

Eine neue Methode als Entscheidungshilfe bei unklarer wissenschaftlicher Evidenz: Test-Implementierung und Simulation

Zusammenfassung

Hintergrund: Viele Gesundheitsfachpersonen kennen Ergebnisse von Reviews wie etwa die der Cochrane Collaboration. Manche Ergebnisse dieser Reviews zeigen, dass (noch) kein Vorteil für eine von zwei vielversprechenden Interventionen erkannt wird. In diesem Fall schlagen Beck-Bornholdt und Dubben den Algorithmus einer modifizierten never-change-a-winning-team Strategie vor. Ähnliche Methoden werden für die Gruppenzuteilung in Studien oder für die Anpassung des Studiendesigns im Verlauf einer Studie genutzt.

Forschungsfrage: Ist die von Beck-Bornholdt und Dubben (2003) vorgeschlagene Methode für den Praxisalltag der Hebammen tauglich, bei der Wahl zwischen zwei in der Literatur als gleichwertig erscheinenden Interventionen eine sinnvolle Entscheidungshilfe zu bieten?

Methodik: Die Anwendung des Algorithmus wird für eine mögliche Anwendung durch Gesundheitsfachpersonen simuliert. Er bezieht alle Erfahrungen ein, die mit den beiden zu vergleichenden Interventionen gemacht werden, um die Intervention für die nächste zu behandelnde Person festzulegen.

Ergebnisse: Für verschiedene Szenarien mit unterschiedlichen Erfolgswahrscheinlichkeiten der beiden Interventionen wurden Simulationen durchgeführt. Die durchschnittliche Erfolgsrate ist dabei in allen Szenarien bereits ab der zweiten zu behandelnden Person besser als die mittlere Erfolgsrate der beiden Interventionen.

Schlussfolgerung: Die vorliegenden Ergebnisse sollen als Diskussionsgrundlage für die Anwendbarkeit der vorgeschlagenen Methode dienen. Bei unklarer wissenschaftlicher Evidenz aus Metaanalysen kann der Algorithmus die Entscheidung von Gesundheitsfachpersonen für eine von zwei Behandlungsmöglichkeiten gewinnbringend unterstützen. Der Subjektivität des eigenen Settings (Klientel, Durchführungsweise der Behandlung) wird dabei stets Rechnung getragen.

Schlüsselwörter: Algorithmus, Intervention, Hebammenforschung

Christine Loytved¹
Rebecca Erdin¹

¹ Institut für Hebammen,
Department Gesundheit,
Zürcher Hochschule für
Angewandte Wissenschaften,
Winterthur, Schweiz

Hintergrund

In der Hebammenarbeit rückt die Frage der Evidenzbasierung immer weiter in den Vordergrund. Die Berücksichtigung der drei Säulen (wissenschaftliche Evidenz, Erfahrung der Hebamme und Einstellung der Klientinnen) wird von den Hebammen selbst gefordert und von dritter Seite verlangt – sei es von Seiten anderer Berufsgruppen oder von Seiten der Klientinnen. Doch gibt es zu manchen Themen, wie beispielsweise Sodbrennen in der Schwangerschaft, keine klare Aussage zur wissenschaftlichen Evidenz, die eine einzige Intervention, welche aus einem Handeln oder auch aus einem Nicht-Handeln bestehen kann, favorisiert. Evidenzbasierte Leitlinien wie etwa die des National Institute for Health and Care Excellence (NICE) oder systematische Übersichtsarbeiten wie die der Cochrane Collaboration bieten zu manchen Themenfel-

dern Empfehlungen, die unterschiedliche Interventionen in einer bestimmten Situation zulassen. Dies ist insbesondere dann der Fall, wenn aufgrund der Studienlage für keine der getesteten Interventionen eine Überlegenheit zu erkennen ist. So ist bei Sodbrennen in der Schwangerschaft ungeklärt, ob Akupunktur hilft oder eine Umstellung der Ernährungsgewohnheiten anzuraten ist [8]. Gerade bei Themen, die keine schwerwiegende Intervention wie eine Oxytocingabe bei Geburtseinleitung, sondern weniger invasive Maßnahmen betreffen, ist auch nicht zu erwarten, dass sich die Studienlage zeitnah verbessern wird. In solchen Fällen, in denen Reviews bzw. Leitlinien zum Schluss kommen, dass das eine wie das andere Verfahren zum Erfolg führen kann, sollten Hebammen ein Hilfsmittel an die Hand bekommen, welches ihnen die Entscheidung für eine der Interventionen in ihrem Arbeitsalltag erleichtert. Da die Studienlage in solchen Fällen

nur eine eingeschränkte Entscheidungshilfe bietet, ist es angeraten, die in der eigenen bisherigen Praxis erfahrene, interne Evidenz für die Entscheidung zu nutzen. Das heißt, die Erfolge und Misserfolge, die im Zusammenhang mit einer Intervention erlebt werden, sollten für zukünftige Entscheidungen sinnvoll mit einbezogen werden.

Beck-Bornholdt und Dubben [2] schlagen als mögliche Lösung einen Algorithmus vor, eine abgewandelte „never-change-a-winning-team“ Strategie. Die Idee dieses Algorithmus ist es, die Entscheidung, welche Intervention bei der nächsten Klientin mit einer betreffenden Diagnose angewendet werden soll, nach einem bestimmten Schema zu treffen: Es soll jeweils aufgrund aller bisher (mit den beiden zur Wahl stehenden Interventionen) gemachten Erfahrungen entschieden werden. Diese Idee, auch bezeichnet als „play-the-winner rule“, bezieht sich auf Werke von Bayes [1] und Wold [11], wobei ähnliche Vorgehensweisen bereits früher beschrieben worden sind [12]. Das Theorem von Bayes wurde für folgende Anwendungen vorgeschlagen:

1. für die Erstellung einer Regel für den Fall, dass es bei einer Zwischenergebnislage ethisch erforderlich ist, eine randomisiert kontrollierte Studie zu stoppen;
2. für die Ermittlung der geeigneten Medikamentendosen für eine randomisiert kontrollierte Studie sowie
3. für die Interpretation der Evidenz aus der Studienlage [10].

Ähnliche Vorgehensweisen werden für die Gruppenzuordnung von Studienteilnehmer/innen [6], die Anpassung des Studiendesigns im Verlauf einer Studie [4] und für Evaluationen im Gesundheitswesen diskutiert [10]. Beck-Bornholdt und Dubben [2] nutzen das Theorem von Bayes, indem sie die bisher beobachteten Erfolge und Misserfolge als Entscheidungshilfe bei der Wahl zwischen zwei als gleichwertig erachteten Interventionen miteinbeziehen. Eine solche Anwendung als Entscheidungshilfe in der Praxis der Gesundheitsberufe ist bislang nicht bekannt. Die Anwendung des Algorithmus für die abgewandelte „play-the-winner rule“ sollte so benutzerfreundlich wie möglich gestaltet werden. Zu denken wäre an die Entwicklung einer App, in die nur die jeweiligen Interventionen und deren Erfolge eingetragen werden müssten. Ihre Anwendung sollte jedoch nicht starr sein und keinesfalls das Selbstbestimmungsrecht der Klientin einschränken. Sie könnte lediglich ein wirksames Mittel sein, den von der Hebamme sonst ohne systematische Verarbeitung gewonnenen Erfahrungsschatz auf manchen Gebieten schneller und effektiver zu erzeugen.

Ziel und Fragestellung

Wenn aufgrund von systematischen Übersichtsarbeiten und evidenzbasierten Leitlinien die Aussage abzuleiten ist, dass zwei Interventionen gleichwertig vielversprechend sein könnten, sollen Hebammen (und andere Berufe im Gesundheitswesen) eine Entscheidungshilfe nach dem aktuellsten Wissensstand erhalten, um ihrer Klientel

die bestmögliche Betreuung zu bieten. Das Selbstbestimmungsrecht der Klientin soll dabei nicht angetastet werden. Auch die individuelle Situation der Klientin soll berücksichtigt werden. Wenn nach Abwägung aller Faktoren keine eindeutige Entscheidung für die eine Intervention und gegen die andere gefällt wurde, soll die Hebamme eine Hilfestellung erhalten. Es geht demnach darum, welche Empfehlung die Hebamme im Rahmen der gemeinsamen, informierten Entscheidung („informed shared decision making“) abgibt. Wird diese Empfehlung umgesetzt, gewinnt die Hebamme wieder eine Erfahrung für ihre zukünftigen Empfehlungen hinzu.

Die Fragestellung der vorliegenden Arbeit lautet daher: Ist die von Beck-Bornholdt und Dubben [2] vorgeschlagene Methode für den Praxisalltag der Hebammen tauglich, bei der Wahl zwischen zwei in der Literatur als gleichwertig erscheinenden Interventionen eine sinnvolle Entscheidungshilfe zu bieten?

