

2.2.3 Darstellung von Häufigkeiten

Einleitung

Der Erfolg evidenzbasierter Praxis hängt von der klaren und effektiven Kommunikation statistischer Informationen ab (1). Ziel ist, dass die Informationen verstanden, Nutzen und Risiken von Interventionen korrekt eingeschätzt und letztlich informierte Entscheidungen ermöglicht werden. Für die Erstellung einer Information muss sowohl eine inhaltliche Auswahl der zu kommunizierenden Häufigkeiten stattfinden, als auch eine Überprüfung, ob Zahlen belastbar genug sind, dass sie kommuniziert werden können. Wurden Zahlen ausgewählt, ist es für das Gelingen der Kommunikation wichtig, die vorhandene Evidenz zu verschiedenen Darstellungen im Hinblick auf Wirksamkeit und auch auf mögliche adverse Effekte hin zu überprüfen.

Für die Darstellung von statistischen Informationen (Wahrscheinlichkeiten, die Güte diagnostischer Tests sowie Nutzen, Schaden und Nebenwirkungen medizinischer Maßnahmen) werden verbale Deskriptoren eingesetzt. Verbale Deskriptoren sind mehr oder weniger konkrete sprachliche Umschreibungen von Häufigkeiten, z.B. *selten*, *gelegentlich*, *häufig*, *sicher* oder *wahrscheinlich*. Untersuchungen haben allerdings gezeigt, dass sich die Interpretationen sprachlicher Beschreibungen und die resultierende Wahrnehmung von Risiken sowohl interindividuell als auch zwischen medizinischen Laien und Professionellen stark unterscheiden (2). Verbale Angaben zu Nebenwirkungen führen zu Überschätzungen der Wahrscheinlichkeit ihres Auftretens (2).

Ein erster Ansatz, die verbale Beschreibung von Risiken zu standardisieren, erfolgte durch die Europäische Kommission im Jahr 1998 (3). In der Leitlinie zur Lesbarkeit von Informationen zu Medizinprodukten wurde fünf verbalen Deskriptoren jeweils eine definierte numerische Häufigkeit oder Häufigkeitsspanne zugeordnet (3). Auch das Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM) fordert für Angaben zu Häufigkeiten von Nebenwirkungen in Produktinformationen die Verwendung der festgelegten sprachlichen Beschreibung zusammen mit der numerischen Angabe (4). In einer in Deutschland durchgeführten Befragung von Ärztinnen, Ärzten, Apothekerinnen, Apothekern, Juristinnen und Juristen konnten diese den verbalen

Wahrscheinlichkeitsangaben für Nebenwirkungen nicht korrekt die entsprechenden Prozentwerte zuordnen (5).

Für die numerische Darstellung stehen verschiedene Formate zur Verfügung: natürliche Häufigkeiten, Prozentangaben, absolute Risikoreduktion (ARR), relative Risikoreduktion (RRR), *Number Needed to Treat / Screen / Harm* (NNT, NNS, NNH). Bei Angabe natürlicher Häufigkeiten werden unterschiedliche Bezugsgrößen (Denominatoren) gewählt: 1 von 100, 1 von 1000, 1 von 10000. Die Frage nach den Effekten der genannten Formate wurde bereits in mehreren systematischen Übersichtsarbeiten untersucht (1, 6-8). Akl et al. beziehen dabei erstmalig explizit den Endpunkt *persuasiveness* (Überredung) mit ein, der anhand von hypothetischen Entscheidungen gemessen wird (1).

Die Verwendung natürlicher Häufigkeiten galt sehr lange den Prozentangaben überlegen und wurde auch im Kontext der evidenzbasierten Medizin stark befürwortet (9). Studien haben untersucht, wie häufig der positiv prädiktive Wert eines Tests korrekt geschätzt wird, wenn Angaben zur Prävalenz der gesuchten Erkrankung sowie zu Sensitivität und Falsch-positiv-Rate gemacht werden (10-13). Sie haben ergeben, dass der Anteil korrekter Antworten, selbst bei Darstellung der genannten Größen als natürliche Häufigkeiten, sehr niedrig war. Deshalb sollten positiv prädiktive Werte und andere Testgüteparameter direkt angegeben werden, ohne dass die Adressaten der Information die entsprechenden Berechnungen selbst durchführen müssen (11).

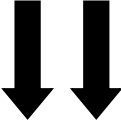
Ein möglicher adverser Effekt der Darstellung von Risiken als natürliche Häufigkeiten kann durch den *denominator neglect*, die Vernachlässigung des Nenners, entstehen: In der Wahrnehmung liegt der Fokus auf der Zahl von beobachteten Ereignissen (Zähler), sei sie noch so klein, nicht auf der Bezugsgröße (Nenner) (14). Beim Vergleich von Risiken – etwa mit und ohne eine Therapie – wird die Wahrnehmung weiter verzerrt, wenn die Bezugsgrößen unterschiedlich gewählt sind (z.B. 80 von 800 vs. 20 von 100). Die Nachteile dieser Darstellung werden schon lange diskutiert. Bei der Verwendung unterschiedlicher Bezugsgrößen gilt es als schwierig, verschiedene Angaben zu vergleichen und die Höhe von Risiken korrekt einzuschätzen (15), wobei es zu einer Überschätzung der Risiken kommt (16).

Fragestellungen

1. Welche Effekte hat die verbale Darstellung von Risiken, Nutzen und Schaden im Vergleich zur numerischen Darstellung?
2. Welchen Effekt hat die Darstellung von Nutzen und Schaden als absolute Darstellung (ARR) im Vergleich zur relativen Darstellung (RRR)?
3. Welche Effekte hat die Darstellung von Zahlen in natürlichen Häufigkeiten im Vergleich zur Darstellung in Prozent?
4. Welche Effekte hat die Darstellung von NNT (NNS, NNH) im Vergleich zur Darstellung als ARR (Prozent oder natürliche Häufigkeiten)?
5. Welche Effekte haben die Darstellungen mit konstanter Bezugsgröße (z.B. x von 1000) im Vergleich zu Darstellungen mit unterschiedlichen Bezugsgrößen (z.B. 2 von 100; 5 von 1000)?

