

Kapitel 4

Diskussion

4.1 Methodik

In dieser Arbeit wurde ein umfassender Überblick über die Methoden bei der Bewertung diagnostischer Tests gegeben. Insbesondere wurde die in der Praxis oft nicht berücksichtigte Problematik bei der Kombination diagnostischer Tests dargelegt.

Da es nicht möglich ist, durch Kombination zweier Tests sowohl die Sensitivität als auch die Spezifität zu verbessern, muss man sich entscheiden, ob durch Verwenden der BTP-Regel die Sensitivität oder durch Verwenden der BTN-Regel die Spezifität gegenüber den Einzeltests verbessert werden soll. Auch die Wahl der zu kombinierenden Tests ist entscheidend. Es konnte gezeigt werden, dass die „naive“ Kombination des Tests mit der höchsten Sensitivität mit dem Test mit der höchsten Spezifität schlechte Ergebnisse liefert. Stattdessen sollte, wenn eine höhere Sensitivität erwünscht ist, die BTP-Regel verwendet werden und zwei Tests mit hoher Spezifität kombiniert werden, damit der Spezifitätsverlust möglichst gering ausfällt. Analog

sollten zwei Tests mit hoher Sensitivität mittels der BTN-Regel kombiniert werden, falls eine Vergrößerung der Spezifität angestrebt wird. Da der Spezifitäts- bzw. Sensitivitätsverlust umso größer wird, je mehr Tests kombiniert werden (Odysseus-Syndrom), ist es zudem meist nicht sinnvoll, mehr als zwei Tests zu kombinieren.

Eine Verbesserung sowohl des positiven als auch des negativen prädiktiven Wertes durch Kombination ist möglich, allerdings nur unter speziellen Voraussetzungen. Bei Verwenden der BTP-Regel können sich beide prädiktive Werte verbessern, wenn bei den Gesunden eine hohe Abhängigkeit zwischen den Tests besteht, bei den Kranken die beiden Tests jedoch (nahezu) unabhängig sind. Analog führt die BTN-Regel nur dann sowohl zu einer Verbesserung des positiven wie auch des negativen prädiktiven Wertes, wenn bei den Kranken eine hohe Abhängigkeit der Tests besteht, aber bei den Gesunden die beiden Tests (nahezu) unabhängig sind.

Abschließend wurden in dieser Arbeit die Zusammenhänge zwischen der statistischen Entscheidungstheorie und der klinischen Entscheidungsanalyse ausführlich an einem Beispiel dargelegt und gezeigt, dass die klinische Entscheidungsanalyse als ein Spezialfall der statistischen Entscheidungstheorie aufgefasst werden kann.

4.2 Vergleich und Kombination der Modalitäten

Als erster Schritt der Auswertung wurde die diagnostische Leistung der sechs verschiedenen bildgebenden Verfahren bei der Diagnose und Beurteilung der Resektabilität bei Patienten mit Verdacht auf ein Pankreaskarzinom miteinander verglichen.

Die Ergebnisse zeigen, dass bislang keine Modalität alleine in der Lage ist, alle diagnostischen Fragen mit hinreichender Genauigkeit zu beantworten. In unserer Studie erreichten ERCP und MR mit 95% bzw. 91% die höchste Sensitivität bezüglich der Beurteilung der Dignität. Allerdings war die Spezifität von MR und ERCP mit 76% bzw. 71% nicht zufriedenstellend und schlechter als die von CT und EUS mit 82% und 78%.

Bezüglich des zweiten Diagnoseschritts war die Spezifität, d.h. die Beurteilung der lokalen Irresektabilität, bei allen Modalitäten sehr niedrig und erreichte nur bei CT mit 57% einen Wert über 50%. Bei den anderen drei Modalitäten werden mehr als 50% der Patienten mit irresektablem Karzinom als resektabel diagnostiziert. Die Sensitivität, d.h. Feststellung der lokalen Resektabilität, war bei allen Modalitäten gut, bei MR und EUS mit 98% bzw. 96% sogar ausgezeichnet.