Methode

Der Algorithmus für die abgewandelte „play-the-winner rule“ geht folgendermaßen für die Wahl der anzuwendenden Intervention vor:

1. Bei der ersten Klientin wird eine der beiden Interventionen zufällig als Startintervention bestimmt (bspw. Münzwurf). Diese sei hier I_S genannt.
2. Solange I_S erfolgreich ist, wird sie für die jeweils nächste Klientin beibehalten.
3. Bei erstmaligem Misserfolg von I_S wird zur anderen Intervention gewechselt. Diese sei hier I_A genannt.
4. I_A wird wiederum für die jeweils nächste Klientin beibehalten, solange sie erfolgreich ist.
5. Ab dem erstmaligen Misserfolg von I_A wird für jede weitere Klientin die Wahlwahrscheinlichkeit α für eine der beiden Interventionen berechnet, hier wiederum α_S und α_A genannt.

$$\alpha_S = \frac{\beta_S}{\beta_S + \frac{1}{\beta_S}} \quad \text{und} \quad \alpha_A = \frac{\beta_A}{\beta_A + \frac{1}{\beta_A}}$$

mit

$$\beta_S = \left(\frac{\text{Success rate } I_S + 0.28}{\text{Success rate } I_A + 0.28} \right)^w \quad \text{und} \quad \beta_A = \left(\frac{\text{Success rate } I_A + 0.28}{\text{Success rate } I_S + 0.28} \right)^w$$

wobei w die Anzahl Wechsel zwischen den beiden Methoden sei und sich die Erfolgsrate (Success rate) der einzelnen Interventionen folgendermaßen berechnet (Anzahl der Erfolge geteilt durch Anzahl der Anwendungen):

$$\text{Success rate } I_S = \frac{\text{Number of successes with } I_S}{\text{Number of applications of } I_S} \quad \text{und}$$

$$\text{Success rate } I_A = \frac{\text{Number of successes with } I_A}{\text{Number of applications of } I_A}$$

Mithilfe einer Zufallszahl zwischen 0 und 1, hier z genannt, und einer der beiden Wahlwahrscheinlichkeiten α_S und α_A wird dann die diesmal zu wählende Intervention bestimmt. Wir erläutern hier die Vorgehensweise anhand von α_S ; genauso gut könnte α_A verwendet werden (da immer gilt: $\alpha_A = 1 - \alpha_S$ entspricht $\alpha_S = 1 - \alpha_A$):

Tabelle 1: Fiktiv angenommene Erfolgsraten der beiden Interventionen Quark und Kohl in sechs Simulationsszenarien

Intervention	Erfolgsraten klein		Erfolgsraten mittel		Erfolgsraten hoch	
	Quark	Kohl	Quark	Kohl	Quark	Kohl
Unterschied in Erfolgsrate klein (5 Prozentpunkte)	8%	3%	55%	50%	97%	92%
Unterschied in Erfolgsrate gross (17 Prozentpunkte)	20%	3%	57%	40%	97%	80%

if $z \leq \alpha_s \rightarrow$ select I_s for next client

if $z > \alpha_s \rightarrow$ select I_A for next client

In Worte gefasst bedeutet dies also: Zunächst einmal sammelt der Algorithmus mit beiden Interventionen erste Erfahrungen. Sobald Informationen über die Erfolgswahrscheinlichkeiten der beiden Methoden vorliegen, werden diese sich vermehrenden Informationen für jede nächste Entscheidung mit einbezogen. Je erfolgreicher sich eine Intervention in den gemachten Erfahrungen im Vergleich zur anderen Intervention gezeigt hat, desto häufiger wird sie gewählt. Und diese Präferenz wird mit zunehmenden Erfahrungen mit den beiden Methoden (Anzahl Wechsel) verstärkt. Dank der Vorgehensweise mit der Zufallszahl für die Wahl der Intervention wird jedoch auch die bisher weniger erfolgreiche Intervention ab und zu wieder einmal zum Zug kommen. Dieses Verhalten ist wichtig, da es durchaus vorkommen kann, dass die eigentlich erfolgreichere Intervention sich bei den ersten Erfahrungen zufälligerweise als weniger erfolgreich zeigt. Durch die wiederholten Anwendungen beider Interventionen werden sich die beobachteten Erfolgsraten im Verlauf der Zeitreihe immer mehr den wahren Erfolgsraten annähern. Für eine sinnvolle Anwendung dieses Algorithmus müssen folgende Bedingungen erfüllt sein:

- Es ist unklar, welche der beiden Interventionen in der betreffenden Situation vorzuziehen ist; es gibt keine wissenschaftliche Evidenz dafür, dass die eine der beiden Interventionen besser geeignet sein könnte als die andere.
- Alle Klientinnen haben nach der Beurteilung der Hebamme die gleiche Ausgangslage, die sich für die Durchführung der ausgewählten Interventionen eignen würde. Sie zeigen in der Regel dieselben im Voraus definierten Symptome.
- Der Erfolg/Misserfolg der Intervention muss klar benennbar sein und immer nach denselben, von der Hebamme im Voraus definierten Kriterien bestimmt werden. Hier fließen subjektive Faktoren sowohl von Seiten der Klientin als auch von Seiten der Hebamme mit ein. Diese sind jedoch Teil der Methode, denn es geht um die Frage: Welche Intervention ist in meiner Praxis erfolgversprechend?
- Die Klientinnen kommen zeitlich nacheinander und der Erfolg/Misserfolg der vorhergehenden Klientin ist bei der nächsten Klientin bereits bekannt.

Selbstverständlich kann es sich bei den beiden zu vergleichenden Interventionen auch um eine Intervention versus eine Nicht-Intervention (beispielsweise bei Sodbrennen

in der Schwangerschaft: Akupunktur versus keine Behandlung) handeln, oder um dieselbe Intervention in verschiedenen Dosierungen (beispielsweise verschiedene Beratungsinhalte zum Kaffeekonsum in der Schwangerschaft [5]).

Für die Simulation in der vorliegenden Arbeit wird die Entscheidung zwischen Quark und Kohl bei verstärkter initialer Brustdrüsenanschwellung in der Stillzeit als Beispielszenario verwendet. Die Sicherheit in der Diagnosestellung wird dabei vorausgesetzt. In der entsprechenden Leitlinie [3] wird konsentiert:

„Aufgrund langjähriger Erfahrungen aus der Praxis können Auflagen in Form von Kühlkissen, Kohlblättern oder Quark sowie die Anwendung einer Tiefdruckmassage zur symptomatischen Behandlung eingesetzt werden.“

Ein Cochrane Review [7] bezieht sich ebenfalls auf verschiedene Maßnahmen bei Brustdrüsenanschwellung, darunter auch unterschiedliche Kohlapplikationen: Bei Vergleichen von Kohl in Raumtemperatur versus gekühlten Kohlblättern, von gekühlten Kohlblättern versus kalten Packungen und von Kohlcreme versus Placebocreme wurden keine Unterschiede festgestellt. Alle Maßnahmen erzielten eine gewisse Linderung. Zu Quark wurde keine Studie eingeschlossen.

Um die Funktionsweise des Algorithmus simulieren zu können, nehmen wir an, dass wir die wahren Erfolgsraten von Quark und Kohl kennen würden und bestimmen willkürlich, dass Quark die erfolgreichere Methode sei. Der Erfolg sei für unser Simulationsbeispiel als vollständiger Rückgang der Rötung und Schwellung nach 24 Stunden Behandlung festgelegt. Wir testen sechs unterschiedliche Szenarien mit je einer erfolgreicheren (Quark) und einer weniger erfolgreichen Intervention (Kohl). Die gewählten Erfolgsraten (siehe Tabelle 1) werden auf verschiedenen Niveaus angesiedelt (im unteren, mittleren und oberen Prozentbereich) und die Differenz in der Erfolgsrate je einmal mit 5 und einmal mit 17 Prozentpunkten angesetzt. Die verwendeten Erfolgsraten haben nichts mit den tatsächlichen (unseren Wissens nach bislang unbekannt) Erfolgsraten von Quark und Kohl bei verstärkter initialer Brustdrüsenanschwellung in der Stillzeit zu tun, sondern sind frei erfunden für die Veranschaulichung des Algorithmus.

Für jedes der sechs Simulationsszenarien mit den vorgegebenen Erfolgsraten wurden 10.000 Simulationen für eine Reihe von 300 aufeinanderfolgenden Klientinnen gestartet: Für jede Klientin wurde als Behandlungserfolg eine Zufallszahl ausgegeben („gewürfelt“), wobei die

Wahrscheinlichkeit, mit der ein Erfolg verbucht werden konnte, der vorgegebenen Erfolgsrate entsprach. Alle Simulationen und Analysen in dieser Studie wurden mit der Statistik Software R durchgeführt [9].

Ergebnisse

In den folgenden beiden Abschnitten werden die Ergebnisse der Simulation von Klientinnenreihen mit dem Algorithmus beschrieben. Im ersten Abschnitt zeigen wir drei Beispiele von simulierten Reihen, im zweiten Abschnitt werden die mittleren Erfolgsraten aus allen 10.000 Simulationen für die sechs verschiedenen Szenarien präsentiert und verglichen.

Beispiele von Simulationsreihen

Abbildung 1 zeigt ein Beispiel einer Simulation mit dem Algorithmus für die ersten 100 Klientinnen. Es ist eine der 10.000 gerechneten Simulationen für das Szenario mit den mittleren Erfolgsraten und dem größeren Unterschied zwischen den Erfolgsraten der beiden Interventionen, in dem die Erfolgsrate für Quark bei 57% und diejenige für Kohl bei 40% angesetzt ist.

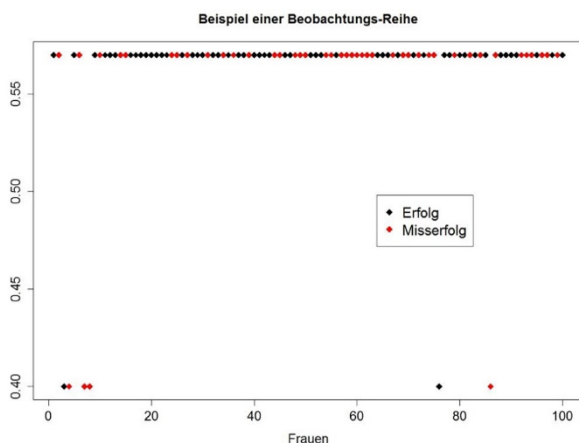


Abbildung 1: Beispiel einer simulierten Reihe des Szenarios (1 Behandlung = 1 Raute) mit Erfolgsrate Quark (bei 57% oben im Diagramm) und Erfolgsrate Kohl (bei 40% unten im Diagramm) bei 100 Frauen

x-Achse: Reihe der Klientinnen von 1 bis 100; y-Achse: Erfolgsrate; schwarz: Erfolg; rot: Misserfolg.