Empfehlungen

1. Verbale Darstellungen

	<p>Empfehlung</p> <p>„Die alleinige verbale Darstellung von Risiken, Nutzen und Schaden soll nicht eingesetzt werden.“</p> <p>Zustimmung: 14, Ablehnung: 0, Enthaltung: 1</p> <p>Qualität der Evidenz: mittlere Qualität</p>
<p>Kommentar der LEG zur Empfehlung:</p> <p>Die Empfehlung bezieht sich auf den Vergleich von alleinigen verbalen Darstellungen mit numerischen Darstellungen.</p> <p>Die Studienlage zeigt, dass die verbale Darstellung von Risiken, Nutzen und Schaden der numerischen Darstellung unterlegen ist. Die verbale Darstellung führt zu Fehleinschätzungen von Risiken. Die numerischen Darstellungen ermöglichten in fünf von sechs Studien eine realistischere Risikoabschätzung / besseres Verstehen und in drei von vier Studien besseres Wissen. Beim Endpunkt Verständlichkeit gibt es keinen Unterschied. Alle sechs Studien zeigen eine höhere Intention, eine Maßnahme durchzuführen.</p> <p>Die Wirksamkeit auf die affektiven Endpunkte (11 Studien) ist uneinheitlich.</p>	

Zusammenfassung der Ergebnisse

Charakteristika der eingeschlossenen Studien

Zu diesem Vergleich wurden 15 Studien mit insgesamt 3531 Teilnehmerinnen und Teilnehmern eingeschlossen. Die Teilnehmerzahlen lagen zwischen 116 und 480, die Altersspanne reichte von 16 bis 82 Jahre. Die Studien wurden in den USA (17-21), Großbritannien (18, 22-28), Kanada (29), Australien (30) und Singapur (31) durchgeführt. Eingeschlossen wurden schwangere Frauen und Mütter (27, 28), Studentinnen und Studenten (21, 24), Patientinnen und Patienten (20, 25, 29-31), Bürgerinnen und Bürger (17, 18, 22, 23, 26) und Betreuerinnen und Betreuer (19).

Die Interventionen bestanden aus Szenarien zu Nebenwirkungen von Antibiotika (23, 24), Schmerzmitteln (18, 22, 31), Statinen (25), Tamoxifen (26) und Krebstherapien (17, 18, 20, 30), zu medizinischen Testergebnissen (27), zu den Wahrscheinlichkeiten bestimmter Ereignisse bei Säuglingen (28), zum Schlaganfallrisiko (29) und zu einer fiktiven Arzt-Angehörigen-Besprechung (19).

Ergebnisse zu den relevanten Endpunkten

Zu den Endpunkten Verstehen, Risikowahrnehmung und Wissen wurde positive Effekte für die numerische Darstellung gezeigt (18, 22-27, 29). Für den Endpunkt Verständlichkeit konnte kein Unterschied gezeigt werden (30). Zu den Endpunkten Akzeptanz, Attraktivität und Glaubwürdigkeit sind die Ergebnisse nicht eindeutig, aber es zeigt sich eine positive Tendenz für numerische Darstellungen (17, 18, 20-25, 28, 30, 31). Die Intention, eine Maßnahme durchzuführen, ist bei numerischen Darstellungen höher (18, 22-26).

Verstehen

Für den Endpunkt Verstehen konnte in einer Studie ein positiver Effekt für die numerische Darstellung gezeigt werden. Der Effekt wurde anhand einer Multiple-Choice Frage erhoben. Im Follow-up war dieser Effekt nicht mehr vorhanden (27).

Risikowahrnehmung

Die Risikowahrnehmung wurde mit einer Frage (22-24) oder zwei Fragen erhoben (Angabe eines Wertes von 0-100%) (19) bzw. es sollte eine Rangfolge zum Schlaganfallrisiko erstellt werden (29). Bei der Vermittlung von Risiken führten verbale Darstellungen gegenüber numerischen Darstellungen (1 von 100 oder %) in vier Studien zu einer schlechteren Risikowahrnehmung (22-24, 29), in einer war kein Unterschied erkennbar (19).

Wissen (recall)

Das Wissen wurde mit Fragen, die die Eingabe einer Schätzung in Prozent erforderten (25, 26), anhand von 24 Aussagen, die mit richtig, falsch und unsicher gekreuzt werden konnten (29) bzw. anhand einer Likert-Skala erhoben (18). In drei von 4 Studien führten verbale gegenüber numerischen Darstellungen zu einem geringeren Wissen (18, 25, 26); eine Studie zeigte keinen Unterschied (29).

Verständlichkeit

Die Erhebung der Verständlichkeit erfolgte anhand einer Likert-Skala. Es wurde kein Unterschied gezeigt (30).

Akzeptanz

Die Erhebung der Akzeptanz erfolgte anhand von Likert-Skalen in Rahmen von qualitativen Studien. In allen vier Studien zeigte sich eine Präferenz für die numerische Darstellung (20, 21, 28, 31).

Attraktivität

Attraktivität (hier Zufriedenheit) wurde anhand von Likert-Skalen erhoben. Drei Studien zeigten einen positiven Effekt für die numerische Darstellung (22-24), drei zeigten keine Unterschiede (18, 25, 30).

Glaubwürdigkeit

In einem Survey wurde ein Unterschied zugunsten der numerischen Darstellungen gezeigt (17).

Intention zur Einnahme der Medikation

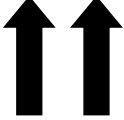
Die Intention, eine Maßnahme durchzuführen, ist in allen sechs Studien bei numerischen Darstellungen aufgrund einer realistischeren Risikoabschätzung höher (18, 22-26). Die Intention wurde jeweils mit einer Likert-Skala erhoben.

Begründung für die Empfehlung

Insgesamt zeigen sich positive Effekte für numerische Darstellungen im Vergleich zu verbalen Deskriptoren. Numerische Darstellungen tragen dazu bei, dass die Intention, eine Maßnahme durchzuführen, auf einer realistischeren Einschätzung des Risikos beruht. Daher hat sich die Leitlinien Entwicklungsgruppe (LEG) für eine starke Empfehlung gegen die alleinige Verwendung verbaler Darstellung ausgesprochen.

Weitere Informationen zu den Ergebnissen und Erhebungsmethoden können den Evidenztabellen und den Zusammenfassungen der Studien (*study fact sheets*) entnommen werden.

