AIDS and Behavior (2019) 23:2185–2189
https://doi.org/10.1007/s10461-018-02387-3

ORIGINAL PAPER

Debrief Reports to Expedite the Impact of Qualitative Research: Do They Accurately Capture Data from In-depth Interviews?

Jane M. Simoni[1,2] · Kristin Beima-Sofie[2] · K. Rivet Amico[3] · Sybil G. Hosek[4] · Mallory O. Johnson[5] · Barbara S. Mensch[6]

Published online: 21 January 2019

Abstract

"Debrief reports" (DRs) use structured forms to capture key concepts from in-depth interviews and focus group discussions. They are completed by interviewers and rapidly disseminated to key team members to facilitate identification of potential problems with study procedures, recruitment, or participant engagement and to inform critical adjustments, which can be especially pertinent in intervention studies. Their reliability and validity have yet to be formally evaluated. To assess the accuracy of DRs in capturing key content, raters analyzed a random sub-sample of 20 pairs of de-identified transcripts and their linked DRs from the VOICE-D trial. Analyses generally supported the accuracy of DRs; however, pertinent information from transcripts was occasionally missed or recorded with discrepancies or lack of detail. Longer transcripts and DR sections describing complex topic areas were more likely to involve discrepancies. Recommendations are offered for further research and optimizing the use of DRs.

Keywords Debrief reports · HIV/AIDS · Medication adherence · Africa · Qualitative data

Resumen

Los *debrief reports* (DR) son recursos emergentes, pero no examinados, que utilizan cuestionarios estructurados para capturar temas importantes, derivada de entrevistas en profundidad y discusiones de grupos focales. Se completan por los entrevistadores y son diseminados rápidamente a los miembros de investigación, para facilitar la identificación de problemas potenciales en cuanto al procedimiento de estudios, el proceso de reclutamiento, o el nivel de participación de los participantes, para que los investigadores puedan informarles de cambios critícales. Esto es útil para desarrollar estudios de intervenciones. Su fiabilidad y validez aún no han sido evaluadas formalmente. Para evaluar la precisión de los DRs en la captura de contenido clave, los evaluadores analizaron una sub-muestra de 20 pares de transcripciones sin identificadas y sus DR vinculados del ensayo VOICE-D. Los análisis generalmente apoyaron la precisión de los DRs, sin embargo, a veces la información pertinente de las transcripciones fue omitida, o escrita con errores, o escrita sin detalles. Las transcripciones más largas, y secciones de los DRs que describen áreas temáticas complejas, eran más probables de involucrar discrepancias. Se ofrecen recomendaciones para más investigación y para optimizar la implementación de los DRs.

Part of this project was presented at the Adherence 2017 conference in Miami in June, 2017.

✉ Jane M. Simoni
jsimoni@uw.edu

[1] Department of Psychology, University of Washington, 3909 Stevens Way NE Campus Box 351525, Seattle, WA 98195-1525, USA

[2] Department of Global Health, University of Washington, Seattle, WA, USA

[3] Department of Health Behavior Health Education, School of Public Health, University of Michigan, Ann Arbor, MI, USA

[4] Stroger Hospital of Cook County, Chicago, IL, USA

[5] University of California, San Francisco, San Francisco, CA, USA

[6] Population Council, New York, NY, USA

Introduction

Qualitative methodologies have yielded rich information in all stages of HIV research and clinical trials, including on the topic of medication adherence [1]. Qualitative data can sometimes further illuminate quantitative findings [2] or help to explain unexpected outcomes [3]. However, the time required to code and iteratively review lengthy transcripts has limited their impact during clinical trial implementation. Full analysis of qualitative data involves intensive procedures that include categorizing concepts and content and identification of themes that emerge from the data, often completed by multiple coders [4]. Transcription, possible translation, and iterative and exhaustive review of interviews or group discourse require substantial time and resources. Typically, this renders qualitative data nested within larger trials difficult to use in real time, limiting their impact. Numerous examples in the literature suggest that the insight gained from qualitative work completed after a trial could have been enormously valuable to the conduct of the trial had it been available in a more timely fashion. For example, qualitative exploration of experiences of women in PrEP trials has revealed critical information about concerns over drug safety, social pressures to avoid use of drug, and fears over dismissal from the study if non-adherence were reported [5–8].

