

设计一份有效的伤口护理问卷

关键词 分析、指标、问卷设计、回复率、验证

文献引用 Stephenson J. Designing an effective questionnaire in wound care .WCET[®] Journal 2022;42(2):24-29

DOI <https://doi.org/10.33235/wcet.42.2.24-29>

提交日期: 2022年4月29日, **接收日期:** 2022年5月6日

引言

通过问卷调查收集定量数据是伤口护理的常见做法。问卷调查是一种相对低成本且快速的数据收集方式,并且在收集数据时不一定要求研究人员在场。很多时候,这是收集所需数据的唯一可行方法。问卷调查(包括向临床医护人员、患者或两者发放的问卷)在伤口护理中的常见用途,包括:

- 评估临床培训计划为提高医护人员对某种疾病的认识方面的有效性。
- 评估特定敷料在特定临床环境中的使用程度。
- 评估新设备。
- 监测新治疗方案下的伤口愈合情况。
- 评估与患者相关的结局,如疼痛、生活质量或对所接受治疗的满意度。

虽然许多经过充分验证的问卷“现成”可用,但伤口护理研究人员可能会发现,这些问卷所体现的具体指标与他们提议的研究目标不匹配,因此可能需要设计定制的量表。基于问卷的研究包括仔细考虑研究样本的选择、最大限度提高回复率、确定要评估的指标、制定问卷项目内容和评分、拟定用于分析的项目、考虑结局指标和项目评分,以及进行试点问卷调查。

问卷将发给谁?

可推论性的概念——能够从样本数据(完成问卷调查的人)推断出通常更广泛的全及总体——是大多数定量研究的关键。这需要有代表性的受访者样本。创建一个在所有方面

都能准确反映其应该代表的群体的样本几乎不可能。需要临床知识来确定重要特征,例如工作水平、患者共病或伤口类型,这些特征在不同的研究中会有所不同。要确定样本是否在对研究最重要的特征上确实反映了全及总体,可能需要至少了解相关群体中单位类别的大致分布:例如,典型组织机构中的一个典型组织活性护理团队的构成可能是已知的,研究人员可能会试图在受邀完成调查问卷的人员中反映出这种构成。如果未能确保样本在某些重要方面与其声称代表的总体没有差异,可能会导致选择偏倚,这可能会削弱或否定研究结果。

一些特定特征适用于许多伤口护理研究中收集的数据。首先,通常必须同时收集临床医护人员和患者的数据。例如,对社区护理团队的病例数量进行研究,在该研究中,对护士及其患者都将进行调查;通常,护士和患者适用不同的问卷项目。这通常会导致数据聚集,因为一名医护人员会治疗多名患者。其次,正如在临床科学其他分支中所常见的,在伤口护理研究中并不总以人为分析单位。分析单位可能是伤口,例如压力性损伤,并且一名患者可能会为同一研究提供多个伤口。同样,这将再次导致数据聚集的问题;此例中是因为压力性损伤聚集在个体患者中。

最大限度提高回复率

通过问卷收集的数据特别容易受到应答偏倚的影响,这种偏差是由选择填写问卷的人与未填写问卷的人之间的特征差异所引起的。虽然有计算方法可用于填补缺失数据值,但 these 方法可能并非在所有情况下都可行,通常最好是最大限度地提高潜在回复者中实际回复者的比例,以及回复者中给出完整回复的比例。低回复率也会导致降低分析能力,即检测可能存在的任何影响的能力。

有一些明显的方法可以提高回复率和完成率:

John Stephenson

PHD、FRSS (GradStat)、CMath (MIMA)

生物医学统计学高级讲师

英国哈德斯菲尔德大学

电子邮箱: J.Stephenson@hud.ac.uk

- 使用电子格式代替纸质问卷，或两种同时投放（可在适当的时间间隔给未回复者发送礼貌的电子邮件提醒）。
- 避免问卷包含的项目过多。所有包含的项目都应出于特定目的：每个多余的项目都会增加受访者不能正确完成问卷的可能性。例如，不应要求受访者直接提供诸如BMI等数量信息，研究人员可根据受访者提供的其他信息计算这些信息。
- 避免措辞含糊的项目。回答时，应提供选项选择或视觉模拟量表，而不是要求自由文本回答。提供有条件的项目可能会引起混淆，应加以限制。
- 确保参与者的匿名性，如果这适合于所收集的信息。