2. Absolute Risikomaße

	<p>Empfehlung</p> <p>„Nutzen und Schaden sollen durch absolute Risikomaße dargestellt werden.“</p> <p>Zustimmung: 15, Ablehnung: 0, Enthaltung: 0</p> <p>Qualität der Evidenz: mittlere Qualität</p>
<p>Kommentar der LEG zur Empfehlung:</p> <p>Die Empfehlung bezieht sich auf den Vergleich der Darstellung als ARR mit der alleinigen Darstellung der RRR in Gesundheitsinformationen.</p> <p>Für den kognitiven Endpunkt Verstehen / Risikowahrnehmung zeigten zwei von drei Studien, dass die ARR bei fehlenden Angaben des Basisrisikos präzisere Schätzungen ermöglicht und auch meistens der RRR überlegen ist, wenn die Basisrisiken genannt werden. RRR führt dagegen zu Überschätzungen des Nutzens einer Therapie. Zu den kognitiven Endpunkten Wissen und Verständlichkeit zeigten zwei Studien keine Effekte.</p> <p>Die Effekte auf affektive Endpunkte (zwei Studien) sind uneinheitlich.</p>	

Zusammenfassung der Ergebnisse

Charakteristika der eingeschlossenen Studien

Zu diesem Vergleich wurden fünf Studien mit insgesamt 4314 Teilnehmern und Teilnehmerinnen eingeschlossen. Die Teilnehmerzahlen lagen zwischen 209 und 2978, die Altersspannen reichten je nach Zielgruppe von 18 bis 74. Die Studien wurden in Großbritannien (32), USA (33-36), Kanada (33), Deutschland (33) und Norwegen (33) durchgeführt. Eingeschlossen waren Personen aus der allgemeinen Öffentlichkeit (30, 34), Frauen (34), Patienten (35) und Beschäftigte in Reservaten und Angehörige der First Nation (36). Die Interventionen bestanden aus Szenarien zur Influenzaimpfung (32), zu fiktiven Erkrankungen (35, 36), zur Einnahme von Statinen bei erhöhtem Cholesterinspiegel (33) sowie von Risikoinformationen zum Mammographiescreening (34).

Ergebnisse zu den relevanten Endpunkten

Zu den Endpunkten Verstehen / Risikowahrnehmung konnte ein Effekt für die ARR gezeigt werden (32, 34). Zu den Endpunkten Wissen und Verständlichkeit / Lesbarkeit konnte kein Unterschied gezeigt werden (33, 36). Zu dem Endpunkt Akzeptanz / Attraktivität sind die Ergebnisse uneinheitlich (32, 33).

Verstehen / Risiko-wahrnehmung

In einer Studie wurde Verstehen / Risikowahrnehmung anhand von zwei offenen Fragen erhoben (32). Wenn das Basisrisiko nicht angegeben war, ermöglichte die ARR in zwei von vier Vergleichen eine präzisere Abschätzung des Nutzens, die RRR führte zu starker Überschätzung (32). Mit Angabe des Basisrisikos führte die Darstellung als ARR zu signifikant präziseren Abschätzungen des Nutzens (34). Auch in dieser Studie wurde die Risikowahrnehmung anhand von zwei Fragen (Werte mussten angegeben werden) erhoben. In einer dritten Arbeit wurde kein Unterschied zwischen den Formaten beobachtet, bei insgesamt nur 13% korrekten Antworten (35). Die Erhebung erfolgte mit einer Frage (Werte mussten angegeben werden).

Wissen

Beim Endpunkt Wissen konnte kein Unterschied gezeigt werden. Wissen wurde anhand von drei Fragen erhoben (Werte mussten angegeben werden) (36).

Verständlichkeit / Lesbarkeit

Der Endpunkt Verständlichkeit / Lesbarkeit war sekundärer Endpunkt in der Studie. Es konnte kein Unterschied zwischen den beiden Formaten gezeigt werden. Die Erhebung erfolgte durch Selbsteinschätzung der Probandinnen und Probanden (33).

Akzeptanz / Attraktivität


Hinsichtlich der affektiven Endpunkte Akzeptanz / Attraktivität („satisfaction“) wurde die Darstellung als ARR bei Angabe des Basisrisikos in einer Studie bevorzugt. Der Endpunkt wurde mit einer Likert-Skala erhoben (32). Ohne die Angabe des Basisrisikos gab es keine Unterschiede zwischen den Darstellungen (32). In einer anderen Studie, in der der Endpunkt ebenfalls anhand einer Likert-Skala erhoben wurde, konnte kein Unterschied gezeigt werden (33).

Begründung für die Empfehlung

Zu den entscheidenden Endpunkten wurden positive Effekte für absolute Risikodarstellungen gezeigt. Insbesondere ermöglichen diese eine realistischere Abschätzung der Risiken. Die Nutzung relativer Risiken führt eher zu einer Überschätzung. Daher hat sich die LEG für eine starke Empfehlung für die Verwendung absoluter Risikomaße ausgesprochen. Diskutiert wurde, dass relative Risikomaße zusätzlich verwendet werden können.

Weitere Informationen zu den Ergebnissen und Erhebungsmethoden können den Evidenztabelle und den Zusammenfassungen der Studien (*study fact sheets*) entnommen werden.

3. Natürliche Häufigkeiten

	<p>Empfehlung</p> <p>„Die Darstellung in Prozent kann statt der Darstellung in natürlichen Häufigkeiten bei Wahrscheinlichkeiten >1% eingesetzt werden.“</p> <p>Zustimmung: 14, Ablehnung: 0, Enthaltung: 0</p> <p>Qualität der Evidenz: hohe Qualität</p>
<p>Kommentar der LEG zur Empfehlung:</p> <p>Die Empfehlung bezieht sich auf den Vergleich der Darstellung in Prozent mit der Darstellung in natürlichen Häufigkeiten in Gesundheitsinformationen.</p> <p>Zu dem Endpunkt Verstehen / Risikowahrnehmung wurde in einer Studie gezeigt, dass bei Wahrscheinlichkeiten >1% die Darstellung in Prozent der Darstellung in natürlichen Häufigkeiten überlegen ist. Für Wahrscheinlichkeiten <1% wurde kein Unterschied gezeigt. Insbesondere für Wahrscheinlichkeiten <0,1% kann keine Aussage getroffen werden.</p> <p>Für alle anderen Endpunkte (Wissen zwei Studien, Verständlichkeit / Lesbarkeit 2 Studien, Akzeptanz / Attraktivität eine Studie) konnte kein Unterschied gezeigt werden.</p>	

Zusammenfassung der Ergebnisse

Charakteristika der eingeschlossenen Studien

Zu diesem Vergleich wurden drei Studien mit insgesamt 3365 Teilnehmerinnen und Teilnehmern eingeschlossen. Die Teilnehmerzahlen lagen zwischen 136 und 2944, das mittlere Alter betrug je nach Zielgruppe 39 bis 61 Jahre. Die Studien wurden in den USA (9, 18, 37) und England (18) durchgeführt. Eingeschlossen wurden Personen aus der allgemeinen Öffentlichkeit (9), Veteranen (37) und Besucherinnen und Besucher einer Webseite (18).

