächst weitere Sensitivitätsanalysen an. Tanniou et al. [30] haben Bias und Variabilität unadjustierter Schätzer untersucht und empfehlen auf Grund ihrer Ergebnisse Verfahren, wie z.B. empirische Bayesianische Shrinkage Methoden, um eine potenzielle Überschätzung der Subgruppeneffekte zu vermeiden. Anderenfalls besteht die Gefahr, wegen vermeintlich großer Effekte, die auf unadjustierten Schätzern beruhen, eine Studie zu planen, die keine ausreichende statistische Power auf Grund der Fallzahlplanung erzielt.

Andere sinnvolle Anwendungsbeispiele stammen, anstelle der Wirksamkeitsuntersuchungen, aus dem Bereich der Nebenwirkungen. Bei der Analyse und Beschreibung von Nebenwirkungsprofilen wäre es nicht wünschenswert allein Shrinkage-Verfahren anzuwenden oder andere adjustierte Schätzer zu berechnen, da die Gefahr besteht einen Subgruppeneffekt zu übersehen, der unadjustiert auffällig wäre. Aus Sicht der Prävention kann auch hier das Subgruppenscreening einen Beitrag leisten, umfassend und frühzeitig Signale zu entdecken, die sonst unter Umständen erst später oder überhaupt nicht entdeckt würden.

Für die komplexen biologischen Zusammenhänge zwischen Krankheiten und Faktoren, die im epidemiologischen Bereich auch von Interesse sind, ist das Subgruppenscreening ebenso von besonderer Bedeutung. Vermutlich haben hier die möglichen Interaktionen mehrerer Faktoren noch mehr Relevanz als in einer klinisch kontrollierten Studie mit klar geregelten Ein- und Ausschlusskriterien. Mit der stärkeren Einbeziehung von Kombination verschiedener Faktorstufen wächst die Anzahl beobachtbarer Subgruppen enorm, wie etwa in der beschriebenen Studie der zweiten ausgewählten Publikation [10], in der über eine Million Subgruppen allein durch die Kombination der 1-3 faktoriellen Stufen entstehen.

4.5.2 Zukünftige Forschung

Eine wissenschaftliche Frage ist die nach der Verlässlichkeit und der Nutzung der Ergebnisse aus dem Subgruppenscreening. Bereits zu früheren Zeitpunkten wurden bei Versionierungen des R-Paktes ‚subscreen‘ Ideen diskutiert, ob unter den Erweiterungen auch Bewertungskriterien eingefügt werden sollen. Darunter befand sich auch der Vorschlag, dem jeweiligen Funnel Plot ein ‚Konfidenzband‘ anzupassen, um die Relevanz auffälliger Subgruppen leichter einzuordnen. Dabei handelt es sich nicht um ein punktweises oder

simultanes Konfidenzband, das als Funktion eines unabhängigen Parameters Überdeckungswahrscheinlichkeiten darstellt [31]. Vielmehr wäre es ein, in Abhängigkeit von den Subgruppengrößen, fortlaufendes Konfidenzintervall.

Alternativ oder zusätzlich hierzu könnten alle Subgruppen mit einem p-Wert versehen werden, das post-hoc nominal oder adjustiert Informationen zur Realisierungswahrscheinlichkeit der jeweiligen Subgruppe, gegeben der Annahme, dass kein Effekt vorliegt, liefert. Die Berechnung zu einem adjustierten p-Wert und Konfidenzintervall könnten etwa auf Bootstrap Verfahren basieren [32].

Jedoch kann eine solche Bereitstellung von p-Werten einer ergebnisoffenen Diskussion in einem interdisziplinären Team entgegenstehen. Subgruppen mit einem ‚nicht-signifikanten‘ p-Wert könnten aus der Betrachtung fallen, obwohl sie anderenfalls als interessant oder auffällig diskutiert würden. Unabhängig vom tatsächlichen unbekanntem Effekt solcher Subgruppen, können Diskussionen über solche potenziellen Subgruppenfaktoren für die Planung weiterer, ähnlicher Studien vorteilhaft sein. Darüber hinaus ist das Subgruppenscreening nicht dafür konzipiert eine konfirmatorische Aussage zu treffen. Die Verlässlichkeit der explorativen Subgruppenresultate sollte erst durch weitergehende Analysen mit weiteren Daten und Hypothesen bestätigt oder widerlegt werden. Nichtsdestotrotz ist der Ansatz zur Berechnung von post-hoc adjustierten p-Werten und Konfidenzintervallen für den ‚Subgroup Explorer‘ von mindestens theoretisch-statistischem Interesse.

Unter Berücksichtigung dieser Problematiken, ist zu klären, wie Homogenität bzw. Heterogenität in Bezug auf den Einfluss von Subgruppenfaktoren deklariert werden können. Das Interesse hieran wird vordringlich in der EMA-Richtlinie [1] beschrieben, ist jedoch auch über klinische Studien hinaus von Interesse, um die Zusammenhänge in den Daten einer Studie zu verstehen. Während in [1] zur Bestimmung von Heterogenität wenige Maße exemplarisch genannt (aber wegen fehlender statistischer Power nicht ausdrücklich empfohlen) werden, liefern Tanniou et al. [33] die in Studien verwendete Analysemethoden zur Untersuchung von Konsistenz im Subgruppenkontext. Allerdings auch hier mit dem Ergebnis eines fehlenden allgemeingültigen Ansatzes zur Überprüfung der Konsistenz. Bisher ist, abgesehen von einer optischen Überprüfung der Homogenität der Daten im Funnel Plot des ‚Subgroup Explorers‘, noch kein Ansatz zur allgemeinen Konsistenzüberprüfung in diesem R-Paket enthalten. Ähnlich zur obigen Diskussion, ob Konfidenzintervalle und p-Werte hilfreich beim Subgruppenscreening sind, ist auch hier zumindest ein theoretisch-statistisches Interesse für weitere Forschungen zu nennen.

5 Schlussfolgerungen

Für die Studien in der vorliegenden Arbeit konnten viele Subgruppen als bekannte Einflussfaktoren bestätigt werden, aber auch mit neuen Erkenntnissen verbunden werden. Bisher unbekannte Auffälligkeiten lieferten Anregungen für weitere wissenschaftliche Diskussionen in den entsprechenden Fachdisziplinen und trugen zur Generierung neuer Hypothesen und gezielten Empfehlungen des zukünftigen Forschungsbedarfs bei.

Die vorliegende Arbeit konnte zeigen, dass die Durchführung des Subgruppenscreenings mittels ‚Subgroup Explorer‘ einfach und effizient zu realisieren ist. Die unterstützende Auswertung bei der Bewertung von Subgruppen einer klinischen oder epidemiologischen Studie ist für ein besseres Datenverständnis routinemäßig zu empfehlen. Hierbei können große Datenmengen verarbeitet werden, um interaktiv graphische Darstellungen und deskriptive Statistiken für sehr viele Subgruppen zu ermöglichen. Die Nachteile, insbesondere Multiplizitätsprobleme, sollten durch multidisziplinäre inhaltliche Diskussionen und Sensitivitätsanalysen ausreichend berücksichtigt, sowie durch konfirmatorische Analysen abgeschlossen werden.

Neben inhaltlichen klinisch orientierten Diskussionen ist biostatistische Expertise essentiell, um die explorativen post-hoc Analysen zu interpretieren. Das Ausmaß der mit dem Paket durchführbaren Subgruppenanalysen ist enorm, ist aber häufig nur ein Zwischenstand zur Generierung von Hypothesen, die mit weiteren Planungen durch Biostatistiker:innen begleitet werden müssen. Hierbei handelt es z.B. um weitere Sensitivitätsanalysen, prospektive geplante und konfirmatorische Datenausanalysen und adjustierbare statistische Modelle, die eine Bestätigung oder Widerlegung der Auffälligkeiten des Subgruppenscreenings erbringen können.

Literaturverzeichnis

1. European Medicines Agency (EMA) / Committee for Medicinal Products for Human Use (CHMP). Guideline on the investigation of subgroups in confirmatory clinical trials. 2019; EMA/CHMP/539146/2013
2. Hemmings R. Comment. *Statistics in Biopharmaceutical Research*. 2015; 7:4, 305-308
3. Dmitrienko A, Muysers C, Fritsch A, Lipkovich I. General guidance on exploratory and confirmatory subgroup analysis in late-stage clinical trials. *J Biopharm Stat*. 2016; 26(1): 71-98.
4. Kreienbrock L, Schach S. *Epidemiologische Methoden*. Heidelberg: Spektrum Akademischer Verlag; 4. Auflage 2005; ISBN-10: 3-8274-1528-4
5. Paz-Ares L, Hirsh V, Zhang L, de Marinis F, Yang JC, Wakelee HA, Seto T, Wu YL, Novello S, Juhász E, et al. Monotherapy Administration of Sorafenib in Patients With Non–Small Cell Lung Cancer (MISSION) Trial. *J Thorac Oncol*. 2015; 10:1745-1753.
6. Nave R, Höchel J, Mellinger U, Kohnke A, Elliesen J, Schmitz H. Pearl Index study with levonorgestrel-releasing intravaginal rings: using pharmacokinetic results to investigate treatment compliance as a potential contributor for contraceptive failure. *Hum Reprod*. 2020; 1;35(11):2515-2523. DOI:10.1093/humrep/deaa195
7. Food and Drug Administration (FDA) / Center for Drug Evaluation and Research. Establishing Effectiveness and Safety for Hormonal Drug Products Intended to Prevent Pregnancy Guidance for Industry. Draft Guidance 2019
8. European Medicines Agency (EMA) / Committee for Medicinal Products for Human Use (CHMP). Guideline on the clinical investigation of steroid contraceptives in women. 2005; EMEA/CPMP/EWP/519/98 Rev 1.
9. Muysers C, Dmitrienko A, Kulmann H, Kirsch B, Lippert S, Schmelter T, Schulz A, Mentenich N, Schmitz H, Schaefers M, et al. A Systematic Approach for Post Hoc Subgroup Analyses With Applications in Clinical Case Studies. *Ther Innov Regul Sci*. 2020; 54, 507–518, DOI:10.1177/2168479019853782
10. Muysers C, Messina F, Keil T, Roll S. A novel concept of screening for subgrouping factors for the association between socioeconomic status and respiratory allergies. *Journal of Exposure Science & Environmental Epidemiology*. 2022; 32, 295–302 DOI:10.1038/s41370-021-00365-x
11. Kurth BM. Der Kinder- und Jugendgesundheitsurvey (KiGGS): ein Überblick über Planung, Durchführung und Ergebnisse unter Berücksichtigung von Aspekten eines Qualitätsmanagements. *Bundesgesundheitsbl—Gesundheitsforsch—Gesundheitsschutz*. 2007; 50:533–546. DOI:10.1007/s00103-007-0214-x
12. Winkler J, Stolzenberg H. Der Sozialschichtindex im Bundesgesundheitsurvey. *Gesundheitswesen*. 1999; 61:178–183.