一些研究要求在多个场合收集基于问卷的数据，例如，监测慢性伤口患者的生活质量或疼痛。这里涉及的一个常见问题是，在每个数据收集时间点，完成问卷的比例通常会降低。当失访者在某种程度上与回复问卷者存在系统性差异时，这可能引发进一步的偏倚，即失访偏倚。虽然在随访期间，患者搬走或死亡的情况几乎无法解决，但通过在问卷邮寄的频率、问卷的长度或复杂性方面不让受访者负担过重，仍然可以将失访损失降到最低。

验证/评估指标

设计适当的项目以有效地囊括所关注结局指标通常是有效问卷设计中最困难的部分。通常，最好使用经验证的可在类似参与者身上实施的问卷。然而，全面验证是一个广泛的过程：Price和Harding¹报告了一份调查问卷的开发和验证，该问卷用于衡量慢性伤口（下肢溃疡和糖尿病足溃疡）对患者健康相关生活质量（HRQoL）的影响，并确定患者关注的方面。这涉及一个三阶段的过程：焦点小组访谈和一系列半结构化访谈，以生成问卷项目；问卷调查的试点过程，并通过因素分析对数据进行分析；以及在3个月的随访期内评估所得量表的信度、效度和再现性。

虽然自行设计的问卷的全面验证是一项重大的任务，可能不在需要在有限时间内设计、实施和分析数据的临床医生的资源范围内，但一些常见的验证步骤是可行的。通常，这将涉及临床医生专家小组对项目措辞的输入，措辞的清晰度可通过焦点小组访谈或其他方式进行评估。其目的是得出一系列项目，每个项目都有助于了解所关注结局的不同方面，并且当相互结合进行评估时，提供对总体结局的有意义的衡量。可能需要专家建议来确认项目确实有助于衡量预期构

念，而非其他构念。Barakat Johnson等人²开发并评价了一种量表的心理测量学特性，该量表用于评估临床医生对失禁相关皮炎的了解，并使用临床医生专家小组的输入作为三阶段过程的第一阶段进行项目设计；然后，通过对临床医生和利益相关者的调查，对量表的内容效度进行评价，并采用试点多站点横断面调查设计，以确定综合信度。

在开发过程中，还应解决内容和构念效度问题。应避免使用自身过于相似的项目。这些项目并不是分别体现所关注构念的一个独特方面，而是重复体现了同一个方面，因此这一方面被重复计算，并且受访者很可能会以同样的方式对这两个项目做出回答。然而，与之相反，彼此差异较大的项目可能根本无法测量相同的构念。另一个常见问题是由不同项目体现的构念方面的“重叠”。Barakat Johnson等人²在其工具开发的后续阶段，使用公认的汇总指标和统计方法评价内容和构念效度。

项目制定和评分

通过问卷获得定量数据需要“封闭式”回答（数字或类别）；“开放式”回答通常不适用于定量报告。封闭式问卷项目可通过多种方式制定。一些较常见的项目包括：

- 直接引出数字值的项目，例如“您的年龄是多少？”
- 间接产生数字值的项目，要求受访者在视觉模拟量表上进行回答，随后由研究人员进行处理。其中一个典型示例可能是给出一条给定长度（比如10 cm）的线，两端明确标记为表示极值；例如：“一点也不痛”和“能想象到的最强烈的痛”；并附上指示，如“请在这条线上标出您的伤口今天给您造成的疼痛程度”。
- 允许受访者从提供的可能选项列表选择一个选项的项目。
- 允许受访者从提供的可能选项列表中选择尽可能多的选项的项目。

前两种类型的项目得到数值型回答；后两种类型的项目得到分类型回答。这两种类型的回答都可以用于后续分析，问卷的格式应确保受访者可以根据情况针对特定项目提供数值型回答，或从选项列表中选择。