In diesem ersten Beispiel wurde mit der Zufallsauswahl (Schritt 1 des Algorithmus, vgl. Abschnitt Methode) Quark als Startintervention gewählt. Die Quarkanwendung führte bei der ersten fiktiven Klientin zum Erfolg, bei der zweiten nicht, weshalb für die dritte Klientin zu Kohl gewechselt wurde (Schritt 3 des Algorithmus, vgl. Abschnitt Methode). Auch bei Kohl zeigt sich zunächst ein Erfolg, danach bei der vierten fiktiven Klientin ein Misserfolg. Ab der fünften Klientin wurde also jede weitere Entscheidung aufgrund der jeweiligen aktuellen Wahlwahrscheinlichkeiten α für die beiden Interventionen Quark und Kohl und einer Zufallszahl gefällt (Schritt 5 des Algorithmus, vgl.

Abschnitt Methode). Kohl hatte unter den ersten vier Anwendungen nur einmal einen Erfolg gezeigt, Quark hingegen zwei Mal. Quark wird deshalb in der Folge viel öfter gewählt als Kohl. Aber auch Kohl kommt im Laufe der Reihe noch zwei Mal zum Zug, was dem Mechanismus mit der Zufallszahl zu verdanken ist.

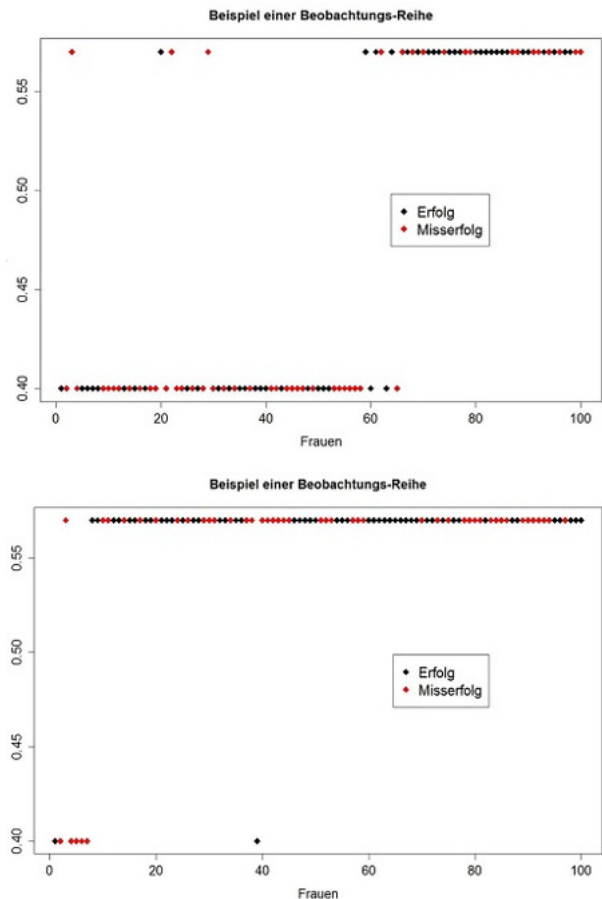


Abbildung 2: Zwei weitere Beispiele von simulierten Reihen des Szenarios (1 Behandlung = 1 Raute) mit Erfolgsrate Quark (bei 57% oben im Diagramm) und Erfolgsrate Kohl (bei 40% unten im Diagramm) bei 100 Frauen

x-Achse: Reihe der Klientinnen von 1 bis 100; y-Achse: Erfolgsrate; schwarz: Erfolg; rot: Misserfolg

Die simulierten Reihen – wie auch tatsächliche Anwendungen eines solchen Algorithmus – sehen durch die Natur des Zufalls jedes Mal anders aus. Für Abbildung 2 haben wir zwei weitere Reihen desselben Simulationsszenarios wie in Abbildung 1 (fiktive Erfolgsraten für Quark 0.57 und für Kohl 0.40) aus den 10.000 Simulationen gewählt zur Darstellung dieser Unterschiedlichkeit. Bei beiden wurde Kohl als Startintervention gewählt und bei beiden zeigte dieser einen Erfolg, gefolgt von einem Misserfolg. Die darauffolgende Quark-Anwendung zeigte in beiden Fällen keinen Erfolg, worauf in beiden Reihen Kohl wieder zum Zug kommt und einen Misserfolg zeigt. Trotz dieser gleichen Ausgangssituation unterscheiden sich die Reihen ab der fünften Klientin beträchtlich. In der Simulationsreihe von Abbildung 2 zeigt Kohl per Zufall am Anfang viele Erfolge und wird deswegen bis etwa zur 60. Klientin vorwiegend gewählt. Hier zeigt sich die

Wichtigkeit des Mechanismus mit der Zufallszahl, der dazu führt, dass trotz der vermeintlichen Überlegenheit von Kohl auch Quark hin und wieder gewählt wird. Dadurch kann sich – früher oder später, in diesem Beispiel nach ca. 60 Anwendungen – die tatsächliche Überlegenheit von Quark zeigen und der Algorithmus entscheidet auf lange Frist trotz dieser ersten Erfahrungen vorwiegend für Quark. In der Simulation auf dem unteren Diagramm von Abbildung 2 hingegen zeigt Kohl gerade andersherum per Zufall viele Misserfolge in den ersten Anwendungen und wird deshalb schon nach wenigen Klientinnen nur noch in Ausnahmefällen gewählt.

Mittlere Erfolgsraten in den Simulationsszenarien

Die Beispiele im vorhergehenden Abschnitt sollen illustrieren, wie einzelne Anwendungen des Algorithmus aussehen und wie verschiedenen Anwendungsreihen des gleichen Szenarios (gleiche wahre Erfolgsraten) sein können. Um etwas über die Güte des Algorithmus aussagen zu können, betrachten wir nun in diesem Abschnitt nicht mehr einzelne Simulationen, sondern die mittleren Erfolgsraten über 10.000 Simulationen desselben Szenarios.

Abbildung 3 zeigt die mittlere Erfolgsrate des Algorithmus für das bereits im vorhergehenden Abschnitt verwendete Szenario mit den Erfolgsraten der beiden Interventionen im mittleren Bereich. Die grünen horizontalen Linien markieren die zugrundegelegten fiktiven Erfolgsraten der beiden Interventionen: 0.57 für Quark und 0.40 für Kohl. Die blaue Linie bei 0.485 entspricht der mittleren Erfolgsrate, wenn man die Hälfte der Klientinnen mit der einen, die andere Hälfte mit der anderen Intervention behandelt, also z.B. abwechselnd Quark und Kohl verwenden würde. Die maximal erreichbare mittlere Erfolgsrate entspricht in diesem Szenario 0.57, der Erfolgsrate von Quark. Die Punkte in Abbildung 3 zeigen die tatsächlich beobachteten Erfolgsraten für die erste bis dreihundertste Klientin gemittelt über die 10.000 durchgeführten Simulationen. Für die erste Klientin liegt diese mittlere Erfolgsrate per Definition auf der blauen Linie, bei 0.485, da ja in der Hälfte der Fälle Quark und in der anderen Hälfte der Fälle Kohl als Startintervention gewählt wird. Bereits ab der zweiten Klientin liegt die mittlere Erfolgsrate jedoch über dieser durchschnittlichen Erfolgsrate der beiden Interventionen. Die beobachteten mittleren Erfolgsraten in diesem Simulationsszenario streuen um eine stetig steigende Kurve, die sich auf lange Sicht dem Optimum von 0.57 annähert. Der Anstieg ist anfangs steil und wird dann immer flacher. Die Streuung entsteht durch Zufallsschwankungen, weil hier „nur“ 10.000 Simulationen durchgeführt wurden. Je mehr Simulationen man verwenden würde, desto kleiner würde diese Streuung werden.

In andere Worte gefasst: Bereits ab der zweiten Klientin kann der Algorithmus die Erfahrung, die man mit der ersten Klientin gemacht hat, gewinnbringend einsetzen. Und mit jeder weiteren Klientin lernt der Algorithmus

weiter, so dass die Erfolgchance für jede weitere Klientin etwas weiter ansteigt.

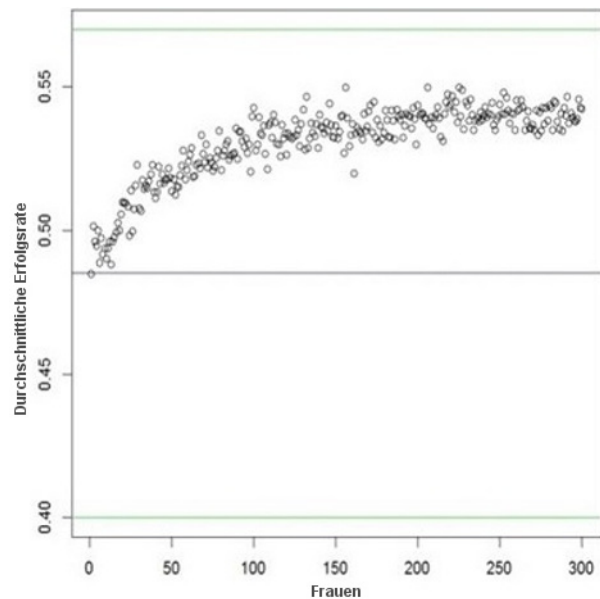


Abbildung 3: Durchschnittliche Erfolgsrate über 10.000 Simulationen des Szenarios mit Erfolgsrate Quark 57% und Erfolgsrate Kohl 40%

In gleicher Weise wie in Abbildung 3 sind in Abbildung 4 für alle sechs Simulationsszenarien (siehe Tabelle 1) die mittleren Erfolgsraten aus den 10.000 Simulationen pro Szenario berechnet und für eine Reihe von 300 Klientinnen dargestellt.