13. Dmitrienko A, Tamhane AC, Bretz F (Eds.). Multiple Testing Problems in Pharmaceutical Statistics. Chapman and Hall/CRC; 1st ed. 2009.
DOI:10.1201/9781584889854
14. Dmitrienko A, Lipkovich I, Dane A, Muysers C. Data-Driven and Confirmatory Subgroup Analysis in Clinical Trials. In: Ting N, Cappelleri J, Ho S, Chen DG (Eds.). Design and Analysis of Subgroups with Biopharmaceutical Applications. Springer International Publishing. 2020; pp. 33-91. DOI:10.1007/978-3-030-40105-4
15. Lipkovich I, Dmitrienko A, Muysers C, Ratitch B. Multiplicity issues in exploratory subgroup analysis. *Journal of Biopharmaceutical Statistics*. 2017
DOI:10.1080/10543406.2017.1397009
16. Bretz F, Maurer W, Brannath W, Posch M. A graphical approach to sequentially rejective multiple test procedures. *Statistics in medicine*. 2009; 28(4):586–604
17. Ballarini NM, Chiu YD, Koenig F, Posch M, Jaki T. A Critical Review of Graphics for Subgroup Analyses in Clinical Trials. *Pharmaceutical Statistics*. 2020; pp. 1539-1604
18. Kirsch B, Jeske S, Lippert S, Schmelter T, Muysers C, Kulmann H. Systematic Screening of Study Data for Subgroup Effects. <https://cran.r-project.org/web/packages/subscreen/index.html> CRAN Reference Manual. 2021; uploaded February 10: 1-7.
19. ISIS-2 (Second International Study of Infarct Survival) Collaborative Group. Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17,187 cases of suspected acute myocardial infarction. *Lancet*. 1988; 2:349-360.
20. Moreira E, Susser E. Guidelines on how to assess the validity of results presented in subgroup analysis of clinical trials. *Rev.Hosp. Clín. Fac. Med. S. Paulo*. 2002; 57 (2):83-88
21. Rothwell PM, Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. *Lancet*. 2005; 365: 176–86
22. Stallones RA. The use and abuse of subgroup analysis in epidemiological research. *Prev Med*. 1987; 16(2):183-94. DOI:10.1016/0091-7435(87)90082-x.
23. Consolidated Standards of Reporting Trials (CONSORT 2010) Statement - Methods for additional analyses, such as subgroup analyses and adjusted analyses
<http://www.consort-statement.org/checklists/view/32--consort-2010/97-additional-analyses> Accessed 2021, 17.Aug
24. Zylka-Menhorn V. Onkologische Studien: Ringen um „optimale“ Endpunkte. *Dtsch Arztebl*. 2017; 114(11)
25. Yamazakia M. Effect of obesity on the effectiveness of hormonal contraceptives: an individual participant data meta-analysis. *Contraception*. 2015; 92:445-452.
26. Boulesteix AL, Janitza S, Kruppa J, Koenig IR. Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. *WIREs Data Min Knowl Discov*. 2012; 2:493–507.

27. Sievert, Carson. Interactive web-based data visualization with R, plotly, and shiny. CRC Press, 2020.
28. Kuhn M, Johnson K. Applied predictive modeling. 1st ed. Springer. 2013
29. Sahlmann J. Divide et calcula - Ein SAS-Makro zur systematischen explorativen Berechnung von Subgruppen (https://www.ksfe-ev.de/2020/abstracts/Divide_et_calcula_Ein_SAS_Makro.pdf; https://www.ksfe-ev.de/2020/programm_donners-tag.html). KSFE; Wiesbaden 5.3.2020. Shaker; 1. Edition ISBN-10 3844075895
30. Tanniou J, van der Tweel I, Teerenstra S, Roes KCB. Estimates of subgroup treatment effects in overall nonsignificant trials: To what extent should we believe in them? *Pharma Stat.* 2017; 16: 280-295.
31. Liu W, Lin S, Piegorsch WW. Construction of Exact Simultaneous Confidence Bands for a Simple Linear Regression Model. *International Statistical Review.* 2008; 76 (1): 39–57. DOI:10.1111/j.1751-5823.2007.00027.x
32. Bluhmki T, Schmoor C, Dobler D, Pauly M, Finke J, Schumacher M, Beyersmann J. A wild bootstrap approach for the Aalen–Johansen estimator. 2018; *Biometrics*, 74(3), 977-985.
33. Tanniou J, van der Tweel I, Teerenstra S, Roes KC. Subgroup analyses in confirmatory clinical trials: time to be specific about their purposes. *BMC Med Res Methodol.* 2016; 16:20. DOI:10.1186/s12874-016-0122-6

Eidesstattliche Versicherung

„Ich, Christoph Muysers, versichere an Eides statt durch meine eigenhändige Unterschrift, dass ich die vorgelegte Dissertation mit dem Thema: „Innovativer Ansatz zur simultanen Analyse zahlreicher Subgruppen und Identifikation von Subgruppeneffekten in klinischen und epidemiologischen Studien“ (Innovational Approach to Simultaneously Analyze Numerous Subgroups and to Identify Subgroup Effects in Clinical and Epidemiological Studies) selbstständig und ohne nicht offengelegte Hilfe Dritter verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel genutzt habe.

Alle Stellen, die wörtlich oder dem Sinne nach auf Publikationen oder Vorträgen anderer Autoren/innen beruhen, sind als solche in korrekter Zitierung kenntlich gemacht. Die Abschnitte zu Methodik (insbesondere praktische Arbeiten, Laborbestimmungen, statistische Aufarbeitung) und Resultaten (insbesondere Abbildungen, Graphiken und Tabellen) werden von mir verantwortet.

Ich versichere ferner, dass ich die in Zusammenarbeit mit anderen Personen generierten Daten, Datenauswertungen und Schlussfolgerungen korrekt gekennzeichnet und meinen eigenen Beitrag sowie die Beiträge anderer Personen korrekt kenntlich gemacht habe (siehe Anteilserklärung). Texte oder Textteile, die gemeinsam mit anderen erstellt oder verwendet wurden, habe ich korrekt kenntlich gemacht.

Meine Anteile an etwaigen Publikationen zu dieser Dissertation entsprechen denen, die in der untenstehenden gemeinsamen Erklärung mit dem/der Erstbetreuer/in, angegeben sind. Für sämtliche im Rahmen der Dissertation entstandenen Publikationen wurden die Richtlinien des ICMJE (International Committee of Medical Journal Editors; www.icmje.org) zur Autorenschaft eingehalten. Ich erkläre ferner, dass ich mich zur Einhaltung der Satzung der Charité – Universitätsmedizin Berlin zur Sicherung Guter Wissenschaftlicher Praxis verpflichte.

Weiterhin versichere ich, dass ich diese Dissertation weder in gleicher noch in ähnlicher Form bereits an einer anderen Fakultät eingereicht habe.

Die Bedeutung dieser eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unwahren eidesstattlichen Versicherung (§§156, 161 des Strafgesetzbuches) sind mir bekannt und bewusst.“

Datum

Unterschrift

Anteilserklärung an den erfolgten Publikationen

Christoph Muysers hatte folgenden Anteil an den folgenden Publikationen:

Publikation 1:

Muysers C, Dmitrienko A, Kulmann H, Kirsch B, Lippert S, Schmelter T, Schulz A, Mentenich N, Schmitz H, Schaefers M, Meinhardt G, Keil T, Roll S. A Systematic Approach for Post Hoc Subgroup Analyses With Applications in Clinical Case Studies. *Ther Innov Regul Sci.* 2020; 54, 507–518, DOI:10.1177/2168479019853782

Beitrag im Einzelnen zur Publikation 1:

Vorverarbeitung der Datenbasis zu den beiden klinischen Studien sowie Simulation einer weiteren Studie; Aufbereitung aller Datenstrukturen für das R-Paket ‚subscreen‘; Durchführung aller Analysen im R-Paket ‚subscreen‘ sowie aller weiteren Analysen, die in der Publikation berichtet wurden.

Publikation 2:

Muysers C, Messina F, Keil T, Roll S. A novel concept of screening for subgrouping factors for the association between socioeconomic status and respiratory allergies. *Journal of Exposure Science & Environmental Epidemiology* 2022; 32, 295–302, DOI:10.1038/s41370-021-00365-x

Beitrag im Einzelnen zur Publikation 2:

Vorverarbeitung der Datenbasis aus den Daten der epidemiologischen Studie des Robert Koch-Instituts sowie Aufbereitung der Datenstruktur für das R-Paket ‚subscreen‘ (u.a. Kategorisierung stetiger Variablen); Selektion der potenziell relevanten Einflussfaktoren und deren deskriptive Aufbereitung zur Auswahl und Diskussion in Zusammenarbeit mit den Koautoren; Durchführung aller Analysen im R-Paket ‚subscreen‘ sowie aller weiterer Analysen, die in der Publikation berichtet wurden, mit Ausnahme des zusätzlichen Machine-Learning basierten Algorithmus ‚Random Forest‘ (durchgeführt vom Koautor Fabrizio Messina).

Beitrag zu beiden Publikationen:

Der Promovend extrahierte die relevanten Ergebnisse aus der statistischen Analyse zur Diskussion für die Koautoren (Epidemiologen und Epidemiologinnen, Kliniker:innen, weitere Statistiker:innen). Er erstellte alle Abbildungen und Tabellen für beide Publikationen mit der Ausnahme von Abbildung 4 der zweiten Publikation. Diese Abbildung 4 entstand aus der ‚Random Forest‘ Analyse des Koautors Fabrizio Messina. Der Promovend erstellte den ersten Entwurf der Manuskripte vollständig (mit der genannten Ausnahme von Abbildung 4 der zweiten Publikation) und arbeitete im weiteren Verlauf die Kommentare der Koautoren ein. Er finalisierte die Manuskripte und führte die Einreichungs- und Publikationsprozesse durch. Im Rahmen der Peer-Review-Prozesse erstellte er eigenständig die Entwürfe der Antworten an die Reviewer, beantwortete und arbeitete alle relevanten Vorschläge in die Manuskripte und Point-to-point-Antworten ein und formulierte Anschreiben zum Wiedereinreichen der Manuskripte.