可直接或间接获得数值型回答的项目可能是后续分析程序中最直接的。然而，随后的数据预处理可以通过优化一个项目的措辞来简化，这样受访者就不会觉得有必要添加不必

要的词：例如“您在这个组织工作了多久？”可能会得到各种回答，如“不到1年”、“18个月”、“5年左右”等，大多数计算机软件会将其处理为文本，而不是数值型回答，所以需要进行大量编辑才能用于分析。简单的重新措辞，如“请说明您在该组织工作的年数（四舍五入到最近的一年）”，可能会节省大量的预处理时间。此外，提供简单的说明，将任何不适用的项目或受访者无法给出正确回答的项目留空，可以节省更多删除各种“不适用”、“不知道”、“不确定”等情况的时间。

在产生数值数据的项目中引入人工分类是常见的做法。例如，要求受访者回答其年龄的项目可提供年龄范围选项的选择：“18-30”、“31-40”、“41-50”等。通常不建议采用此类方法：首先，同一年龄段内不同年龄的受访者之间的差异信息丢失（18岁和30岁的受访者的回答可能存在很大差异）；其次，分组变量中的多个类别意味着分析中需要进行多重比较（18-30岁与31-40岁之间的结果比较，18-30岁与41-50岁之间的结果比较，等等），这可能会导致技术问题和解释问题。

然而，如果一个构念的确是在分类层面进行测量，那么对体现这一构念的项目，除了提供选项列表供受访者选择之外，别无他法。所提供的选项列表应详尽无遗。例如，如果要求受访者提供其在组织中的角色，但受访者发现其角色在所提供的选项中没有体现，受访者可能会对其参与研究能否得到对自身观点或情况的准确记录失去信心，并且可能不太倾向于准确完成问卷的其余部分。

当选项重叠时，也会出现类似的问题。如果“您每周的病例数量是多少？”的选项是，比如说，“10个或更少”、“10-20”、“20-30”等等，那么病例数量为10名或20名患者的受访者将不知道他们应该选择哪个选项。另一个例子是受访者实际上有两个或多个角色时，被要求从选项列表中选择一项他们的工作角色。这种情况可以通过更好的问题措辞来避免，例如：“请从以下列表中选择与您的主要工作角色最接近的一项”。

在制定此类项目时，可能很想允许受访者进行自由文本回答。这可以防止意外遗漏受访者的首选选项，或防止因多个选项与受访者希望做出的回答相似但不完全相同而引起的混淆。然而，这种允许可能需要对自由文本数据进行大量的后续预处理，并将其划分为不同定义的组，如果受访者的自由文本回答不够明确，这可能会不太容易。这种情况通常可以通过在选项列表中提供“其他”选项来避免。

分类项目提供的选项可以是无序分类（无基本顺序；在这种情况下，选项的顺序不重要）或有序分类（在这种情况下，选项应按逻辑顺序呈现）。李克特项目是“经典”的有序问卷项目，它是最简单的，而且在某种程度上是最流行的问卷项目形式，在许多（如果不是大多数）问卷中都可以看到。李克特项目是一个问题，通常要求受访者从有序排列的五个选项中选择一个选项，代表对某一特定陈述（例如“产品X是肉芽组织过度增生的有效治疗方法”）的同意程度。此类项目的典型选项可能是“强烈不同意”、“不同意”、“二者都不是”、“同意”和“强烈同意”。其他Likert项目可能会要求受访者评估事件的频率或程度，例如，“伤口周围是否肿胀？”在这里，典型的选项可能是“完全没有”、“有一点”、“适中”、“相当多”、“很多”。