Die Interventionen bestanden aus Szenarien zu Cholesterinsenkern und Sodbrennen im Format von Faktenboxen (9), zu kardiovaskulärem Risiko (37), Chemotherapie (18) und Schmerztherapie (18).

Ergebnisse zu den relevanten Endpunkten

Zum Endpunkt Verstehen / Risikowahrnehmung wurde in einer Studie von hoher Qualität ein Effekt für die Darstellung in Prozent gezeigt (9). Für die Endpunkte Wissen, Verständlichkeit / Lesbarkeit, Akzeptanz / Attraktivität konnte kein Unterschied gezeigt werden (9, 18, 37).

Verstehen / Risikowahrnehmung

Für Wahrscheinlichkeiten >1% wurde ein Effekt für die Darstellung in Prozent gezeigt (9). Bei der Kommunikation geringer Wahrscheinlichkeiten (<1%) gab es keine Unterschiede (9). Die Erhebung erfolgte anhand von 18 Multiple-Choice-Fragen.

Wissen

Wissen wurde mit einer bzw drei Fragen erhoben (Multiple Choice bzw. Werte mussten angegeben werden) (18, 37). Zudem erfolgte anhand zusätzlicher Fragen eine Einschätzung auf einer Likert-Skala. Es konnten keine Effekte gezeigt werden.

Verständlichkeit / Lesbarkeit

Verständlichkeit / Lesbarkeit wurden mit einer Frage (Multiple Choice) (8) bzw. drei Fragen (Likert-Skala) erhoben (37). Es wurde kein Unterschied zwischen den Darstellungen gezeigt.

Akzeptanz / Attraktivität


Zur Erhebung von Akzeptanz / Attraktivität wurde eine Frage (Likert-Skala) eingesetzt (18). Es konnte kein Unterschied gezeigt werden.

Begründung für die Empfehlung

Zwischen der Darstellung in Prozent und der Verwendung natürlicher Häufigkeiten konnten aus Sicht der LEG nur geringe Unterschiede gezeigt werden, so dass bei Darstellungsformen als gleichwertig betrachtet werden. Daher wurde eine offene Empfehlung ausgesprochen.

Weitere Informationen zu den Ergebnissen und Erhebungsmethoden können den Evidenztabellen und den Zusammenfassungen der Studien (study fact sheets) entnommen werden.

4. Number Needed to Treat (NNS, NNH)

	<p>Empfehlung</p> <p>„Die Darstellung in Number Needed to Treat (NNT), Number Needed to Screen (NNS), Number Needed to Harm (NNH) sollte nicht eingesetzt werden.“</p> <p>Zustimmung: 11, Ablehnung: 0, Enthaltung: 2</p> <p>Qualität der Evidenz: mittlere Qualität</p>
<p>Kommentar der LEG zur Empfehlung:</p> <p>Die Empfehlung bezieht sich auf den Vergleich der Darstellung als NNT (NNS, NNH) mit der Darstellung als absolute Risikoreduktion.</p> <p>Die Studienlage zu dem Endpunkt Verstehen / Risikowahrnehmung zeigt, dass die Darstellung in NNT (NNS, NNH) der Darstellung der absoluten Risikoreduktion unterlegen ist. Insbesondere wenn keine Basisrisiken angegeben sind, führte die NNT in zwei Studien zur Überschätzung des Nutzens bzw. Risikos.</p> <p>Für alle anderen Endpunkte (Verständlichkeit / Lesbarkeit eine Studie, Akzeptanz / Attraktivität zwei Studien) konnte kein Unterschied gezeigt werden.</p>	

Zusammenfassung der Ergebnisse

Charakteristika der eingeschlossenen Studien

Zu diesem Vergleich wurden drei Studien mit insgesamt 3653 Teilnehmerinnen und Teilnehmern eingeschlossen. Die Teilnehmerzahlen lagen zwischen 268 und 2978, die Altersspanne reichte von 18 bis über 60 Jahre. Die Studien wurden in den USA (33, 35), England (38), Kanada (38), Deutschland (38) und Norwegen (38) durchgeführt. Eingeschlossen wurden Passantinnen in einem Stadtzentrum (38), Patientinnen und Patienten (35) und Personen aus der allgemeinen Öffentlichkeit (33).

Die Interventionen bestanden aus Szenarien zur Pille (38), zu fiktiven Erkrankungen (35) und zur Einnahme von Statinen bei erhöhtem Cholesterinspiegel (33).

Ergebnisse zu den relevanten Endpunkten

Für den Endpunkt Verstehen / Risikowahrnehmung wurde ein Effekt für die Darstellung der ARR gezeigt (35, 38). Für die Endpunkte Verständlichkeit / Lesbarkeit und Akzeptanz / Attraktivität konnte kein Unterschied gezeigt werden (33, 38).

Verstehen / Risikowahrnehmung

Die Erhebung erfolgte mit zwei (offenen) Fragen (38) bzw. mit einer Frage (Wert musste angegeben werden) (35). Es wurde ein Effekt für die Darstellung der ARR gezeigt, insbesondere wenn keine Basisrisiken angegeben waren. Die NNT führte zu Überschätzungen des Nutzens bzw. Risikos (35, 38).

Verständlichkeit / Lesbarkeit

Verständlichkeit / Lesbarkeit wurde mit einer Frage (Likert-Skala) erhoben (33). Es konnte kein Unterschied zwischen den Darstellungsformen gezeigt werden.

Akzeptanz / Attraktivität


Akzeptanz / Attraktivität wurde mit jeweils einer Frage (Likert-Skala) erhoben (33, 38). Es wurde kein Unterschied gezeigt. Aus den Arbeiten lässt sich ableiten, dass die Angabe des Basisrisikos sehr bedeutsam ist.

Begründung für die Empfehlung

Hinsichtlich des entscheidenden Endpunktes Verstehen / Risikowahrnehmung zeigt sich die Darstellung als absolute Risikoreduktion überlegen. Insbesondere ohne Angabe eines Basisrisikos führt die NNT zu einer Überschätzung des Nutzens bzw. Risikos. Daher stimmte die LEG einer abgeschwächten Empfehlung gegen die Verwendung von NNT / NNH / NNS zu.

Weitere Informationen zu den Ergebnissen und Erhebungsmethoden können den Evidenztabelle und den Zusammenfassungen der Studien (*study fact sheets*) entnommen werden.