Die vorliegende Studie ist bislang die größte diagnostische Studie zum Vergleich bildgebender Verfahren in der Pankreasdiagnostik. Zudem ist sie die erste, in der die sechs meist verwendeten Modalitäten prospektiv direkt miteinander verglichen werden. Aufgrund unterschiedlicher Studiendesigns, teil-

weise sehr kleinen Fallzahlen, verschiedener Einschlusskriterien für die Patienten und abweichender Kriterien für die Bestimmung der Resektabilität ist ein Vergleich mit anderen Studien schwierig. Das gute Abschneiden von MR und ERCP bezüglich der Sensitivität bei der Beurteilung der Dignität bestätigt aber die Ergebnisse einer vergleichbaren prospektiven Studie [21]. Auch die schlechte diagnostische Leistung der Verfahren bei der Beurteilung der Resektabilität ist bekannt [25, 20, 21].

Die Ergebnisse zeigen, dass für alle Modalitäten die Differentialdiagnose zwischen chronischer Pankreatitis und einem Pankreaskarzinom bzw. die Beurteilung der Resektabilität schwierig bleibt. Insbesondere die Feststellung der Irresektabilität bereitet Schwierigkeiten. Deswegen wurde im zweiten Teil der Auswertung untersucht, ob es Kombinationen gibt, die die diagnostische Leistung verbessern. Da es nicht möglich ist, sowohl die Sensitivität als auch die Spezifität durch Kombination zweier Tests zu verbessern, führt die Verwendung eines weiteren diagnostischen Tests allerdings nicht unbedingt zu besseren Ergebnissen.

Die „naive“ Kombination des Verfahrens mit der höchsten Sensitivität (ERCP mit 95%) mit dem Verfahren mit der größten Spezifität (EUS mit 82%) veranschaulicht die Problematik. Zwar erhöht sich bei Verwenden der BTN-Regel die Spezifität auf 85%, allerdings auf Kosten einer deutlich verringerten Sensitivität von 53%. Bei der empfohlenen Kombination werden - kontraintuitiv - die Tests mit den höchsten Sensitivitäten (ERCP mit 95% und MR mit 91%) kombiniert, um den Sensitivitätsverlust möglichst gering zu halten. In diesem Fall beträgt die Sensitivität der Kombination 72% und die Spezifität 91%.

Auch bei der Beurteilung der Resektabilität kann durch Kombination mittels BTN-Regel die Spezifität deutlich erhöht werden. Da alle vier Einzelsensitivitäten über 80% liegen, hält sich auch der Sensitivitätsverlust in Grenzen.

4.3 Entscheidungsanalyse

Als letzter Schritt der Auswertung wurde ein Entscheidungsbaum modelliert, in dem 44 verschiedene Strategien zur Diagnose und Beurteilung der Resektabilität bei Patienten mit Verdacht auf ein Pankreaskarzinom miteinander verglichen wurden.

In dieser Analyse waren alle sechs Strategien mit MR als erster Modalität unter den besten acht und erzielten etwas bessere Resultate als Strategien mit CT als erster Modalität. Die beste diagnostische Leistung erbrachte die Strategie „MR gefolgt von ERCP für positive Ergebnisse“, die 78% der Patienten richtig in benigne, maligne/resektabel und maligne/irresektabel klassifizierte. Die zweitbeste Strategie war „nur MR“, mit 77% richtig klassifizierten Patienten.

Dies stimmt mit den Ergebnissen verschiedener Studien überein, die gezeigt haben, dass die diagnostische Leistung des MR bei der Detektion und dem Staging des Pankreaskarzinoms mit der des CT vergleichbar oder besser ist [21]. Einige Autoren haben Strategien mit CT als erster Modalität empfohlen, gefolgt von EUS oder ERCP [33, 34]. Diese Strategien schnitten in dieser Untersuchung etwas schlechter ab als diejenigen mit MR als erster Modalität. Das mag dadurch verursacht sein, dass in dieser Studie die MR Untersuchun-

gen MR Cholangiopankreaticographie (MRCP) und MR Angiographie mit einschlossen, was die Leistungsfähigkeit des MR verbessert.