Die Diagramme oben in Abbildung 4 zeigen die beiden Szenarien mit den kleinen Erfolgsraten für beide Interventionen, die mittleren Diagramme die Szenarien mit den mittleren Erfolgsraten und die unteren beiden Diagramme die Szenarien mit den hohen Erfolgsraten. In den Diagrammen auf der linken Seite sind die drei Szenarien mit dem größeren Unterschied von 17 Prozentpunkten, in den Diagrammen auf der rechten Seite diejenigen mit dem kleineren Unterschied von 5 Prozentpunkten zwischen den Erfolgsraten der beiden Interventionen dargestellt. Der Lerneffekt des Algorithmus bereits ab der zweiten Klientin und der stete Anstieg der mittleren Erfolgsrate mit Annäherung an das Optimum sind bei allen Szenarien zu beobachten. Der Vergleich der linken mit den rechten Diagrammen in Abbildung 4 zeigt, dass bei den Szenarien mit nur kleinem Unterschied zwischen den Erfolgsraten der beiden Interventionen die Streuung deutlich stärker ist. Vor allem im Szenario mit den mittleren Erfolgsraten ist sie so groß, dass die eigentliche Kurve nur noch undeutlich erkennbar ist. Außerdem zeigt der Vergleich auch, dass der Algorithmus bei den Szenarien mit den größeren Unterschieden in der Erfolgsrate deutlich schneller lernt als bei denjenigen mit den kleineren Unterschieden. Dies ist auch intuitiv einleuchtend: Wenn Quark in Wirklichkeit sehr viel erfolgreicher ist als Kohl, ist das schneller feststellbar, als wenn Quark nur ein bisschen erfolgreicher ist als Kohl.

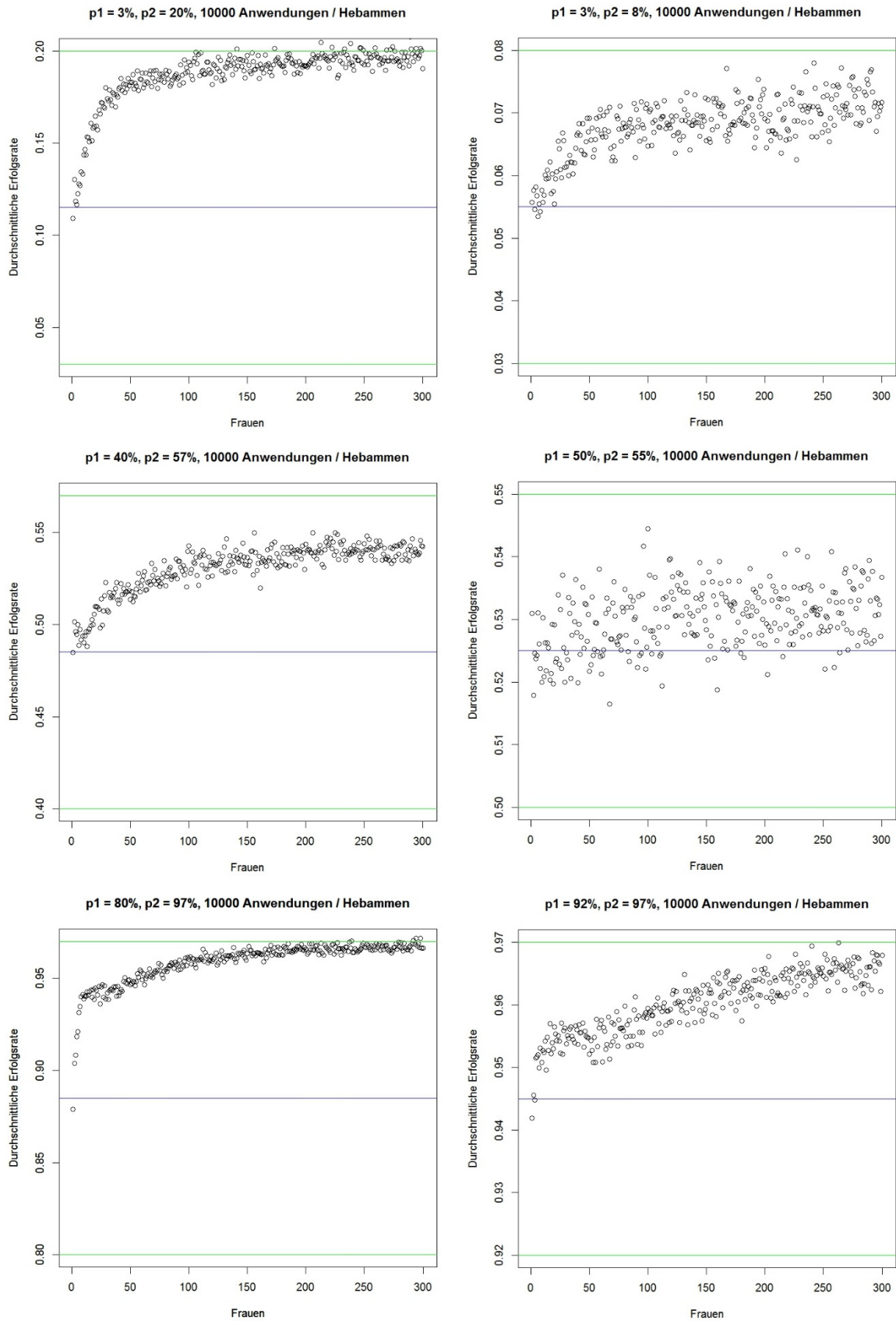


Abbildung 4: Durchschnittliche Erfolgsraten über 10.000 Simulationen für die sechs Simulationsszenarien: Erfolgsraten klein (oben), mittel (Mitte), hoch (unten) bei jeweils 300 Frauen
 x-Achse: Reihe der Klientinnen von 1 bis 300; y-Achse: Durchschnittliche Erfolgsrate.

Der Vergleich zwischen den Szenarien mit den kleinen, mittleren und hohen Erfolgsraten der beiden Interventionen (Abbildung 4) zeigt, dass der Lerneffekt in den Szenarien mit den kleinen und hohen Erfolgsraten schneller ist als in denjenigen mit den mittleren. Das kommt daher, dass bei Unterschieden in festen Prozentpunkten, wie wir sie für die Szenarien gewählt haben, der relative Unterschied im mittleren Prozentbereich am kleinsten ist. Bei kleinen Erfolgsraten ist er groß, bezogen auf die Erfolge; bei großen Erfolgsraten bezogen auf die Misserfolge. Und diese vergleichsweise großen relativen Unterschiede in den Erfolgsraten bei den Szenarien im tiefen und hohen Prozentbereich führen dazu, dass der Algorithmus schneller lernen kann, welche der beiden Interventionen die erfolgreichere ist.

Diskussion

Die Frage danach, ob es eine für den Praxisalltag der Hebammen taugliche Methode gibt, die zwischen zwei als gleichwertig erscheinenden Interventionen Entscheidungshilfe bietet, kann positiv beantwortet werden.

Es wird immer Themen geben, für die Reviews und Leitlinien (noch) keinen eindeutigen Vorteil einer Maßnahme gegenüber einer anderen möglichen Intervention darstellen können und sie als gleichwertig empfehlen. Wahrscheinlichkeitsberechnungen wie der hier vorgestellte Algorithmus könnten Hebammen in solchen Situationen eine sinnvolle Entscheidungshilfe bieten. Der Subjektivität des eigenen Settings (Klientel, Durchführungsweise der Behandlung) wird dabei stets Rechnung getragen. Die gemeinsame, informierte Entscheidung wird dadurch nicht beeinträchtigt, sondern eher dadurch bereichert, dass die Hebamme die von ihr gemachten Erfahrungen in Zahlen fassen kann.

Interventionsalternativen mit großen Unterschieden sind nicht zu erwarten, denn sie wären bereits in einer Metaanalyse aufgefallen und in die wissenschaftliche Evidenz eingegangen. Je nach Anzahl der Fälle und damit der gemachten Erfahrungen, aber auch je nach Überlegenheit einer Intervention wird den ratsuchenden Frauen etwas später oder etwas früher vorwiegend die überlegene Intervention angeboten. Wie erwähnt, bevorzugt der Algorithmus bereits ab der zweiten Anwendung häufiger die erfolgreichere Intervention. Eine Hebamme benötigt daher in ihrer eigenen Praxis nicht so viele Fälle wie in den gezeigten Simulationen.