李克特项目不一定必须提供五个选项，但通常提供奇数个选项，其中五个可能是最常见的数量，以允许“中立”的中间选项。虽然选项数量较多的项目看起来提供了更细化的回答，但对于受访者来说，量表上各点之间的区别可能更加难以辨别（“部分时间”、“很多时间”、“大部分时间”、“几乎所有时间”等）。李克特项目的视觉等价项是一个类似如下措辞的问题：“在0到10的范围内，你的伤口在多大程度上妨碍了你完成日常家务？”。这是一个11点项目：一个常见的错误是让此类问题的量表范围处于1到10（而不是0到10）。在这种情况下，中立回答是5.5，而非5；尽管许多用5来回答这类问题的人，无疑是想用一个正好在可用量表中心的值进行回答。具有多种有序选项的项目在某些方面的表现类似于通过视觉模拟量表间接产生数值型回答的项目。

要求受访者选择“尽可能多的选项”的问题是可以接受的，但与只要求选择单个选项的相应问题相比，此类问题可能更难分析。例如，像“您经常使用以下哪种伤口敷料-请选择所有适用的选项”这样的问题，然后列出了26个选项（产品A、产品B、产品C……产品Z），从分析角度来看，实际上相当于一系列的26个问题：“您是否经常使用伤口敷料产品A？是或否？”；“您是否经常使用伤口敷料产品B？是或否？”……“您是否经常使用伤口敷料产品Z？是或否？”。这一系列的问题可能会产生各种各样的回答组合，并产生数十组成对比较，而所有这些都很难解释。

拟定用于分析的项目

典型的调查问卷可以从一些基本的人口统计学项目开

始，引出受访者的人口统计学和生活方式属性，如年龄、性别、家庭状况等；和/或与他们的健康状况（存在各种精神或身体健康状况、原有伤口的持续时间）或就业状况（服务年限、员工等级等）有关的项目。其中一些项目可能有助于说明样本的多样性或特征，但不会进一步在分析本身中发挥作用。

在合理的范围内，测量此类“背景变量”的项目，通常是引发数值型或分类型回答的事实性问题，而不是李克特式或类似的项目，可以用任何需要的方式记录。旨在描述数据，但不涉及任何类型的推论分析（即从样本数据推论到全及总体）的问卷，可能仅限于此类问题。此类研究通常旨在评估某一数量的普遍性或比例，如确定使用特定伤口护理产品的护士比例的研究，或对皮肤变红等视觉提示作出反应的临床医护人员比例的研究。Brown和Sneddon³实施了一份问卷调查，主要由提供有序选项的“独立”项目组成，以了解淋巴水肿服务在英国的资金来源和服务提供方式及其资源水平。问卷数据得出了比例（例如，接受调查的临床医生中治疗开放性伤口的医生比例）的估计值，但研究人员并未试图推论样本数据之外的情况。

然而，推论分析通常大多数属于定量研究的范围，因此，大多数得出定量数据的问卷将包括后续推论分析所需的项目。例如，关于某个或多个结局，可能需要对有经验的工作人员和新手医护人员，或定期翻身的ICU患者和不翻身的ICU患者，或新设备和标准设备进行比较。这些分析是比较研究的例子，在这些研究中，两个或多个组相互比较：许多标准研究设计，如队列研究、病例对照研究和随机对照设计，都属于这一范畴。Ousey等人⁴使用基于问卷调查的数据，在一系列患者体验指标（舒适度、温度和睡眠质量）方面，将新型设计的床垫与标准床垫进行了比较。研究人员使用标准的推论统计方法，根据床垫类型定义组别，并比较效果的显著性和程度。

这些研究中用于定义分组变量的项目是分类项目。正如在Ousey等人⁴的研究中所体现的，只有两个类别（或有时被称为“层次”）的分类变量被称为二元变量。一些分组变量可能包含两个以上的类别。例如，一项比较体重不足、正常体重、超重、肥胖或病态肥胖患者的结局的研究，可能会使用分组变量“肥胖状况”将每个问卷调查对象分为上述五个类别之一。

应谨慎指定此类多类别分组变量；虽然二元分组变量可以进行单一分析（例如，男性的结果对比女性的结果），但随着多层次分组变量的引入，所需的分析数量迅速增加。限制多层次分组变量的另一个原因是，虽然记录分组变量的项目通常应允许受访者选择任何可能的项目，但研究人员应做好准备，以防数据在多个类别中分散得太稀疏，导致一些组别数量过小，无法进行有意义的分析。在这种情况下，可能需要在分析之前将某些类别合并在一起。