Wenn die Kosten der einzelnen diagnostischen Verfahren berücksichtigt werden, zeigt sich, dass alle Strategien außer „nur US“, „nur CT“, „nur MR“ und „MR gefolgt von ERCP für Positive“ dominiert werden. Die Strategie „nur US“ ist mit erwarteten Kosten von 37,90 Euro pro Patient mit Abstand am billigsten. Bei der Strategie „nur CT“ liegen die erwarteten Kosten pro Patient um 185,40 Euro höher, dafür werden ca. 10% mehr Patienten richtig diagnostiziert. Dies führt zu erwarteten Kosten von 1.911 Euro für jeden zusätzlich richtig diagnostizierten Patienten. Bei der Strategie „nur MR“ liegen die erwarteten Kosten pro Patient um weitere 154,70 Euro höher als bei der Strategie „nur CT“, bei 4% mehr richtig diagnostizierten Patienten. Die Kosten für jeden im Vergleich zu CT zusätzlich richtig klassifizierten Patienten belaufen sich auf 3.516 Euro. Wenn positive Ergebnisse des MR mit ERCP verifiziert werden (Strategie „MR gefolgt von ERCP für Positive“), erhöht sich der Anteil der richtig klassifizierten Patienten um weitere 1,5%, allerdings steigen die erwarteten Kosten um beträchtliche 587,10 Euro. Die Kosten für jeden zusätzlich richtig diagnostizierten Patienten im Vergleich zur Strategie „nur MR“ betragen 13.490 Euro. Berücksichtigt man, dass es sich bei ERCP um ein invasives Verfahren handelt, das Komplikationsraten von 5-10% aufweist, scheint es nicht nur wenig kosteneffektiv, sondern klinisch auch nicht sinnvoll, für den marginalen Zugewinn an diagnostischer Leistung alle Patienten mit positivem MR den Risiken einer ERCP auszusetzen. Vielmehr sollte die Entscheidung abhängig von den Patientencharakteristika und klinischen Erwägungen getroffen werden.

Diese Entscheidungsanalyse unterscheidet sich von den bereits publizierten Entscheidungsanalysen in mehreren Punkten. Erstens wurde die diagnostische Leistung der Verfahren sowohl hinsichtlich der Beurteilung der Dignität als auch hinsichtlich der Beurteilung der Resektabilität untersucht. In den anderen Arbeiten wurden die Verfahren nur bezüglich der Beurteilung der Resektabilität verglichen. Zweitens wurden in dieser Arbeit andere Modalitäten betrachtet. Die der Arbeit zugrunde liegende Studie war die erste, bei der die sechs meist verwendeten bildgebenden Verfahren für die Diagnose und das Staging des Pankreaskarzinoms verglichen wurden. Das Fehlen von Angiographie und Feinnadel-Biopsie (FNA) erscheint als kein Nachteil, da beide Methoden aus klinischen Gründen nicht mehr empfohlen werden: Eine konventionelle Angiographie ist sowohl einer Spiral-CT mit adäquater Kontrastmittelgabe als auch der MR-Angiographie unterlegen und heute nicht mehr indiziert [7]. Aufgrund der Invasivität der FNA wird diese nur im Falle eines irresektablen Karzinoms zur Bestätigung vor einer palliativen (Chemo)Therapie empfohlen [72].

Der wichtigste Unterschied ist, dass in dieser Arbeit im Gegensatz zu den anderen Studien nicht angenommen werden musste, dass die Verfahren bedingt unabhängig sind (siehe Seite 2.4.2). Die ϕ -Koeffizienten und p-Werte des exakten Tests von Fisher zeigen, dass die Annahme der bedingten Unabhängigkeit zumindest für einige Kombinationen (z.B. US und EUS, aber auch CT und MR) nicht haltbar ist. Da in den anderen Studien ebenfalls bildgebende Verfahren verglichen werden, ist es wahrscheinlich, dass die Befunde auch in diesen Fällen zumindest durch die Größe der Läsion korrelieren und die Annahme der bedingten Unabhängigkeit auch dort nicht haltbar ist.