Unterschrift, Datum und Stempel der erstbetreuenden Hochschullehrerin

Unterschrift des Doktoranden

Auszug aus der Journal Summary List

Publikation 1 (Manuskript Januar 2019 eingereicht):

Therapeutic Innovation & Regulatory Science

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
18	Health Information Management Journal	320	1.742	0.000390
19	JOURNAL OF CLINICAL PRACTICE EVALUATION IN INTERNATIONAL JOURNAL OF TECHNOLOGY ASSESSMENT IN HEALTH CARE	4,039	1.536	0.005120
20	Applied Clinical Informatics	2,143	1.418	0.002140
21	Informatics for Health & Social Care	664	1.306	0.002050
22	CIN-COMPUTERS INFORMATICS NURSING	285	1.218	0.000470
23	METHODS OF INFORMATION IN MEDICINE	836	1.029	0.001120
24	Biomedical Engineering- Biomedizinische Technik	1,330	1.024	0.001760
25	Therapeutic Innovation & Regulatory Science	1,007	1.007	0.001320
26	Regulatory Science	371	0.901	0.001600

Copyright © 2019 Clarivate Analytics

Journal Data Filtered By: **Selected JCR Year: 2018** Selected Editions: SCIE, SSCI
 Selected Categories: "MEDICAL INFORMATICS"
 Selected Category Scheme: WoS
Gesamtanzahl: 26 Journale

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
1	JOURNAL OF MEDICAL INTERNET RESEARCH	13,602	4.945	0.030580
2	JMIR mHealth and uHealth	2,576	4.301	0.007920
3	JOURNAL OF THE AMERICAN MEDICAL ASSOCIATION	9,319	4.292	0.019480
4	IEEE Journal of Biomedical and Health Informatics	4,082	4.217	0.010320
5	ARTIFICIAL INTELLIGENCE IN MEDICINE	2,462	3.574	0.002960
6	COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE	7,147	3.424	0.009350
7	JMIR Serious Games	269	3.351	0.000660
8	JMIR Medical Informatics	384	3.188	0.001480
9	JOURNAL OF BIOMEDICAL INFORMATICS	7,431	2.950	0.010300
10	MEDICAL DECISION MAKING	5,281	2.793	0.009000
11	INTERNATIONAL JOURNAL OF MEDICAL INFORMATICS	4,765	2.731	0.006720
12	JOURNAL OF MEDICAL SYSTEMS	4,680	2.415	0.006220
13	STATISTICAL METHODS IN MEDICAL RESEARCH	4,156	2.388	0.012230
14	Health Informatics Journal	691	2.297	0.001450
15	BMC Medical Informatics and Decision Making	3,578	2.067	0.008490
16	MEDICAL & BIOLOGICAL ENGINEERING & COMPUTING	5,904	2.039	0.004380
17	STATISTICS IN MEDICINE	24,925	1.847	0.034040

Publikation 2 (Manuskript Dezember 2020 eingereicht):

Journal of Exposure Science & Environmental Epidemiology

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
20	EPIDEMIOLOGIC REVIEWS	3.838	6.222	0.002790
21	Travel Medicine and Infectious Disease	5.034	6.211	0.003430
22	Environmental Health	7.667	5.984	0.009280
23	INTERNATIONAL JOURNAL OF HYGIENE AND ENVIRONMENTAL HEALTH	7.425	5.840	0.008110
24	INDOOR AIR	7.488	5.770	0.005780
25	DRUG SAFETY	6.817	5.608	0.006840
26	Journal of Exposure Science and Environmental Epidemiology	5.227	5.583	0.004600
27	BMJ Global Health	4.992	5.568	0.018080
28	Evolution Medicine and Public Health	758	5.425	0.001880
29	International Journal of Transgenderism	1.381	5.333	0.002130
30	AMERICAN JOURNAL OF PREVENTIVE MEDICINE	28.400	5.043	0.037310
31	SCANDINAVIAN JOURNAL OF WORK ENVIRONMENT & HEALTH	6.700	5.024	0.004530
32	JOURNAL OF ADOLESCENT HEALTH	20.914	5.012	0.024870
33	AMERICAN JOURNAL OF EPIDEMIOLOGY	43.617	4.897	0.031670
34	Antimicrobial Resistance and Infection Control	3.629	4.887	0.008170
35	EPIDEMIOLOGY	16.437	4.822	0.018740
36	Health Reports	1.781	4.786	0.001770
37	Clinical Epidemiology	4.754	4.780	0.010780
38	PALLIATIVE MEDICINE	7.332	4.762	0.009100
39	SOCIAL SCIENCE & MEDICINE	57.968	4.634	0.052130
40	ENVIRONMENTAL GEOCHEMISTRY AND HEALTH	6.557	4.609	0.004550
41	GeoHealth	381	4.529	0.001190

Journal Data Filtered By: **Selected JCR Year: 2020 Selected Editions: SCIE**
Selected Categories: "PUBLIC, ENVIRONMENTAL and OCCUPATIONAL HEALTH" Selected Category Scheme: WoS
Gesamtanzahl: 203 Journale

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
1	MMWR Surveillance Summaries	3,806	58.769	0.012980
2	MMWR Recommendations and Reports	3,288	55.857	0.011230
3	Lancet Global Health	16,662	26.763	0.051650
4	Annual Review of Public Health	9,927	21.981	0.013560
5	Lancet Public Health	5,793	21.648	0.014840
6	Lancet Planetary Health	2,540	19.173	0.008590
7	MMWR-MORBIDITY AND MORTALITY WEEKLY REPORT	31,127	17.586	0.078080
8	BULLETIN OF THE WORLD HEALTH ORGANIZATION	19,297	9.408	0.015980
9	AMERICAN JOURNAL OF PUBLIC HEALTH	51,398	9.308	0.056870
10	ENVIRONMENTAL HEALTH PERSPECTIVES	48,245	9.031	0.028380
11	JOURNAL OF TRAVEL MEDICINE	5,260	8.490	0.004900
12	EUROPEAN JOURNAL OF EPIDEMIOLOGY	11,622	8.082	0.017840
13	TOBACCO CONTROL	10,279	7.552	0.015380
14	INTERNATIONAL JOURNAL OF EPIDEMIOLOGY	30,944	7.196	0.045590
15	Current Environmental Health Reports	1,967	7.122	0.004720
16	ENVIRONMENTAL RESEARCH	28,576	6.498	0.034280
17	JOURNAL OF CLINICAL EPIDEMIOLOGY	36,224	6.437	0.028380
18	JOURNAL OF TOXICOLOGY AND ENVIRONMENTAL HEALTH-PART B-CRITICAL REVIEWS	2,226	6.393	0.001410
19	Current Pollution Reports	1,107	6.373	0.001940

Druckexemplare der Publikationen

Publikation 1:

Muysers C, Dmitrienko A, Kulmann H, Kirsch B, Lippert S, Schmelter T, Schulz A, Mentenich N, Schmitz H, Schaefers M, Meinhardt G, Keil T, Roll S. A Systematic Approach for Post Hoc Subgroup Analyses With Applications in Clinical Case Studies. *Ther Innov Regul Sci.* 2020; 54, <https://doi.org/10.1177/2168479019853782>

Publikation 2:

Muysers C, Messina F, Keil T, Roll S. A novel concept of screening for subgrouping factors for the association between socioeconomic status and respiratory allergies. *Journal of Exposure Science & Environmental Epidemiology* 2022; 32, 295–302
DOI:10.1038/s41370-021-00365-x

ARTICLE OPEN



A novel concept of screening for subgrouping factors for the association between socioeconomic status and respiratory allergies

Christoph Muysers^{1,2}, Fabrizio Messina³, Thomas Keil^{1,4,5} and Stephanie Roll¹

© The Author(s) 2021

BACKGROUND: The new subgroup screening tool “subscreen” aims to understand the unclear and complex association between socioeconomic status (SES) and childhood allergy. This software R package has been successfully used in clinical trials but not in large population-based studies.

OBJECTIVE: To screen and identify subgrouping factors explaining their impact on the association between SES and respiratory allergies in childhood and youth.

METHODS: Using the national German childhood and youth survey dataset (KiGGS Wave 2), we included 56 suspected subgrouping factors to investigate the association between SES (low vs. high) and allergic rhinitis and/or asthma in an exploratory manner. The package enabled a comprehensive overview of odds ratios when considering the SES impact per subgroup and analogously all disease proportions per subgroup.

RESULT: Among the 56 candidate factors, striking subgrouping factors were identified; e.g., if mothers were younger and in the low SES group, their children had a higher risk of asthma. In addition children of the teen’s age were associated with increased risks in the low SES group. For the crude proportions, factors such as (parental) smoking or having had no “contact with farm animals” were identified as strong risk factors for rhinitis.

SIGNIFICANCE: The “subscreen” package enabled the detection of notable subgroups for further investigations exemplarily for similar epidemiological research questions.

Keywords: “Subscreen” R package; Epidemiology survey; Asthma; Rhinitis

Journal of Exposure Science & Environmental Epidemiology (2022) 32:295–302; <https://doi.org/10.1038/s41370-021-00365-x>

INTRODUCTION

Since the consistency of study results across all participants cannot generally be assumed, analyses considering subgroups of participants are frequently used and discussed, with remarkably varying approaches [1–4]. However, conclusions about the true relationship between factors and a study endpoint can remain incomplete if not analyzed comprehensively. With increasing computer power and statistical state-of-the-art methodology, even a huge list of subgrouping factors can be investigated in an explorative manner. Various options for subgroup analyses, including graphical approaches, were presented in a systematic review of Ballarini et al. [5]. Among these, the subgroup screening with the R package “subscreen” [6] has demonstrated high efficiency and usefulness in the exploration of otherwise overlooked factors in clinical trials [7]. Any kind of statistical analysis, which was for instance pre-specified in an analysis plan such as difference in means or time-to-event analyses, can be done with the “subscreen” package for all subgroups simultaneously in one go and comprehensively

visualized for further investigations. In contrast to e.g., statistical models with interaction terms, the basic approach of “subscreen” is to repeat the primary analysis for each subgroup. With the novel concept, we recommend fostering discussions in interdisciplinary teams (e.g., epidemiologists, clinicians, and statisticians) focusing on striking subgroups and evaluate interactively the relevance from a medical and statistical point of view.

