

Datenaufbereitung mittels nichtlinearer Klassifikationsverfahren zur Erhöhung der statistischen Aussagekraft von odds ratio Analysen

Bitterlich N¹, Schneider J²

¹Medizin & Service GmbH Chemnitz, Deutschland

²Institut und Poliklinik für Arbeits- und Sozialmedizin der Universitätskliniken Gießen/Marburg, Deutschland
bitterlich@medizinservice-sachsen.de

Einleitung Leben und Gesundheit vieler Teile der arbeitenden Bevölkerung sind durch krebserzeugende Gefahrstoffe der Arbeitswelt in vielfältiger Weise bedroht. Die Zahl entschädigungspflichtiger Berufskrebserkrankungen steigt besorgniserregend an. Allein die Gruppe Lungenkrebs-erzeugender Arbeitsstoffe im geltenden Berufskrankheitenrecht umfasst zurzeit nicht weniger als neun Stoffgruppen. Ein zukunftssträchtiges Forschungsgebiet stellt beim arbeitsbedingten Lungenkrebs die synkanzerogene Wirkungsverstärkung einerseits durch Koexposition infolge kanzerogener Gefahrstoff-Einwirkung am Arbeitsplatz und andererseits im Hinblick auf die Zigarettenrauchgewohnheiten dar. Für die Sekundärprävention folgt aus dem Zeitablauf bei Berufskrebserkrankungen als Paradigma der jahrzehntelangen Latenzschäden das Postulat, durch frühestmögliche Diagnose und nachfolgende Interventionsstrategien die zukünftige Sterblichkeit der betroffenen arbeitenden Bevölkerung zu minimieren. Als zentrales Anliegen der arbeitsmedizinischen Onkologie kann die Eingrenzung von Hochrisikogruppen durch die Definition der genetischen Prädisposition und die Identifizierung spezifischer exogener Schädigungsfolgen auf molekularer Basis angesehen werden. Entscheidend für Art und Zeitpunkt der Krankheitsausprägung ist die Interaktion exogener Schädigungsfaktoren bei gegebenem Gendefekt. Als so genannte Suszeptibilitätsmarker haben sich die Faktoren der genetischen Prädisposition, die am Stoffwechsel der Gefahrstoffe beteiligt sind, erwiesen.

Die Analyse solcher Zusammenhänge mittels multipler logistischer Regression und daraus abgeleitet der odds ratio weist zwar auf erwartete Abhängigkeiten hin, kann aber im Allgemeinen selbst bei Subgruppen nicht statistisch signifikant nachgewiesen werden. Adjustierung auf bekannte Effekte wie Alter und Rauchgewohnheiten verbessern die Aussagekraft nur wenig. Eine weitere Homogenisierung des zu betrachtenden Datenmaterials mittels Ein-/Ausschlusskriterien bewirken tendenziell eine Unterstützung der Arbeitsthese, jedoch verringern sich die verfügbaren Fallzahlen so drastisch, dass deshalb die statistische Signifikanz nicht erreicht werden kann.

Betrachtet man allerdings die Klassifikationsgüte der zugrunde liegenden Regressionsmodelle, so bleiben diese weit hinter den Erwartungen zurück. Mithilfe der berücksichtigten Einflussfaktoren lassen sich offensichtlich keine Klassenzuordnungen ausreichend begründen. Verwendet man dagegen nichtlineare Klassifikationsverfahren, so kann die Trefferquote der Zuordnung, gemessen an den ROC-Kurven, teilweise erheblich verbessert werden. Es wird deshalb untersucht, in wie weit die verwendeten Klassifikationsverfahren genutzt werden können, um die Einflussfaktoren zu transformieren und damit die Aussagekraft der odds ratio Analyse zu steigern.

Material und Methoden: Von über 1.000 Personen eines Hochrisikokollektivs wurden Polymorphismen bestimmt. Als arbeits- und umweltmedizinisch besonders relevant können hierbei die Cytochromoxydasen (CYP) und die Glutathion S-Transferasen (GST) angesehen werden. Die DNA-Extraktion erfolgte nach den herkömmlichen Standardmethoden, die Analyse der Genprodukte mittels real-time-PCR-Analyse.

Die statistische Basisanalyse der Einzelparameter wird geschichtet nach dem histologischen Tumortyp zunächst mit Hilfe der multiplen logistischen Regressionsanalyse durchgeführt. Als Einflussgrößen werden das Alter, das Geschlecht, der Rauchkonsum und die bekanntermaßen Lungenkrebs-erzeugenden Gefahrstoffe berücksichtigt. Eine systematische Subgruppen-Untersuchung zeigt Tendenzen auf, die in einer weiteren Analyse zu bestätigen sind. Allerdings sinkt dabei die statistische Aussagekraft wegen der abnehmenden Fallzahlen schnell ab. Die Grenzen der Klassifikationsleistung können auch anhand der zugehörigen ROC-Kurven demonstriert werden.

Ergebnisse Als alternativer Lösungsansatz werden zunächst fuzzy-basierte Klassifikationsverfahren auf die Parameter der betrachteten Einflussfaktoren angewandt. Dabei werden nicht nur die Diskussionsansätze aus der Regressionsanalyse bestätigt, sondern deutliche Sensitivitätssteigerungen bei gleichbleibender Spezifität für die Klassenzuordnungen erreicht. Die Fuzzifizierung der Eingangsparameter als ein Verfahrensschritt der verwendeten Methoden eignet sich zur Transformation. Werden nun anstelle der ursprünglichen Messwerte die entsprechend klassifikatorischer Prinzipien transformierten Kenngrößen (so genannte Entscheidungssicherheiten für Profilauswertungen) der multiplen logistischen Regression zugeführt, reduzieren sich die p-Werte der odds ratio. Ausgewählte Arbeitsthese können auf diese Weise als signifikant nachgewiesen werden. Das Verfahren zeichnet sich dabei durch hohe Interpretationsfähigkeit der Vorverarbeitung aus und eignet sich für eine breite Klasse von Parametern. Die Vorgehensweise und die Ergebnisse werden anhand konkreter Analysen zur Diskussion gestellt.