4.4 Limitationen

Diese Arbeit hat verschiedene Limitationen. Erstens basierte die Entscheidungsanalyse auf der Annahme, dass unbeschränkter Zugang zu allen in die Untersuchung aufgenommenen Modalitäten besteht. Dies mag nicht in allen Situationen zutreffen. Allerdings wird für die (laut dieser Arbeit) optimale Strategie nur ein MR und evtl. anschließend ein ERCP benötigt.

Zweitens konnten nicht alle diagnostischen Tests bei allen Patienten protokollgemäß durchgeführt werden, da bei einigen Patienten bereits Ergebnisse externer Tests vorlagen oder Gerätedefekte Untersuchungen unmöglich machten. Um einen möglichen Selektionsbias auszuschließen, wurden die Analysen sowohl für alle 193 Patienten als auch für die Gruppe der 133 Patienten durchgeführt, die alle sechs Untersuchungen komplett absolviert haben. Dabei zeigten sich nur geringe Unterschiede, es ist also anzunehmen, dass es nicht zu einem Selektionsbias kam.

Drittens waren die Studienpatienten hoch selektierte Patienten eines tertiären Zentrums, was die Generalisierbarkeit der Ergebnisse z.B. auf eine gastroenterologische Facharztpraxis, in der die Prävalenz deutlich geringer ist, einschränkt. Zwar zeigen die Ergebnisse der Sensitivitätsanalyse, dass auch bei geringerer Prävalenz die Ergebnisse stabil sind, allerdings werden sich bei dem zu erwarteten anderen Patientenspektrum die Sensitivitäten und Spezifitäten der einzelnen Modalitäten möglicherweise ändern.

Viertens hängt die diagnostische Leistung der einzelnen Verfahren von verschiedenen lokalen Faktoren wie z.B. den Protokollen vor Ort und der Erfahrung der Reader ab. Zusätzlich verändert sie sich ständig durch die fortlau-

fende technische Entwicklung und dadurch einhergehende Verbesserung der Modalitäten. Eine diagnostische Studie kann somit immer nur eine Momentaufnahme sein. Da es aber meist einige Zeit dauert bis technische Neuerungen der Mehrheit der Patienten zur Verfügung stehen, ist anzunehmen, dass die in dieser Arbeit untersuchten Modalitäten die Praxis recht gut widerspiegeln.

Fünftens konnte nur bei 90 der 119 Patienten mit Malignom der Goldstandard der lokalen Resektabilität ermittelt werden. Dies könnte zu einem so genannten Verifikationsbias geführt haben. Bei knapp der Hälfte ($n=13$) der 29 Patienten ohne Goldstandard erfolgte aufgrund von mangelnder Compliance, anderweitiger Operationsrisiken oder Loss-to-follow-up keine Laparotomie. Diese Gründe können aufgrund von klinischer Erfahrung als vom Ergebnis der Laparotomie unabhängig betrachtet werden, es ist nicht anzunehmen, dass dadurch ein Bias induziert wurde. Von den restlichen 16 Patienten wurden bei 12 Patienten von mindestens drei der Modalitäten Fernmetastasen detektiert und bei vier Patienten eine lokale Irresektabilität festgestellt. Da auch keine anderen medizinischen Gründe für eine Laparotomie (z.B. Einsatz eines Stents) vorlagen, konnte bei diesen Patienten aus ethischen Gründen der Goldstandard nicht ermittelt werden. Im Allgemeinen führt der Verifikationsbias zu einer Überschätzung der Sensitivität und einer Unterschätzung der Spezifität. Aber selbst wenn im Rahmen einer Worst-Case-Analyse angenommen wird, dass bei allen vier Patienten die jeweilige Modalität die lokale Irresektabilität richtig erkannt hätte, würde die Spezifität nur auf Werte zwischen 43 und 75% ansteigen und weiterhin sehr niedrig sein. Entsprechend würde, wenn die 12 Patienten mit Fernmetastasen alle lokal resektabel gewesen wären, aber von den Modalitäten fälschlicherweise alle als irresektabel diagnostiziert worden wären, die Sensitivität auf Werte zwischen 69 und 82%

fallen. Da aber die meisten Patienten mit Fernmetastasen ein bereits lokal fortgeschrittenes Karzinom haben, dürfte die für den Verifikationsbias korrigierte Sensitivität weitaus höher und damit in der Nähe der unkorrigierten Werte liegen.