From 2003 to 2006, a nationally representative survey was conducted on children and adolescents (later on often “children” for easier reading). This “German health interview and examination survey for children and adolescents (KiGGS)” aimed to improve the information available on the health of the up and -coming generation in Germany [8]. The subsequent KiGGS Wave 2 aimed to conduct a follow-up survey and had a data release in 2019 [9]. A huge amount of demographic, health-related, laboratory and environmental parameters (called “factors” in the following) including the socioeconomic status (SES) were collected [10]. Many potential risk and protective factors have already been

¹Institute of Social Medicine, Epidemiology and Health Economics, Charité—Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany. ²Statistics and Data Insights, Bayer AG, Berlin, Germany. ³Data Science, Bayer Pharmaceutical PLC, Reading, UK. ⁴Institute for Clinical Epidemiology and Biometry, Julius-Maximilians-University Würzburg, Würzburg, Germany. ⁵State Institute for Health, Bavarian Health and Food Safety Authority, Bad Kissingen, Germany. email: christoph.muysers@bayer.com

Received: 8 December 2020 Revised: 28 June 2021 Accepted: 29 June 2021
Published online: 12 July 2021

investigated in various studies worldwide including socio-economic variables [11, 12] indicating a higher risk for asthma and rhinitis in the case of a higher SES [13]. On the other hand, the results of KiGGS Wave 2 reported by Kuntz et al. [14] demonstrated that children with a low SES have a poorer level of general health and were more affected by asthma compared to their peers with a higher SES.

Even though the SES impact was less profound in the 12-month prevalence of bronchial asthma and allergic rhinitis compared to general health, we aimed to assess the scientific question of whether the effect of SES on asthma, rhinitis, or both is consistent among children when considering subgrouping factors. Additionally, the impact of these subgrouping factors on the crude proportions of these allergic diseases was investigated. For these aims, the R package “subscreen” as a graphical and analytical tool was used in this large epidemiological data set.

METHODS

R package “subscreen” to investigate SES (in-)consistency

The second version of the “subscreen” package was illustrated with examples of clinical trials in Muysers et al. [7]. Several enhancements and additions were implemented with version 3 [6]. Basically, the “subscreen” package requires the specification of a user-defined statistical model which can be applied to one group, e.g., for time-to-event or proportion analyses, or to two groups, e.g., for the difference in means. In our case, the odds ratio (OR) describing the association between the disease (outcome) and the two SES groups (“low” or “high”) was calculated using logistic regression. The disease outcome (present or not) was asthma, rhinitis, or both. For each of the three outcomes and within each factor level combination (subgroup) the ORs were individually estimated.

This would be possible for all combinations of factor levels but for interpretational reasons, it was restricted to combinations of up to two-factor levels only. For example, 2 sex levels combined with 4 age groups resulted in 8 subgroups. Subsequent SES comparisons in each of these 8 groups in addition to the 6 single factor levels of gender (2) and age group (4), resulted in 14 ORs derived for the investigation and repeatedly used for each of the three disease outcomes. Empty cells are the result of a constellation of factor levels without children and consequently no OR were calculated.

For our presented purpose, a “striking factor” was defined on subjective assessments supported by various indicators. In line with the exploratory approach, we intentionally excluded the calculation of *p*-values and confidence intervals to avoid strict categorizations such as conventional significant and non-significant factors, instead of leaving room for interdisciplinary discussions of clinical relevance.

R package “subscreen” to investigate crude proportions

To broaden the insight into the relationship of the investigated factors and the allergic diseases, crude proportions of all disease outcomes ignoring the SES status were additionally derived using the “subscreen” package. One element of the “subscreen” package is the variable importance (VIMP), which is based on the machine learning algorithm random forest [15–17]. VIMP was implemented to compensate for the intentional lack of confirmatory testing of subgrouping factors in “subscreen” [6]. While the VIMP feature within the “subscreen” package offers a ranked factor list of the importance only, to offer better insight, a more comprehensive analysis and result reporting outside the “subscreen” package was conducted in R. The random forest runs a series of decision trees, collect them and average their characteristics. An additional feature of the Random Forest is that it can collect the out-of-bag (OOB) sample [18]: at each tree iteration, the random forest leaves out observations and uses them to perform cross-validation and improve accuracy in the forest. VIMP was used to determine the factor (predictor) contribution in predicting the disease outcome (dependent variable) of interest. Minimal depth [19] was also used to compare the VIMP findings. Minimal Depth identifies the high impact predicting variables as the ones which partition the largest samples of the population. For the analysis, training and testing datasets were created, respectively using 80% and 20% of the observations.

Exemplary further investigation of striking factors

While the explorations for both the SES-related and the crude proportions-related screening supported the finding of striking factors, a conclusion

about their causal impact is not possible because of the post-hoc type of analyses and multiplicity reasons [20, 21]. Consequently, it is recommended to conduct further investigations, e.g., after the setting of a hypothesis, focusing on such striking factors. Since the term “striking” is in contrast to a “significance” approach with no strict classification, different hypotheses, and factor selections are obvious. For the purpose of this paper, a multivariable logistic regression analysis was derived exemplarily for a certain subjective selection of factors including interaction terms with SES to derive ORs and 95% CIs. As sensitivity analysis stepwise forward, backward, and bidirectional selection were conducted using the Akaike’s information criterion (AIC) as described in Harrell [22].

Software references

SAS® version 9.4 and R [<https://cran.r-project.org/>] version 3.6.1 were used for the data set-up and supportive analyses. The R package “subscreen” version 3.0 was used for the subgroup screening [6]. The random forest was implemented by using the “randomForestSRC” R package [23].

Study data

The KiGGS wave two studies comprised a nationwide, representative cross-sectional sample of children the age of 0–17 years from the Robert Koch Institute (RKI) [24]. A comprehensive overview of data architecture, collection, and management has been published [25]. Of the 15,023 children in the KiGGS Wave 2 data set, 4686 children were available for the analysis. Children aged 0–3 years (here an allergic disease diagnosis is mainly not reliable), children having a medium SES, children with missing information for asthma, rhinitis, or SES were excluded from the analysis dataset.

Missing values were handled as follows: (i) if asthma or rhinitis disease status or the SES category was missing, the child was withdrawn from the analyses, (ii) if a subgroup factor level (or the combination of two levels) did not have observations, it was set to the new category “no data” in order to avoid too many empty subgroups and to become aware of potential systematic missingness.

In the KiGGS survey, weighting variables were calculated to compensate for imbalances in certain variables with respect to the general German population of children and adolescents [25]. Since it is not the aim of the present analysis to estimate a population prevalence and because the sampling design is assumed to be non-informative, i.e., survey selection probabilities are not correlated with the outcome of factors [26], the weights were not used in our analyses.

Outcomes

Three outcomes were investigated: (i) current asthma symptoms only (i.e., in last 12 months), (ii) current allergic rhinitis symptoms only (in last 12 months), (iii) both current asthma and current rhinitis (in last 12 months). This approach aimed to minimize inferences with overlapping associations for the three groups. The common control group included children who had neither current asthma nor current rhinitis. A “current” disease refers to the time frame “within the last 12 months” or “12-month prevalence” based on the KiGGS survey terminology [24]. The term “only” indicates the absence of the alternate disease. For improved readability throughout this article, “asthma” describes “current asthma only” and analogously for “rhinitis”. For each outcome the dichotomized status “yes” or “no” was taken from the source data of the RKI [24].

SES

The assessment of the SES is based on a broad assessment of the parents’ education, occupational status, and income [10, 27]. From this, participants were categorized into three groups: “low”, “medium”, or “high” SES. In order to concentrate on presumed larger effects and for better usability of the ‘subscreen’ package, we compared children with low vs. high SES only in all analyses.

Factors and subgroups

The selection of potentially relevant factors was not only based on the literature but also on any suspected association of the disease with SES and of their availability in the KiGGS dataset. In total, 87 factors were selected, including information on general health, quality of life [28], allergies, smoking, socio-demography and earnings, and breastfeeding. For simplicity, some factors were combined into new factors. Specifically, continuous factors or factors with numerous levels were categorized into

Table 1. Frequencies and percentages of disease outcomes (asthma and/or rhinitis) with crude odds ratios (OR) and Wald's 95% confidence intervals (CI) for each socioeconomic status (SES) level.

	Asthma (n = 78)	Rhinitis (n = 361)	Asthma & Rhinitis (n = 72)	Control Group (n = 4175)	Total (n = 4686)
SES low n (%)	35 (44.9 %)	104 (28.8%)	24 (33.3%)	1361 (32.6%)	1524 (32.5%)
SES low n (%)	43 (55.1 %)	257 (71.2%)	48 (66.7%)	2814 (67.4%)	3162 (67.5%)
OR (95% CI)	1.68 (1.07–2.64)	0.84 (0.66–1.06)	1.03 (0.63–1.70)	–	

few classes since the “subscreen” approach requires an ordinal or a nominal scale [7]. As a result, 56 factors comprising 221-factor levels were available. For a comprehensive overview, Supplementary Table S1 shows all used and modified factors. A “subgroup” is used to denote one specific level of a factor (e.g., “girls”) or one combination of two-factor levels (e.g., “girls of the youngest age group”).

RESULTS

Among the 4686 children, 78 (1.7 %) reported asthma only, 361 (7.7 %) rhinitis only, 72 (1.5 %) both asthma and rhinitis and 4175 (89.1%) neither disease. Crude ORs between outcomes and SES are presented in Table 1.

Subgroup screening regarding SES

For the subgroup screening, 221 1-factorial subgroups and 22,704 1–2 factorial subgroups (combinations of the 221-factor levels) were considered based on the 56 selected factors. Fig. 1 shows the surface of the “subscreen” package for the 1-level subgroups, with a first striking result of the factor “mother’s age at childbirth”. More than one million 1–3 factorial subgroups could be technically handled; however, already the 1–2 factorial subgroups in Fig. 2 offered a good display of the subgroup dispersion while 1–3 factorial subgroups would result in too much overlap of dots. On the other hand, striking subgroups consisting of a combination of three or even more factor levels are rather difficult to interpret and based often on very small group sample sizes; three or more factorial subgroups were thus omitted.