To harness qualitative data in an expedited fashion, so-called "debrief reports" (DRs) involve interviewers, note takers, or group facilitators summarizing overall impressions, main themes, or content specific to a relevant topic immediately following an interview or discussion. Some data extraction methods use "process forms." The strategy of employing DRs is grounded in the concepts of peer debriefing, which has long been recommended in qualitative research [9, 10]. Peer debriefing involves discussing experiences and findings with other interviewers or research teams throughout the process of gathering interview data largely to enhance interviewer self-reflection and build confidence in the data being collected. Summarizing main findings in DRs is akin to the commonly used qualitative method of taking field notes, a formal way to record important findings [11]. DRs are distinct from peer debriefing and field notes, however, in that they use a structured question-and-answer format, seek to highlight specific issues or pieces of information, and are circulated among study investigators and partners for immediate review. As such, DRs are presumed to assist in the identification of potential problems with study procedures, recruitment, or participant engagement and guide trial implementation with ongoing course corrections that might better utilize study resources. Their use is increasing in HIV prevention studies and network-sponsored trials.

The utility and value of DRs depends on how closely they capture the key content of full interviews or discussions. However, we could locate no studies evaluating the reliability of validity of any sort of structured data extraction techniques. The extent to which DRs will yield themes and content similar to those emerging when the data are comprehensively analyzed remains unexamined. Accuracy might be eroded by interviewers' bias in inclusion or exclusion of highlights, their inability to retain and convey important information, and other factors that ultimately distance the data in the DR from the actual conversations recorded verbatim in complete transcripts. Determining the accuracy of DRs in reliably summarizing qualitative data is critical before their more widespread adoption in HIV clinical trials. To address this gap in the literature, we evaluated qualitative data collected from the VOICE-D study, in which each interview was summarized in a DR.

Methods

Data Source

VOICE-D was a two-stage qualitative study conducted in Durban, South Africa (2 sites); Kampala, Uganda; and Harare, Zimbabwe after VOICE—a phase 2B, placebo-controlled, randomized trial testing daily use of an antiretroviral tablet (tenofovir or Truvada) or daily use of tenofovir gel in 5029 women [5, 6, 12]. In the first stage, whence the data for this paper arise, 88 participants were randomly selected after stratification by study arm, reporting of anal sex, and HIV status. Five interviewers across the four sites conducted in-depth interviews (IDI) with the selected participants according to a semi-structured interview guide. Topics included two potential sources of vaginal gel efficacy dilution in the trial (i.e., heterosexual anal sex and non-adherence to the product) as well as original motivations for joining the trial. The interviewers were assisted by a note-taker and sometimes by a secondary interviewer, an observer, or both.

The DR form contained separate sections eliciting (1) data on subjective impressions of the participant as well as information about the interview context and experience; (2) the most important themes/ideas discussed in relation to (a) motivation to join trial (including risk perception and life events), (b) adherence (including discussion around measures and adherence reporting), and anal sex (including lubricant/gel use); and (3) any other important issues. The completed DR underwent site-level quality control review and, within a week, was sent to study investigators. Quality control focused on readability, coherence and completeness, with re-formatting or queries for additional clarifying information if warranted. There was no cross check with the recording or transcript. In fact, in most cases, the recorded

IDI was not transcribed or translated into English until well after the DR was completed and shared with study teams.

All VOICE-D study interviewers had backgrounds in qualitative research, although they had varying degrees of experience. A 2-day training workshop included content on qualitative research methods, how to build rapport and probe, as well as a detailed review of the study guides. There was discussion of the importance of and strategies for taking notes. IDI role playing was also incorporated. Other areas focused on transcription and translation as well as coding and analysis. For the DR training, the purpose and content of the form were discussed. The interviewers were instructed to elaborate on topic areas included in the DR, finish their notes immediately following the interview, complete the DR the same day, and have the co-interviewer review and add to the report before circulating for quality control. The DRs were written the same day of the interview and sent to trial leadership within a week so that any needed adjustments could be implemented in a timely manner.