5. Bezugsgrößen

	<p>Empfehlung</p> <p>„In Gesundheitsinformationen sollten gleiche Bezugsgrößen eingesetzt werden.“</p> <p>Zustimmung: 9, Ablehnung: 0, Enthaltung: 0</p> <p>Qualität der Evidenz: hohe Qualität</p>
<p>Kommentar der LEG zur Empfehlung:</p> <p>Die Empfehlung bezieht sich auf den Vergleich der Darstellung mit gleicher Bezugsgröße mit der Darstellung mit unterschiedlichen Bezugsgrößen.</p> <p>In einer Studie zeigte die Darstellung mit gleicher Bezugsgröße in einer Faktenbox einen positiven Effekt auf den Endpunkt Verstehen Risikowahrnehmung. Für den Endpunkt Verständlichkeit / Lesbarkeit wurde in dieser Studie kein Unterschied gezeigt. Für alle anderen Endpunkte liegen keine Ergebnisse vor.</p> <p>Nutzen und Schaden sollten soweit möglich mit der gleichen Bezugsgröße dargestellt werden. Bei Abweichungen von der Empfehlung sollte dieses gut begründet und transparent dargelegt werden, dass ein Wechsel stattfindet.</p>	

Zusammenfassung der Ergebnisse

Charakteristika der eingeschlossenen Studien

Zu diesem Vergleich wurde eine Studie mit insgesamt 1181 Teilnehmerinnen und Teilnehmern eingeschlossen (9). Das mittlere Alter betrug 47 Jahre. Die Studie wurde in den USA durchgeführt. Eingeschlossen waren Personen aus der allgemeinen Öffentlichkeit. Die Interventionen bestanden aus Szenarien zu Cholesterinsenkern und Sodbrennen, die jeweils in einer Faktenbox präsentiert wurden (9).

Ergebnisse zu den relevanten Endpunkten

Zu dem Endpunkt Verstehen wurde für die Darstellung mit durchgehend gleicher Bezugsgröße ein positiver Effekt gezeigt (9). Zu dem Endpunkt Verständlichkeit / Lesbarkeit konnte kein Unterschied gezeigt werden (9).

Verstehen

Bezüglich des Endpunkts Verstehen wurde für die Darstellung mit durchgehend gleicher Bezugsgröße (z.B. x von 1000) im Vergleich zur Darstellung mit wechselnder Bezugsgröße (z.B. x von 100; x von 1000; x von 10000) ein positiver Effekt gezeigt (9). Die Erhebung des Endpunkts erfolgte mit 18 Fragen. Die Darstellung in Prozent (siehe *Darstellung von Häufigkeiten*, Frage 3) wäre ebenfalls möglich.

Verständlichkeit / Lesbarkeit

Für den Endpunkt Verständlichkeit (erhoben mit einer Frage) wurde kein signifikanter Unterschied gezeigt (9).

Für andere Endpunkte liegen keine Ergebnisse vor.

Begründung für die Empfehlung

Da nur aus einer Studie Evidenz vorliegt, wurde zunächst diskutiert, keine Empfehlung auszusprechen. Die Studie hat den Effekt gleicher und ungleicher Bezugsgrößen nur im Kontext einer Faktenbox untersucht, somit erscheint eine Generalisierbarkeit der Ergebnisse schwierig. Zudem wurde nur zu einem der zwei Endpunkte ein Effekt gezeigt. Allerdings ist Verstehen / Risikowahrnehmung ein entscheidender Endpunkt und hat somit mehr Gewicht als Verständlichkeit / Lesbarkeit. Die vorliegende Evidenz ist zudem von hoher Qualität. Bedenken gab es, dass keine Empfehlung als negative Empfehlung missverstanden werden könnte und gleiche Bezugsgrößen nicht genutzt würden.

Ein weiterer Vorschlag war, eine offene Empfehlung auszusprechen, da der Unterschied als gering, das Verständnis aber mit beiden Formaten als hoch bewertet wurde. Gegen die offene Empfehlung sprechen die Forschungsergebnisse zum *denominator neglect* (siehe Einleitung) und die Gefahr einer missbräuchlichen Verwendung unterschiedlicher Bezugsgrößen, um eine Beeinflussung zu erreichen.

Da sich die LEG auch bei schwacher Evidenz für eine Empfehlung entscheiden kann, wurde aus den oben genannten Gründen eine abgeschwächte Empfehlung für gleiche Bezugsgrößen formuliert. Diese Empfehlung bezieht sich trotz der Studienlage auf Gesundheitsinformationen allgemein und nicht nur auf Faktenboxen. Ausschlaggebend war für diese Entscheidung ist die Gefahr der Beeinflussung (*persuasiveness*) durch unterschiedliche Bezugsgrößen. Der LEG ist bewusst, dass unter bestimmten Voraussetzungen ein Abweichen von der Empfehlung notwendig

sein kann. Die Verwendung unterschiedlicher Bezugsgrößen sollte dann gut begründet und transparent dargelegt werden.

Weitere Informationen zu den Ergebnissen und Erhebungsmethoden können den Evidenztabelle und den Zusammenfassungen der Studien (*study fact sheets*) entnommen werden.

Evidenztabelle

Tabelle 4: Evidenztabelle „Verbale versus numerische Darstellung“

Bewertung der Qualität der Evidenz						Zusammenfassung der Ergebnisse				
						Anzahl der Teilnehmer		Effektschätzer		
Anzahl der Studien je Endpunkt	Design	Risk of bias	Konsistenz	Direktheit	Präzision	Intervention	Kontrolle	Effekte	Qualität	Bedeutsamkeit des Ergebnisparameters
Vergleich verbale versus numerische Darstellung										
Verstehen [n=1] Marteau (27)	RCT	Schwerwiegendes Risiko für Bias (-1)	Keine Inkonsistenz	Keine Indirektheit	Keine Impräzision	(verbal) N= 112	(numerisch) N= 97	Effekt zugunsten der numerischen Darstellung direkt nach der Kommunikation, nach vier Monaten gibt es keinen Unterschied (27).	Mittlere Qualität	Entscheidend
Risikowahrnehmung [n=5] Berry (Studie 2) (24) Berry (22) Berry (23) Lee Char (19) Man-Son-Hing (29)	RCT	Sehr schwerwiegendes Risiko für Bias (-2)	Keine Inkonsistenz	Keine Indirektheit	Keine Impräzision	N= 335 + N= 480 (Verteilung auf die Gruppen unklar.)	N= 332	In 4 Arbeiten (5 Studien) Effekte zugunsten der numerischen Darstellung (22-24, 29). Massive Überschätzungen bei den verbalen Darstellungen, bei den numerischen Darstellungen weniger starke Überschätzungen. In einer Studie kein Effekt (19).	Niedrige Qualität	Entscheidend