Zu den Vorteilen dieses Algorithmus gehört es, dass er durch das laufende Einbeziehen aller verfügbaren Erfahrungen seinen Vorschlag zur Intervention wählt. Es kann allerdings nicht dargelegt werden, bei welchen Interventionen in welchen Zielgruppen wie viele Fälle benötigt werden, um festzulegen, welche Intervention durchgängig empfohlen werden sollte. Wie erläutert, werden sich die beobachteten Reihen von Entscheidungen für die eine oder andere Intervention und von Erfolgen und Misserfolgen bei jeder Anwendung unterscheiden. Dies liegt zum einen in der Natur des Zufalls, da jedes Mal wieder ande-

re Klientinnen mit anderen Voraussetzungen nacheinander beobachtet werden, zum anderen in der Art der Anwendung durch verschiedene Hebammen. Diese Aspekte können als Limitationen der Methode verstanden werden. Ein solches Vorgehen eignet sich auch nicht für alle Fragen in der Geburtshilfe, denn Erfolg bzw. Misserfolg der Intervention müssen eindeutig zu beurteilen sein und zwar relativ zeitnah, bevor die nächste Klientin mit vergleichbarem Behandlungsbedarf erscheint. Als vorteilhaft einzustufen ist es, dass dieses Vorgehen auch bei Maßnahmen geeignet ist, die nur kleine Erfolgsraten aufweisen.

Schlussfolgerungen

Im Sinne der allgemeinen Bemühungen, evidenzbasiert zu arbeiten, könnte eine solche Methode für die eigene Praxis eine sinnvolle Ergänzung zu den vorhandenen Leitlinien, Reviews und Studienergebnissen darstellen, wenn in Metaanalysen Interventionsalternativen als wirksam beschrieben werden. Sie hilft bei der Auswahl der Alternative, welche in der Praxis der Hebamme die größere Wirksamkeit hat. Die Hebamme wird nie mit letzter Sicherheit sagen können, dass Intervention A, die in manchen Fällen auch die gekonnte Nichtintervention bedeuten kann, grundsätzlich wirksamer ist als Intervention B, aber ihr Vorgehen beruht immer auf den bisherigen Erkenntnissen. Diese Anwendung könnte auch auf ganze Hebammen- oder Klinikteams übertragen werden. Mit diesem Vorgehen würde eine Entscheidungsgrundlage in der Praxis geschaffen werden, die jedoch nicht als statisch anzusehen ist, sondern sich laufend an die Empfehlungen aktueller Leitlinien und Reviews anpassen soll. Für Hebammen und andere Berufe im Gesundheitswesen, die ihre eigenen Erfahrungen strukturiert erfassen und systematisch umsetzen wollen, könnte dies ein geeignetes Hilfsmittel sein. Die Methode wurde allerdings bislang noch nicht in der Praxis erprobt. Reaktionen der Leserschaft auf diesen Artikel werden daher von den Autorinnen mit besonderem Interesse aufgenommen.

Anmerkungen

Anmerkung der Herausgeberin

Auf Anregung beider Reviewerinnen wurde Herr Dubben angeschrieben und gebeten, einen Kommentar zum Einsatz der von ihm und Herrn Beck-Bornholdt vorgestellten Methode zu verfassen. Leider konnte der Kommentar bis zum Zeitpunkt der Publikation nicht eingeholt werden.

Interessenkonflikt

Die Autorinnen erklären, dass keine Interessenkonflikte vorliegen.

Literatur

1. Bayes M, Price M. An Essay towards Solving a Problem in the Doctrine of Chances. By the Late Rev. Mr. Bayes, F.R.S. Communicated by Mr. Price, in a Letter to John Canton, A.M.F.R.S. Philosophical Transactions. 1763;53:370–418. DOI: 10.1098/rstl.1763.0053
2. Beck-Bornholdt HP, Dubben HH. Der Schein der Weisen: Irrtümer und Fehltritte im täglichen Denken. 7th ed. Reinbeck: Rowohlt Taschenbuch Verlag; 2003. German.
3. Deutsche Gesellschaft für Gynaekologie und Geburtshilfe (DGGG). S3 Leitlinie: Therapie entzündlicher Brusterkrankungen in der Stillzeit; 2013. [Zugriff/access Jun 2018] Verfügbar unter/available from: http://www.awmf.org/uploads/tx_szleitlinien/015-0711_S3_Therapie_entz%C3%BCndlicher_Brustentz%C3%BCndungen_Stillzeit_2__2013-02_01.pdf. German.
4. Huskins WC, Fowler VG, Evans S. Adaptive Designs for Clinical Trials: Application to Healthcare Epidemiology Research. Clin Infect Dis. 2018;66(7):1140–6. DOI: 10.1093/cid/cix907
5. Jahanfar S, Jaafar SH. Effects of restricted caffeine intake by mother on fetal, neonatal and pregnancy outcomes. Cochrane database Syst Rev. 2015;(6):CD006965. DOI: 10.1002/14651858.CD006965.pub4
6. Liang Y, Carriere KC. Stratified and randomized play-the-winner rule. Stat Methods Med Res. 2008;17(6):581–93. DOI: 10.1177/0962280207081606
7. Mangesi L, Zakarija-Grkovic I. Treatments for breast engorgement during lactation. Cochrane database Syst Rev. 2016;(6):CD006946. DOI: 10.1002/14651858.CD006946.pub3
8. Phupong V, Hanprasertpong T. Interventions for heartburn in pregnancy. Cochrane database Syst Rev. 2015;(9):CD011379. DOI: 10.1002/14651858.CD011379.pub2
9. R Core Team. R - A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2014 [Zugriff/access Jun 2018]. Verfügbar unter/available from: <http://www.R-project.org/>
10. Spiegelhalter DJ. Incorporating Bayesian Ideas into Health-Care Evaluation. Statistical Science. 2004;19(1):156–74.
11. Wold HOA. A study in the analysis of stationary time series. Stockholm: Almqvist & Wiksell; 1938.
12. Zelen M. Play the Winner Rule and the Controlled Clinical Trial. Journal of the American Statistical Association. 1969;64(325):131–46. DOI: 10.1080/01621459.1969.10500959

Korrespondenzadresse:

Dr. rer. medic. Christine Loytved
ZHAW, Technikumstrasse 71, CH-8401 Winterthur
christine.loytved@zhaw.ch

Bitte zitieren als

Loytved C, Erdin R. Eine neue Methode als Entscheidungshilfe bei unklarer wissenschaftlicher Evidenz: Test-Implementierung und Simulation. GMS Z Hebammenwiss. 2018;5:Doc01. DOI: 10.3205/zhwi000011, URN: urn:nbn:de:0183-zhwi0000111

Artikel online frei zugänglich unter

<http://www.egms.de/en/journals/zhwi/2018-5/zhwi000011.shtml>

Eingereicht: 20.09.2017

Angenommen: 16.04.2018

Veröffentlicht: 28.12.2018

Copyright

©2018 Loytved et al. Dieser Artikel ist ein Open-Access-Artikel und steht unter den Lizenzbedingungen der Creative Commons Attribution 4.0 License (Namensnennung). Lizenz-Angaben siehe <http://creativecommons.org/licenses/by/4.0/>.

A new method supporting decision-making in case of unclear scientific evidence: Test implementation and simulation

Abstract

Background: Many health workers are aware of the results of reviews like those of the Cochrane Collaboration. Some results of these reviews show, at that stage of the research, no advantage for one of two promising interventions. In these cases, Beck-Bornholdt and Dubben propose a modified, never-change-a-winning-team algorithm. Similar algorithms are used in cases of study group assignments or adjustments to the design of a study in progress.

Research question: Is the method proposed by Beck-Bornholdt and Dubben in 2003 helpful for the daily work of midwives when they have to choose between two interventions with similar evidence of success?

Methodology: The application of the algorithm is being simulated for possible use by health workers. This includes all existing experiences made with both interventions to decide on the intervention for the next person to be treated.

Results: Simulations were carried out for various scenarios with different likelihood for success with regard to both interventions. It can be demonstrated that the average success rate in all scenarios is already improved starting with the second person treated, in comparison to the average success rate for both interventions.

Conclusions: The results can serve as a basis for discussion for the applicability of the suggested method. If the evidence is unclear, the algorithm can support the decision of health workers for one of two possible treatments, with positive effect. The special conditions of the setting in question (clientele, treatment realization) are hereby taken into account in each case.

Keywords: algorithm, intervention, midwifery research

Christine Loytved¹

Rebecca Erdin¹

¹ Institute of Midwifery, Health Department, School of Health Professions, Winterthur, Switzerland

Background

In midwifery, the issue of evidence-based practice has increasingly come to the fore. Midwives themselves insist on the three pillars (scientific evidence, the midwife's experience and the client's preference) being taken into account and this is something that is also required by other parties involved – whether by other healthcare professionals or by clients. On some topics however, such as heartburn during pregnancy, the scientific evidence does not clearly favour a single intervention, which can comprise of either action or non-action. For some topic areas, evidence-based guidelines such as those published by the National Institute for Health and Care Excellence (NICE) or systematic reviews such as those by the Cochrane Collaboration make recommendations supporting different interventions in a specific situation. This particularly applies in situations where, based on the current state of research, there is no apparent advantage for any

of the interventions tested. In the case of heartburn during pregnancy, for instance, it remains unclear whether acupuncture alleviates the condition or whether clients should be advised to adjust their eating habits [8]. Especially for issues that do not involve a significant intervention, such as administering oxytocin for induction of labour, but rather less invasive measures, it is also unlikely that the state of research will improve in near future. In cases like this where reviews or guidelines conclude that either method could result in a successful outcome, midwives need access to a tool supporting the decisions they have to make in their daily work between one of two interventions. Given that the available studies only support decision-making to a limited extent in these instances, it is advisable for midwives to use the internal evidence based on their own practical experience up to now to make their decisions. In other words, the successes and failures experienced in connection with an intervention should be effectively incorporated into future decision-making.