For the current analysis, we selected a random sample, $N = 20$, of the 88 DRs and their linked transcripts. We chose four DR-transcript pairs from each of the five interviewers. The data were stripped of participant-identifying information. The mean interview length was 78 min (range = 54–165 min).

Data Analysis

Prior to the main analysis, each of the six authors ("raters"), all of whom are social and behavioral scientists with content expertise, analyzed the same three DR-transcript pairs (results not shown). After group discussion to develop a review process, all raters were involved in the subsequent analysis of 20 DR-transcript pairs for concordance/discordance (two raters for each pair). The analysis focused on two of the three topics that were listed in the DB form: motivations for joining the VOICE trial and study product adherence. The raters, who did not have access to the full IDI guide prior to analysis, created a DR-transcript Concordance Form specifically for this project. Raters identified in the transcript and recorded on the form content on reasons for joining the trial and adherence and cross-checked that information with data in the DR. This process varied by rater. Some read the DR first and then the transcript, whereas others read the transcript first, highlighting relevant content, and then compared this to the DR. Content that was in the transcript and of relevance was coded as (1) accurately represented in the DR, (2) noted in the DR but the emphasis, prominence, or thematic qualities differed from what was in the transcript, or (3) missing completely from the DR. Also, content could be (4) in the DR but not found anywhere in the transcript. Raters then evaluated the overall quality of the DRs according to the amount of discordance and missing or added information in the DR, coding each DR-transcript pair as *superior* (none or almost no discrepant, missing, or added information); *acceptable* (minor discrepancies between the DR and transcript); or *unacceptable* (considerable differences between the two).

Results

Of 20 DR-transcript pairs analyzed, 13 (65%) originally received the same concordance quality score from both raters. Final scores for the 7 with discrepant concordance score ratings were determined by consensus after a group discussion. In six cases, the scores of 1 and 2 were determined to be a 2; in the seventh case the scores of 3 and 2/3 were determined to be a 3. According to the final ratings, almost all DR-transcript pairs were considered *superior* ($n = 8$) or *acceptable* ($n = 11$); only 1 was deemed *unacceptable*. The interviews for the *superior* pairs were comparable in duration to the *acceptable* ones (respective means = 70 and 76 min); however, the one DB rated as *unacceptable* was associated with the longest interview (165 min).

In *superior* pairs, the DRs consistently and accurately captured main themes from the transcripts. Raters noted very little transcript material that was missing from the DR. If something was missing or discrepant, it tended to be a minor detail (e.g., one transcript quoted a participant as saying she told her family about her trial participation but could not explain a lot to younger siblings while the DR notes her having told only her mother) or not related to the main categories under investigation.

In the 11 *acceptable* DRs, the main themes from the full transcripts also were generally recorded in the DR, but the DR summaries might have omitted some information the raters considered worthy of mention. For example, in one transcript a participant discussed HIV stigma in the community as a barrier to study product adherence, yet stigma was not reported as one of the barriers in the DR. In another transcript, the participant discussed her husband's infidelity as a motivator for joining the study, but the DR did not mention partner infidelity when reporting motivations to join the study. One transcript described both "forgetting" and "vaginal leakage" as reasons for non-adherence to the study product, but the DR only reported "forgetting". In some pairs rated *acceptable*, there was content in the DB not in the transcript, as in the DR that stated that the participant's partner often convinced her not to use a condom though this language could not be found in the transcript. The examples of discordance for *acceptable* pairs fell more often in the category of content that was noted in the DR but its emphasis, prominence, or thematic qualities differed from that of the transcript. For example, one DR noted that a participant had no problems with the study staff or clinic except for the long

times spent at the clinic. The rater noted that the DR did not mention a section quote in the paired transcript describing how this participant added she was bored by the staff as well.

The one DR-transcript pair rated *unacceptable* failed to include in the DR several themes from the transcript that the raters considered important. For example, in the transcript, the participant discussed missing/late doses of study product and reported crowded living conditions and fears of vaginal dryness as barriers to adherence. However, none of those concerns were written into the DR. Moreover, the DR included some content not present at all in the transcript. For example, the DR noted "needed the money" as motivation to join the trial but, in the transcript, the participant discussed needing money as justification for commercial sex and having anal sex but never as a motivation to join trial. As noted, the transcript in this pair was exceptionally long (165 min transcribed onto 52 pages), perhaps complicating the task of summarizing it in a DR.