Throughout this subgroup screening, ORs greater than 1 indicates a stronger association of the children with low SES with the disease outcome in the respective subgroup. It happened that no observations occurred for a specific disease outcome and SES within a children subgroup. Such constellations but also subgroups with small numbers triggered extreme OR values. Such ORs were calculated but were truncated in the visualization in the “subscreen” tool at values 0.3 and 15 (individual user-specification) for an optimized view for all outcomes outside this range.

The new “interaction plot” feature in version 3.0 of “subscreen” [6] was helpful in identifying the factor “mother’s age at childbirth” as striking. The corresponding line plot at the top right in Fig. 1 shows a clear decreasing trend for the asthma outcome (n.b. the “No Data” group should not be considered for the trend). As age increases, the ORs decrease, reflecting a higher risk for the children of young women with a low SES to have asthma. In Fig. 1, the table below the dot plot shows simultaneously the numerical results and trend for the disease group both asthma and rhinitis (lgAR12). In addition, the rhinitis results (lgR12) are shown, though an OR trend is not available.

The line plots for the other disease outcomes and factors can be generated in a user-friendly way (not shown). For instance, we noticed high ORs for the low SES group for the asthma outcome for 11–13 years old children group (OR = 6.58) and for rhinitis in the 14–17 years old children group (OR = 1.13, whereas all other age groups have a consistent similar OR equal or below 0.62 for rhinitis).

Fig. 2 shows the extension of the subgroup screening to 1–2 level factorial subgroups, represented by 22,704 dots (cropped of sidebars for simplicity). A funnel shape is formed by the dispersion

of the subgroup dots around the blue line representing overall asthma OR. Low sample size groups with high variation are shown on the left, whilst high sample size groups with low variation are shown on the right side of the graph. Deviations from this funnel shape indicate striking subgroups.

In Fig. 2, the red dot represents the high “Education level of parents” combined with no “current contact with farm animals”. An OR of 7.8 indicating a remarkable higher risk for asthma in children with a low SES is observed.

Other relevant factors (data not shown) interfering with the relationship of SES and all three disease outcomes were the allergy factors SX1-screening, air/pollen, and food allergy. These factors showed high ORs in either direction for different levels but lack a clear trend-making hypothesis generation difficult. In addition, the subgroup “Berlin” residency showed a lower risk for the low SES group for all disease outcomes; though, this subgroup had only a few observations.

Overall, the subgroup screening showed larger OR dispersion for the subgroups for asthma compared to rhinitis indicating a general stronger impact of the various factors on the relationship between asthma and SES than on rhinitis and SES. This can easily be seen with the “Compare” feature of the “subscreen” package showing a “Bubble Plot” (Supplementary Fig. S1). The OR dispersion of subgroups of “both asthma and rhinitis” is in-between asthma’s and the rhinitis OR dispersion (data not shown).

Subgroup screening investigating crude disease proportions

Crude proportions, independent of the SES interaction and the OR plotting, were simultaneously visualized for the subgroups for asthma and for rhinitis using the “Compare” feature. As seen in Fig. 3 the “yes”-level of the factor “Any smoke exposure” (green dots) shows a remarkably higher proportion for rhinitis (9.5%) but only a moderately increased proportion for asthma compared to their averages of 8.0% and 1.8%, respectively. In a smaller subset, “Active smoke exposure” (red dots) showed for rhinitis an extremely higher proportion (12%) if the corresponding youths were smoking. Similar results were observed for the factor “Smoking during breastfeeding” as reported by their mothers (data not shown).

Furthermore, the “no”-level for the factor “contact with farm animals in the age of 0–6 years” resulted in a remarkably high proportion of 10.3% of children having rhinitis (not highlighted in the figure).

Random forest

Some variables were highly correlated and were thus excluded from the analyses (“correloplot” in Supplementary Fig. S2). For each disease outcome, a separate Random Forest model was run. Missing observations were omitted and not imputed because a lower OBB error rate estimate and a higher variance explained were observed in a sensitivity analysis.

The most important variables regarding VIMP identified by the Random Forest associated with rhinitis were “General Health”, “Active smoke exposure”, “Atopic Dermatitis” and “Age group” (Fig. 4, left). However, the Random Forest does not give an indication of the direction of the effect. Minimal depth showed for rhinitis similar results to VIMP, whereas the highest impact variables to partition the data were found to be “Animal contact between 0 and 6 years” and “Atopic Dermatitis” (Fig. 4, right). The

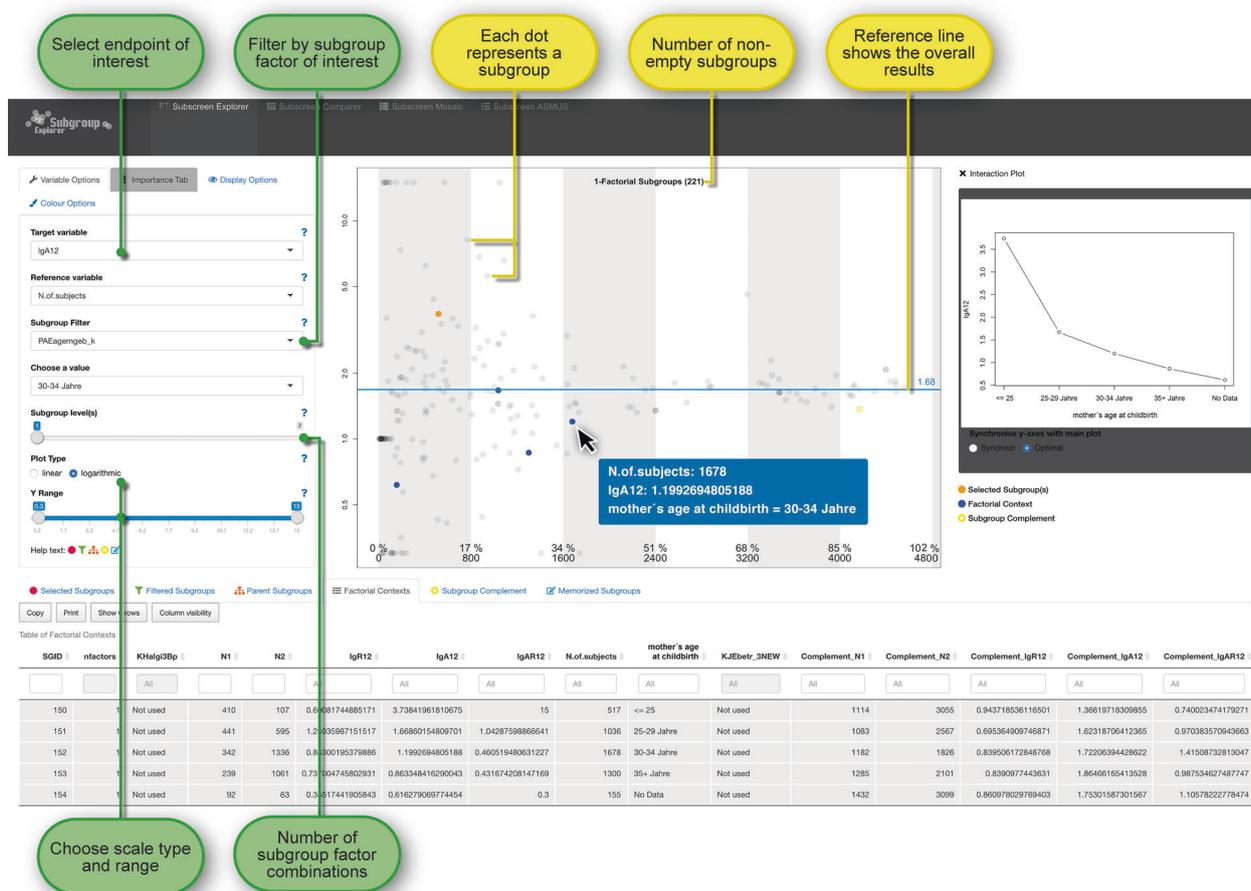


Fig. 1 Each subgroup is represented by a dot in the central plot where the position is determined by the subgroup's sample size (horizontal axis) and statistical measure of the treatment effect (vertical axis). The blue dots in the plot and numerical tabulated results below the plot for the factor "mother's age at childbirth" appear after clicking the "Factorial Context" tab for a previously actively selected subgroup (orange). While the factorial context with blue dots highlights the individual levels of the factor "mother's age at childbirth", the yellow complementary group highlights the subgroup of all combined subjects excluding the subjects belonging to the selected (orange) subgroup. The interaction plot (line plot) at the right can be extended with another factor if desired.

corresponding results for asthma and both asthma and rhinitis are shown in Supplementary Figs. S3 and S4, respectively.

Exemplary further investigation of striking factors

A multivariable logistic regression model was run for the factors "Mother's age at birth", "Age group", "Current contact with farm animals", "Smoking during breastfeeding", and "Residence", including their interaction with SES. Numerically similar results were found for (i) using the forward selection and the full model as well as (ii) using the backward elimination and the bi-directional approach. We thus report only the forward selection and the backward elimination models (Supplementary Table S2) representing all four approaches. The reported p -values for each outcome describe the given model test compared with the intercept-only model.

Asthma ($p = 0.106$): ORs with a nominal p -value below 5% were observed only for the SES interactions of age groups with an OR of 0.20 (0.04–0.98; 95% CI) for "14–17 years" and 0.19 (0.04–0.88) for "7–10 years" where the "11–13 years" group was used as the reference group.

Rhinitis ($p < 0.001$): one OR with a nominal p -value below 5% was 2.22 (1.23–4.01) observed for "Mother's age at birth" for the age group "up to 24 years" with "25–29 years" as reference.

Both asthma and rhinitis ($p < 0.001$): no single OR with a nominal p -value below 5% was seen.

DISCUSSION

Main findings

Our results based on the crude OR indicated a considerably increased asthma risk for children with a low SES, whereas, for rhinitis, we found an increased risk for children with a high SES. No such risk trend was observed for children with both asthma and rhinitis, most probably since the odds were outbalanced by the diverging trends of asthma and rhinitis.