Wissen (recall) [n=4] Knapp (18)(16) Knapp (25)(26) Knapp (26)(27) Man-Son-Hing (29)(28)	RCT	Schwerwiegendes Risiko für Bias (-1)	Keine Inkonsistenz	Keine Indirektheit	Keine Impräzision	N= 324	N= 400	In 3 Studien Effekte zugunsten der numerischen Darstellung (18, 25, 26), in einer Studie kein Effekt (29).	Mittlere Qualität	Entscheidend
Verständlichkeit [n=1] Hagerty (30)	Qualitative Arbeiten	keine GRADE Bewertung				N= 126		Beide Darstellungen werden als vergleichbar verständlich eingeschätzt (30).	Evidenz aus qualitativen Arbeiten	Evidenz aus Survey
Akzeptanz [n=4] Cheung (31) Mazur (20) Wallsten (21) Shaw 1990 (28)	Qualitative Arbeiten	keine GRADE Bewertung				N= 991		In allen vier Studien zeigt sich eine Präferenz für die numerische Darstellung (20, 21, 28, 31).	Evidenz aus qualitativen Arbeiten	Geringe Bedeutung
Attraktivität [n=6] Berry (Studie 2) (24) Berry (22) Berry (23) Knapp (18) Knapp (Szenario 1+2) (25) Hagerty (30)	RCT	Sehr schwerwiegendes Risiko für Bias (-2)	Schwerwiegende Inkonsistenz (-1)	Keine Indirektheit	Schwerwiegende Impräzision (-1)	N= 260	N= 399 + N= 606 (Verteilung auf die Gruppen unklar.)	In drei Studien statistisch signifikant höhere Zufriedenheit mit numerischer Darstellung (22-24), in drei Studien kein Unterschied (18, 25, 30).	Sehr niedrige Qualität	Geringe Bedeutung

Glaubwürdigkeit [n=1] Gurmankin (17)	Survey	keine GRADE Bewertung				N= 115	Statistisch signifikanter aber fraglich relevanter Unterschied zugunsten der numerischen Darstellung (17).	Evidenz aus einem Survey	Geringe Bedeutung	
Intention zur Einnahme der Medikation [n=6] Berry (Studie 2) (24) Berry (22) Berry (23) Knapp (18) Knapp (25) Knapp (26)	RCT	Sehr schwerwiegendes Risiko für Bias (-2)	Keine Inkonsistenz	Keine Indirektheit	Schwerwiegende Impräzision (-1)	N= 375	N= 451	In sechs Arbeiten höhere Intention zur Einnahme der Medikation bei numerischen Darstellungen (18, 22-26).	Sehr niedrige Qualität	Nicht definiert

Tabelle 5: Evidenztabelle „Absolute versus relative Risikoreduktion“

Bewertung der Qualität der Evidenz						Zusammenfassung der Ergebnisse				
						Anzahl der Teilnehmer		Effektschätzer		
Anzahl der Studien je Endpunkt	Design	Risk of bias	Konsistenz	Direktheit	Präzision	Intervention	Kontrolle	Effekte	Qualität	Bedeutsamkeit des Ergebnisparameters
Vergleich absolute versus relative Risikoreduktion										
Verstehen / Risikowahrnehmung [n=3] Natter (32) Schwartz (34) Sheridan (35)	RCT	Schwerwiegendes Risiko für Bias (-1)	Keine Inkonsistenz	Keine Indirektheit	Keine Impräzision	(ARR) N=365	(RRR) N=347	Ohne Angabe des Baseline Risikos konnten der Nutzen in einer Studie mit der Darstellung ARR präziser eingeschätzt werden. RRR führte zu Überschätzungen des Nutzens. Mit Angabe der Baseline Risiken führte die Darstellung ARR zu einer signifikant genaueren Abschätzung des Nutzens (34). In der zweiten Studie gab es in 2 von 4 Gruppen einen Vorteil für die ARR Gruppen (32). In einer Studie kein Effekt (35).	Mittlere Qualität	Entscheidend

Wissen [n=1] Sprague (36)	RCT	Schwerwiegendes Risiko für Bias (-1)	Keine Inkonsistenz	Keine Indirektheit	Keine Impräzision	N=46	N=54	In einer Studie kein Effekt (36).	Mittlere Qualität	Entscheidend
Verständlichkeit Lesbarkeit [n=1] Carling (33)	RCT,	Kein Risiko für Bias	Keine Inkonsistenz	Keine Indirektheit	Keine Impräzision	N=505	N=508	In einer Studie kein Effekt (der Endpunkt wurde von den TN selbst eingeschätzt, sek. Endpunkt) (33).	Hohe Qualität	Wichtig, aber nicht entscheidend
Akzeptanz / Attraktivität [n=2] Natter (32) Carling (33)	RCT	Schwerwiegendes Risiko für Bias (-1)	Keine Inkonsistenz	Keine Indirektheit	Keine Impräzision	N=615	N=618	Mit Angabe des Baseline Risikos wurde das absolute Format in einer Studie signifikant bevorzugt. Ohne die Angabe des Baseline Risikos gab es keine Unterschiede zwischen den Darstellungen (32). Eine weitere Studie zeigte ebenfalls keinen Unterschied in der Zufriedenheit (33).	Mittlere Qualität	Geringe Bedeutung

Tabelle 6: Evidenztabelle „Natürliche Häufigkeiten versus Prozent“

Bewertung der Qualität der Evidenz						Zusammenfassung der Ergebnisse				
Anzahl der Studien je Endpunkt	Design	Risk of bias	Konsistenz	Direktheit	Präzision	Anzahl der Teilnehmer		Effektschätzer		
						Intervention	Kontrolle	Effekte	Qualität	Bedeutsamkeit des Ergebnisparameters
Vergleich Natürliche Häufigkeiten (NH) versus Prozentangaben										
Verstehen / Risikowahrnehmung [n=1] Woloshin (9)	RCT	Kein Risiko für Bias	Keine Inkonsistenz	Keine Indirektheit	Keine Impräzision	(NH) N=590	(Prozent) N=591	Positiver Effekt für Prozen-te, bei niedrigen Wahrscheinlichkeiten (<1%) kein Unterschied (9).	Hohe Qualität	Entscheidend
Wissen [n=2] Ruiz (37) Knapp (2 Experimente) (18)	RCT	Schwerwiegendes Risiko für Bias (-1)	Keine Inkonsistenz	Keine Indirektheit	Keine Impräzision	N=135	N=134	In zwei Studien kein Effekt (18, 37).	Mittlere Qualität	Entscheidend
Verständlichkeit / Lesbarkeit [n=2] Woloshin (9) Ruiz (37)	RCT,	Schwerwiegendes Risiko für Bias (-1)	Keine relevante Inkonsistenz	Keine Indirektheit	Keine Impräzision	N=630	N=631	In zwei Studien kein Effekt (9, 37).	Mittlere Qualität	Wichtig, aber nicht entscheidend
Akzeptanz / Attraktivität [n=1] Knapp (18)	RCT	Schwerwiegendes Risiko für Bias (-1)	Keine Inkonsistenz	Keine Indirektheit	Keine Impräzision	N=244	N=245	In einer Studie kein Effekt (18).	Mittlere Qualität	Geringe Bedeutung