Die sechste und wichtigste Einschränkung ist, dass nicht zwischen den unterschiedlichen Arten der Fehlklassifikation unterschieden wird, sondern nur eine dichotome Verlust- bzw. Nutzenfunktion verwendet wurde. So unterscheiden sich beispielsweise die klinischen Konsequenzen bei einem Patienten mit resektablem Karzinom, der fälschlicherweise als irresektabel diagnostiziert wird, von denjenigen eines Patienten mit benigner Erkrankung, der fälschlicherweise als maligne, aber resektabel klassifiziert und operiert wird. Zudem wurden nur die Kosten der Diagnostik berücksichtigt und weitere Kosten, z.B. durch Therapie oder indirekte Kosten durch Arbeitsausfall außer Acht gelassen. Für eine differenziertere Verlust- bzw. Nutzenfunktion werden Informationen über die Lebensqualität und Lebenserwartung und die entstehenden Kosten benötigt. Es ist geplant, dementsprechende Studien durchzuführen und deren Ergebnisse in weiteren Analyseschritten in das Modell zu integrieren.

4.5 Schlussfolgerung

Diese Arbeit hat gezeigt, dass es vor der Durchführung umfangreicher diagnostischer Tests sinnvoll ist, sich Gedanken über das Vorgehen bei der Kombination zu machen und sich, um das Odysseus-Syndrom zu umgehen, auf die Tests zu beschränken, deren Ergebnisse für die Therapieentscheidung not-

wendig sind.

Statistische Verfahren und die Methoden der klinischen Entscheidungsanalyse können dazu beitragen, aus der Menge der zur Verfügung stehenden Modalitäten und ihrer Kombinationen diejenigen auszuwählen, die unter Berücksichtigung des jeweiligen Kontexts am besten geeignet ist.

Allgemein lässt sich feststellen, dass die „naive“ Kombination des Verfahrens mit der höchsten Sensitivität mit dem Verfahren mit der größten Spezifität nicht sinnvoll ist. Die empfohlene, kontraintuitive Kombination der Tests mit den höchsten Sensitivitäten mittels BTN-Regel bzw. der Tests mit den höchsten Spezifitäten mittels BTP-Regel führt zu deutlich besseren Ergebnissen.

Bezogen auf die Pankreasdiagnostik zeigt sich aufgrund der erfolgten Untersuchungen, dass MR im Vergleich zu den anderen Modalitäten in dieser Studie die beste diagnostische Leistung bei der Diagnose und dem Staging des Pankreaskarzinoms erbringt. MR liefert sowohl bei der Sensitivität und der Spezifität als auch bei den Likelihood Ratios bezüglich der Differentialdiagnose und der Beurteilung der Resektabilität die besten Ergebnisse. Zudem ist es im Gegensatz zu EUS und insbesondere ERCP nicht-invasiv und erfordert im Gegensatz zu CT auch keine Kontrastmittelgabe.

Da aber alle Modalitäten insbesondere bei der Beurteilung der Resektabilität Defizite hatten, muss weiter an der Verbesserung der Pankreasdiagnostik gearbeitet werden. Ebenso ist es notwendig, die Folgen von Fehlklassifikationen zu untersuchen und die Auswirkungen der Behandlung auf die Lebensqua-

lität und Lebenserwartung der Patienten zu berücksichtigen.

Bis neuere Untersuchungen bzw. Erkenntnisse vorliegen, wird aufgrund der Ergebnisse dieser Arbeit empfohlen, bei Patienten mit Verdacht auf Pankreaskarzinom MR als ersten diagnostischen Test durchzuführen. Ein positiver Befund kann anschließend, abhängig von Patientencharakteristika und klinischen Erwägungen, mittels ERCP verifiziert werden.