The subgroup screening was applied to find amplifying or weakening factors that might interfere with the association of disease (asthma, rhinitis, or both) with SES. Asthma demonstrated higher variation across all subgrouping factors and more striking factors were identified compared to rhinitis. One finding triggered the hypothesis of a higher risk of children's asthma for less privileged mothers at a younger age. This finding was based on both (i) observed subgroups at the edge of the funnel-shaped plot and (ii) the monotonicity trend in the interaction plot. The combined disease outcome, i.e., both current asthma and rhinitis, showed huge variations (picture not shown), and no clear striking factors were identified.

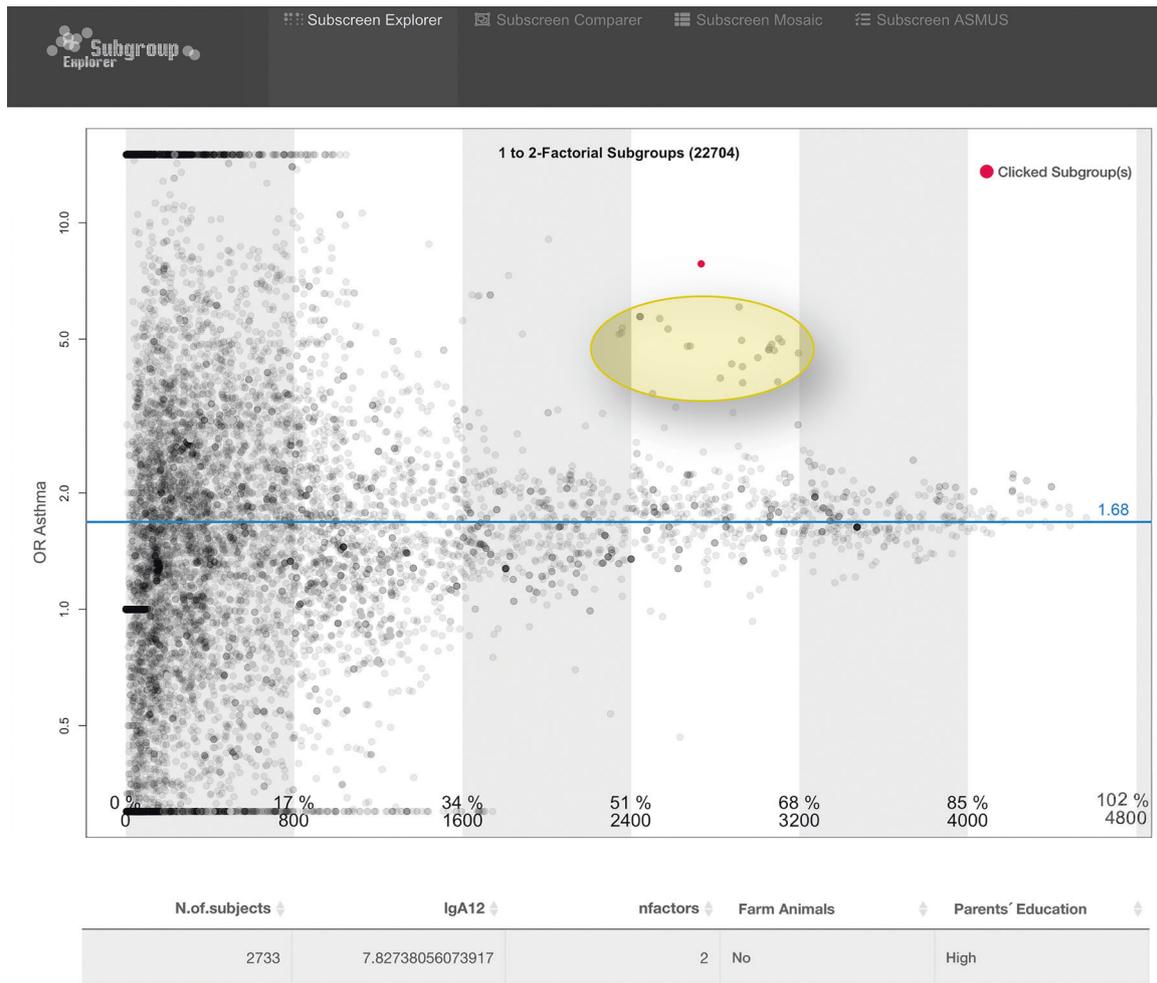


Fig. 2 The 1–2 factorial subgroups for outcome asthma. The degree of dot transparency reflects the number of factor levels: the brighter the dot, the more factors are combined in that group. Darker dots can be also caused by the overlap of multiple subgroups. A visual striking subgroup is actively highlighted (red dot). A dot cluster (figure edited with a yellow oval) appears below the red highlighted subgroup. The striking cluster yellow oval added appears because other factor level combinations' odds ratio, which includes parent's education, are highly correlated with the red highlighted subgroup; while the combination with no "current contact with farm animals" yielded the highest odds ratio in that area.

Furthermore, remarkably increased ORs were observed for the low SES group: (i) for the asthma outcome for 11–13-year old children (OR = 6.58) and (ii) for rhinitis outcome for the 14–17-year old children group (OR = 1.13, whereas all other age groups have an OR equal or below 0.62). A lack of compliance, in terms of appropriate disease management, could be hypothesized for the youths of the low SES group and requires further investigations.

High "Education level of parents" in combination with no "current contact with farm animals" was found to increase the risk of asthma for the low SES group. However, considering the two factors, an interaction is questionable from a content point of view and requires further investigation; furthermore, the factor level of high education as opposed to the low SES group.

The crude proportions were simultaneously inspected with the subgroup screening for asthma and for rhinitis. Smoking in different variants (any or active exposure and smoking during breastfeeding) seems to increase the proportion of children experiencing rhinitis. Similarly, when no "contact with farm animals in the age of 0–6 years" was experienced, a remarkably higher proportion of children having rhinitis was observed. Random forests presented graphically which factors contribute more in predicting the dependent variables. Though not showing the direction of the most important variable found, the random forest was able to confirm striking factors from the subgroup

screening, e.g., for smoking and animal contact-related factors. Furthermore, other factors were derived as important for the prediction of rhinitis, e.g., "General health" or "Atopic Dermatitis". Noteworthy, all three random forest models had a poor predictive ability and showed low variance explained.

Interesting factors from the explorative investigations were used for a statistical adjusted logistic regression model including interaction terms with the SES. For asthma, the forward selection logistic model yielded considerable ORs for the SES interactions with age groups in line with observations for a higher risk for children in the age group 11–13 years and the low SES. A meaningful OR was observed for "Mother's age at birth" in the youngest age group for rhinitis with the forward selection model as already seen in the subgroup screening. Therefore, the logistic regression model partially confirmed what was found in the subgroup screening.

Strengths and limitations

The "subscreen" package had already demonstrated its usefulness and efficiency to detect potentially overlooked factors in clinical trials [7]. It was conveniently extended here to a large population-based epidemiologic study given the tool's property to handle huge data sets with millions of subgroups. The usually lengthy and tedious explorative search for striking subgroups was replaced by an interactive, comprehensive, and coherent screening.

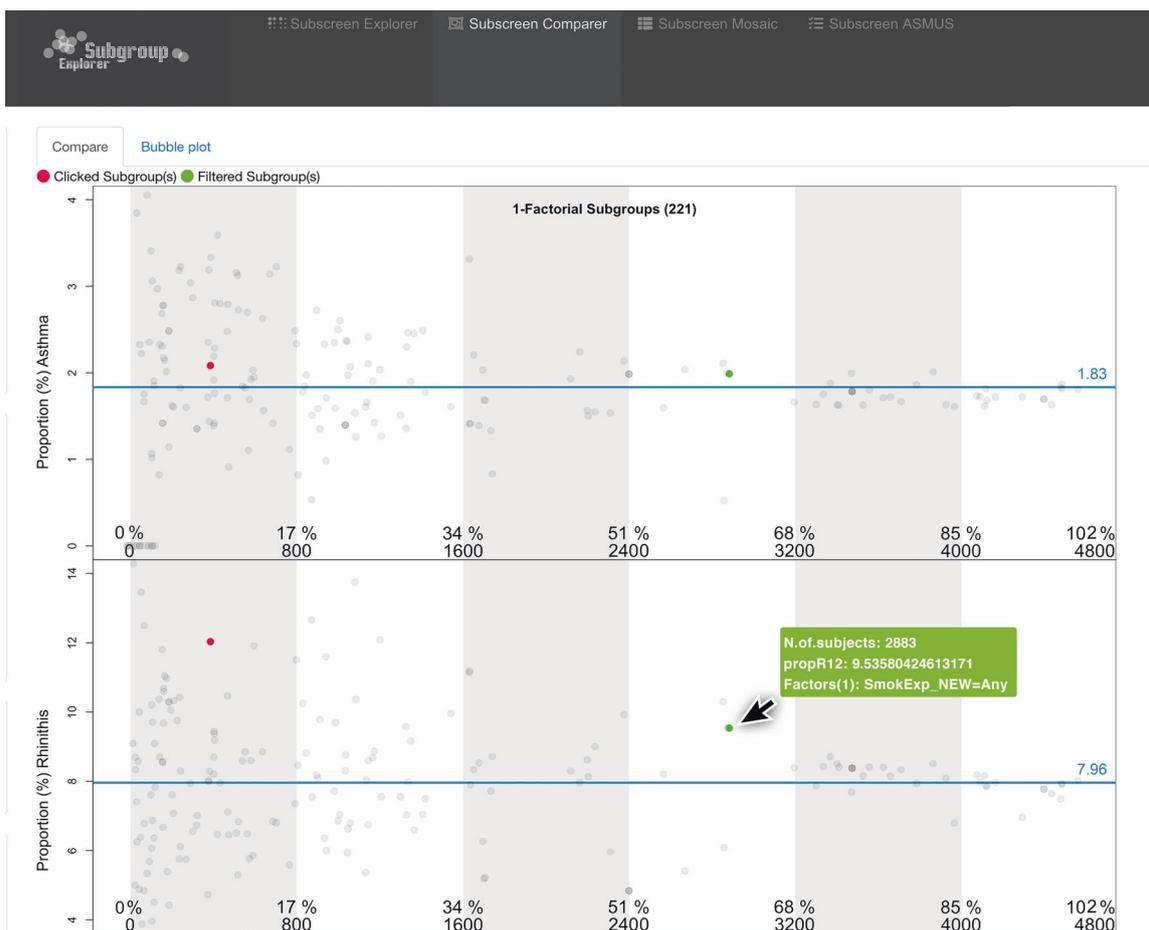


Fig. 3 The proportions of children with asthma (upper panel) and rhinitis (lower panel) are simultaneously shown for all 221-factor levels and compared to the overall proportions (blue lines). Factors “Any smoke exposure” (green dots) and “Active smoke exposure” (red dots) are selected.