结局指标

在大多数问卷中，大部分项目都与结局指标的引出有关。许多结局是分类的，且通常是二元的，例如，伤口在治疗后30天内达到50%愈合的可能性；或多类别的，例如，伤口床中的主要组织类型。这些结局通常可以很容易地在带有单个二元或有序项目的问卷中得到。Dhoonmoon⁵通过一份反馈问卷调查了56名专业医护人员（HCP）使用清创垫的经验。大多数项目，包括与清创垫性能相关的项目（清除腐肉碎片、清创等），都使用分类项目进行评估，选项从“优秀”到“差”。这些指标自然有助于进行有序的分类评估。为了便于分析或其他目的，许多有序结局被二分化——例如，Ousey等人⁴研究（睡眠质量）中的一项测量结果从最初的五个选项（“优秀”、“非常好”、“好”、“一般”、“差”）处理为二分测量，将“优秀”或“非常好”的回答与其他回答进行比较。数值型结局（例如患者痊愈的百分比，或疼痛水平达到某个预先规定数值的时间）也有，但在伤口护理基于问卷的分析中不太常见。

项目评分

问卷通常用于评价不存在简单客观指标的数量。在伤口护理研究中，这些数量可能是，例如，临床医生对新的压力重新分配床垫的评价，或患者对其伤口在多大程度上妨碍他们完成日常工作的看法。单个项目通常不能囊括此类数量；可能需要一系列项目，所有这些项目都涉及到所关注的概念。例如，最近完成了皮炎主题研讨会的实习护士的皮炎知识，或慢性伤口患者的生活质量。通常，这些组成项目可能是李克特式或类似的。在这种情况下，人们的兴趣几乎无一例外地集中在一组项目的处理得分上，而不是任何单个项目本身。因此，虽然从理论上讲，一个问卷项目中的每个项目都可以代表一个单一的指标，但一个典型问卷上体现的不同指标的数量通常远远少于问卷中的项目数量，一些项目用于对每个概念进行评估。

对结局的数量进行限制通常是可取的：以饼状图等形式广泛展示单个结局的结果，可能无法深入了解各种结局的相对重要性。还有一些分析问题可能会导致大量主要结局不理想。就像通过其他方式收集数据的研究一样，理想的调查问卷可能会体现关于单个预先指定的主要结局和少量次要结局的信息。

需要对有助于评估特定指标的所有项目进行评分。通常，李克特5点项目的评分非常简单，从“*强烈不同意*”的1分到“*强烈同意*”的5分，中间选项的对应相应的评分。具有其他选项数量的李克特项目以类似的方式进行评分。许多研究人员更喜欢使用以下编码：“*强烈不同意*”-2分，“*不同意*”-1分，依此类推，最多到“*强烈同意*”+2分，这可能是考虑到消极措辞的回答需要得负分。此编码与上述1-5编码完全相同——因为在整体得分中，每个选项对应的分数都会降低3分。只要使用的评分方法始终一致，在两种评分系统下的推断都是一样的。

通常假设项目的分数是可相加的，即通过把对同一指标有贡献的各个项目的分数相加来得出一个总分是有意义的。如果项目的表述具有一致性，那么这种假设通常更容易得到证明。如果一系列项目的选项数量从2到3到5到7不等，那么如何得出总分就不明显了。如果将每个项目的回答都简单地编码为1，直到选项数的值为止，那么选项数最多的项目的得分将淹没选项数较少的项目的得分。

如果选项不同，即使每个项目中的选项数量相同，也很难证明将多个项目的分数相加就能得到一个有意义的衡量。如果一组项目提供了“*强烈不同意*”、“*不同意*”……“*强烈同意*”的选项，而另一组项目提供了“*完全没有*”、“*有一点*”……“*很多*”选项，那么很难证明这两组项目的得分可以有意义地结合起来。