Beck-Bornholdt and Dubben [2] propose a potential solution in form of a modified never-change-a-winning-team algorithm. Based on this algorithm, the decision on which intervention to use for the next client presenting with a relevant diagnosis is made according to a specific formula: in each case, the decision should be made based on all experiences of both intervention options so far. This idea, also referred to as the “play-the-winner rule”, draws on the works of Bayes [1] and Wold [11], although similar approaches had already been described in earlier studies [12]. Bayes’ theorem was proposed for the following applications:

1. To create a rule identifying the interim results that would make it ethically necessary to stop a randomised controlled trial,
2. to determine the appropriate drug dosage for a randomised controlled trial and
3. to interpret evidence from trials conducted to date [10].

Similar methods were discussed for use in the assignment for trial participants to groups [6], adjustments to the design of a trial when it is in progress [4] and for evaluations in the healthcare sector [10]. Beck-Bornholdt and Dubben [2] use Bayes’ theorem by drawing on the successes and failures observed to date to help individuals decide between two interventions considered to be similarly promising. To date, this method has not been used as a decision-making tool in the day-to-day practice of healthcare professionals. The application of the algorithm for a modified “play-the-winner rule” should be as user-friendly as possible. One plausible option would be to develop an app where the user only has to enter the relevant interventions and whether or not they were successful. The application of the algorithm should not be rigid, however, and in no way should it restrict the client’s right to self-determination. The algorithm should only serve as an instrument for the midwife to create a pool of experience more effectively and rapidly in some areas. These experiences would otherwise still be acquired but not systematically processed.

Objective and research question

If, based on systematic reviews and evidence-based guidelines, the conclusion can be drawn that two interventions are equally promising, midwives (and other healthcare professionals) should have access to a decision-making tool based on current scientific knowledge to enable them to provide their clients with the best possible care. This should not infringe the client’s right of self-determination, however. The client’s individual circumstances should also be taken into account. If, after weighing up all the relevant factors, no clear-cut decision can be made in favour of one intervention and against a second, the midwife should receive support. The purpose of the tool described here is to help the midwife to decide which recommendation to make in the process of in-

formed shared decision making. If her recommendation is implemented, the midwife acquires further experience, which will, in turn, help her to make future recommendations.

This study therefore seeks to address the following question: Is the method proposed by Beck-Bornholdt and Dubben [2] a helpful decision-making tool for the day-to-day work of midwives when they are faced with a choice between two interventions shown in the literature as having similar evidence of success?

Method

To facilitate the decision as to which intervention should be used, the modified “play-the-winner rule” algorithm is applied as follows:

1. For the first client one of the two interventions is defined at random as the starting intervention (e.g. by tossing a coin). This is termed I_S .
2. Provided that I_S is successful, the intervention is repeated for the next client.
3. When I_S fails for the first time, the midwife switches to the other intervention. This is termed I_A .
4. Again, I_A is then continually repeated for each subsequent client as long as it remains successful.
5. When I_A fails for the first time, for each subsequent client the probability of selection (α) is calculated for one of the two interventions (termed α_S and α_A).

$$\alpha_S = \frac{\beta_S}{\beta_S + \frac{1}{\beta_S}} \quad \text{and} \quad \alpha_A = \frac{\beta_A}{\beta_A + \frac{1}{\beta_A}}$$

and

$$\beta_S = \left(\frac{\text{Success rate } I_S + 0.28}{\text{Success rate } I_A + 0.28} \right)^w \quad \text{and} \quad \beta_A = \left(\frac{\text{Success rate } I_A + 0.28}{\text{Success rate } I_S + 0.28} \right)^w$$

where w is the number of switches between the two methods and the success rate of the individual interventions is calculated using the following formula:

$$\text{Success rate } I_S = \frac{\text{Number of successes with } I_S}{\text{Number of applications of } I_S} \quad \text{and}$$

$$\text{Success rate } I_A = \frac{\text{Number of successes with } I_A}{\text{Number of applications of } I_A}$$

Using a random value between 0 and 1 (here termed z) and one of the two probabilities α_S and α_A , the intervention to be chosen this time is then determined. Here, we outline the approach based on α_S but we could equally have used α_A (because the following is always true: $\alpha_A = 1 - \alpha_S \equiv \alpha_S = 1 - \alpha_A$ would have applied):

- if $z \leq \alpha_S \rightarrow$ select I_S for next client
- if $z > \alpha_S \rightarrow$ select I_A for next client

Put into words: first, the algorithm gathers initial experiences from both interventions. As soon as information is available about the probability of success of each of the two methods, this growing body of information is then incorporated into the next decision. The more successful an intervention appears to be (based on experiences ac-

Table 1: Fictitious success rates of the two interventions 'quark' and 'cabbage' in six simulation scenarios

Intervention	Small success rates		Medium success rates		High success rates	
	quark	cabbage	quark	cabbage	quark	cabbage
Small difference in success rates (5 percentage points)	8%	3%	55%	50%	97%	92%
Large difference in success rates (17 percentage points)	20%	3%	57%	40%	97%	80%

quired), as compared to the other intervention, the more frequently this intervention will be selected. And, with increasing experiences of both methods (number of switches), this preference is reinforced. However, due to the technique of using a random value to select the intervention, the intervention which has so far been less successful is also occasionally given another chance. This approach is important since it is distinctly possible that the intervention which is actually the more successful appears, by chance, to be less successful during the initial experiences. By repeatedly applying both interventions, the success rates observed over the course of the time series converge on the true success rates.

For the application of this algorithm to be useful, the following conditions must be met.

- It is unclear which of the two interventions should be favoured in the given situation; there is no scientific evidence that one intervention might be more suitable than the other.
- According to the midwife's assessment, all clients have the same starting situation, meaning the implementation of the interventions selected would be suitable for them. They generally have the same predefined symptoms.
- The success/failure of the intervention must be clearly identifiable and always determined according to the same criteria, predefined by the midwife. Here, subjective factors both from the side of the client and the midwife play a role. However, these are part of the methodology since the question we are seeking to address is: Which is the most promising intervention for my daily work?
- The clients come one after the other and the success/failure of the intervention for the preceding client is already known by the time the next client is treated.

Obviously the two interventions to be compared can also comprise intervention versus non-intervention (for instance, in the case of heartburn during pregnancy: acupuncture versus no treatment), or the same intervention with different dosages (for example, different advice on coffee consumption during pregnancy [5]).

For the simulation in the present study, we use the decision between applying quark (a form of curd cheese) and cabbage in the case of excessive initial breast engorgement during the lactation period as an example scenario. We assume that the diagnosis is correct. According to the relevant guideline [3], the consensus is:

'Based on many years of experience from midwifery practice, the conclusion is that cold compresses in the

form of cold packs, cabbage leaves or quark as well as deep tissue massage can be applied for symptomatic treatment.'

A Cochrane Review [7] also refers to different treatments for breast engorgement, including different cabbage leaf applications: trials looking at cabbage leaves showed no difference between room temperature and chilled cabbage leaves, between chilled cabbage leaves and gel packs and between cabbage cream and the inactive cream. However, all forms of treatment provided some relief. There was no trial looking at the application of quark.

In order to be able to simulate the functioning of the algorithm, we assume that we know the true success rates of quark and cabbage and so arbitrarily define that quark is the more successful of the two methods. For our simulation example, success is defined as complete elimination of redness and swelling within 24 hours of treatment. We tested six different scenarios, each with the more successful (quark) and less successful (cabbage) intervention. We selected success rates at different levels (in the lower, mid and upper percentage range) (see Table 1) and the difference between the success rates was set once at 5 percentage points and once at 17 percentage points for each intervention. The success rates used have nothing to do with the true success rates of quark and cabbage in excessive initial breast engorgement during the lactation period, which, to our knowledge are not yet known, but rather are generated for the purpose of illustrating the algorithm.

For each of the six simulation scenarios with the predetermined success rates, 10,000 simulations were started for a series of 300 consecutive clients: to indicate treatment success, each client was allocated a random value ("a roll of the dice"), where the probability of a successful outcome corresponded with the predetermined success rate.

All simulations and analyses in this study were conducted using the statistical software R [9].

Results

In the next two sections, we describe the results of the simulation for the client series using the algorithm. In the first section, we show three examples of simulated series and, in the second section, the mean success rates across all 10,000 simulations for the six different scenarios are presented and compared.

Examples from the simulation series

Figure 1 shows an example of a simulation using the algorithm for the first 100 clients. It is one of the 10,000 simulations calculated for the scenario with the mid-range success rates and the greater difference between the success rates of the two interventions, where the success rate for quark is set at 57 percent and for cabbage at 40 percent.

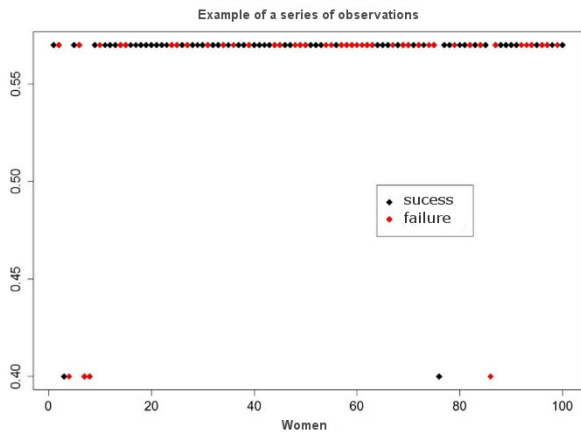


Figure 1: Example of a scenario's simulation series (1 treatment = 1 diamond) in case of quark's success rate (at 57%; see top of diagramme) and cabbage's success rate (at 40%; see bottom of diagramme) in 100 women

x-axis: series of clients from 1 to 100; y-axis: success rate; black: success; red: failure

In this first example, quark is selected at random (step 1 of the algorithm, see Methodology section) as the starting intervention. For the first hypothetical client, the quark compress resulted in a success, for the second, it did not and thus, for the third client, a switch was made to the cabbage treatment (step 3 of the algorithm, see Methodology section). Similarly, the cabbage treatment was initially successful but then, for the fourth hypothetical client, it failed. Therefore, from the fifth client on, each successive decision was made based on the respective current selection probability α for both the quark and cabbage interventions and a random value (step 5 of the algorithm, see Methodology section). Over the first four applications, cabbage was only successful once, whereas quark demonstrated two successful outcomes. Quark was therefore subsequently selected much more frequently than cabbage. But, over the course of the series, cabbage also had two more turns owing to the random value mechanism.