Tabelle 7: Evidenztabelle „NNT / NNH versus ARR“

Bewertung der Qualität der Evidenz						Zusammenfassung der Ergebnisse				
Anzahl der Studien je Endpunkt	Design	Risk of bias	Konsistenz	Direktheit	Präzision	Anzahl der Teilnehmer		Effektschätzer		
						Intervention	Kontrolle	Effekte	Qualität	Bedeutsamkeit des Ergebnisparameters
Vergleich NNT / NNH versus ARR										
Verstehen / Risikowahrnehmung [n=2] Berry (38) Sheridan (35)	RCT	Schwerwiegendes Risiko für Bias (-1)	Keine Inkonsistenz	Keine Indirektheit	Keine Impräzision	(NNT) N=192	(ARR) N=200	Effekt für die ARR (hier in %) ohne Angabe des Baseline Risikos. NNH führte zur Überschätzung. Mit Angabe der Baseline Risiken kein Unterschied (38). In der zweiten Studie gab es ebenfalls einen Effekt für ARR (35).	Mittlere Qualität	Entscheidend
Verständlichkeit / Lesbarkeit [n=1] Carling (33)	RCT	Kein Hinweis auf Bias	Keine Inkonsistenz	Keine Indirektheit	Keine Impräzision	N=484	N=505	NNT und ARR (%): kein Effekt (33).	Hohe Qualität	Wichtig, aber nicht entscheidend
Akzeptanz / Attraktivität (satisfaction, preference) [n=2] Berry (38) Carling (33)	RCT	Schwerwiegendes Risiko für Bias (-1)	Keine Inkonsistenz	Keine Indirektheit	Keine Impräzision	N=576	N=597	Kein Effekt, jedoch war bei beiden Formaten die Zufriedenheit höher, wenn die Baseline Risiken auch dargestellt wurden (33, 38).	Mittlere Qualität	Geringe Bedeutung

Tabelle 8: Evidenztabelle „Gleiche versus ungleiche Bezugsgrößen“

Bewertung der Qualität der Evidenz						Zusammenfassung der Ergebnisse				
						Anzahl der Teilnehmer		Effektschätzer		
Anzahl der Studien je Endpunkt	Design	Risk of bias	Konsistenz	Direktheit	Präzision	Intervention	Kontrolle	Effekte	Qualität	Bedeutsamkeit des Ergebnisparameters
Vergleich Gleiche versus ungleiche Bezugsgrößen (x in 100; x in 1000; x in 10000)										
Verstehen [n=1] Woloshin (9)	RCT	Kein Risiko für Bias	Keine Inkonsistenz	Keine Indirektheit	Keine Impräzision	N=590	N=591	Effekt für die die Darstellung mit gleicher Bezugsgröße (x in 1000) im Vergleich zur Darstellung mit wechselnder Bezugsgröße in einer „drug facts box“ (9).	Hohe Qualität	Entscheidend
Verständlichkeit / Lesbarkeit [n=1] Woloshin (9)	RCT	Kein Risiko für Bias	Keine Inkonsistenz	Keine Indirektheit	Keine Impräzision	N=590	N=591	In einer Studie kein Effekt (9).	Hohe Qualität	Wichtig, aber nicht entscheidend

Referenzen

1. Akl EA, Oxman AD, Herrin J, Vist GE, Terrenato I, Sperati F, et al. Using alternative statistical formats for presenting risks and risk reductions. The Cochrane database of systematic reviews. 2011;3:CD006776. Epub 2011/03/18.
2. Berry DC, Knapp P, Raynor DK. Provision of information about drug side-effects to patients. Lancet. 2002; <http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/036/CN-00558036/frame.html> (Zugriff am 11.10.2016).
3. European Commission D-GI. A guideline on the readability of the label and package leaflet of medical products for human use. 1998. http://www.paint-consult.com/fileadmin/editorial/downloads/guidelines_behoerden/anzneimittelgesetze/PAINT-Consult_A_Guideline_on_the_readability_1998.pdf (Zugriff am 11.10.2016).
4. Bundesministerium für Arzneimittel und Medizinprodukte (BfArM). Wie sollen die Häufigkeiten für Nebenwirkungen in der Produktinformation angegeben werden? 2015; <http://www.bfarm.de/SharedDocs/FAQs/DE/Arzneimittel/pal/ja-ampal-faq.html> (Zugriff am 11.10.2016) .
5. Ziegler A, Hadlak A, Mehlbeer S, König IR. Comprehension of the description of side effects in drug information leaflets: a survey of doctors, pharmacists and lawyers. Deutsches Arzteblatt international. 2013;110(40):669-73. Epub 2013/10/30.
6. Bunge M, Mühlhauser I, Steckelberg A. What constitutes evidence-based patient information? Overview of discussed criteria. Patient Education and Counseling. 2010;78(3):316-28. Epub 2009/12/17.
7. Covey J. A Meta-analysis of the effects of presenting treatment benefits in different formats. [References]. Medical Decision Making. 2007;27(5):638-654.
8. Trevena LJ, Davey HM, Barratt A, Butow P, Caldwell P. A systematic review on communicating with patients about evidence. Journal of evaluation in clinical practice. 2006;12(1):13-23. Epub 2006/01/21.
9. Woloshin S, Schwartz LM. Communicating data about the benefits and harms of treatment: a randomized trial. Annals of internal medicine. 2011;155(2):87-96. Epub 2011/07/20.