The combination of disease outcomes both asthma and rhinitis seemed to increase the dispersion of the investigational outcomes (ORs) but striking factors leading to a hypothesis generation were not detected. This might be due to the outbalancing of the observed diverging effect of risk trends in asthma and rhinitis. While this composite group appears not informative, the disjunction of the disease groups asthma and rhinitis were beneficial to avoid overlapping vanishing effects for the investigations in these two disease outcome groups.

In order to report a prevalence and to make conclusions about the whole target, population weighting is generally recommended. Since the sampling design for the KiGGS survey is assumed to be not informative, the weights were not essential for our analyses [26] and therefore omitted. Furthermore, we did not aim to describe the prevalence in German childhood but to increase the understanding of the data and allow hypothesis generation.

To find further explaining factors for the association between allergy outcomes and SES, a discriminating approach was used. For this, only the low and high SES groups were considered while children of the medium SES were dismissed. A U-shaped distribution, i.e., the extreme outcome of the medium SES group, deemed not existing based on the monotonicity of ORs for asthma in conjunction with SES [14]. Monotonicity of ORs for rhinitis was not observed [14], however, the focus of our investigation was on the screening for additional and interacting factors independent of the OR's monotonicity for the SES groups. Alternatives to dismissing the medium group, such as dichotomization, would imply less comparability with other publications.

Due to missing data and/or few occurrences of diseases within a factor level, e.g., for the factor “Mildew in rooms”, the number of outcome events was very small. These factors were nevertheless kept in the screening in order to detect extreme abnormalities. However, low associations seen in this data do not necessarily mean that no association exists.

CONCLUSIONS

The easy-to-use “subscreen” tool can handle large epidemiologic study data with millions of subgroups fast and efficiently. This has been enhanced in a new version of the tool. It now includes further supportive graphical and analytical features such as interaction plots facilitating also trend analyses. With the novel concept, we propose to make use of this fast and efficient tool in an interdisciplinary team. The common discussion from different fields of expertise of striking subgroups can lead directly to the generation of new hypotheses or confirmation of previous assumptions. This exploratory approach yields more numerical and graphical insight into the data compared to the restricted use of a few selected subgroups.

For rhinitis, our approach confirmed findings for known risk factors (such as smoking) and protective factors (such as contact with farm animals). For rhinitis and for asthma, unexpected findings were seen for factors in conjunction with SES such as the age of both children and mothers. Specifically for asthma, our findings could trigger the hypothesis of a higher risk for children from less privileged, younger mothers. However, further investigations are needed to support this hypothesis.

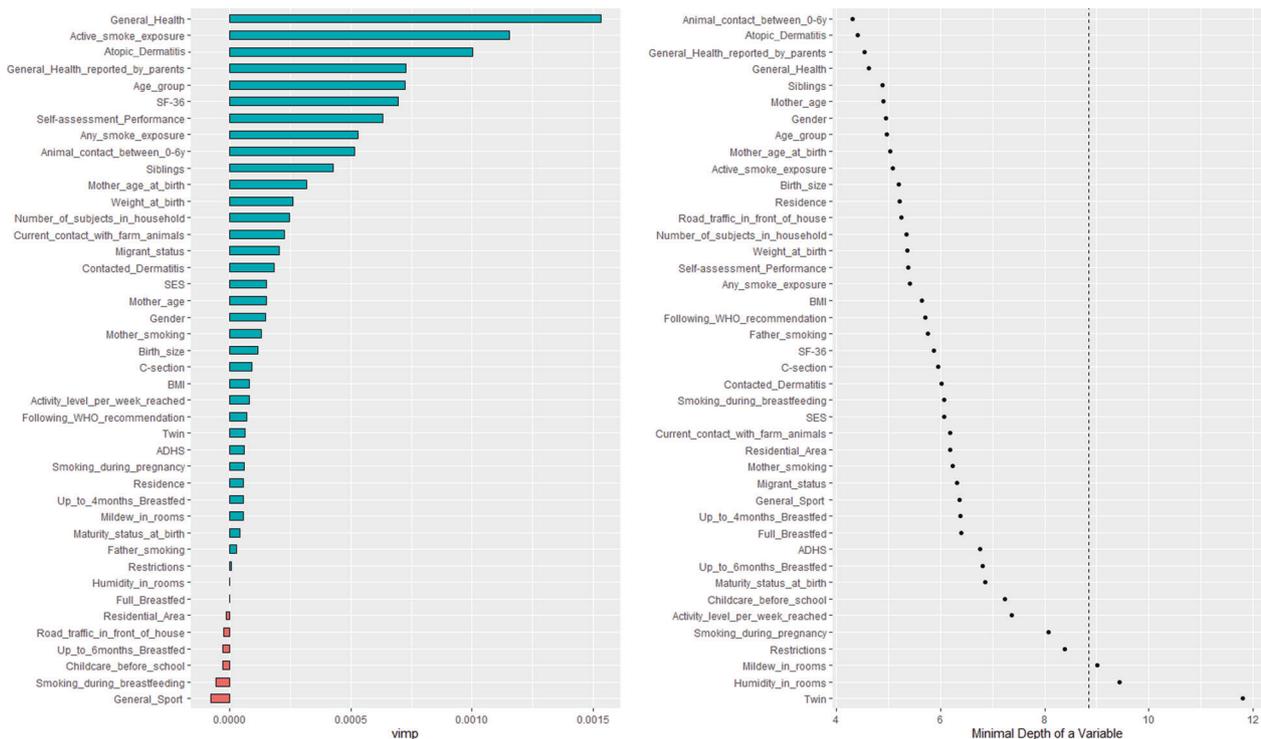


Fig. 4 Variable Importance (VIMP) for the best-fitted random forest (left) is presented for rhinitis. Blue bars indicate the importance of variables (positive VIMP) relative to their lengths, red bars indicate presumably noise variables (negative VIMP). The best combination for the hyper-parameters was found to be $nodelist = 10$, $nsplit = 2$, and $mtry = 3$ with the lowest OBB error estimate of 2%. Minimal depth (right) refers to the topology of the forest. Low minimal depth indicates important variables. The dashed line represents the threshold of the maximum value for variable selection.

In summary, the “subscreen” tool can be applied routinely to clinical and epidemiological study data for future discussions and hypothesis generation requiring further analyses.

DATA AVAILABILITY

Used source data, i.e., The German Health Survey for Children and Adolescents (KiGGS Wave 2), were provided by the Robert Koch Institute, Department of Epidemiology and Health Monitoring [24]; data are available from the Robert Koch Institute.

CODE AVAILABILITY

The software SAS® Version 9.4 and R Version 3.6.1 with the free of charge “subscreen” package version 3.0 [6] was used. Programming code is available for the Journal upon request by CM.

REFERENCES

- Paquette M, Alotaibi AM, Nieuwlaat R, Santesso N, Mbuagbaw L. A meta-epidemiological study of subgroup analyses in Cochrane systematic reviews of atrial fibrillation. *Syst Rev*. 2019;8:241 <https://doi.org/10.1186/s13643-019-1152-z>
- Fletcher J. Subgroup analyses: how to avoid being misled. *BMJ*. 2007;335:96–97. <https://doi.org/10.1136/bmj.39265.596262.AD>
- Dmitrienko A, Muysers C, Fritsch A, Lipkovich I. General guidance on exploratory and confirmatory subgroup analysis in late-stage clinical trials. *J Biopharm Stat*. 2016;26:71–98.
- Pocock SJ, Assmann SF, Enos LE, Kasten LE. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. *Stat Med*. 2002;21:2917–30.
- Ballarini NM, Chiu YD, Koenig F, Posch M, Jaki T. A critical review of graphics for subgroup analyses in clinical trials. *Pharmaceut Stati*. 2020; 1539–1604.
- Kirsch B, Lippert S, Schmelter T, Jeske S, Muysers C, Kulmann H. Systematic screening of study data for subgroup effects. *CRAN Manual*. 2020. <https://cran.r-project.org/web/packages/subscreen/index.html> Accessed 13 Aug 2020.
- Muysers C, Dmitrienko A, Kulmann H, Kirsch B, Lippert S, Schmelter T, et al. A systematic approach for post hoc subgroup analyses with applications in clinical case studies. *Ther Innov Regul Sci*. 2020;54:507–518. <https://doi.org/10.1177/2168479019853782>
- Kurth BM. Der kinder- und jugendgesundheitsurvey (KiGGS): ein überblick über planung, durchführung und ergebnisse unter berücksichtigung von aspekten eines qualitätsmanagements. *Bundesgesundheitsbl—Gesundheitsforsch—Gesundheitsschutz*. 2007;50:533–546. <https://doi.org/10.1007/s00103-007-0214-x>
- Kurth B. M. Editorial: new information from and about the KiGGS study. *J Health Monit*. 2018; 3 <https://doi.org/10.17886/RKI-GBE-2018-020>
- Winkler J, Stolzenberg H. Der sozialschichtindex im bundes-gesundheitsurvey. *Gesundheitswesen* 1999;61:178–183.
- Mitchell EA, Robinson E, Black PN, Becroft DM, Clark PM, Pryor JE, et al. Risk factors for asthma at 3.5 and 7 years of age. *Clin Exp Allergy*. 2007;37:1747–55.
- Bjerg A, Sandström T, Lundbäck B, Rönmark E. Time trends in asthma and wheeze in Swedish children 1996–2006: prevalence and risk factors by sex. *Allergy* 2010;65:48–55.
- Norbäck D, Lu C, Wang J, Zhang Y, Li B, Zhao Z, et al. Asthma and rhinitis among Chinese children—indoor and outdoor air pollution and indicators of socio-economic status (SES). *Environ Int*. 2018;115:1–8. <https://doi.org/10.1016/j.envint.2018.02.023>
- Kuntz B, Rattay P, Poethko-Müller C, Thamm R, Hölling H, Lampert T. Social inequalities in health of children and adolescents in Germany. Results of the cross-sectional KiGGS Wave 2 study. *J Health Monit*. 2018; 3, <https://doi.org/10.17886/RKI-GBE-2018-083>
- Breiman L. Random Forests. *Mach Learn*. 2001;45:5–32. <https://doi.org/10.1023/A:1010933404324>
- Boulesteix AL, Janitzka S, Kruppa J, Koenig IR. Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. *WIREs Data Min Knowl Discov*. 2012;2:493–507.
- Kuhn M, Johnson K. *Applied predictive modeling*. 1st ed. Springer; 2013
- Breiman L. Out-of-bag estimation. Technical report, Dept. of Statistics, Univ. of Calif., Berkeley. 1996. <https://www.stat.berkeley.edu/~breiman/OOBestimation.pdf>
- Ilshwaran H, Kogalur UB, Gorodeski EZ, Minn AJ, Lauer MS. High-dimensional variable selection for survival data. *J Am Statist Assoc*. 2010; 205–217