The simulated series – much like real applications of an algorithm like this – are different every time due to the very nature of chance. To show this diversity, in Figure 2, we selected another two series from the same simulation scenario as in Figure 1 (hypothetical success rates of 0.57 for quark and of 0.40 for cabbage) from among the 10,000 simulations. In both series, cabbage was selected as the starting intervention and also in both series the intervention was initially successful and then unsuccess-

ful. The subsequent quark treatment failed in both cases which meant that, in both series, it was the turn of the cabbage treatment again, which then also failed in both cases. Despite the fact that the starting point is the same in these two series, starting with the fifth client, they differ considerably. In the upper simulation series in Figure 2, by chance, cabbage initially demonstrates many successes and is thus the intervention predominantly selected until around the 60th client. Here we can see the importance of the random value mechanism, as this means that, despite the supposed dominance of cabbage, quark is still occasionally selected. Thus, sooner or later (in this example after around 60 applications), the true dominance of quark comes to light and, in the long term, the algorithm settles on quark, irrespective of these initial experiences. In the simulation displayed in the lower diagram in Figure 2, however, precisely the opposite occurs, and, by chance, the cabbage intervention frequently fails during the early applications and thus, after being used for just a few clients, is then only rarely selected.

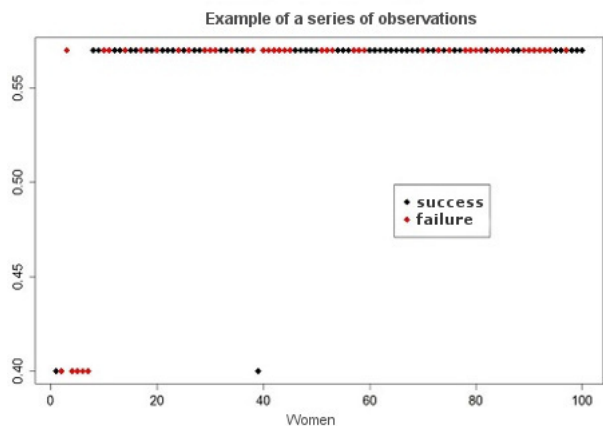
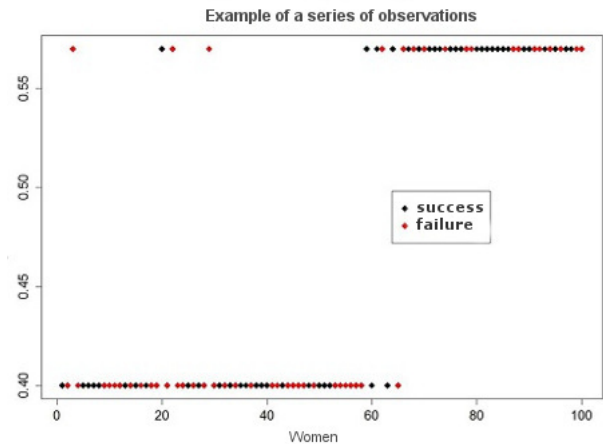


Figure 2: Two other examples of a scenario's simulation series (1 treatment = 1 diamond) in case of quark's success rate (at 57%; see top of diagramme) and cabbage's success rate (at 40%; see bottom of diagramme) in 100 women

x-axis: series of clients from 1 to 100; y-axis: success rate; black: success; red: failure

Mean success rates in the simulation scenarios

The aim of the examples presented in the previous section was to illustrate the various individual applications of the algorithm and show the possible different series of applications of the same scenario (same true success rates). In order to be able to draw any conclusions on the benefits of the algorithm, in this section we will now no longer examine the individual simulations but rather the mean success rates across 10,000 simulations of the same scenario.

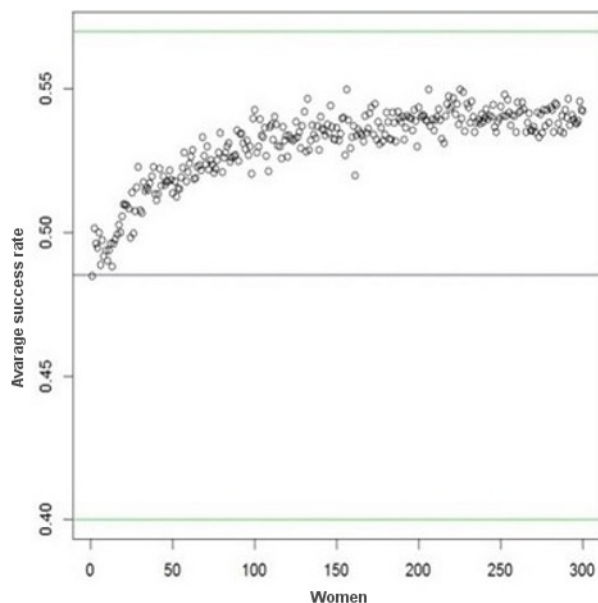


Figure 3: Average success rate of 10,000 simulations of the scenario in case of quark's success rate at 57% and cabbage's success rate at 40%

Figure 3 shows the mean success rate of the algorithm for the scenario already used in the previous section where the success rates of both interventions are in the mid-range. The green horizontal lines indicate the hypothetical success rates used as a basis for both interventions: 0.57 for quark and 0.40 for cabbage. The blue line at 0.485 corresponds to the mean success rate if we were to treat half of the clients using one intervention and the other half using the second, in other words, alternating between quark and cabbage, for example. In this scenario, the maximum achievable mean success rate is 0.57, the success rate for quark. The dots in Figure 3 show the success rates that were actually observed for the first to the 300th client averaged across the 10,000 simulations conducted. For the first client, this mean success rate is, by definition, located on the blue line (with a value of 0.485), since quark was selected as a starting intervention in half of the cases and cabbage in the other half. Already from the second client on, the average success rate has exceeded the mean success rate of the two interventions. The mean success rates observed in this simulation scenario are dispersed along a steadily ascending curve, which, in the long term, con-

verges on the optimum value of 0.57. The increase is initially steep but then levels out. The dispersion occurs due to random fluctuations because we “only” carry out 10,000 simulations. The more simulations conducted, the smaller the dispersion.

In other words: already starting with the second client, the algorithm can profitably use the experience acquired with the first client. Moreover, with each subsequent client, the algorithm continues to learn, meaning that the chance of success continues to increase for each subsequent client.

As in Figure 3, also in Figure 4 the mean success rates are calculated for all six simulation scenarios (see Table 1) based on 10,000 simulations per scenario and shown for a series of 300 clients.

The two diagrams at the top in Figure 4 show the two scenarios with the low success rates for both interventions, the diagrams in the centre of the illustration show the scenarios with the mid-range success rates and the diagrams on the bottom show the scenarios with the high success rates. The diagrams on the left show the three scenarios with the biggest difference between the success rates of the two interventions (17 percentage points) and the diagrams on the right depict the scenarios with the smaller difference (5 percentage points). The learning effect of the algorithm that already starts with the second client as well as the steady increase in the mean success rate converging towards the optimum value can be observed in all the scenarios. A comparison of the diagrams on the left and on the right in Figure 4 shows that for the scenarios where there is just a small difference between the success rates of the two interventions, the dispersion is considerably larger. Particularly in the scenario with the mid-range success rates, the dispersion is so large that the actual curve is barely still recognisable. The comparison also shows that in the scenarios with the larger differences in success rates, the algorithm learns much more rapidly than in the scenarios with the smaller differences. This also intuitively makes sense: if, in reality, quark was far more successful than cabbage, this would be determined more quickly than if quark was only a slightly more successful treatment than cabbage.

A comparison of the scenarios with the low, mid-range and high success rates of the two interventions in Figure 4 shows that the learning effect is stronger in the scenarios with the low and high success rates than in those with the mid-range success rates. This is because with differences in fixed percentage points, such as those we selected for the scenarios, the relative difference is smallest in the mid percentage range. In the case of low success rates, the difference is large relative to the successes, and in the case of high success rates, relative to the failures. And these comparatively large relative differences between the success rates in the scenarios in the lower and higher percentage ranges enables the algorithm to learn more rapidly which of the two interventions is the more successful.