Although our relatively small sample of interviewers and DR-transcript pairs limited our ability to compare the performance of individual interviewers, raters observed some tended to record detailed and extensive content in the DR, while others recorded sparse content.

Raters noted that brief answers to specific questions from the interview guide tended to be faithfully captured in the DR. For example, the opening section on the participant's motivations to join the trial often yielded a list of specific reasons that were accurately reflected in the DR. On the other hand, longer and more detailed discussions were harder to capture in the DR format. The discussions on how participants interpreted the adherence assessment items in the trial surveys, for instance, were sometimes recorded in the DRs without the specificity and nuance of the original transcript. This was exacerbated perhaps by the DR form, which offered one open-ended section for recording content on this topic even though the prompts on this section in the interview guide spanned multiple domains. This open-ended section in the DR likely made it more difficult for the person completing the DR to discern which aspects of the participant's response should be recorded.

Discussion

This post-trial, secondary analysis of qualitative data from VOICE-D indicated that DRs were generally accurate in capturing thematic content from full transcripts of IDI. However, there are some limitations to our analysis. Specifically, our findings are based on a relatively small sample of DR-transcript pairs from one qualitative study, in which the text had been translated from the original language. Also, while we have no reason to believe that the interviewers in VOICE-D are more or less qualified than other qualitative interviewers, we have no way of knowing if their ability to accurately fill out a DR is typical of other interviewers. In addition, we only assigned two reviewers to each DR-transcript pair. Multiple reviewers for each DR-transcript pair might have detected more disagreement. Finally, we set no benchmarks for our overall rating of concordance as *superior*, *acceptable* or *unacceptable* prior to conducting the analysis, although we did have clear coding criteria for concordance. Thus, there is an element of subjectivity in our overall evaluation.

Despite these limitations, we believe our study offers some useful information about DRs for HIV researchers, especially those conducting trials in which qualitative data can be collected concurrently. Findings suggest that although DBs cannot provide a complete and in-depth account that perfectly mirrors a transcript, they can provide a sampling of responses in key areas of inquiry and signal possible concerns that can be validated with further investigation. Investigators can determine the utility of this method based on the particular needs or demands of their research. Below we offer guidance for the use of DRs aimed to enhance their accuracy and utility:

- Identify a clear and specific purpose for use of the DR. Ideally, the process should focus on collecting critical information that can be rapidly obtained and acted upon.
- A DR is best for capturing specific data from semi-structured interviews on circumscribed topics (e.g., reasons for joining or dropping out of a trial, comprehension of risks and study requirements, participant interactions with study staff members, extent of adherence to some aspect of the experimental protocol, and community beliefs around the study and/or study product). DRs may less effectively capture nuanced responses to complex, multi-level, or broad open-ended questions. This type of qualitative inquiry, as well as research utilizing phenomenological methods, is better suited to a comprehensive and in-depth qualitative analysis.
- The DR should mirror the interview guide in content and sequence. It would optimally be limited to the most critical questions asked as the inclusion of too many sections may increase burden and reduce reliability. Minimally, it should include clearly labelled sections and sub-sections on targeted topics in the order they are expected to appear in the interview. As noted, the DR used in VOICE-D combined several subtopics into overarching section headings, which likely diminished the detail recorded.
- In the extreme, a written version of the interview guide printed with ample spacing could form the basis of the DR. Notes from the interview (perhaps written during the interview by the primary interviewer or an assistant) could be recorded directly onto this form. This form could be further annotated after the interview and con-

stitute the formal DR or used as a basis for completing a more detailed DR.
- Deliver detailed training on how to complete a DR, with close supervision and corrective feedback early in the process. Retraining over time may be warranted.
- Consider having a trained note-taker present during interview. Although a burden on resources, a note-taker could enhance recall for completion of the DR.
- Complete the DR during and/or immediately after the interview to facilitate accuracy.
- In the case of more open-ended interviews, researchers might consider reviewing the first few DR-transcript pairs and revising the DR guide in an iterative fashion to ensure it is capturing the types of information arising from the interviews with the desired level of detail.
- Conduct ongoing quality assessment of the DR process by periodically comparing DR-full transcript pairs (perhaps according to the protocol described for this study).