10. Bramwell R, West H, Salmon P. Health professionals' and service users' interpretation of screening test results: experimental study. *British Medical Journal (Clinical research ed)*. 2006;333(7562):284. Epub 2006/07/15.
11. Galesic M, Gigerenzer G, Straubinger N. Natural frequencies help older adults and people with low numeracy to evaluate medical screening tests. *Medical decision making: an international journal of the Society for Medical Decision Making*. 2009;29(3):368-71. Epub 2009/01/09.
12. Garcia-Retamero R, Hoffrage U. Visual representation of statistical information improves diagnostic inferences in doctors and their patients. *Social science & medicine*. 2013;83:27-33. Epub 2013/03/08.
13. Siegrist M, Keller C. Natural frequencies and Bayesian reasoning: The impact of formal education and problem context. [References]. *Journal of Risk Research*. 2011;14(9):1039-55.
14. Reyna VF, Brainerd CJ. Numeracy, ratio bias, and denominator neglect in judgments of risk and probability. *Learning and Individual Differences*. 2008;18(1):89-107.
15. Zikmund-Fisher BJ. Time to retire the 1-in-X risk format. *Medical Decision Making: an international journal of the Society for Medical Decision Making*. 2011;31(5):703-4.
16. Pighin S, Savadori L, Barilli E, Cremonesi L, Ferrari M, Bonnefon J-F. The 1-in-X effect on the subjective assessment of medical probabilities. *Medical Decision Making: an international journal of the Society for Medical Decision Making*. 2011;31(5):721-9.
17. Gurmankin AD, Baron J, Armstrong K. The effect of numerical statements of risk on trust and comfort with hypothetical physician risk communication. *Medical decision making: an international journal of the Society for Medical Decision Making*. 2004;24(3):265-71. Epub 2004/05/25.
18. Knapp P, Gardner PH, Carrigan N, Raynor DK, Woolf E. Perceived risk of medicine side effects in users of a patient information website: a study of the use of verbal descriptors, percentages and natural frequencies. *British journal of health psychology*. 2009;14(3):579-94. Epub 2008/11/11.
19. Lee Char SJ, Evans LR, Malvar GL, White DB. A randomized trial of two methods to disclose prognosis to surrogate decision makers in intensive care units. *American journal of respiratory and critical care medicine*. 2010;182(7):905-9. Epub 2010/06/12.

20. Mazur DJ, Hickam DH, Mazur MD. How patients' preferences for risk information influence treatment choice in a case of high risk and high therapeutic uncertainty: asymptomatic localized prostate cancer. *Medical decision making: an international journal of the Society for Medical Decision Making*. 1999;19(4):394-8. Epub 1999/10/16.
21. Wallsten TS, Budescu DV, Zwick R, Kemp SM. Preferences and reasons for communicating probabilistic information in verbal or numerical terms. *Bulletin of the Psychonomic Society*. 1993;31(2):135-8.
22. Berry D, Raynor T, Knapp P, Bersellini E. Over the counter medicines and the need for immediate action: a further evaluation of European Commission recommended wordings for communicating risk. *Patient Education and Counseling*. 2004;53(2):129-34. Epub 2004/05/14.
23. Berry DC, Raynor DK, Knapp P. Communicating risk of medication side effects: an empirical evaluation of EU recommended terminology. [References]. *Psychology, Health & Medicine*. 2003;8(3):251-63.
24. Berry D, Raynor T, Knapp P, Bersellini E. Is 15 per cent very common? Informing people about the risks of medication side effects. *International Journal of Pharmacy Practice*. 2002;10(3):145-51.
25. Knapp P, Raynor DK, Berry DC. Comparison of two methods of presenting risk information to patients about the side effects of medicines. *Quality & safety in health care*. 2004;13(3):176-80. Epub 2004/06/04.
26. Knapp P, Raynor DK, Woolf E, Gardner PH, Carrigan N, McMillan B. Communicating the risk of side effects to patients: an evaluation of UK regulatory recommendations. *Drug safety: an international journal of medical toxicology and drug experience*. 2009;32(10):837-49. Epub 2009/09/03.
27. Marteau TM, Saidi G, Goodburn S, Lawton J, Michie S, Bobrow M. Numbers or words? A randomized controlled trial of presenting screen negative results to pregnant women. *Prenatal diagnosis*. 2000;20(9):714-8. Epub 2000/10/04.
28. Shaw NJ, Dear PR. How do parents of babies interpret qualitative expressions of probability? *Archives of disease in childhood*. 1990;65(5):520-3. Epub 1990/05/01.
29. Man-Son-Hing M, O'Connor AM, Drake E, Biggs J, Hum V, Laupacis A. The effect of qualitative vs. quantitative presentation of probability estimates on patient decision-making: a

randomized trial. *Health expectations: an international journal of public participation in health care and health policy*. 2002;5(3):246-55. Epub 2002/08/30.

30. Hagerty RG, Butow PN, Ellis PA, Lobb EA, Pendlebury S, Leighl N, et al. Cancer patient preferences for communication of prognosis in the metastatic setting. *Journal of Clinical Oncology*. 2004;22(9):1721-30.

31. Cheung YB, Wee HL, Thumboo J, Goh C, Pietrobon R, Toh HC, et al. Risk communication in clinical trials: a cognitive experiment and a survey. *BMC medical informatics and decision making*. 2010;10:55. Epub 2010/09/28.

32. Natter HM, Berry DC. Effects of presenting the baseline risk when communicating absolute and relative risk reductions. *Psychology, Health & Medicine*. 2005;10(4):326-34.

33. Carling CL, Kristoffersen DT, Montori VM, Herrin J, Schunemann HJ, Treweek S, et al. The effect of alternative summary statistics for communicating risk reduction on decisions about taking statins: a randomized trial. *PLoS medicine*. 2009;6(8):e1000134. Epub 2009/08/27.

34. Schwartz LM, Woloshin S, Black WC, Welch HG. The role of numeracy in understanding the benefit of screening mammography. *Annals of internal medicine*. 1997;127(11):966-72. Epub 1997/12/31.

35. Sheridan SL, Pignone MP, Lewis CL. A randomized comparison of patients' understanding of number needed to treat and other common risk reduction formats. *Journal of general internal medicine*. 2003;18(11):884-92. Epub 2003/12/23.

36. Sprague D, LaVallie DL, Wolf FM, Jacobsen C, Sayson K, Buchwald D. Influence of graphic format on comprehension of risk information among American Indians. [References]. *Medical Decision Making*. 2011;31(3):437-43.

37. Ruiz JG, Andrade AD, Garcia-Retamero R, Anam R, Rodriguez R, Sharit J. Communicating global cardiovascular risk: Are icon arrays better than numerical estimates in improving understanding, recall and perception of risk? *Patient Education and Counseling*. 2013;93(3):394-402.

38. Berry DC, Knapp P, Raynor T. Expressing medicine side effects: assessing the effectiveness of absolute risk, relative risk, and number needed to harm, and the provision of baseline risk information. *Patient Education and Counseling*. 2006;63(1-2):89-96. Epub 2005/10/26.