为了确保一个有意义的总分，如果某些项目与其他项目的意义相反，则可能需要颠倒上述编码，例如，如果使用上述1-5分量表对“*我的伤口迫使我限制与他人的活动*”和“*伤口影响我的睡眠*”等李克特5点项目进行编码，“*强烈不同意*”的回答得1分，“*强烈同意*”的回答得5分，这意味着分数越高，结局越差。因此，如果要包括同一量表中的其他项目，例如“*我能够毫无困难地完成日常工作*”，则可以对该项目进行编码，使“*强烈同意*”获得5分，“*强烈不同意*”获得1分，并相应地对该量表的其他选项进行评分，以与其余量表项目保持一致。

试点问卷调查

实施试点调查可成为完善问卷项目的有用工具，并可揭示可能影响后续回复率和回复信度的问题，如项目措辞不清晰或完成问卷所需时间过长。如果调查问卷包含一组李克特式或类似项目，这些项目旨在探查相同的构念，那么可以使用大多数统计软件轻松快速地评估试点调查中对这些项目的回答的内部一致性。该过程可以识别声称与其他项目测量相同构念，但未以类似方式回答的项目，因此可能需要修改其措辞（如果措辞不清楚或被受访者误解），从问卷中删除，或可能转移至另一个构念的测量。如果需要，试点调查阶段通常是进行此类修改的唯一机会。

总结

好的问卷设计是由研究问题以及由此产生的分析所驱动的。事实上，对终点的考虑通常就是起点。要考虑的问题包括确定要测量的结局；如何测量它们；结局是否是客观的指标，可以通过得出简单数值型回答或分类型回答的项目来充分体现，或者需要多个项目来体现指标的一系列特定方面。

还必须确定进行分析的层次——在伤口护理研究中，对患者、临床医生或伤口水平的分析都很常见。还必须确定结局是否与任何其他变量相关联，以及所需的比较组别是否作为分组变量出现在项目中，以对分析单位（患者、临床医生或伤口）进行适当分类。

通过调查问卷收集数据的方式应与通过医疗器械或其他方式收集数据的方式相同——有必要确保数据收集量表符合目的。这意味着在验证过程中要采取尽可能多的步骤（假设未使用预先已验证的量表），以确保我们通过仔细措辞的项目分组和适当评分来测量我们认为正在测量的结局。应注意，仅使用必要的项目来体现人口统计、其他背景信息和结局指标。有必要确保受访者尽可能是总体中有代表性的样本，以便对其进行推论。就像使用其他数据收集方式所做的那样，通过使项目尽可能清晰，并尽可能降低受访者完成问卷所需的时间和精力，可以最大限度地提高回复率。

虽然促进有效的问卷数据收集所需的努力很容易被低估，但如果进行得当，基于问卷的数据收集可以是一种非常有效的数据收集手段，并为研究奠定良好的基础。

利益冲突声明

作者声明无利益冲突。

资助

作者未因该项研究收到任何资助。

参考文件

1. Price P, Harding K. Cardiff Wound Impact Schedule: the development of a condition-specific questionnaire to assess health-related quality of life in patients with chronic wounds of the lower limb. *Int Wound J*. 2004 Apr;1(1):10-17.
2. Barakat-Johnson M, Beeckman D, Campbell J, Dunk AM, Lai M, Stephenson J, Coyer F. Development and Psychometric Testing of a Knowledge Instrument on Incontinence-Associated Dermatitis for Clinicians: The Know-IAD. *J Wound Ostomy Continence Nurs*. 2022 Jan-Feb 01;49(1):70-77.
3. Brown L, Sneddon MC. Lymphoedema service provision across the UK: a national survey. *J Lymphoedema*. 2020;15(1):16-21.
4. Ousey K, Stephenson J, Fleming L. Evaluating the Trezzo range of static foam surfaces: results of a comparative study. *Wounds UK* 2016;12(4):66-73.
5. Dhoonmoon L. Experiences of healthcare professionals using Prontosan® debridement pad. *Wounds UK* 2021;17(1):118-123.