20. Lipkovich I, Dmitrienko A, Muysers C, Ratitch B. Multiplicity issues in exploratory subgroup analysis. *J Biopharm Stat.* 2018;28:63–81.
21. Dmitrienko A, Lipkovich I, Dane A, Muysers C. Data-driven and confirmatory subgroup analysis in clinical trials. In: Ting N, Cappelleri J, Ho S, Chen DG, editors. *Design and Analysis of Subgroups with Biopharmaceutical Applications.* Springer International Publishing; 2020. pp. 33–91. <https://doi.org/10.1007/978-3-030-40105-4>
22. Harrell FE. *Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis.* 2nd ed. Springer-Verlag, New York; 2015
23. Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS. Random survival forests. *Ann Appl Stat.* 2008; 841–860
24. Robert Koch Institute. The German Health Survey for Children and Adolescents (KiGGS Wave 2). Department of Epidemiology and Health Monitoring Public Use File first version. 2019; <https://doi.org/10.7797/17-201417-1-1-1>. Accessed 27 Oct 2019.
25. Mauz E, Gößwald A, Kamtsiuris P, Hoffmann R, Lange M, Schenck U, et al. New data for action. Data collection for KiGGS Wave 2 has been completed. *J Health Monit.* 2017;2:2–27. <https://doi.org/10.17886/RKI-GBE-2017-105>
26. Lavallée P, Beaumont JF. Why we should put some weight on weights. *Survey Insights: Methods from the Field, Weighting: Practical Issues and 'How to' Approach, Invited article.* 2015. <https://surveyinsights.org/?p=6255>. Accessed 30 Jan 2020
27. Lampert T, Hoebel J, Kuntz B, Müters S, Kroll LE. Socioeconomic status and subjective social status measurement in KiGGS Wave 2. *J Health Monit.* 2018; 3, <https://doi.org/10.17886/RKI-GBE-2018-033>
28. The KIDSCRREN Group Europe (Ed). *The KIDSCREEN questionnaires: quality of life questionnaires for children and adolescents.* Handbook. Pabst Science Publishers, Lengerich; 2020

ACKNOWLEDGEMENTS

Data access of KiGGS Wave 2 was kindly provided by the national German public health institute Robert Koch-Institute (RKI). Steffen Jeske, Chrestos Concept, Germany, supported on behalf of Bayer AG in collaboration with Bodo Kirsch, Bayer AG, Germany the development of the current version 3.0 of the R package “subscreen”.

FUNDING

There was no funding for this work, except Open Access funding enabled and organized by Projekt DEAL.

ETHICS APPROVAL FOR THIS STUDY NO ADDITIONAL ETHICS APPROVAL WAS REQUIRED

For the original study and investigations, the Hannover Medical School's ethics committee assessed the ethics of the study and provided its approval (No. 2275-2014).

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41370-021-00365-x>.

Correspondence and requests for materials should be addressed to C.M.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021

Lebenslauf

Mein Lebenslauf wird aus datenschutzrechtlichen Gründen in der elektronischen Version meiner Arbeit nicht veröffentlicht.

Komplette Publikationsliste

Es wird der Impact Factor (IF) gemäß „InCites Journal Citation Reports“ von Web of Science für das Jahr der Veröffentlichung angegeben.

1. **Muysers C**, Messina F, Keil T, Roll S. A novel concept of screening for subgrouping factors for the association between socioeconomic status and respiratory allergies. *Journal of Exposure Science & Environmental Epidemiology* 2022; 32, 295–302 DOI:10.1038/s41370-021-00365-x (IF (2021): 6,371)
2. **Muysers C**, Dmitrienko A, Kulmann H, Kirsch B, Lippert S, Schmelter T, Schulz A, Mentenich N, Schmitz H, Schaefers M, Meinhardt G, Keil T, Roll S. A Systematic Approach for Post Hoc Subgroup Analyses With Applications in Clinical Case Studies. *Ther Innov Regul Sci.* 2020; 54, 507–518, DOI:10.1177/2168479019853782 (IF: 1,778)
3. Kirsch B, Jeske S, Lippert S, Schmelter T, **Muysers C**, Kulmann H. Systematic Screening of Study Data for Subgroup Effects. <https://cran.r-project.org/web/packages/subscreen/index.html> CRAN Reference Manual. 2021; uploaded February 10: 1-7. (kein Impact Factor)
4. Lipkovich I, Dmitrienko A, **Muysers C**, Ratitch B. Multiplicity issues in exploratory subgroup analysis. *Journal of Biopharmaceutical Statistics*; 2017 DOI:10.1080/10543406.2017.1397009 (IF: 0,638)
5. Kulmann H, **Muysers C**, Dmitrienko A, Röhmel J. The Biostatistics Innovation Center at Bayer. *Biopharmaceutical Report.* 2016; Fall: 1-5. (kein Impact Factor)
6. Dmitrienko A, **Muysers C**, Fritsch A, Lipkovich I. General guidance on exploratory and confirmatory subgroup analysis in late-stage clinical trials. *Journal of Biopharmaceutical Statistics*; 2016. 26:1, 71-98, DOI:10.1080/10543406.2015.1092033 (IF: 0,994)
7. Endrikat J, Vilos G, **Muysers C**, Fortier M, Solomayer E, Lukkari-Lax E. The levonorgestrel-releasing intrauterine system provides a reliable, long-term treatment option for women with idiopathic menorrhagia. *Arch Gynecol Obstet.* 2012; 285(1):117-21. DOI:10.1007/s00404-011-1902-1. (IF: 1,330)

8. Kaunitz AM, Bissonnette F, Monteiro I, Lukkari-Lax E, **Muysers C**, Jensen JT. Levonorgestrel-releasing intrauterine system or medroxyprogesterone for heavy menstrual bleeding: a randomized controlled trial. *Obstet Gynecol.* 2010 Sep;116(3):625-32. DOI:10.1097/AOG.0b013e3181ec622b. (IF: 4,392)
9. Schaefers M, **Muysers C**, Alexandersen P, Christiansen C. Effect of microdose transdermal 17A-estradiol compared with raloxifene in the prevention of bone loss in healthy postmenopausal women: a 2-year, randomized, double-blind trial. *Menopause: The Journal of The North American Menopause Society.* 2009; Vol. 16, No. 3, 559-565. DOI:10.1097/gme.0b013e31818ebfba (IF: 3,082)

Wissenschaftliche Vorträge [Auszug] mit Bezug zum Subgruppenthema

1. Lippert S, Kirsch B, **Muysers C**. *'Subgroup Explorer' zur systematischen Analyse von Subgruppen - einsetzbar bei der Nutzenbewertung?*
64. Deutsche Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie (GMDS) Jahrestagung; Dortmund. 11.9.2019
2. **Muysers C**, Kirsch B. *A Systematic Approach for Post Hoc Subgroup Analyses With Applications in Clinical Case Studies*
The PSI & DIA's Journal Club presents: Subgroup Analyses; Webinar. 5.11.2019
3. **Muysers C**, Kulmann H. Biostatistics Innovation Center (BIC) - *An all-included approach.*
CEN ISBS conference; Vienna, Austria. August 2017
4. **Muysers C**. *Do Subgroup Analyses harm? Usually not. Are they helpful? It depends. A regulatory and practical review.*
At the X. International Symposium On Biopharmaceutical Statistics (ISBS); Beijing, China. 29.06–01.07.2015
5. **Muysers C**. *Key Subgroup Analysis Issues in Clinical Trials*
Joint Statistical Meetings (JSM); Seattle, United States. 8–13.08.2015

Buchbeitrag

Dmitrienko A, Lipkovich I, Dane A, **Muysers C**. *Data-driven and confirmatory subgroup analysis in clinical trials*. In: Ting N, Cappelleri J, Ho S, Chen DG, editors. *Design and Analysis of Subgroups with Biopharmaceutical Applications*. Springer International Publishing; 2020. pp. 33–91. DOI:10.1007/978-3-030-40105-4

Danksagung

An dieser Stelle möchte ich mich ganz herzlich bei Frau PD Dr. Stephanie Roll und Herrn Prof. Dr. med. Thomas Keil für die außerordentlich hervorragende Betreuung und Unterstützung bei diesem Dissertationsthema bedanken. Ganz besonderer Dank gilt Frau PD Dr. Stephanie Roll für die Idee und kontinuierliche Motivation zu dieser Promotionsarbeit sowie für ihre Großzügigkeit, Hilfsbereitschaft und Geduld, die sie mir trotz ihrer Einbindung in ihren weiteren Forschungsprojekten entgegengebracht hat. Die immer wieder unglaublich schnelle aber auch gründliche Unterstützung von Herrn Prof. Dr. med. Thomas Keil soll hierbei nicht unerwähnt bleiben.

Zu großem Dank fühle ich mich auch Herrn Dr. Hermann Kulmann verpflichtet. Seine inspirierenden und innovativen Ideen biometrische Aufgaben durch alternative und moderne Lösungen zu bewältigen, haben eine wesentliche Grundlage auch für diese Arbeit gelegt.

Alle zuvor genannten haben mich bei diesem Projekt stets unverbindlich und umfassend unterstützt und somit zum Gelingen dieser Promotionsarbeit beigetragen.

Ebenso danke ich dem gesamten Team, das mit vielen Erweiterungen und Verbesserungen aus dem R-Paket ‚subscreen‘ eine professionelle Applikation entwickelt hat. Diese Applikation war das zentrale analytische Element dieser Promotionsarbeit. Zu diesem Team zählen neben Dr. Hermann Kulmann vor allem auch Bodo Kirsch, Susanne Lippert, Dr. Thomas Schmelter und Steffen Jeske.