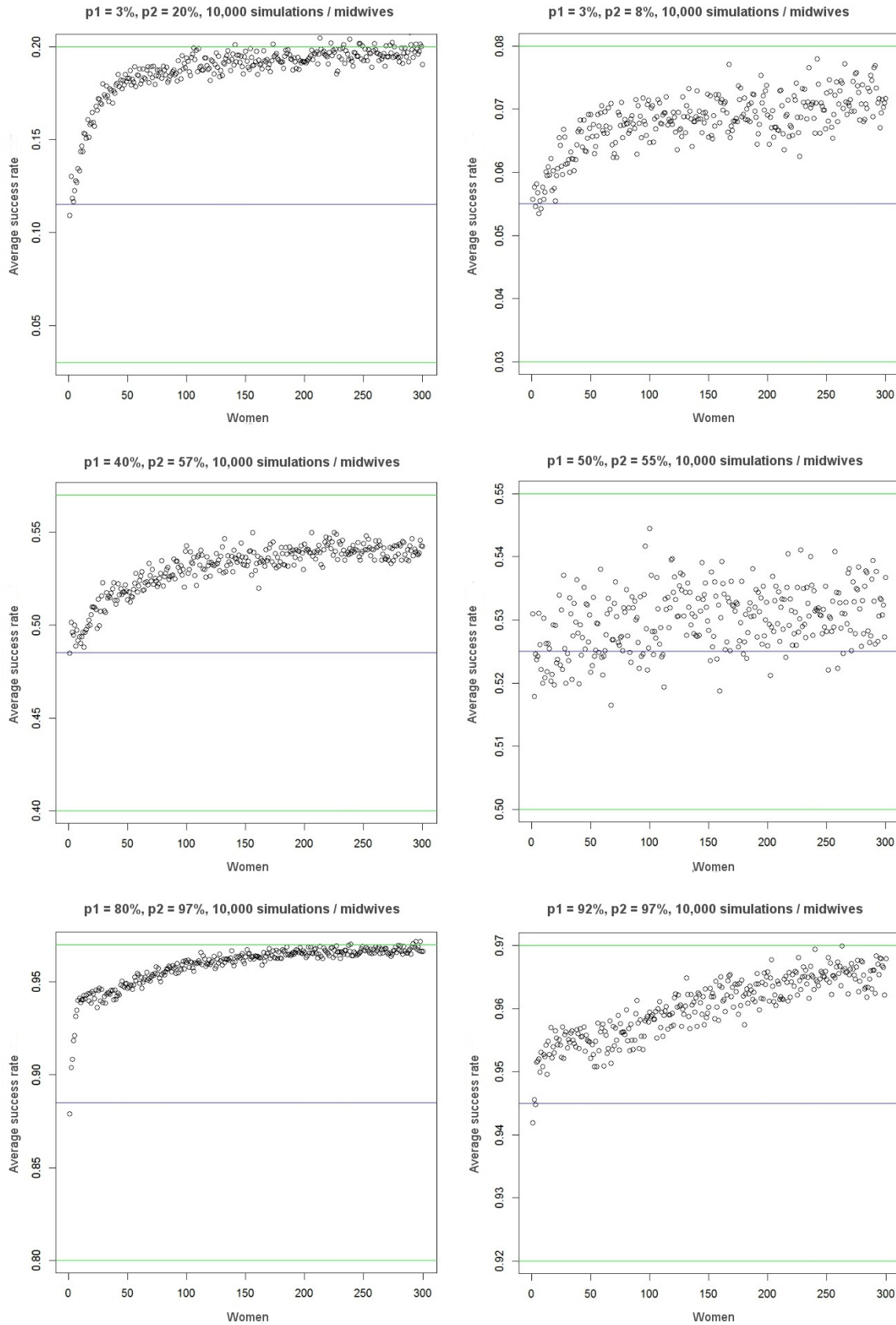


Figure 4: Average success rate of 10,000 simulations relating to six simulation scenarios: small success rate (at the top); medium success rate (central); high success rate (at the bottom); in 300 women each
 x-axis: series of clients from 1 to 300; y-axis: average success rate

Discussion

The answer to the question as to whether there is a helpful method to assist midwives in deciding between two equally promising interventions in their day-to-day work is thus affirmative.

There will always be areas where reviews and guidelines are (as yet) unable to present a clear advantage for one intervention over another and therefore recommend both equally. In situations like this, probability calculations such as the algorithm presented here can provide midwives with a useful decision-making tool. The specific conditions of each individual setting (clientele, how the treatment is implemented) will always be taken into account. The joint shared decision-making will not be compromised by the use of such a tool but rather enriched by it as it enables the midwife to capture her experiences in numbers. Intervention alternatives with significant differences are not to be anticipated as these would have already been apparent in a meta-analysis and incorporated into the scientific evidence. Depending on the number of cases and the experiences acquired through them, but also depending on the dominance of an intervention, the women seeking advice will primarily be offered the dominant intervention slightly later or earlier. As described, the algorithm already more frequently favours the more successful intervention from the second application on. A midwife would therefore not require as many cases in her own day-to-day practice as in the simulations shown. One of the advantages of this algorithm is that, by continuously incorporating all available experiences, it suggests an intervention. However, we are unable to identify the number of cases that the algorithm would require for which intervention and in which target group in order to determine which of the two interventions should always be recommended. As already outlined, the series of decisions for one or the other intervention and the series of successes and failures will be different for each application. This is partly down to the nature of chance because each time a different client with different requirements will be observed and it is partly due to the way different midwives carry out the treatment. These aspects could be seen as limitations of the method. The approach described is also not suitable for all issues arising in the field of midwifery because a prerequisite for the algorithm is that the success or failure of an intervention can be clearly determined relatively quickly before the next client with similar treatment needs appears. On the positive side, this approach is suitable for treatments showing only low success rates.

Conclusion

In line with more general endeavours to work in an evidence-based manner, for practicing midwives, a method like this could be a useful complement to the existing guidelines, reviews and study findings in cases where meta-analyses describe different intervention options as

effective. The method helps midwives to select which alternative is the more effective in their day-to-day practice. A midwife will never be able to say with absolute certainty that intervention A, which, in some cases, might also mean skilled non-intervention, is fundamentally more effective than intervention B but her course of action is always based on her existing knowledge and experiences to date. This application could also be used for entire midwifery teams. The approach would establish a decision-making basis in midwifery practice, which, however, should not be seen as static but rather as a method which should be continuously adapted to the recommendations of current guidelines and reviews. For midwives and other healthcare professionals who would like to capture their experiences in a structured manner and systematically put this experience into practice, this could be a suitable tool. The method has, however, not yet been tested in practice. The readers reactions to this article will therefore be received with particular interest.

Notes

Editor's note

At the suggestion of both reviewers, I wrote to Mr Dubben requesting a comment on the application of the method proposed by him and Mr Beck-Bornholdt. Unfortunately, however, given the short time available, this did not prove possible.

Acknowledgements

We would like to thank Carla Welch, qualified translator, for assisting with the English translation of the manuscript.

Competing interests

The authors declare that they have no competing interests.

References

1. Bayes M, Price M. An Essay towards Solving a Problem in the Doctrine of Chances. By the Late Rev. Mr. Bayes, F.R.S. Communicated by Mr. Price, in a Letter to John Canton, A.M.F.R.S. *Philosophical Transactions*. 1763;53:370–418. DOI: 10.1098/rstl.1763.0053
2. Beck-Bornholdt HP, Dubben HH. *Der Schein der Weisen: Irrtümer und Fehlurteile im täglichen Denken*. 7th ed. Reinbeck: Rowohlt Taschenbuch Verlag; 2003. German.
3. Deutsche Gesellschaft für Gynaekologie und Geburtshilfe (DGGG). S3 Leitlinie: Therapie entzündlicher Brusterkrankungen in der Stillzeit; 2013. [Zugriff/access Jun 2018] Verfügbar unter/available from: http://www.awmf.org/uploads/tx_szleitlinien/015-071I_S3_Therapie_entz%C3%BCndlicher_Brustentz%C3%BCndungen_Stillzeit_2__2013-02_01.pdf. German.

4. Huskins WC, Fowler VG, Evans S. Adaptive Designs for Clinical Trials: Application to Healthcare Epidemiology Research. *Clin Infect Dis.* 2018;66(7):1140–6. DOI: 10.1093/cid/cix907
5. Jahanfar S, Jaafar SH. Effects of restricted caffeine intake by mother on fetal, neonatal and pregnancy outcomes. *Cochrane database Syst Rev.* 2015;(6):CD006965. DOI: 10.1002/14651858.CD006965.pub4
6. Liang Y, Carriere KC. Stratified and randomized play-the-winner rule. *Stat Methods Med Res.* 2008;17(6):581–93. DOI: 10.1177/0962280207081606
7. Mangesi L, Zakarija-Grkovic I. Treatments for breast engorgement during lactation. *Cochrane database Syst Rev.* 2016;(6):CD006946. DOI: 10.1002/14651858.CD006946.pub3
8. Phupong V, Hanprasertpong T. Interventions for heartburn in pregnancy. *Cochrane database Syst Rev.* 2015;(9):CD011379. DOI: 10.1002/14651858.CD011379.pub2
9. R Core Team. R - A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2014 [Zugriff/access Jun 2018]. Verfügbar unter/available from: <http://www.R-project.org/>
10. Spiegelhalter DJ. Incorporating Bayesian Ideas into Health-Care Evaluation. *Statistical Science.* 2004;19(1):156–74.
11. Wold HOA. A study in the analysis of stationary time series. Stockholm: Almqvist & Wiksell; 1938.
12. Zelen M. Play the Winner Rule and the Controlled Clinical Trial. *Journal of the American Statistical Association* 1969;64(325):131–46. DOI: 10.1080/01621459.1969.10500959

Corresponding author:

Dr. rer. medic. Christine Loytved
 ZHAW School of Health Professions, Technikumstrasse
 71, CH-8401 Winterthur
 christine.loytved@zhaw.ch

Please cite as

Loytved C, Erdin R. Eine neue Methode als Entscheidungshilfe bei unklarer wissenschaftlicher Evidenz: Test-Implementierung und Simulation. *GMS Z Hebammenwiss.* 2018;5:Doc01.
 DOI: 10.3205/zhwi000011, URN: urn:nbn:de:0183-zhwi0000111

This article is freely available from

<http://www.egms.de/en/journals/zhwi/2018-5/zhwi000011.shtml>

Received: 2017-09-20

Accepted: 2018-04-16

Published: 2018-12-28

Copyright

©2018 Loytved et al. This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 License. See license information at <http://creativecommons.org/licenses/by/4.0/>.