Future research should explore which training strategies best facilitate comprehensive and accurate debrief reporting and which formats for DRs promote the collection of concise and accurate data from interviewers. A replication of the current research with a larger sample and pre-specified coding criteria also might serve to build more support for the utility of the approach.

In conclusion, the data generated from VOICE-D provided a unique opportunity to evaluate the accuracy of DRs in reference to a complete transcript. Our findings suggest that some confidence can be placed in these kinds of reports and the rapid information sharing they enabled may provide valuable and cost-saving data in trial implementation. HIV researchers, especially if they adhere to the recommended guidelines, may find many ways DRs can expedite their qualitative research.

Acknowledgements We thank Miriam Hartmann, Ariana Katz, Elizabeth Montgomery, and Ariane van der Straten for patiently answering our questions and the Microbicides Trial Network for access to the data.

Funding This work was supported in part by the Office of HIV/AIDS Network Coordination (HANC) which is funded in whole or in part with Federal funds from the Division of AIDS, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, grant number UM1 AI068614, entitled Leadership Group for a Global HIV Vaccine Clinical Trials (Office of HIV/AIDS Network Coordination) with additional support from the National Institute of Mental Health. The MTN 003-D study was designed and implemented by the Microbicide Trials Network (MTN) funded by the National Institute of Allergy and Infectious Diseases through individual grants (UM1AI068633, UM1AI068615 and UM1AI106707), with co-funding from the Eunice Kennedy Shriver National Institute of Child Health and Human Development and the National Institute of Mental Health, all components of the U.S. National Institutes of Health (NIH).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

1. Sankar A, Golin C, Simoni JM, Luborsky M, Pearson C. How qualitative methods contribute to understanding combination antiretroviral therapy adherence. J Acquir Immune Defic Syndr. 2006;43:S54–68.
2. Chen WT, Shiu CS, Yang JP, et al. Fatigue and sleep disturbance related to perceived stress in Chinese HIV-positive individuals: a mixed methods study. J AIDS Clin Res. 2013. https://doi.org/10.4172/2155-6113.1000214.
3. Simoni JM, Huh D, Frick PA, et al. Peer support and pager messaging to promote antiretroviral modifying therapy in Seattle: a randomized controlled trial. J Acquir Immune Defic Syndr. 2009;52(4):465–73.
4. Creswell JW, Poth CN. Qualitative inquiry and research design: choosing among five approaches. Thousand Oaks: SAGE Publications; 2017.
5. Montgomery ET, Mensch B, Musara P, et al. Misreporting of product adherence in the MTN-003/VOICE trial for HIV prevention in Africa: participants' explanations for dishonesty. AIDS Behav. 2017;21(2):481–91.
6. van Der Straten A, Montgomery ET, Musara P, et al. Disclosure of pharmacokinetic drug results to understand nonadherence. AIDS. 2015;29(16):2161–71.
7. Corneli AL, McKenna K, Perry B, et al. The science of being a study participant: FEM-PrEP participants' explanations for overreporting adherence to the study pills and for the whereabouts of unused pills. J Acquir Immune Defic Syndr. 2015;68(5):578–84.
8. Amico KR, Wallace M, Bekker LG, et al. Experiences with HPTN 067/ADAPT study-provided open-label PrEP among women in Cape Town: facilitators and barriers within a mutuality framework. AIDS Behav. 2017;21(5):1361–75.
9. Creswell JW. Qualitative inquiry and research design: choosing among five traditions. Thousand Oaks: Sage Publications; 1998.
10. Creswell JW. Research design: qualitative, quantitative, and mixed methods approaches. 3rd ed. Thousands Oaks: Sage Publications; 2009.
11. Emerson RM, Fretz RI, Shaw LL. Writing ethnographic fieldnotes. 2nd ed. Chicago: University of Chicago Press; 2011.
12. Musara P, Montgomery ET, Mgodi NM, et al. How presentation of drug detection results changed reports of product adherence in South Africa, Uganda and Zimbabwe. AIDS Behav. 2018;22(3):877–86.